Evolution and Development of the Olfactory Neurons in the Olfactores Clade

Thesis presented for the degree of Doctor of Philosophy in Zoology

Guillaume Poncelet
Linacre College, University of Oxford
Michaelmas 2021

Evolution and Development of the Olfactory Neurons in the Olfactores Clade.

Thesis presented for the degree of Doctor of Philosophy in Zoology

Guillaume Poncelet
Linacre College, University of Oxford
Michaelmas 2021

ABSTRACT

In vertebrates, specialised olfactory systems and associated neurons are crucial to sense the outside chemical world. They also play a notable part in reproduction with the role of GnRH-producing neurons, derived from the olfactory placode, in sexual maturation and fertility. Despite the vital importance of these olfactory cell types, our knowledge of their evolutionary origin remains limited. In this work, a comparative approach is undertaken to study the development of olfactory neurons and provide insights into their evolutionary history. Two main organisms are used for comparison: the lamprey *Lampetra planeri* and the ascidian *Ciona intestinalis*. These model species are used to investigate the expression, regulation, and function of specific genes to trace olfactory cell type relationships across vertebrates and their chordate relatives. First, the embryonic origin and specification of hypothalamic GnRH secretory neurons in lamprey is examined to develop a pan-vertebrate view of the origin and specification of these crucial neurosecretory cells. The potential association during embryogenesis with the developing olfactory system is investigated and this study provides support for the hypothesis that lamprey hypothalamic GnRH cells do not derive from the nasal placode, despite some specification mechanisms appearing conserved with olfactory-derived GnRH neurons of jawed vertebrates. Second, it is revealed that the palp axial columnar cells (ACCs) of *Ciona* have homologies with olfactory sensory and secretory neurons of vertebrates. The regulation, expression, and function of *MS4A* and *GnRH* genes in the palp ACCs, combined with data from vertebrates, suggest a common evolutionary origin at cell-type level to explain how the olfactory-derived cell types in vertebrates derived from an ACC-like GnRH/chemosensory multifunctional precursor by segregation of functions. Finally, a microfluidic chip for immobilisation and controlled stimulation of *Ciona* larvae is designed and built to interrogate whether the larva responds to chemical stimuli. The detection by live cell calcium imaging of a CO_2-evoked response upon chemical stimulation represents the first experimental evidence that ascidian larva sense chemical cues in their environment and provides a valuable tool for further investigations.

ACKOWLEGDMENT

This thesis represents the outcome of many years of hard-work, dedication, and the pinnacle of my University education. I like to thank Dr Sebastian Shimeld for the opportunity and trust he gave to do my D.Phil. in his lab. This research has been marked with numerous personal and external hardships, due to several building relocations or the coronavirus pandemic to name a few. Therefore, this thesis completion, would have been impossible without moral support. For that, I want to thank my family, my beloved parents Dr Etienne Poncelet and Ms Genevieve Peeters as well as my two brothers Dr Pierre-Antoine Poncelet and Dr Jean-Benoit Poncelet for the continuous love and assistance given throughout the years. I mainly dedicate this thesis to you. There are also personalities that have marked my life at Oxford and who have made this whole adventure humanly rich and worth it. Special thanks to my labmate Dr Vasileios Papadogiannis with whom I have shared this adventure since the start. I will not forget your kind help and joviality. Thanks also to all the members of Holland's lab: Dr Amy Royall, Dr Yichen Dai, and Dr Sonia Trigueros for your friendship and kindness. Special thanks also to all the friends that I have made at Linacre College and to Dr Boyoung Lee and Olive, for the good moments we shared. I liked to also thank my College for the financial support, through the EPA Cephalosporin Scholarship, the Linacre Trust Writing fund and Hardship fund. Also, thanks to Dr Jay Willis, to have chosen me to work on his research project funded by the Leverhulme Trust which enabled me to ensure constant funding during my thesis as well as all the knowledge acquired on fish keeping. Finally, special thanks to Dina, with whom I had the happiness to share the last moment of my thesis. Thanks for your unconditional love, patience, and support during the long writing process. Overall, it is a time that I will remember for sure as it has pushed me to my limits, changed me and made me grow immensely. I hope this work can honour the devotion that has been given to make those years a personal success.

Contents

Chapter 1 General Introduction

Work Declaration

The literature review in this Chapter and associated Figures were performed by me. The majority of it has been published in the peer-reviewed journal Open Biology with co-authorship with my supervisor (see Appendix for the full publication).

1.1 Aim and layout of the thesis

The central aim of this research is the study of the evolutionary origins of olfactory neuron development in the Olfactores lineage. Comparative genetics and development, using lampreys and ascidians as models, is used to interrogate potential conserved expression, regulation, and function of selected gene markers with demonstrated cell type function and/or developmental roles in olfactory neurons. In particular I focus on tracing cell type evolutionary relationships to uncover ancestral expression traits connected to olfactory chemosensory/GnRH secretory neurons.

Chapter 2 aimed to develop a better understanding on the origin and specification of hypothalamic GnRH secretory neurons in lamprey. To achieve this, three main lines of research were undertaken. First the expression of chosen gene markers known to be important for the differentiation and migration of GnRH neurons in the olfactory placode of gnathostomes was studied in the lamprey, *L. planeri*. Second, the specification of lamprey GnRH neurons was examined by testing its susceptibility to manipulation of the FGF and RA signalling pathways. Third, cell-labelling was done to test whether migratory neurons could arise from the nasohypophyseal placode. Work described in this chapter

helps to generate a pan-vertebrate vision for the origin and specification of those crucial neurosecretory cells.

Chapter 3 explores putative homologies between vertebrate olfactory neurons and the palp cells of *Ciona intestinalis*. First, I investigated the function of the regulatory region of *CiGnRH1* in *Ciona* tadpoles by successive trimming constructs to identify important sequences and infer putative transcription factor binding sites and compare the gene regulatory network with that of olfactory-derived GnRH neurons of vertebrates. Second, I report the identification in the *Ciona* palp axial columnar cells of a member of the proposed chemoreceptors of the *MS4A* family. Evidence of positive selection in the extracellular domains of the *Ciona* MS4A protein supports its putative role as a chemoreceptor in ascidians, as seen with mammals.

Chapter 4 describes an attempt to develop new technology for the detection of chemosensory activity in the nervous system of *Ciona* swimming larvae. A system was designed which integrates the use of customised microfluidic chip and live calcium-imaging through the expression of a genetically-encoded calcium sensor by transgenesis. This provides the first direct evidence of sensory reaction to chemical stimulation in a living ascidian larva and indicates that they can sense CO_2.

Finally, Chapter 5 reviews how this study advances our understanding of the evolutionary origin of vertebrate olfactory neurons and proposes the next steps in this comparative exploration.

1.2 Review: The evolutionary origins of the vertebrate olfactory system

1.2.1 Abstract

Vertebrates develop an olfactory system that detects odorants and pheromones through their interaction with specialized cell surface receptors on olfactory sensory neurons. During development, the olfactory system forms from the olfactory placodes, specialized areas of the anterior ectoderm that share cellular and molecular properties with placodes involved in the development of other cranial senses. The early-diverging chordate lineages amphioxus, tunicates, lampreys and hagfishes give insight into how this system evolved. Here, we review olfactory system development and cell types in these lineages alongside chemosensory receptor gene evolution, integrating these data into a description of how the vertebrate olfactory system evolved. Some olfactory system cell types predate the vertebrates, as do some of the mechanisms specifying placodes, and it is likely these two were already connected in the common ancestor of vertebrates and tunicates. In stem vertebrates, this evolved into an organ system integrating additional tissues and morphogenetic processes defining distinct olfactory and adenohypophyseal components, followed by splitting of the ancestral placode to produce the characteristic paired olfactory organs of most modern vertebrates.

1.2.2 Introduction: olfaction and chemosensation

Olfaction is a form of chemosensation. It is colloquially equated to the sense of smell, the specific sensing of chemicals in the air via the nose and the relaying of this information to the brain via olfactory nerves. However, the precise evolutionary and developmental delineation of the olfactory system becomes blurry when one considers the details. Many vertebrates have a related chemosensory system in the vomeronasal organ, which shares

a developmental origin with the main olfactory system but has generally been thought to be devoted to chemical communication between conspecifics. In aquatic vertebrates, such as fish and amphibians, a homologous olfactory system to that of terrestrial vertebrates detects waterborne rather than airborne chemicals, while insects possess a well-described system in their antennae that senses airborne chemicals and is usually called an olfactory system, but is convergently evolved at the system level. Furthermore, the development of the vertebrate olfactory system includes the formation of cells associated with other functions, including that of the pituitary, and there are many vertebrate chemosensory cells that relay information to the brain but that are not part of olfactory systems in the conventional sense. Taste is an obvious example.

Sensing chemicals on the outer side of the cell membrane is a fundamental feature of all cells and sensing environmental chemicals has obvious adaptive advantages. It is therefore not surprising that a diversity of chemosensory mechanisms and systems have evolved in animals. We will not attempt to cover this diversity here, but will focus specifically on the evolution of the vertebrate olfactory system. We will combine two levels of comparison: first, the types of neural cells that develop in the olfactory system (both chemosensory and neurosecretory cell types). Second, the mechanisms that control the specification of olfactory cells and organs. Since vertebrate olfactory cells develop from an ectodermal placode that shares a developmental and evolutionary history with other such placodes, we will also consider placode development and evolution more broadly.

We first summarize what is known about this in jawed vertebrates (also known as gnathostomes). We then compare this to olfactory systems and related cells and

structures in other chordates (Fig. 1.1): the jawless vertebrates (represented by lampreys and hagfishes, collectively the cyclostomes), the tunicates (including sea squirts and their allies) and the cephalochordates (represented by amphioxus). We will finish with a model for how the olfactory system in living vertebrates evolved.

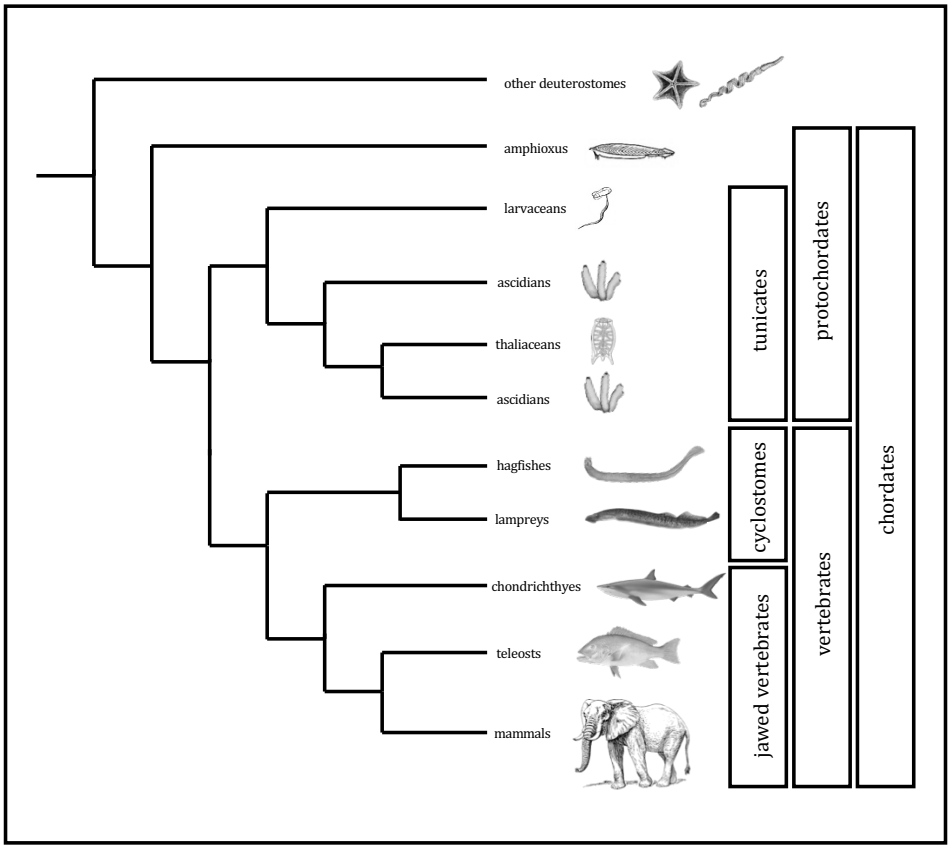


Figure 1.1 A phylogeny of the chordates showing the relationships of the major lineages discussed in this review. Note that the ascidians appear twice in the tree as they are paraphyletic.

1.2.3 The olfactory systems of jawed vertebrates and their neural cell-type derivatives

There are two major olfactory subsystems in jawed vertebrates: the main olfactory system (MOS) and the accessory olfactory system (AOS) (Fig. 1.2), and relevant neuronal cell types are summarized in Table 1.1 The MOS is historically said to be important for the detection of odorants and the AOS to mainly sense pheromones. However, the systems may overlap functionally and act synergistically (Suarez et al., 2012). When a chemical stimulus flows into the MOS or AOS, it is detected by olfactory sensory neurons (OSNs) through specific membrane chemoreceptors. Nearly all these chemoreceptors are coupled to a specific G protein subunit α, encoded by genes of the GNAL and GNAS families (Oka et al., 2011). The different Gα proteins mediate signal transduction pathways that open cyclic nucleotide-gated (CNG) ion channels in the MOS or transient receptor potential (TRP) channels in the AOS. These channels trigger a calcium influx in the olfactory neuron cytosol, promoting the opening of calcium-gated chloride channels. The combined effect of calcium and chloride efflux triggers OSN depolarization (Buck and Axel, 1991). OSNs expressing the same chemoreceptor gene send their axon projections via the olfactory nerve to a specific glomerulus in the olfactory bulb. The OSNs of the MOS transmit the chemosensory signal through the main olfactory bulb, which then connects to higher brain centres for the processing of a behavioural response. The OSNs of the AOS target their axons to the accessory olfactory bulb in the rostral telencephalon, which then projects towards the amygdala and hypothalamus, which are involved in aggression and mating behaviours (Dulac and Torello, 2003). The MOS includes the main olfactory epithelium, which is composed of typical ciliated OSNs. The cilia of the OSNs harbour seven-transmembrane domain G-protein-coupled receptors from the olfactory receptor

family (OR), or receptors of the trace amine-associated receptor family (TAAR), on their surface (Bear et al., 2016). The ORs are the largest gene family in vertebrates and there may be more than a thousand different genes in the genomes of some species (Niimura et al., 2014). The TAARs and ORs are coupled to G protein subunit Gαolf. In the recesses of the mammalian main olfactory epithelium, there is an expression of chemoreceptors from the guanylate cyclase D receptor family, and the MS4A gene family, in the cilia of a specific group of OSNs known as the necklace OSNs. These latter receptor families are not coupled to specific G proteins (Greer et al., 2016).

The AOS is also known as the vomeronasal system in tetrapods and includes a distinct sensory epithelium containing OSNs bearing microvilli instead of cilia. Like ciliated MOS OSNs, these microvillous OSNs express seven-transmembrane domains G-protein-coupled receptors, but in the AOS, these are chemoreceptors of the vomeronasal type 1 (V1R) or type 2 (V2R) families. V1Rs and V2Rs are associated with the G protein subunits Gαi and Gαo, respectively. In rodents, some microvillous sensory neurons of the vomeronasal organ also express members of the formyl peptide receptor family associated with the identification of pathogens and infections (Riviere et al., 2009). The description of an accessory olfactory system in lungfish suggests that all sarcopterygians primitively had such a dual system (Gonzalez et al., 2010); however, the vomeronasal system is not preserved in all tetrapod groups and is absent or vestigial in some lineages like archosaurs (birds and crocodilians) and higher primates (Suarez et al., 2012). While molecular studies have yet to extend across the diversity of tetrapods, there is a long history of histological and ultrastructural studies of tetrapod olfactory and vomeronasal systems covering many different species. There is insufficient space to detail these studies here and they have been well-reviewed by Eisthen (Eisthen, 1992). It is important

to note, however, that the tetrapod group may harbour system diversity beyond the simple division of ciliated OSNs in the MOS and microvillous OSNs in the AOS familiar from mammals. For example, both types of OSN are found in the MOS of some urodele amphibians and ciliated OSNs in some lizard and bird species may have their cilia surrounded by microvilli (Eisthen, 1992). In teleost fishes, there is no proper vomeronasal organ but a single olfactory epithelium showing morphological features of both the main and accessory systems and with intermingled ciliated and microvillous OSNs (Hussain, 2011). Teleost ciliated OSNs express the ORs and TAARs associated with Gαolf, similar to the main olfactory epithelium of tetrapods. Similarly, the teleost microvillous OSNs express V1Rs and V2Rs associated with Gαi and Gαo, as seen in the vomeronasal organ of mammals. It was also demonstrated in zebrafish that microvillous OSNs form a neural circuit via the dorsomedial olfactory bulb and intermediate ventral telencephalic nucleus (the putative teleost medial amygdala) to the tuberal hypothalamus, similar to the vomeronasal circuit of tetrapods (Biechl et al., 2017). Teleosts have a third class of OSNs, the crypt cells, which possess microvilli and cilia and express V1R-type (ORA) receptors (Oka and Korsching, 2011; Saraiva and Korsching, 2007). In zebrafish, there is also a fourth type of OSN, the cap (kappe) cell, whose receptor type is unidentified but known to associate with Gαo (Ahuja et al., 2014). In Chondrichthyes, it was observed that the sense of smell relies primarily on microvillous OSNs coupled to Gαo (Ferrando et al., 2010). The chemosensory receptor repertoire of cartilaginous fishes is dominated by the expanded V2R family, although there are also a few OR, TAAR and V1R genes (Sharma et al., 2019). As in teleosts, there is also the presence of crypt neuron-like cells but their exact receptor is unknown (Ferrando et al., 2006). This peculiar feature and the absence of ciliated OSNs suggest that the cartilaginous fish olfactory system is just an accessory system (Ferrando and Gallus,

2013). Some neurosecretory cells also delaminate from the olfactory placode of gnathostomes. The most well-known are the gonadotropin-releasing hormone (GnRH) neurons involved in the reproductive axis (Aguillon et al., 2018; Sabado et al., 2012). There are three distinct populations of GnRH neurons in the brain of gnathostomes expressing one of the three existing GnRH genes: GnRH1, GnRH2 and GnRH3. Some species have lost some of these paralogues, but most vertebrates express at least two. Neurons expressing either GnRH1 or GnRH2 have been identified in most gnathostomes, while GnRH3-expressing neurons are fish-specific (Gaillard et al., 2018; Duan and Allard, 2020). The prevailing view is that the GnRH1 neurons and GnRH3 neurons originate from the olfactory placodes, while the GnRH2 neurons are of non-placodal origin and develop within the central nervous system, mainly the midbrain (Umatani and Oka, 2019; Northcutt and Muske, 1994). The GnRH1 neurons are key regulators of fertility as an essential part of the hypothalamic–pituitary–gonadal axis (HPG). These cells migrate along axons of the terminal nerve/olfactory pathways up to the forebrain where they settle inside the pre-optic and hypothalamic areas (Cho et al., 2019). Once settled within the hypothalamus, GnRH1 neurons send their axons to the median eminence and secrete GnRH1 into the portal vessels, where it travels to the adenohypophysis (Wierman et al., 2011). Here, GnRH1 activates specific receptors of the gonadotrope cells, which then release two crucial hormones for sexual maturation and reproduction, luteinizing hormone and follicle-stimulating hormone (Wray, 2002). In chick, it was shown that the neuropeptide Y (NPY) neurons also derive from the olfactory placodes and migrate to the hypothalamus along with GnRH1 neurons. NPY neurons control the secretion of GnRH1 by acting directly on GnRH1 neurons (Hilal et al., 1996). GnRH3 neurons migrate from the olfactory placodes and become components of the terminal nerve whose processes extend anteriorly to the nasal cavity and posteriorly to various brain regions, mediating

chemosensory processing and reproduction (Wirsig-Wiechmann, 2001). The GnRH3 neurons of the terminal nerve (TN- GnRH3) have neuromodulatory effects on the OSNs (Eisthen et al., 2000; Kawai et al., 2009). In addition, the TN-GnRH3 neurons of zebrafish have been demonstrated to be chemosensory, detecting CO_2 in order to avoid incoming predators (Koide et al., 2018). This latter finding suggests that olfactory-derived neurons with dual GnRH/chemosensory abilities could exist in vertebrates (as is advocated for the aATENs of ascidians, see §1.2.10 below for details) (Abitua et al., 2015). In species that lack either GnRH1 or GnRH3, the remaining gene compensates functionally for the lost paralogue by being expressed by the relevant cells. For example in zebrafish, where GnRH1 is lost, the GnRH3 gene is expressed in the pre-optic area and hypothalamus (the GnRH1 territory) in addition of the terminal nerve (the GnRH3 territory). It needs to be made clear that the olfactory epithelia of the MOS and AOS also contain non-neural cells that surround the OSNs, such as supporting cells and basal cells. The supporting cells provide physical and metabolic support to the olfactory epithelium. The basal cells are stem cells used to constantly replenish the olfactory epithelium as they can differentiate into either OSNs or supporting cells. The MOS also contains the mucus-producing olfactory (Bowman's) glands whose proteinaceous secretion allows solubilization of odorants in the nasal cavity (Cho et al., 2019; Katoh et al., 2011). These cells are important for vertebrate olfactory system function but will not be the focus of this review as they are not easily compared between vertebrates and other chordates.

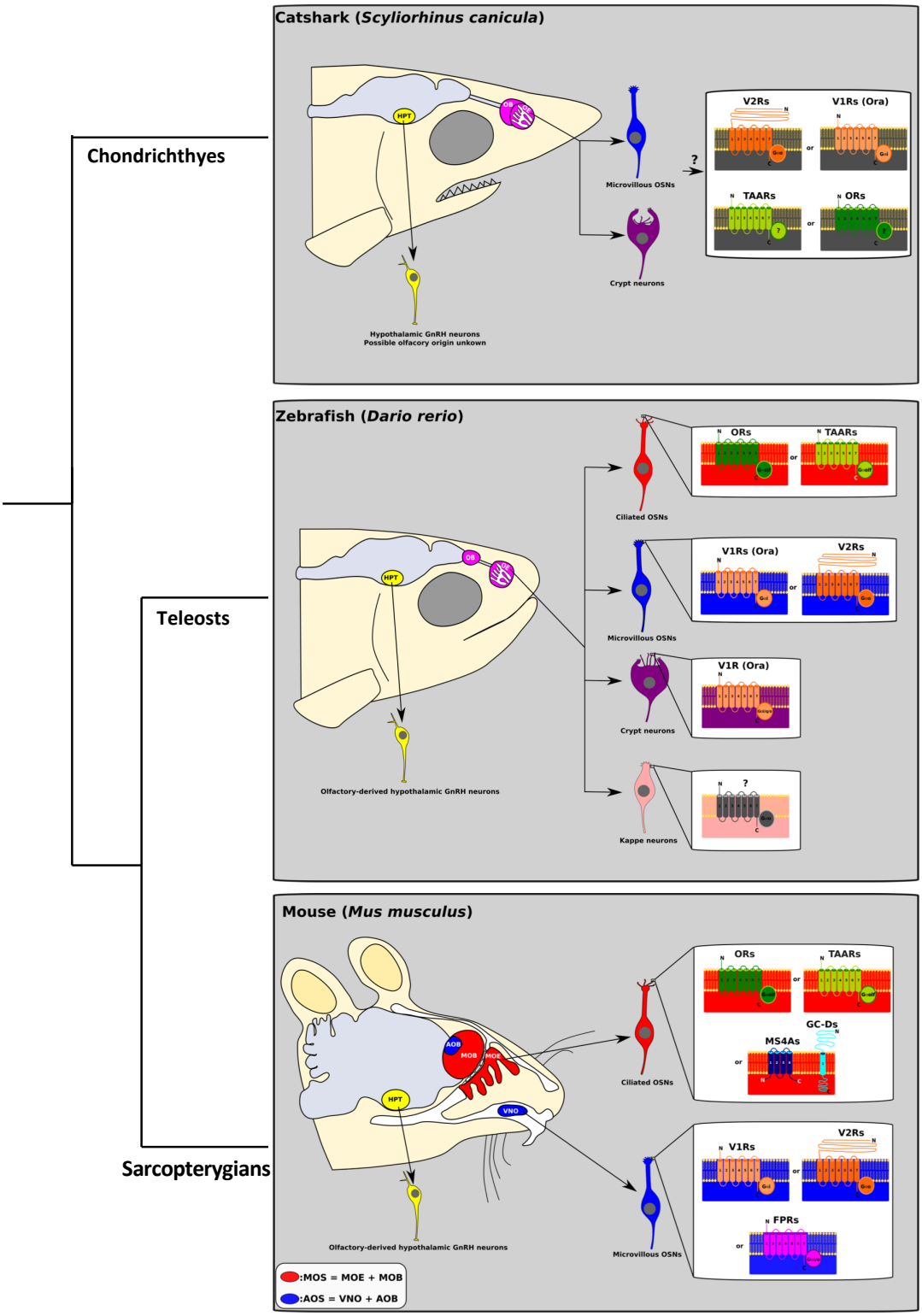


Figure 1.2. Schematic diagrams of adult organisation of the main and accessory olfactory systems in shark, teleost, and mammal lineages, including some of the neural cell types discussed in the text and shown in Table 1.1. Note that olfactory-derived GnRH neurons of the terminal nerve and NPY neurons are not represented here.

Table 1.1 Summary of neural cell types developing from the olfactory placode and its derivatives. Abbreviations: MOS, main olfactory system; AOS, accessory olfactory system; FPR, formyl peptide receptor; GC-D, guanylate cyclase D; GnRH, gonadotropin-releasing hormone; OSN, olfactory sensory neuron; OR, olfactory receptor; TAAR, trace amine-associated receptor; MS4A, membrane-spanning 4A receptor; V1R, vomeronasal type 1 receptor; V2R, vomeronasal type 2 receptor.

taxonomic group	location/ embryo origin	sensory cell types and receptors	G protein used	additional comments
sarcopterygians	MOS	ciliated OSNs expressing ORs	Gαolf	
		ciliated OSNs expressing TAARs	Gαolf	
		ciliated OSNs expressing GC-D family receptors and MS4As family receptors	not G protein coupled	demonstrated only in rodents
	AOS	microvillous OSNs expressing V1Rs	Gαi	AOS lost in some tetrapod lineages
		microvillous OSNs expressing V2Rs	Gαo	
		microvillous OSNs expressing FPRs family	Gαi/o	demonstrated only in rodents
teleosts	single epithelium	ciliated OSNs expressing ORs	Gαolf	
		ciliated OSNs expressing TAARs	Gαolf	
		microvillous OSNs expressing V1Rs	Gαi	
		microvillous OSNs expressing V2Rs	Gαq	
		crypt cells with cilia and microvilli expressing V1Rs	Gαi/o/q	
		cap cells, receptor unknown	Gαo	
chondrichthyes	AOS only	microvillous OSNs likely expressing V2Rs	Gαo	dominance of V2Rs, with few TAARs, V1Rs and ORs genes. It is not precisely known which receptor is expressed by each sensory cell type
		crypt cells with cilia and microvilli likely expressing V1Rs	unknown	
possibly all gnathostomes	olfactory placode	neuropeptide Y neurons	N/A	migrate from placode to hypothalamus, but this has only been demonstrated in chicken
		GnRH1 neurons	N/A	migrate from olfactory placode to hypothalamus and pre-optic area. It has been demonstrated in osteichthyes but not to date in chondrichthyes. In some species, they also migrate to form the terminal nerve if GnRH3 is absent (functional compensation of paralogue)
		GnRH3 neurons	N/A	migrate from olfactory placode to form the terminal nerve. In some species, they also migrate to the hypothalamus and pre-optic area if GnRH1 is absent (functional compensation of paralogue)

1.2.4. Development of the jawed vertebrate olfactory system from the olfactory placodes

The vertebrate cranial placodes are transient ectodermal thickenings of embryonic head and contribute to the developing cranial sensory systems. In jawed vertebrates, the olfactory system develops from a pair of cranial placodes, the olfactory placodes. Other cranial placodes such as the lens, vestibulo-acoustic, trigeminal, epibranchial and lateral line placodes contribute to other cranial senses, including sight, hearing, balance, somatosensation, gustation and internal physiological monitoring: their respective cell types and functions have been extensively reviewed elsewhere and will not be further considered here (Baker and Bronner-Fraser, 2001; Patthey et al., 2014; Schlosser, 2005; Schlosser et al., 2014). However, one other placode develops in intimate association with the olfactory placodes and warrants further mention. The adenohypophyseal placode forms between the paired olfactory placodes, in front of the extreme anterior of the neural plate. During subsequent development, it becomes internalized through the mouth (Saint-Jeannet and Moody, 2014), eventually forming the adenohypophysis and thus having a direct functional connection with olfactory placode-derived GnRH neurons in the hypothalamus. All the cranial placodes arise from the pre-placodal ectoderm (PPE), a U-shaped cell field around the edge of the anterior neural plate (Fig. 1.3). The PPE is specified by fibro- blast growth factor (FGF) signalling and bone morphogenetic protein (BMP)- and Wingless-related integration site (Wnt)- antagonism (Schlosser, 2006). It is characterized by the expression of pre-placodal competence factors such as the transcription factors (TFs) Six1/2, Six4/5, Eya1-4, Dlx3/5/6, Gata3 and Foxi1 (Aguillon et al., 2016). During the development, the PPE subdivides in specific regions along the

anteroposterior axis to give rise to individual cranial placodes. In particular, the lens, adenohypophysis and olfactory placodes are defined anteriorly through the combinatorial expression of TFs such as Dmrt4, Otx2/5, Pax6, FoxE, Six3/ 6 and Pitx1/2 (Toro and Varga, 2007). BMP signalling promotes specification of an olfactory fate, while extended BMP exposure time promotes lens fate (Sjodal et al., 2007). FGF signalling from the anterior neural plate is known to block expression of the lens marker gene Pax6 and to promote the expression of Dmrt4, an olfactory placodal gene (Bailey et al., 2006; Huang et al., 2005). Olfactory placodes are further characterized by the expression of the transcription factor genes Emx2 and COE2 (Ebf2), among others (Saint-Jeannet and Moody, 2014). In addition to the ectodermal cells from the PPE, the olfactory placodes become associated with migratory neural crest cells (Cho et al., 2019; Baker and Bronner-Fraser, 2001). These give rise to the olfactory ensheathing cells, which are glial cells that envelop the bundles of olfactory axons (Katoh et al., 2011; Forni et al., 2011; Barraud et al., 2010). The possible contribution of neural crest to other cell types in the olfactory placode such as the GnRH neurons is debated and controversial, and there is no coherent vision on the lineage origin of the major neural cell types associated with the olfactory sensory epithelia of vertebrates (Sabado et al., 2012; Forni et al., 2011; Whitlock et al., 2003). Precise lineage cell-tracing data in zebrafish argue against a contribution from the neural crest and support the view that all the different neuronal populations within the olfactory epithelium originate from overlapping pools of progenitors in the PPE (Aguillon et al., 2018). However, a recent analysis of GnRH1 neurons in the mouse olfactory placode argued for a heterogeneous origin, with neural crest and PPE-derived GnRH1 neurons (Shan et al., 2020). It is possible these apparently conflicting reports reflect genuine differences between species.

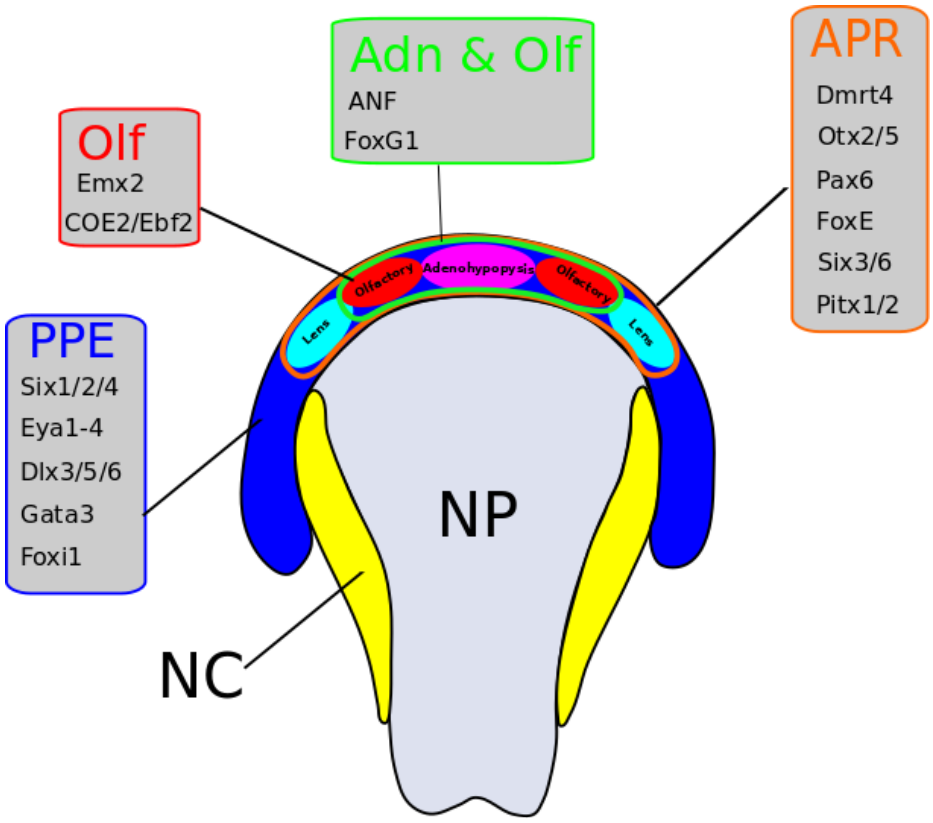


Figure 1.3. Anterior ectodermal patterning and origin of individual anterior placodes from the pre-placodal ectoderm (PPE) in jawed vertebrates. The PPE is specified by FGF signalling, BMP antagonism and Wnt antagonism, all coming from the underlying mesoderm. These induce transcription factors within the PPE (blue), which specify precursor regions for multiplacodal areas (coloured outlines) and individual placodes (coloured ovals). Known transcription factors for each are shown (for a more detailed discussion of these genes and genes marking other placodes see (Schlosser et al., 2014). Abbreviations: And, adenohypophyseal placode. APR,

1.2.5 The olfactory system in jawless vertebrates

There are only two surviving lineages of jawless vertebrates, the lampreys and the hagfishes (Fig 1.1), though fossils show a much wider diversity of extinct lineages (Johanson et al., 2019). There are many similarities in head development between jawed and jawless vertebrates, including sensory systems and the central nervous system. For example, most jawed vertebrate placodes are clearly identifiable in lampreys (Kuratani et al., 1997; Modrell et al., 2014); cranial nerve organization is similar (Kuratani et al., 1997) and gross brain organization well-conserved (Parker et al., 2014; Sugahara et al., 2016). There is, however, an important difference in the olfactory system. Jawed vertebrates have paired nostrils leading to paired olfactory sacs and derived from the paired olfactory placodes. Lampreys and hagfish have a single median nostril, a condition known as monorhiny. This develops to form a single anterior medial placode known as the nasohypophyseal placode, which combines characters of both olfactory and adenohypophyseal placodes and produces a single, medial nasal (olfactory) sac. Despite this difference, both lampreys and hagfishes have well-developed olfaction. Adult sea lamprey uses odours from conspecific larvae (including dihydroxylated tetrahydrofuran fatty acids and some bile acids) to select the best streams for spawning based on their larval population (Bjerselius et al., 2000; Li et al., 2018; Li et al., 1995; Siefkes and Li, 2004; Zielinski et al., 1996). Hagfishes are usually found in deep water and their olfactory organ seems particularly efficient as they have been shown to be among the first fish species to locate chemical signals of decaying prey (Glover et al., 2019; Martinez et al., 2011). Electrophysiological recordings indicate that their olfactory sensory neurons are particularly sensitive to amino acids (Døving and Holmberg, 1974). In addition to the conventional olfactory system, hagfishes have specialized chemosensory structures

named 'Schreiner organs' all over the body epidermis (Braun, 1998). Their ecological role is not known, though we can speculate they may help with directional chemosensation in the absence of paired olfactory membranes.

1.2.6. The olfactory system of lamprey and its neural cell-type derivatives

In larval and adult lamprey, the single olfactory organ is composed of three elements: the nasal duct, the nasal sac and the nasopharyngeal pouch (Fig 1.4) (Leach, 1951). The nasal tube opens externally as a single nostril on the dorsal head surface. In adult lampreys, the nasal tube contains a valve that serves to introduce and expel water into the entrance of the nasal sac, the chemosensory part of the organ (Kleerekoper and van Erkel, 1960). When larval lamprey metamorphoses into adults, the olfactory organ extends and changes from an epithelial lined tube to a nasal sac with lamellar folds (VanDenbosshe et al., 1997). In the sea lamprey *Petromyzon marinus*, the nasal sac wall is divided into 25 folds (Kleerekoper and van Erkel, 1960). Each fold is lined with the main olfactory epithelium, which is mainly composed of tall, narrow, ciliated OSNs (Thornhill, 1967; VanDenbosshe et al., 1995). However, OSNs in the main olfactory epithelium display at least three distinct morphotypes, with some not necessarily ciliated but with microvillar-like protrusions. These different OSNs have been proposed to be similar to the ciliated OSNs, microvillous OSNs and crypt cells found in teleost fishes (Laframboise et al., 2007). The OSNs in the main olfactory epithelium express the three chemoreceptor gene families identified in the sea lamprey genome, the ORs, TAARs and V1Rs; the V2R gene family seems to be gnathostome-specific as it is apparently absent in lamprey (Freitag et al., 1999; Grus and Zhang, 2006; Libants et al., 2009; Ubeda-Banon et al., 2011). The OSNs of the main olfactory epithelium send projections mainly to the non-medial region of the

olfactory bulbs, but also send some projections to its medial region (Green et al., 2013). In the caudoventral portion of the peripheral olfactory organ, there is an accessory olfactory organ (Scott, 1887), which is covered with the accessory olfactory epithelium containing short, rounded, ciliated neurons (de Beer, 1924; Ren et al., 2009). The accessory olfactory epithelium sends projections exclusively to the medial olfactory bulb, which connects to the posterior tuberculum creating a motor response from olfactory inputs (Ren et al., 2009; Derjean et al., 2010). It has been hypothesized based on anatomical and molecular evidence that the lamprey accessory olfactory epithelium, coupled with the dorsomedial telencephalic neuropil, is the putative homologue of the tetrapod vomeronasal system (Chang et al., 2013; Hagelin et al., 1955). However, it remains questionable whether the so-called 'main' and 'accessory' olfactory epithelia of lamprey are indeed homologous to the main and accessory olfactory epithelia of sarcopterygians. Characters that point to homology are that the lamprey main and accessory olfactory epithelia have differences in their respective pathways and that distinct G protein subtypes are used for signal transduction, with Gαolf located only in the OSNs projecting to the non-medial olfactory bulb (Frontini et al., 2003). Hence, different types of G proteins are used from those of OSNs projecting to the medial glomeruli, a similarity shared with the vomeronasal organ. However, there is a notable difference in that ORs, TAARs and V1Rs are not differently expressed between the lamprey olfactory epithelia as opposed to the tetrapod vomeronasal organ and main olfactory epithelium (Ren et al., 2009; Chang et al., 2013). This difference could represent an intermediate and ancestral condition before the exclusive shift to vomeronasal receptor genes, as seen in the AOS of gnathostomes. In lampreys, current evidence suggest that GnRH neurons of the pre-optic area and hypothalamus are not derived from the nasohypophyseal (NHP) placode, contrary to what is observed in jawed vertebrates

(Muske, 1993; Tobet et al., 1996). Immunohistochemical investigation concluded that lamprey GnRH neurons were never seen in association with the NHP placode during embryonic development, and it was hence hypothesized that GnRH neurons originate within proliferative zones of the diencephalon in developing lamprey, not in the olfactory system (Tobet et al., 1996). However, other data supporting this difference to jawed vertebrates are lacking.

1.2.7. The olfactory system of hagfish and its neural cell-type derivatives

As in lampreys, the olfactory system of adult hagfish is composed of three main parts: a nasal duct, a nasal sac and a nasopharyngeal duct (Fig. 1.4) (Døving, 1998; Theisen, 1973). Hagfishes also possess a single nostril just above the mouth, surrounded by two nasal tentacles on each side and by a dorso-median lip. The nasal duct leads to the nasal sac anterior to the brain (Theisen, 1973; Theisen, 1976). A valve is present in an oblique position inside the nasal duct and serves to manage water flow in the duct towards the nasal sac (Holmes et al., 2011). The latter receives a continuous flow of water during respiration as the water flows from the nostril and is ejected through the gill openings as, unlike in lamprey, the nasopharyngeal duct does not end blindly but opens into the pharynx (Theisen, 1973; Theisen, 1976). The nasal sac is divided into seven olfactory laminae and the olfactory epithelium is composed of two types of OSNs, ciliated or microvillous. In adult hagfishes, GnRH neurons have been identified in the diencephalon (Braun et al., 1995; Sower et al., 1995, Blähser et al, 1989). However, the embryonic development of these cells has not been investigated and nothing is known about the potential association or shared origin of hagfish GnRH neurons with the olfactory system.

1.2.8. Olfactory system development in jawless vertebrates

The developmental trajectories of lamprey and hagfish systems are shown in Figure 1.4. Early in development the single median nasohypophyseal placode is characterized by orthologous molecular markers to those seen in the gnathostome olfactory placode. The entire nasohypophyseal placode territory expresses Six3/6A and Soxb1 in hagfish (Oisi et al., 2013) and Pax6 in lamprey (Derobert et al., 2002; Murakami et al., 2001), consistent with the expression of the gnathostome orthologues Pax6, Six3/6 and Sox2/3 in both the olfactory and adenohypophyseal placodes (Schlosser, 2005). At the late neurula stage of lamprey and hagfish embryos, the anterior part of the nasohypophyseal placode becomes the likely olfactory territory as it expresses FGF8/17, the orthologue of FGF8 in gnathostomes. Reciprocally, the posterior part becomes the adenohypophyseal territory as it shows PitxA expression (Oisi et al., 2013; Uchida et al., 2003). At the pharyngula stage in lamprey, the nasohypophyseal placode forms a thickened area of the ventral ectoderm, anterior to the mouth cavity. Morphogenesis of this region is coordinated with that of two ectomesenchymal processes (Fig 1.4) and has been described elsewhere (Oisi et al., 2013). Important points to note are that the nasohypophyseal placode extends an epithelial cell process posteriorly to establish close contact with the definitive hypothalamic region, thus forming a pituitary similar to that of jawed vertebrates in combining central nervous system and placode-derived parts. The anterior part of the nasohypophyseal placode, which will form the future olfactory epithelium, is characterized by the expression of olfactory developmental gene markers such as OtxA, CoeA, CoeB, EmxA and EmxB (Murakami et al., 2001; Uchida et al., 2003; Lara-Ramirez et al., 2017; Myojin et al., 2001). As lamprey embryos approach the larval stage, the anterior part of the nasohypophyseal placode differentiates as the nasal sac, composed of a thick

columnar epithelium and covering the rostral aspect of the telencephalon, whereas the posterior (as the future adenohypophysis and consisting of an epithelium of a few cell layers) extends caudally to the level of the optic chiasma (Uchida et al., 2003). In late hagfish embryos, the nasohypophyseal duct and oral cavity grow posteriorly relative to the position of the adenohypophysis. The presumptive nasohypophyseal duct is tilted inward in an oblique position unlike that of lamprey, which is situated more vertically. In addition, the posterior end of the nasohypophyseal duct in hagfishes ruptures into the pharynx (Fig. 1.4). Thus, the nasohypophyseal duct in hagfishes opens secondarily into the pharynx (Oisi et al., 2013). To summarize, there are many similarities in development, gene expression and cell-type derivatives between the olfactory systems of jawless and jawed vertebrates. There are also some important differences. Most notably, jawless vertebrate systems develop from a single medial placode combining olfactory and adenohypophyseal progenitors that separate during morphogenesis, and form a single medial olfactory system and not the paired systems of jawed vertebrates. Fossil data suggest monorhiny is the ancestral condition (Gai et al., 2011; Janvier, 1996; Kuratani et al., 2001), so a single medial olfactory/adenohypophyseal placode as seen in living jawless fish is probably also ancestral (though note the description of paired nasal sacs in some vertebrate stem lineage fossils, and that despite monorhiny, lampreys and hagfishes have a pair of olfactory nerves like gnathostomes, meaning that some questions remain over this inference (Shu et al., 2004; Moris and Caron, 2014; Pombal and Megías, 2019). It is also unclear whether GnRH neurons are olfactory placode derivatives in jawless fish like GnRH1 and GnRH3 neurons in gnathostomes, or are born within the central nervous system like gnathostome GnRH2 neurons.

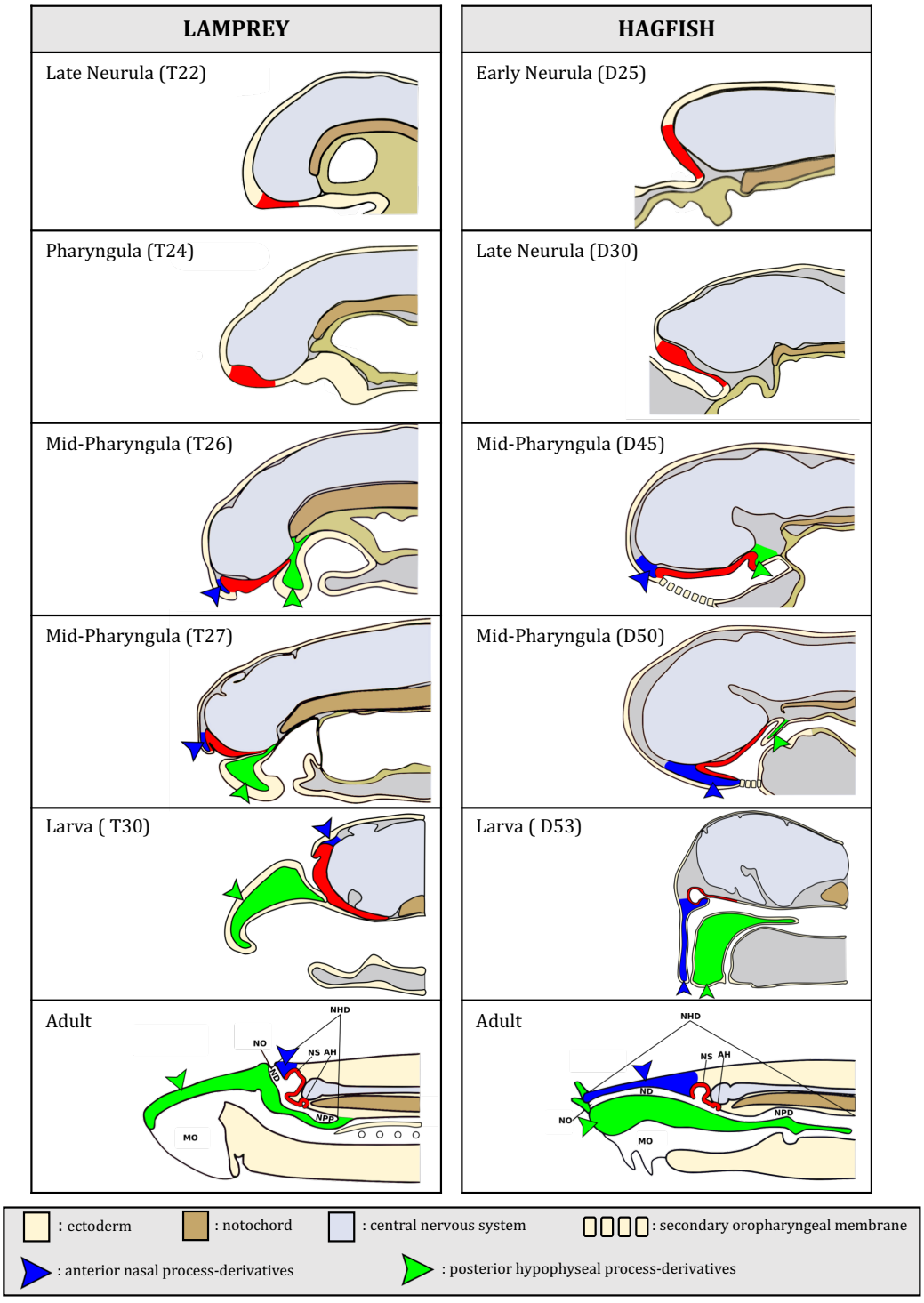


Figure 1.4. Comparative development of nasohypophyseal placode (NHP) in lampreys and hagfishes. The NHP is labelled in red. At the mid-pharyngula stage when development is the most similar in cyclostomes, the NHP is rostrocaudally bordered by ventral growth of two ectomesenchymal processes, the anterior nasal process (blue) and the posterior hypophyseal process (green). Abbreviations: AH, adenohypophysis. MO, mouth. ND, nasal duct. NHD, nasohypophyseal duct. NO, nostril. NPD, nasopharyngeal duct. NPP, nasopharyngeal pouch. NS, nasal sac. Figure adapted from (Oisi et al., 2013) with permission. T and D denote lamprey and hagfish embryo staging respectively (Tahara, 1988; Dean, 1899).

1.2.9. Potential olfactory system homologues in protochordates

The tunicates and cephalochordates are collectively known by the paraphyletic term 'protochordates' (Fig. 1.1). They are usually considered not to have cranial placodes in a strict sense (Gans and Northcutt, 1983). However, they do have some ectodermal patterning mechanisms that appear conserved with those of vertebrate placodes. These have been best studied in the ascidians, which have two areas of ectoderm postulated to be placode homologues: one just anterior to the neural plate and a potential homologue of the olfactory and adenohypophyseal placodes (discussed more below) and the other paired and lateral to a more posterior neural plate. Data on ectodermal patterning have been recently reviewed in detail elsewhere (Patthey et al., 2014; Schlosser et al., 2014; Schlosser, 2017) and the reader is referred here for more discussion on this aspect. Protochordates also have several types of morphologically distinct ectodermal sensory cells. Some have been proposed to be chemosensory, though this is mainly based on cytological features and the expression of marker genes, and no cell advocated as chemosensory has had this experimentally evaluated (Schlosser, 2017; Caicci et al., 2010; Gropelli et al., 2003; Imai and Mainertzhagen, 2007; Torrence and Cloney, 1982, 1983). Amphioxus species are all quite similar in gross morphology. Tunicates are more disparate. The majority of species fall into the 'Ascidiacea', a paraphyletic grouping united by an ascidian type life cycle, with a motile tadpole larva and a sessile adult that may be solitary or may form colonies by asexual reproduction (Delusc et al., 2018). Other tunicates are motile as adults. This includes the larvaceans, which maintain the tadpole body plan throughout life, and thaliaceans (Fig 1.1, Fig. 5A–C).

1.2.10. Putative olfactory cell homologues in Ascidiacea

Several studies have reported a failure to identify orthologues of the vertebrate olfactory receptor genes in tunicate genomes, and orthologues of chemosensory receptor genes used by insects and nematodes could also not be found (Churcher and Taylor, 2009; Kamesh et al., 2008; Niimura, 2009). Tunicate orthologues of V1R and V2R genes are also missing (Grus and Zhang, 2006; Libants et al., 2009; Nordstorm et al., 2008). Furthermore, although TAAR-like genes have been proposed to be present in tunicates and amphioxus (Libants et al., 2009), these authors did not give details and others have disagreed with this as they failed to identify TAAR orthologues (Grus and Zhang, 2006; Eyun et al., 2017). It would be surprising if tunicates do not sense any chemical cues, especially with an active swimming larval stage that in most species needs to choose an appropriate settlement site for metamorphosis. Further, in adult tunicates, there is evidence that the oral siphon can sense chemicals as sensitivity to acids, bases, salts and anaesthetics has been shown (Hecht, 1918). It is therefore probable that alternatives to olfactory receptor genes are used in these organisms. Recently, some seven-transmembrane protein encoding genes have been identified and proposed to fulfil the role (discussed more below), though protein function is not established (Abitua et al., 2015; Johnson et al., 2020). The model ascidian *Ciona intestinalis* is by far the best-studied species in this field. During neurulation, the border region to the neural plate gives rise to several types of sensory cell, two of which are relevant to the discussion of olfactory placode evolution: a subset of anterior trunk epidermal neurons (the aATENs) and the palp sensory cells (PSCs) (figure 5a). Both arise from the cells just anterior to the neural-plate, in the region mentioned above as a prospective olfactory/adenohypophyseal homologue. The developmental pathway leading to the specification of these sensory cells has been expertly reviewed elsewhere (Liu and Satou,

2020), and the reader is referred here for a detailed description of how they form and then acquire distinct identities. The aATENs are four ciliated primary sensory neurons that in Ciona larvae come to be located in the epidermis dorsal to the cavity of the sensory vesicle, their simple equivalent of the brain (Imai and Meinertzhagen, 2007; Yokoyama et al., 2014). These neurons express GnRH, a cyclic nucleotide-gated channel (CNGA) and, as mentioned above, two seven-transmembrane G-protein-coupled receptors: a relaxin-3 receptor (RXFP3) and a somatostatin/ opioid/galanin/chemokine-like receptor (SOG/Chemokine- like) (Abitua et al., 2015). This led the authors to propose that the aATENs had dual chemosensory and neurosecretory activities, combining functions of vertebrate olfactory-derived OSNs and GnRH neurons in a single cell (Abitua et al., 2015). However, caveats to this are that functional chemosensory activity has not been experimentally shown, and that the two seven-transmembrane G-protein-coupled receptors identified in these cells are not orthologous to the vertebrate olfactory receptors. Testing sensory cells for chemical stimulation has been an experimental challenge in a developmental system like Ciona as the small cells make electrophysiology difficult. The recent development of fluorescent cell activity reporters will probably resolve this technical bottleneck (Johnson et al., 2020; Okawa et al., 2020). Ascidian tadpoles also have ciliated primary sensory neurons in each of the palps, the anterior adhesive organs by which larvae appear to sense and bond to attachment sites (Torrence and Cloney, 1983; Zeng et al., 2019a). Like aATEN cells, the palps develop from the area anterior to the neural plate postulated to be an olfactory/adenohypophyseal placode homologue and express many regulatory genes that are important for olfactory placode development in vertebrates, like Eya, COE, Dmrt, FoxC, FoxG, FGF, Sp8, Dlx and Isl (Cao et al., 2019; Liu and Satou, 2019; Mazet et al., 2005; Wagner et al., 2014). It is possible that the palp sensory neurons are involved in tadpole settlement site selection via

chemoreception, though again this is not experimentally validated, and others have suggested that a different cell type in the palps may be chemosensory (Johnson et al., 2020). The palps are also known to produce GnRH in cells likely to be neuronal as they seem to possess long axons (Kusakabe et al., 2012), suggesting an evolutionary link to the olfactory placode-derived GnRH neurons in gnathostomes.

In addition to these larval cell types, the oral siphon primordium (OSP; Fig. 1.5A) develops in the same region, with its progenitors sandwiched between those that give rise to the aATENs and those that give rise to the palps. The primordium maintains the expression of placode marker genes including Six1/2, Six3/6 and Pitx into the larval stage and at metamorphosis forms the oral siphon, which in adults includes sensory cells in many ascidian species. Some of these cells may be chemosensory (Hecht, 1918) but their developmental origin has yet to be traced so it is still possible they do not derive from the primordium. A structure called the ciliated funnel opens into the oral siphon and connects to a gland associated with a ganglion, the combination of which is termed the neural complex. This dual structure is reminiscent of the pituitary and homology has been considered (Manni et al., 2005), with the ciliated funnel sensing water entering the oral siphon and perhaps chemosensory. As for other ascidian sensory cells, this has not been experimentally validated.

1.2.11. Putative olfactory cells in larvaceans

It has been suggested that the ventral organ (Fig 1.5B) of the larvacean tunicate Oikopleura dioica is homologous to the vertebrate olfactory organ. The ventral organ possesses about 30 primary sensory cells with cilia that protrude externally into sea water. These neurons are located in an ectodermal slit-like pocket and send their axons to the rostral-most CNS (Bollner et al., 1986). Furthermore, developmental genes like Eya, Pitx and Six1/2, which are important in the development of vertebrate olfactory and adenohypophyseal placodes, are expressed in the primordia of the larvacean ventral organ. Therefore, the *Oikopleura* ventral organ placode was said to be homologous to the ectoderm of the ascidian palps based on gene expression and structure (Bassham and Postlethwait, 2005). Again, the sensory function of the cells has not been tested.

1..2.12. Putative olfactory cells in thaliacea

Thaliaceans comprise the pelagic tunicates salps, doliolids and pyrosomes. There has been less research in thaliaceans in comparison to other tunicate classes. In salps and doliolids, the cells located around the oral lips (Fig. 1.5C) have been suggested to be chemoreceptors based on the observation that salps respond to chemical stimuli positioned in proximity of the oral opening (Madin, 1995). Another possible chemosensory structure is the ciliated funnel, as discussed above with respect to ascidians and which is present in thaliaceans. In the thaliacean *Thalia democratica*, it has been suggested that the ciliated funnel could possibly collect odorants from the environment (Pennati et al., 2012). Developmental and genetic confirmation of this hypothesis is currently lacking, however.

1.2.13. A summary of olfactory system homology in tunicates

Experiments show tunicates respond to chemical stimuli, and position and developmental genetic data point to the ectoderm just anterior to the neural plate as the homologue of the vertebrate olfactory and adenohypophyseal placodes. This area also produces sensory neurons, at least some of which express GnRH. It remains to be experimentally shown whether cells are chemosensory, and if so what receptors they use considering homologues of vertebrate receptor families are lacking. However, assuming they are chemosensory, the data strongly support the contention that the ancestor of the tunicates and vertebrates had a chemosensory system developing from the ectoderm alongside the anterior neural plate and that has evolved into the systems we see today in living tunicates and vertebrates. It is less clear how complicated this ancestral system was. In tunicate larvae, it consists of scattered sensory cells that may combine multiple functions, rather than a larger organ system with the morphogenesis and specialized cell types of vertebrates. This points to a simple grade of organization in the common ancestor. Adult tunicates have morphologically more complex organs with more cells, which could point to a more complex ancestral state. However, cell functions are again unknown, developmental genetics poorly understood and cell lineages unclear. This makes discriminating between conservation and parallelism or convergence challenging. Until additional data are available, the analysis of outgroups like amphioxus provides the alternative route to inferring ancestral states.

1.2.14. Putative olfactory cells in amphioxus

Amphioxus are known to exhibit sensitivity to chemicals dissolved in sea water, with most triggering an escape response (Parker, 1908; Zieger et al., 2018). Ectodermal territories with the gene expression, morphogenetic processes or focused neurogenesis characteristic of vertebrate placodes have not been identified, with the possible exception of Hatschek's pit (see below). Ectodermal sensory cells are present but are broadly scattered throughout the general epidermis (Fig. 1.5D), and at least some of these cells develop in the ventral ectoderm of the early embryo under BMP signalling (Lu et al., 2012). It is not known which sensory modalities are mediated by each cell but some have been suggested to be chemosensory based on cytology and molecular markers reviewed in (Patthey et al., 2014; Schlosser, 2017; Holland, 2005; Lacalli, 2004). There are two major subtypes of these epidermal sensory cells: type I and type II. The type I sensory cells are primary sensory cells with a long cilium surrounded by microvilli. The type II sensory cells are secondary neurons and have a short cilium encircled by a collar of microvilli (Baatrup, 1981; Lacalli and Hou, 1999; Stokes and Holland, 1995). The type I cells have been proposed to be mechano- and/ or chemoreceptors (Bone and Best, 1995). The population of type I sensory neurons is molecularly heterogeneous and subsets express orthologues of transcription factor genes seen during vertebrate olfactory placode development such as COE, Islet, POU4, SoxB1c, Six1/2, Six4/5 and Eya (Meulemans and Bronner-Fraser, 2007; Candiani et al., 2006; Jackman and Langeland, 2000; Kozmik et al., 2007; Mazet et al., 2004). Some may be chemosensory, for instance, the expression of COE by type I sensory cells located caudally along the flanks of the amphioxus body might suggest this as COE genes are expressed in the olfactory placodes of vertebrates and chemosensory neurons of the organism such as *Caenorhabditis elegans*

(Lara-Ramirez et al., 2017; Kim et al., 2005). In particular, the anterior of amphioxus is interesting as it includes type I ciliated primary sensory neurons coming from an ectodermal region expressing olfactory placodal markers like Pax6, Six3/6, Ngn and POU4 (Candiani et al., 2006; Jackman and Langeland, 2000; Kim et al., 2005). Furthermore, in the amphioxus *Branchiostoma floridae* genome, there are more than 50 genes orthologous to the vertebrate ORs (Churcher and Taylor, 2009; Niimura, 2009; Churcher and Taylor, 2011; Niimura, 2009), and some of these anterior type I cells express at least one of these genes suggesting they may be cell-type homologues of OSNs (Satoh, 2005). Type II cell morphology, with a collar of external projecting microvilli supported by a modified cilium, also suggests chemosensory rather than mechanosensory function. The cells could be homologous to vertebrate primary olfactory sensory cells, which means they would have secondarily lost their axons. They could also be homologous to secondary vertebrate chemosensory cells like the taste buds or solitary chemoreceptor cells. However, the molecular identity of the cells is unknown as their late appearance in larval development (Lacalli and Hou, 1999) has so far hindered molecular studies. All these hypotheses also need to be confirmed by physiological studies as none of these cells has had their sensory modality validated. In addition of type I and type II epidermal sensory cells, there are other specialized sensory cell types in amphioxus that have been inferred to be chemosensory; the cells from the corpuscles de Quatrefages, the oral spine cells and Hatschek's pit cells (Fig. 1.5D). The corpuscles de Quatrefages are located in the rostrum and form two clusters of specialized ciliated primary sensory neurons that have one to four sensory cells with two cilia each, surrounded by up to seven sheath cells (Baatrup, 1982; de Quatrefages, 1845). They have been speculated to form a mechanosensory organ but could well be chemosensory (Baker and Bronner-Fraser, 1997). The ciliated oral spine cells around the mouth opening do not possess axons or

microvilli, though they do express Pitx, POU4 and SoxB1c, which are vertebrate anterior placodal markers (Meulemans and Bronner-Fraser, 2007; Boorman and Shimeld, 2002). These specialized cells have been suggested to be mechanoreceptors but also proposed to be homologous to vertebrate taste cells (Lacalli et al., 1999). In adult amphioxus, Hatschek's pit is a structure located in the roof of the oral cavity and sends a projection dorsally around the notochord to contact the ventral brain, an organization similar to the relationship between adenohypophysis and hypothalamus of the vertebrate pituitary system (Patthey et al., 2014; Hatscheck, 1881). The idea that Hatschek's pit and pituitary may be homologous organs is over 100 years old, and molecular analysis has added some support to this since the developmental precursor to Hatschek's pit (known as the pre-oral pit) expresses the vertebrate anterior placode markers Six1/2, Eya, Pitx, Pax6 and Six3/6. There has also been a suggestion that Hatschek's pit includes chemosensory cells because they carry microvilli and cilia, and are exposed to water flowing into the mouth and so are in contact with potential odorants (Nozaki and Gorbman, 1992). This hypothesis requires experimental corroboration as the cytoarchitecture of the Hatschek's pit cells suggests that they are neurosecretory, and they do not have axons.

Immunochemistry with antibodies raised to vertebrate GnRH proteins has been used to suggest that GnRH neuro- secretory cells are present in the amphioxus neural tube, and possibly also Hatschek's pit (Nozaki and Gorbman, 1992; Chang et al., 1985). Neural tube GnRH cells are candidates for homologues of the olfactory-derived GnRH neurons of vertebrates, but nothing is currently known about their development: they might migrate in from an ectodermal territory like vertebrate GnRH1/3 cells, but also could be born within the central nervous system like vertebrate GnRH2 cells. The expression of GnRH in Hatschek's pit has also been questioned as it could not be detected in a later study, which only identified GnRH immunoreactivity in the central canal of the anterior nerve

cord (Roch et al., 2014). This difference in the detection of GnRH could be owing to genuine biological causes for example, a difference in the reproductive state of the animals examined might affect GnRH expression), or might reflect experimental artefacts from fixation or antibody cross reactivity. Additional work is needed to clarify this.

Overall, Hatschek's pit appears similar to the vertebrate adenohypophysis but not the olfactory system. However, even here there is an important difference in that Hatschek's pit derives from the pre-oral pit of the larva, which itself develops from an anterior head cavity, an endodermal derivative. The adenohypophysis is an ectodermal placode derivative. The lineages of the relevant Hatschek's pit cell types through this developmental process need to be confirmed for certainty, but if they too derive from endoderm then homology would imply that the capacity to build this organ and its associated cell types has transferred from one germ layer to another in chordate evolution. Such a shift in cell fate could have occurred through the shifting in the expression domains of the relevant transcription factors, as proposed by Schlosser (2005). This possibility has had recent support from cell lineage studies in zebrafish, which have demonstrated that in this species some adenohypophysis cells may naturally derive from the endoderm during normal development (Fabian et al., 2020). While this could be a derived character of zebrafish and not reflect the ancestral condition, it irrespectively shows confinement of pituitary cell fates by germ layer is not as strict as historically imagined.

1.2.15. A summary of olfactory system homology in amphioxus

In general, the scattered ectodermal sensory neurons of amphioxus are poor candidates for olfactory cell homologues. Most are not anteriorly located and while some express genes that mark vertebrate placode cells, there are also substantial differences in their developmental programmes. Most importantly their early induction and regulation are different as they form far from the neural plate in ventral ectoderm under high Bmp signalling (Lu et al., 2012). As such they more resemble another type of sensory neuron in tunicates, those in the ventral tail fin (Liu and Satou, 2020), and not aATENs, palp cells or vertebrate placode cells, which all originate in the neural plate boundary. Sensory cells in the anterior of amphioxus larvae may be the exception to this owing to their possible expression of amphioxus OR genes and anterior location. However, as yet their developmental genetics, cell lineage and sensory functions have not been determined and these data would be needed to convincingly test a hypothesis of homology. Hatschek's pit remains the best candidate for an adenohypophysis homologue, though this too has counterarguments as discussed above, and the expression of GnRH here remains to be convincingly established. With respect to inferring the organization of olfactory sensation in the common ancestor of amphioxus and other chordates, the data do not support the presence of morphogenetic processes building an olfactory organ at this stage in evolution. Rather, chemosensation may have been mediated by scattered sensory cells. Amphioxus OR gene expression data suggest assigning some of these as cell-type homologues of vertebrate OSNs but more work on their development is needed to support this. It would be especially important to know what specifies OR-expressing cells and if their lineage traces back to the anterior neural plate border.

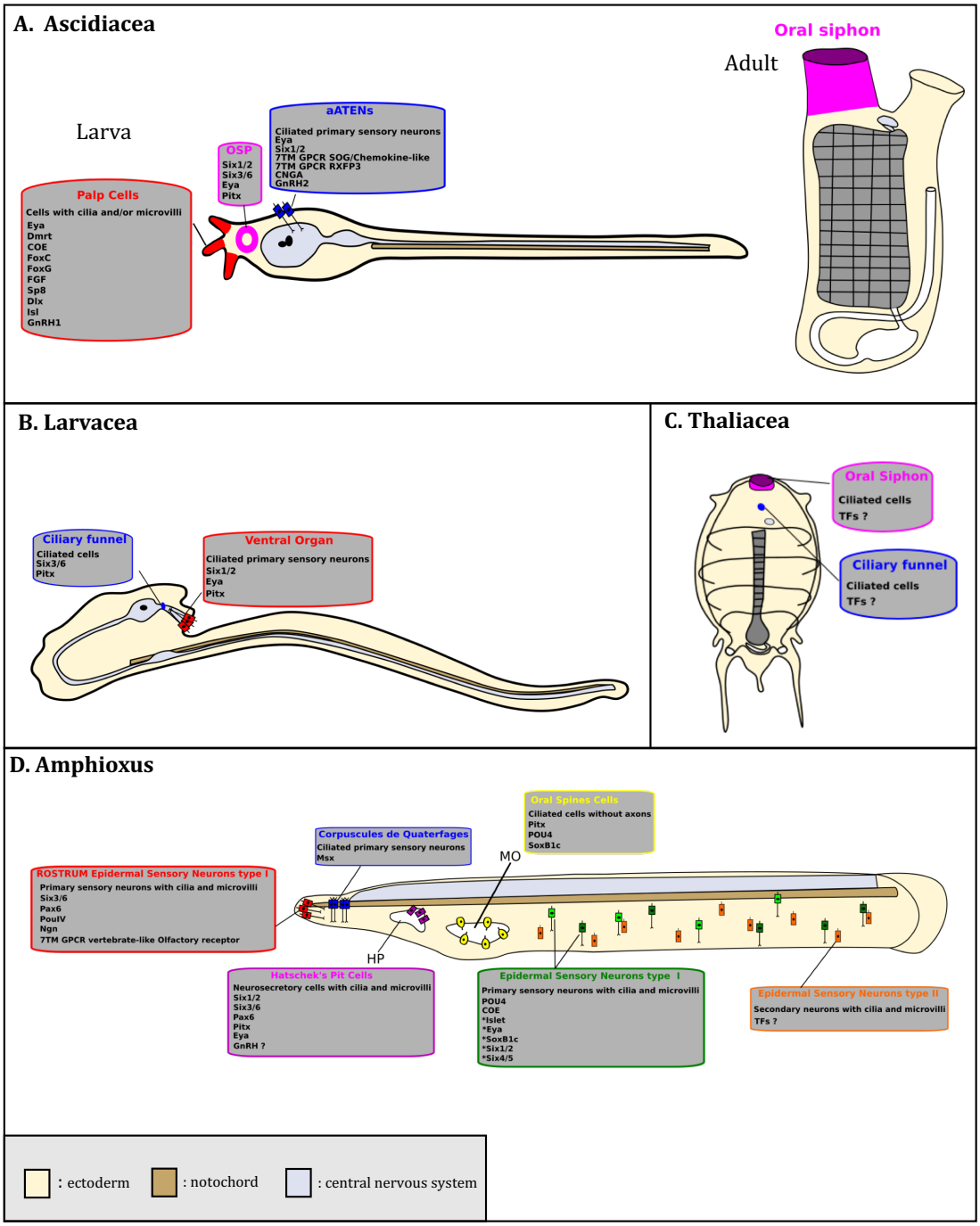
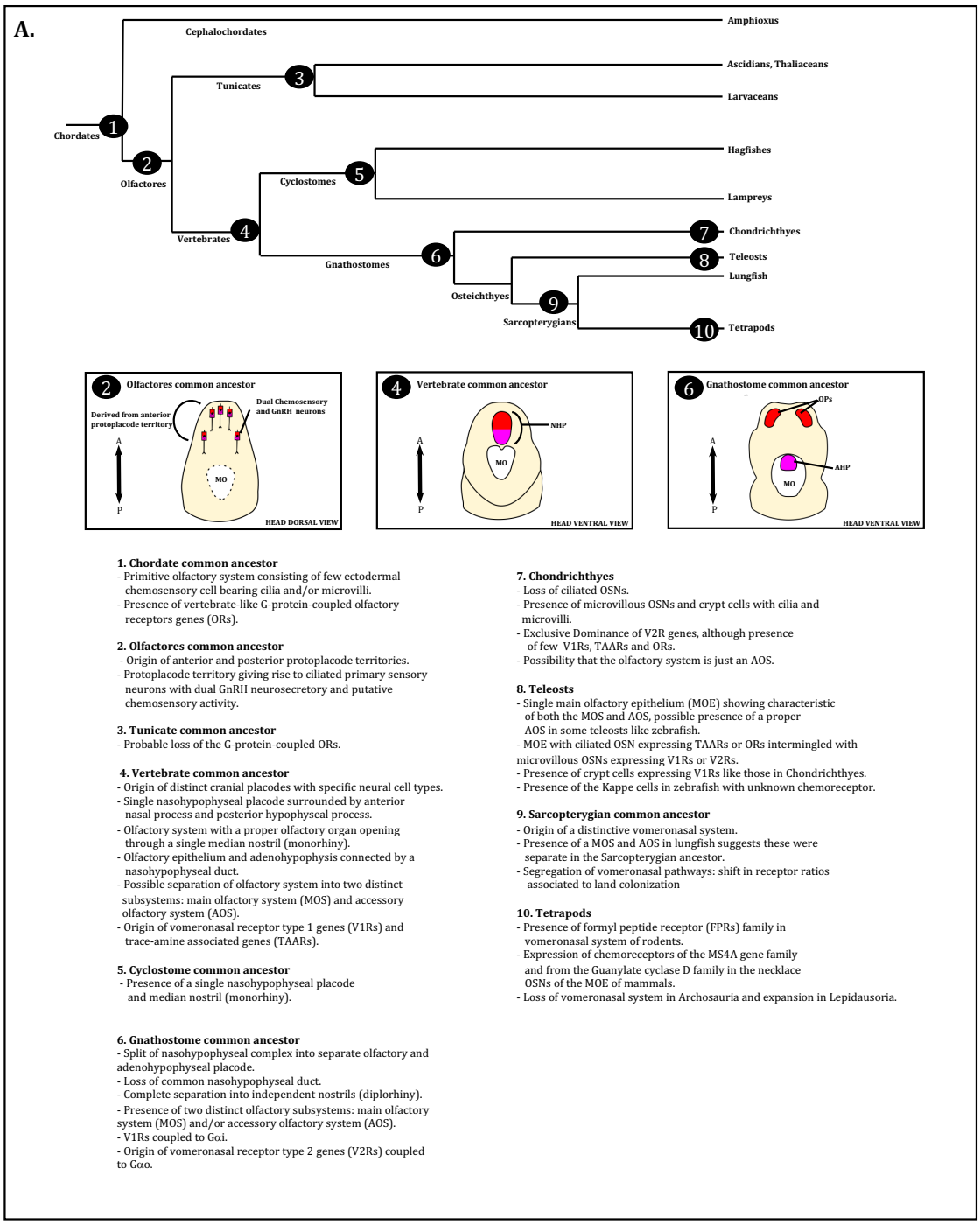


Figure 1.5. Possible olfactory sensory cells in protochordates and the genes they express. **A**. Schematic drawing of an ascidian larva and adult of *C. intestinalis*. The ascidian swimming larva has three regions suspected to hold or give rise to putative chemosensory cells: the palp cells (red), the oral siphon primordium (OSP, pink) which becomes the oral siphon in adults, and the anterior trunk epidermal neurons (aATENs, blue). **B.** Schematic drawing of the adult larvacean, *O. dioica*. Two main regions are believed to host chemosensory cells: the ciliary funnel (blue) and the ventral organ (red). **C.** Schematic drawing of the adult thaliacean, *T. democratica*. Chemosensory cells are thought to be present in two places: the oral siphon (pink) and ciliary funnel (blue). **D.** Schematic drawing of an amphioxus larva and the cells suspected to be chemosensory. The types of cells and relevant gene expression is indicated. Note that the Type 1 sensory cells are heterogeneous as indicated by different colours. Genes listed with an asterisk are only expressed in subsets of sensory cells. For more detail see (Schlosser, 2017; Jackman et al., 2000; Meulemans and Bronner-Fraser, 2007). Abbreviations: HP, Hatschek's Pit. MO, mouth. Panel D adapted from (Patthey et al., 2014) with permission.

1.2.16. A model for vertebrate olfactory system evolution

By plotting genes, developmental processes, cell types and tissues onto a phylogeny of the chordates we arrive at a model for how the olfactory systems of living vertebrates evolved (Fig 1.6). Chemosensation is an ancient sense and was likely mediated by scattered sensory cells in the epidermis of the chordate common ancestor (Fig. 1.6–1). Some of these cells probably expressed orthologues of the vertebrate OR genes, as still seen in living amphioxus. By the common ancestor of the tunicates and vertebrates (collectively the Olfactores), specialized ectodermal territories homologous to vertebrate placodes had evolved. We have not discussed the detailed evidence for this here as it has been recently and extensively reviewed (Patthey et al., 2014; Schlosser et al., 2014; Schlosser, 2017); however, the data suggest the Olfactores ancestor had two placode-like territories and that the anterior of these is the source of the olfactory and adenohypophyseal placodes of vertebrates (Fig. 1.6–2). In living tunicates, the anterior placode territory persists and is the source of GnRH and probably chemosensory cells, though the tunicate lineage also lost conventional OR genes. By the common ancestor of the vertebrates (Fig. 1.6–4) a well-defined anterior placode combining olfactory and adenohypophyseal progenitors had evolved, as well as the morphogenetic processes by which it built distinct olfactory and adenohypophyseal systems. It is possible this included distinct main and accessory olfactory system components, including the evolution and deployment of V1R receptor genes. We do not yet know whether it also included the progenitors of GnRH neurons that then migrated to the pre-optic area and hypothalamus as this has not been determined in cyclostomes. Key innovations that map to the origin of vertebrates therefore include: (i) the evolution of new types of receptor gene; (ii) as proposed by Schlosser (2017) for placode evolution more generally, changes

in the control of progenitor proliferation turning single neurons into a neurogenic organ; (iii) mechanisms for specifying subpopulations of OSNs expressing different types of receptor; (iv) changes in morphogenesis including the incorporation of neural crest cells and interactions between placode ectoderm and cranial ectomesenchyme. In the common ancestor of jawed vertebrates (Fig. 1.6–6), this combined placode separated into a medial adenohypophyseal placode and paired olfactory placodes, possibly facilitating the evolution of paired nostrils as seen in all living jawed vertebrates. Subsequently the loss of main or accessory olfactory systems has occurred in some vertebrate lineages, with concomitant shifts in dependence on ORs, V1Rs and V2Rs (Fig 1.6–7, 1.6–10). It is also interesting that studies of model systems have now identified additional olfactory chemosensory mechanisms beyond these well- known receptors, such as the MS4As, TAARs and FPRs. The evolutionary ancestries of these are less well known and worthy of more study. These findings come from just a handful of species, raising the possibility that the many thousands of less well-studied vertebrate species may harbour additional surprises on this front and that the diversity of olfactory mechanisms deployed by vertebrates may be far higher than currently understood.



1. Chordate common ancestor
- Primitive olfactory system consisting of few ectodermal chemosensory cell bearing cilia and/or microvilli.
- Presence of vertebrate-like G-protein-coupled olfactory receptors genes (ORs).

2. Olfactores common ancestor
- Origin of anterior and posterior protoplacode territories.
- Protoplacode territory giving rise to ciliated primary sensory neurons with dual GnRH neurosecretory and putative chemosensory activity.

3. Tunicate common ancestor
- Probable loss of the G-protein-coupled ORs.

4. Vertebrate common ancestor
- Origin of distinct cranial placodes with specific neural cell types.
- Single nasohypophyseal placode surrounded by anterior nasal process and posterior hypophyseal process.
- Olfactory system with a proper olfactory organ opening through a single median nostril (monorhiny).
- Olfactory epithelium and adenohypophysis connected by a nasohypophyseal duct.
- Possible separation of olfactory system into two distinct subsystems: main olfactory system (MOS) and accessory olfactory system (AOS).
- Origin of vomeronasal receptor type 1 genes (V1Rs) and trace-amine associated genes (TAARs).

5. Cyclostome common ancestor
- Presence of a single nasohypophyseal placode and median nostril (monorhiny).

6. Gnathostome common ancestor
- Split of nasohypophyseal complex into separate olfactory and adenohypophyseal placode.
- Loss of common nasohypophyseal duct.
- Complete separation into independent nostrils (diplorhiny).
- Presence of two distinct olfactory subsystems: main olfactory system (MOS) and/or accessory olfactory system (AOS).
- V1Rs coupled to Gαi.
- Origin of vomeronasal receptor type 2 genes (V2Rs) coupled to Gαo.

7. Chondrichthyes
- Loss of ciliated OSNs.
- Presence of microvillous OSNs and crypt cells with cilia and microvilli.
- Exclusive Dominance of V2R genes, although presence of few V1Rs, TAARs and ORs.
- Possibility that the olfactory system is just an AOS.

8. Teleosts
- Single main olfactory epithelium (MOE) showing characteristic of both the MOS and AOS, possible presence of a proper AOS in some teleosts like zebrafish.
- MOE with ciliated OSN expressing TAARs or ORs intermingled with microvillous OSNs expressing V1Rs or V2Rs.
- Presence of crypt cells expressing V1Rs like those in Chondrichthyes.
- Presence of the Kappe cells in zebrafish with unknown chemoreceptor.

9. Sarcopterygian common ancestor
- Origin of a distinctive vomeronasal system.
- Presence of a MOS and AOS in lungfish suggests these were separate in the Sarcopterygian ancestor.
- Segregation of vomeronasal pathways: shift in receptor ratios associated to land colonization

10. Tetrapods
- Presence of formyl peptide receptor (FPRs) family in vomeronasal system of rodents.
- Expression of chemoreceptors of the MS4A gene family and from the Guanylate cyclase D family in the necklace OSNs of the MOE of mammals.
- Loss of vomeronasal system in Archosauria and expansion in Lepidausoria.

Figure 1.6. Schematic phylogeny of chordate lineages. Numbers relate to the key below and indicate where key events in the evolution of the olfactory system are likely to have occurred. Abbreviations: AHP, adenohypophyseal placode. MO, mouth. NHP, nasohypophyseal placode. OP, olfactory placode.

Chapter 2 Embryonic origin and specification of Lamprey GnRH neurons

Work Declaration

The experimental work and subsequent data analyses presented in this Chapter have performed by me, Guillaume Poncelet.

2.1 Introduction

2.1.1 Research context

Understanding how pluripotent progenitors give rise to different neuronal subtypes and what are the main evolutionary changes explaining this diversity are fundamental questions in evolution and developmental biology. One such source of neuronal diversity is the olfactory placodes (OPs) of jawed vertebrates. During development, the OPs produce numerous neuronal cell types including olfactory sensory neurons (OSNs) and the neuroendocrine GnRH3 and GnRH1 neurons (cf. Chap1). The latter are most well-known for their fundamental role in reproduction as essential components of the hypothalamus-pituitary-gonadal axis. Despite the importance of GnRH cells in neurobiology many gaps remain in our understanding of their ontogeny and regulation (Duan and Allard, 2020). In fact, despite more than 30 years from the identification of the olfactory placodes as a source of forebrain GnRH cells, a pan-vertebrate vision of the cell lineage origin and specification of GnRH neurons remains unclear (Schwanzel-Fukuda and Pfaff, 1989; Wray et al., 1989). In particular, it is ambiguous whether GnRH neurons are also nasal placode derivatives in more basal vertebrates such as Chondrichthyes and cyclostomes, or are just born within the central nervous system. It is also unclear if GnRH

neurons in those 'primitive' organisms are specified through similar molecular mechanisms as observed in bony vertebrates (Poncelet and Shimeld, 2020). In this regard, lampreys are of particular interest for evolutionary studies because their basal phylogenetic position makes them key to reveal the molecular basis of the important morphological and physiological innovations that characterise vertebrates.

2.1.2 The vertebrates: characteristics, evolution, and development

The vertebrates represent the overwhelming majority of the phylum Chordata with an estimated ~70,000 extant species (IUCN, 2020). The evolutionary success of vertebrates can be explained in part by the appearance of key distinctive features such as neural crest, placodes, brain segmentation, an adaptive immune system, and an endoskeleton. The appearance of early vertebrates like *Myllokunmingia* and *Haikouichtys* can be traced back to the Cambrian, a period where all major animal phyla start to appear in the fossil record (Shu et al., 1996). All these early vertebrates lacked a proper jaw and were probably filter-feeders. Vertebrates can be divided into two groups: gnathostomes (jawed vertebrates) and agnathans (jawless vertebrates). Gnathostomes are characterised by the presence of hinged jaws, paired appendages, and a mineralized skeleton. They diverged from agnathans 430 to 520 million years ago. Today, they represent more than 99% of the living vertebrates, demonstrating the advantage the jaw has been as an evolutionary innovation (Ziermann, 2019).

Agnathans are jawless vertebrates occupying an important position at the base of vertebrate evolution (Shimeld and Donoghue, 2012). Today, they are represented by the surviving lamprey (Petromyzontiformes) and hagfish (Myxiniformes) species which

together form a monophyletic group known as the cyclostomes (Heimberg et al., 2010). Although previously debated, the cyclostome monophyly is supported by some morphological data and numerous molecular studies based on coding genes, microRNAs, and ribosomal RNA (Delarbre et al., 2002; Takezaki et al., 2003; Mallatt and Winchell, 2007; Heimberg et al., 2010; Kuraku, 2013; Miyashita et al., 2019). All cyclostomes possess a cranium, retain an adult notochord, and have cartilaginous elements homologous to vertebrae (Ota et al., 2011). In addition, they have a distinct adaptive immunity to gnathostomes, i.e., the variable lymphocyte receptor system (Rast and Buckley, 2013). However, one of the most prominent differences between gnathostomes and cyclostomes is not necessarily the existence or lack of a jaw but the presence of a single median nostril (monorhiny) in lampreys and hagfishes (Haeckel, 1874). In fact, the counterpart to the gnathostomes tripartite system with independent paired olfactory and adenohypophysis placodes is a single medial nasohypophysial placode (NHP) in cyclostomes (Gorbman and Tamarin, 1985). The NHP anterior part differentiates into the olfactory epithelium, while the posterior part of this placode forms the adenohypophysis. In consequence, the development of the olfactory system in cyclostomes is closely linked to the adenohypophysis (Uchida et al., 2003).

In the cyclostomes, there are 38 species of lampreys described so far (Gee, 2018). Lampreys are filter-feeders inhabiting freshwater rivers at the larval stage, known as an “ammocoete” (Gee, 2018). After finding a suitable settlement site, the ammocoetes burrow themselves in and feed on filtered organic matter for usually three to seven years (Applegate, 1950; Dawson et al., 2015). This filter-feeding lifestyle is considered to be a conserved protochordate character (Gans and Northcutt, 1983). The ammocoete has indeed a similar lifestyle to amphioxus, living burrowed in the sediment with their front

end exposed to the water for particle filtration. After metamorphosis to their adult form, lampreys acquire a circular mouth with concentric rows of teeth. Adult lampreys can be parasitic or nonparasitic depending on the species. The parasitic lampreys attach themselves into the body of passing teleost fish and feed on their host tissues and fluids. The nonparasitic lampreys do not feed during their post-larval life and have a smaller adult size (Potter, 1980). During reproduction, sexually mature lampreys migrate back to the river streams from which they were born to spawn and lay eggs. On the other hand, the hagfishes are opportunistic deep-sea scavenger feeding on dead prey lying on the seabed. The hagfish feeding apparatus contains teeth used to rip flesh off carcasses, but this is not a proper jaw. Overall, hagfishes have been far less studied than lampreys mainly because it is harder to collect and gather large quantity of embryos for developmental studies (Shimeld and Donoghue, 2012).

It is assumed that all living vertebrates evolved from an amphioxus-like ancestor by developing a more complex head which fostered the transition from filter feeding to active predation in ancestral vertebrates. For this transition to happen it is argued that the genome of vertebrates has distinguished itself from invertebrate chordates, including amphioxus and tunicates, by two rounds of whole genome duplication (2R hypothesis) (Ohno, 1970). In fact, numerous gene families such as those encoding transcription factors and signalling molecules were obviously expanded by gene duplication in the vertebrate stem lineage. The increased numbers of gene copies are one of the major events underlying increased morphological complexity under developmental control (Satoh et al., 2014). New gene paralogues can diversify from their original copy by specialisation, neo-functionalisation, or sub-functionalisation. In particular, evolution of developmental transcription factor paralogs may allow the appearance of new cell types

and the evolution of new structures via rewiring of existing gene regulatory networks. Following 2R, subsequent gene gains and losses differentially shaped the genomes of the diverse vertebrate classes leading to the existing paralog counts (Cañestro et al., 2013). It has to be mentioned that there is also a third event of genome duplication (3R) in the ray finned fishes (Actinopterygii, specifically in teleosts) and even a fourth one (4R) in salmonids (Meyer and Van De Peer, 2005; Lien et al., 2016; Robertson et al., 2017). The number, mechanism, and precise timing of the 2R is still elusive and subject to intensive debate. The most recent analysis looking at conserved syntenies in chordates at near chromosome-scale concluded that lamprey and gnathostomes definitely share one round of whole-genome duplication (1R) but that the second round of whole-genome duplication (2R) is not shared (Simakov et al., 2020). This might explain why it is so difficult to assign a clear one-to-one orthology between cyclostome and gnathostome paralogs. In fact, molecular phylogenetic analyses of gene families usually cluster cyclostome paralogs together and usually fail to group them further with specific gnathostome paralogs (Kuraku et al., 2013; Smith and Keinath, 2015).

2.1.3 Chapter aims and experimental approach

The aim of this chapter was to gain insight into the embryonic origin and mechanism of GnRH specification in lampreys and investigate their possible association with the olfactory system during development. In simple terms, two main research questions are addressed here: (1) Do the GnRH neurons of lamprey develop through similar signalling pathways and TFs as seen in jawed vertebrates? (2) Are the GnRH neuron progenitors of lamprey associated with the NHP during development?

In this chapter, I investigated the timing and location of GnRH neurons differentiation in lamprey. To do so, three lines of research were undertaken. First, I characterised the embryonic expression of a carefully selected set of transcription factors and signalling pathway genes known to underlie cytodifferentiation of olfactory-derived GnRH cells in jawed vertebrates. Secondly, I examined by pharmacological manipulation if FGF and RA signalling levels impact GnRH embryogenesis in lamprey, as it does in jawed vertebrates. Finally, a cell lineage tracing experiment was performed on the nasohypophyseal placode (NHP) to track possible cell migration of GnRH neurons from that region.

2.1.4 Marker selection

The molecular mechanism that gives rise to the GnRH neurons remain largely unresolved. Nevertheless, certain TFs and signalling pathways are known to be important for their development. Taking advantage of published data, I carried out a careful analysis of the literature to select marker genes with established roles in the specification and/or migration of GnRH neurons form the olfactory placodes. A set of transcription factors and signalling genes were chosen for their ability to regulate downstream genes and infer potential regulatory networks. In the embryonic lamprey, I examined the spatial relationship of GnRH-containing neurons to regions of the brain expressing ISLET, FGF8, COE and NELF genes. My intention was to use the expression of those genes as markers to identify the timing and location of GnRH neuron development in lamprey and see if any can serve as molecular guideposts for olfactory placode GnRH precursors at different stages. A short description for each of these genes follows to highlight why orthologues of these gene families were chosen to interrogate GnRH neuron origin(s) and specification in lamprey.

2.1.5 GnRH (Gonadotropin Releasing Hormone)

The isolation and structural elucidation of the GnRH decapeptide produced in the hypothalamus became a Nobel prize discovery for medicine or physiology in 1977 (Matsuo et al., 1971; Burgus et al., 1972). Since then, GnRH-producing neurons have kept fascinating scientists and remain a trending topic in neurodevelopmental research. In vertebrates, there are three different groups of paralogous GnRH genes and populations of GnRH neurons: GnRH1, GnRH2 and GnRH3. The consensus is that GnRH1 and GnRH3 (GnRH1/3) neurons derive from the olfactory placodes although their exact embryonic origin is still controversial and debated (Duan and Allard, 2020). The mature locations of GnRH1 neurons are the preoptic area and hypothalamus where they have an hypophysiotropic role controlling pituitary-gonadal function (Zhao et al., 2013). The GnRH3 neurons are present in the terminal nerve and are neuromodulatory without direct action on reproduction (Takahasi et al., 2016; Umatani and Oka, 2019). Teleost GnRH3 neurons are also present in the thalamus and trigeminal nerve (Okubo et al., 2006; Abraham et al., 2008). Note that if the GnRH1 or GnRH3 gene is missing (as occurs in some species), the remaining GnRH takes on the localisation and function of the lost paralogue. The GnRH2 neurons on the other hand are said to be of non-placodal origin and are thought to derive from the neural plate instead (Northcutt and Muske, 1994). The GnRH2 cells have the most conserved expression in the midbrain tegmentum, but are also found in the hindbrain, spinal cord and extra-hypothalamic regions (Xia et al., 2014; Kusakabe et al., 2012; Kaufmann, 2004). Not surprisingly, GnRH peptides or mRNAs are the best markers to determine GnRH cell identity. However, the reliance on GnRH for identification of those cells is problematic for the study of their embryonic origins. In fact, progenitor cells cannot be easily identified as they are no known markers prior to GnRH

expression. The lack of specific early marker genes for GnRH progenitors is one of the reasons why the field is so divisive about the exact embryonic origins of vertebrate GnRH neurons. So far, the accumulation of data on the origin of forebrain GnRH1/3 neurons in different jawed vertebrate species resulted in three competing hypotheses: (1) GnRH1/3 neurons are derived from PPE-derived precursors cells residing in the olfactory placodes (Murakami et al., 1992; Dellovade et al., 1998; Sabado et al., 2012; Aguillon et al., 2018). (2) GnRH1/3 neurons are derived from both PPE-derived and NC-derived precursors cells present in the olfactory placodes (Forni et al., 2011; Forni and Wray, 2012; Shan et al., 2020). (3) GnRH1/3 neurons are derived from NC-derived precursors cells and PPE-derived precursors cells residing in the adenohypophyseal placode then transiently associate with the olfactory system while migrating to the forebrain (Whitlock et al., 2003; Whitlock, 2005). In conclusion, it seems that precursor cells of the "new-born" GNRH1/3 neurons could have some other point of origin before they are localised in the olfactory placodes. The lack of consensus might just reflect actual differences between species but regardless of their exact origins, the different types of GnRH neurons have conservation in their final brain localisation and function which support their homology across different vertebrates.

2.1.6 ISL1 (ISL LIM Homeobox 1)

Islet genes encode for LIM-homeodomain transcription factors implicated in many developmental processes. They are considered as early differentiation markers of many embryonic cells (Ericson et al., 1992). The transcription factor Isl1 have recently gained much attention for the study of GnRH origin as they localise in subsets of neurons within the OPs and the nasal migratory mass (Taroc et al., 2020). In fact, forebrain GnRH1/3 cells

in zebrafish, chick, mouse and human co-localise perfectly with Isl1/2 immunoreactivity in the olfactory placodes and migratory neurons before their mature localisation to the terminal nerve and/or hypothalamus/preoptic area (Aguillon et al., 2018; Palaniappan et al., 2019; Shan et al., 2020; Taroc et al., 2020; Lund et al., 2020). In consequence, Isl1 is one of the best candidates to identify placodal derived neurons in the developing nasal area and may provide an earlier marker than GnRH itself.

2.1.7 NELF (Nasal Embryonic Luteinizing Hormone-Releasing Factor)

NELF (also known as NSMF or Jacob) is a nuclear protein first identified in migrating GnRH neurons from the OPs (Kramer and Wray, 2000; 2001). Due to its nuclear localisation, it was thought to be a transcription factor. However, NELF has no DNA binding motifs and is thought to have a role in transcriptional regulation instead by being part of a multimeric transcriptional complex (Ko et al., 2018). In support of this hypothesis, is the finding that NELF has been found in interaction with RNA II polymerase and DNA (Dieterich et al., 2008). In addition, NELF is known to regulate downstream genes of the JAK/STAT signalling pathway as well as cell migration proteins and proteins involved in transcription, all of which have essential function for the development and migration of GnRH neurons from the olfactory placodes (Ko et al., 2018). NELF was identified as a protein which translocate from synapse to the nucleus with NMDA receptor activation then interacts with the cAMP-response element-binding protein (CREB) to regulate transcription (Karpova et al., 2013; Spilker et al., 2016). Knockdown experiments of *NELF* reduced the number of GnRH neurons and affected complexity and length of olfactory nerve fibres (Kramer and Wray, 2000). NELF silencing in zebrafish and mouse affect GnRH1 migration to the forebrain (Palevitch et al., 2009; Xu et al., 2010).

In fact, mutations in the *NELF* gene have been identified in human with normosmic idiopathic hypogonadotropic hypogonadism (nIHH) and Kallmann syndrome, in other words deficient for GnRH1 neurons setting up the HPG axis and thus with an impaired pubertal development and associated subfertility (Miura et al., 2004; Quaynor et al., 2015).

2.1.8 FGF8 (Fibroblast Growth Factor 8)

FGF8 is a member of the fibroblast growth factor (FGF) family which have various functions during vertebrate development such as embryonic patterning and neuronal development (Schlosser, 2005). During the formation of placodes, FGF8 originates from the anterior neural ridge and promotes the formation of olfactory placodal fate over lens placodal fate (Bailey et al., 2006). In mouse, *FGF8* is expressed in the respiratory epithelium adjacent to the OP and define a morphogenetic centre necessary for olfactory neurogenesis and nasal cavity development (Kawauchi et al., 2005; Forni et al., 2013). Loss or decrease in FGF8 signalling components results in disruption of the olfactory epithelium and also affects GnRH development (Falardeau et al., 2008; Chung and Tsai, 2010). In chick, FGF8 signalling is sufficient to induce the specification of GnRH1 neurons while retinoic acid antagonises their formation (Sabado et al., 2012). In human, mutations in FGF8 or its receptor FGFR1 have been detected in person suffering from Kallmann syndrome (Pitteloud et al., 2007). Furthermore, human stem cells can be differentiated into GnRH cells via dual SMAD inhibition and treatment with FGF8 and DAPT, a Notch inhibitor (Lund et al., 2016; 2020).

2.1.9 COE2 (Collier, Olf and EBF Transcription Factor 2)

COE2 (or Ebf2) is a helix-loop-helix transcription factor expressed in various region of the embryonic brain of vertebrates but also prominently expressed by olfactory placode neurons (Dubois and Vincent, 2001). Intriguingly, a COE gene in the nematode *Caenorhabditis elegans* is expressed by chemosensory neurons, leading to the suggestion that COE may primitively mark chemosensory neurons (Prasad et al., 1998). In mouse, the COE2 gene is expressed in GnRH neurons migrating from the future vomeronasal part of the olfactory placodes. Knockout mice for *COE2* retain GnRH neurons as a cluster in the nasal mesenchyme and are hypogonadic (Corradi et al., 2003). The effects seem directly linked to GnRH neurons as the development of the olfactory system was not impaired. As the expression of *COE* genes in lamprey embryos has been previously characterised (see Appendix), a complementary work to *in situ* hybridization was made in this chapter to understand the synteny relationship of COE genes and GnRH genes in chordates (Lara-Ramírez et al., 2017).

2.2 Material and Methods

2.2.1 Embryo collection and fixation

Lamprey embryos of *Lampetra planeri* were collected from a shallow river in the New Forest National Park, United Kingdom, with permission from Forestry England. Embryos were brought to the laboratory and placed in Petri dishes with filtered river water (FRW) from the same river where they were caught. They were kept at 15 °C and fixed at different stages of development following the Tahara staging system (Tahara, 1988). All experiments were performed under local ethical approval. When necessary, embryo chorions were removed with fine forceps before fixation. Embryos were fixed in 4% paraformaldehyde (PFA) pH 7.5 in phosphate-buffered saline (PBS). PFA was cooled on ice before use. Embryos were fixed in an approximately 10× excess volume of 4% PFA-PBS with respect to river water at 4 °C overnight or longer. After fixation, embryos were washed twice in diethyl pyrocarbonate (DEPC)-treated 1× PBS for 10 min each and then dehydrated through a graded series of PBS:methanol (25, 50 and 75% of methanol in 1× PBS) once for 10 min each. Finally, they were washed twice in 100% methanol for 10 min each and stored in fresh methanol at −20 °C.

2.2.2 Bioinformatic analysis

A reciprocal BLAST approach was used to identify orthologous genes in the lamprey *Lampetra planeri* with the longest isoform available of annotated *Homo sapiens* amino acid sequences as queries (NCBI RefSeq database). Multiple transcriptomic data were searched: *Lampetra planeri* transcriptome data (Lara- Ramírez et al., 2017), *Lethenteron camtschatium* transcriptome (PRJNA269952) and *Petromyzon marinus* transcriptome (PRJNA294488). The matching transcript sequences to the protein queries were matched

back to the NCBI RefSeq database to give additional support for their identity. The selected transcripts through this procedure were then mapped to the genome assembly of *Lethenteron camtschaticum* to identify individual gene loci (Mehta et al., 2013). The whole operation has been automated with python scripts developed by Dr Vasileios Pappadiogannis.

2.2.3 Molecular phylogenetic analysis

Alignments with jawed vertebrate paralogs and molecular phylogenies were built to test the orthology of identified family members. Selected sequences of major vertebrate groups were chosen for the following representatives : mammal (*Homo sapiens, Mus musculus, Rattus norvegicus, Phascolarctos cinereus, Ovis aries, Ailuropoda melanoleuca, Orcinus orca, Ornithorhynchus anatinus, Sarcophilus harrisii*); lepidosaur (*Anolis carolinensis, Eublepharis macularius, Zootoca vivipara, Gekko japonicus, Podarcis muralis, Python bivittatus*); archosaur (*Gallus gallus, Sturnus vulgaris, Parus major, Anas platyrhynchos, Meleagris gallopavo, Pelecanus crispus, Falco peregrinus, Aptenodytes forsteri, Alligator mississippiensis, Alligator sinensis)*; turtles (*Chrysemys picta bellii, Pelodiscus sinensis, Chelonia mydas*); amphibian (*Xenopus tropicalis, Rana catesbelana, Rana temporaria, Bufo bufo)*; sarcopterygian fish (*Latimeria chalumnae*); actinopterygian (*Lepisosteus oculatus, Acipenser sinensis*); teleost (*Anguilla anguilla*, *Orizias latipes, Danio rerio, Scomber japonicus, Oerochromis niloticus, Cichiasoma dimerus, Dicentrarchus labrax, Odontesthes bonariensis, Engraulis japonicus, Carassius auratus, Scleropages jardinii, Coregonus clupeaformis, Salmo salar, Cyprinus carpio*); elasmobranch (*Rhincodon typus, Scyllorhinus canicula, Callorhinchus milli, Amblyraja radiata*); lamprey (*Petromyzon marinus, Lethenteron camtschaticum, Lampetra planeri)*. Sequences of amphioxus

Branchiostoma floridae were used as an outgroup. Accession numbers for sequences used for phylogenetic analysis are shown in Appendix. Multiple sequence alignments were performed using MAAFT v7 (Katoh et al., 2019) and trimmed with TRIMAL using gappyout (Capella-Gutierrez et al., 2009). For phylogenetic tree construction, maximum likelihood (ML) was conducted using RAxML (GUI, v.2.0.3). For ML, the JTT +G amino acid substitution matrix was used and 1000 bootstrap replicates to obtain support values at each node.

2.2.4 Gene probes cloning and synthesis

Lamprey gene fragments were cloned via PCR amplification from *L. planeri* cDNA made-from total RNA extracted from mixed embryonic stages Tahara 23 to 28. The PCR reactions used specific primers (see Table 2.1) and the amplified products were cloned in the pCR™-II vector using TA Cloning™. The genes *LpISLA* and *LpISLC* were cloned by gene synthesis (Twist Bioscience) in the vector pTwist Amp High Copy. Probe synthesis was performed as in (Lara-Ramirez, 2015). An extra step of probe hydrolysis was performed for the *LpNELF* and *LpGNRH1* probes to make fragment size of 250-500 bp as described in the protocol (Ferrandiz and Sessions, 2008). All clones were verified by Sanger sequencing. The complete plasmid constructs are available as electronic .dna files (see Appendix).

Table 2.1. PCR primers used for lamprey gene cloning.

Gene	Primers	Probe Size
LpGNRH-I	5'-AGCACTACTCCCTGGAATGGA-3' 5'-ATGCACAACACCTCCTTGTTA-3'	532 bp
LpGNRH-II	5'-CAAACGGAGCGTTCAGGAG-3' 5'-CGGGAGTGACGGATCATGTTTA-3'	454 bp
LpGNRH-III	5'-CAAACGGAGCGTTCAGGAG-3' 5'-TGGCCTGTTCGTGACCAATAA-3'	664 bp
LpISLB	5'-GACCCGGATCGAACTCCAC-3' 5'-CATTCTTGCACTTTCCCCAGC-3'	701 bp
LpNELF	5'-AATTAGTAAGATTTGTCACGTA-3' 5'-AGGCAAGAAGATGATAGAAAC-3'	2396 bp
LpFGF8/17	5'-AGGATTAACTACCCCGCTGT-3' 5'-CCTGGATGAACTTGCGCTC-3'	638 bp

2.2.5 *in situ* hybridisation

Embryos were first treated with a mixture of hydrogen peroxide and methanol in a 1:1.5 ratio overnight, and later washed three times with PBS-DEPC for 5 min each at room temperature (RT). Embryos were then treated with 0.2 N hydrochloric acid in PBS-DEPC for 10 min static, followed by a treatment with 1 $\mu g ml^{-1}$ Proteinase K (Roche) in PBS-DEPC, static, at 37°C for 18 min for all stages. Samples were post-fixed with 4% PFA, 0.2% glutaraldehyde in PBS-DEPC for 20 min on ice, then washed three times with PBS-DEPC for 10 min each at RT, and twice in hybridization buffer (50% formamide, 5x standard saline citrate [SSC] pH 7.0, 0.1% Tween 20, 0.1% CHAPS, 1 $mg ml^{-1}$ total yeast RNA, 100 $\mu g ml^{-1}$ heparin in DEPC-H_2O) 10 min each at 65°C. Embryos were then prehybridized overnight at 65°C in hybridization buffer and later incubated with 600–700 $ng ml^{-1}$ of digoxigenin (DIG)-labelled sense or antisense riboprobe in prewarmed hybridization buffer at 65°C for 48–60 hr. After hybridization, embryos were washed twice with wash A (50% Formamide, 5x SSC pH 4.5, 1% SDS in DEPC-H_2O) for 30 min each at 65°C, then substituted gradually with 50% and 75% TST (10 mM Tris-HCl pH 7.5, 0.5 M NaCl, 0.1% Tween 20) in wash A for 10 min each, and twice with 100% TST for 10 min each at 65°C. RNase A was then added at 0.05 $mg ml^{-1}$ concentration in TST for 30 min at RT, followed by washes with wash B (50% Formamide, 2x SSC pH 4.5), wash C (2x SSC pH 4.5, 0.3% CHAPS), and wash D (0.2x SSC pH 4.5, 0.3% CHAPS) at 65°C each twice for 30 min. For immunological detection, samples were blocked overnight at 4°C with sheep serum previously heat-treated at 65°C for 30 min then combined with TST in a 4:1 ratio, followed by incubation with anti-DIG-AP Fab fragments antibody (1:3,000 dilution, Roche) overnight at 4°C. Embryos were later washed with TST five times for 60 min each at RT. Embryos were stained with 6 $\mu l ml^{-1}$ of NBT/BCIP (Roche) stock solution in NTMT

buffer (100 mM NaCl, 100 mM Tris-HCl pH 9.5, 50 mM $MgCl_2$, 1% Tween 20). Embryos were left in NBT/BCIP for up to 3 days. Once specific staining had developed, embryos were washed with PBT (1x PBS-DEPC, 0.1% Tween 20) three times for 10 min each at RT, then fixed overnight with 4% PFA in PBS-DEPC at 4°C and washed again with PBT three times 10 min each. Finally, embryos were cleared with a graded series of methanol 25%, 50%, 75% in PBS-DEPC, and 100% at RT. Then they were transparentised and stored in a BABB solution (a 1:2 mixture of benzyl alcohol and benzyl benzoate) at 4°C. To observe the embryos under the microscope, they were first mounted on a glass microscope slide immersed in BABB and observed with a Carl Zeiss Axioskop 2 plus microscope. Pictures were taken with a Zeiss AxioCam HRc camera (for a detailed protocol see thesis of Lara-Ramirez, 2015). Images of x 40 Focus were photomerged and focus stacked using Photoshop CS6 to do a collage from multiple smaller images of the same embryo, with the plane of focus adjusted between shots to keep the labelled structures in focus.

2.2.6 Synteny analysis

COE genomic loci were identified in the genomes of *B. floridae* (amphioxus), *L. camtschaticum* , *C. milii* (elephant shark) and *H. sapiens* by BLAST. We chose *H. sapiens* as an extensively annotated vertebrate genome and *C. milii* as a member of the earliest diverging jawed vertebrate lineage; together, they encompass extant jawed vertebrate diversity. To map synteny and paralogy relationships, genes adjacent to the COE loci in each species were searched by TBLASTN of their predicted proteins across the other genomes. The top chromosomal or scaffold hits were recorded to predict orthology. Human paralogues were extracted from Ensembl predictions of paralogy.

2.2.7 Drug treatments of FGF and RA signalling pathways

To inhibit FGF signalling, lamprey embryos were treated with 100 μM of the FGFR antagonist SU5402 (Sigma, SML0443). To inhibit RA signalling, the embryos were treated with 0.1 μM of RA inverse agonist BMS493 (Sigma, B6688) and to enhance RA signalling, the embryos were treated with 0.05 μM All-Trans-RA agonist (Sigma, R2625). Stocks of both drugs were made in undiluted DMSO (see Table 2.2). The drug treatments were performed in FRW. Controls were performed by incubating sibling larvae in FRW with DMSO added to the concentration of the experimental drug treatments. Embryos were treated in groups of 25 in 4-ml Petri dishes. Specific time windows were used for each drug (see Fig. 2.17). Dead larvae, if any, were removed and solutions were changed every 48h. After the appropriate treatment time was over the embryos were washed three times in 4ml of FRW to wash out the drugs so it does not interfere with further development. The embryos were let to develop until Tahara stage 29 and were then fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 1 h at room temperature for subsequent *in situ* hybridisation.

Table 2.2.

Pathway/activity	Drug name	Stock solution	Working Concentration
FGFR antagonist	SU5402	10mg/ml in DMSO	100 μM
RA agonist	ALL-TRANS-RA	20mg/ml in DMSO	0.05 μM
RA inverse agonist	BMS493	20mg/ml in DMSO	0.1 μM

2.2.8 CM-DiI dye labelling

Lamprey embryos at Tahara stage 23 were manually dechorionated with fine forceps. The embryos were placed in agarose-coated Petri dishes. In the agarose coating a narrowing channel was dug with a razor blade and the embryos were carefully immobilised and oriented in that trench with the nasohypophyseal placode (NHP) facing upward for dye injection (Fig. 2.1). Embryos were pressure-injected (Femtojet 4i, Eppendorf) using glass capillary tubes (Femtotips II, Eppendorf) filled with 0.5 $mgml^{-1}$ of CellTracker™ CM-DiI Dye (Invitrogen) diluted in 0.3 M sucrose (from a 5 $mgml^{-1}$ stock diluted in ethanol). The filled capillary tube was place on top of the NHP while making sure not to go through the surface ectoderm. After dye injection, the embryos were individually place to an uncoated 96 well plate and allowed to recover and develop at 18°C. Embryos were periodically checked and imaged during development with a ZOE™ Fluorescent Cell Imager (Bio-Rad). For imaging, the embryos were transferred in a separate petri dish and anesthetised with MS-222 (0.1%) to avoid embryo movement artefacts. The embryos were imaged on both lateral sides. After each imaging session, the embryos were removed from the anaesthetic solution and placed back into a well with FRW. The developing labelled embryos were then permanently fixed at Tahara stage 29 with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 1 h at room temperature and snap frozen in OCT media (Tissue-Tek®) if further sectioning was required.

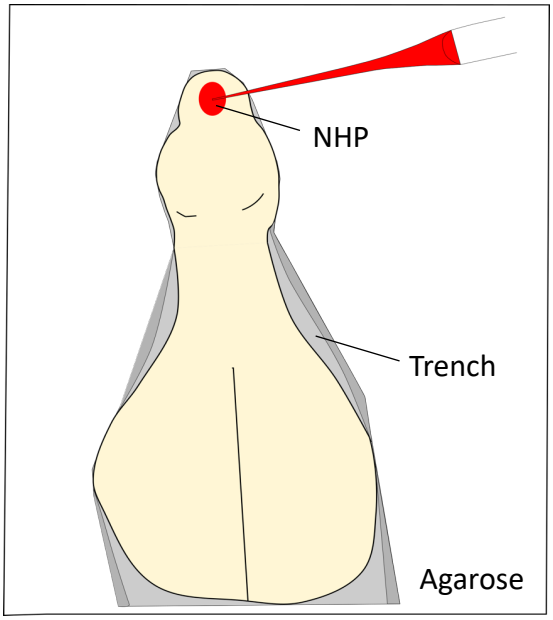


Figure 2.1. Schematic drawing of stage T23 embryo placed in a trench dug with a razor blade in an agarose plate, with the snout facing upwards for an optimal access to inject the NHP with CM-Dil dye (red) through a thin capillary tube.

2.2.9 Immunostaining of lamprey larva

Embryonic lamprey at T29 were fixed overnight with 4% paraformaldehyde (PFA) in 1× phosphate buffered saline (PBS) overnight, washed three times with PBS, then gradually dehydrated to 100% methanol and stored at −20 °C. After stepwise rehydration to 100% PBS, samples were permeabilized with 1% (w/v) Triton X-100 in PBS (PBS-T). Non-specific antibody binding was blocked with 10% (w/v) heat treated sheep serum (HTSS) in PBS-T for >1 h at room temperature and samples were then incubated 24-48 hours with primary antibodies in HTSS-PBS-T at 4 °C. Primary antibodies were: rabbit-anti-GnRH-I polyclonal (diluted 1/200) (King et al., 1988), rabbit-anti-GnRH-II polyclonal (diluted 1/40) (Kavanaugh et al., 2008), rabbit-anti-GnRH-III polyclonal (diluted 1/200) (Nozaki et al., 2000). After five washing-steps with PBS-T at room temperature, they were

incubated with secondary antibody, a goat anti-rabbit HRP conjugate diluted 1:500 in BSA-PBS-T. After 5 washes in PBS-T, a tyramide signal amplification (TSA) was performed by addition of tyramide solution made of TSA buffer (2M NaCl, 100mM Borate buffer pH 8.5), 0.5% H_2O_2, 0.1% 4-Iodophenylboronic acid (4IBPA) and 0.2 % 5-TAMRA fluorescent dye (Abcam, ab145438) or AF488 (ThermoFisher, B40953) for 30 min. Embryos were washed 5 times in PBS-T for 30 min and once overnight. Samples were washed with PBS several times, mounted with 50% (w/v) Glycerol in PBS on a coverglass-bottom petri-dish. The AF488 and 5-TAMRA fluorescence were excited at 488 and 559 nm, respectively, and recorded with a ZOE™ Fluorescent Cell Imager (Bio-Rad). DAB staining was made by using the HRP substrate, 3,3'-diaminobenzidine (0.25 $mgml^{-1}$) in TBST with 0.01% H_2O_2. The colour reaction was stopped by 4% PFA in PBST and the embryos were then gradually dehydrated in EtOH to be transparentized in BABB (a 1:2 mixture of benzyl alcohol and benzyl benzoate). The embryos were then imaged with a Zeiss AxioCam HRc camera.

2.3 Results

The gene expression data presented here are all the stages from which expression was detected in brook lamprey embryos, but expression was assessed from T23 to T28 for all genes via *in situ* hybridisation (see details in material and methods). The list of the identified loci for each marker gene is available in the Table 2.S1 in Appendix and contains the loci positions in the Japanese lamprey genome (*L. camtschaticum)* as well as the set of mapped lamprey transcripts whose sequences are available in the Appendix.

2.3.1 Identification of GnRH genes in *Lampetra planeri* and phylogenetic analysis

There have been three *GnRH* paralogues described so far in the sea lamprey (*Petromyzon marinus*), known as *GnRH-I*, *GnRH-II* and *GnRH-III* (Kavanaugh et al., 2008; Silver et al., 2004; Suzuki et al., 2000). Hereafter, I also confirm the presence of three *GnRH* genes in the brook lamprey, *Lampetra planeri*: *LpGnRH-I*, *LpGnRH-II* and *LpGnRH-III*. Those genes are located on two different loci with *GnRH-I* and *GnRH-III* on the same locus next to one another suggestive of a tandem duplication event. *L. planeri* GnRH transcripts were only detected for *LpGnRH-II* in the available transcriptome data, however gene amplification using other lamprey species (*P. marinus* and *L. camtschaticum)* transcripts for reference revealed the existence of three paralogues in that species. Molecular phylogeny based on amino acid alignment of 80 chordate GnRH precursor sequences segregated gnathostome GnRH sequences into three paralogous clades of the GnRH1, GnRH2 and GnRH3 families (Fig. 2.2). The phylogenetic tree group lamprey GnRH-II sequences with member of the gnathostome GnRH2 family in agreement with previous proposed orthology relationships (Gwee et al., 2009; Decatur et al., 2013). On the other hand, lamprey GnRH-I and -III sequences while being possibly related to gnathostome GnRH1 and/or GnRH3 did not cluster with those clades but group together at the base of the GnRH2 clade. The

bootstrap support values for each GnRH clades were extremely low (20% for GnRH1, 15% for GnRH2 and 55% for GnRH3), therefore we cannot convincingly assume any orthology relationships between lamprey and gnathostomes GnRH sequences based solely on molecular phylogeny.

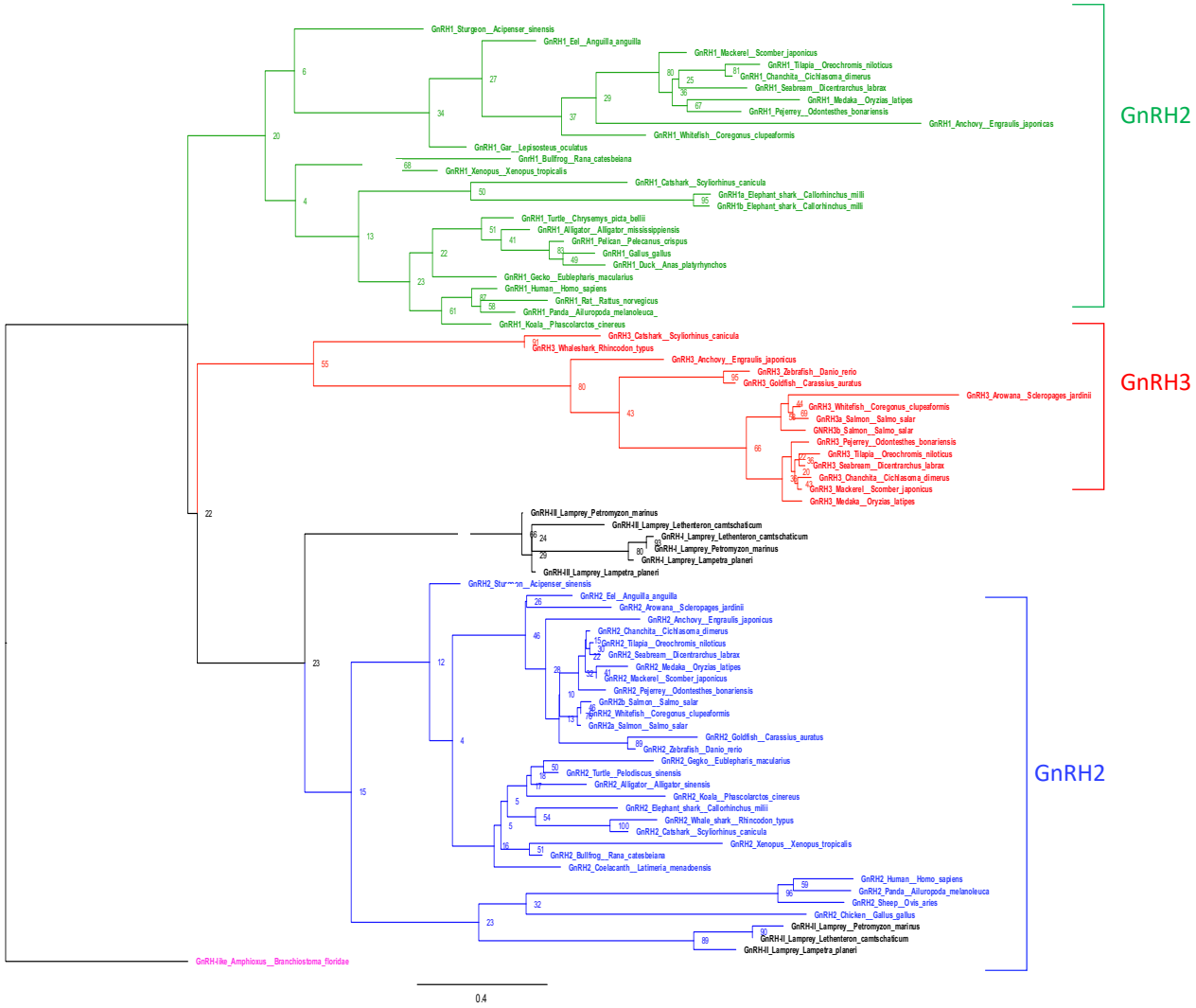


Figure 2.2. Phylogenetic tree of GnRH sequences in chordates. Phylogenetic analysis of 80 chordate GnRH amino acid sequences (79 sequences from vertebrate species plus the amphioxus prepro-GnRH sequence, used as outgroup) was constructed by Maximal Likelihood, with numbers below nodes are percentage bootstrap support, with 1000 replicates. Only values above 15% are indicated. Branches and names in the tree are in green if member of the GnRH1 family, in blue if member of the GnRH2 family or in red if member of the GnRH3 family. The lamprey sequences are in black. The tree is rooted with an Amphioxus GnRH sequence as an outgroup in magenta. Sequence references and alignment are given in Appendix.

2.3.2 GnRH gene family expression in the developing Lamprey, *Lampetra planeri*

While RNA expression of lamprey gonadotropin-releasing hormones (GnRH-I, -II, and -III) was described in the brains of larval, parasitic phase and adult sea lampreys (*P. marinus*) (Van Gulick et al., 2018), it has never been portrayed during embryogenesis, although it has been studied in developing sea lamprey for GnRH-I and -III via immunocytochemistry (Tobet et al., 1996). Here, I report the RNA expression of *LpGnRH-I*, *-II*, and *-III* in developing *L. planeri* embryos. No GnRH expression were detected at stage T23 to T26. *LpGnRH-II* was the first GnRH paralogue with expression detected in two discrete and specific bilateral spots in the hypothalamus at T27 (Fig. 2.3A-B). At T28, the expression of *LpGnRH-II* intensified to expand into two bilateral dense lines in the hypothalamus. In a lateral view the *LpGnRH-II* neurons generally appears as a single line (Fig. 2.3C), but if the embryos are tilted two bilateral lines become clearly visible in a lateral view (Fig. 2.3D). At T28, LpGnRH-III expressing cells first appeared in two bilateral distinct and specific dots in the presumptive preoptic area of the hypothalamus close to the diencephalic-telencephalic boundary (Fig. 2.4E). At T29, the number of LpGnRH-III positive cells increased and extended posteriorly by stage 30 (Fig. 2.4F-G). At stage 28 and 29, *LpGnRH-I* expression paralleled the development of LpGnRH-III cells, albeit at lower levels (Fig. 2.4H-I). The similitude of the expression territory of these genes suggests that *LpGnRH-III* and *-I* transcripts may co-localize in the same cells. Thus, while GnRH-I and GnRH-III cells are in the same location, they appear anteriorly in the hypothalamus compared to LpGnRH-II cells. It is worth mentioning that GnRH-expressing cells were never detected within the nasohypophyseal placode at any stage tested (T23-T30) by *in situ* hybridization even when letting the embryos overstain to detect possible low level of expression.

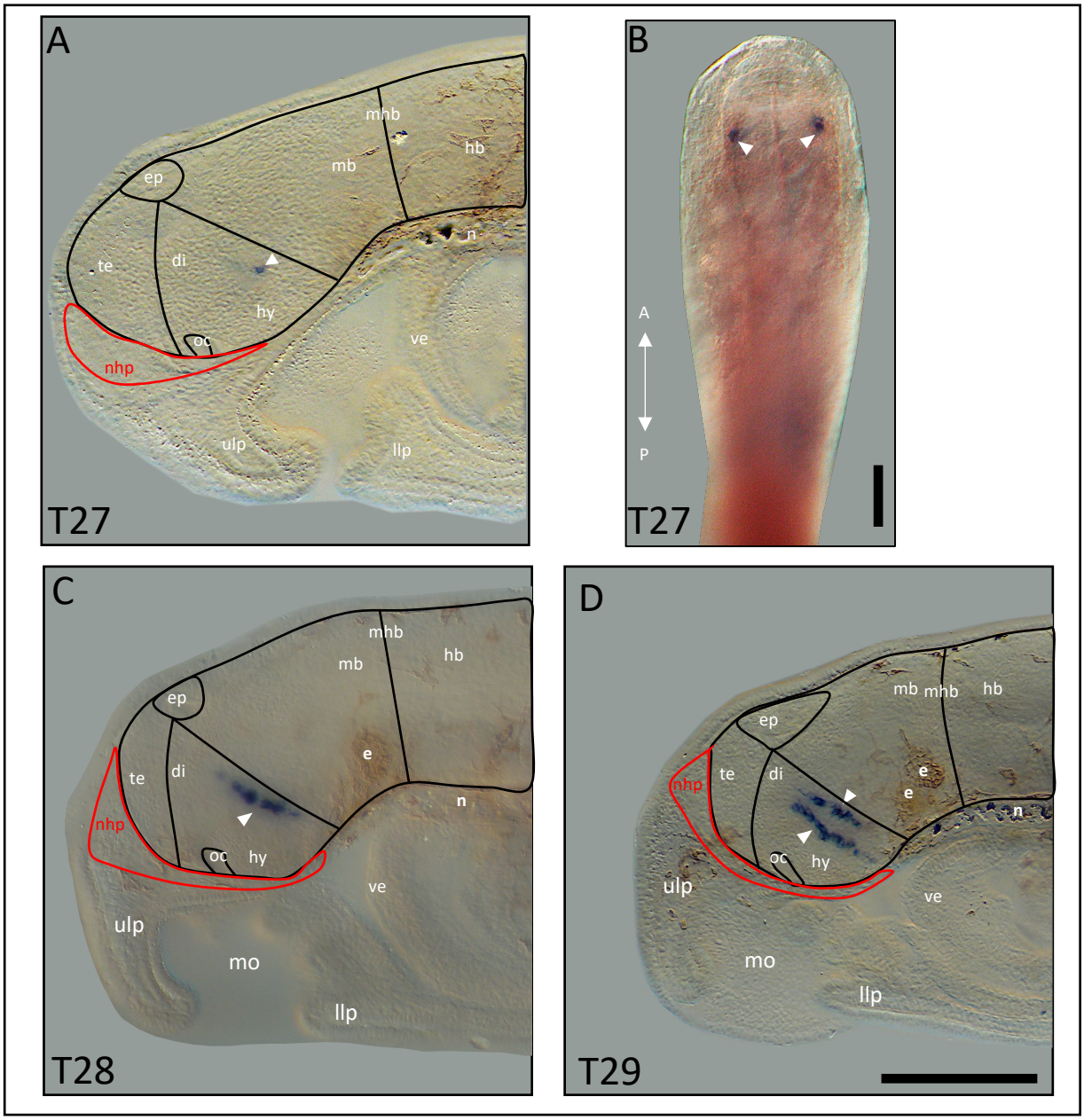


Figure 2.3. *LpGnRH-II* (A-D) expression during *L. planeri* development at stages 27-30. All images are lateral views, except for (B) a dorsal view. (A-B) At stage 27, *LpGnRH-II* expression is seen in two discrete and specific bilateral spots (white arrowheads) in the hypothalamus (hy). (C-D) At stage 28 and 29, *LpGnRH-II* expression intensified to expand into two bilateral dense lines in the hypothalamus (hy). In a lateral view the *LpGnRH-II* neurons generally appears as a single line (C) but if the embryos are tilted two bilateral lines become clearly visible (D). The black outline represents major brain subdivisions based on the lamprey prosomeric model established from neurogenin gene expression (Lara-Ramírez et al., 2015). The red outline represents the nasohypophyseal placode (nhp). Additional abbreviations: a, anterior; e, eye; ep, epiphysis; hb, hindbrain; llp, lower lips; mhb, midbrain-hindbrain boundary; n, notochord; oc, optic chiasma; p, posterior; te, telencephalon; ulp, upper lips; ve, velum. Scale bars: 200 µm

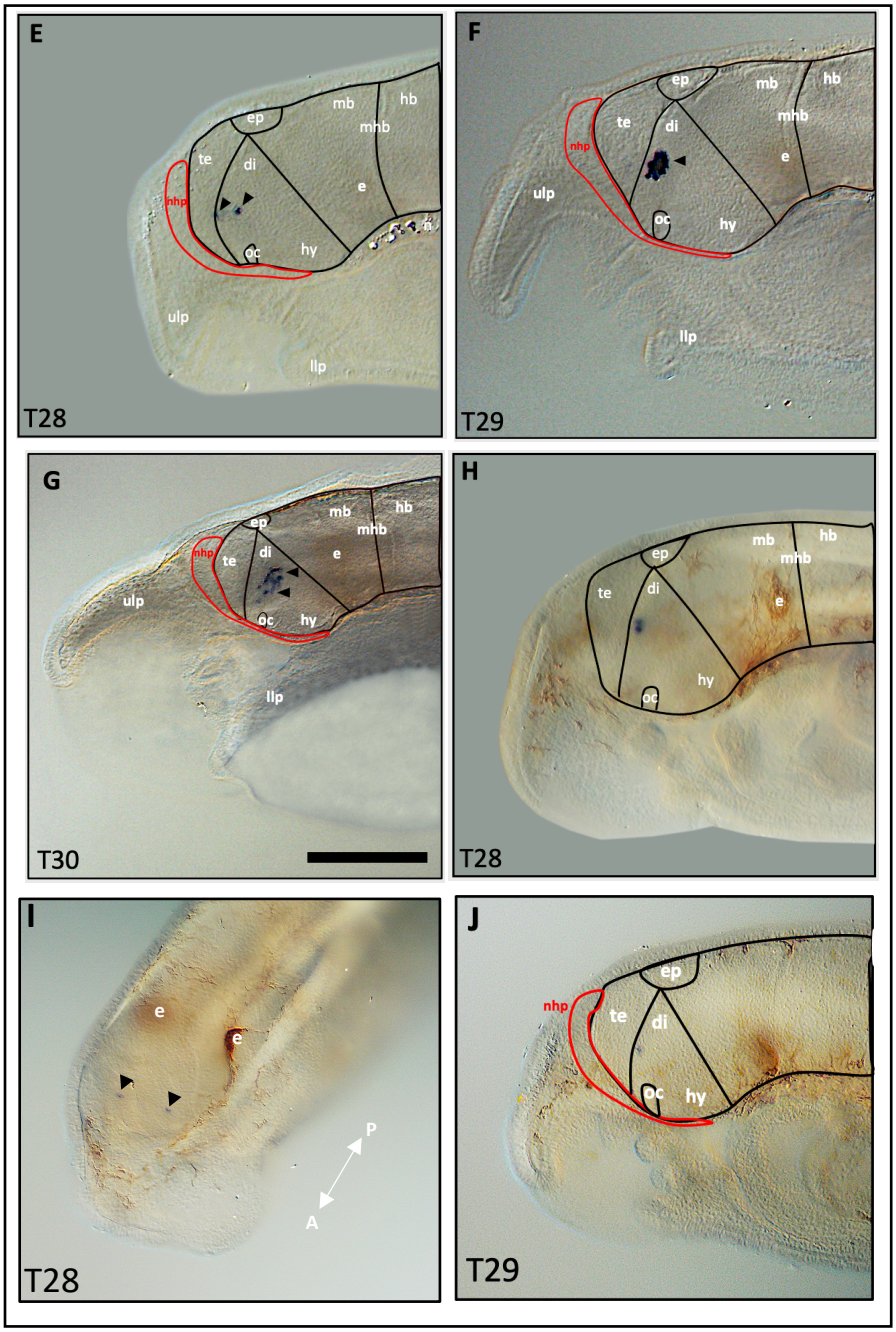


Figure 2.4. *LpGnRH-III* (E-G) and *LpGnRH-I* (H-J) expression during *L. planeri* development at stages 27-30. All images are lateral views, except for (I) a dorsal view. At stage 28, LpGnRH-III expressing cells first appeared in two bilateral distinct and specific dots (black arrowheads) in the presumptive preoptic area of the hypothalamus close to the diencephalic-telencephalic boundary. At stage 29 the number of LpGnRH-III positive cells dramatically increased and extended posteriorly by stage 30. At stage 28 and 29, the *LpGnRH-I* expression paralleled the development of LpGnRH-III cells, albeit at lower levels. The similitude of the expression territory of these genes suggests that *LpGnRH-III* and *-I* transcripts may co-localize in the same cells. The black outline represents major brain subdivisions based on the lamprey prosomeric model established from neurogenin gene expression (Lara-Ramírez et al., 2015). The red outline represents the nasohypophyseal placode (nhp). Additional abbreviations: a, anterior; e, eye; ep, epiphysis; hb, hindbrain; llp, lower lips; mhb, midbrain-hindbrain boundary; n, notochord; oc, optic chiasma; te, telencephalon; ulp, upper lips; p, posterior; ve, velum. Scale bar: 200 µm.

2.3.3 Synteny analysis of COE and GnRH gene loci in lamprey

Two COE genes in lamprey were identified in *L. planeri*, termed *LpCOEA* and *LpCOEB*. The expressions of those genes were thoroughly studied by *in situ* hybridization and published in Lara-Ramirez et al. (2017) (publication in Appendix). The main information to retain is that both genes are expressed in the NHP and more specifically in the anterior part that would give rise to the future olfactory epithelium at T23 to T28. Since those genes are important for GnRH development and because jawed vertebrates COE genes localise on the same loci as GnRH paralogues, I further examined the relationships between lamprey and other chordate *COE* loci. To do so, I examined the synteny surrounding *COE* loci of human, elephant shark, lamprey, and amphioxus. Jawed vertebrate *COE1–4* lie in paralogous regions of the genome (Fig. 2.5), with sufficient similarity in neighbouring genes to conclude the loci evolved by block duplication. For example, human *COE2* and *COE4* are linked to *GnRH* paralogues, while *COE1* and *COE3* are linked to *FoxI* paralogues. Moreover, *COE1–4* were predicted to form a family of whole genome duplication paralogues in a pre-computed whole genome assessment based on synteny (Singh et al. 2015). I found weak evidence for syntenic organisation of these regions with the amphioxus *COE* locus. Genes linked to amphioxus *COE* on scaffold 381 had orthologues on the same chromosome as human *COE* loci, though the genomic distance was relatively large. Both lamprey *COE* loci also showed evidence of similarity in organisation to jawed vertebrate *COE* loci. For example, both are linked to *GnRH* paralogues, lamprey *COE-B* and elephant shark *COE1* are linked to *CLNT1A*, and lamprey *COE-A* and human *COE4* are linked to *NOP56*. However, there is no clear one-to-one relationship between the lamprey loci and the four jawed vertebrate loci such that orthology can be deduced for COE genes or GnRH genes (Fig. 2.5). At the time of the

analysis, it was uncertain whether the GnRH-like gene in amphioxus was in the same locus as its COE gene, because only a scaffold level assembly was available. Recent chromosome level assembly (Marlétaz et al., 2018) confirmed that they are on the same chromosome, as in vertebrates. We hence conclude that both jawed vertebrate and lamprey COE and GnRH regions evolved by block duplications from a single ancestral COE-GnRH locus as seen in amphioxus, but we cannot determine whether these are shared duplications or occurred in parallel.

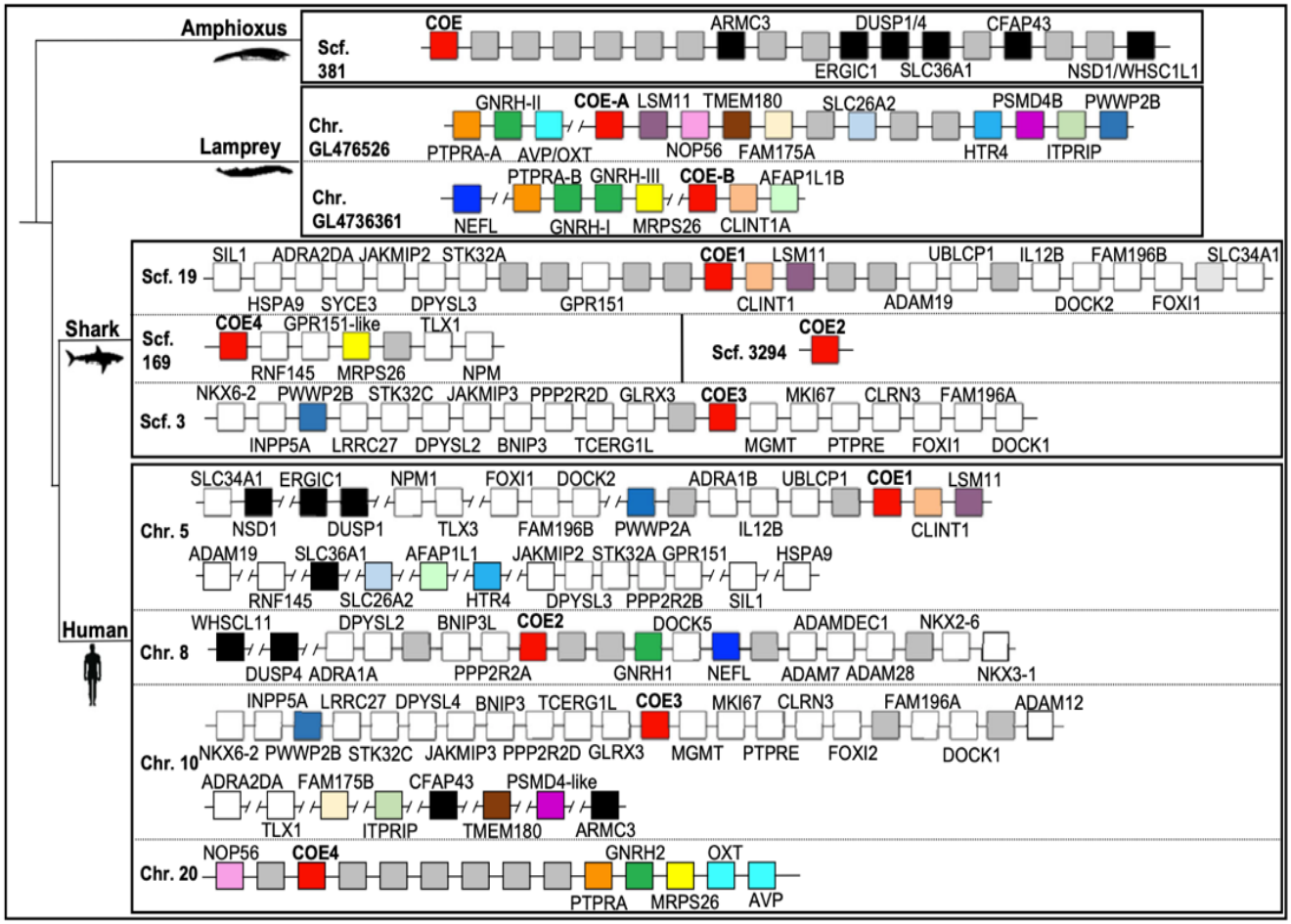


Figure 2.5. Schematic maps of COE locus paralogy and synteny in jawed vertebrate, lamprey, and amphioxus genomes. COE genes are in red, GnRH genes are in green, colour coding of other genes is as follows: Grey, genes with no orthologues or paralogues identified in the analysed regions. White, genes with syntenic orthologues and/or paralogues in human and shark. Black, genes linked to COE in amphioxus and their orthologue positions in other species. Other colours, genes linked to lamprey COE genes and their orthologues and/or paralogues in other species. Discontinuities shown as angled bars indicate where genes map to the same chromosome arm or scaffold but are separated by multiple intervening genes which are not shown. The figure is published in Lara-Ramirez et al., 2017 (see Appendix).

2.3.4 Identification of ISL genes in *Lampetra planeri* and phylogenetic analysis

I report three LIM-homeobox Islet paralogues in the Lamprey *Lampetra planeri,* named *LpISLA, LpISLB* and *LpISLC.* Amino acid alignment confirmed that ISL sequences of the lamprey species *L. planeri*, *P. marinus* and *L. camtschaticum* were similar with the presence of two protein-interacting LIM domains and a DNA-binding homeodomain characteristic of the LIM-homeodomain protein family (Fig. 2.6).

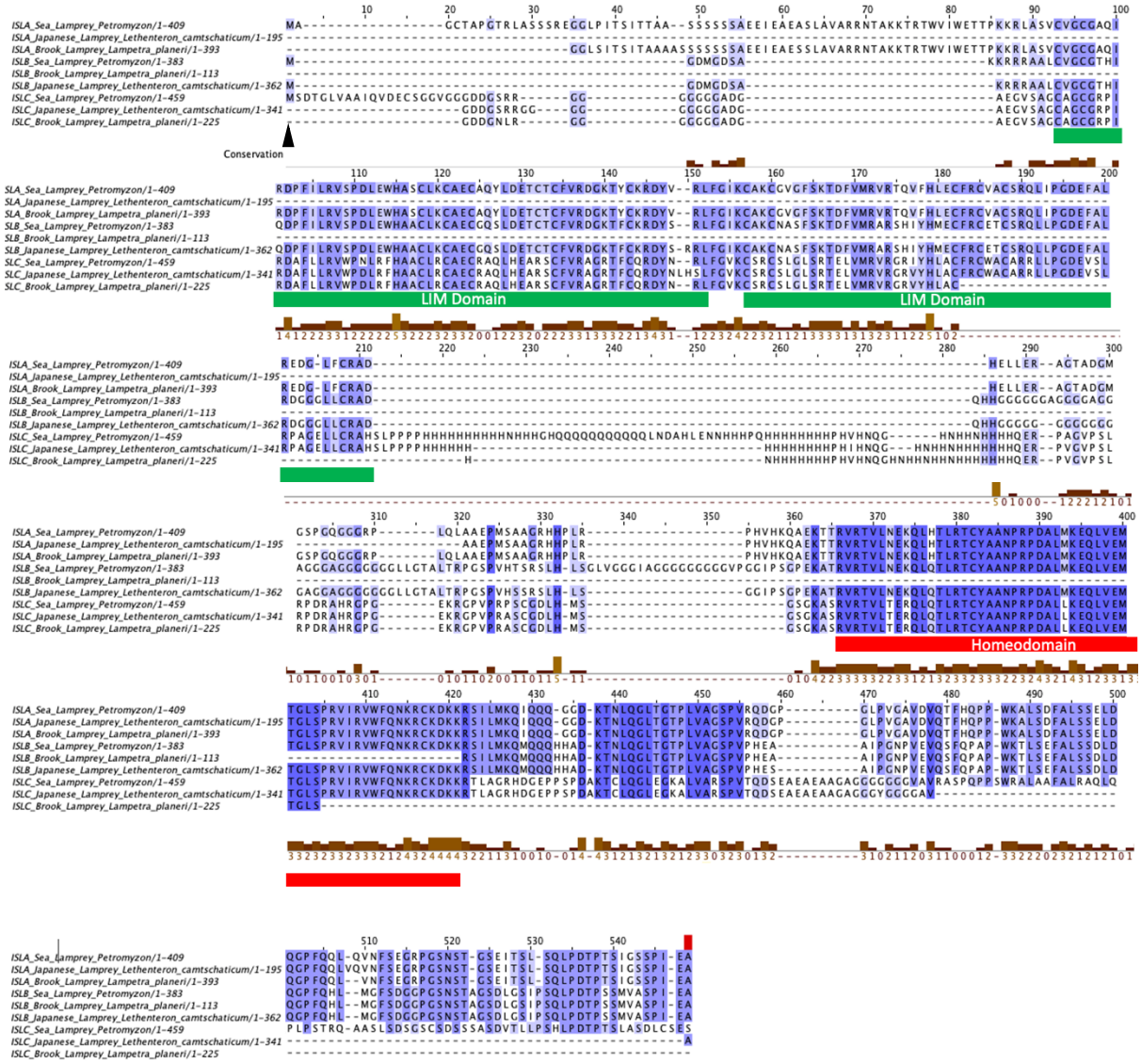


Figure 2.6. Alignment of predicted amino acid sequences of lamprey *ISLA*, *ISLB*, *ISLC* fragments. The two LIM domain are underlined in green while the homeodomain is in red as indicated. Predicted methionine start codons are indicated with black arrowhead.

Given that only partial coding region of *L. planeri* ISL sequences were retrieved, only the complete *P. marinus* ISL sequences were used for the subsequent molecular phylogeny. Maximum likelihood grouped *P. marinus* ISLC with jawed vertebrate ISL2 sequences (Fig. 2.7). The *P. marinus* ISLA and ISLC proteins on the other hand group at base of gnathostome ISL1 sequences. However, as is often the case for phylogenetic tree analyses of other gene families including cyclostome sequences (Kuratani et al., 2001), I could not obtain a robust tree to support the orthology of ISLA, ISLB and ISLC to representative of the ISL1 and ISL2 families in jawed vertebrates as bootstrap values are low (19% for ISL1 family and 31% for ISL2 family). Therefore, caution is advised upon interpretation of these proposed orthologies between lamprey and gnathostome Islet homeobox paralogues.

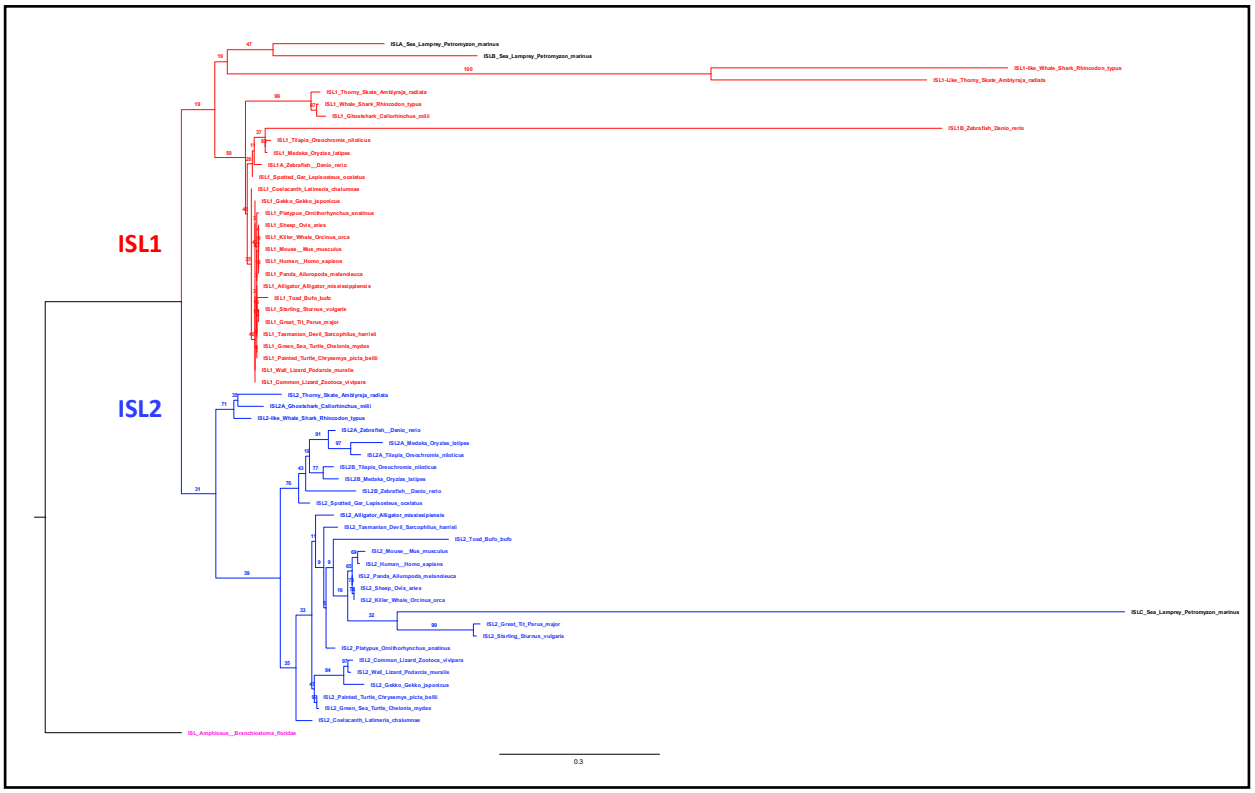


Figure 2.7. Phylogenetic tree of ISL sequences in chordates. Phylogenetic analysis of 58 chordate GnRH amino acid sequences. The tree was constructed by Maximal Likelihood, with numbers below nodes are percentage bootstrap support, with 1000 replicates. Branches and names in the tree are in red if member of the ISL1 family, in blue if member of the ISL2 family. The lamprey sequences are in black. The tree is rooted with the Amphioxus ISL sequence as an outgroup in magenta. Sequence references and alignment are given in Appendix.

2.3.5 ISL gene family expression in the developing Lamprey, *Lampetra planeri*

Isl1 has been recently used as a molecular marker for GnRH cells localized and specified within the olfactory placodes of jawed vertebrates. Thus, I investigated the expression of the lamprey Isl1/2 orthologs, *LpISLA*, *LpISLB* and *LpISLC* in lamprey embryos to see if any hint on the specification events of lamprey GnRH neurons could be found. At T23, *LpISLA* expression is seen in the ventral telencephalon, the ophthalmic and maxillomandibular trigeminal placodes. *LpISLA* transcripts are also present in the pharyngeal region, heart, spinal cord and tailbud (Fig. 2.8A-C). At T24, *LpISLA* are expressed in the same territories as T23 but also clearly appears in the hypothalamus (Fig. 2.8D-F). At T25, expression is seen in hypothalamus, epiphysis, pharyngeal region, spinal cord, tail bud, midbrain tectum and tegmentum. Staining was also visible in all cranial sensory ganglia: the ophthalmic trigeminal (profundal) ganglion, maxillomandibular trigeminal ganglion, the geniculate ganglion, anterior lateral line ganglion, nodose ganglion, petrosal ganglion, posterior lateral line ganglion and possibly hypobranchial ganglia (Fig. 2.8G-I).

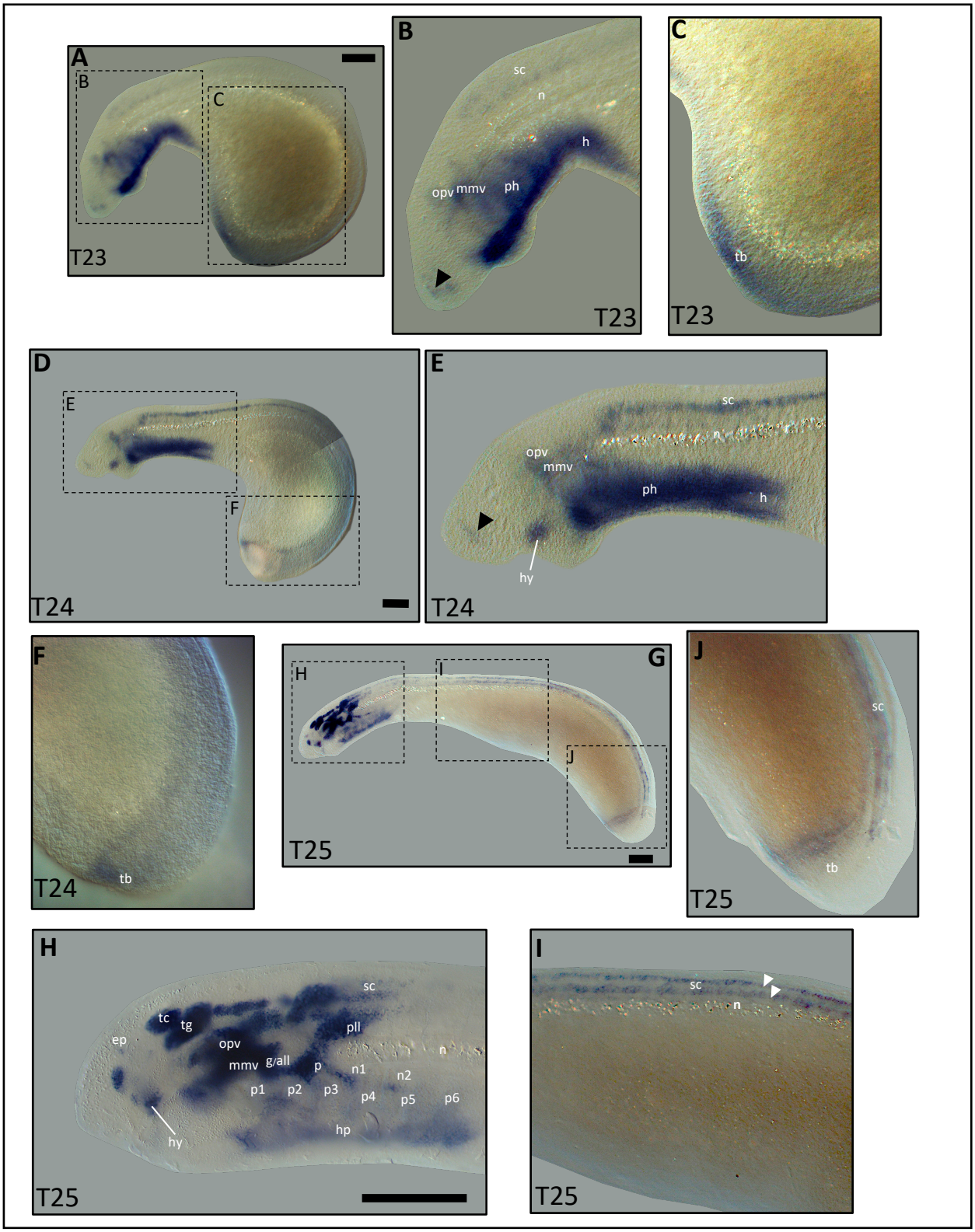


Figure 2.8. *LpISLA* expression during *L. planeri* development at stages 23-25. All images are lateral views and anterior is to the left. B, C, E, F H, I and J: magnification views with higher contrast as outlined by the dotted lines. (A-C) At stage 23, expression is seen in the ventral telencephalon (black arrowhead), the ophthalmic (opV) and maxillomandibular trigeminal (mmV) placodes. *LpISLA* transcripts are also present in the pharyngeal region (ph), heart (h), spinal cord (sc) and tailbud (tb). (D-F) At stage 24, *LpISLA* are expressed in the same territories as T23 but also clearly appears in the hypothalamus (hy). At stage 25, expression is seen in hypothalamus, epiphysis (ep), pharyngeal region, spinal cord (white arrowheads), tail bud, midbrain tectum (tc) and tegmentum (tg). Staining was also visible in all cranial sensory ganglia: the ophthalmic trigeminal (profundal) ganglion (opV), maxillomandibular trigeminal (mmV) ganglion, the geniculate ganglion (g), anterior lateral line ganglion (all), nodose ganglion (n), petrosal ganglion (p), posterior lateral line ganglion (pll) and possibly hypobranchial ganglia (hp). Scale bars: 200 µm.

At T26 and T27, transcripts in the forebrain are visible in the two areas in the hypothalamus, but also in all cranial sensory ganglia, the dorsal isthmus, caudal rhombencephalic alar plate, velum, spinal cord, and pharyngeal region (Fig. 2.9J-K). At T28, *LpISLA* expression is seen in the same territories as T27 except the staining in the pharyngeal region disappeared (Fig. 2.9L). In the hypothalamic preoptic area, expression is seen as two bilateral dots in a place and shape reminiscent of *LpGnRH-I and -III* (Fig. 2.9M) hinting for a likely co-localisation of those markers, as seen in jawed vertebrates.

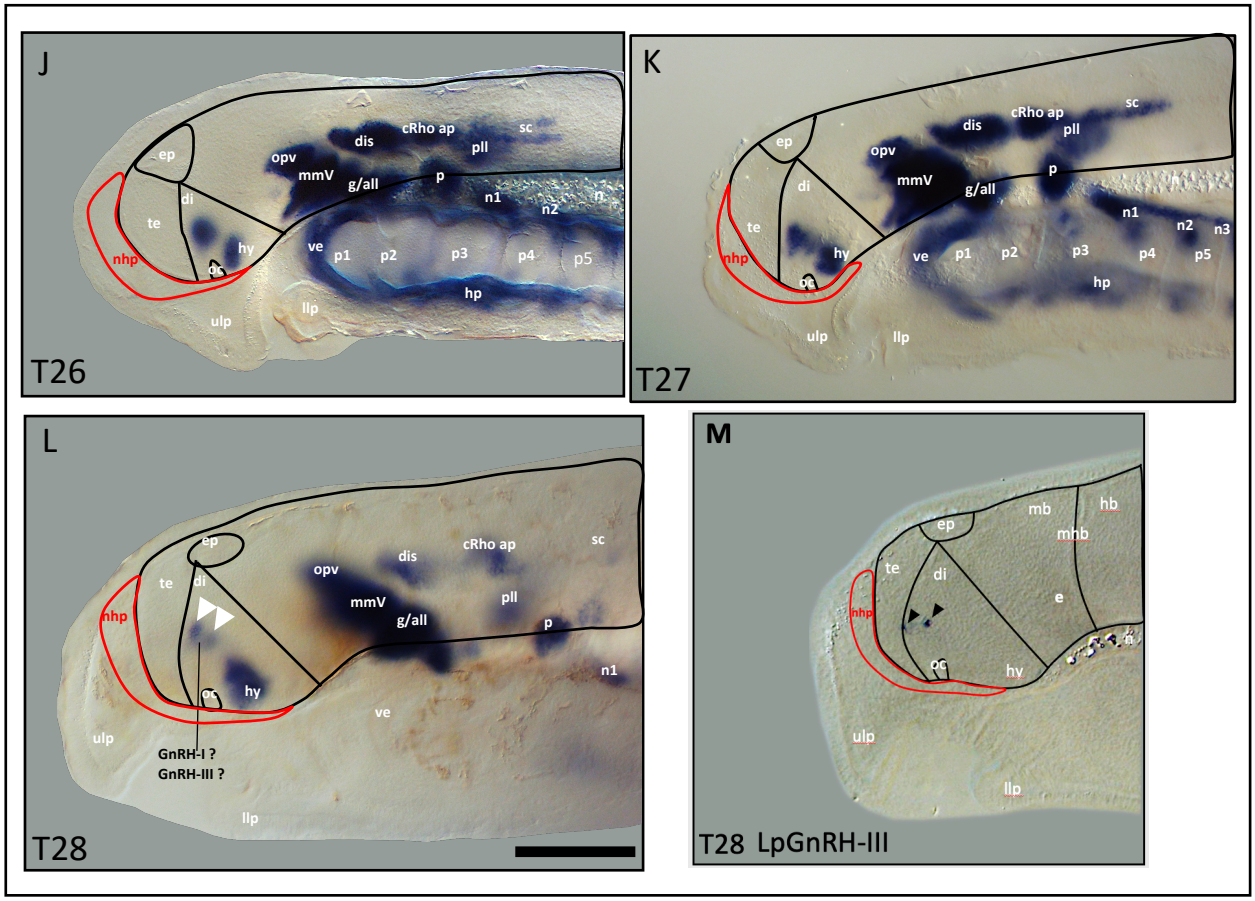


Figure 2.9. *LpISLA* expression during *L. planeri* development at stages 26-28. All images are lateral views and anterior is to the left. At stage 26 and 27, transcripts in the forebrain are visible in the two areas in the hypothalamus (hy), but also in all cranial sensory ganglia, i.e., the ophthalmic trigeminal (profundal) ganglion (opV), maxillomandibular trigeminal (mmV) ganglion, the geniculate ganglion (g), anterior lateral line ganglion (all), nodose ganglion (n), petrosal ganglion (p), posterior lateral line ganglion (pll) and possibly hypobranchial ganglia (hp). *LpISLA* RNA was also observed in the dorsal isthmus (dis), caudal rhombencephalic alar plate (cRho ap), velum, spinal cord (sc) and pharyngeal region. At stage 28, *LpISLA* expression is seen in the same territories as T27 except the staining in the pharyngeal region disappeared. In the hypothalamic preoptic area, expression is seen as two bilateral spots (white arrowhead) in a place and shape reminiscent of *LpGnRH-I and -III* (black arrowhead) (M) hinting for a possible co-localisation of those markers. The black outline represents major brain subdivisions based on the lamprey prosomeric model established from neurogenin gene expression (Lara-Ramírez et al., 2015). The red outline represents the nasohypophyseal placode (nhp). Additional abbreviations: oc, optic chiasma; hb, hindbrain; llp, lower lips; mhb, midbrain-hindbrain boundaries; n, notochord; ot, otic; p1-p5; pharyngeal pouches te, telencephalon; ulp, upper lips; ve, velum. Scale bar: 200 µm.

At T23, *LpISLB* expression is seen in the ophthalmic trigeminal and maxillomandibular trigeminal placodes, hindbrain, spinal cord, and heart region (Fig. 2.10A). At T24, expression is still present in the ophthalmic trigeminal and maxillomandibular trigeminal placodes and hindbrain (Fig. 2.10B). At T25, *LpISLB* transcripts are seen in the epiphysis, the ventral telencephalon (subpallium), the hindbrain, the ophthalmic trigeminal ganglion, maxillomandibular trigeminal ganglion, geniculate ganglion, petrosal ganglion, heart tube and spinal cord (Fig. 2.10C) as previously reported for that stage in the lamprey *L. camtschaticum* (Sugahara et al., 2011).

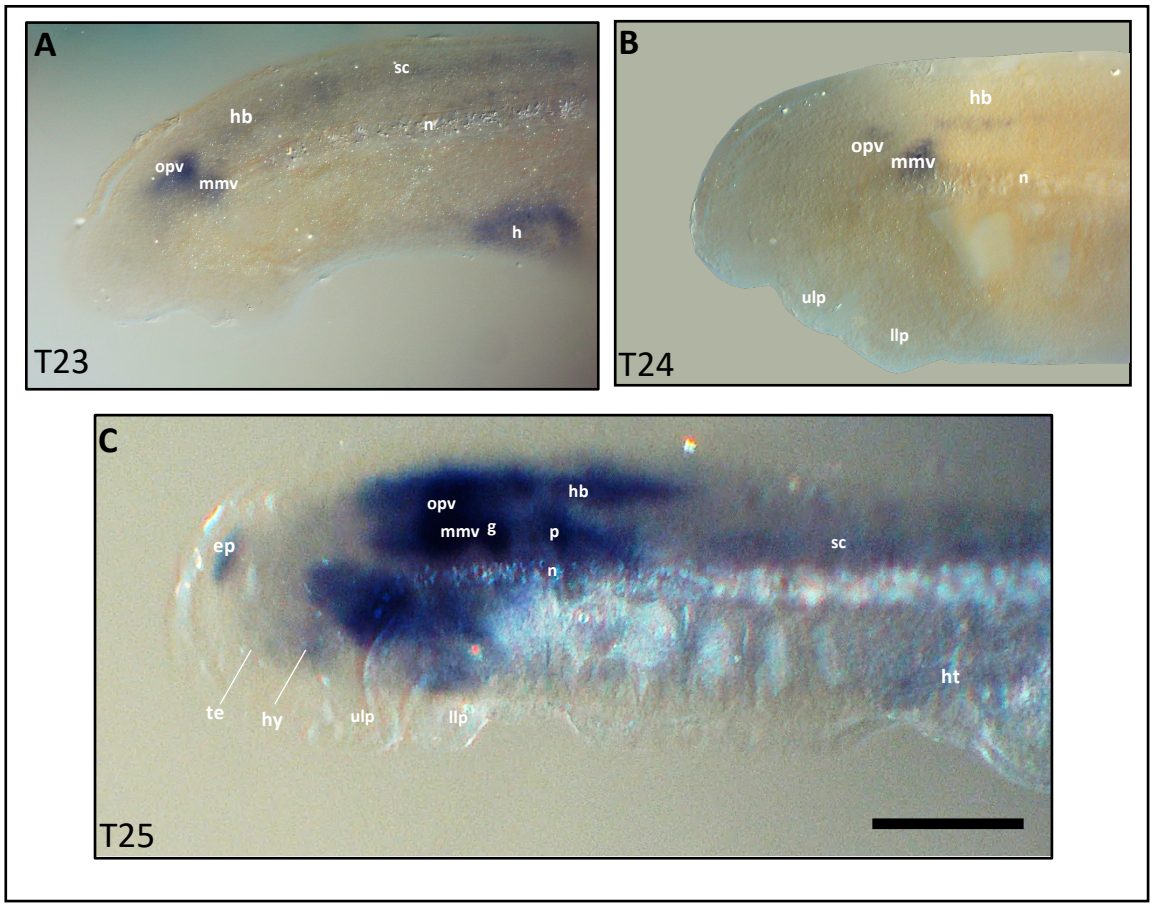


Figure 2.10. *LpISLB* expression during *L. planeri* development at stages 23-25. All images are lateral views and anterior is to the left. (A) At stage 23, expression is seen in the ophthalmic trigeminal (opV) and maxillomandibular trigeminal (mmV) placodes, hindbrain (hb), spinal cord (sc) and heart region (h). (B) At stage 24, expression is still present in the ophthalmic trigeminal and maxillomandibular trigeminal placodes and hindbrain. (C) At stage 25, *LpISLB* transcripts are seen in the epiphysis, the ventral telencephalon (subpallium), the hindbrain, the ophthalmic trigeminal ganglion, maxillomandibular trigeminal ganglion, geniculate ganglion, petrosal ganglion, heart tube and spinal cord. Additional abbreviations: llp, lower lips; n, notochord; ulp, upper lips. Scale bar: 200 µm.

At T23, *LpISLC* expression is seen in the nasohypophyseal placode, midbrain, hindbrain, pharyngeal region, stomodeum, and spinal cord (Fig. 2.11A). At T24, expression is in the nasohypophyseal placode, upper lips, lower lips, pharyngeal arches, and the petrosal placode (Fig. 2.11B). At T25, *LpISLC* transcripts are seen in the same territories as T24 but also appears in the hypothalamus (Fig. 2.11C). At T26, expression only remain in the anterior pharyngeal arches (Fig. 2.11D).

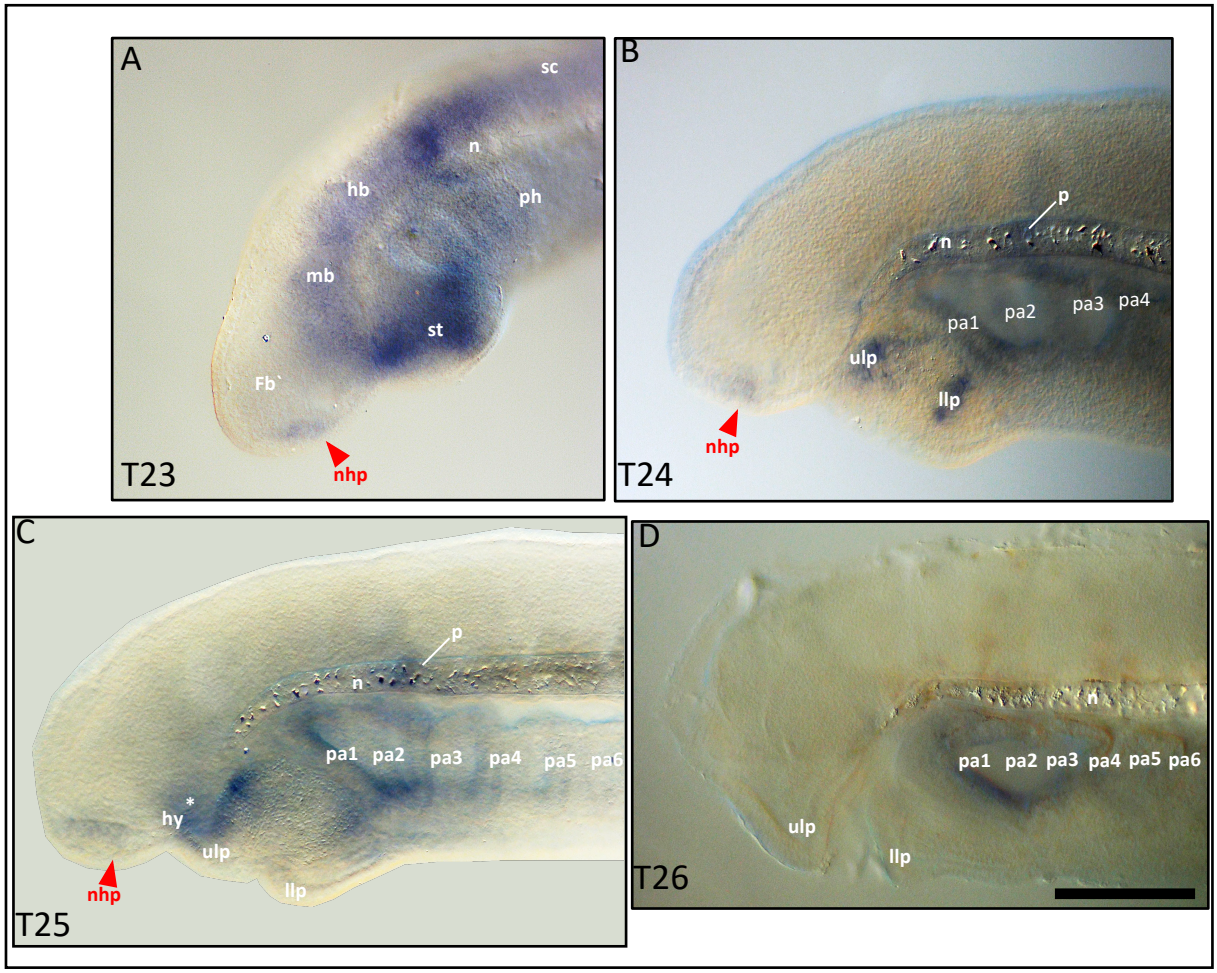


Figure 2.11. *LpISLC* expression during *L. planeri* development at stages 23-26. All images are lateral views and anterior is to the left. (A) At stage 23, expression is seen in the nasohypophyseal placode (nhp) (red arrowhead), midbrain (mb), hindbrain (hb), pharyngeal region (ph), stomodeum (st) and the spinal cord (sc). (B) At stage 24, expression is in the nasohypophyseal placode, upper lips (ulp), lower lips (llp), pharyngeal arches (pa1-pa4) and the petrosal placode (p).(D-E) At stage 25, *LpISLC* transcripts are seen in the same territories as T24 but also appears in the hypothalamus (white asterisk). At stage 26, expression only remain in the anterior pharyngeal arches. Additional abbreviations: n, notochord. Scale bar: 200 µm.

2.3.6 Identification of NELF gene in *Lampetra planeri* and phylogenetic analysis

As seen in other jawed vertebrates, I report the presence of a single NELF gene in the Lamprey *Lampetra planeri,* termed *LpNELF*. Maximum likelihood clustered the lamprey NELF sequences of *L. planeri, P. marinus* and *L. camtschaticum* together at the base of gnathostome NELF sequences (Fig. 2.12).

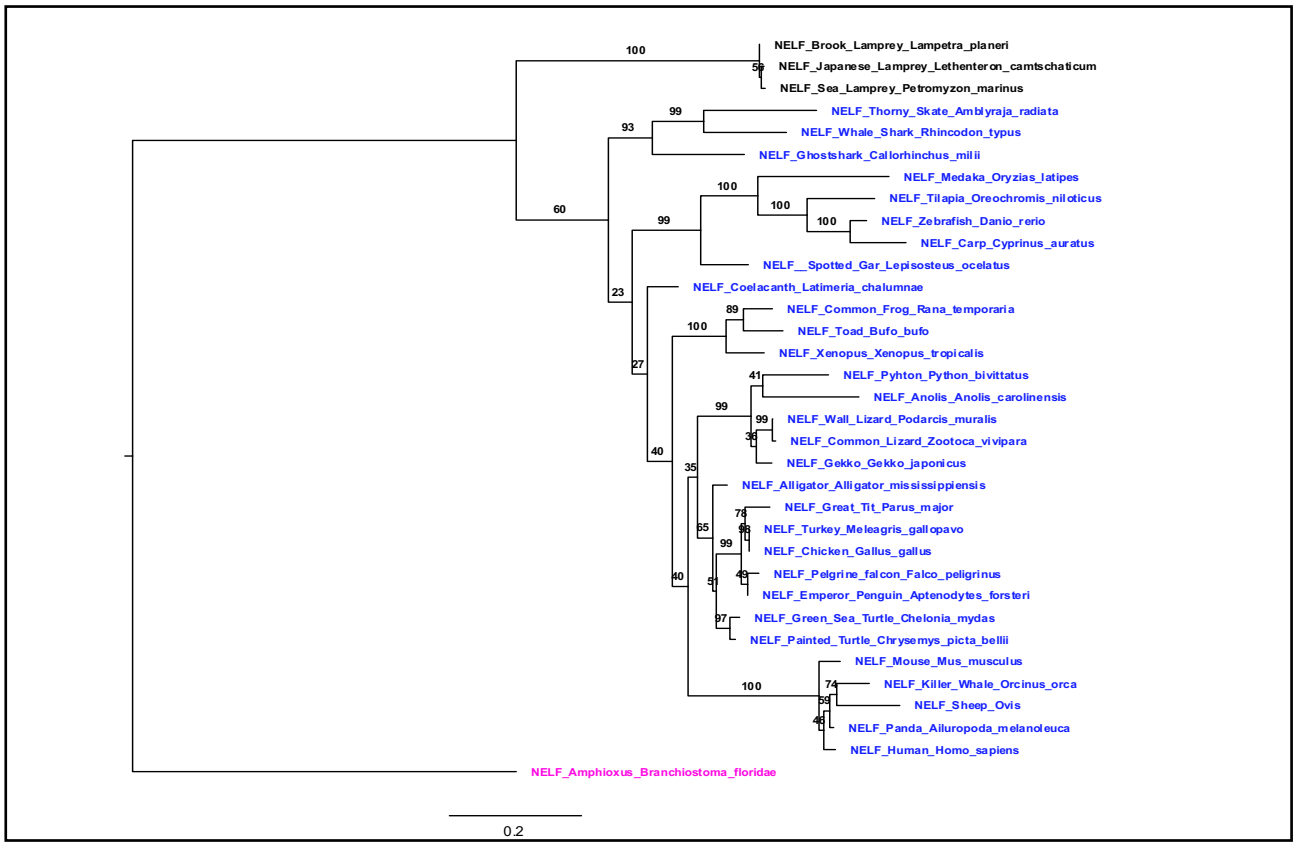


Figure 2.12. Phylogenetic tree of NELF sequences in chordates. Phylogenetic analysis of 34 chordate NELF amino acid sequences. The tree was constructed by Maximal Likelihood, with numbers below nodes are percentage bootstrap support, with 1000 replicates. Branches and names in the tree are in blue if NELF sequences of jawed vertebrates while the lamprey sequences are in black. The tree is rooted with an Amphioxus NELF sequence as an outgroup in magenta. Sequence references and alignment are given in Appendix.

2.3.7 NELF gene expression in the developing Lamprey, *Lampetra planeri*

NELF expression in jawed vertebrates has been used as a marker of GnRH neuron migration from the olfactory placode to the hypothalamus. Therefore, this marker is of particular interest to assess the origin of GnRH neurons in lamprey. *LpNELF* expression was not detected at T23 and T24. At T25, expression is seen in the maxillomandibular trigeminal, the posterior lateral line, the geniculate and anterior lateral line placodes. Staining is also seen in the spinal cord and pharyngeal region (Fig. 2.13A). At T26, some expression remained in the hindbrain, spinal cord, and pharyngeal region. Also, the first *LpNELF* forebrain expression become visible in the hypothalamus (Fig. 2.13B). The weak signal in the hypothalamus was better revealed under high contrast (Fig. 2.13C). At T27, *LpNELF* transcripts are seen in the upper lip, hypothalamus, optic chiasma, midbrain tegmentum, hindbrain, and spinal cord (Fig. 2.13D and 2.13F). At T28, *LpNELF* expression is seen in the same territories as T27 except that some *LpNELF* staining in the hypothalamic area is seen as a prominent transverse line localized in a place and shape reminiscent of *LpGnRH-II* at that stage, suggesting a possible co-localisation of those markers as seen in jawed vertebrates (Fig. 2.13E, 2.13G and 2.13H).

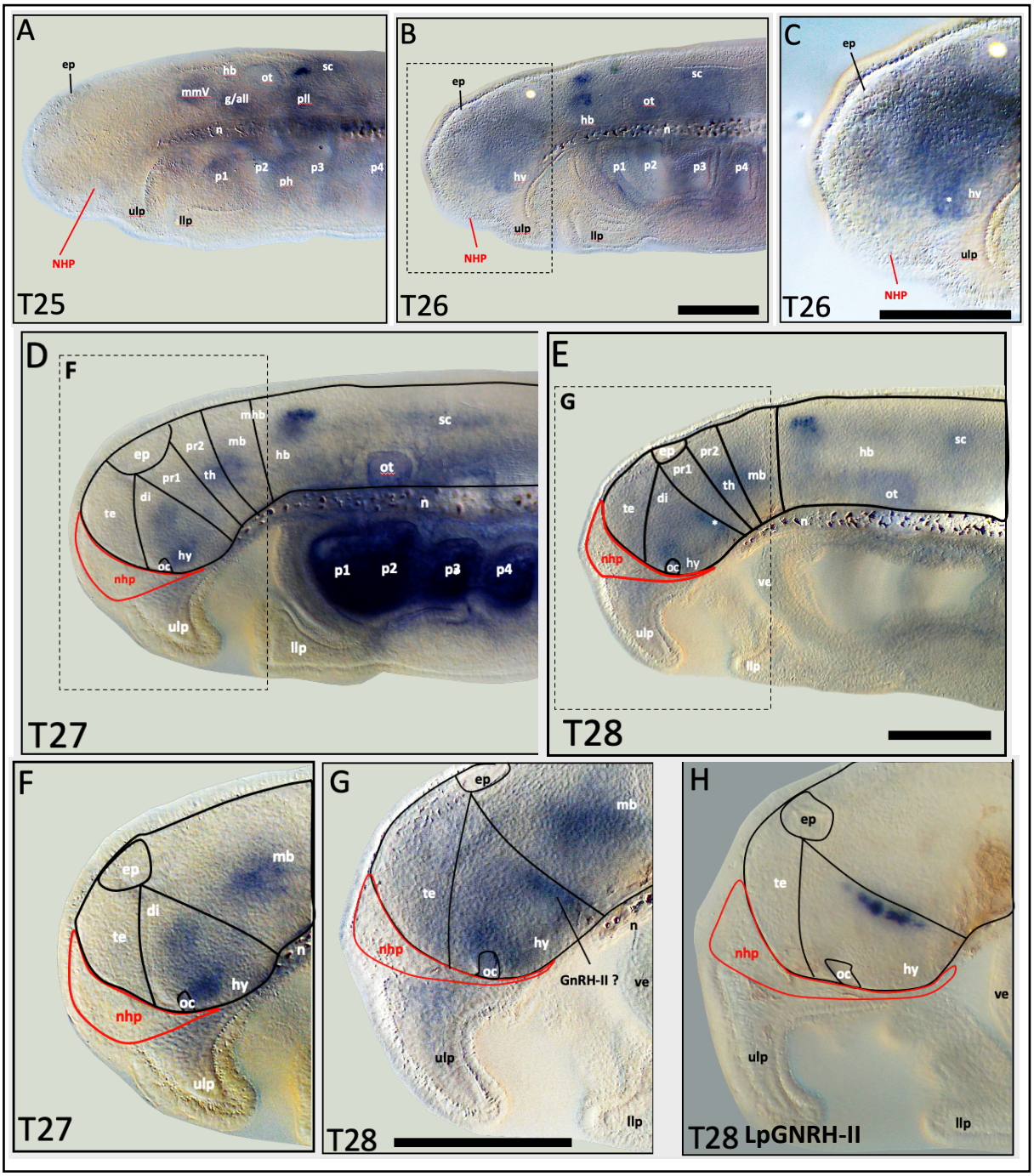


Figure 2.13. *LpNELF* expression during *L. planeri* development at stages 25-28. All images (A-H) are lateral views and anterior is to the left. C, E and G are magnification views with higher contrast as outlined by the dotted lines. (A) At stage 25, expression is seen in the maxillomandibular trigeminal (mmV), the posterior lateral line (pll), the geniculate and anterior lateral line (g/all) placodes. Staining is also seen in the spinal cord (sc) and pharyngeal region (ph). (B-C) At stage 26, At T26, some expression remained in the hindbrain, spinal cord, and pharyngeal region. Also, the first *LpNELF* forebrain expression become visible in the hypothalamus (white asterisk). (D-E) At stage 27, *LpNELF* transcripts are seen in the upper lip, hypothalamus, optic chiasma, midbrain, thalamus, hindbrain, and spinal cord. At stage 28 *LpNELF* expression is seen in the same territories as T27 except that some *LpNELF* staining in the hypothalamus is seen as a prominent transverse line localized in a place and shape reminiscent of *LpGnRH-II* at that stage (H) suggesting a possible co-localisation of those markers. The black outline represents major brain subdivisions based on the lamprey prosomeric model established from neurogenin gene expression (Lara-Ramírez et al., 2015). The red outline represents the nasohypophyseal placode (nhp). Abbreviations: ch, optic chiasma; ep, epiphysis; hb, hindbrain; hy, hypothalamus; llp, lower lips; mhb, midbrain-hindbrain boundaries; mmV, maxillomandibular trigeminal placode; mt, midbrain tegmentum; nhp, nasohypophyseal placode; n, notochord; ot, otic; p1-p5; pharyngeal pouches; pr1; prosomere 1; pr2; prosomere 2; pll, posterior lateral line placode; sc, spinal cord; te, telencephalon; ulp, upper lips; ve, velum. Scale bars: 200µm.

2.3.8 Identification of FGF8/17 gene in *Lampetra planeri*

Orthologues to FGF8 have previously been identified in the lampreys *L. camtschaticum, P. marinus* and *L. fluviatilis* and the phylogenetic analyses revealed that they sit at the base of jawed vertebrate FGF8 and FGF17 sequences (Shigetani et al., 2002; Hammond and Whitfield, 2006; Jandzik et al., 2014). Amino acid alignment confirmed that the Fgf8/17 orthologue sequence identified in *L. planeri* was 90.28% identical to its *P. marinus* cognate (Fig. 2.14).

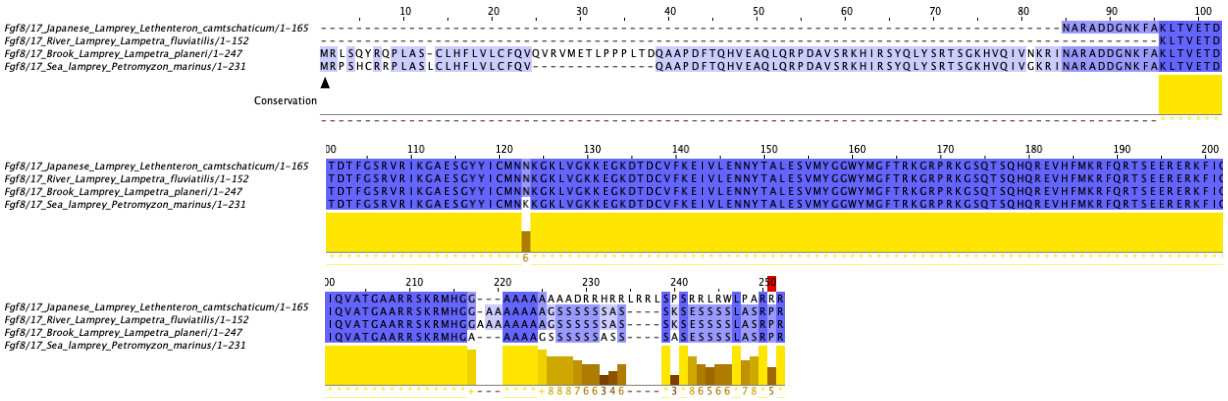


Figure 2.14. Alignment of predicted amino acid sequences of the Fgf8/17 sequences in four lamprey species (*L. camtschaticum, L. fluviatilis, L. planeri* and *P. marinus*). Predicted methionine start codons are indicated with the black arrowhead

2.3.9 FGF8/17 gene expression in the developing Lamprey, *Lampetra planeri*

To focus on the inductive events possibly involved in the development of the lamprey GnRH neurons at a molecular level, I further examined the expression patterns of the signalling molecule-encoding gene *LpFgf8/17*. At T23, *LpFgf8/17* expression is seen in part of the anterior ectoderm that corresponds to the future nasohypophyseal placode. Expression is also observed in the midbrain-hindbrain boundary and hindbrain expression was seen in placodal territories such as the trigeminal, otic, and posterior lateral line placodes. *LpFgf8/17* RNA was also seen in the developing pharyngeal arches, upper lips, lower lips and tailbud (Fig. 2.15A-C). At T24, expression is in the nasohypophyseal placode, ventral telencephalon, and hypothalamus. Expression is also present in the midbrain-hindbrain boundary, pharyngeal arches, upper lips, lower lips, tailbud, otic and posterior lateral line placodes (Fig. 2.15D-F). At T25, expression is in the epiphysis, nasohypophyseal placode, ventral telencephalon, two distinct areas of the hypothalamus, pharyngeal region, velum, upper lips, lower lips and tailbud (Fig 2.15G-I).

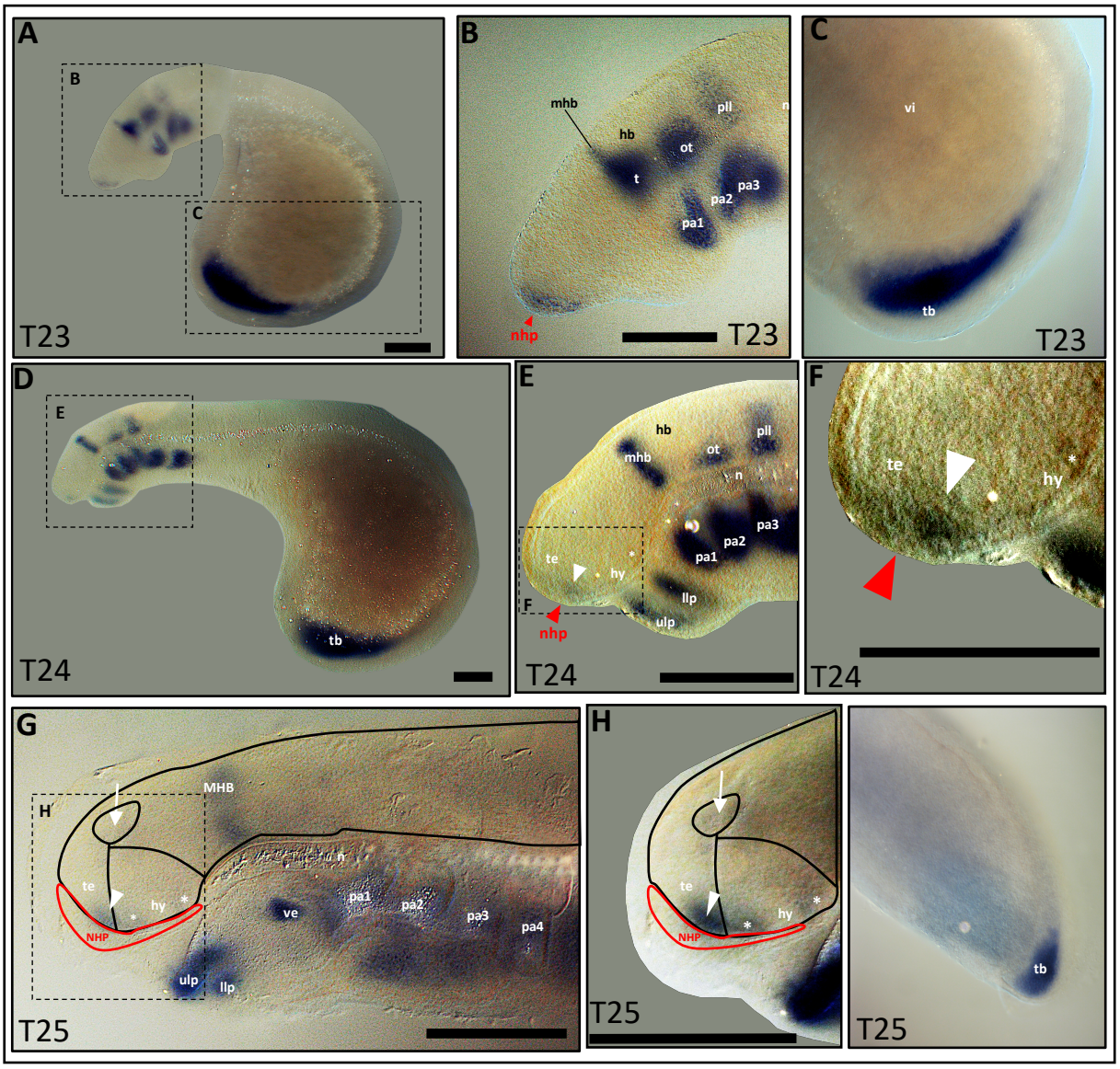


Figure 2.15. *LpFGF8/17* expression during *L. planeri* development at stages 23-25. All images are lateral views and anterior is to the left. B, C, E, F and H: magnification views with higher contrast as outlined by the dotted lines. (A-C) At stage 23, expression is seen in part of the anterior ectoderm that corresponds to the future nasohypophyseal placode (nhp) (red arrowhead). Expression is observed in the midbrain-hindbrain boundary (mhb) and hindbrain (hb) but also in placodal territories such as the trigeminal (t), otic (ot), and posterior lateral line placodes (pll). *LpFgf8/17* RNA was also seen in the developing pharyngeal arches (pa), upper lips (ulp), lower lips (llp) and tailbud (tb). (D-F) At stage 24, expression is in the nasohypophyseal placode (red arrowhead), ventral telencephalon (te) (white arrowhead) and hypothalamus (white asterisk). Expression is also present in the midbrain-hindbrain boundary, pharyngeal arches, upper lips, lower lips, tailbud, otic and posterior lateral line placodes. (G-I) At stage 25, expression is in the epiphysis (white arrow), nasohypophyseal placode, ventral telencephalon (white arrowhead), two distinct areas of the hypothalamus (white asterisks), pharyngeal region, velum (ve), upper lips, lower lips and tailbud. Additional abbreviations: n, notochord. Scale bars: 200 µm.

At T26 brain expression was only seen in a faint region of the midbrain-hindbrain boundary, but expression was still present in the velum, upper lips, lower lips and tailbud (Fig. 2.16J-K). At T27, expression persisted in the velum, upper lips, and lower lips (Fig. 2.16L-M). At T28, *LpFgf8/17* transcripts were only strongly seen in the velum and a very faint staining was see in the lower lips (Fig. 2.16N-O).

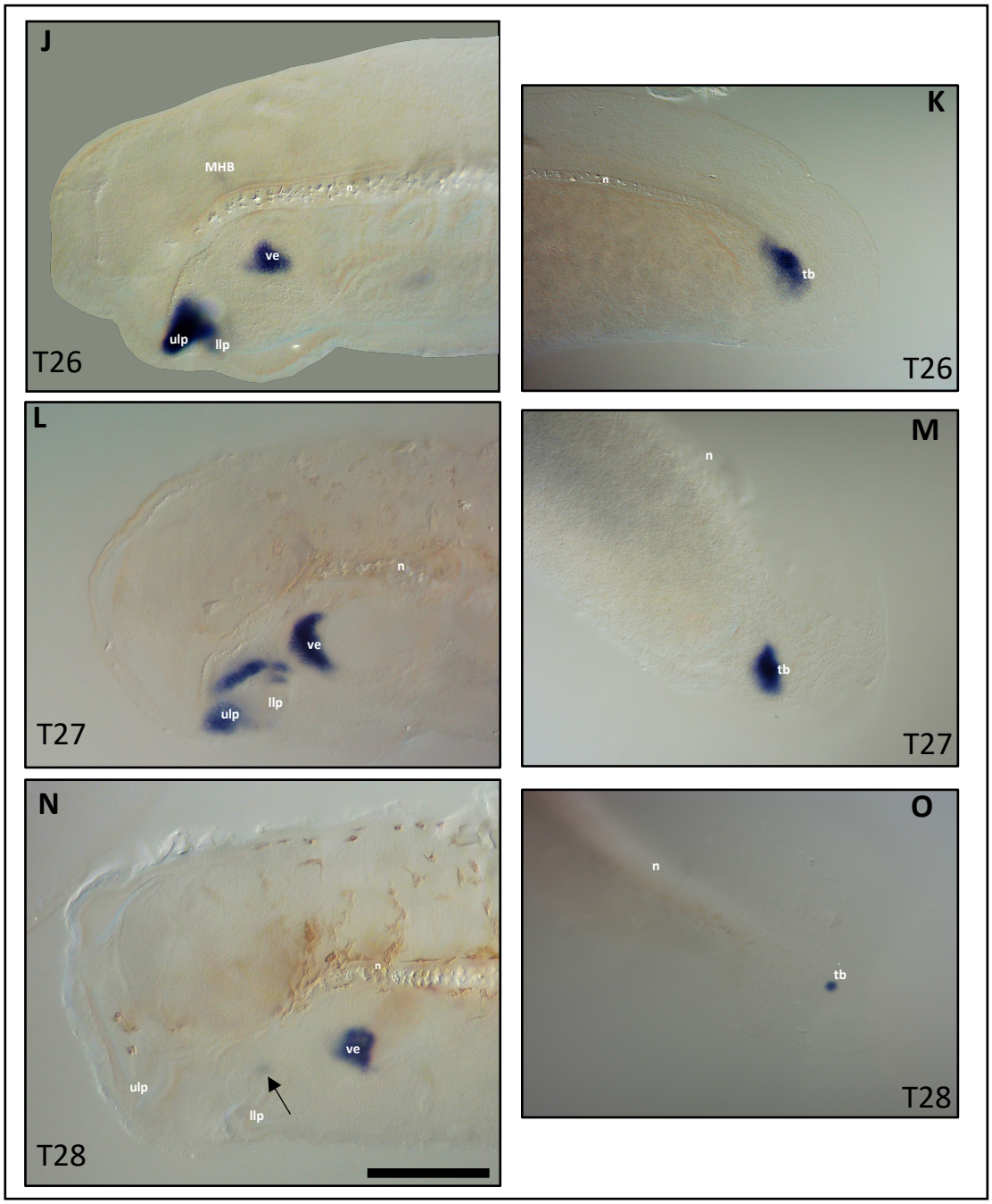


Figure 2.16*. LpFGF8/17* expression during *L. planeri* development at stages 26-28. All images are lateral views and anterior is to the left. (J-K) At stage 26, brain expression was only seen in a faint region of the midbrain-hindbrain boundary, but expression was still present in the velum (ve), upper lips (ulp), lower lips (llp) and tailbud (tb). (L-M) At stage 27, expression persisted in the velum, upper lips, lower lips and tailbud. (N-O) At stage 28, *LpFgf8/17* transcripts were only strongly seen in the velum and a faint staining remained in the lower lips (black arrow) and tailbud. Additional abbreviations: n, notochord. Scale bar: 200 μm.

2.3.10 FGF and RA signalling act in specific time windows to specify LpGnRH-III neurons

In chick, the fibroblast growth factor (FGF) signalling is known to act during a short time window to specify GnRH1 neurons in the olfactory placode via Fgf8 and its receptor Fgfr1 (Sabado et al., 2012). To test whether FGFs such as *LpFgf8/17* induce the specification of lamprey GnRH neurons, I applied SU5402, a potent FGF receptor inhibitor onto embryos (Mohammadi et al., 1997). The FGF antagonist SU5402 has inhibitory effect on the action of tyrosine kinases as it competitively binds to the ATP-binding site on the fibroblast growth factor receptor. It was chosen because it has a known ability to inhibit FGF signalling in lamprey and other jawed vertebrates (Sabado et al., 2012; Sugahara et al., 2011; Jandzik et al., 2014). The SU5402 was applied during specific developmental time windows and embryos fixed when they reached T29, a stage where GnRH neuronal populations are specified and GnRH paralogue expression visible in the diencephalon. These drug-treated embryos were probed for *LpGNRH-III* and *LpGNRH-II* transcripts. In addition, control embryos were run in parallel for all time windows studied with DMSO (see material and methods for details). The SU5402-treated embryos have lost the expression of *LpGNRH-III* when the FGF pathway was inhibited during the following time windows; between T23 to T26 or T25 to T27 (Fig.2.17). In addition, the inhibition of FGF signalling in later time windows between T26 to T28 or T27 to T29 did not prevent LpGNRH-III neuron production. The fact that inhibition of FGF during the time of LpGnRH-III neuron appearance in the brain did not prevent the expression of *LpGnRH-III* indicates that the FGF pathway is required for the early phase of LpGNRH-III neuron specification, but not for their maintenance. In other words, LpGNRH-III cells are independent of sustained FGF signalling once they are specified and the LpGNRH-III

precursors are certainly specified prior to *LpGNRH-III* expression at T28. On the contrary, the FGF inhibition with SU5402 did not affect the expression of *LpGNRH-II* in every developmental time windows tested, implying that different mechanisms of specification exist for the distinct GnRH neuronal populations in lampreys. Because FGF signalling is important for LpGNRH-III neuron specification, I also tested whether retinoic acid (RA) signalling interferes with *LpGNRH-III* expression during the same developmental time windows. In fact, in chick FGF and RA pathways regulate and antagonize each other in a brief time window to specify olfactory-derived GnRH1 neurons. While FGF induces the formation of GnRH1 neurons, RA suppresses them. To test the effect of RA signalling in lamprey, I applied All-*trans*-RA, an agonist ligand of retinoic acid receptors into embryos. Interestingly, this drug-treatment resulted in a lack of LpGNRH-III neurons during the same time windows as with the FGF inhibitor SU5402, i.e., between T23 to T26 or T25 to T27 (Fig. 2.17). The latter time windows between T26 to T28 or T27 to T29 did not inhibit LpGNRH-III neuron formation like the DMSO-treated controls. I also inhibited RA signalling with the drug BMS493, an inverse agonist of pan-retinoic acid receptors which work in jawed vertebrates and amphioxus (Comai et al., 2020; Zhong et al., 2020). The production of LpGnRH-III cells was unaffected by the presence of this RA signalling inhibitor like in the DMSO-treated controls. Overall, these results demonstrate that RA represses the formation of LpGNRH-III precursors prior to *LpGnRH-III* expression at T28 but that it does not interfere with their subsequent development once these cells are specified. In conclusion, the result points that both FGF and RA signalling act during specific time window between T23-T26 and/or T25-27 to affect LpGnRH-III neuron precursors but not more mature LpGnRH-III expressing cells. If we account for common developmental time between the windows T23-T26 and T25-27, we can deduce that the specification of LpGnRH-III precursors likely take place between T25 and T26.

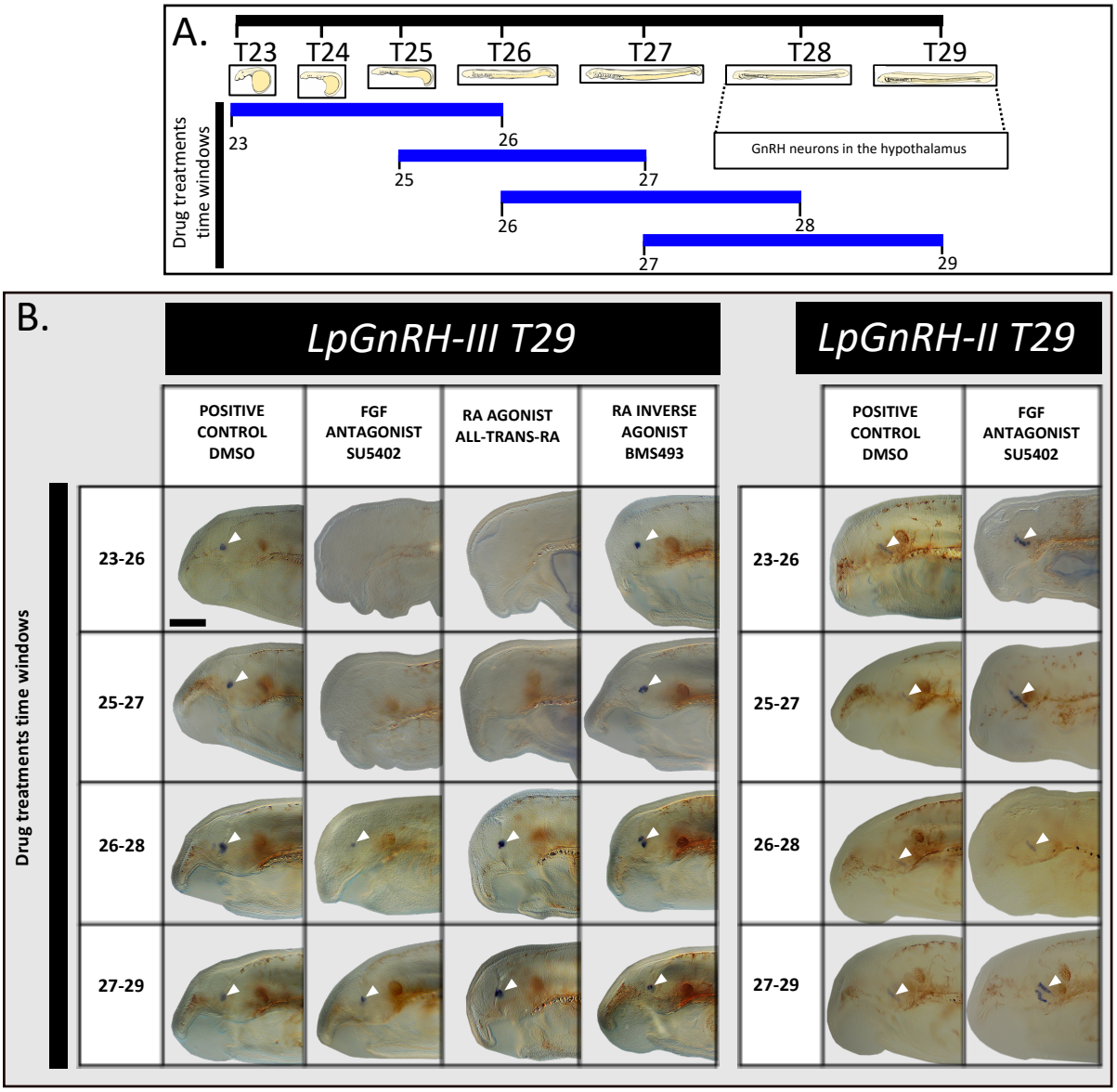


Figure 2.17. Effect of FGF and RA signalling drugs on the specification of *LpGnRH-III and LpGnRH-II* neurons during specific time windows. (A) Schematic summary of the different time windows (blue bars) used for each drug treatments performed. (B) Untreated larvae with DMSO or drug-treated larvae with SU5402, All-Trans-RA or BMS493 were fixed at T29 and probed for *LpGnRH-III or LpGnRH-II* transcripts. The larvae exposed to SU5402 or All-Trans-RA from Tahara st. 23- 26 or from st. 25-27 have apparent loss of *LpGnRH-III* expression in the hypothalamus. The FGF antagonist SU5402 has no apparent effect on the specification of *LpGnRH-II* neurons. White arrowhead on *LpGnRH-III or LpGnRH-II* positive cells. n = 6 embryos for each condition tested. Scale bar: 200 µm.

2.3.11 No cells migrate out of the developing NHP in lamprey

I re-considered the possibility that lamprey GnRH neurons arise from the nasohypophyseal placode (NHP) like the GnRH1/3 neurons of gnathostomes migrating out the olfactory placodes. In fact, despite the lack of GnRH-expressing cells in the NHP during development, it remained possible that GnRH precursors are specified within the NHP, and only express GnRHs transcripts once settled in the diencephalon thus after a discrete migratory phase. Several elements in this study hinted that such nasal migratory mechanism may have an ancestral origin in vertebrates such as the expression of *LpISLC*, *LpCOEA*, *LpCOEB* in the NHP and of *LpFGF8/17* in the NHP and anteroventral telencephalon (Lara-Ramirez et al., 2017; this study). Also, the effect of FGF and RA signalling pathways on the specification of LpGnRH-III cells has strong resemblance to the FGF-RA antagonistic mechanism that take place during the specification of olfactory-derived GnRH1 neurons in chick (Sabado et al., 2012). To assess the putative placodal origin of GnRH neurons in lamprey, I made focal injections of Chloromethyl Dil-dye in the NHP of living lamprey embryos to identify possible migrating cells that could arise from that region. This fluorescent dye is taken up preferentially but not selectively by epidermal receptors and often stains sensory axons to their full extent (Holland and Yu, 2002). The dye was initially placed at T23, when the nasohypophyseal plate starts to form and embryos where imaged throughout development until T29 when GnRH neurons are well established in their mature hypothalamic locations (Fig 2.18A). The Dil-dye cell lineage tracing showed no labelled cells migrating out of the NHP (Fig 2.18B) which is in accordance with previous analysis by immunochemistry showing that embryonic GnRH neurons are not associated with the NHP in developing lampreys (Tobet et al., 1996). Attempts were made to confirm the presence of GnRH-I, -II and -III proteins in those Dil-

labelled embryos with lamprey-specific GnRH antibodies known to mark GnRH neurons in thin brain section (~40 μM) (Sower, 2015) and mouse ISL1/2 antibody which marks GnRH neurons originating from the olfactory placodes of gnathostome (Aguillon et al., 2018), unfortunately these antibodies did not work on whole-mount embryos (see Supplementary data, Fig. 2.S1). Nevertheless, the last developmental stage where the embryos were permanently fixed is at T29, a time in development where the GnRH neuron populations are prominent (see Fig. 2.3D and 2.4F-J) therefore Dil-labelled cells should have been visible if any migration took place at earlier stages. In conclusion, the present data support a previous statement by Tobet et al (1996) who hypothesised that GnRH neurons in developing lamprey originate within proliferative zones of the diencephalon rather than the olfactory system.

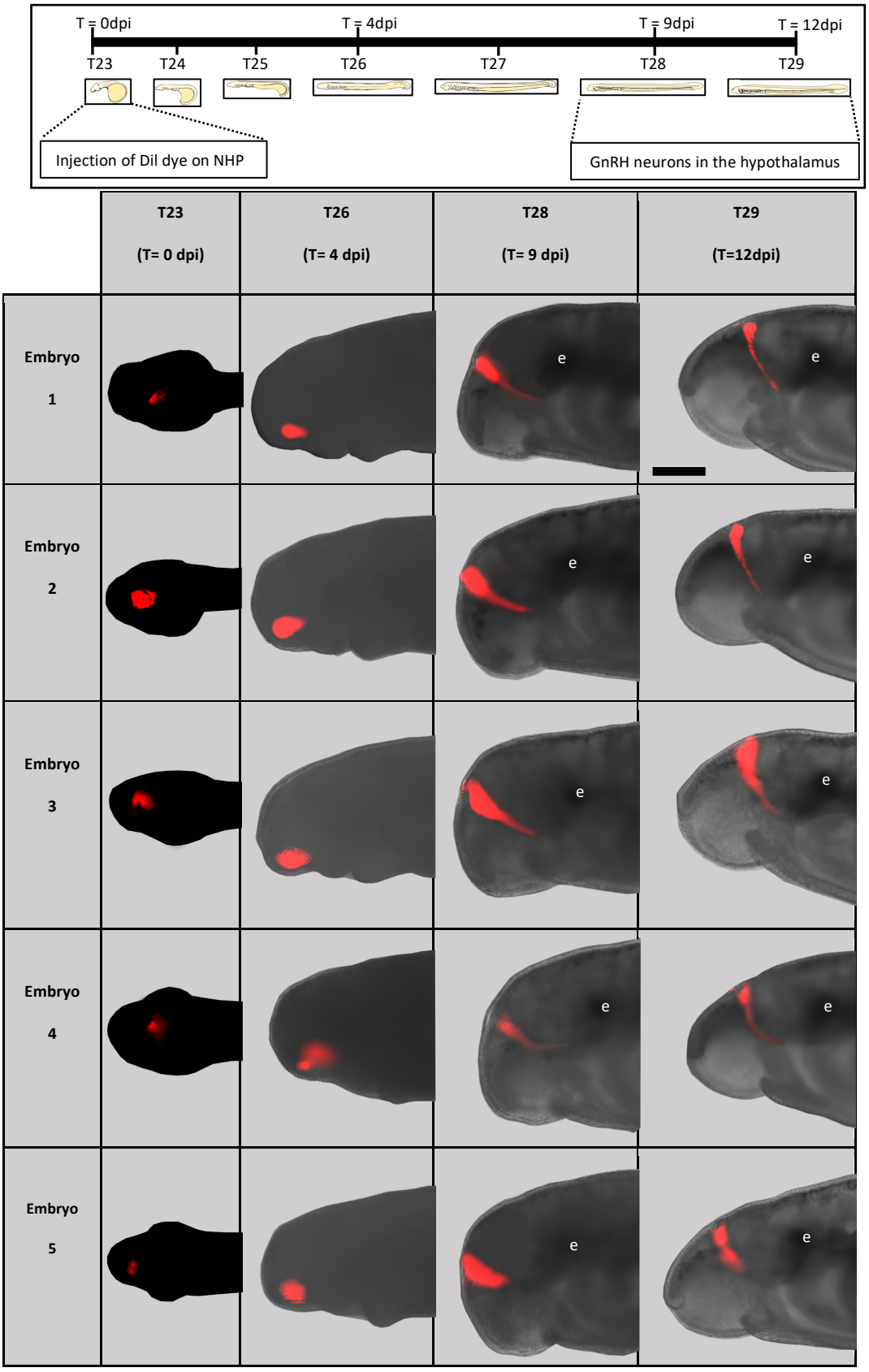


Figure 2.18. Fate-mapping of NHP placode-derived neurons reveals no apparent cell migration to the brain of developing lamprey embryos. (A) Schematic drawing showing lamprey development from stage T23 at the point of injection to T29, where GnRH neurons are present in the diencephalon. (B) The embryos were injected with Dil dye at stage Tahara 23 into the presumptive NHP territory which possibly contain GnRH precursors. The same embryos were followed through development and pictured at stage T25, T28 and T29. The results show Dil labelling in the NHP but no migrating cells to the brain were detected. Abbreviations: dpi, days post-injection; e, eye; t, time. Scale bar: 200μm.

2.3.12 Supplementary Data

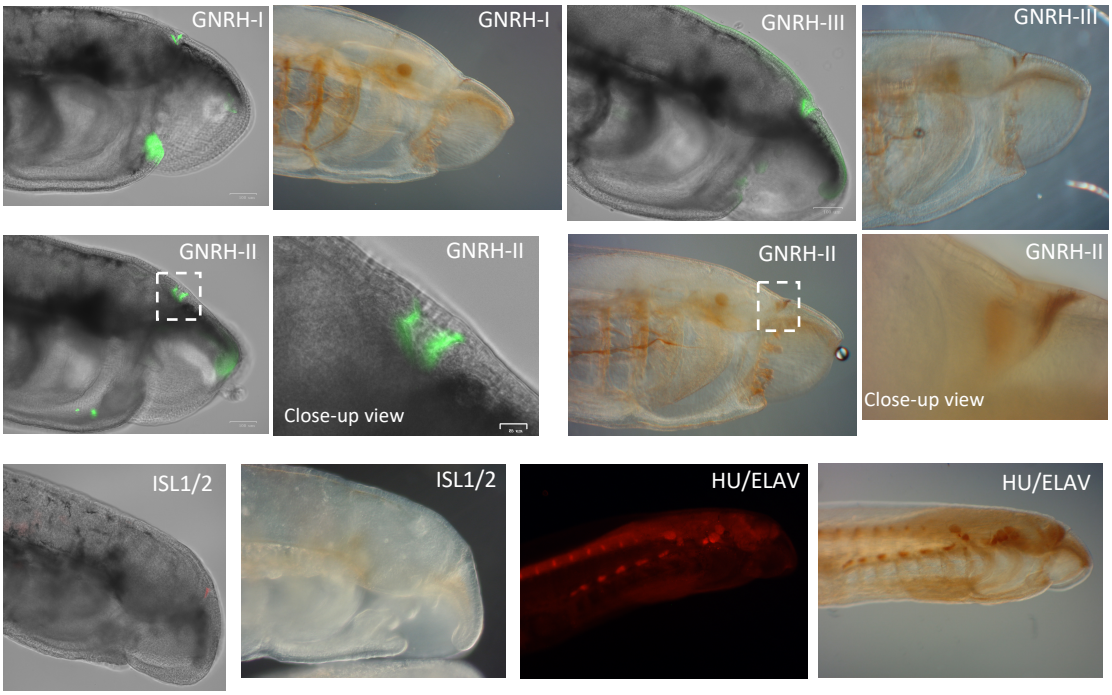


Figure 2.S1. Lamprey-specific GnRH and mouse ISL1/2 antibodies do not work on whole-mount *L. planeri* embryos. Immunolabelling of LpGnRH-I, LpGnRH-II, LpGnRH-III, and ISL1/2 on whole-mount embryos at T29/30 with fluorescence or DAB staining. Some staining was seen in the nasal opening but most likely represent background trapping from this cavity rather than real expression as it does not correspond to any RNA *in situ* hybridization or immunocytochemistry data. Controls with the pan-neuronal Hu/ELAV antibody was run alongside as positive control with either fluorescence or DAB staining and show labelled neurons in the olfactory epithelium and cranial sensory ganglia. Scale bars : 100 µm.

2.4 Discussion

2.4.1 GnRH neuronal expression in lamprey

The marker genes studied here were selected based on their role in the development and migration of hypothalamic GnRH1 neurons and knowledge of their role in the developing olfactory system of jawed vertebrates. The conservation of orthologous gene expression in the same embryonic territories across vertebrates is likely indicative of conserved function. Although the lamprey genome appears to contain the basic set of regulatory genes required for normal patterning of hypothalamic GnRH neurons from the nasal placode, the orthologous genes were not expressed in the patterns that comparative embryology between this animal and gnathostome would necessarily predict. In fact, the lamprey data shown here identify notable difference in the expression of those marker genes as well as similarities. This study provides the first *in situ* hybridization data for all three *GnRH* genes (*GnRH-I*, *-II* and *-III*) in embryonic lampreys. Previously, the only report of GnRH expression in embryonic lamprey was made by immunocytochemistry with lamprey-specific antisera for GnRH-I and GnRH-III, but not for GnRH-II as it was not yet discovered (Tobet et al., 1996). This earlier study stated that GnRH-I and GnRH-III neurohormones mirrored each other in localisation and distribution with the first detectable expression caudal to the telencephalon in the preoptic area and anterior hypothalamus at Piavis stage 17 (equivalent to Tahara stage 29). Sectioning revealed that the immunoreactive GnRH cells were restricted to the cell dense periventricular zone of the diencephalon (Tobet et al., 1996). Here, the RNA expression gathered for LpGnRH-I and -III support those former immunostaining data as we also report expression for both genes in the preoptic area of the hypothalamus starting with a discrete expression at stage 28. These observations bring into line previous dual-label *in situ* hybridization in

adult lampreys which demonstrated that GnRH-I and -III mRNA co-localized in the same cells inside the preoptic nucleus/hypothalamic regions (Root et al., 2005; Van Gulick et al., 2018). Previous study by immunocytochemistry in adult lampreys did not determine whether GnRH-II co-localizes in GnRH-I and/or -III cells, or presents in different GnRH cells, because the anti-GnRH-II used had a slight cross-reactivity to both GnRH-I and -III (Kavanaugh et al., 2008). Here, the *in situ* data clearly demonstrated that *LpGnRH-I* and -*III* localize to a different territory than *LpGnRH-II* expressing neurons during embryogenesis. The fact that *L. planeri* embryos have their three GnRH genes expressed in the hypothalamus contrasts with jawed vertebrates' embryos which have distinct GnRH neuronal populations located in the preoptic area of the hypothalamus (GnRH1 neurons), the midbrain (GnRH2 neurons) and the terminal nerve ganglion near the olfactory bulb (GnRH3 neurons). Another major difference between lamprey and jawed vertebrates is the apparent heterochrony in the timing of appearance of the hypothalamic GnRH neurons as in lamprey GnRH transcripts appear much later in development compared to jawed vertebrates, at T28 and directly inside the hypothalamus rather than the nasal placode. However, in adult lampreys the distribution and localization of GnRH-I, -II, -III is far more complex than during embryogenesis with numerous localizations which vary depending on the sex of the individual and stage of the life cycle (amnocoete larva, parasitic and adults) (Van Gulick et al., 2018). Considering this in detail, over the past three decades there have been several studies assessing the localisation of GnRH neurons in larval, parasitic, and adult lampreys with lamprey-specific GnRH antisera or via single, dual, or even triple *in situ* hybridization. The summary Table 2.3 illustrates the major brain territories observed to express each type of lamprey GnRH paralogues from current and past studies.

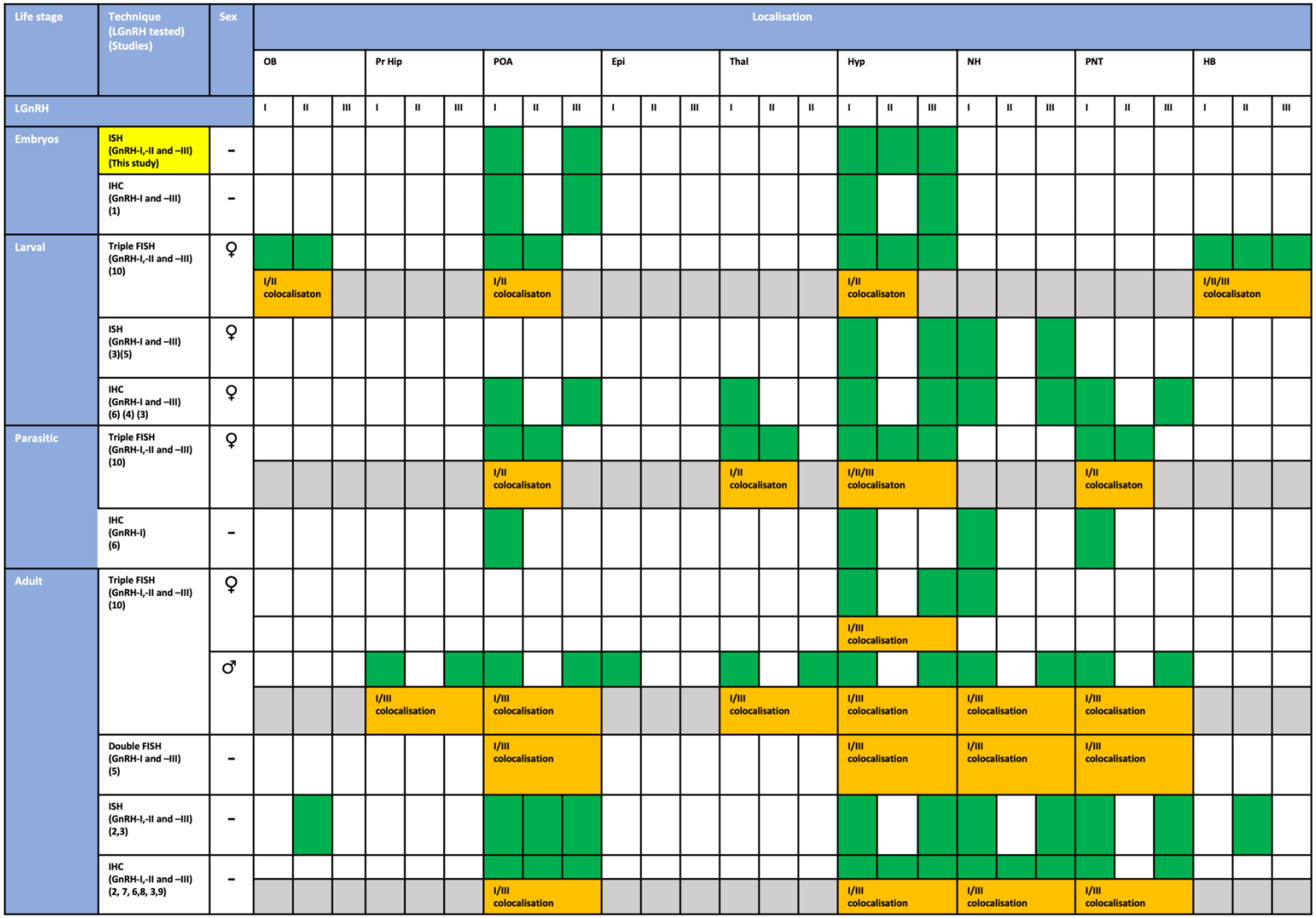


Table.2.3. Summary of expression results from current (highlighted in yellow) and past studies ([1]Tobet et al., 1996; [2]Kavanaugh et al., 2008; [3]Reed et al., 2002; [4]Tobet et al., 1995; [5]Root et al., 2005; [6]Wright et al., 1994; [7]King et al., 1988; [8]Nozaki et al., 2000; [9]Youson et al., 2006; [10]Van Gulick et al., 2018). Presence of expression is in green and existence of colocalisation in the same cells in orange. Abbreviations: FISH, fluorescent in situ hybridization; ISH, in situ hybridization; IHC, immunohistochemistry; LGnRH, Lamprey gonadotropin-releasing hormone; OB, olfactory bulb; Pr.Hip, primordial hippocampus; POA, preoptic area; Epi, epithalamus; Thal, thalamus; Hyp, hypothalamus; NH, neurohypophysis (posterior pituitary); PNT, preoptico-neurohypophyseal tract; Hb, Hindbrain. Figure adapted from Van Gulick et al., 2018.

2.4.2 Evolution of the developmental processes of vertebrate hypothalamic GnRH neuron specification: a hypothesis

The hypothalamic GnRH neurons, consisting of cells feeding hormones to the adenohypophysis (hypophysiotropic), are found in all vertebrates and form the reproductive hypothalamus-pituitary-gonadal axis, a complex system typically not found in invertebrate chordates. In mammals, the hypothalamic GnRH neurons are derived from progenitors within the olfactory placodes including neural crest cells in some species (Shan et al., 2020). The patterning programs acting during the development of the hypothalamic GnRH neurons in gnathostomes raise an intriguing question as to how such programs might have been acquired in evolution. To address this, it was first necessary to understand the possible primitive state of GnRH development in vertebrates by studying the lamprey. In the present study, some of the developmental elements found in the patterning of olfactory neurons in jawed vertebrates appeared to be conserved during lamprey development. For example, in *L. planeri* some of the *ISL* and *COE* paralogues, i.e., *LpISLC* and *LpCOE-A*/*-B* have restricted expression to the anterior part of the NHP placode, corresponding to the future olfactory epithelium (Lara-Ramírez et al., 2017; this study). In mice, at embryonic stage E11.5, *ISL1* mRNA is mainly confined to the GnRH neurogenic area in the ventral portion of the developing olfactory pit proximal to the vomeronasal organ but is only sparsely detected in the olfactory epithelium. *ISL1* transcripts and proteins co-localise with migrating GnRH neurons on their route from the olfactory placode to the brain (Taroc et al., 2020, Shan et al., 2020). Moreover, at that embryonic stage in mouse, *COE2* is expressed broadly in the olfactory epithelium but also co-localises with migrating GnRH neurons (Corradi et al., 2003; Magdaleno et al., 2006). In the brook lamprey, *LpFGF8/17*, the *FGF8* orthologue is expressed in the NHP and

immediate surroundings of it in the anteroventral telencephalon, suggesting its potential involvement in GnRH differentiation from this region like in gnathostomes, where *FGF8* is expressed in the respiratory epithelium next to the OPs and plays a role in OSNs neurogenesis and GnRH neurons specification (Kawauchi et al., 2005; Chung et al., 2016). Finally, the drug tests revealed that the lamprey hypothalamic LpGnRH-III neurons are sensitive to FGF and RA signalling during specific time windows with antagonistic effects, similar to the specification mechanism of hypothalamic GnRH1 cells in the olfactory placode of chick (Sabado et al., 2012). All these similitudes may lead to the impression that some hypothalamic GnRH neurons also arise in and migrate from the nasal placode in lampreys, however several lines of evidence collected during this study point otherwise. First, in *L. planeri LpNELF is* never seen in association with the olfactory system contrary to jawed vertebrates like mice where *NELF* transcripts are observed in the olfactory epithelium and within migrating GnRH neurons exiting the olfactory pit to the forebrain (Kramer and Wray, 2000). However, *LpNELF* is expressed in same territories as LpGnRH-II neurons at T28 which suggest that they may also be implied in the development of some hypothalamic GnRH cell population in lampreys. Based on the expression at earlier stages at T26-27, I would hypothesise that LpNELF-expressing cells may migrate from the ventral hypothalamus to the mature location of LpGnRH-II neurons. Second, in human, pluripotent stem cells differentiated into GnRH1 neurons under FGF8 signalling have upregulated expression of *ISL1* (Lund et al., 2020). In lamprey, *LpISLA* is a gene marker expressed in the same position as LpGnRH-I and -III neurons at late pharyngula (T28), i.e., in the preoptic area of the hypothalamus, but was never seen expressed in the NHP at any stages. In fact, the *LpISLA* expressing cells in the forebrain are first detected at early pharyngula (T23-24) in the anteroventral telencephalon likely under the influence of LpFGF8/17 signalling in the NHP and

anteroventral telencephalon at those stages. I would thus speculate that GnRH-I/-III neurons arise from the anteroventral telencephalon upon induction of FGF8/17 signalling. In fact, a telencephalic origin for GnRH neurons has been described in medaka however for two non-hypothalamic GnRH1 population: a dorsal preoptic population that migrates from the dorsal telencephalon and a medial ventral telencephalic population that migrates from the anterior telencephalon (Okubo et al., 2006).

Differences in gene expression could be caused by heterochronic changes, for example lamprey cells might be specified in the olfactory placode and migrate into the brain, but only become visible in gene expression studies when their delayed expression of these marker genes commences. However, in addition of these differences in the expression of GnRH-related gene markers, DiI-dye cell lineage tracing data indicated that GnRH neurons in lamprey are not derived from NHP precursor cells as no migration from the placode was detected. This corroborates previous investigation in sea lamprey embryos via immunocytochemistry which observed overlap between the olfactory and GnRH systems at the level of fibre projections only, as no cells of apparent placodal origin were seen entering the region of the preoptic area or hypothalamus that contained GnRH neurons (Tobet et al., 1996). Whether the presence of migratory hypothalamic GnRH cells from the nasal placode is a gnathostome novelty or its absence a lamprey specificity remains an open question (though see consideration of tunicate data below). However, if we agree to the postulate that lampreys embody the ancestral vertebrate state, then it supposes that a heterotopic change happened in jawed vertebrates with a shift from a putative CNS origin to a placodal origin for hypothalamic GnRH neurons. It should be noted that it remains possible that neural crest cells contribute to the appearance of hypothalamic GnRH neurons in lamprey which has not been investigated in this study.

Overall, the observed differences in gene expression patterns between gnathostomes and lamprey may possibly be linked to the differences in embryonic development of hypothalamic GnRH cells. Irrespective of such differences in developmental patterns, apparently homologous sets of genes are involved in olfactory/GnRH development in both animal groups, even though some are expressed in different places and at different times.

Tunicates, the sister group of vertebrates, share common basic features with vertebrates, not only in larval morphology, but also in embryonic developmental mechanisms (Passamaneck and Di Gregorio, 2005). From a comparison of the gene expression patterns between tunicate and vertebrates, the anterior ectoderm in *Ciona* may possibly contain the origin of the sensory placode that is homologous to the gnathostome olfactory placodes. In fact, in the ascidian *C. intestinalis*, some GnRH neurons present in the larva have striking similarities with nasal placode-derived hypothalamic GnRH neurons (Poncelet and Shimeld, 2020). For instance, the GnRH cells localized in the palps, i.e., the primary sensory neurons (PSNs) and the axial columnar cells (ACCs) arise from an anterior proto-placodal territory (cf. Chapter 3). In particular, the ACCs have a dual GnRH neurosecretory and chemosensory function (cf. Chapter 3). Furthermore, these GnRH palp cells require induction by FGF signalling for their development (Wagner et al., 2014) while RA repress their formation as observed for some hypothalamic GnRH neurons in jawless and jawed vertebrates (Yagi and Makabe, 2002; Sabado et al., 2012; this study). Furthermore, the *Ciona* cognates of ISL and COE TFs are necessary for the proper development of the GnRH-expressing palp ACCs (Wagner et al., 2014; Mazet et al., 2005; Johnson et al., 2020). NELF is absent in *Ciona* but likely due to secondary loss as it is present in the genome of amphioxus, a more basal chordate. Overall, the spatiotemporal

expression patterns of these regulatory genes hint that the basic developmental patterning mechanisms of the anteriormost GnRH/olfactory neurons may have already been established at molecular levels in the common ancestor of Olfactores. If we assume that lamprey and *Ciona* represent respectively the ancestral state of the vertebrate and the Olfactores common ancestors, we can then draw a model for the evolution of vertebrate hypothalamic GnRH neurons. The phylogenetic tree below indicates the probable timing of the acquisition of morphological characteristics in relation to different expression patterns studied (Fig. 2.19).

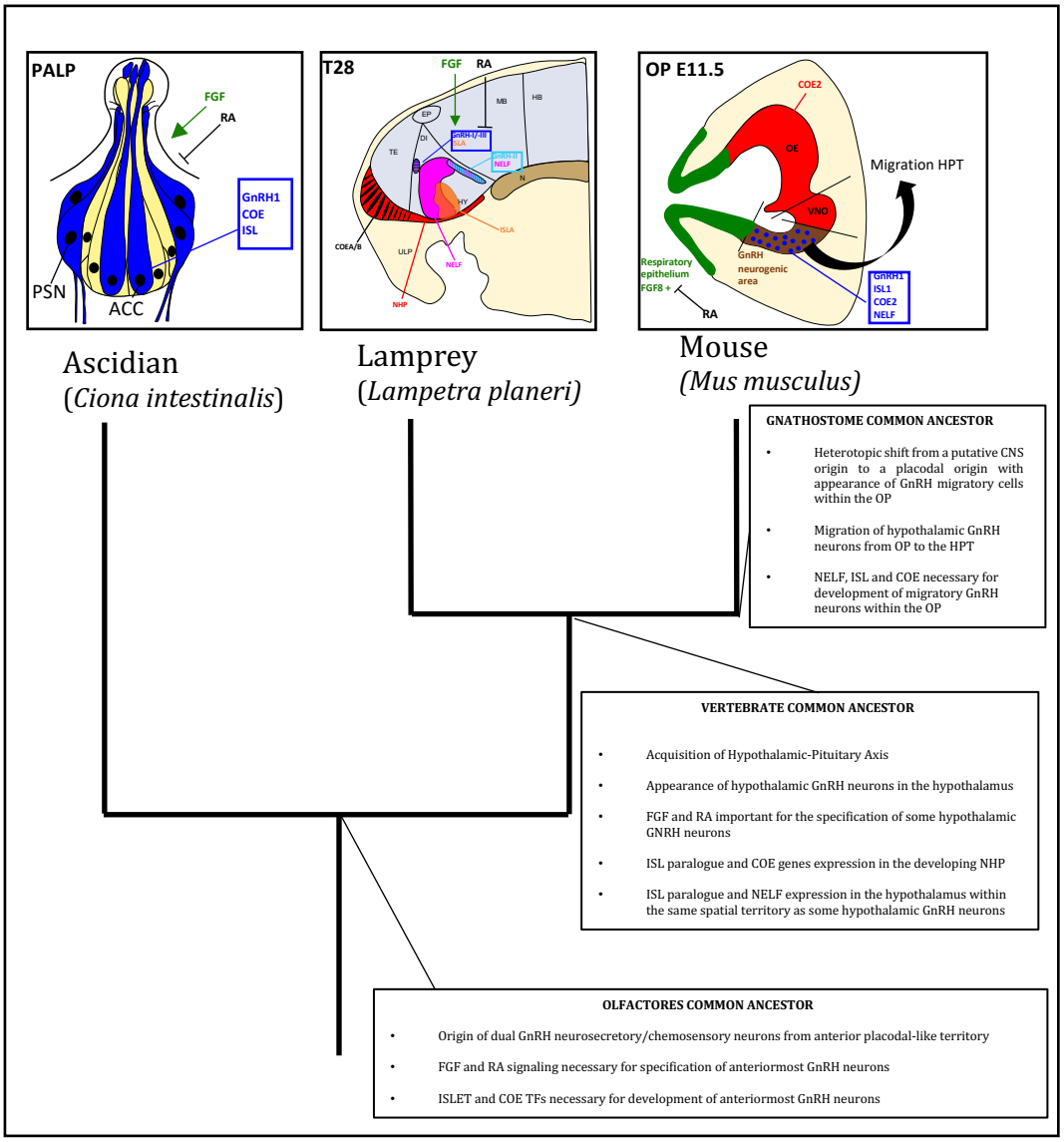


Fig. 2.19. Evolution of the developmental processes of vertebrate hypothalamic GnRH neuron specification: a hypothesis. Comparison of the developmental process of the hypothalamic GnRH neurons in mouse (*Mus musculus*) and lamprey (*Lampetra planeri*) as well as the anteriormost palp GnRH neurons in the ascidian *Ciona*. In *Ciona*, the dual GnRH/chemosensory ACCs localised in the palps (blue) arise from an anterior proto-placode territory and require FGF and RA signalling molecules as well as the regulatory TF genes ISL and COE for their proper development. In lamprey, GnRH neurons appear *in situ* within the hypothalamus not in the NHP. NELF is present in the same spatial territory as GnRH-II (light blue) neurons while ISLA is present in same spatial territory as GnRH-III neurons (dark blue) hinting a possible colocalisation in the same cells. Also, FGF and RA are required for the specification of GnRH-III neurons. In mouse, hypothalamic GnRH1 neuron progenitors are specified within the nasal placode upon transient FGF8 signalling which can be inhibited by RA. The specified GnRH neurons migrate secondarily to their mature location in the hypothalamus. The migrating GnRH neurons express the marker COE2, NELF and ISL1. Abbreviation: di, diencephalon; ep, epiphysis; hb, hindbrain; hpt, hypothalamus; hy, hypothalamus; mb, midbrain; n, notochord; oe, olfactory epithelium; op, olfactory placode; ulp, upper lips; te, telencephalon; vno, vomeronasal organ.

The model above highlights the apparent conundrum to explain the evolutionary origin of hypothalamic GnRH neurons in vertebrates. Data in *Ciona* embryonic larva suggest that the anteriormost GnRH neurons with molecular signatures similar to olfactory-derived hypothalamic GnRH cells, i.e., expressing GnRH, COE, ISL and susceptible to FGF and RA signalling, already had a placodal-like origin in the last common ancestor of Olfactores as seen in modern jawed vertebrates. Although, some GnRH-expressing neurons also derive from the neural plate in *Ciona* as *CiGnRH1* and *CiGnRH2* positive cells are present in the brain vesicle, a homologous structure to the vertebrate forebrain/midbrain (Kusakabe et al., 2012). In addition, ISL and COE TFs are also expressed in the brain vesicle, although it is unknown whether these markers co-localise with brain vesicle GnRH neurons like seen in the palp $GnRH^+$ ACCs (Giuliano et al., 1998; Mazet et al., 2005). Lampreys represent an oddity in extant vertebrates as all hypophysiotropic GnRH cells seem to appear directly in the CNS within the hypothalamus. In tetrapods, such as in mammals, birds and amphibians, the forebrain GnRH populations all originate from precursors in the olfactory placodes (Murakami et al., 1992; Yamamoto et al., 1996; Shan et al., 2020). However, exception to this peripheral origin can be found in teleosts. For example, in medaka, although the hypophysiotropic GnRH neurons localised in the preoptic area emerge also from the olfactory placodes, there also exists some non-migratory hypothalamic GnRH neuron populations that appear within the hypothalamus (Takahashi et al., 2016). Also, in cichlid fish, the population of hypophysiotropic GnRH neurons in the preoptic area of the hypothalamus appears to originate directly from the rostrobasal preoptic recess, although it was not clear whether they migrate from other brain or peripheral territories prior to GnRH expression (Pandolfi et al., 2002). If lampreys really embody the ancestral vertebrate state, then it would mean that the forebrain GnRH cells would have lost their initial placodal-type origin previously existing

in the Olfactores ancestor via the co-option of some components of the anteriormost GnRH specification network involving FGF, RA, COE and ISL in tissue of the hypothalamus. This new CNS origin would have been reverted to a placodal one in the ancestor of gnathostomes again via the co-option of a regulatory network involving FGF8, RA, ISL1, COE2 and NELF, maybe facilitated by the fact that orthologues of genes like FGF8, COE and ISL where already expressed in the nasal placode of the vertebrate ancestor as suggested by their expression there in lamprey. This complex evolutionary scenario is not a parsimonious model as it implies some sort of back-and-forth tissue origin from a placode territory to CNS and then back to some anterior placodal origin with a subsequent complex migration process prone to errors, as observed in Kallmann syndrome. The existence of anterior GnRH cells either derived from the olfactory placodes or the neural plate in ascidian and teleost may rather suggest that such dual origin for forebrain GnRH neurons was present in the last common ancestor of vertebrates, with the loss of a placodal population explaining the state in lamprey. Further investigation would be necessary to confirm that this model is a genuine reflect of past evolutionary events. More discussion on how this enigma could be addressed in further research can be found in the General Discussion (see Chapter 5).

Chapter 3 GnRH and MS4A trace a common origin of *Ciona* ACCs and vertebrate olfactory neurons

Work Declaration

The experimental work and subsequent data analyses presented in this Chapter have performed by me, Guillaume Poncelet.

3.1 Introduction

3.1.1 Research Context

Comparative genomics, gene expression studies and the findings of conserved gene regulatory networks provide valuable tools for inferring the evolutionary history of physiological systems. This is particularly true when this information is difficult or even impossible to ascertain by morphological traits only. One such example is the vertebrate olfactory system, the nasal chemosensory system that is responsible for detecting environmental odorants as well as intraspecific pheromonal cues. The morphological components of the olfactory system are found only in vertebrates, but some genetic components of the system have been found in protochordates, in addition to vertebrates (cf. Chapter 1). To this day it is still unclear how the olfactory systems of vertebrates originated and how their different neuronal cell types diversified during evolution.

Are the different vertebrate olfactory cell types recent innovations, or do they predate them? Answering such questions is key to understanding overall chordate evolutionary history, since olfactory-derived sensory and secretory cells are assumed to have played a crucial role in the evolution of vertebrates (Poncelet and Shimeld, 2020). Therefore, identifying and characterising potential olfactory cell homologs in tunicates and

cephalochordates is paramount to further resolving the evolutionary origins of the vertebrate olfactory systems.

3.1.2 Protochordates: characteristics, evolution, and development

The origin and evolution of chordates is an exciting source of scientific research as it can help us to understand the appearance of complex vertebrates such as ourselves. The phylum Chordata are defined by five anatomical features at some period of their life cycle: a notochord, a dorsal nerve cord, pharyngeal slits, an endostyle and a post-natal tail (Rychel et al., 2006). Furthermore, chordates are divided in three major subphyla: cephalochordates, tunicates, and vertebrates (Haeckel, 1874), the two first grouping as the paraphyletic "protochordates".

The cephalochordates are represented today by the ~35 extant amphioxus species. Amphioxi are typical filter-feeders which live burrowed in sand or mud. Amphioxus anatomy includes typical chordate characteristics such as a dorsal nerve cord, notochord, pharyngeal slits, myomeres and a post-natal tail. The dorsal nerve cord is slightly enlarged in the anterior head, forming a brain-like blister (Satoh et al., 2014). Overall, the amphioxus body plan is reminiscent of the oldest chordate fossils known-to-date from the early Cambrian period, i.e., earlier than 500 million years ago (Ma) such as *Pikaia* and *Cathaymyrus* (Conway Morris, 1982; Shu et al., 1996). Therefore, amphioxi are crucial research models because they have presumably retained most of the developmental pathways found in the extinct ancestors of chordates. Phylogenetic analysis has indeed classified amphioxus as evolutionary more distant to vertebrates than the fast-evolving tunicates despite what might be suggested by cephalochordate anatomy at first glance

(Delsuc et al., 2006). As a result, the basal position of amphioxus in the chordate phylogeny makes it the best living proxy to what the vertebrate-chordate ancestor might have looked like.

The tunicates (or urochordates) are animals enclosed with a "tunic" made of tunicin, a type of cellulose (Lamarck, 1816). The capacity for cellulose synthesis was most-likely acquired through horizontal gene transfer from a bacterial cellulose synthase (Nakashima et al., 2004). Most tunicates show a chordate body plan only as a larval form, where they have a mobile larva with a notochord resembling a tadpole (Satoh, 1994; Satoh, 2009). The earliest fossil of tunicates such as *Shankouclava* are from the early Cambrian period about ~520 Ma (Chen et al., 2003). Today, they are approximately ~3,000 extant species of tunicate which are all specialist marine filter feeders classified into three distinct classes: Larvacea and Thaliacea and Ascidiacea (Satoh et al., 2014).

The Larvacea (or Appendicularia) are solitary planktonic tunicates found in the surface water of all oceans. They are the only tunicates which retain a larval form with chordate characteristic even at the adult stage (Satoh et al., 2014). The larvaceans are highly derived among tunicates and went through massive genome reduction. In fact, *Oikopleura dicoia*, the current molecular model in the larvacean class has the smallest genome of any chordate known to date at about 50-65 megabases (Seo et al., 2001). This massive genome reduction can be explained by gene loss, rarity of mobile elements and shortening of intergenic distances. *O. dicoia* is actually considered the fastest evolving metazoan recorded so far (Berna and Alvarez-Valin, 2014). The tendency to genome reduction and reorganisation is not only seen in larvaceans but also in ascidians (Dehal et al., 2002).

Members of the Thaliacea are colonial planktonic tunicates with pelagic individuals moving by jet propulsion. They lack a free-swimming larval stage and have a distinct direct development (Bone, 1998). Molecular phylogenetics suggests that thaliaceans and the ascidians form a monophyletic group while the fast-evolving larvaceans are the sister group of all the other tunicates (Kocot et al., 2018).

The Ascidiacea (or ascidians) are extremely diverse and form the overwhelming majority of tunicate species. The ascidians can be either solitary or colonial. The Ascidiacea have a planktonic larva (tadpole-like) which is composed of ~2,500 cells. Larvae rapidly metamorphose into a sessile adult after finding a suitable settlement site (Satoh, 2009). Despite being highly derived organisms with a very different body plan in the adult stage, the ascidians such as the research model *Ciona intestinalis* are invaluable to study the evolution and development of key vertebrate innovations such as the appearance of placodes and neural crest cells (Abitua et al., 2015). In fact, the simplicity of ascidian embryos and genomes makes them well-suited to decipher the gene regulatory network controlling embryonic development in chordates (Kubo et al., 2009). In addition, the tunicates and vertebrates form together a sister group known as the "Olfactores" (Jefferies, 1991). This latter terminology is a bit misleading because it is suggestive that olfaction only arises in this group of chordates, which is not what current evidence suggest in cephalochordates.

3.1.3 The ascidian larva: a swimming journey for a life of settlement

When solitary or colonial ascidians reproduce sexually, they release eggs and sperm from their atrial siphons and subsequent fertilisations occur in the water column (although, note that in some colonial species, fertilisation and development of eggs can occur directly inside the atrium (Gasparini and Ballarin, 2018)). Once fertilised, the eggs develop into swimming tadpole larvae which last for a brief planktonic period (Holland et al., 2016). The whole life purpose of an ascidian larva is to swim to an appropriate settlement site to metamorphose into its final form, a sessile filter-feeding adult. This peculiar process is described as "retrograding metamorphosis" as quoted by Charles Darwin in his book, the Descent of Man: "We should thus be justified in believing that an extremely remote period a group of animals existed, resembling in many respects the larvae of our present Ascidians, which diverged into two great branches- the one retrograding in development and producing the present class of Ascidians, the other rising to the crown and summit of the animal kingdom by giving birth to the Vertebrata" (Darwin, 1871).

To find an appropriate site for settlement, the ascidian larval head is equipped with multiple sensory structures such as the sensory vesicle, which includes the otolith, an organ used to sense gravity, and the ocellus, a pigmented light-sensing organ (Sakurai et al., 2004). The larval head has also, at its anteriormost part, three protruding organs in a triangular arrangement known as the palps (or papillae) (Fig. 3.1-A). The palps are of particular interest because of their critical role in the settlement of the larva on a chosen substrate and its role in the initiation of metamorphosis (Cloney, 1982). Despite various organisational differences in different ascidian species, each palp contains three distinct

cell-types: collocytes (CCs), primary sensory neurons (PSNs) and axial columnar cells (ACCs) (Dolcemascolo et al., 2009). The CCs are elongated secreting cells known to produce adhesive substances so the larva can fix itself into a chosen substrate. This cell type is subject to recent investigation for the creation of bioglue and for strategies to create antifouling agents (Zeng et al., 2019b). The ACCs have long microvilli emerging from the apical edge and extending throughout the hyaline cap. The PSNs have a cilium at the apical side and an axon proceeding from the basal side (Zeng et al., 2019a). The precise functions of the ACCs and PSNs are unknown. It is likely that some of the cells in the palps are involved in testing out substrates before settlement, perhaps by mechano- and/or chemo-reception (Pennati et al., 2009). In support of this hypothesis, are the findings that ascidian larvae have preference for surfaces coated with different texture and chemicals, some of which are produced by specific marine bacteria (Szewzyk et al., 1991; Holmström et al., 1992; Groppeli et al., 2003; Flores and Faulkes, 2008; Zeng et al., 2019b). Once a suitable surface is found, the larva everts its palps and attaches itself with adhesive secretion from the CCs. This event triggers an irreversible metamorphosis in which various organs are lost, such as the tail and fins, or rearranged in their final adult position (Gasparini and Balarin, 2018)

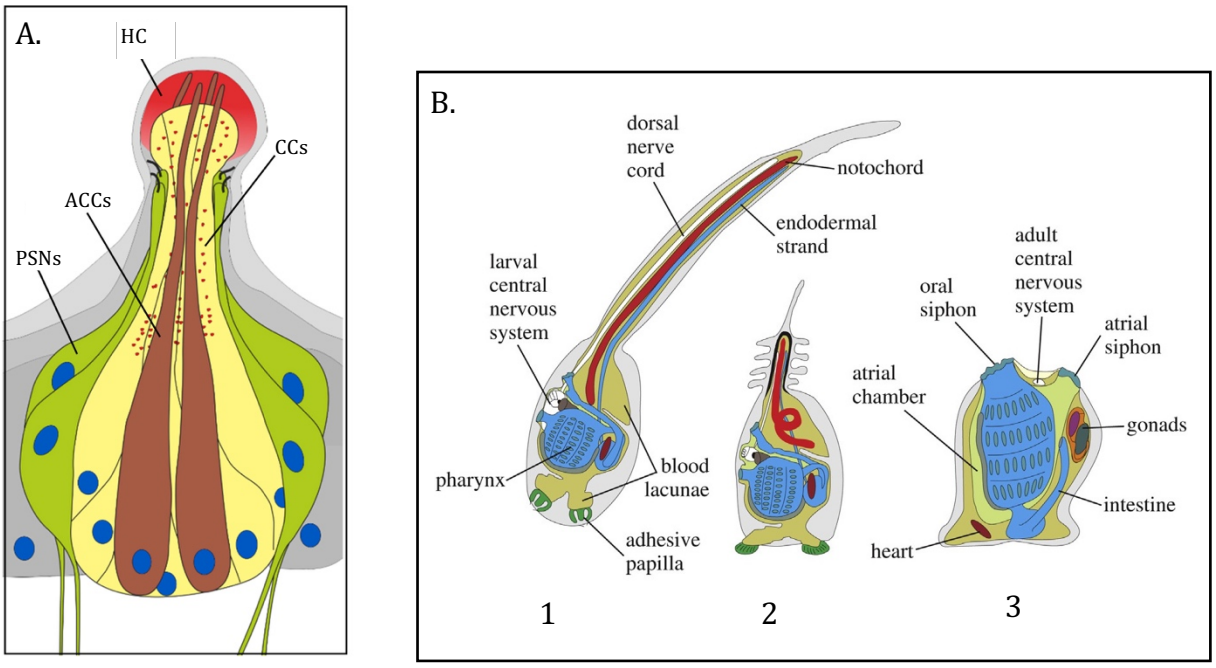


Figure 3.1. (A) Scheme of a single palp and its associated cell types in *Ciona intestinalis*. Four ACCs in a mid-central position (brown) are surrounded by CCs (yellow) that contain abundant granules with adhesive (red) that accumulates in the apical hyaline cap (hc, red). The ACCs reach their apical protrusions far into the hyaline cap. ACCs and CCs are flanked by four PSNs (green) with cilia protruding from their apical thickenings (black lines). Epithelial cells (dark gray) surround the base of the organ and most nuclei are in the basal part of the cell, while the entire organ is covered with two layers of tunic. **(B)** Ascidian metamorphosis. A swimming tadpole larva (1) adheres to the substrate (2) and metamorphoses into a sessile oozooid (3). In (2) the tail is retracting, and the axial organs are pushed into the head. Figures are reprinted from (A) Zeng et al., 2019a and (B) Gasparini and Ballarin, 2018, with permission from Elsevier.

3.1.4 Elusive chemosensory cells in tunicates

The ability to detect and discriminate chemical cues is vital for animals. This is the reason metazoans have occasionally evolved complex chemosensory organs such as the olfactory system of vertebrates, which express a large range of chemosensory receptors (Bear et al., 2016). Although well characterized in some animals, chemosensory receptors remain largely undescribed in protochordates and more generally in many metazoan lineages (Marquet et al., 2020). In vertebrates, the different subtypes of OSN can be identified by the type of chemoreceptors they express. As said in Chapter 1, the

orthologues of vertebrate's ORs, TAARs and VR1/2s family genes could not be identified in tunicate genomes. The apparent lack of chemoreceptor genes in tunicates is rather unexpected and has hindered the discovery of putative chemosensory cells in that lineage. This oddity suggests that chemosensory cells of tunicates, if they exist, harbour an as-yet unrecognised receptor type, whose identification could reveal key features of the functional organisation and neural logic that govern larval settlement and subsequent metamorphosis (Churcher and Taylor, 2009; Eyun et al., 2017; Grus and Zhang, 2006; Kamesh et al., 2008; Libants et al., 2009; Niimura, 2009; Nordstrom et al., 2008).

3.1.5 GnRH systems in ascidians

In the solitary adult ascidian, the GnRH immunoreactive (GnRH-ir) cell bodies and nerve fibres are located near or within the cerebral ganglion, neural gland, visceral nerves, dorsal strand, dorsal blood sinus, dorsal strand plexus, gonoducts and gonads (Mackie, 1995; Powell et al., 1996; Tsutsui et al., 1998). In fact, GnRH peptides are released in the circulation and act directly on the gonads to induce spawning (Terakado, 2001). Thus, there is a conserved role of GnRH for reproduction in Olfactores which appeared before the appearance of the vertebrate pituitary (Powell et al., 1996). The release of GnRH from neurons was shown to be light-regulated which ensures a precise biological clock for reproduction in ascidians (Tsuda et al., 2001). It was reported that GnRH neurogenesis occurs in the dorsal strand epithelium. Speculation was made that GnRH neurons delaminate and migrate from the dorsal strand epithelium up to the cerebral ganglion via the visceral nerve. This hypothesis would mean that GnRH migration to the brain from a peripheral organ predates vertebrates. In fact, the dorsal strand was said to possibly be homologous to vertebrate olfactory placode as this peripheral organ is derived from the

anterior region of the embryonic neural plate (Terakado, 2009). However, no cell lineage tracing data exist to support that scenario of GnRH migration in ascidians.

In the larva of ascidian such as *Ciona intestinalis*, there are two *GnRH* genes (*CiGnRH1* and *CiGnRH2*) which have some differential expression, suggesting distinct GnRH-producing cell populations (Fig. 3.2) (Adams et al., 2003; Kusakabe et al., 2012). Note that *CiGnRH1* and *CiGnRH2* are not necessarily orthologous genes to vertebrate GnRH1 and GnRH2, respectively. Nevertheless, a non-reproductive role for GnRH cells was said to be conserved between tunicates and vertebrates. In fact, *CiGnRH2* was found expressed in the motor ganglion and nerve cord, homologous structures to the vertebrate hindbrain and spinal cord, respectively. Homologies between the CiGnRH2 system and the GnRH2 system of vertebrates was strengthened by the discovery that *GnRH2* of medaka (*Oryzias latipes*), is expressed in the hindbrain and spinal cord during development (Kusakabe et al., 2012). However, the anterior apical trunk epidermal neurons (aATENs) also express *CiGnRH2* but homologies have been suggested to the olfactory-derived GnRH1/3 neurons of vertebrate instead (cf. Chap 1) (Abitua et al., 2015). However, little is known about putative homologies that could exist between the CiGnRH1-producing cells of *Ciona* and the vertebrate GnRH neurons. In *Ciona* larva, the GnRH peptides have been shown to induce metamorphosis upon GABA induction which is a trait conserved with vertebrates (Hozumi et al., 2020). In fact, GABA is an important excitatory regulator of GnRH1/3 neurons in the hypothalamus of vertebrates (Watanabe et al., 2014).

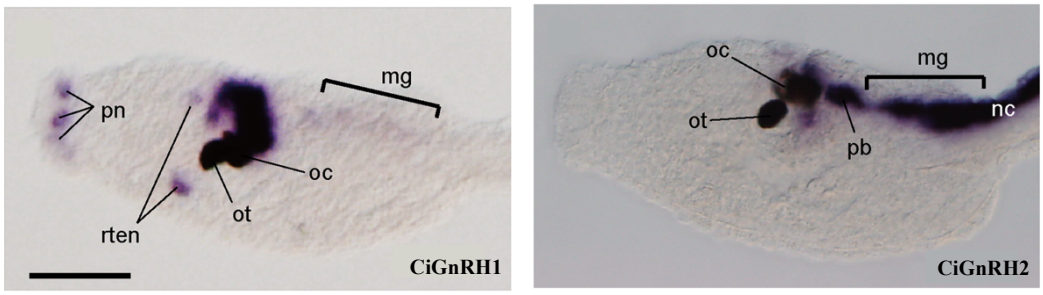


Figure 3.2. Expression by *in situ* hybridization of *CiGnRH1* and *CiGnRH2* in the *Ciona intestinalis* larva. Abbreviation: bv, brain vesicle. nc, nerve cord. mg, motor ganglion. oc, ocellus. ot, otolith. pb, posterior brain. pn, palp, neuron. rten, rostral trunk epidermal neuron. Scale bar, 50 μm. Images taken from Kusakabe et al (2012).

3.1.6 Chapter Aim and Summary

As the palps are one of the best candidate organs for hosting the *Ciona* olfactory cells, this Chapter aimed to identify regulatory and functional homologies between the ascidian palp cells and the vertebrate olfactory neurons. To this end, I characterised the regulatory region and expression of chosen members of the *Ms4a* and *GnRH* gene families in *Ciona*. These molecular markers were selected because they label specific subsets of chemosensory and neurosecretory neurons arising from the vertebrate olfactory system. First, I dissected the ~4.7kb regulatory region of *CiGnRH1* gene known to be expressed in the palps of *Ciona*. I report the presence of a core region of 190bp, which is sufficient to drive expression in all the territories seen by the initial ~4.7kb cis-regulatory element. A conserved gene regulatory network (GRN) was then inferred by prediction of putative transcription factor binding sites in that short sequence and comparison was made to a robust GRN for the specification of olfactory-derived GnRH1/3 neurons in vertebrates. Moreover, I report the expression of this newly characterised enhancer in two distinct palp cell-types, the PSNs and unexpectedly the ACCs. Second, I characterised the

expression of a MS4a regulatory region, and report its expression in ACC apical digitiform protrusions extending in the hyaline cap. The ascidian Ms4a proteins show evidence of diversifying selection in their extracellular loops, strongly suggestive of a role as chemoreceptors. Furthermore, I show evidence of a close genomic linkage between *Ms4a* gene clusters and *OR* gene clusters in all chordate genomes investigated, except neopterygians. In conclusion, I report that gene family members encoding GnRH and Ms4a are both expressed in the ACCs of the palps of the swimming larvae, revealing that some genetic components of the present-day olfactory derived chemosensory and secretory neurons existed in the common ancestor of all extant chordates. These findings are important for understanding the evolution of vertebrate olfactory systems and illustrate the utility of the study of tunicates in uncovering the evolution of vertebrate-specific traits.

3.2 Material and Methods

3.2.1 Animals

Animals previously classified as *Ciona intestinalis* belong to two species, previously defined as *Ciona intestinalis* type A and type B, but which are nowadays renamed *Ciona robusta* and *Ciona intestinalis* respectively. The animals used here are *Ciona intestinalis* in the current usage of the name. Adults of *Ciona intestinalis* were collected from Sparkes Yacht Haven, UK and kept in an aquarium with circulating and oxygenated sea water at 16°C until usage. Gametes were liberated by dissection and cross fertilisations performed *in vitro* for 15 min.

3.2.2 Plasmid construction, electroporation, and staining

DNA fragments were amplified from the vector pCi-proGnRH1 -4687/1::EGFP (provided by Dr. Takehiro Kusakabe) by PCR using Phusion polymerase (New England BioLabs). The oligonucleotides were designed according to the sequence of the vector cis-regulatory region and contained desirable restriction sites (Table 3.1). The amplified fragments were cloned inside the pBluescript-eGFP cut with SacI/BamHI or the PCES-LacZ reporter vector (Harafuji et al., 2002) cut with SphI/XhoI, PstI/BamHI or SalI/NcoI. The genomic DNA fragments containing an upstream region of *Ms4a* (KH.C1.484) and *Delta-like* (KH.L.50.6) were amplified from *Ciona* sperm DNA by PCR using Phusion polymerase (New England BioLabs) with a pair of primers containing 15 bp extremities complementary to the vector ends (Table 3.1). The PCR products were combined by In-Fusion Cloning (Takara Bio) using the ISLET>mCherry vector (kindly provided by Dr. Laurence Lemaire) linearized with BssHII/NotI. All the plasmids were used at a final concentration of 50-100 µg per 50 µl in electroporation, as described previously (Kari et al., 2016). X-gal staining of electroporated sea squirt larvae was done as described

(Christiaen et al., 2009). For reporter gene assays, each construct was tested in three independent electroporations, with parallel positive (GnRH1-A>LacZ) and negative controls (empty LacZ vector). The complete plasmid constructs are available as electronic .dna files (see Appendix).

Table 3.1. PCR primer pairs used to amplify the different regulatory regions.

Construct Name	Relative Position to the TSS (bp)	Restriction sites	Length of regulatory region	Primer Pairs
GnRH1-A>LacZ	-4603/+83	SphI/XhoI	4687 bp	Fw 5'-AGA**CTCGAG**TGTAACACACGCATGTT-3' Rv 5'-GTA**GCATGC**ATTGTGAAGTTGGATAAGCGA-3'
GnRH1-B>LacZ	-473/-328	PstI /BamHI	146 bp	Fw 5'-T**CTGCAG**GAGTTGGTTGTCGAGTGCAATTA-3' Rv 5'-AC**GGATCC**TTTTATAATGCAACACAAC-3'
GnRH1-C>LacZ	-2274/+83	SalI/NcoI	2358 bp	Fw 5'-CAA**GTCGAC**AGCGTCCCCCACAGTAATATA-3' Rv 5'-GTA**CCATGG**ATTGTGAAGTTGGATAAGCGA-3'
GnRH1-D>LacZ	-4603/-2275	PstI/BamHI	2329 bp	Fw 5'-AGA**CTGCAG**ATGCTGTAACACACGCATGTT-3' Rv 5'- AC**GGATCC**CTATGGGGCACCTTTATAGT-3'
GnRH1-E>LacZ	-2769/-2275	PstI/BamHI	583 bp	Fw 5'-AGA**CTGCAG**ATGCTGTAACACACGCATGTT-3' Rv 5'-GC**GGATCC**GTGGTTGTTGAATCAGTTAGTTGCAAGACA-3'
GnRH1-F>LacZ	-3440/-2746	PstI/BamHI	604 bp	Fw 5'-AGA**CTGCAG**GCATTTAACTGATGAAGTTTCGCACACC-3' Rv 5'-AC**GGATCC**CTAAAGATGTGTCGTCTTCTCCCCCCTACT-3'
GnRH1-G>LacZ	-4024/-3418	PstI/BamHI	695 bp	Fw 5'-AGA**CTGCAG**GGAGAAGACGACACATCTTTAGCACATATT-3' Rv 5'-GG**GGATCC**GTTGCCCATTTCTTCAGTTCATACATTTTC-3'
GnRH1-H>LacZ	-4603/-4021	PstI/BamHI	495 bp	Fw 5'-GGCA**CTGCAG**TATGAACTGAAGAAATGGGCAACATAAAA-3' Rv 5'-AC**GGATCC**CTATGGGGCACCTTTATAGT-3'
GnRH1-I>LacZ	-2769/-2425	PstI/BamHI	345 bp	Fw 5'-AGA**CTGCAG**GTATGAACTGAAGAAATGGG Rv 5'-GG**GGATCC**AAAAGAAATGAATAAGTGTCC
GnRH1-J>LacZ	-2769/-2580	PstI/BamHI	190 bp	Fw 5'-AGA**CTGCAG**GTATGAACTGAAGAAATGGG-3' Rv 5'-GG**GGATCC**TTAAAAAACCTCAACGGTCC-3'
GnRH1-J>eGFP	-2769/-2580	SacI/BamHI	190 bp	Fw 5'-AGA**GAGCTC**GTATGAACTGAAGAAATGGG-3' Rv 5'-GG**GGATCC**TTAAAAAACCTCAACGGTCC-3'
Ms4a >mCherry	-1102/+45	N/A	1128 bp	5'-**ATTAATTAAGGCGCG**TTCAATAAAACAACCCAGTCACC-3' 5'-**GCCATGGTTGCGGCCG**TTGTTTTTAGGCTACCTTTACAAA-3'
Delta-like>mCherry	-1741/+195	N/A	1937 bp	5'-**ATTAATTAAGGCGCG**GTTTCCGCTATTTCAATCTTAAGCG-3' 5'-**GCCATGGTTGCGGCCT**CGCAACGCTGCAGAGAAG-3'

Nucleotides in bold black represent restriction sites and nucleotides in bold red represent the 15 bp complementary to the vector end for In-Fusion cloning.

3.2.3 Identification of transcription factor binding sites

The analysis of conservation among ascidian putative *GnRH1* regulatory regions was performed on genomic regions from *Ciona savignyi*, *Phallusia mammillata* and *Halocynthia roretzi* using the mVISTA tool (Ratnere and Dubchak, 2009), the *GnRH1* regulatory sequence of *Ciona robusta*, previously known as *Ciona intestinalis* type A was employed as reference. I used LAGAN (global pairwise and multiple alignment of finished sequences) and the following parameters: Conservation identity (65%), minimum Conservation Width for CNS (100 bp), Minimum Y value (50%), Minimum Length for CNS

(100 bp) for the plot. To predict putative transcription factor binding sites (TFBS) in the surveyed 190 bp cis-regulatory region of *CiGnRH1*, I utilized CIS-BP (http://cisbp.ccbr.utoronto.ca) employing *Ciona intestinalis* DNA-binding-domain classes database (Weirauch et al., 2014). I cross-confirmed the presence of those TFBS with the vertebrate JASPAR database (http:// jaspar.genereg.net/) (Khan et al., 2018) and available Selex-seq data on ANISEED which describe the *in vitro* DNA-binding specificity of 139 of the estimated 500 transcription factors encoded in the *Ciona robusta* genome (Nitta et al., 2019).

3.2.4 Synteny analysis

The identification of homologous Ms4A genes in chordates was done by doing a TBLASTN search on the available transcriptome and genome assemblies on NCBI. The Ms4a proteins used for orthologue searches were always from the closest known relative. For larvacean and thaliacean an additional search was done using a hidden markov model with HMMER (https://www.ebi.ac.uk/Tools/hmmer/). Protein family membership of newly discovered Ms4a proteins was confirmed using InterproScan and the characteristic tetraspanning topologies were predicted via TOPCONS (Tsirigos et al., 2015). To map synteny relationships, genes adjacent to the different Ms4a loci in each species were searched using available genome assemblies on the NCBI Genome Data Viewer.

3.2.5 Amino acid conservation analyses of ascidian Ms4a proteins

FASTA format sequences of the specified *Ciona intestinalis and Ciona savignyi* MS4A proteins were downloaded from the ANISEED database. A multiple alignment with a corresponding score for amino acid conservation was obtained using the PRALINE program with default settings where the least conserved amino acids were given a 0 score

and the most conserved amino acids were assigned a 10 (Simossis and Heringa, 2005). Topographical representations of the MS4A family member that was used on the top line of the alignment were made using the Protter program by manually entering the topographical orientation of the protein as predicted by TOPCONS.

3.2.6 Diversifying selection analyses of ascidian *Ms4a* genes

Ascidian Ms4a amino acid and nucleotide coding sequences were retrieved from Aniseed (see Appendix Table 3.S3). When a gene had more than one predicted isoform, the sequence that contained the longest open-reading frame was selected. The consensus topology of each MS4A amino acid sequences was assessed with TOPCONS and only those with a characteristic tetraspanning topology were selected (Tsirigos et al., 2015). A multiple alignment of amino acid sequences was done with Muscle using default parameters. Manual trimming of the alignment was done in AliView (Larsson, 2014) to remove columns containing gaps in majority of sequences. The N- and C-terminal domains predicted via TOPCONS were also removed because of uncertainty over alignment in those regions. Maximum likelihood phylogenetic trees were built in RAxML (GUI, v2.0.3) with the best substitution model suggested (CPREV+G+F) and 100 bootstrap to support nodes. The final dataset consisted of 130 sequences and 100 aligned characters (see in Appendix). Species-specific clades containing recently diverging paralogues were identified in the tree and considered when there were at least four sequences. In each clade selected, the complete coding sequences of the Ms4a genes were re-aligned with Clustal Omega using default parameters and a corresponding codon alignment was obtained using PAL2NAL with all gaps removed (Suyama et al., 2006). The N- and C- terminal regions of the codon aligned MS4a sequences were too variable and hence not included for the subsequent statistical tests. Nucleotide-based phylogenetic

trees were done with MEGAX using Kimura-2 parameter model to obtain branch lengths. The codon alignments (available in Appendix) and corresponding trees were fit to codon-based models implemented in EasyCodeML to estimate non-synonymous to synonymous substitution rate (d_N/d_S) ratios (ω) across paralog sequences (Yang and Swanson, 2002; Gao et al., 2019). These models are: M0 (one average ratio ω), M1a (neutral; codon values of ω fitted into two discrete site classes between 0 and 1), M2a (positive selection; like M1a but with one additional class allowing ω>1), M7 (neutral; value of ω following a β distribution with ω=1 maximum), M8 (positive selection; like M7 with one additional class allowing ω>1). These different models, all assume that ω ratio is the same across branches of the phylogeny but different among sites (codons or amino acids) in the alignment. Likelihood-ratio tests were used to compare the fit of these models to the sequence data. Support for positive selection was identified when M2a provided better fit than M1a or when M8 provided better fit than M7 (Yang and Swanson., 2000). The M1a-M2a comparison is the most stringent but can lack power to detect signatures of diversifying selection compared to the M7-M8 models comparison, which imposes less constraints on the distribution of ω, but may have a higher rate of false positives (Swanson et al., 2003). EasyCodeML was run in preset mode using default settings. The Bayes empirical Bayes (BEB) procedure executed in EasyCodeML was used to identify amino acid residues evolving under positive selection when the likelihood-ratio test had a significant result for any of the pairwise comparisons of codon models. The standard threshold for determining amino acid sites under selection is a posterior probability of 0.95 (Scheffler and Seoighe, 2005). The predicted sites under positive selection were mapped across the predicted Ms4a protein domains using the PROTTER program by manually entering the topographical orientation of the protein predicted via TOPCONS. To identify sites that have experienced purifying selection (p-value ≤ 0.10), the Fixed

Effects Likelihood (FEL) method present in the HyPhy package was used (Kosakovsky Pond and Frost, 2005).

3.2.7 Immunofluorescence staining of larvae

Transgenic swimming larvae (22-24hpf) generated by electroporation were fixed with 4% paraformaldehyde (PFA) in 1× phosphate buffered saline (PBS) for 30 min, washed three times with PBS, then gradually dehydrated to 100% methanol and stored at −20 °C. After stepwise rehydration to 100% PBS, samples were permeabilized with 0.2% (w/v) Triton X-100 in PBS (PBS-T). Non-specific antibody binding was blocked with 20% (w/v) heat treated sheep serum (HTSS) in PBS-T for >1 h at room temperature and samples were then incubated with primary antibody in HTSS-PBS-T. Primary antibodies were: rabbit-anti-Ci-ßγ-Crystallin (Shimeld et al., 2005), chicken-anti-GFP polyclonal (Abcam, ab13970), mouse-anti-mCherry monoclonal (Antibodies.com, A104343) all diluted 1:500 at 4 °C overnight. After several washing-steps with PBS-T at room temperature, they were incubated in secondary antibody, either goat anti-rabbit Alexafluor 488 (Molecular Probes, A11008) or goat anti-chicken Alexafluor 488 (Abcam, Ab150169) or goat anti-rabbit HRP conjugate or goat anti-mouse HRP conjugate diluted 1:500 in BSA-PBS-T. After 5 washes in PBS-T, a tyramide signal amplification (TSA) was performed by addition of tyramide solution made of TSA buffer (2M NaCl, 100mM Borate buffer pH 8.5), 0.5% H_2O_2, 0.1% 4-Iodophenylboronic acid (4IBPA) and 0.2 % 5-TAMRA fluorescent dye (Abcam, ab145438) for 30 min. Embryos were washed 5 times in PBS-T for 30 min and once overnight. Nuclear DNA was stained with DAPI (1:1000 dilution). DAPI, Alexa 488 and 5-TAMRA fluorescence were excited at 405, 488 and 559 nm, respectively, and recorded with an Olympus Fluoview FV1000 confocal microscope equipped with 60X, 100X oil-immersion objective. Samples were washed with PBS several times, mounted

with 50% (w/v) Glycerol in PBS on a coverglass-bottom petri-dish. Stacks of optical sections from 0.3 μm to 0.8 μm were acquired sequentially (one dye at a time) and z-projected. Images were analysed with Fiji (Version 2.0.0).

3.3 Results

3.3.1 Reporter assay identify a 190 bp regulatory region sufficient to drive *CiGnRH1* expression in the palps, brain vesicle and motor ganglion

Previous study done by Kusakabe et al. (2012) characterised a 4687 base pairs (bp) regulatory region upstream of *CiGnRH1* (KH.S1051.1), which is one of the two *GnRH* genes of *Ciona intestinalis*. The identified cis-regulatory element region was fused to enhanced green fluorescent reporter (eGFP) in frame with the predicted *CiGnRH1* start codon. This construct was able to drive eGFP expression in the same territory as observed by whole-mount *in situ* hybridisation, i.e., the palps, the brain vesicle and the motor ganglion cells of *Ciona intestinalis* swimming larvae (Kusakabe et al., 2012). More precisely, the original eGFP construct includes 4687 bp of *Ciona* sequence 5' to the point of fusion with eGFP. Of this, 4603 bp lies at 5' to the Transcription Start Site (TSS) inferred from the 5'extreme of the *CiGnRH1* cDNA sequence found in the ANISEED database. Hereafter, I refer to the TSS as 0 bp, with sequence 5' to this denoted with a minus sign (Fig. 3.3C). To define the precise regulatory sequence within this 4603 bp region, I fragmented successive sections of the original construct and tested these via electroporation into *Ciona* zygotes (Fig. 3.3C). The dissection of the *CiGnRH1* regulatory element aimed to reveal important part of this sequence to infer putative transcription factor binding sites and make hypotheses of possible regulatory mechanisms conserved with hypothalamic GnRH1 cells of vertebrate (Fig 3.3A-D).

First, I confirmed that the 4687 bp regulatory region (called GnRH1-A here) could operate independently as a classical enhancer by cloning it upstream of a basal promoter from a different gene, FoxA and driving β-galactosidase, an enzyme of the gene lacZ, in

the vector pCES-lacZ (Harafuji et al., 2002). The reporter construct was electroporated into *Ciona* zygotes. At the larval stage, the resulting transgenic embryos confirmed that lacZ reporter expression faithfully reproduced endogenous protein localisation and original transgene expression as expression was observed in the brain vesicle (55%, 50 larvae out of 91), motor ganglion (43%, 39/91) and palps (22%, 20/91) (Fig. 3.3A). From here onwards, I refer the larval stage as being 22 hours post fertilisation (hpf), which is the time in development the larva was fixed for subsequent analysis. As cis-regulatory elements are often evolutionary conserved, my initial point of investigation was to look for some conserved regions with another *Ciona* species, *Ciona savignyi* (Small et al., 2007). The genome of this species is sufficiently distant from *Ciona intestinalis* to allow constrained non-coding sequences to be identified by sequence comparison. The *Ciona* cross-species comparison of 5kb upstream their respective *GnRH1* genes pointed to a sequence of 146 bp (-328 to –473 bp) with 66.4% of interspecific conservation. This level of conservation made it likely to be a regulatory element therefore I proceeded to clone this conserved region that I named GnRH1-B directly in front of the plasmid reporter PCES-lacZ. However, this short conserved non-coding region did not show any reporter activity and probably constitute another type of cis-regulatory element, maybe a silencer or an enhancer driving expression at a different life cycle stage, although this was not investigated further. I also verified that the GnRH1-B conserved region was not a repeat element. There are 4 different repeat elements according to ANISEED in the full 4687 bp regulatory region. Indeed, there are three miniature inverted-repeat transposable elements (MITEs), known as *Cimi-1*, which have typical tandem inverted repeat on each side. In addition to the MITEs, there is one long interspersed retrotransposable element (LINEs) (Simmen and Bird, 2000). Due to their widespread distribution in the genome of *Ciona*, these transposable elements are unlikely to act in the regulation of *CiGnRH1*,

although transposable elements are sometimes pervasively co-opted for the regulation of host genes (Chuong et al., 2017). Conservation with distantly related ascidians of the genera *Phallusia* and *Halocynthia* was also looked at but did not show any conserved sequence.

As conservation in sequence does not seem the best strategy to identify the important part of the regulatory region in this case, I shifted strategy and tested successive fragment of the full 4687 bp regulatory region. I first split it roughly into two halves, GnRH1-C (2329 bp long, -2274 to +83) and GnRH1-D (2357 bp long, -4603 to -2275). I tested these two elements activity by reporter gene assay following lacZ expression at the larval stage of *Ciona*. Interestingly, the construct with GnRH1-C, which contain the TSS did not show any reporter expression. In contrast, the other half, GnRH1-D, showed expression in the same territory as the full 4687 bp sequence with transgenic larva displaying expression in brain vesicle (57%, 49/85), motor ganglion (45%, 38/85) and palps (25%, 21/85). These results hint that GnRH1-D contains all the transcription factor binding sites necessary to activate *CiGnRH1* expression. To further analyse the minimal functional element driving reporter expression, I further divided the GnRH1-D element in four distinct fragments: GnRH1-E (495 bp, -2769 to -2275), GnRH1-F (695 bp, -3440 to -2746), GnRH1-G (605 bp, -4024 to -3418), and GnRH1-H (583 bp, -4603 to -4021). I tested all of them by reporter gene assay. GnRH1-F reported some expression in the brain vesicle exclusively, but the percentage of positive larva was low (8%, 5/70). Only GnRH1-E, was able to drive LacZ expression in the palps (19%, 14/76), brain vesicle (47%, 36/76) and motor ganglion (37%, 28/76) suggesting this region contains all the functional cis-regulatory elements to activate transcription on its own in a reporter assay**.** The element GnRH1-E was then trimmed 150 bp from the 3' end and this new fragment

named GnRH1-I (345 bp, -2769 to -2425). The subsequent transgenic larvae showed again expression in the palps (15%, 10/72), the brain vesicle (50%, 36/72) and the motor ganglion (42%, 30/72). Then, GnRH1-I was also trimmed 155 bp from the 3' end to leave a sequence of 190 bp named GnRH1-J which also displayed expression in the palps (11%, 7/64), the brain vesicle (52%, 33/64) and the motor ganglion (39%, 25/64) same territories as the initial 4687 bp regulatory region, GNRH1-A (Fig 3.3D).

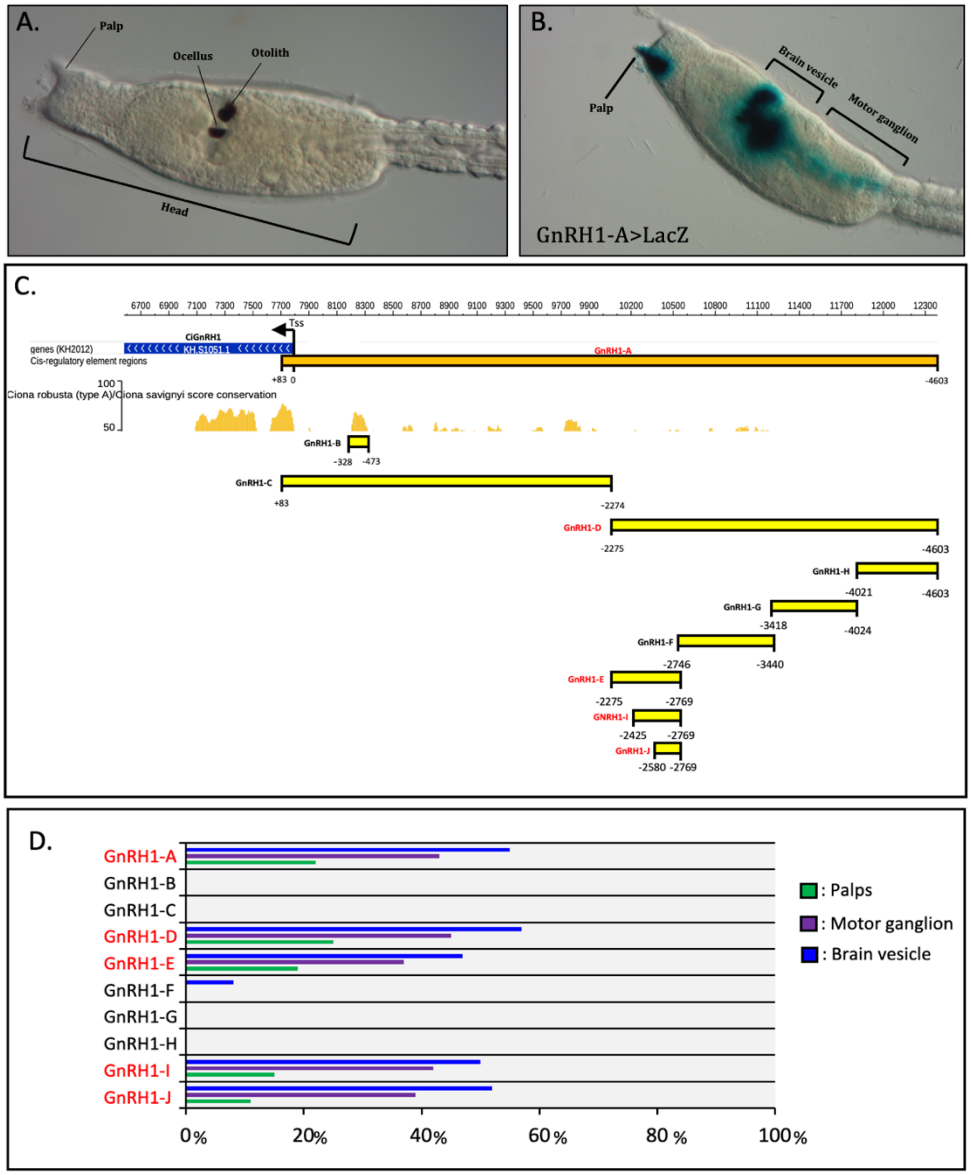


Figure 3.3. The *CiGnRH1* cis-regulatory region. (A) Ciona larval head 22 hpf (B) Side view of larva embryo electroporated with GnRH1-A > LacZ construct, showing expression in the brain vesicle, motor ganglion and a single palp only, a common occurrence reflecting transgenesis mosaicism. (C) Schematic representation of *CiGNRH1* (KH.S1051.1) (blue rectangle) and upstream regulatory region (GnRH1-A) (orange rectangle). The conservation score of between *Ciona robus*ta and *Ciona savignyi GnRH1 loci* employing WASHU browser is represented underneath, as well as the different fragments tested by gene reporter assay (yellow rectangles). The TSS indicates the transcription start site and numbering start from this point. (D) Percentage of larvae expressing LacZ reporter in palps, brain vesicle and motor ganglion; each bar represents the combined number of larvae counted during at least three trials; n > 60 embryos scored for transgene expression. The fragments driving expression in brain vesicle, motor ganglion and palps are highlighted in red.

3.3.2 Prediction of transcription factors binding sites and hypothesis for a conserved mechanism for GnRH1 regulation in Olfactores

The new core 190 bp regulatory region identified here has a reasonable size to make transcription factor binding site (TFBS) prediction as it limits greatly the number of potential candidate regulators in comparison to the initial ~4.7 kb. The regulators that are binding those TFBS could be either activators or repressors. They should be co-expressed with *CiGnRH1* if activators although not necessarily if repressors. To identify the best candidate regulators, I looked for transcription factor genes (TFs) in recently published Single-cell RNA-seq data of the palp axial columnar cells (ACCs) in *Ciona* (Sharma et al., 2019; Johnson et al., 2020). The palp ACCs have been shown in this study to express the CiGnRH1 neurohormone (see section 3.3.5). Whilst *GnRH1* regulation has been studied in a range of vertebrate models, the immortalised GnRH1 mouse cell lines (GN11, GT1-7, NLT) has enabled a detailed characterisation of the transcription factors regulating the 5 kb upstream region of the mammalian *GnRH1* gene (Hoffman and Mellon, 2018). Hence, to guide interpretation of potential homologies for the regulation of GnRH genes in the *Ciona* ACCs and in the olfactory-derived GnRH1 neurons of mammals, I made a recapitulative table of the TFs known to interact with the mammalian GnRH1 regulatory region from the literature (Table 3.2). In total, 18 transcription factors are known to physically interact with the GnRH1 regulatory region in mammals. Out of these a staggering 13 TFs have homologous genes that have enriched transcripts in the ACCs according to Single-cell RNA-seq (Johnson et al., 2020), which support the possibility of a conserved regulatory mechanism (Table 3.3).

Table 3.2. TF genes known to physically interact with mammalian GnRH1 regulatory region

TF families	Gene Name	Regulator type	Mouse GnRH Cell-line	Publication
HOMEOBOX	**POU2F1 / OCT1**	Activator	GT1-7	Clark and Mellon, 1995
	POU3F1 / OCT6	Repressor	GT1-7	Wierman *et al.*, 1997
	POU3F2 / BRN2	Activator	GN11, NLT	Wolfe et al., 2002
	MSX1/2	Repressor	GT1-7, GN11, NLT	Givens *et al.*, 2005
	OTX2	Activator	GT1-7	Kelley *et al.*, 2000
	DLX2/5	Activator	GT1-7, GN11, NLT	Givens *et al.*, 2005
	VAX1	Activator	GT1-7, GN11	Hoffman *et al.*, 2016
	SIX3/6	Activator	GT1-7, GN11	Larder et al., 2011
	PKNOX1	Activator	GT1-7	Rave-Harel *et al.*, 2004
	NKX2.1	Repressor	GT1-7	Lee et al., 2001
	ZEB1	Repressor	GN11	Messina *et al.*, 2016
	PBX	Activator	GT1-7	Rave-Harel *et al.*, 2004
GATA Zinc finger	GATA4	Activator	GT1-7	Lawson *et al.*, 1996
SRY-BOX	**SOXC**	Activator	GT1-7	Kim *et al.*, 2011
NHR	**RXR**	Activator	GT1-1	Cho *et al.*, 2001
	COUP-TFI / NR2F1	Repressor	GT1-7	Gillespie et al., 2004
Zinc fingers C2H2-type	**EGR1**	Activator	GN11	DiVall et al., 2007
Basic leucine zipper	**CEBPB**	Repressor	GT1-7	Belsham and Mellon, 2000

List of TF genes regulating the 5 kb upstream of the mammalian GnRH1 gene. The gene name, type of regulator, type of GnRH mouse cell line used, and the publication sources are detailed. The genes with homologues in the GnRH1^{+} palp ACCs of Ciona are in bold (see table 3.3).

Table 3.3. Selected TF genes with transcripts enriched in the axial columnar cells compared to neural cells

TF families	NAME	ANISEED ID	KH ID	Top BLASTP hit in human
HOMEOBOX	Pou-like	Cirobu.g00008662	KH.C7.782	POU2F3; POU4F1; POU6F1
	PouIV	Cirobu.g00004616	KH.C2.42	POU4F1; POU4F2; POU4F3
	Msxb	**Cirobu.g00005203**	**KH.C2.957**	**DLX1; MSX1; MSX2**
	Otx2	**Cirobu.g00006940**	**KH.C4.84**	**CRX; OTX1; OTX2**
	Nk4	**Cirobu.g00009121**	**KH.C8.482**	**NKX2-3; NKX2-5; NKX2-6**
	Zeb1/2	Cirobu.g00001049	KH.C1.777	ZEB1; ZEB2; ZNF22
	Pbx	**Cirobu.g00014220**	**KH.S215.9**	**PBX1; PBX2**
	Six1/2/6	**Cirobu.g00005742**	**KH.C3.553**	**SIX1; SIX2; SIX6**
NHR	**Coup**	**Cirobu.g00011734**	**KH.L17.15**	**NR2F1; NR2F2; NR2F6**
	Rar	**Cirobu.g00010140**	**KH.C9.580**	**RARA; RARB; RARG**
SRY-BOX	**SoxC**	**Cirobu.g00008378**	**KH.C7.523**	**SOX11; SOX12; SOX4**
Zinc fingers C2H2-type	Egr1/2/3	Cirobu.g00011866	KH.L172.16	EGR1; EGR2; EGR3
Basic leucine zipper	CEBPB/E/G	Cirobu.g00005334	KH.C3.176	CEBPB; CEBPE; CEBPG

Selected list of homologous TF genes identified in the ACCs from the Sc-RNA-seq dataset of Sharma *et al.*, 2019, according to the reanalysis of Johnson *et al.*, 2020. The unique gene ID (Cirobu.gxxxxxxxx), KyotoHoya (KH) gene model ID, name given according to CIS-BP or Aniseed databases, and Top BLASTP (after translation) hit in human are detailed. The genes with whose TFBS were detected in GNRH1-J are in bold.

I therefore searched in the 190 bp regulatory region "GnRH1-J" for TFBS of those 13 regulators using the CIS-BP database with *C. intestinalis* motifs (Weirauch et al., 2014) and cross-confirmed the results with the vertebrate JASPAR database (Khan et al., 2018) and SELEX-seq data on Aniseed which describes the DNA-binding specificity and the relatedness of DNA-binding specificity to mammalian orthologs. The search detected TFBS for 8 out of the 13 TFs selected in *Ciona* (Table 3.3, Table 3.S1 in Appendix). Especially, I highlight the presence of four core ATTA sequences which characterise TFBS for homeodomain proteins such as Pbx, Msxb, Nk4, Otx, Six1/2/6 (Fig. 3.4). Homeodomain proteins play a key role in mammalian regulation of *GnRH1* as 12 out of 18 TFs known to interact are from that protein family. I also report the presence of a TFBS for SoxC and five TFBS for Nuclear hormone receptors such as Rar and Coup. Further experimental work is however needed to confirm which of those individual predicted TFBS are functional *in vivo*.

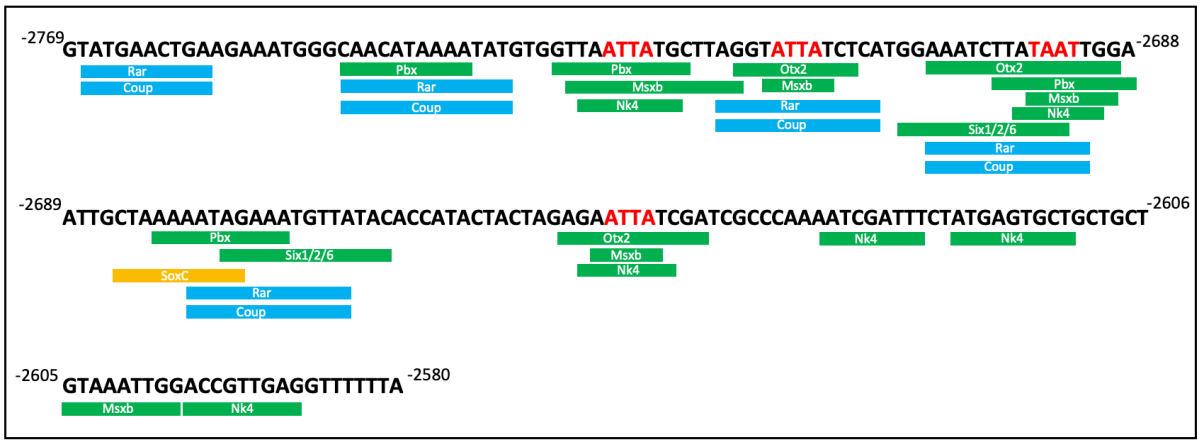


Figure 3.4. Predicted TFBS in GnRH1-J. The 190 bp GnRH1-J regulatory region from *C. intestinalis*, with putative binding motifs in color boxes, with green for homeobox TFs, blue for nuclear hormone receptors and orange SRY-Box TF. The names of these reflect the database entry (CIS-BP or JASPAR) to which they match and do not necessarily mean that the specific factor binds this site in *Ciona*. Core ATTA homeodomain binding site sequences are highlighted in red. Numbering is done according to the 4603 bp upstream of the TSS (not shown) which represent the start point (0 bp).

3.3.3 Identification and genomic organisation of *Ms4a* genes in chordates

The *Ms4a* genes represent a recently described chemoreceptor family present in immune cells but also non-immune cells like the necklace OSNs, an atypical subset of neurons present in the recess of the mammalian olfactory epithelium (Greer et al., 2016). Previous, phylogenetic analysis of the *Ms4a* gene family concluded that those genes first appeared in cartilaginous fish as the most primitive species with clear examples of *Ms4a* genes was the spiny dogfish (*Squalus acanthias)* (Zuccolo et al., 2010). These latter authors said that no evidence of *Ms4a* genes was found in jawless vertebrates or in organisms such as bacteria, protists, fungi, or plants. Therefore, it was concluded that *Ms4a* genes first appeared in Chondrichthyes with expression outside of the immune system and have since diversified in many species into their modern forms with expression and function in both immune and nonimmune cells (Zuccolo et al., 2010). However, since member of *Ms4a* genes are chemoreceptors, they are quite divergent among evolutionary distant organisms as their amino acid composition changes rapidly (Ishibashi et al., 2001). Therefore, I investigated whether *Ms4a* chemoreceptor genes could have a more ancestral origin or if they are just gnathostome-specific as previously stated. Contrary to which was previously stipulated my search reports the existence of a *Ms4a* gene in the lamprey genome of *Petromyzon marinus,* identified as an uncharacterized gene locus by using a Ms4A protein sequence from the whale shark (*Rhincodon typus)* as a search query. An orthologue to the lamprey *Ms4a* gene was identified in the hagfish genome of *Eptatretus burgeri* also as an uncharacterized gene locus. Furthermore, I report the presence of great diversity of *Ms4a* genes in the genomes of protochordates such as in ascidians and amphioxi species. Interestingly, no *Ms4a* homologues or pseudogenes could be identified in the genomes and available

transcriptomes of larvaceans and thaliaceans, even when using more sensitive search methodology using hidden Markov models (see methods).

In mammal genomes, the *Ms4a* genes organise together in a cluster on a single locus in all species investigated. Here, I use the definition of Yi et al., 2007 for what makes a gene cluster: "a cluster of genes is a group of two or more genes found within an organism's genome that encode similar functional proteins and are usually located within a few thousand base pairs of each other". A cluster of genes is a form of organisation typical of chemoreceptor families (Niimura, 2009). In fact, the *Ms4a* clusters are indicative of tandem duplication events which contributed to facilitate the diversification of paralogues by neo-functionalisation (Nei et al., 2008). In the mouse it has been highlighted that there is a cluster of conventional seven transmembrane G-protein-coupled olfactory receptors (7TM GPC ORs) genes immediately telomeric to the *Ms4a* gene cluster (Greer et al., 2016). I examined if the spatial proximity (or linkage) seen in mouse between *Ms4a* and *OR* genes clusters can be found in other chordate genomes. My investigation pointed to a preserved chromosomal linkage between those different chemoreceptor families in all sarcopterygian's genomes scrutinised while in actinopterygians a MS4A-OR linkage was only seen for sterlet but not in spotted gar (*Lepisosteus oculatus*) or teleosts. In Chondrichthyes, a preserved Ms4a-OR linkage was observed in the skate despite the paucity of *OR* genes present in that class (*Amblyraja radiata*). In lamprey, the *Ms4a* gene is close from a cluster of three *OR* genes on the small chromosome 59. In protochordates like amphioxus, the *Ms4a* genes are all linked to orthologous genes of vertebrates *ORs* except for one *Ms4a* cluster on an unplaced scaffold. In *Ciona intestinalis*, despite the highly fragmented genome of ascidian, some *Ms4a* genes are found near a cluster of orphan 7 TM receptors. While not being genuine orthologues,

these orphan 7 TM receptors have recently been suggested to be ORs expressed in the palp axial columnar cells (Johnson et al., 2020). It should be noted that in chordates there can be more than one locus of *Ms4a* genes on different chromosomes, and all are not linked to *ORs* but there is always one *Ms4a* gene linked to *OR*, except in neopterygians (teleosts + gars + bowfin). Also, *Ms4a* genes do not always organise as tight clusters and can be more scattered on a same chromosome. To further examine their genomic relationships, I looked for additional evidence of ancestral synteny surrounding *Ms4a* genes linked to *OR* genes in human, sterlet, skate, lamprey, Ciona and amphioxus. Additional conserved genes provided further evidence that I was comparing homologous loci in different chordate genomes (Fig. 3.5). Among those are genes used in the vertebrate olfactory system, such as *Mucin5ac,* a gel-forming protein produced in high quantity by the olfactory glands and vomeronasal receptors, *V2R*, another type of chemoreceptor. There is also conserved synteny with developmental genes like *Notch1* and the oocyte zinc finger *XiCOF6*, in some species. The identified *Ms4a* genes represented on Figure 3.5 are in Table 3.S2 of the Appendix.

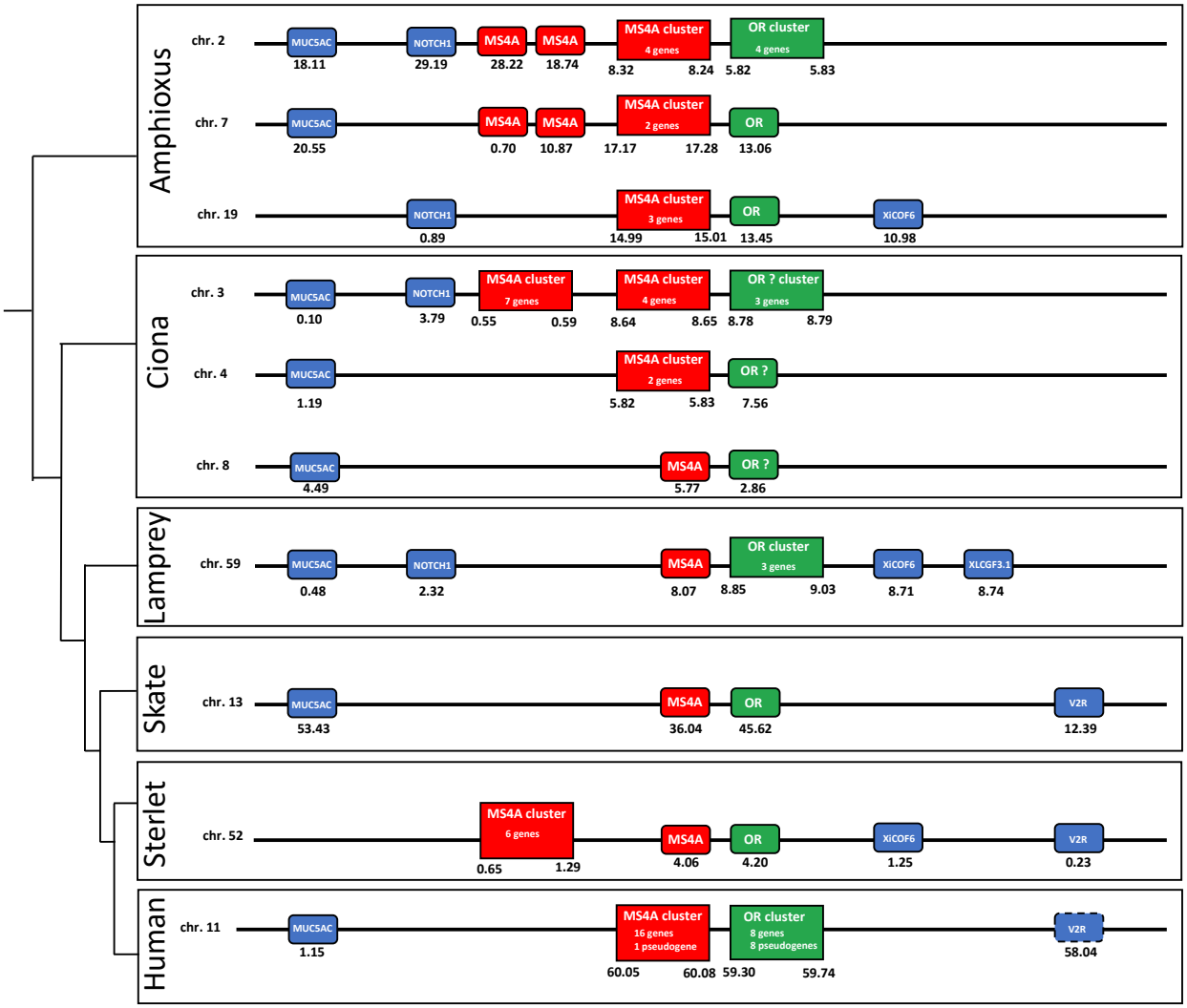


Figure 3.5. Synteny of *Ms4a* genes linked to *OR* genes in amphioxus, Ciona, lamprey, skate, sterlet and human. The neighbouring genes are aligned for ease of comparison, but the distance (shown below in Mb) and order may vary. *Ms4a* genes are in red, colour coding of other genes is as follows: Green, *OR* genes. Blue colour, genes linked to MS4A-OR genes on the same chromosome and their orthologues and/or paralogues in other species. Dotted outlines represent pseudogene.

The presence of *Ms4a-like* genes was briefly explored outside chordates, and I report their presence in evolutionary distant organisms. *Ms4a* genes are present in hemichordates and echinoderms and some organised in clusters are linked to rhodopsin-like G protein-coupled receptors, some of which also organise as clusters and could well be olfactory receptors although this was not investigated further. Furthermore, the antiquity of chordate *Ms4a* genes receptors are revealed by their existence in the sea anemone *Nematostella vectensis.* This latter organism is also known to possess chordate *OR-like* receptors (Churcher and Taylor, 2011). I therefore examined if they also could be found in proximity of those genes, but I could not find linkage with *Ms4a* genes. Although, those cnidarian *Ms4a* genes organised in clusters are close to other 7TM GPCR genes some of which could well be ORs but have not been labelled as such by the NCBI annotation. In conclusion, the evolution of the *Ms4a* genes is complex but interesting as they display various numbers highlighting that these loci are extremely dynamic, with expansion or shrinkage which reflect the necessity for each lineage to maintain, extend or pseudogenise those chemoreceptors according to the need to detect specific chemicals in their environment. Clusters of *Ms4a* chemoreceptor genes are present in all chordate genomes investigated (with the apparent exception of thaliacean and larvacean) which emphasises the important role of these genes in olfactory system evolution. The conserved linkage of MS4A and OR clusters across chordate genomes might help to infer the chemoreceptor function of uncharacterised MS4A and 7TM genes, such as those in *Ciona*.

3.3.4 Evidence of positive Darwinian selection in recently diverged ascidian *Ms4a* paralogues support a functional role as chemoreceptor

While numerous *Ms4a* genes were identified in ascidian's genomes, there are no data to support that they may have a functional role as chemoreceptors like in mammals. Metazoan chemoreceptor genes such as the vertebrate's *OR*, *TARR, VR* and *Ms4a* genes are known to be shaped by positive selection in the motifs responsible for ligand-binding (Chen et al., 2010; Hussain et al., 2009; Greer et al., 2016; Yang et al., 2019). The variability in sequence in ligand-binding domains is what enables different related chemoreceptors to detect distinct chemical cues. Multiple sequence alignments revealed substantial intra-species (Fig. 3.6A) and inter-species (Fig. 3.6B) diversity among ascidian Ms4as, particularly within the two extracellular domains of the proteins, whose lengths are variable and are alleged to be responsible for ligand-binding.

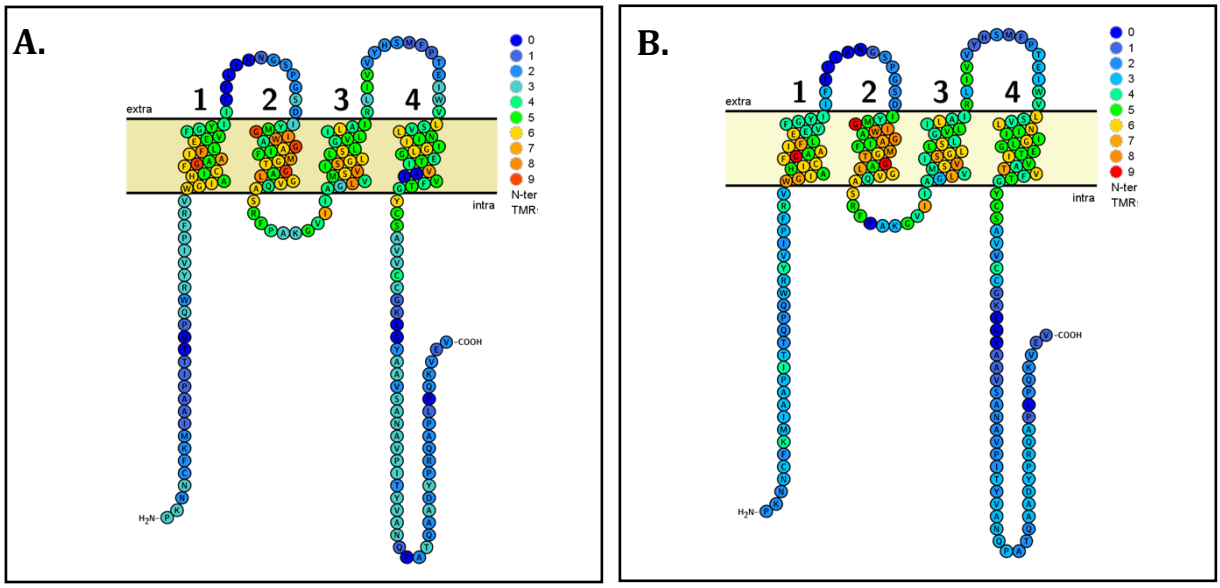


Figure 3.6. Topographical representations of *Ciona robusta* MS4A (KH.C1.484) primary sequence with amino acid conservation heat mapped. In (A) the schematic represents amino acid conservation between all Ms4a paralogues in *Ciona robusta* while in (B), it is a representation of amino acid conservation between all MS4A homologs in *Ciona robusta* and *Ciona savignyi*. Residues that are more conserved are shown in warmer colours, whereas residues that are less conserved are depicted in colder colours (conservation scores were determined using PRALINE). The extracellular domains reveal the greatest sequence diversity with additional diversity in the intracellular N-terminus and C-terminus.

These findings raised the possibility that each Ms4a within a given ascidian species has diversified through positive Darwinian selection to interact with distinct chemical cues. Hence, I explored the role of adaptive evolution in ascidian *Ms4a* genes diversification using a method that allows me to discuss functional aspects of such diversification. To do so, I applied a codon substitution model in a phylogenetic context to the find distribution of selective constraints across different domains in the ascidian Ms4a proteins and evaluated the impact of positive selection in their evolution. A useful statistic to measure the strength and mode of natural selection acting on protein-coding genes is the ratio (ω) of non-synonymous (dN) to synonymous substitutions (dS), ω= dN/dS. Non-synonymous substitutions are nucleotide changes that alter the protein sequence, while synonymous substitutions do not. Typically, when this ratio has a value of ω>1, it indicates positive (adaptive) selection, whereas ω<1 suggest negative (purifying selection) and when ω =1, it signifies neutral evolution (Nei and Gojobori, 1986). Positive selection happens when there is good incentive to change to increase the organism's fitness, while negative selection occurs generally in conserved regions of protein-coding genes when there is no incentive to change as it can be deleterious. The methodological approach used herein to detect sign of positive selection relies on sequence variation and its accuracy is reduced if it is applied to very divergent homologs (Almeida et al., 2015). The fact is that ascidian genomes evolve quickly, for instance the evolutionary distance between the genomes of congeneric species like *Ciona intestinalis* and *Ciona savignyi* is considered greater than the distance between human and chick, providing a very low background of unconstrained conservation (Johnson et al., 2004). In consequence, the power of such diversifying selection analysis is greatly reduced when considering gene sequences of such fast evolvers. To circumvent this problem, I identified phylogenetic clades of recently diverging Ms4a paralogues in ascidians (Fig. 3.7A). The colonial ascidian

Botryllus schlosseri turned out to be the most interesting for such analysis as molecular phylogeny revealed the presence of three clades characterising recent paralogous expansions. Statistical analyses were thus performed within six clades of paralogous genes in *B. schlosseri* (Fig. 3.7B). To maximise the output of the likelihood analysis, the Ms4a proteins Boshl_g00014523 and Boshl_g00060878 were excluded in clades 4 and 5 because they did not align robustly. Also, the N- and C-terminal regions were removed in all clades because of uncertainty over the alignment in those regions. The present methodology kept genetic divergence minor and preserved the robustness of the alignments while enabling the examination of a satisfactory number of sequences.

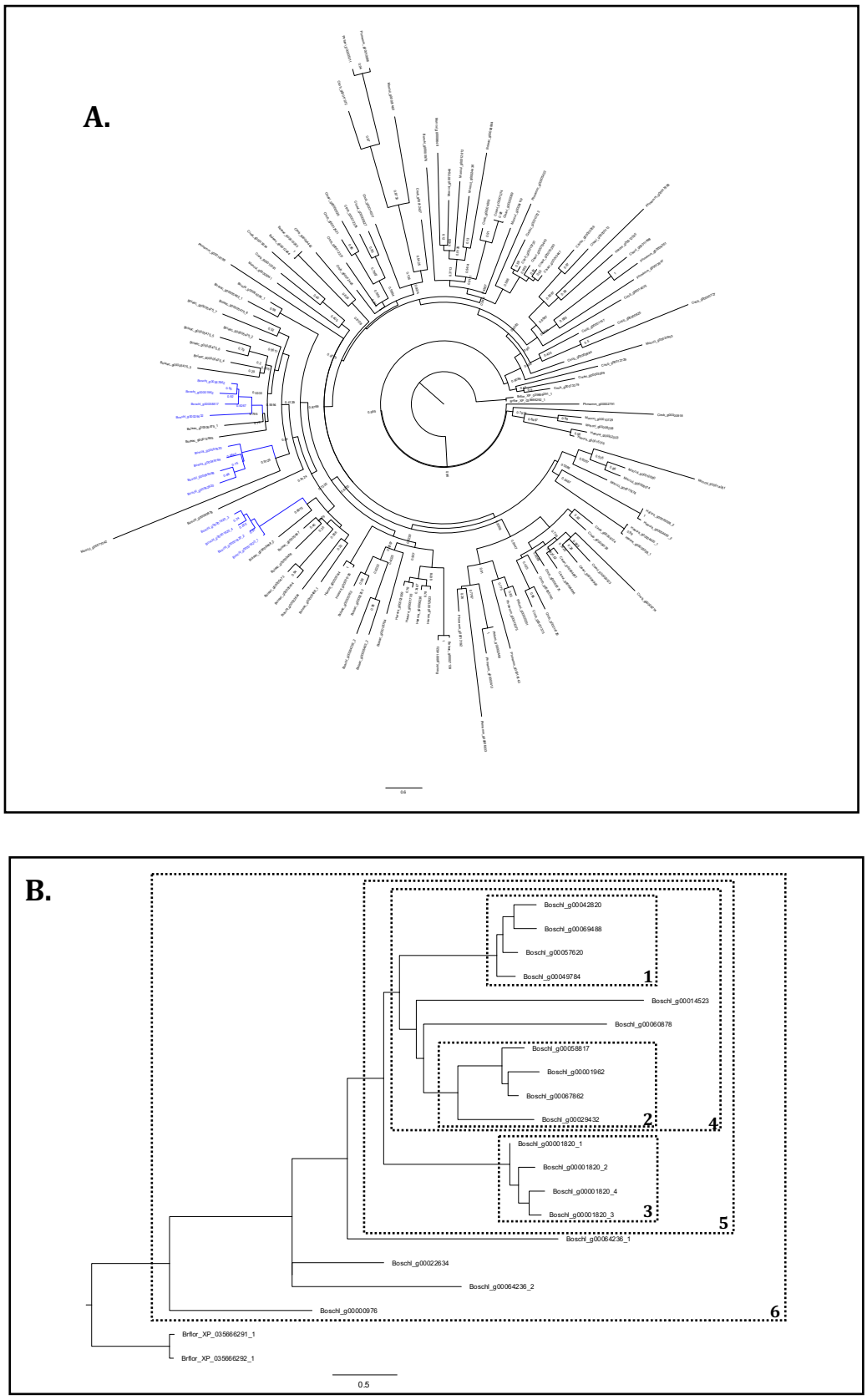


Figure 3.7. (A) Circular maximum-likelihood phylogenetic tree of 128 ascidian MS4A proteins. Specific MS4a gene expansions of *B. schlosseri* are in blue. the rest of ascidian gene names and branches are in black. (B) Clades of *B. schlosseri* paralogous genes analysed in EASYCODEML for dN/dS ratio. Each clade is boxed by a dashed line, and the numbering correspond to the result summarized in Table 3.4. For both trees a bootstrap analysis was performed 100 times. All trees are rooted with amphioxus (*Br. floridae*) Ms4a proteins as an outgroup. The scale bars represent the number of substitutions per sites.

The number of genes per clade varied from 4 to 18 and the total tree length in number of substitutions per codon varied from 0.62 to 9.38 (Table 3.4). The average ω-values across alignments (model M0) never reached 1, being the biggest for Clade 1 (ω= 0.8141) (Table 3.4). The lowest ω-value (ω=0.3149) was observed in Clade 6, which was the largest clade with the longest branch lengths separating its members (Fig. 3.7B). These relatively high ω-values observed in *B. schlosseri* MS4a paralogues agree with a relaxation of purifying selection expected for duplicated genes and detected in different chemosensory genes families of other species (Qiu et al., 2019; Almeida et al., 2015; Chen et al., 2010). These average ω-values below one indicates that most codons in these genes are evolving under purifying selection although some codons might undergo positive selection when selective constraints are relaxed.

Table 3.4. Characteristics of the codon sequence alignments per clade.

CLADES	N	C	T	ω
Clade 1	4	372	0.7	0.8141
Clade 2	4	396	1.3	0.3166
Clade 3	4	372	0.62	0.4874
Clade 4	8	378	3.11	0.3948
Clade 5	12	363	4.96	0.3812
Clade 6	18	291	9.38	0.3149

N, number of sequences; C, number of codons; T, total tree length (in number of substitutions per codon); ω, average dN/dS over all alignment positions analysed (model M0).

Multiple sequence alignments for each clade were fed to the codon substitution models M1a, M2a, M7 and M8 (see method for details). The presence of positively selected sites in the analyzed paralogs was supported by the highly significant likelihood ratio tests (LRT) in all clades when comparing M1a versus M2a or M7 versus M8 (Table 3.5). The bigger the clade was the fewer positive selected sites were found due to the increasing level of divergence detected between paralogues. The strongest signal of positive

selection was in the most recently expanded clade of *B. schlosseri* Ms4a (Clade 3, LRT P-values in Table 3.5). The proportion of sites estimated to have an ω > 1 ranged from p1 = 5.22–15.2 % (according to model M8), which indicates certain levels of positive selection acting during the diversification of these receptors. Given the significant LRT attested the presence of positively selected sites, the Bayes empirical Bayes (BEB) method was then used to predict the location of such sites under the model M2a and M8. The BEB analysis identified the specific codons with a strong posterior probability (PP) to belong to the ω > 1 site class. Table 3.5 present all the positive sites with PP >0.5. Amino acid sites are generally accepted to be under positive selection if they have at least a PP > 0.95 (Scheffler and Seoighe, 2005). Here, the number of sites with PP > 0.95 was low in all clades. For instance, Clade 3 had the highest number of sites with PP > 0.95 with 5 out of 13 sites with probability to be under positive selection and with 3 sites with PP > 0.99.

Table 3.5. Results of likelihood ratio tests and parameters estimates under the best-fitting model for each clade

CLADES	M1a vs M2a LRT (P-value)[a]	M7 vs M8 LRT (P-value)[b]	Parameters estimates[c]	Positive sites[d]
Clade 1	0.000034549	0.000033956	p0= 0.94775 p1= 0.05225 ω= 27.36303	**28 L 0.788**,29 N 0.517,**30 P 0.977*,31 S 0.994****,33 A 0.609,**35 P 0.685**,**36 P 0.955***,37 I 0.507,39 A 0.605,**98 L 0.945**,**101 V 0.983***,**102 A 0.766**,103 V 0.574
Clade 2	0.001722930	0.000021385	p0= 0.84271 p1= 0.15729 ω= 4.17122	**24 S 0.987***,**27 T 0.847**,**28 K 0.814**,**29 R 0.957***,**31 E 0.622**,**33 T 0.714**,**34 Y 0.592**,**35 S 0.964***,**36 Y 0.982***,**40 N 0.629**,41 V 0.533,**100 G 0.905**,**101 L 0.954***,**113 I 0.757**
Clade 3	0.000011343	0.000005698	p0= 0.90223 p1= 0.09777 ω= 29.95506	**16 T 0.586**,**25 I 0.628**,**27 P 0.976***,**28 T 0.685**,**29 N 0.995****,**30 G 0.977***,**31 V 0.960***,**36 G 0.997****,**39 L 0.984***,59 V 0.526,**60 F 0.824**,**68 C 0.878**,**101 E 0.965***,**102 R 0.997****,**104 L 0.928**
Clade 4	0.001946302	0.000001075	p0= 0.90160 p1= 0.09840 ω= 3.16466	26 L 0.589,**27 L 0.893**,**28 N 0.975***,**30 S 0.995****,32 A 0.628,**34 P 0.762**,**35 P 0.960***,**92 T 0.876**,**97 V 0.921**
Clade 5	0.000050308	0.000000000	p0= 0.88319 p1= 0.11681 ω= 2.48143	**23 S 0.780**,25 E 0.612,**26 N 0.999****,27 P 0.823,**28 S 0.995****,29 I 0.642,**32 P 0.987***,**33 P 0.878**,**93 V 0.990****
Clade 6	1.000000000	0.000023868	p0=0.94775 p1= 0.05741 ω= 2.39634	22 Y 0.710**, 23 P 0.992****, 24 G 0.915**

[a,b] P-value obtained with the Likelihood Ratio Test (LRT) after M1a and M2a models comparison or M7and M8 models comparison

[c]Parameters estimated under the model M8: p0 = proportion of sites that follow a beta distribution with 10 classes (0< ω<1); p1= proportion of sites in the extra class with ω>1

[d]Positive site under positive selection according to M7 vs M8 models comparison with a PP>0.50. The sites with PP > 0.95 have a single asterisk while sites with PP > 0.99 have double asterisk. The residues in bold are the one also identified with the M1a vs M2a comparison which is more stringent. The numbers and letters correspond to the positions and relevant amino acids of the sequences that were present on top of the multiple codon alignments.

To examine how selective pressure is spread across functional elements, I mapped the location of all amino acid sites under positive and negative selection in the predicted Ms4a functional domains of Clade1-3. The structure of the Ms4a proteins is characterized by an N-terminal domain (N-ter), two extracellular loops (ECLs) variable in size, one intracellular loop, four transmembrane domains (TM1–TM4) and a C-terminal domain (C-ter) (Ishibashi et al, 2001). The ECLs supposedly recognize specific ligands (Greer et al., 2016). All the MS4A codons predicted to be under positive selection with PP > 0.95 are in the ECLs, strongly suggesting that these structures played a central role in the functional diversification of MS4A paralogs, an adaptation likely driven by different types of chemical ligands. In addition, the vast majority (92.5%) of sites identified as being under purifying selection are in the transmembrane domains (TM1–TM4) and the intracellular loop between TM2 and TM3 hinting that sequence conservation of such domains are essential for protein stability and function.

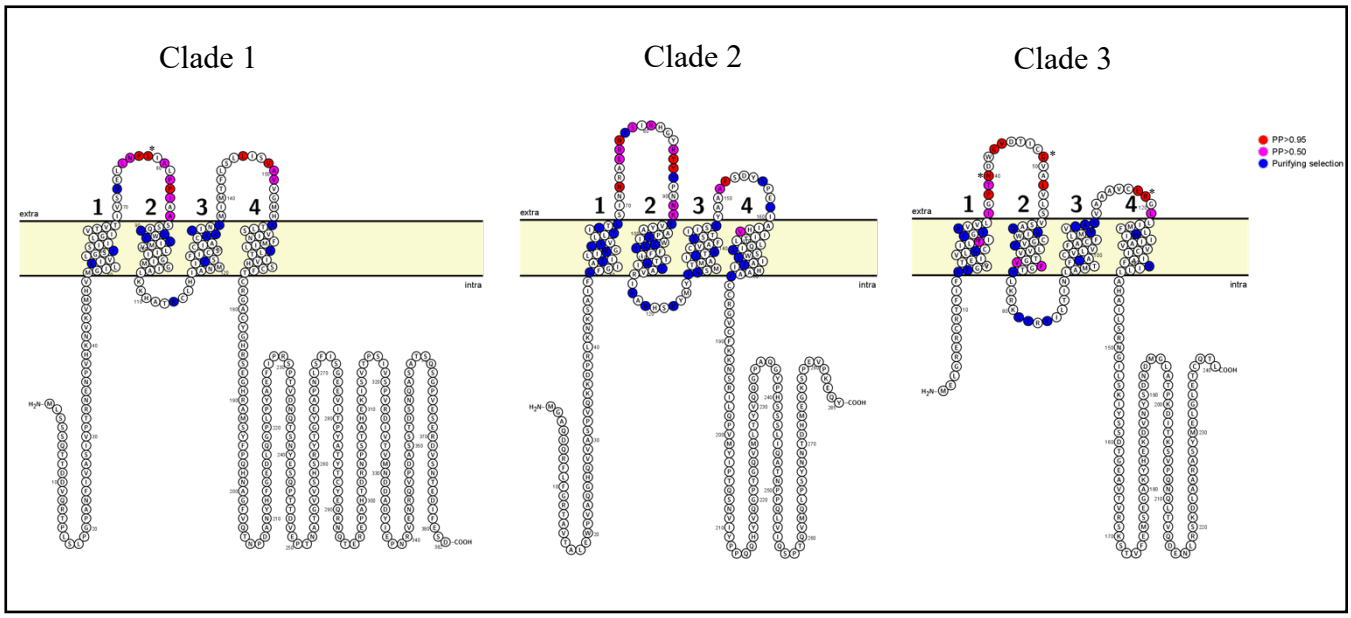


Figure 3.8. Schematic representation of the positively selected sites located on the different MS4A domains in Clade1-3. Amino acids under high probability to be under positive selection are in magenta (PP >0.5) and red (PP>0.95). The sites with a PP>0.99 have an additional asterisk. Sites under purifying selection are in blue. The numerations indicate transmembrane domains 1 to 4.

3.3.5 A GnRH neurohormone and a putative MS4A chemoreceptor are both expressed in the palp ACCs of *Ciona* larva

The ascidian palp cells constitute one of the best candidates for homologous cell type with the olfactory placodal neurons of vertebrates due to their anteriormost positions and transcriptomic profiles (cf. Chap1). In particular, the ACCs have apical finger-like protrusions extending far through the adhesive hyaline cap and the tunic which make them exposed to the outside world in a way that the PSNs and collocytes are not (Fig. 3.1). Previously, Kusakabe *et al*. (2012) suggested that *Ci-GnRH1* is expressed in the PSNs of the palps, due to the detection of axons. However, this was not investigated in depth at cellular resolution, and it remains unclear whether other palp cell types could express that neurohormone. I thus analysed by confocal imaging in which of the three-existing type of palp cells, the GnRH-1 neurohormone is expressed. To do so, I generated a construct where the newly identified 190 bp *CiGnRH1* enhancer drove eGFP and named it, GnRH1-J>eGFP. I then doubly labelled the transgenic larvae with antibody specific to ßγ-crystallin (Shimeld et al., 2005) which mark only the axial columnar cells (ACCs). The resulting transgenic larvae confirmed *CiGnRH1* expression in the PSNs as labelling was present laterally in spindled shape cells with axons typical of PSNs (Fig. 3.9B-E). In addition, co-staining clearly showed that the ACCs are simultaneously labelled by *CiGnRH1* and *ßγ-crystallin*. Indeed, the GFP positive cells and TAMRA red fluorescent dye seemed tightly intermingled in the digitiform protrusions of ACCs (Fig. 3.9; video files 3.1-3.6 in Supplementary Data). Using the same methodology, I characterised the expression of a regulatory region ~1kb upstream of the *Ciona* MS4A gene "KH.C1.484" transcription start site. I fused this newly identified MS4A regulatory region to the mCherry fluorescent reporter. The generated transgenic larvae showed simultaneous

expression of the MS4A>mCherry reporter construct and ßγ-crystallin in the exposed apical tip of the ACCs (Fig. 3.10; video files 3.7-3.21 in Supplementary Data). Some staining was also observed on some extension looking like cilia outside the ACCs, which suggest that MS4A>mCherry may also stain the ciliated PSNs, although this inference was less certain (Fig. 3.10D). Finally, a positive control was made to validate the functionality of the molecular constructs as well as the imaging methodology. I used a previously described regulatory region of the Notch ligand Delta-like (KH.L50.6) (Roure *et al*., 2014) and fused it upstream of the mCherry fluorescent reporter (the same vector as for the Ms4a regulatory region). The Delta-like>mCherry constructs confirmed the method validity as reporter gene expression was clearly seen in the lateral PSNs of *C. intestinalis* larva and distinct from the ßγ-crystallin positive ACCs. Indeed, the PSNs were clearly identified from their spindled shape, axons and cilia (Fig. 3.11; video files 3.22-3.24 in Supplementary Data).

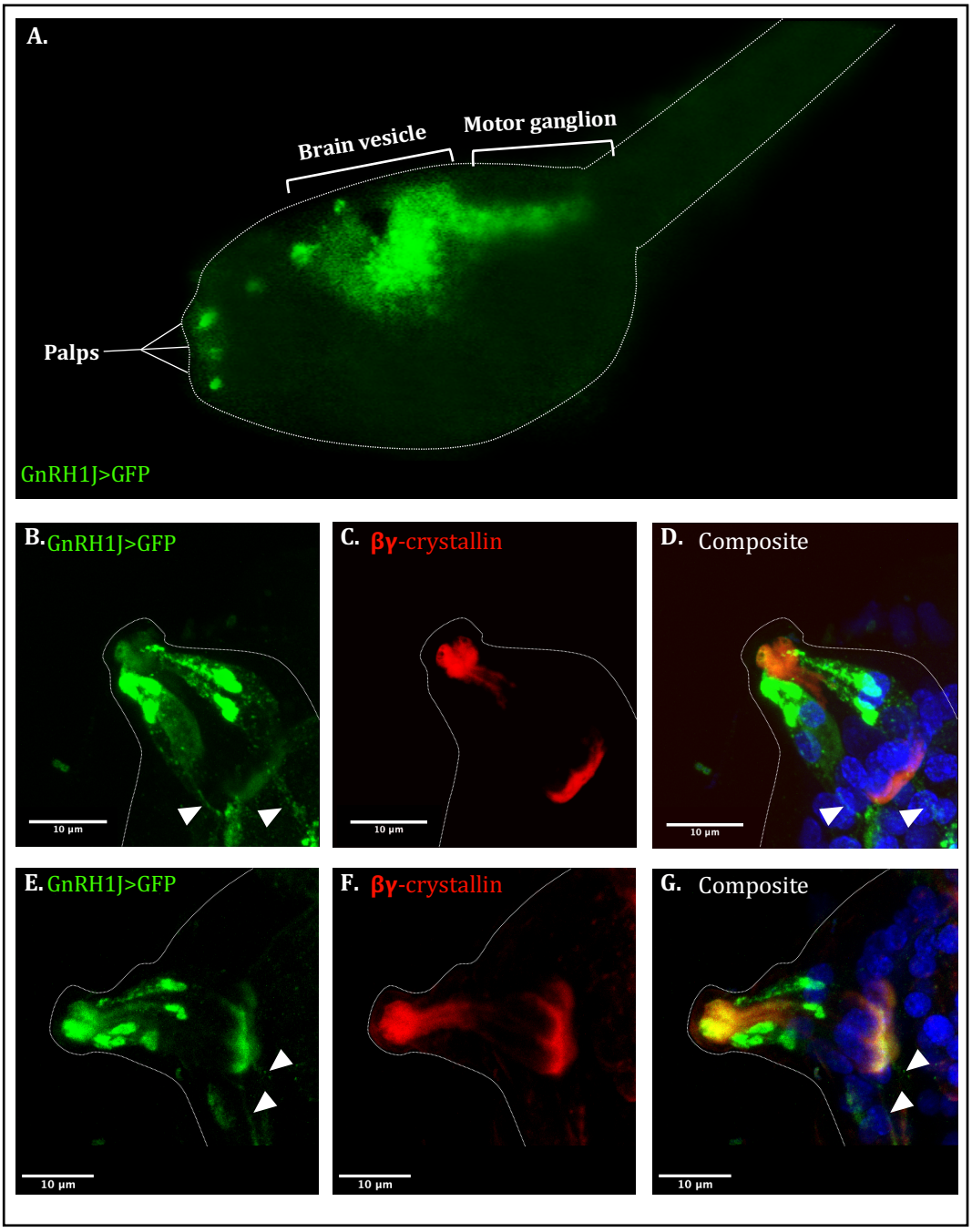


Figure 3.9. GnRH-1 neurohormone is expressed in the PSNs and ACCs. (A) Ciona larva head expressing the GFP reporter under the control of the identified 190 bp GnRH1 enhancer. The construct, GnRH1J>GFP shows reporter expression in the palps, brain vesicle and visceral ganglion. (B-G) Confocal projections of two different individual palp of transgenic larvae (24 hpf) with (B-D) first palp and (E-G) second palp. (B & E) GnRH1J>GFP signal (green) in PSNs and ACCs. White arrowheads show axons of PSNs. (C & F) ßγ-Crystallin antibody marking the ACCs (red). (D & G) Composite image with GnRH1J>GFP in triple staining with ßγ-Crystallin antibody (ACCs, red), and DAPI (nuclei, blue). Video files are in the electronic supplementary material. (B-G) Dotted outline around a single palp. Scale bar: 10 µm.

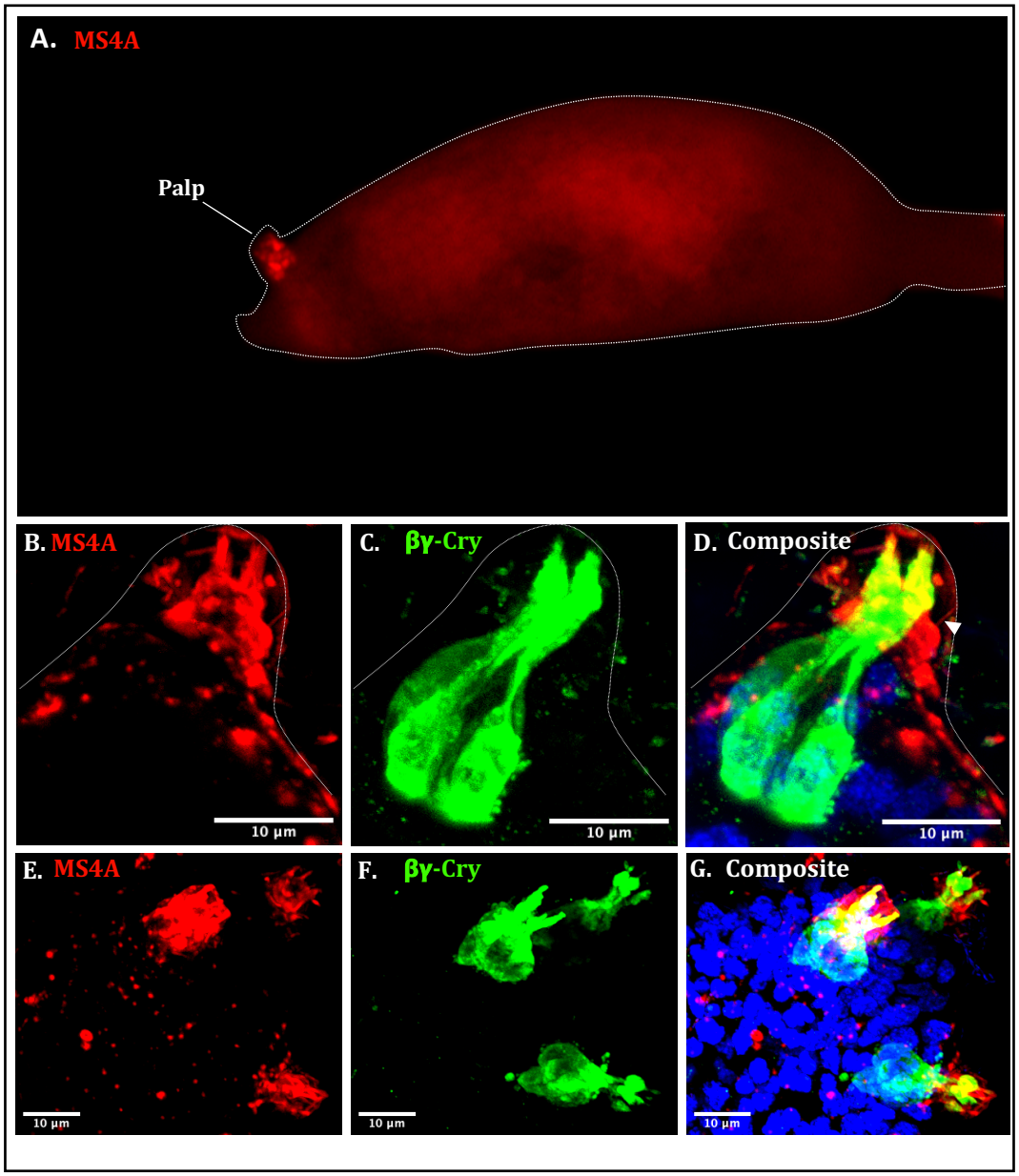


Figure 3.10. Putative MS4A chemoreceptor KH.C1.484 is expressed in the ACCs apical digitiform protrusions. (A) Ciona larva head expressing MS4A reporter, MS4A>mCherry. The mCherry reporter expression is seen clearly in the palp. A single palp is labelled here which is a common occurrence of mosaicism during transgenesis. (B-G) Confocal projections of palp cells of transgenic larvae (24 hpf) with (B-D) individual palp and (E-G) three palps of a single animal. (B & E) MS4A>mCherry signal (red) in digitiform protrusions of ACCs. (C & F) ßγ-Crystallin antibody marking the ACCs (green). (D & G) Composite image with MS4A>mCherry in triple staining with ßγ-Crystallin antibody (ACCs, green), and DAPI (nuclei, blue). The white arrowehead in D represent a putative cilium of PSN. Video files of 3D projections are in the electronic supplementary material. (B-D) Dotted outline around a single palp. Scale bar: 10 µm.

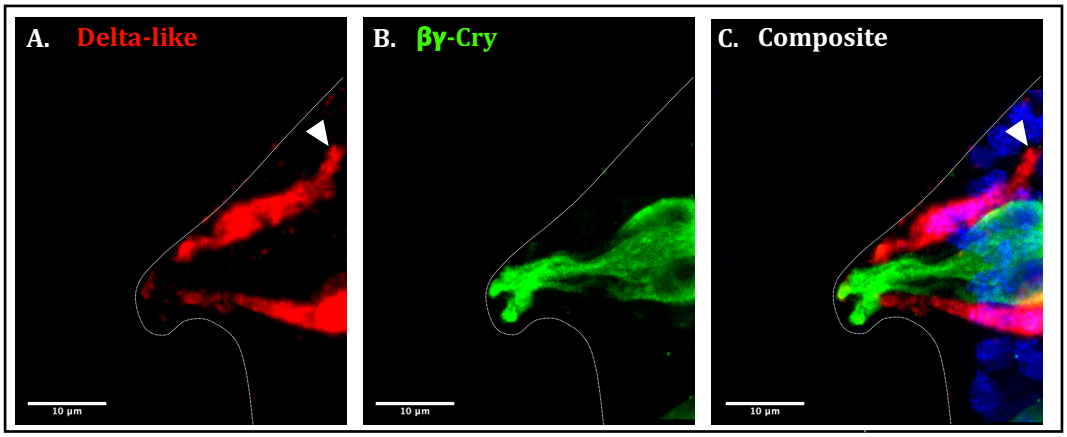


Figure 3.11. The Notch ligand Delta-like is expressed in the PSNs. (A-C) Confocal projections of a palp showing mCherry reporter expression under the known Delta-like regulatory region. The construct Delta-like>mCherry was used as a positive control to validate the tyramide signal amplification methodology in Ciona larvae. (A) Delta-like>mCherry signal (red) in PSNs. White arrowheads show axon of PSNs. (B) ßγ-Crystallin antibody marking the ACCs (green). (C) Composite image with Delta-like>mCherry (red) in triple staining with ßγ-Crystallin antibody (ACCs, green), and DAPI (nuclei, blue). Video files are in the electronic supplementary material. Dotted outline around the palp. Scale bar: 10 µm.

3.3.5 Supplementary Electronic Data

These supplemental data files are available on the "Additional Material" .

- **Video_file3.1** and **Video_file3.4**: GnRH1J>GFP signal (green) in PSNs and ACCs
- **Video_file3.2** and **Video_file3.5**: ßγ-Crystallin antibody marking the ACCs (green)
- **Video_file3.3** and **Video_file3.6**: Composite image with GnRH1J>GFP in triple staining with ßγ-Crystallin antibody (ACCs, red), and DAPI (nuclei, blue).
- **Video_file3.7, Video_file3.10, Video_file3.13, Video_file3.16, Video_file3.19**: MS4A>mCherry signal (red) in digitiform protrusions of ACCs
- **Video_file3.8, Video_file3.11, Video_file3.14, Video_file3.17, Video_file3.20**: ßγ-Crystallin antibody marking the ACCs (green).
- **Video_file3.9, Video_file3.12, Video_file3.15, Video_file3.18, Video_file3.21:** Composite image with MS4A>mCherry in triple staining with ßγ-Crystallin antibody (ACCs, green), and DAPI (nuclei, blue).
- **Video_file3.22**: Delta-like>mCherry signal (red) in PSNs.
- **Video_file3.23**: ßγ-Crystallin antibody marking the ACCs (green).
- **Video_file3.24:** Composite image with Delta-like>mCherry (red) in triple staining with ßγ-Crystallin antibody (ACCs, green), and DAPI (nuclei, blue).

3.4 Discussion

3.4.1 MS4a genes represent the first chemoreceptor family described in ascidian

To date essentially nothing is known about the molecular apparatus of the ascidian's chemosensory system. This chapter represent a first contribution to our understanding, highlighting the presence of *Ms4a* chemoreceptor genes in ascidians. Overall, the presence of Ms4a homologs in most Eumetazoa indicates that the evolution of these genes precedes the advent of other genes used as chemoreceptors in mammals such as for taste (T1/2Rs) and for pheromones (V1/2Rs) (Grus and Zhang, 2009). I have demonstrated that *Ms4a* genes that have recently expanded in number in ascidian such as in *B. schlosseri* have ~90% of sites detected under positive selection in the extracellular loops that are likely responsible for the detection of ligands, which is expected for chemoreceptors. Similar results were indeed obtained in the ligand-binding domain of recently duplicated chemosensory genes in other species such as in centipede, pea aphid and Drosophila (Alemeida et al., 2015; Croset et al., 2010; Smadja et al., 2009). It is therefore tempting to speculate that the Ms4a proteins represent an ancient mechanism to sense relevant chemical cues in the environment. However, it remains to be established that *Ms4a* genes have a chemosensory function in all these species as well as the exact nature of their ligands. Even so, the conservation of the Ms4a receptor repertoire for more than 500 million years of chordate history argues that they have an important and unique role in sensory physiology (Zuccolo et al., 2010). The Ms4a family is characterised by its fast evolution and vast diversity which is a result of lineage- and species-specific duplications and gene loss occurrences, suggesting rapid gene turnover throughout evolution. Synteny data pointed out that the *MS4a* genes mostly group as tight clusters. Eukaryotic genomes tend to cluster genes with similar and/or coordinated expression to facilitate the control of gene expression and related biological processes

(Hurst et al., 2004). Conserved syntenic localization between different chemoreceptor families such as the observed linkage between ORs and Ms4as is likely to represent such regulatory advantage.

3.4.2 ACCs have homologies with vertebrate olfactory neurons

In the palps of *Ciona*, the axial columnar cells (ACCs), collocytes (CCs) and primary sensory neurons (PSNs) share some similarities in their cellular ultrastructure such as basal bodies, various types of granules and vesicles, despite their distinctive morphologies. These subcellular resemblances suggest that the three palp cell-types have specialized from a common ancestral neuroepithelial cell type, maybe a ciliated neurosecretory cell (Zeng et al., 2019a). This study revealed that the ACCs have dual GnRH neurosecretory and Ms4a chemosensory activities. The PSNs may also have such dual GnRH/Ms4a functions but the expression of Ms4a in those cells was not establish with certitude, although cilia have only been noticed in PSNs. Taken together, the GnRH transgene expression points to a close relatedness of ACCs and PSNs consistent with a common ontogeny and/or ancestry of these two sister cell types. It must be noted that shared expression of effector genes (i.e., genes encoding cell-type specific structures and functions) such as GnRH neurohormones and Ms4A chemoreceptors between two different cell types (within the same or between species) are not always indicative of a profound homology through shared ancestry. They can also indicate homoplasies because of convergent evolution/parallelism. Support for homology becomes more robust when it is supported by the expression patterns of conserved “core regulatory complex” (CoRC) which regulate cell-type specific effector genes (Arendt et al., 2016). A CoRC contains the set of transcription factors and their mutual interactions that allow and sustain differential gene expression of a cell-type inside an organism. The evolution

of independent gene expression is what enables new cell types to become different from their evolutionary sister cell types. In theory, the CoRC from homologous cell types is conserved while their phenotypes are more flexible and can acquires lineage-specific differences (Arendt et al., 2016). Recently, the ACCs were advocated to be a non-neural myoepithelial cell type homologous to vertebrate smooth muscle cells, myoepithelial cells, and cardiomyocytes (Johnson et al., 2020). In support of this homology was the observation that the ACCs are contractile and that they express orthologs of vertebrate smooth muscle/myoepithelial cell effectors and share a conserved CoRC with vertebrate cardiomyocyte transcription factors (TFs). However, the presented evidence for such homologies has several caveats. For instance, the homologues of the transcription factors said to form a CoRC in *Ciona* ACCs were expressed at various time points in the precursor cells of the papillae. It was unclear whether these TFs bind DNA altogether as a single complex in the ACCs which would be expected for members of a CoRC. Furthermore, the *Ciona* TFs said to form a CoRC were not necessarily true orthologues to their vertebrate counterparts in the cardiomyocytes. Here, I do not argue against the contractile properties of the ACCs but question their suggested deep homology with vertebrate smooth muscle cells, myoepithelial cells, and cardiomyocytes. In fact, the data presented in this Chapter, suggest an alternative scenario and I rather hypothesize that ACCs have homologies with the placode-derived olfactory neurons of vertebrates. Several lines of evidence appear to support such a model. First, the ACCs are derived from an anterior proto-placode territory and develop through similar developmental genes seen in the anteriormost olfactory placodes of vertebrates (cf. Chap1). These correspondences respect the principle historical criteria for homology, similarity of position. Second, the ACCs have dual GnRH neurosecretory and MS4A chemosensory functions. This chemosensory/GnRH duality was advocated before for the aATENs a type of primary

sensory neurons also derived from the anterior proto-placode (Abitua et al., 2015). It is argued that multifunctionality is a common feature of ancient metazoan cell types (Arendt, 2008). In fact, homologous cell types conserved over long evolutionary distances tend to have multiple and distinct cellular functions. For example, the light-sensitive vasotocinergic cells are said to resemble protoneurons with multifunctional sensory and neurosecretory functions that are considered ancestral to the vertebrate brain (Vigh et al., 2004). As with the increase in organismal complexity during evolution, these multiple functions were subsequently distributed among emergent sister cell types, resulting in cell type specialized descendants. This process known as segregation of functions means that one sister cell type specifically loses a function that the other maintains and is thought to enable the evolution of more complex neuronal circuits (Arendt, 2021). At the gene level, functional segregation implies the selective loss of expression of the corresponding effector genes encoding this function and of some upstream transcription factors. For example, this process of functional segregation is said to have happened during the diversification of the cnidarian myoepithelial cell with sensory, epithelial, and contractile functions into separate muscle cell, motor neuron and sensory cell (Mackie, 1970).

3.4.3 Evolutionary scenario for the diversification of olfactory neurons in Olfactores

Hereafter, we propose an evolutionary scenario to explain how the ascidian ACCs and vertebrate olfactory neurons may have evolved from an ancestral olfactory cell type present in the last common Olfactores ancestor. Indeed, the ACCs may be a good proxy of what the ancestral olfactory neuron would have looked like as they harbour an ancestral characteristic through their multifunctionality with at least neurosecretory, chemosensory, and contractile functions. In the past, the ACCs of *Ciona* have been advocated to be neurons rather than non-neural myoepithelial cells as other ascidian species such as *Phallusia mammallita* and *Clavelina lepadiformis* showed axon-bearing central ACCs in their papillae (Dolcemascolo et al., 2009; Stogia et al., 1998; Pennati et al., 2009). In *Ciona,* the ACCs have always been shown to lack axons or neurites of any kind, which would mean that the axons of ACCs would have been secondarily lost (Zeng et al., 2019, this study). This is not unlikely as the lack of axons is an idiosyncratic feature of several neurons in *Ciona* (Ryan and Meinertzhagen, 2019). In addition, their ultrastructure points towards a hidden neuronal nature as they contain numerous neuroactive vesicles (likely peptidergic, aminergic or cholinergic) (Zeng et al., 2019a). In this study, we showed that ACCs expressed a GnRH neurohormone as well as a 4 transmembrane domains (4TM) Ms4a chemoreceptor (KH.C1.484) on their apical tips. In mammals, the Ms4a proteins also localise to the dendritic apical endings of necklace OSNs. The necklace OSNs do not express one type of chemoreceptor for each neuron as is the case with olfactory receptors (ORs) in conventional OSNs, instead many different Ms4a members are expressed in each necklace OSN. The same pattern seems to occur in

Ciona as at least 10 different Ms4as are expressed at the larval stage in the ACCs according to Sc-RNAseq data (Sharma et al., 2019; Johnson et al., 2020). The ACCs also express an orphan 7 TM receptor (KH.C3.516) that is enriched at their exposed apical tip and has been suggested to be an OR (Johnson et al., 2020). Syntenic data indicate that this gene as well as other members in this orphan 7 TM receptor gene family are linked to *Ms4a* genes, *Mucin5ac* and *Notch1* similar to what is observed with *Ms4as* and *ORs* genes in other chordates (Fig. 3.5). It remains a possibility that these orphan 7 TM receptor genes are indeed orthologues to vertebrate ORs but are not revealed as such by sequence analyses due to an extreme degree of divergence resulting from the fast-evolving nature of those chemoreceptor genes in *Ciona*. The present-day olfactory neurons of mammals deriving from the olfactory placodes could have thus diversified from an ACC-like multifunctional olfactory neuron by segregation of function into three distinct cell types. In fact, the anteriormost olfactory placodes of mammals produce conventional OSNs expressing 7TM G-protein coupled chemoreceptors (ORs), the necklace OSNs expressing 4TM MS4A chemoreceptors, and GnRH neurons (Fig. 3.12).

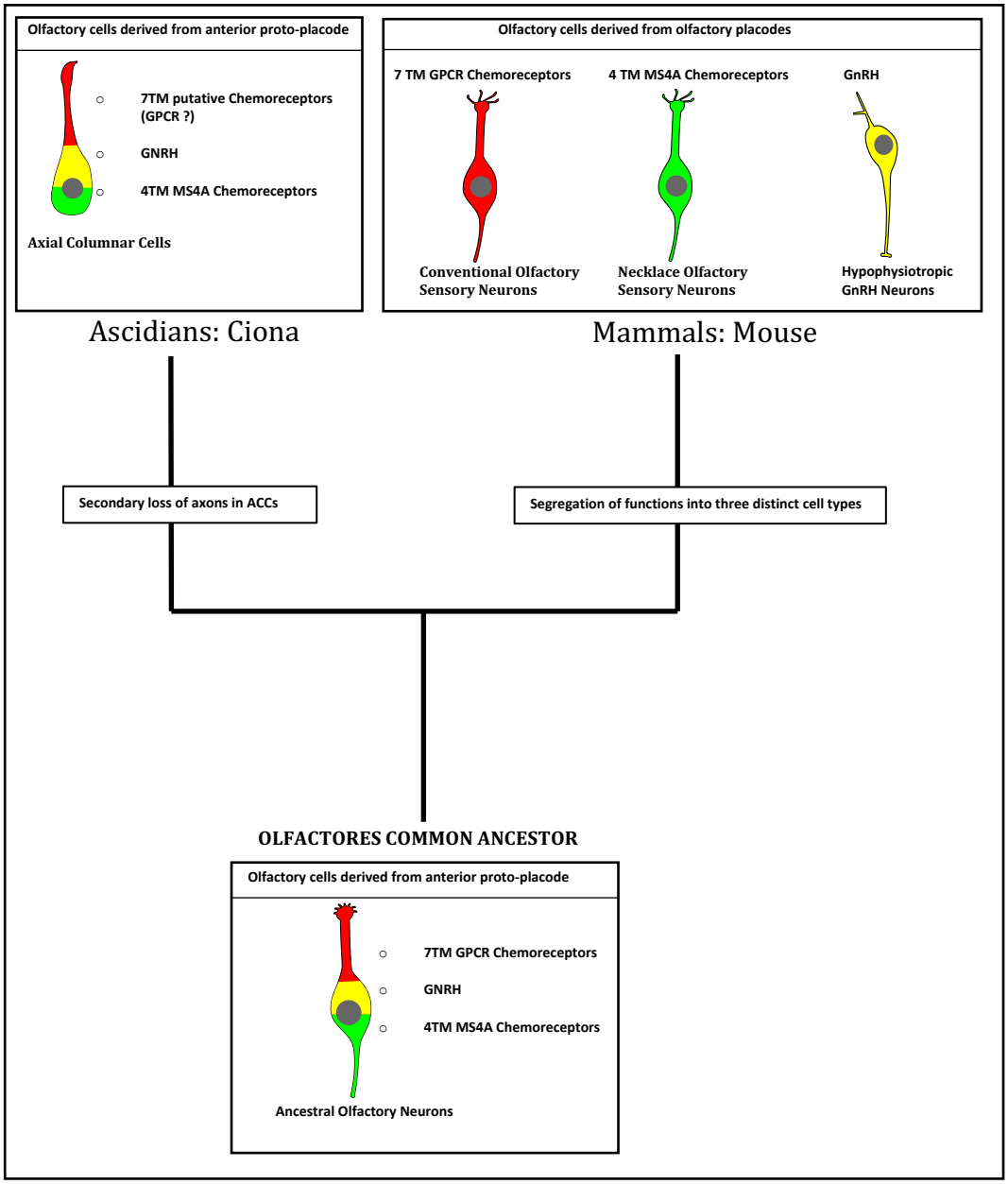


Figure 3.12. Evolutionary model for the origin and diversification of olfactory neurons in the Olfactores lineage: a hypothesis. In the common ancestor of Olfactores, the ancestral olfactory neurons derived from an anterior proto-placode territory and expressed GnRH neurohormone as well as chemoreceptors with 7 TM coupled to G-protein and 4 TM from the MS4A family. This condition is preserved in the palp ACCs of the ascidian *Ciona intestinalis*. It is probable that the ACCs secondarily lost their axons, as it is an idiosyncratic feature of several neurons in *Ciona* and that axons on ACCs have been said to exist in several other ascidian species. The modern-day olfactory neurons of mammals deriving from the olfactory placodes would have diversified from that ancestral olfactory neuron by dividing the labour into three distinct cell types. In fact, mammals have conventional OSNs expressing 7TM GPCR chemoreceptors, the necklace OSNs expressing MS4A chemoreceptors and GnRH neurons, all derived from the anterior olfactory placodes.

Chapter 4 Development of microfluidic technology for the detection of chemosensory neuronal responses in *Ciona intestinalis* larvae

Work Declaration

The experimental work and subsequent data analyses presented in this Chapter have performed by me, Guillaume Poncelet.

4.1 Introduction

4.1.1 Research context

To date, studies on *Ciona* chemosensory behaviour have been restricted to largely subjective observational studies in adult ascidian (Hecht, 1918). While such studies offer useful insight into *Ciona* chemical behaviours they do not necessarily facilitate the objective testing of specific hypotheses, under controlled laboratory conditions. The development of reliable chemical testing technology is therefore critical for developing an improved understanding of *Ciona* ecology and physiology. In fact, despite several structures advocated to be chemosensory in *Ciona* (cf. Chapter 1), there has not been a single cell shown to possess genuine chemoreceptors which are orthologous to the ones detected in the olfactory system of other invertebrates or vertebrates (until this study, cf. Chapter 3). Also, the type of chemical cues leading to the larva locomotion and attraction or to start metamorphosis are unknown. Especially, how they are detected and what physiological responses they trigger remain to be investigated (Poncelet and Shimeld, 2020). Furthermore, no physiological studies have been performed in this model tunicate, due to the small size of the larva and its cells and inherent difficulty this creates for conventional electrophysiological approaches.

4.1.2 Chapter Aim & Strategy

The present chapter aims to develop a solution to the existing limitation that no cells or organs claimed to be chemosensory can be functionally tested in tunicates (or more broadly in any protochordates). The purpose of this study was to design an experimental tool enabling the investigation of the supposed chemosensory systems of *Ciona* larvae. Ideally, the system developed should allow me to test the *in vivo* physiology of chosen cells and be applied to any type of chemical stimuli. Such a system would be essential to provide comprehensive and unbiased knowledge of chemosensation in *Ciona*. To design such technology, my initial point of investigation was elegant methodology developed by Chartier et al. (2018). These authors used an interesting approach to record chemosensory neural responses in larva of the marine polychaete *Platynereis dulmerii* by combining microfluidics and calcium imaging. A similar methodology was thus attempted for *Ciona* inspired by the recent success of innovative experiments using calcium sensors for physiological characterisation of calcium transients during embryogenesis (Hackley et al., 2013; Abdul-Wajid et al., 2015; Akahoshi et al., 2017; Johnson et al., 2020; Okawa et al., 2020). Such experiments have demonstrated the relevance of calcium imaging methods to study the physiology of embryonic cells in *Ciona* larvae. Furthermore, customised microfluidic chips have been shown to be a robust methodology to test chemosensory activity and responsiveness in terrestrial and marine invertebrates and vertebrates alike (Ramanathan et al., 2015; Hu et al., 2015; Candelier et al., 2015; Reilly et al., 2017; Chartier et al., 2018).

To make calcium imaging of chemosensation widely applicable in *Ciona* larvae, two main issues need to be solved: suppression of movement and precise stimulus delivery. In fact, from 17 hours post fertilisation (hpf) onwards, the *Ciona* larva is highly motile due to the beating of the tail endowed with a substantial complement of muscle cells (Satoh, 1994). So far, the immobilisation strategies used for calcium imaging in *C. intestinalis* embryos relied either on the embedding of the animal in low-melting point agarose (Abdul-Wajid et al., 2015) or trapping the larva between a glass-bottom dish and coverslip (Okawa et al., 2020; Johnson et al., 2020). Actually, both techniques preclude repeated chemical stimulus delivery, at most a chemical stimulus can be delivered once but cannot be washed off. In consequence, these two ways of immobilisation limit stimulus delivery to photic and mechanical stimuli but also suffer from low reproducibility which makes precise testing even more challenging. In particular, they do not allow a stimulus to be repeatedly turned on and off while responses are recorded, a "gold standard" in general sensory physiology studies. Therefore, a more elaborated strategy for immobilisation and stimulus delivery is needed to demonstrate that *Ciona* cells respond specifically to a certain chemical stimulus via calcium imaging. Thus, a customised microfluidic trap for *Ciona* could offer a first solution for immobilisation and enable to expose the same larval cells to repeated chemical stimuli together with a control.

4.1.3 Calcium imaging

Neuronal activity causes fluctuations in the concentration of intracellular calcium ions (Ca^{2+}). This principle is broadly used to image neuronal activity with the use of fluorescent calcium sensors. There are two main categories of fluorescent calcium sensors: synthetic chemical sensors and genetically-encoded protein sensors

(Grienberger et al., 2012). The synthetic chemical sensors (e.g., Fura-2, Fluo-3) complex Ca^{2+} ions and respond to this fixation by changing their initial properties of fluorescence, either by shifting their emission or excitation spectra. Their advantages are their great sensitivity and that they are readily available as they pass through cellular membrane. Therefore, synthetic dyes are ideal for cell cultures and tissues and are introduced into cells via electroporation, lipotransfer, microinjection or diffusion from patch clamp pipettes or simple immersion (Paredes et al., 2008). The genetically-encoded protein sensors include those involving Förster resonance energy transfer (FRET) and the single-fluorophore GCaMP (for GFP-CalModulin Probe) (Nakai et al., 2001). The latter GCaMP and its derivatives consist of the circularly permutated green fluorescent protein (cpGFP), the calcium-binding protein calmodulin (CaM) and the M13 peptide. The binding of Ca^{2+} by CaM result in an interaction with M13 (CaM-M13). The conformational change of the CaM-M13 complex induce a change in solvent access and pKa of the GFP chromophore which greatly increase the GCaMP fluorescence brightness (Barnett et al., 2017). This increased in brightness is compared to the basal fluorescence level (prior to neuronal excitation) to assess if there is a neuronal sensory response. Historically, genetically-encoded protein sensors were less sensitive and with slower kinetics than the most commonly used synthetic calcium indicators. Today, the structure-guided optimisation of GCaMPs by protein engineering has led to the production of protein-based indicators that surpass chemical sensors in all aspects. Among these are the sixth generation of GCaMP proteins used in this study, GCaMP6, which are brighter than synthetic dyes (Fig. 4.1). There are three different type of GCaMP6 harbouring slow, medium or fast kinetic properties: GCaMP6s (slow), GCaMP6m (medium), GCaMP6f (fast). The GCaMP6s is the brightest but has the slowest kinetic, while GCaMP6f has the fastest kinetics but is less bright. For the purpose of my experiment, I chose GCaMP6m

which showed strong levels of fluorescence (almost as strong as GCaMP6s) but had a faster kinetics, which seemed more appropriate for the recording of chemosensory responses. It takes in the order of 0.2-0.3s for GCaMP6m to reach its maximum fluorescence intensity when a neuron is stimulated (Chen et al., 2013).

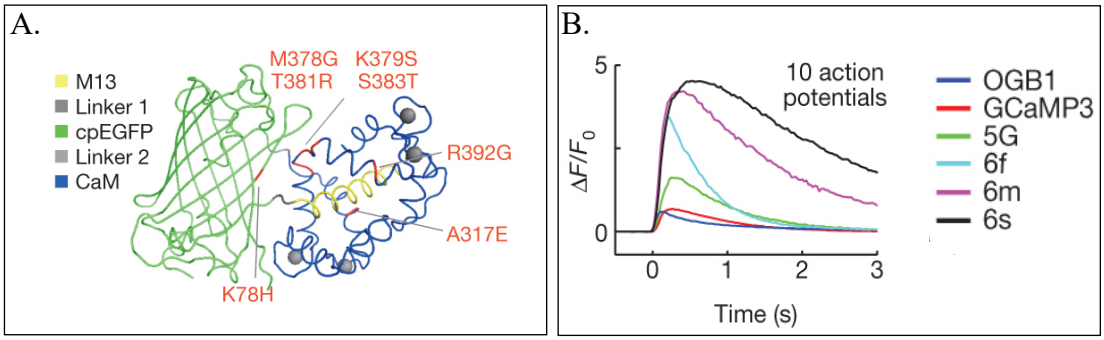


Figure 4.1. (A) GCaMP structure and mutations in different GCaMP6 variants relative to GCaMP5G. The grey balls represent four calcium ions binding calmodulin (CaM). (B) shows the response kinetics of the three GCaMP6 versions (6s, 6m, 6f), compared to the fifth (5G) and third generation (3) GCaMP and to the synthetic dye OGB1. After one or several action potentials, it takes longer for 6s to reach its maximum level of fluorescence, and also to relax to its baseline level**.** Figures are reprinted from Chen et al., 2013, with permission from Springer Nature.

The genetically-encoded protein sensors have an enormous advantage over chemical ones in that they can be permanently expressed through transgenesis and can easily be targeted to cell types and organs of interest with specific regulatory regions. Genetically-encoded protein Voltage sensors such as Voltron also exist, which capture neuronal events with a much better time resolution than GCaMP (Abdelfattah *et al.*, 2019). However, they still require further development before they can be considered an established technique in the field, but they are promising tool worthy to keep an eye upon. Therefore, I chose calcium imaging as a method since my aim was first to localise any existing chemosensory activity in the nervous system of Ciona.

4.1.4 Choice of a chemical stimulus: CO_2 detection hypothesis

Once the type of imaging methodology was chosen, a suitable chemical stimulus had to be selected in order to elicit a potential neuronal response in *C. intestinalis* larvae. The careful analysis of published single cell RNA-seq data and *in situ* data revealed that palp cells express a carbonic anhydrase gene (KH.C13.119) (Kuskabe et al., 2002; Sharma et al., 2019; Johnson et al., 2020). Carbonic anhydrase is a widespread enzyme which plays a role as a CO_2 sensor from prokaryotes to mammals (Bahn and Mühlschlegel, 2006). They are zinc-containing metalloenzymes that mediate the reversible hydration of CO_2 into carbonic acid (H_2CO_3), protons (H^+) and bicarbonate ions (HCO^-). The ability to detect CO_2 is widespread in living-organisms and is monitored in the body of animals to ensure internal and external homeostasis. CO_2 emitted by respiration is also used to detect a food source or as an avoidance cue to stay away from predators (Scott, 2011; Jones, 2013). For instance, mammals can sense environmental CO_2 through the gustatory or olfactory system. The olfactory neurons that detect CO_2 in mammals are known as the necklace olfactory sensory neurons. The necklace OSNs express a complement of signalling molecules for CO_2 detection totally different to those of conventional OSNs. Necklace OSNs target axons to the necklace glomeruli, which are anatomically segregated from other projections of the conventional OSNs. This difference hints to a distinct subsystem to the main olfactory system for CO_2 detection (Greer et al., 2016). The combined presence of MS4A putative chemoreceptor genes and carbonic anhydrase in the palps is suggestive of a functional similarity between the ascidian palp cells and the mammalian olfactory necklace sensory neurons (cf. Chapter 3). In consequence, I developed the hypothesis that *C. intestinalis* and especially the palp cells could sense dissolved CO_2 in sea water.

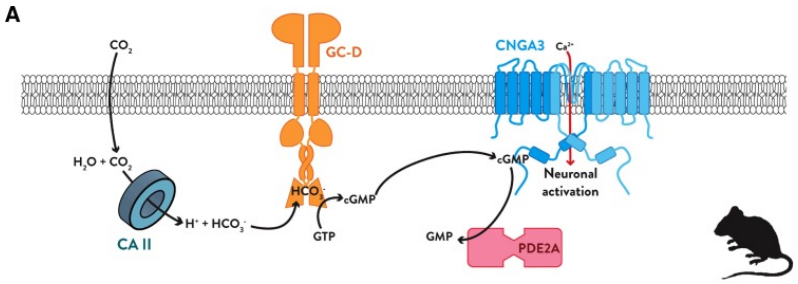


Figure 4.2. (A) In the mouse system, CO_2 diffuses into the necklace olfactory sensory neurons. Its hydration is catalysed by the carbonic anhydrase CAII to form carbonic acid, which immediately dissociates to form protons and bicarbonate. The bicarbonate activates the receptor guanylate cyclase GC-D, which converts bound GTP to cGMP. The cGMP then binds to the cyclic nucleotide fated channel CNGA3 and causes it to open and permit the entry of calcium ions that initiate action potentials that travel to the necklace glomeruli into the olfactory bulb. The response to CO_2 is then terminated when PDE2A converts the cGMP into GMP. Figure is reproduced from Jones, 2013 with permission.

4.1.5 Chapter Summary

The results shown in this chapter are the first direct evidence that the nervous system of *Ciona intestinalis* larvae respond to environmental chemosensory inputs, in particular CO_2, and provide a template for the development of a new technique of great potential interest in the field. As shown in the results section, it is possible to record calcium activity in the nervous system of a 20-24hpf swimming Ciona larvae over several minutes. This constitutes a first milestone in the establishment of a mature functional imaging method and has been enabled using a low-height channel trap in the microfluidic device. However, the current design showed technical limitations and further improvements are suggested in the general discussion. The long-term prospect of such technology is very exciting for the numerous insights it could bring to the field and more generally for our greater understanding of sensory system evolution in chordates.

4.2 Material and methods

4.2.1 Plasmid construction and electroporation

The DMRT regulatory region was amplified from the DMRT>GFP plasmid, which was kindly provided by Dr Laurence Lemaire (Wagner et al., 2014). Amplification was done with Phusion polymerase (New England Biolabs) and a pair of gene specific oligonucleotide primers: 5'-ACGAGATCTTAGTAGGGTGGAGGAAGATGG-3' and 5'-ACGCCTAGGGCCAGTTAAACGAACTGTTTG-3'. The DMRT regulatory region was inserted into the plasmid pAAV-hSyn1-mRuby2-GSG-P2A-GCaMP6m-WPRE-pA cut with AvrII and BglII restriction enzymes. The complete construct plasmid sequence is available as an electronic .dna files (see Appendix).

4.2.2 Validation of GCaMP6m as a tool to record calcium transients in *Ciona*

Ciona larvae were immobilised in low-melting point agarose inside a perfusion slide (µ-Slide Luer, ibidi). The latter was perfused with sea water and recorded for GCaMP6m fluorescence: to do this GCaMP6m was excited at 488 nm wavelength and the fluorescence was detected with a Fluorescence Axiozoom V.16 microscope, equipped with a Plan-NEOFLUAR Z 2.3x (NA 0.57). GCaMP6m fluorescence was recorded for 30 sec (300 frames at 10 fps). The sea water solution was then switched, and the slide was perfused with 1M calcium ionophore A23187 diluted in sea water to increase intracellular Ca^{2+} in embryonic cells. GCaMP6m fluorescence was then also recorded for 30 sec (300 frames at 10 fps). The identification of the responding cells was done using Fiji by calculating the standard deviation through all 300 frames in both conditions, i.e., with or without the addition A23187.

4.2.3 Microfluidic chip fabrication

Standard soft lithography was used to fabricate the mould (Xia and Whitesides, 1998). The photomask was designed with AutoCAD (2019 free student version, Autodesk, Inc.). Premium grade high resolution film photomasks (features down to 5 μm) were printed by an external company (Micro Lithography Services Limited, Chelmsford, UK). The photomask source file is available as an electronic .dwg file (see Appendix). The chip mould with a uniform height of 23.5μm was obtained by spin-coating a silicon wafer (4 inches; Siltronix, France) with a negative photoresist (SU-8 2035, MicroChem Corp., Newton, MA, USA) according to manufacturer instructions. Devices were produced by pouring onto a mould a prepolymer mixture of polydimethylsiloxane (PDMS, Sylgard 184 silicone elastomer kit, Dow Corning Corp.) with a 1:10 ratio of curing agent and curing at 65°C for a minimum of 4 h. PDMS blocks were then irreversibly bound to a 1mm thick standard microscope slide (25mm x 75mm) by a 1 min treatment in a plasma oven. The whole process of PDMS microfluidic chip fabrication was done in the laboratory of Professor Dirk Aarts (Oxford Colloid Group, Physical and Theoretical Chemistry) under the supervision of Dr Lucia Parolini.

4.2.4 Experimental set-up and procedure

Polytetrafluoroethylene tubings (Microbore PTFE, outer diameter 0.030", inner diameter 0.012", Cole-Parmer ltd, UK) were cut to approximately 30cm lengths and fed into the device inlets, while tubing with a bigger diameter was used for the outlet (PTFE, TW24, inner diameter 0.59 mm, Adtech Polymer Engineering Ltd, UK). All the tubings are connected to the chip by forcing them into the PDMS holes with tweezers, ensuring sufficient sealing at the ranges of pressure used here. The inlet tubings connected to Falcon tubes containing the test solutions. Fluid flow through the microfluidic device was driven by the controller MFCS™-EZ (Fluigent) containing three pressure channels alimented by an integrated positive pressure source (0-345 mbar). The MFCS™-EZ received commands from an automated programme built with the Fluigent Microfluidic Automation Tool (SFT-MAT). Water streams were generated in a laminar flow regime**.** Pressure rates of 0 mbar, 60 mbar, 80mbar, 120 mbar were used in the different channels (Fig. 4.9), with the total pressure in the chip kept constant (200 mbar) to minimise pressure change experienced by the animal. Image acquisition and stimulus delivery were synchronized with AUTO MOUSE CLICK (MurGee.com), a software tool for automated mouse actions. A customized chip holder was built by 3D-printing (Bodleian Library 3D printing and scanning services) to hold the device and enable its observation under an upright microscope. The 3D chip holder source file and details are available as a .stl electronic file (see Appendix).

4.2.5 Animal handling and preparation

Adults of *Ciona intestinalis*, were collected from Sparkes Yacht Haven or Northney Marina, UK and kept in an aquarium with circulating and oxygenated sea water at 16°C. Gametes were liberated by dissection and cross fertilisations performed *in vitro* for 15 min. The DMRT>GCaMP6m plasmid was used at a final concentration of 50-100 µg per 50 µl in electroporation, as described previously (Kari et al., 2016). After electroporation, eggs were kept in filtered natural seawater until they reach the larval stage (20 hpf). The microfluidic chip was washed twice with artificial seawater before used. The transgenic larvae were sucked into a PFTE tube plugged to a metallic needle (Microlance #20, 302200, BD, USA) connected to a plastic syringe (Luer Plastipak, BD, USA). The animals inside the PFTE tube were manually fed inside the introduction inlet by applying light pressure on the plastic syringe to push slowly the larvae inside the trapping channel. Calcium imaging experiments were then conducted at room temperature between 20-24 hpf. After the experiments, the animals were flushed out of the device by applying maximal pressure on the syringe filled with artificial sea water.

4.2.6 Test solution preparation

Artificial seawater (ASW) with a salinity level of 35 ppt, a carbonate hardness (dKH) of 8 and a pH of 8 was obtained by mixing 35g of pharmaceutical grade sea salt (PRO-REEF Sea Salt) per litre of ultrapure water. CO_2 was dissolved in ASW using a SodaStream with a built-in CO_2 cylinder. The made-up ASW solution with infused CO_2 (ASW-CO_2) had a lowered pH of 6. This decrease in pH relates to an estimated CO_2 concentration of 129.3 mg/L (ppm) in comparison to non-injected ASW at 1.1 mg/L (ppm) of CO_2, based on

calculations from Mojica Prieto and Millero, 2002. Due to the lowering effect of CO_2 on the pH, a control ASW solution was made with a corrected pH of 6 by gradual addition of 1M HCl. The stock solutions were kept at 18°C in glass bottle for long-term storage and loaded in 15mL Falcon tubes on the day of the experiment to avoid possible detection of dissolved substances from plastic containers.

4.2.7 Calibration experiments

As the different flow streams will be transparent in the final experiment, something was needed to ensure that the automated protocol was moving flow boundaries and enabling chemical stimulation on the larvae. To ensure visibility, the dye Fast Green FCF (E143) was dissolved in the two side streams to visualize their moving boundaries at 0.5 g/mL. Calibration recordings were done with a Carl Zeiss Axioskop 2 plus microscope mounted with a Zeiss AxioCam HRc camera. A single frame was taken every 6.8 msec. The mean grey value through all frames was calculated using Fiji, in a square region of interest, constant in size and position, located in the trapping channel (yellow rectangle in Fig. 4.12). Minimal and maximal pixel values, which according to the Beer–Lambert law corresponded, respectively, to the absence of stimulant and maximum stimulant concentration, were normalized between 0 and 1. The calibration experiments allowed to quantify the beginning and ending of stimulation onsets and offsets. Measurements were made successively 60 times.

4.2.8 Imaging and analysis

GCaMP6m fluorescence excited at 488 nm wavelength was detected with a Fluorescence Axiozoom V.16 microscope, equipped with a Plan-NEOFLUAR Z 2.3x (NA 0.57). 3000 images were taken at 10 fps (100ms exposure time) in 5 min (300 sec). A pixel binning of 5X5 was done to increase signal-to-noise ratio. Movement artefacts on the raw calcium recordings were first corrected in FIJI (v.2.1.0) using the plugin TurboReg with rigid body transformations (Thévenaz et al., 1998). Mean fluorescence intensity was then calculated from regions of interest (ROI) drawn manually in FIJI. The responsive ROI were assessed by eye from raw recordings. The attribution of a ROI to a specific embryonic region relied on its relative position guided by anatomical landmark recognition such as the palps, tail, ocellus, and otolith. Further data analysis was done in Excel, following the guidelines of Akira Muto (http://akiramuto.net/archives/148). Traces were plotted as $\Delta F/F_0=(F_t-F_0)/F_0$, with F_t the fluorescence intensity at time t, and F_0 the mean fluorescence value over a 5 s time window during the initial 30 s resting state, i.e., outside of stimulation periods. The fluorescence change ($\Delta F/F_0$) was calculated after subtracting the background fluorescence. The raw recording files are available as .avi in the electronical additional material.

4.3 Results

4.3.1 Assessment of calcium imaging feasibility with GCaMP6m in *Ciona* larva

The first step was to assess if it was possible to do calcium imaging with the genetically-encoded calcium sensor GCaMP6m in *Ciona* larva. The GCaMP6m was targeted to specific cells in *Ciona* by using the characterised ~1 kb *Dmrt* regulatory region (Wagner and Levine, 2012). This was chosen because it has been shown to express in the correct cells in *Ciona*, and because vertebrate *Dmrt* orthologues are expressed in the anterior placodal domain giving rise to the olfactory placodes and play a role in neurogenesis of the olfactory epithelium (Winkler et al., 2004; Huang et al., 2005; Hong et al., 2007; Parlier et al., 2013). At the mid-gastrula stage in *Ciona*, *Dmrt* is expressed in rows III, IV, V, VI of the neural plate. Rows III and IV contribute to the CNS, by giving rise to the anterior sensory vesicle. On the other hand, rows V and VI contribute to the palps, oral siphon primordium and peripheral nervous system (Tresser et al., 2010; Gainous et al., 2015). For a first examination, a regulatory region such as *Dmrt* enabling the investigation of a broad expression territory was crucial to investigate chemosensory responses in candidate organs such as the palps while still being able to reveal other unknown sensory structures or cells of downstream circuits related to the chemosensory systems. The Dmrt>GCaMP6m construct was electroporated into fertilised zygotes to create transgenic expression of the genetically-encoded calcium sensor throughout development. To determine whether Dmrt>GCaMP6m was correctly expressed, responding to cytosolic Ca^{2+} increase and visible in recordings, I first created a simple system where I immobilised transgenic *Ciona* larvae in low-melting-point agarose in a perfusion slide. Immobilised larvae were recorded before and after the addition of the calcium ionophore A23187. This ionophore transports external Ca^{2+} through cellular

membranes thus creating a sharp increase in cytosolic Ca^{2+}, similar to what happens in vertebrate OSNs when a chemical stimulus hits a chemoreceptor and triggers an increase of intracellular Ca^{2+} through CNG or TRP channels. This proof-of-principle experiment represented a positive control to ensure that the system was responding to sudden calcium influx in *Ciona* and to see if measurement of calcium transients was possible within the scope of this experiment. The responding cells were identified by calculating the standard deviation through all frames to identify the pixels whose value fluctuated the most (Fig. 4.3). This way, it was visually demonstrated that the expression of the calcium reporter GCaMP6m by the *Dmrt* regulatory region was sufficiently high to record calcium transients. The imaging quality obtained with the microscope used in this study does not allow to distinguish individual cells with ease but is sufficient to see enough anatomical details to localise certain regions and measure their calcium activity.

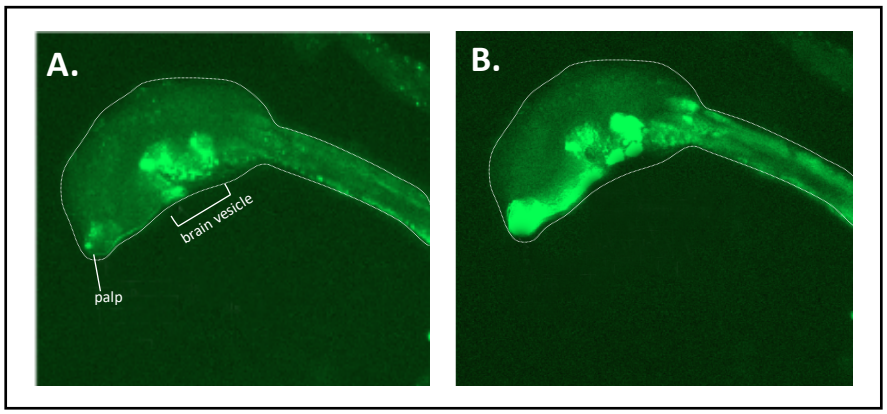


Figure 4.3. Identification of responding cells from GCaMP6m fluorescence by standard deviation through 3000 frames.

A Ciona larva was perfused with sea water (A) or sea water with 1M Ca^{2+} ionophore A23187 (B). A clear increase in fluorescence signalling is seen visually in (B), when the ionophore is present compared to the resting condition in (A). Note that the figures above are produced by averaging across all individual frames and show pixels whose value fluctuated most. They do not represent simple snap shots, i.e., single selected frame.

4.3.2 Design of a microfluidic chip for *Ciona* larva

4.3.2.1 Immobilisation strategy

The challenge in doing tests on live *Ciona* larvae was to be able to immobilise them while submitting the larvae to a controlled sea water flow with or without dissolved chemicals of interest, in this case CO_2. The initial idea was to use the exact same chip design as developed by Thomas Chartier that was applied successfully for six days post-fertilisation larvae of the marine annelid, *Platynereis durmerilii* (Fig. 4.4A-D) (Chartier et al., 2018). This chip enables precise chemical delivery using three parallel flow streams. These three different streams can switch their respective boundaries to expose the animal to the different test solutions (Fig. 4.4D). The chip holds the animal dorso-ventrally and laterally inside a trapping channel (Fig. 4.4B).

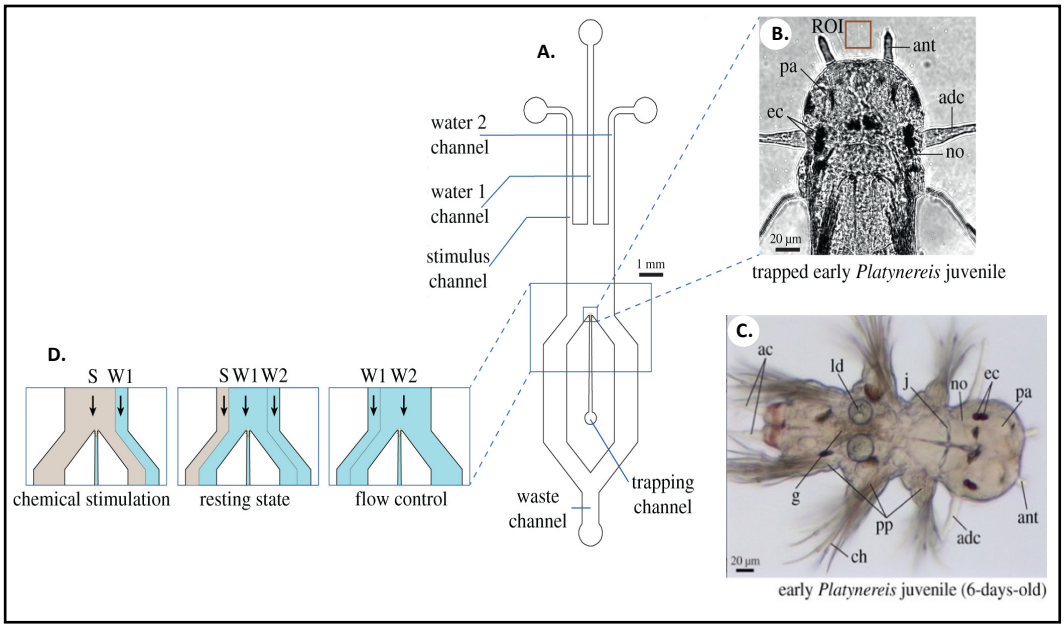


Figure 4.4 Microfluidic device for precise chemical stimulations in the marine larva of *Platynereis dumerilii*. (A) Schematic of the microfluidic device of Thomas Chartier. The three inlet channels are operated by computer-controlled pressure syringe-pumps. A single animal is introduced manually in the trapping channel and immobilized at its end. (B) An immobilized early juvenile, with its head freely exposed to the seawater flows. (C) Light microscopy pictures of *Platynereis* at the early juvenile stages, showing antennae (ant), palps (pa) and tentacular cirri (tc, adc, avc, pdc, pvc). (D) the different flow patterns used to expose the larva to three different test solutions (S, W1, W2). Figure adapted and reprinted from Chartier et al., 2018 with permission.

To our knowledge, no studies prior to this one had ever attempted to introduce a *Ciona* larva inside a microfluidic chip. Therefore, the first step was to identify the relevant dimensions to trap the *Ciona* larva in the PDMS chip and assess how they generally behave inside a microfluidic device. To find the appropriate dimension of a *Ciona* larva to trap it inside the microfluidic chip design of Chartier *et al.* (2018) I first produced a preliminary chip to define the relevant measurements needed to immobilise the larva dorso-ventrally and laterally. This chip consisted of a trapping channel with a constant height of 60 µm, whose width decreased linearly along its main axis from 150 µm (more than the head width) to 20 µm (less than the head width). In *Ciona* larva, the head is the limiting size as it is the widest part of the body, thus knowing precisely what the head dimensions are is an absolute necessity to trap the larva efficiently. The principle of head width determination is illustrated in Fig. 4.5A-B.

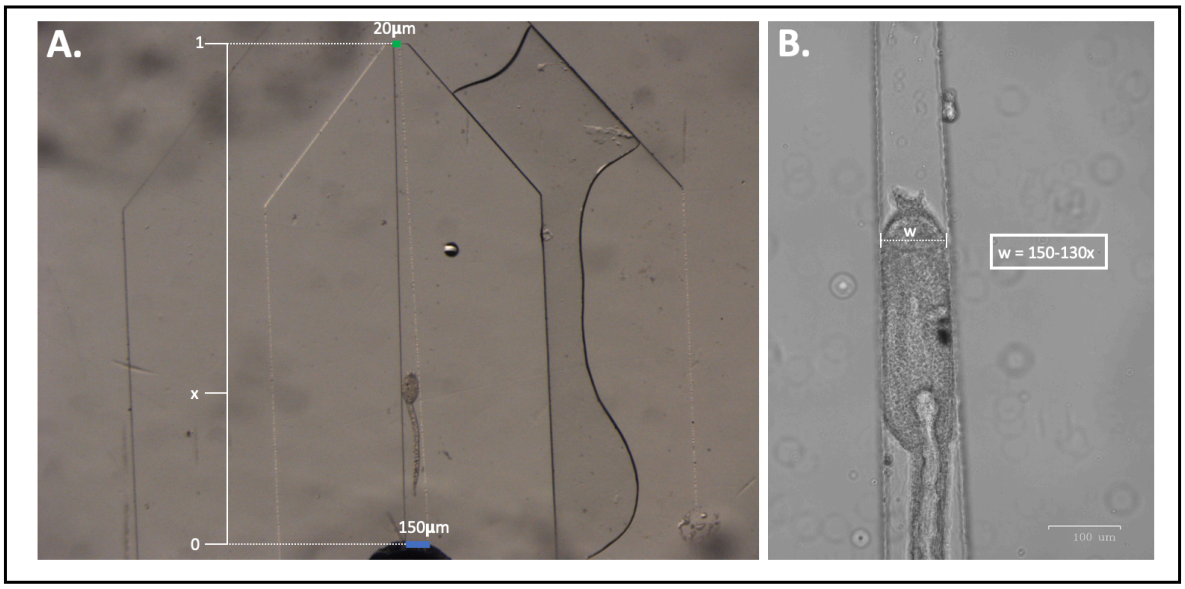


Figure 4.5 Preliminary chip to measure adequate dimensions to trap a Ciona larva. (A) The trapping channel started from a width of 150 µm (blue) to 20 µm (green). If the widest part at the start of the introduction channel is 0 and the narrowest part at the end of the channel is 1. Then the head width (w) can be determined by the equation w= 150-130x if x represent the abscissa of the trapped head along the axis between 0 and 1 (Fig. 4.5). (B) closeup view with scale at 100 µm.

The measurements of 50 animals revealed that the mean head width was about 89 µm with standard deviation of 10 µm, and a 95 % confidence interval bounds at 82.15 to 95.7 µm (See Table 4.S1 in Appendix for full body measurements). There is some variability in size among larvae, but I concluded that a lateral width of 85 µm and a height of 60 µm would be relevant to trap the head of a *Ciona* larva. Unfortunately, the design of Chartier turned out to be unsuitable for *Ciona*. Although the larva was introduced without damage and with relative ease in the chip using a simple P10 micropipette with siliconized tip, problems occurred in the subsequent steps. In fact, the larva was pull back into the introduction channel upon flow pressure and tends to go beyond the trapping zone upon introduction (Fig. 4.6, video 4.S1 in Appendix). This phenomenon happened even when the lowest pressure value possible were used to obtain a laminar flow. In consequence, a completely new design had to be developed for *Ciona*.

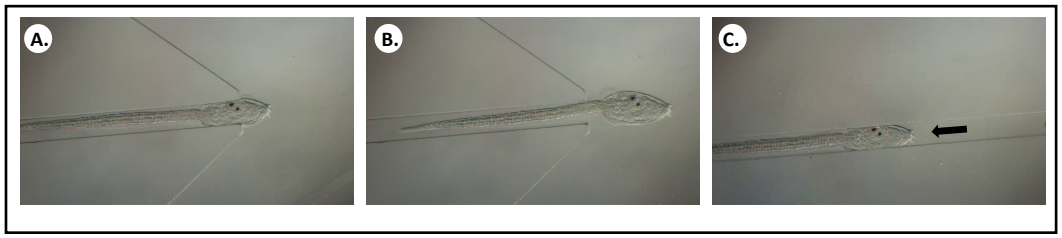


Figure 4.6. The chip design of Chartier for *Platynereis* is not an adequate solution to immobilise *Ciona*. (A) *Ciona* larvae ideally positioned at the anterior end of the trapping channel similar to *Platynereis*. (B) The *Ciona* larva tend to slide off the trap and go beyond the trapping channel over time. (C) When the flow protocol is turned on and laminar flow running, the *Ciona* larva is pushed back into the introduction inlet and is not retained at the end of the trapping channel like in (A). See Video in additional material.

Overall, the solution found in this chapter for *Ciona* came out as a symmetrical microfluidic chip organised around a central trapping channel with a general low-height of 23.5 μm, which holds the animal dorso-ventrally similar to the holding between a slide and a coverslip (Fig. 4.7). The microfluidic chip comprises one trapping channel for the animals, three inlet channels and one outlet channel. The inlet channels are numbered from 1 to 3 and deliver the different test solutions inside the chip. The outflow from the outlet channel was collected in a disposal glass bottle. The introduction inlet is situated in the middle of the trapping channel which has a width of 500 μm and a length of 2 cm. The chamber upstream of the trapping channel has a width of 2 mm, and then splits into two lateral chambers with a width of 1 mm each. The Ciona larva that is trapped inside the chip has its body freely exposed to water streams like natural conditions. The chip allows the introduction and immobilisation of multiple animals at the same time, expanding the number of larvae that can be recorded at once. This is particularly interesting as transgenic larvae can show variation in signal intensity and mosaicisms in the territories expressing GCaMP6m thus the experimenter can select the best larva for recording. This feature avoids the painful process of screening prior to introduction that is necessary if only one larva at a time can be loaded.

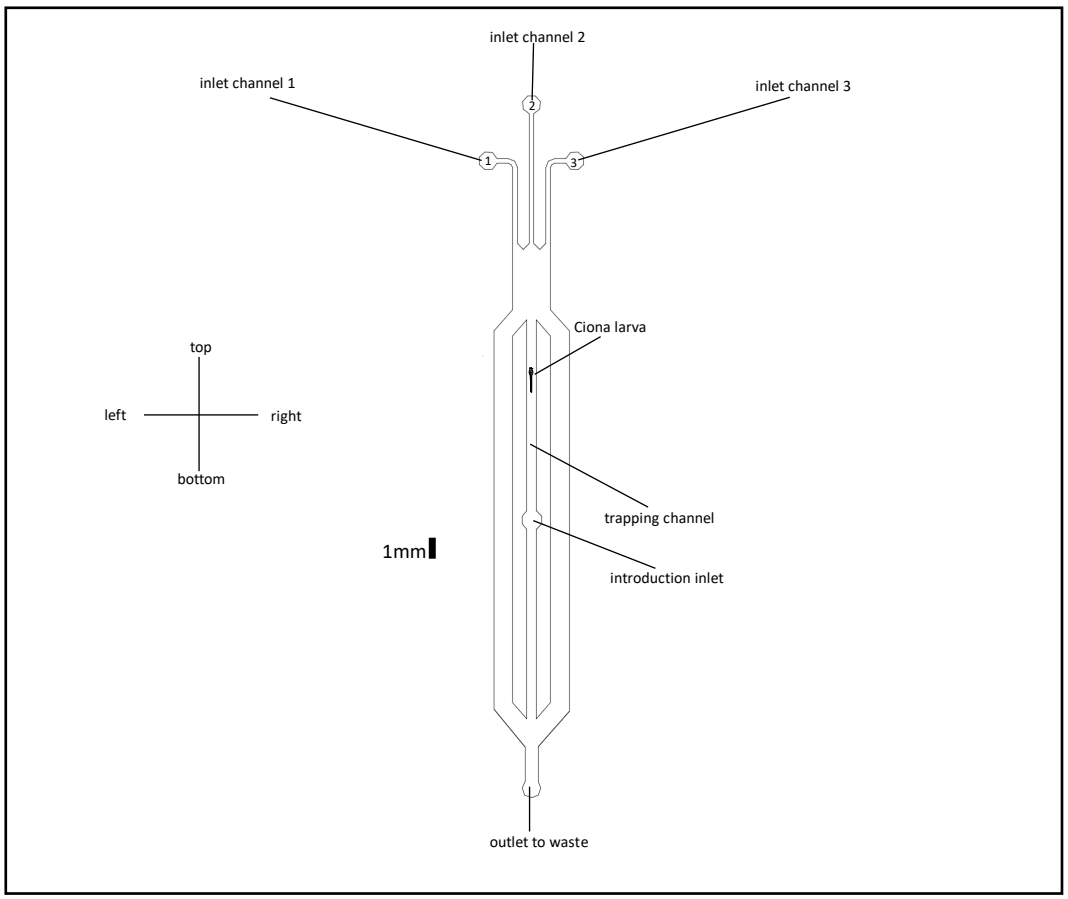


Figure 4.7. Microfluidic chip design used in this study. The Ciona larvae are introduced manually in the trapping channel via the introduction inlet and are immobilised dorso-ventrally in it by the low height (23.5 µm). Scale: 1 mm.

In addition of the immobilisation issue for *Ciona* with the design of Chartier, I noticed that some aspect of the design was easily trapping air bubbles (Fig. 4.8A). This defect in conception is highly problematic as air bubbles represent obstacles interfering with the smooth switch of flow boundaries. Air bubbles are one of the most recurring issues affecting microfluidic and often appear when the chip gets filled with a fluid. They can be very difficult to remove, as was the case here. The use of debubbling/degassing systems was out of question as the chemical cue tested was also a gas. I resolved this issue by altering the design and making the surface triangular instead of flat. This innovative upgrade did not affect the smooth transition of flow boundaries and enabled the air bubbles trapped in the chip to be chased off easily once the flow protocol was running (Fig. 4.8B).

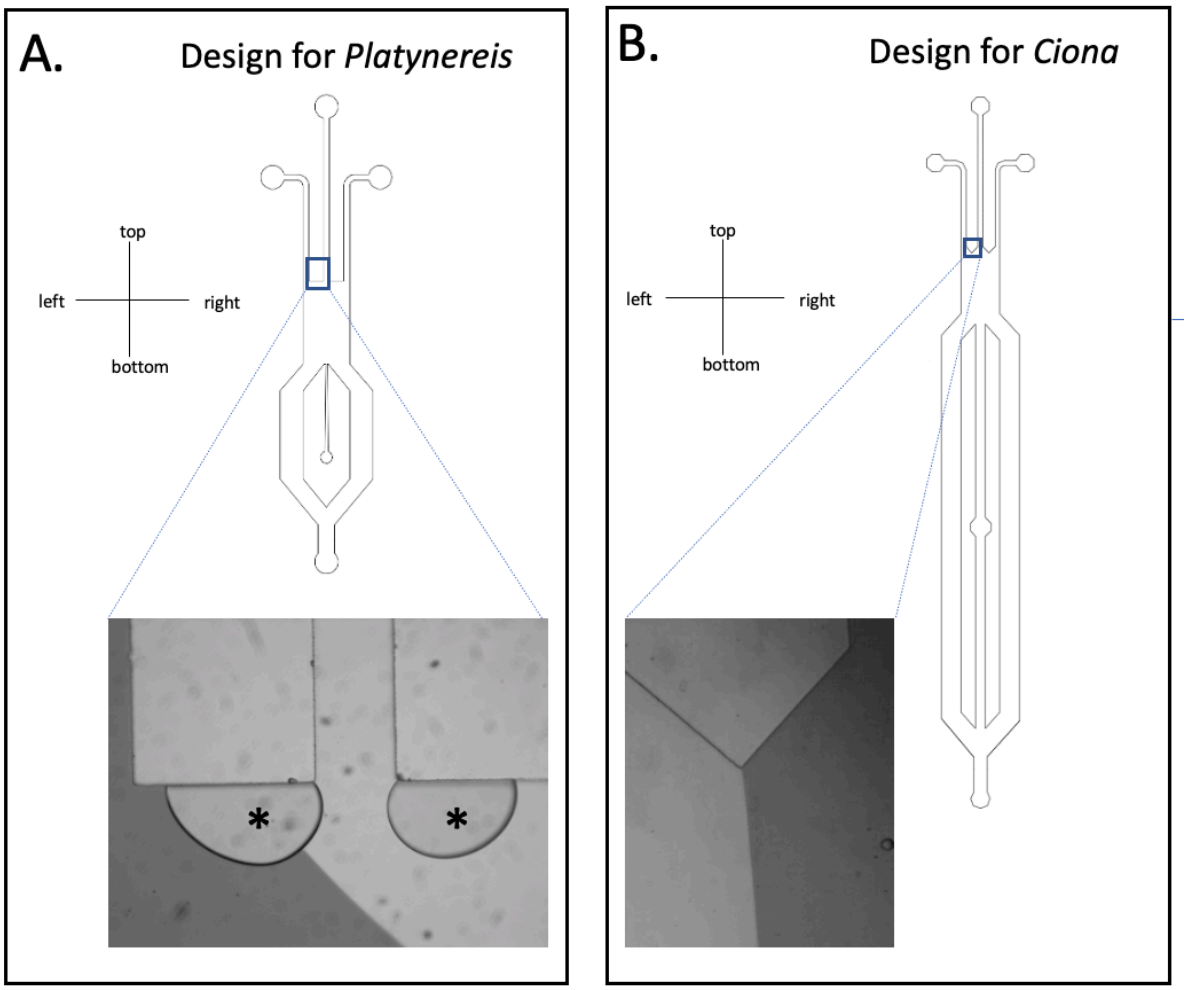


Figure 4.8. (A) The chip design of Chartier *et al.* for *Platynereis* is easily trapping air bubbles (black asterisk) on flat surfaces located before the main chamber which are crucial for the smooth transition of flow boundaries. (B) The chip design used in this study for *Ciona* is chasing away air bubbles as the surfaces are made triangular instead of flat therefore always ensuring a smooth transition of flow boundaries. The flow is vertical from top to bottom on the pictures, and flow boundaries are displaced horizontally. Note the sharp stream boundaries between dyed (dark grey) and non-dyed water (light grey).

4.3.2.2 Experimental flow protocol

A precise flow protocol was elaborated using different pressure rates in each of the three inlet channels. Switches between these pressure rates enable the experimenter to swiftly change the given solution a *Ciona* larva is exposed to when trapped inside the chip. For instance, at resting state (Fig. 4.9B), the chip continuously runs three parallel streams of

sea water in a series of laminar flows. In this resting situation, the animal is exposed to the central stream containing artificial sea water (W) from the inlet channel 2 (Fig. 4.9B). However, the animal can be exposed to the left-side stream (Fig. 4.9C) or the right-side stream on demand (Fig. 4.9D). The left-side stream contains a high CO_2 stimulus solution (S) from the inlet channel 1 while the right-side stream from the inlet channel 3 is a flow and pH control (C), i.e., sea water with a corrected pH value equivalent to the solution with high CO_2. During experiments, the total pressure rates from channel 1, 2 and 3 is kept constant at 200 mbar. For instance, if one pressure rate is reduced or augmented the two others are corrected immediately. The swift movement of stream boundaries without pressure fluctuation is important to avoid mechanical disturbance that arise when the larvae are exposed to a particular stream. When a flow rate is decreased drastically, for instance in channel 1 when it goes from 70 mbar to 0 mbar, some pressure release happens in the corresponding channel since pressure is stored in the tubing and in the PDMS walls of the chip. The slow boundary adjustments between streams due to this pressure release take place on the side far away from the animal as the chip chamber is wide enough. To avoid mixing of the different solutions it was necessary to ensure that the different streams were laminar, meaning they run parallel to each other to guarantee precise transition of stimulus delivery. In fluid mechanics, the Reynolds number is a dimensionless quantity expressing the ratio of inertial forces versus viscous forces. When inertial forces dominate, the flow is turbulent while if viscous forces dominate, the flow is laminar. In resting condition or when the animal is exposed to the left- or right-side stimuli the corresponding Reynolds number were estimated to be around 2 and 3 respectively, which satisfies the conditions for laminar flow.

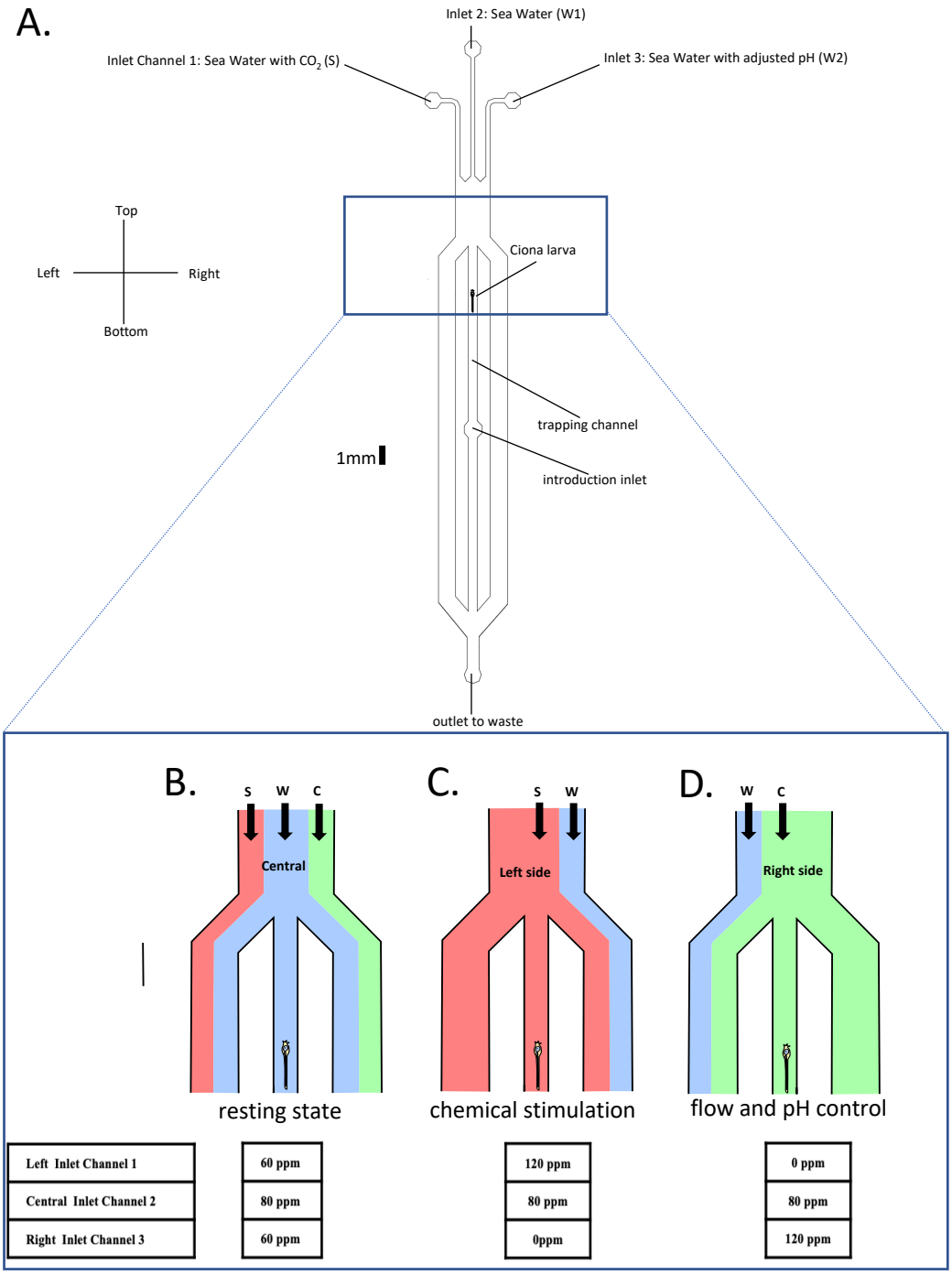


Figure 4.9 (A) Microfluidic chip design for *Ciona*. (B-D) Schematic of the different flow patterns used in the experimental protocol. Flow direction is from top to bottom. (B) In the resting state, the larva is exposed to a continuous flow of artificial sea water (W) while the stimulus (S) and the flow control (C) are flowing on the side streams. (C) During chemical stimulation, the flow boundaries move to hit the larva with the stimulus. (D) A flow control was performed to ensure that the chemical stimulations recorded are not due to change of flow streams or a lower pH. The pressure value for each inlet used to obtain the specific flow patterns is mentioned on the bottom. Scale: 1 mm.

The flow protocol used here started with of an initial resting period of 30 sec, then was followed by three successive cycles made out of 4 steps: (1) a chemical stimulation period of 15s, (2) a resting sate of 30s, (3) a flow and pH control of 15s and (4) a second resting sate of 30s (Fig. 4.10).

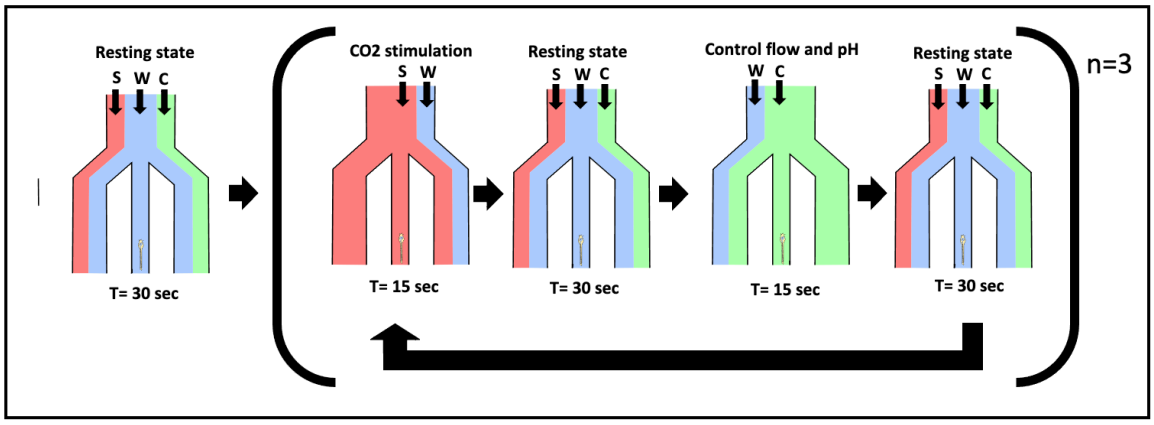


Figure 4.10 (A) Schematic of the different flow patterns used in the experimental protocol. The whole flow protocol lasts for 300 sec (5 min). Flow direction is from top to bottom. In the resting state, the larva is exposed to a continuous flow of artificial sea water (W), the stimulus (S) and the control (C) are flowing on the side streams. During chemical stimulation, the flow boundaries move to hit the larva with the stimulus. A flow control was performed to ensure that the chemical stimulations recorded are not due to change of flow streams or a lower pH. Scale bar: 1 mm.

Before running the automated flow protocol, the chip was prefilled with ASW. Several animals could be used in the same chip successively and/or at the same time. No dye was used when doing the experiments with chemical stimuli. An overview of the whole setup is shown in Fig. 4.11A. The pressure controller MFCS™-EZ (Fluigent) containing three pressure channels was connected to a computer (Fig. 4.11B), and each channel operated one tube filled with a test solution. The MFCS™-EZ was positioned next to the microscope stage, with a constant tubing length of 30cm used to connect tubes and chip (Fig. 4.11A). A customised 3D-printed chip holder was built to maintain the chip and allow its observation under an upright microscope (Fig 4.11C-E).

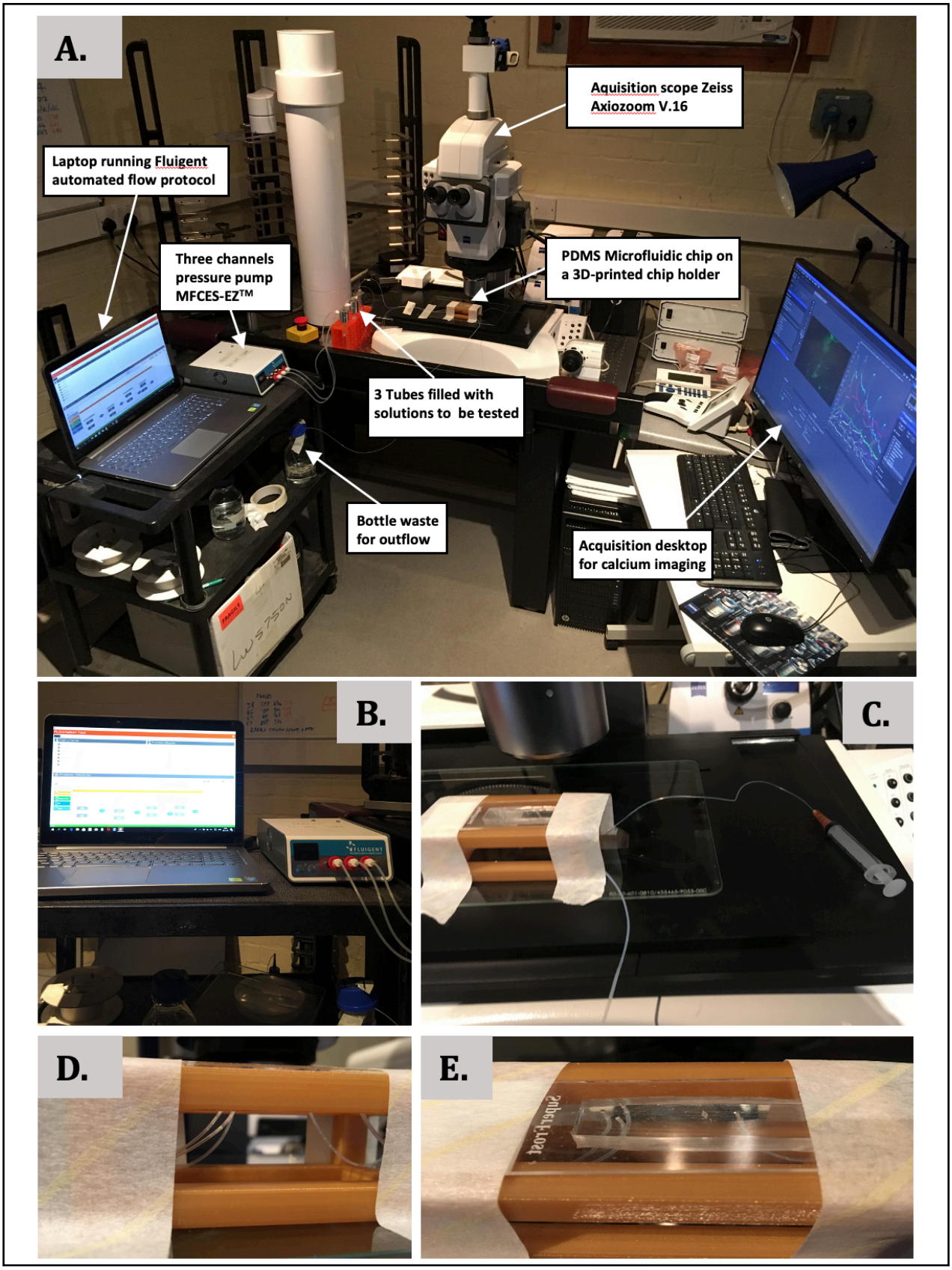


Figure 4.11. (A) Experimental setup overview. The controller MFCES-EZ™ with three pressure channels on the left side of a Fluorescence stereo microscope Zeiss Axio zoom V.16. Each pressure channel operates a tube connected to an inlet of the chip. The bottle on the left bottom for collection of the outflow. (C) The syringe visible on the right side of the microfluidic chip is used for animal introduction. (B-E) closeup view of individual parts. (D-E) shows the customised 3D-printed chip holder (golden colour) holding the PDMS microfluidic chip with tape. (D) The three inlet tubings are visible on the left side while the introduction and outlet tubings are visible on the right side.

4.3.3 Precision of stimulus delivery

When doing such physiological experiments, it is crucial to precisely correlate the period of chemical stimulation with the observed neuronal responses. To do so it is important to know when exactly chemical stimulation occurs. The durations in the stimulus protocol are target values, i.e., the durations in which the three channels operated by the pressure controller (MFCS™-EZ) are switched to a certain state. An experimental quantification of the stimulus timing was performed to assess the actual temporal changes of stimulus concentration experienced by the animals inside the trapping channel.

Temporal quantification of stimulus delivery was achieved by running a series of 10 calibration experiments. Each calibration experiment consisted of the complete flow protocol as shown in Fig. 4.12B, i.e., with three successive cycles presenting in total 6 exposures (2 exposures x 3 cycles) of the stimulus present on each side streams with resting intervals in-between. To visualise the different flow streams boundaries a dye was used during the calibration experiments. The switch from the central stream to a side stream is named “stimulus onset” (a_1, a_2, a_3) while the switch from a side stream to the central stream is named “stimulus offset” (b_1, b_2, b_3). The mean grey pixel values are plotted as graphs and represent variation in stimulus concentration during one calibration experiments normalised between 0 and 1 (Fig. 4.12A).

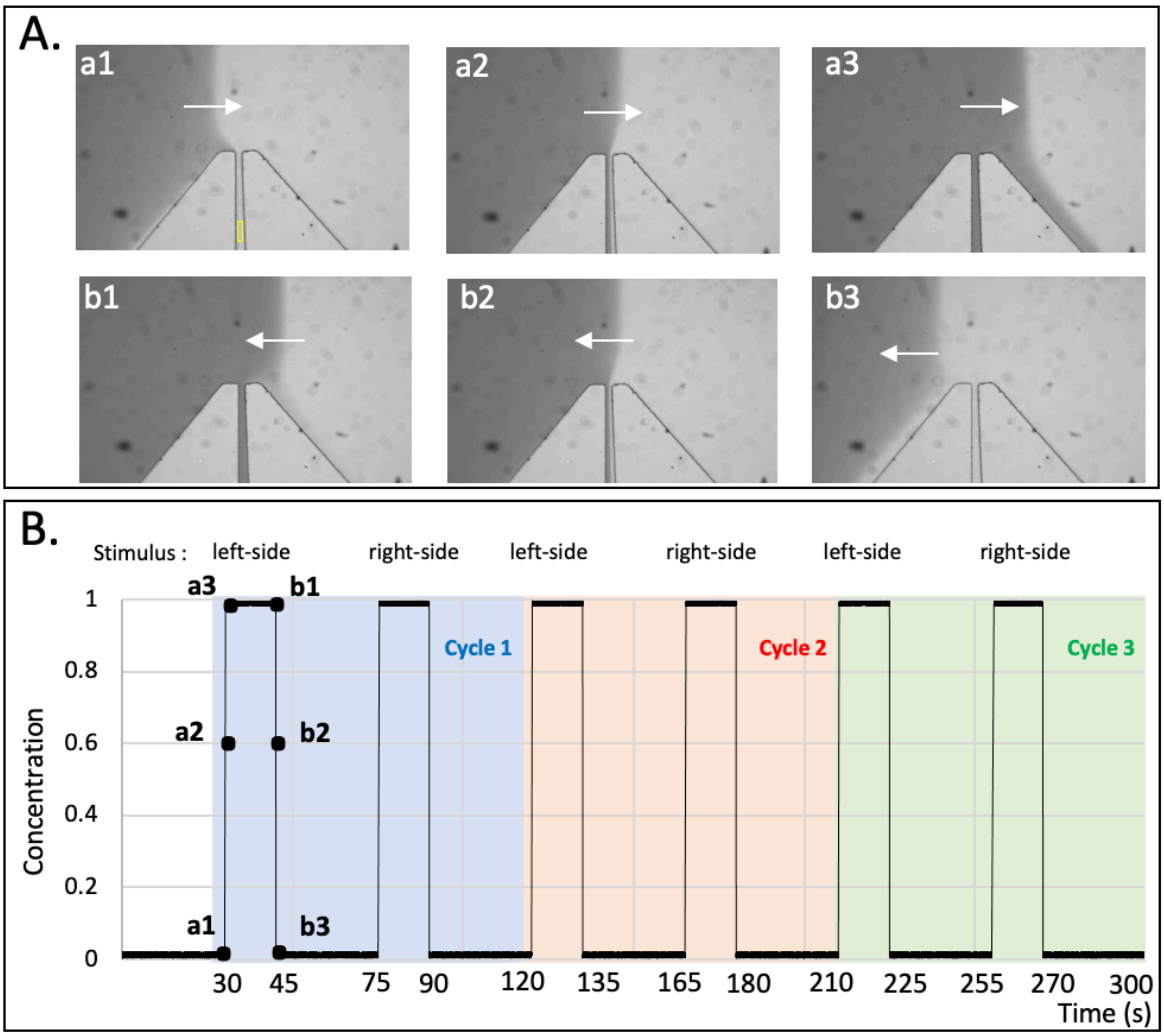


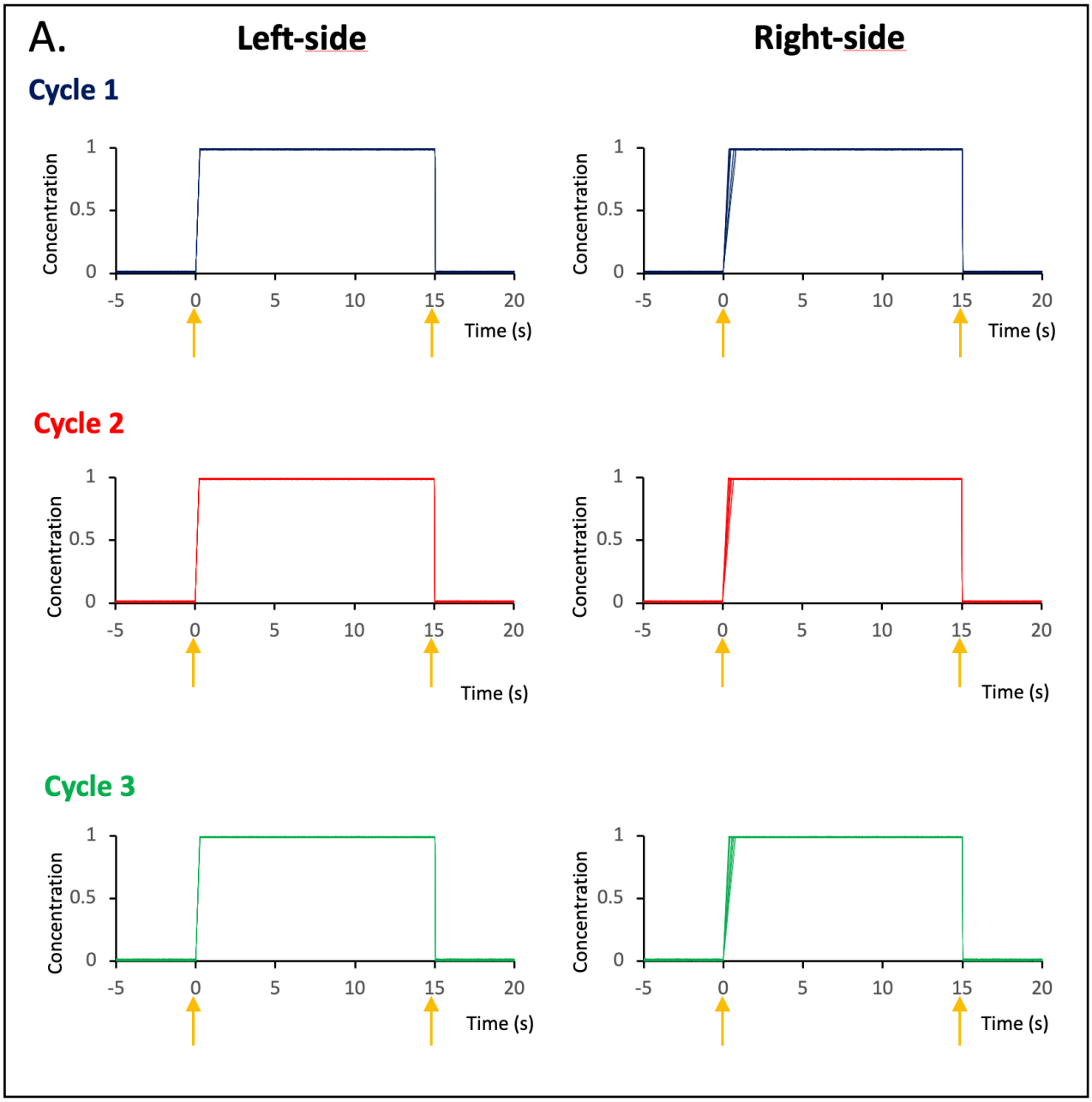
Figure 4.12. Calibration experiment to measure the temporal variations of stimulus concentration. (A) The pictures illustrate the consecutive steps during onset (a_1,a_2,a_3) and offset (b_1,b_2,b_3) of the stimulus located on the left side of the trapping channel. The flow moves from top to bottom and the arrow indicate the horizontal movement of stream boundaries, visualised with a dye. (B) Graph representing the normalised stimulus concentration estimated from mean grey value intensity measured in a region of interest inside the trapping channel (yellow rectangle shown in a1). The graph represents the entire experimental flow protocol used during one calibration experiment. The trapping channel is exposed to either left-side stimulus or right-side stimulus during 15 sec with resting intervals of 30 sec during three repetitive cycles (blue, red, green).

In total, 60 (6 exposures x 10 calibrations) different onset/offset measurements enabled to define the following parameters:

- onset duration = average time needed to go from zero to maximal concentration when the pressure controller switches pressure rates.
- onset duration variability = standard deviation of onset duration
- offset duration = average time needed to go from maximal to 0 concentration when the pressure controller switch pressure rates.
- offset duration variability = standard deviation of offset duration

The 60 different measurements of these parameters have been plotted and superimposed to demonstrate the variability of the stimulus timing with cycles and stimulus sides shown separately (Fig. 4.13A) or together (Fig. 4.13B). It revealed that the stimulus onset takes 390 milliseconds (msec) on average with a duration variability of 197 msec. The offset is almost instant and takes around 27 msec to happen, with no detectable variability in duration during the recordings. The experiment demonstrated the reproducibility of the stimulation onset/offset. It also supports the superiority of pressure-driven systems compared to conventional syringe-pumps systems which were tested in similar way but were much slower, taking several seconds to reach maximal concentration of the stimulus (data not shown). Also, the calibration measurements revealed that there is existence of a slight asymmetry in onset delivery between the two sides. In fact, it became apparent that the onset happens earlier for the left-side than for right-side (Fig. 4.13A). Such difference is unlikely to stem from the chip design as it is symmetrical, it may be due to differences between the pressure channels or tubings used between the left and the right side. Nonetheless, the timing is always reproducible for a

chosen side meaning that they will stay similar during the actual experiments for a given stimulus. In conclusion, as the technical variability for stimulus delivery is minor any recorded variability in the delay of neuronal responses should be accredited to biological causes.



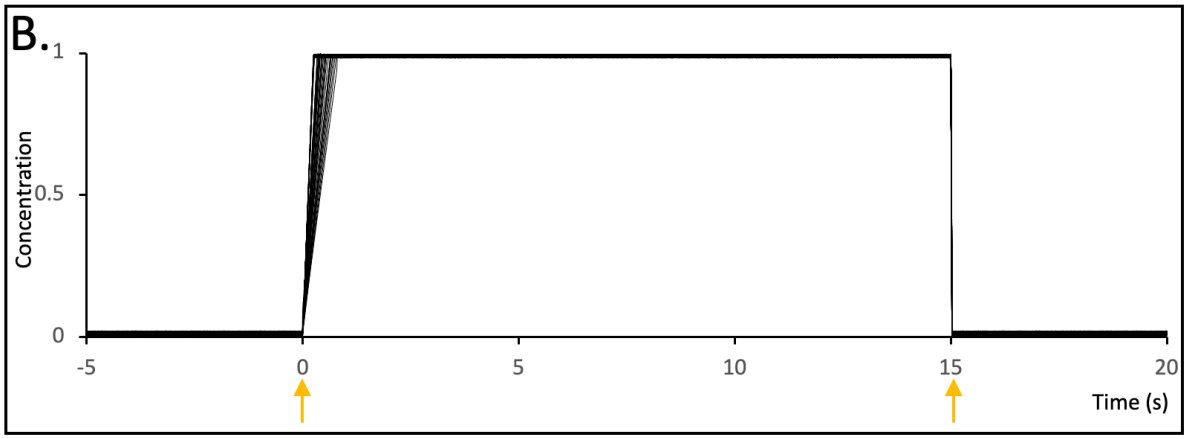


Figure 4.13. (A) Temporal variation of dye concentration for 10 consecutive calibration experiments with cycles and stimulus sides shown separately. The normalised stimulus concentration (from 0 to 1) is estimated from mean grey value intensity. The orange arrow indicates the time point when the pressure controller switches pressure rates starting at time 0 and ending at time 15. (B) Graph representing all 60 measurements of onset/offset with the different cycles and stimulus sides together.

4.3.4 CO_2-evoked responses are detected in Dmrt$^+$-cells of *Ciona* larvae

Calcium imaging with late *Ciona* swimming larvae (20-24 hpf) is possible in this microfluidic setup. In fact, the immobilisation obtained in the microfluidic trap is appropriate for stable recordings over several minutes. There is however some inconvenience as the larvae are oriented in an arbitrary position in the chip. More problematic is that cellular damage sometimes happens upon introduction. Although some tissue damage was apparent on some larvae, it was always possible to observe responsiveness to the microfluidic system (see Fig. 4.S2 in Supplementary data). Overall, the signal quality acquired here does not allow to see many anatomical details such as individual cells or axon processes, but it is high enough to record cellular sensory responses. The calcium imaging experiments presented here constitute the first attempt to study the chemosensory physiology of the model ascidian, *Ciona intestinalis.*

The present microfluidic system has allowed me to test for chemosensory responses to the dissolved gas CO_2 and revealed Dmrt positive ($Dmrt^+$) regions in the swimming larva involved in chemosensation. Due to extreme difficulties in obtaining Ca^{2+} imaging data from *Ciona* larvae inside such microfluidic chip, I was able to take serial fluorescence images from just 7 individuals (Larva 1-7).

Chemosensory responses to CO_2 were confirmed in specific regions of interest (ROI) in and around the brain vesicle area (Fig. 4.14A-G). In fact, groups of $Dmrt^+$ cells near the ocellus and otolith were found to synchronise perfectly with the flow protocol as the detection of the CO_2 stimulus by *Ciona* coincides with a rather strong calcium activity in the ROI selected. The precise identity of these $Dmrt^+$ cells are uncertain from the recordings; however, their relative position suggest they could be epidermal sensory neurons from the peripheral nervous system (Fig. 4.13).

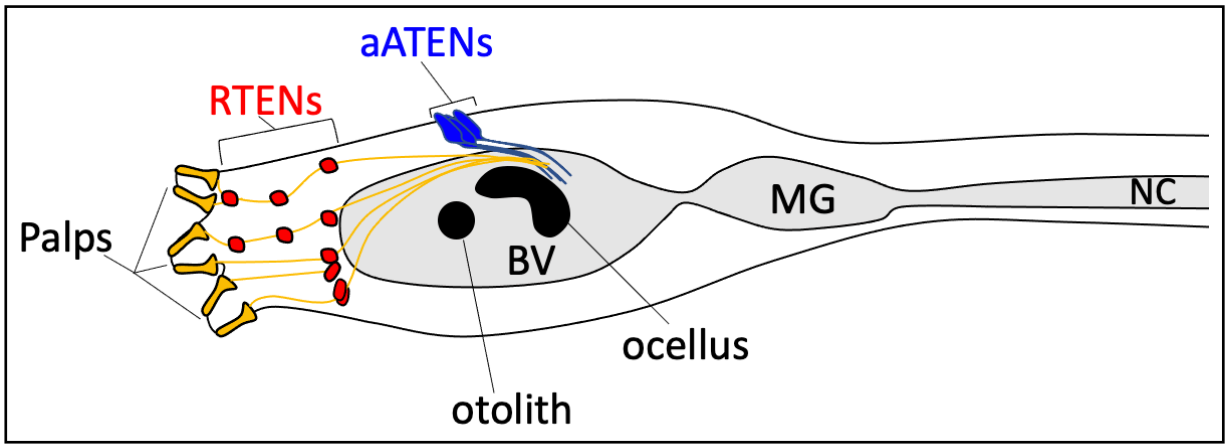
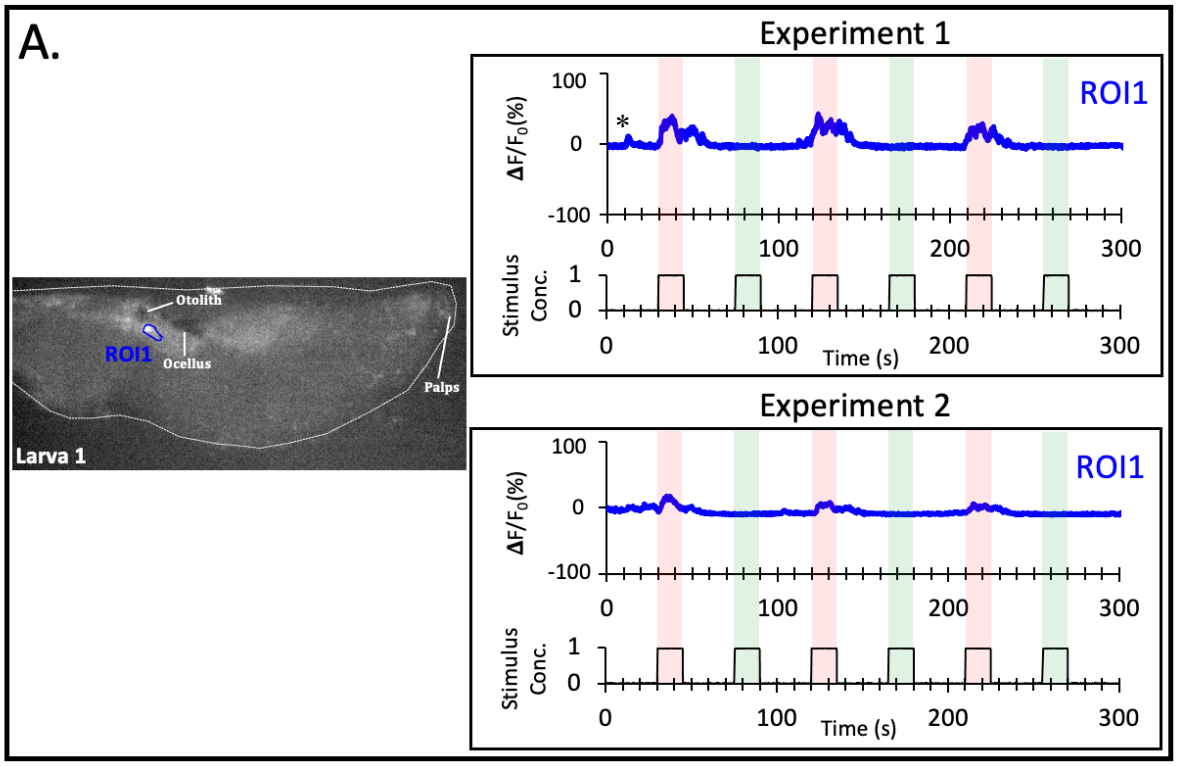
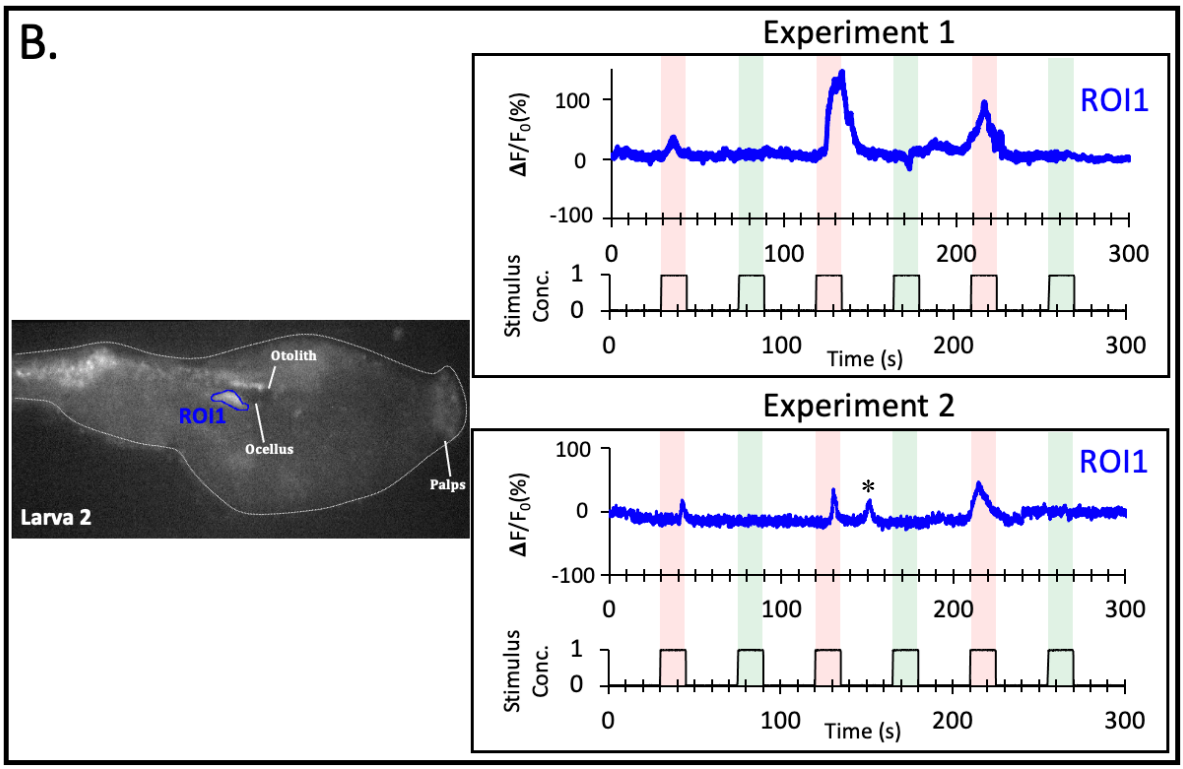


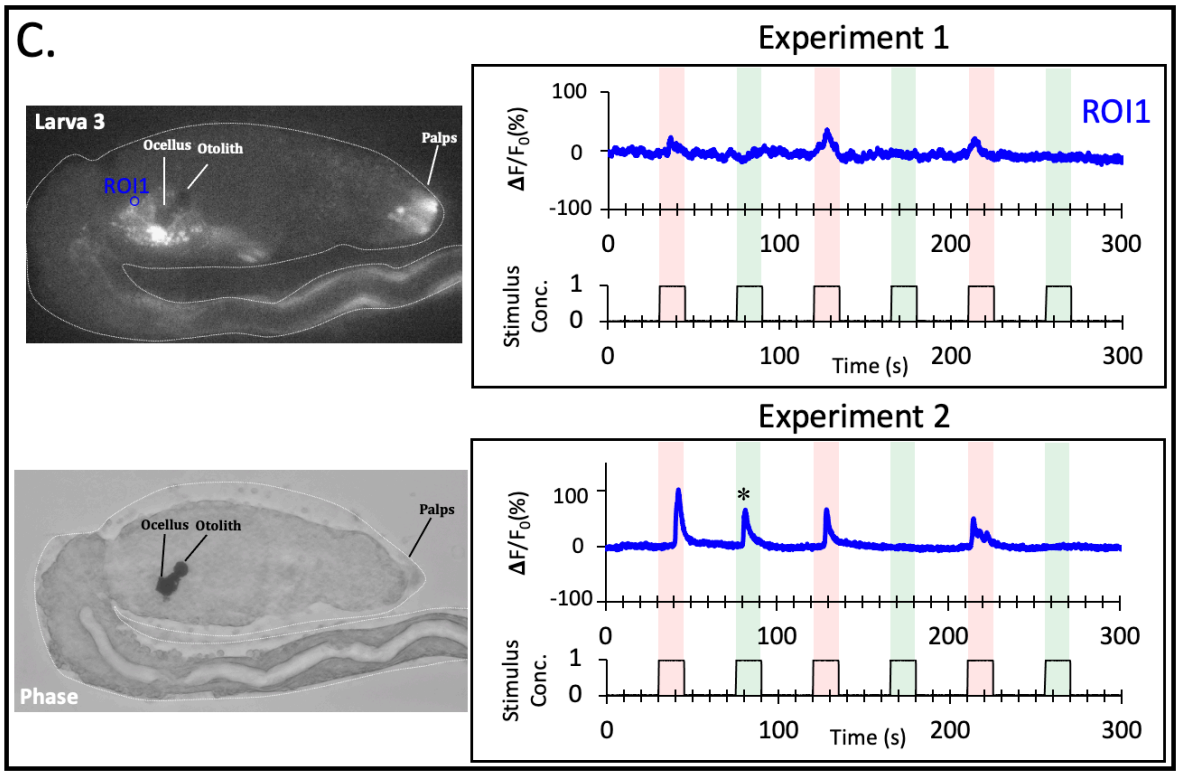
Figure 4.13. Schematic drawing of the central nervous system and Dmrt+ peripheral sensory neurons in *Ciona* larvae. The Dmrt regulatory region express the GCaMP6m calcium sensor in the palps, RTENs and aATENs as well as in the anterior brain vesicle. BV: brain vesicle. MG: motor ganglion. NC: nerve cord.

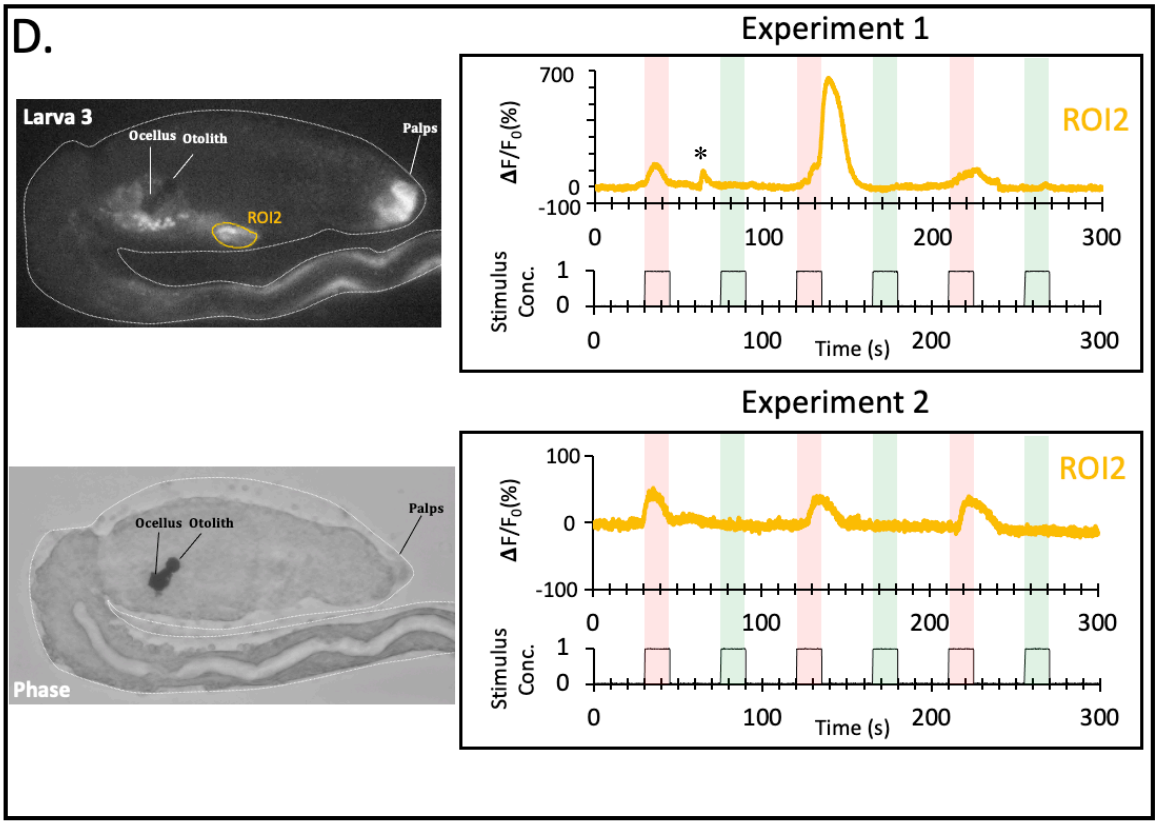
The CO_2-responsive cells (in Fig. 4.14A-G) could be anterior apical trunk epidermal neurons (aATENs), which have been suggested to have dual chemosensory and GnRH activities. These are the only known putative chemosensory cells candidates near the $Dmrt^+$ anterior sensory vesicle. However, it is not clear whether the responding cells correspond to actual chemosensory cells or to their postsynaptic targets such as brain vesicle interneurons integrating chemosensory inputs from the periphery. Also, some responsive cells were seen anterior to the brain vesicle area (Fig. 4.14D), positioned between the palps and brain vesicle which suggest that they might be rostral trunk epidermal neurons (RTENs). A supposed role of RTENs is to serve as a relay for sensory information coming from the palp sensory cells to the brain vesicle, so if the palps contain chemosensory cells as it is likely the case from the presence of putative MS4A chemoreceptors (cf. Chap 3), it is expected that those cells would be activated when transducing chemical information.

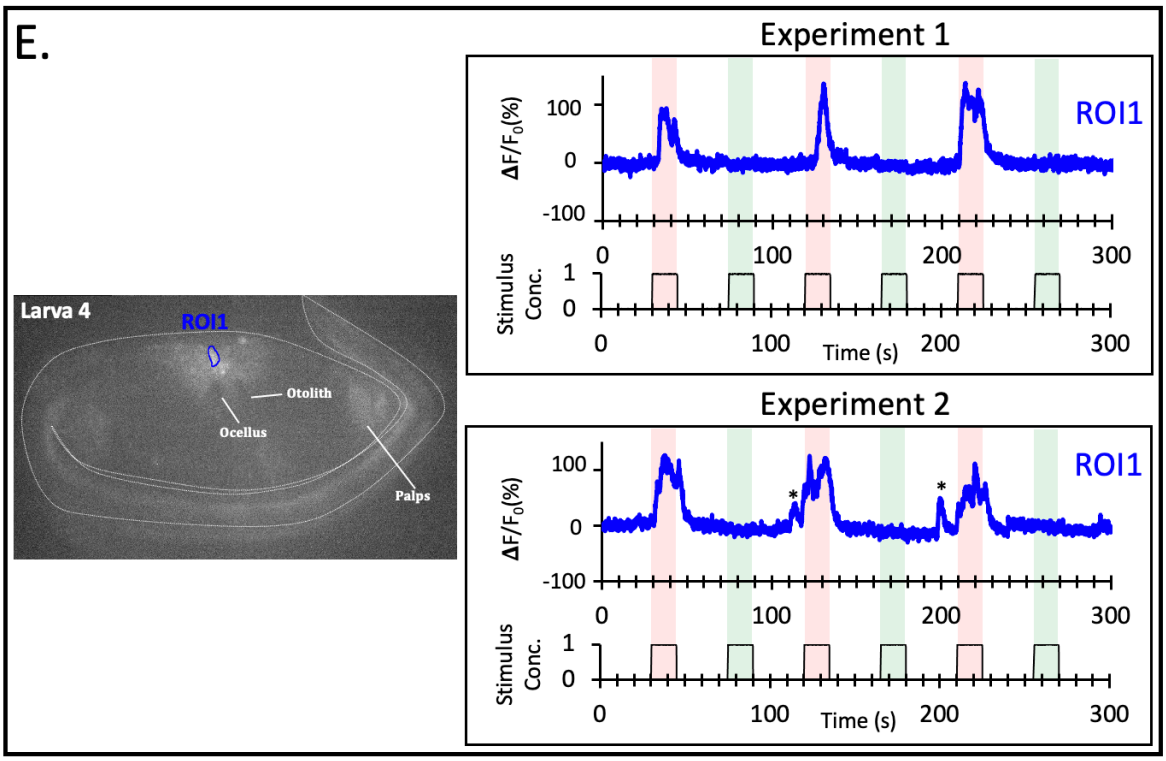
Figure 4.14 A-G. CO_2-evoked responses are registered by calcium imaging in Dmrt+ cells of the brain vesicle and/or epidermal sensory neurons. The fluorescence image shows the coloured region of interests (ROI) used to make the calcium activity traces. The graphs show the temporal patterns of fluorescence intensity for each ROI. The red trace corresponds to exposure to the left side stimulus, i.e., the solution with high CO_2. The green trace is the exposure to the right-side stimulus with the flow and pH control. Asterisks mark random calcium peaks happening outside the zone of chemical exposure (red).

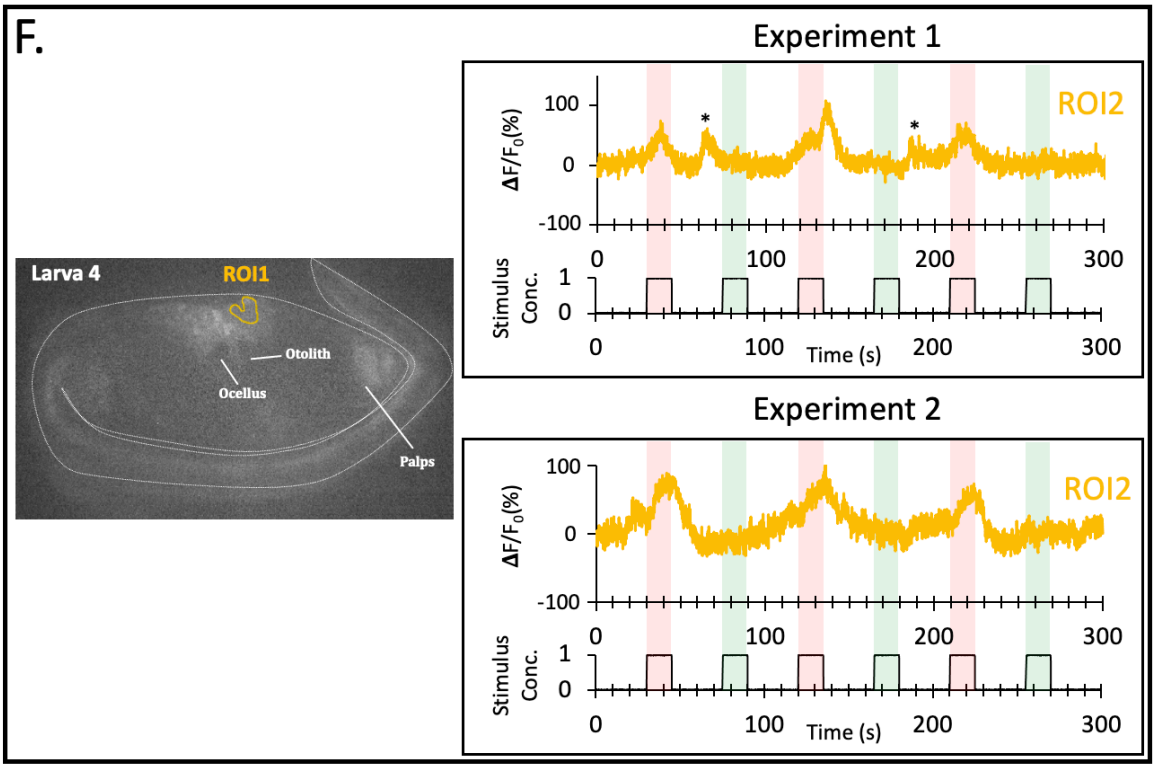

A.
Experiment 1
ROI1
ΔF/F0(%)
100
0
-100
*
0
100
200
300
Stimulus Conc.
1
0
Time (s)
Otolith
ROI1
Ocellus
Palps
Larva 1
Experiment 2
ROI1
ΔF/F0(%)
Stimulus Conc.
Time (s)

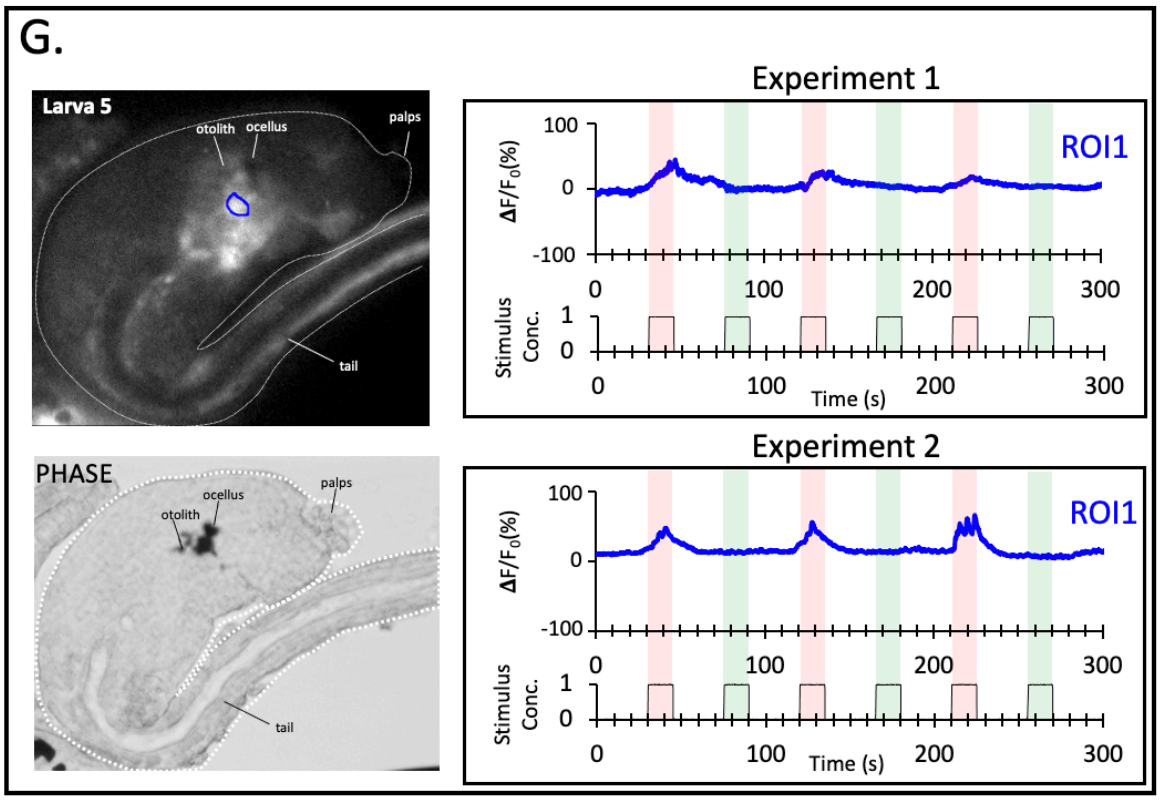

B.
Experiment 1
ROI1
ΔF/F0(%)
100
0
-100
0
100
200
300
Stimulus Conc.
1
0
Time (s)
Otolith
ROI1
Ocellus
Palps
Larva 2
Experiment 2
ROI1
ΔF/F0(%)
*
Stimulus Conc.
Time (s)


C.
Experiment 1
Larva 3
Ocellus
Otolith
Palps
ROI1
ΔF/F0(%)
100
0
-100
ROI1
0
100
200
300
Stimulus Conc.
1
0
Time (s)
Experiment 2
Phase
*


D.
Experiment 1
Larva 3
Ocellus
Otolith
Palps
ROI2
ΔF/F0(%)
700
0
-100
*
Stimulus Conc.
Time (s)
Experiment 2
Phase

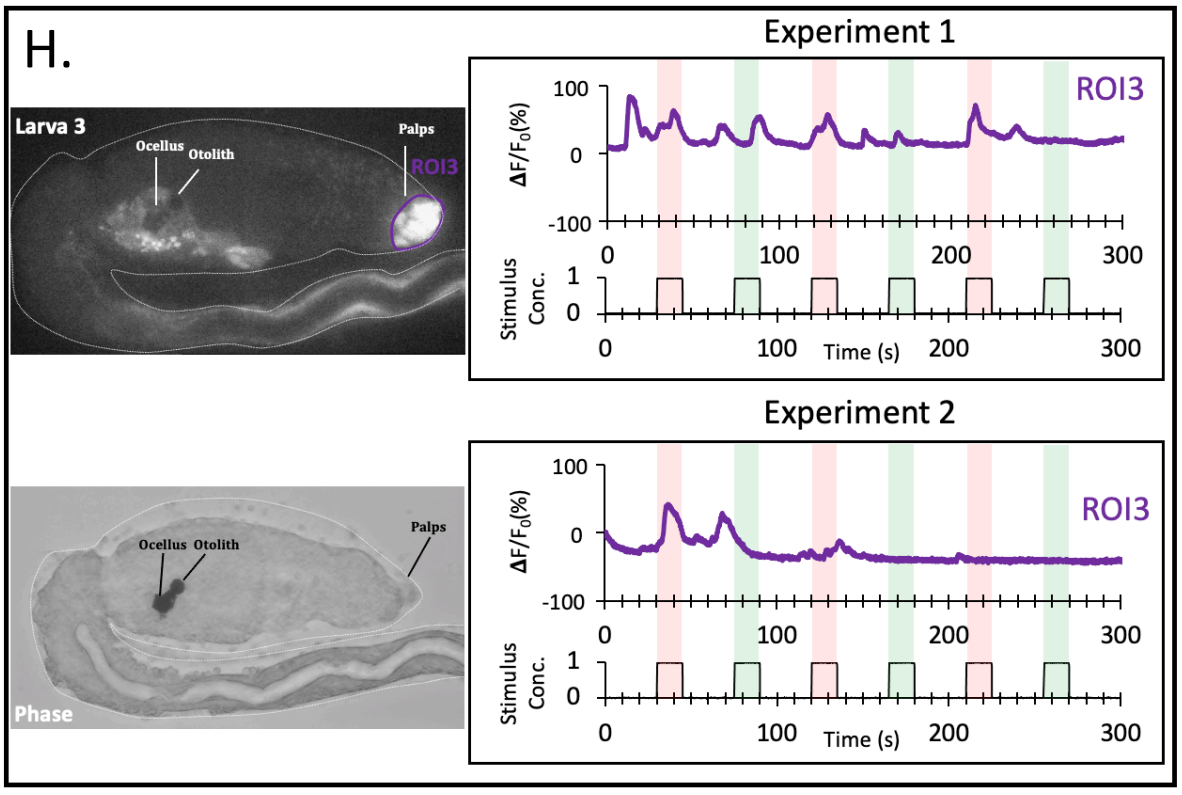

E.
Experiment 1
ROI1
ΔF/F0(%)
100
0
-100
Stimulus Conc.
1
0
0
100
200
300
Time (s)
Larva 4
ROI1
Otolith
Ocellus
Palps
Experiment 2
ROI1
*
*


F.
Experiment 1
ROI2
ΔF/F0(%)
100
0
-100
Stimulus Conc.
1
0
0
100
200
300
Time (s)
*
*
Larva 4
ROI1
Otolith
Ocellus
Palps
Experiment 2
ROI2



When a region of interest was drawn in the palp area (Fig 4.14H-I), there are calcium spikes registered when the solution with increase CO_2 concentration hits the larva but also frequently when the flow and pH control is turned on. These mixed results imply that the palp could sense sudden change in the water flow patterns and/or pH variation. If Ciona detect variation in water flow through the palps, then it strongly supports previous claims that this organ contains mechanosensory cells as it responds to variation in sea water flow inside the microfluidic chip. There is recent convincing experimental proof that the palps contain mechanosensory cells that initiate the metamorphosis process through a two phase Ca^{2+} transient pattern after mechanical touch (Wakai et al., 2021). The *Dmrt* regulatory region was expressing the fluorescent calcium sensor GCaMP6m in all the three palp cell-types. As said earlier, it is likely that the ACCs, PSNs and CCs have different functions for different sensory modalities such as mechanosensory and/or chemosensory inputs. Therefore, the tight juxtaposition of putative mechanoreceptor

cells and chemoreceptor cells in the palps makes it a challenging organ to discriminate the different type of sensory responses with the current regulatory region used for transgenesis. It is also possible that some of the palp cells may have dual mechanosensory and olfactory functions, thus resembling the polymodal ASH neurons in *C. elegans* which sense a variety of stimuli and mediate avoidance to high osmotic, mechanical, and chemical stimuli (Kaplam & Horvitz, 1993). If this is the case, then it would be extremely challenging to distinguish these different sensory modalities, and the current microfluidic chip and testing strategy would be unworkable as the negative control (without chemical stimulus) would also trigger a physiological response.



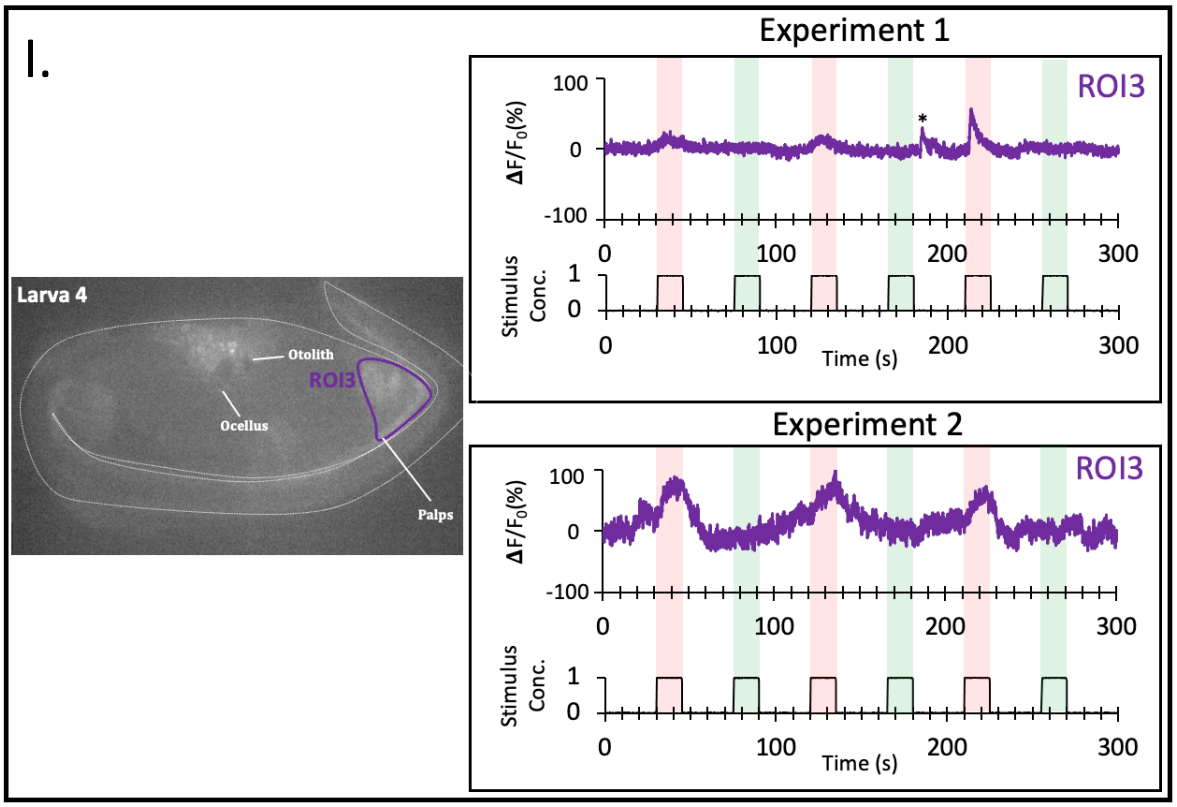


Figure 4.14H-I. Multiple calcium spikes registered in the palps. The fluorescence image shows the region of interests (ROI) used to make the calcium activity traces. The graphs show the temporal patterns of fluorescence intensity for each ROI. The red trace corresponds to exposure to the left side stimulus, i.e., the solution with high CO_2. The green trace is the exposure to the right-side stimulus, i.e., the flow and pH control.

It is expected that ascidian larvae integrate chemosensory inputs to the motor ganglion to control tail muscle contractions and locomotion. In fact, the larvae respond to different physical stimuli such as light/shadows and gravity by altering their swimming through the activity of their muscles (Zega et al., 2006; McHenry and Strother, 2003, Tsuda et al., 2003). The *Ciona* transgenic larva 5 had expression of the calcium reporter outside the expected Dmrt territories in the motor ganglion, which was easily distinguished by its unique form. I therefore checked if Ca^{2+} transients occurred at the same time or independently between the ROI in the brain vesicle responding to CO_2 stimulation and the neuronal response observed in the motor ganglion. Although some calcium spikes were registered during period of chemical stimulation, there was no clear associative relationship that could be observed from those records. Also, it is known that the motor

ganglion has periodic spikes of Ca^{2+} transients even when not subject to chemical stimulation (Okawa et al., 2020).

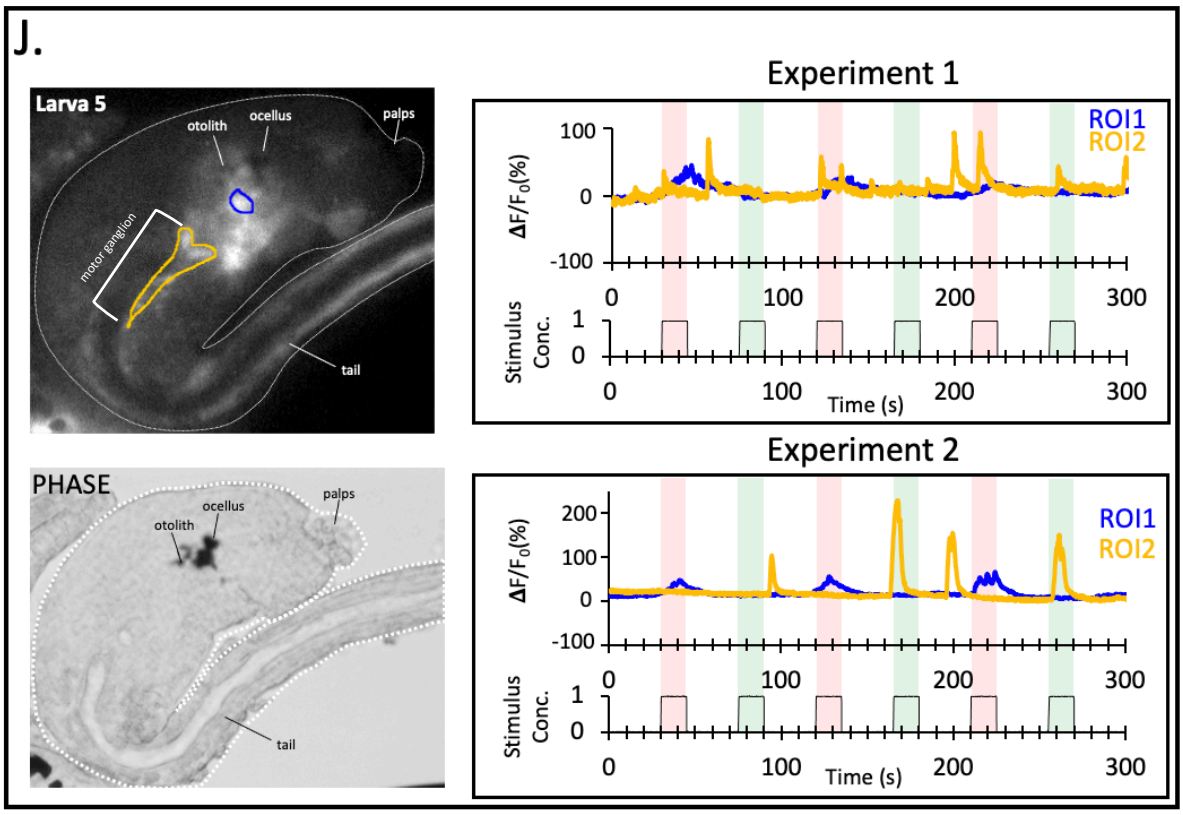
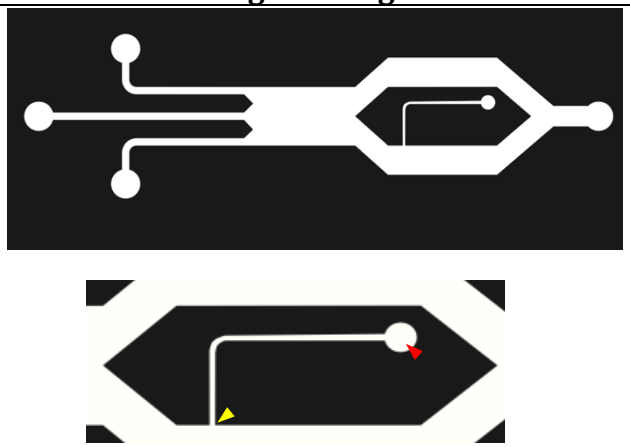
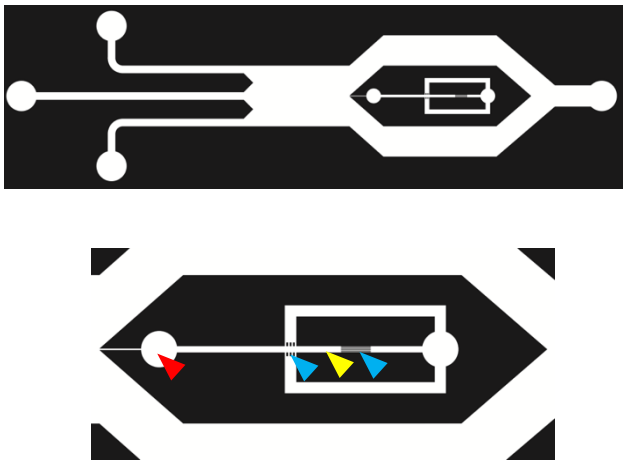
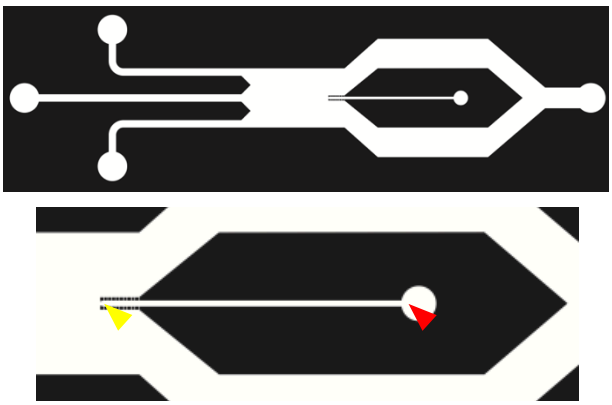


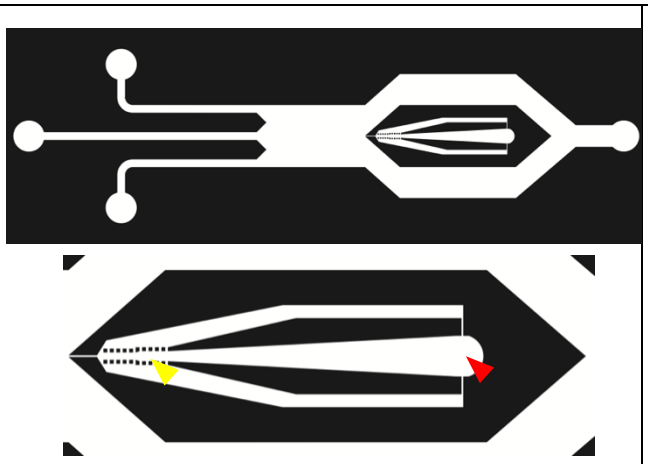
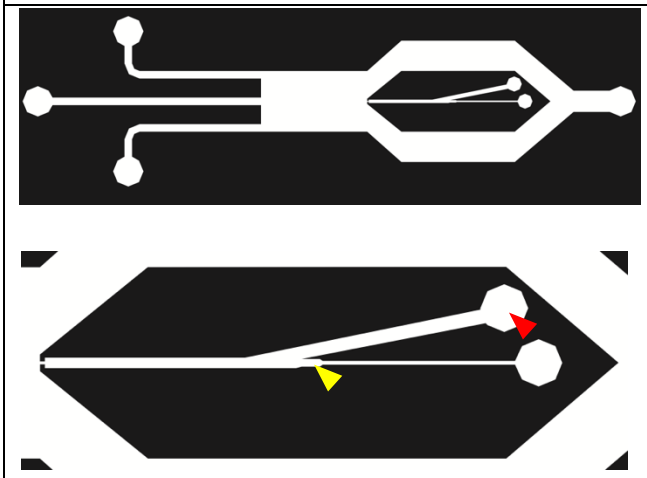
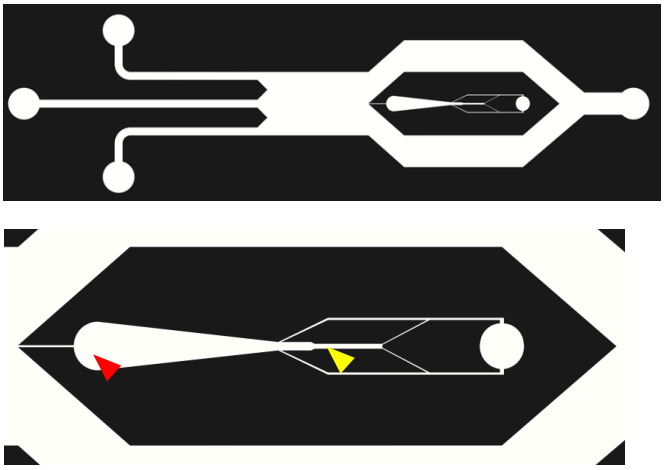
Figure 4.14J. No clear associative relationship observed between GCaMP6m fluorescence in the motor ganglion and in Dmrt-responsive cells of the brain vesicle. The graphs show the temporal patterns of fluorescence intensity for each ROI, with motor ganglion in orange and Dmrt-responsive cells of the brain vesicle in blue. The red trace corresponds to exposure to the left side stimulus, i.e., the solution with high CO_2. The green trace is the exposure to the right-side stimulus, i.e., the flow and pH control.

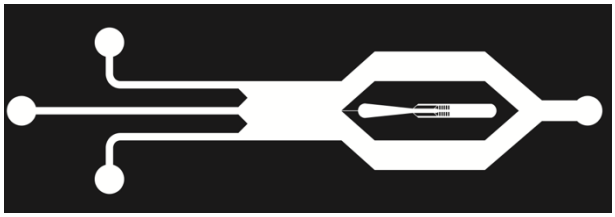
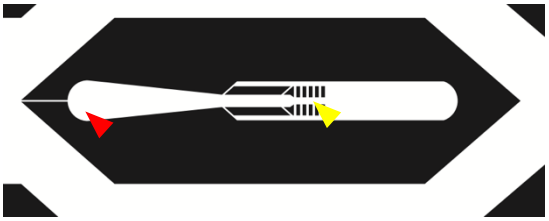
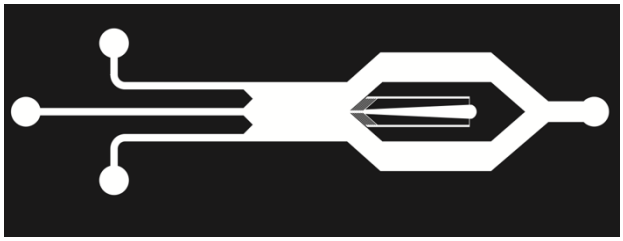
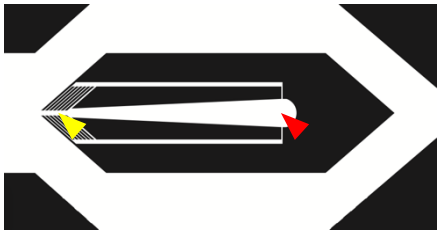
In conclusion, these results indicate that some Dmrt+ cells corresponding to either epidermal sensory neurons and/or components of the brain vesicle can detect and/or integrate chemosensory information such as high concentration of CO_2 dissolved in seawater. The simultaneous activation of cells located at different sites suggest the presence of a neural circuit connecting Dmrt-expressing cells at different locations. Although not yet optimal, these observations validate the possibility of immobilisation strategy with a microfluidic device and show that calcium imaging experiments are feasible in late Ciona larvae with such a setup.

4.3.5 Supplementary data

Figure 4.S1. Sample of the unsuccessful design strategies tried to immobilise *Ciona* larvae.

Unsuccessful design strategies	Comments
	Upon introduction in the introduction inlet (red arrowhead), the larva is placed and immobilised perpendicular (yellow arrowhead) to the direction of the laminar flow. However, this design could not retain the larva well as they slide off the trapping zone.
	In this design, the introduction inlet was at the top (red arrowhead) and the larva was placed tail first in the introduction channel and was pushed to the back by the laminar flow and immobilised below (yellow arrowhead). This PDMS chip had issues resolving the pillar structure (blue arrow) only separated by 5-10 µM, but did show some interesting prospects if only the pillar structures could be better resolved by soft lithography.
	Upon introduction (red arrowhead) the larva was pushed till the end into pillar structures (yellow arrowhead). This chip was very successful at immobilising the larvae although it had problems upon switch of the flow boundaries at the larvae moved into the right or left side. Also, these pillar structures were hard to resolve by soft lithography and tended to damage the larva.

	Upon introduction (red arrowhead) the larva was pushed till the end into pillar structures (yellow arrowhead). This chip was successful at immobilising the larvae and corrected the problem of movement upon the switch of the flow boundaries. Nonetheless, the damage to the larva was too recurrent and the pillar structures were even harder to resolve by soft lithography, so the design was dropped.
	The larva was introduced in a channel on the side (red arrowhead) and immobilised with the laminar flow forces in a bottleneck trap (yellow arrowhead). However, the larva was pushed back into the introduction channel.
	The larva was placed in the introduction inlet at the top (red arrowhead) and was placed tail first in the introduction channel which has a bottleneck shape at the end specific to the larva, with a narrowing for the tail that was supposed to keep the larva in place (yellow arrowhead). Although, the larva was well positioned, the larva was deformed when pushed back by the laminar flow.

 	The introduction inlet was at the top of the design (red arrowhead) and the larva pushed back below to the trapping zone (yellow arrowhead). This design showed partial success, the larva was immobilised but also damaged as the pillar structures were not smooth enough.
 	In this design, upon introduction (red arrowhead), the larva was pushed till the pillar structures (yellow arrowhead). This design did not resolve well by soft lithography, so the structures were too edgy and damaged the larva.

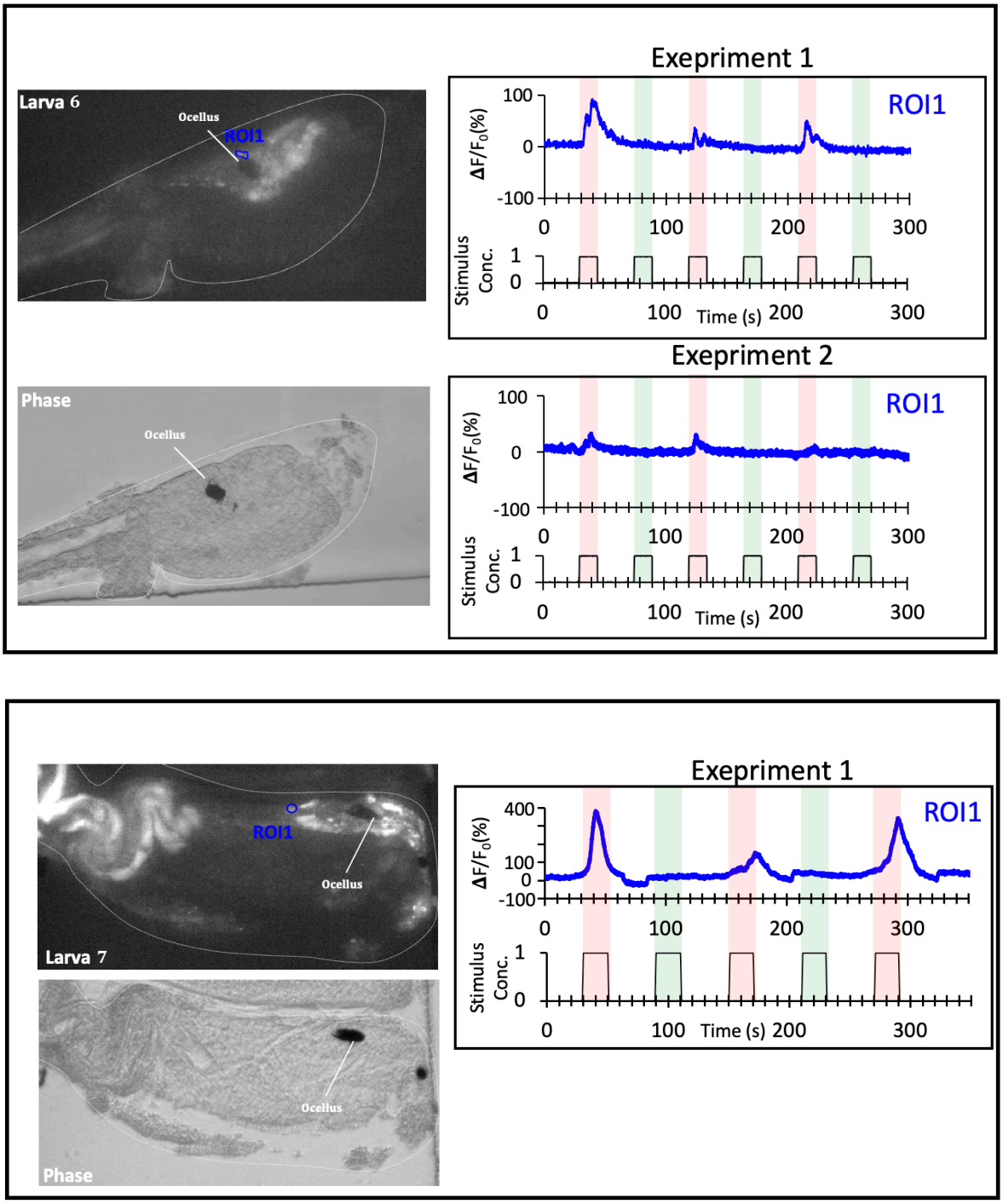


Figure 4.S2. Larva 6 and 7 sustained obvious tissue damage upon introduction in the chip but showed nonetheless obvious responsiveness to the microfluidic system. The graphs show the temporal patterns of fluorescence intensity for each ROI. The red trace corresponds to exposure to the left side stimulus, i.e., the solution with high CO_2. The green trace is the exposure to the right-side stimulus, i.e., the flow and pH control.

4.4 Discussion

4.4.1 Immobilisation

This microfluidic research is the first step towards a new technology for studying *Ciona intestinalis* larvae. The final output provides evidence for the initial hypothesis of CO_2 chemosensory detection in *Ciona* larvae. However, this newly developed technology has room for future improvements, especially to ameliorate ease of practice upon immobilisation of the larva. A new solution for immobilisation and a technological upgrade is suggested in more detail in the General Discussion (see Chapter 5). Overall, it must be stated that *Ciona* swimming larvae are extremely challenging to work with, not because of their small size but because of their inability to remain easily stuck in a microfluidic device submitted to a laminar flow. This experiment has had many more technical issues to troubleshoot than initially anticipated. One main problem is the high deformation capability of *Ciona* larvae under physical stress. They can go through an opening as small as 5 μM if pushed with sufficient pressure. The main issue faced during the elaboration of my microfluidic chip designs was that *Ciona* larvae were not retaining themselves in the trapping channel as the backwards force from the laminar flows drove them away from the trap end extremity. This is particularly problematic in a system only capable of driving positive pressure as only a mechanical constraint could retain the larvae. The issue was most likely avoided in *Platynereis* larva due to the presence of appendages such as the chaeta and anterior dorsal tentacular cirrus which took an important role in the retention of the larvae in the trapping channel and increased resistance to the incoming flow. The fact that *Ciona* larva is just in a streamlined rocket shape with no appendages is possibly the reason why retention in the chip design of Chartier was unsuccessful. Also, the presence of an outer tunic of cellulose is a major

difference with the *Platynereis* outer surface and this also made them different in their retention abilities inside a PDMS chip. Numerous designs with different immobilisation strategies were created and tested. A recapitulative sheet of all those explored designs are in the supplementary data with comments for the reasons they were turned down or if partial success were obtained (Fig. 4.S1). The rate of progress to find a suitable design for recordings was slow cause it relied on trial-errors with successive iterations. Finally, the reliance on mechanical constraint did work when the height was lowered to an extreme of 23.5 μM but this came with some problems. In fact, the low-height strategy used in this study compressed the larvae to the point that cellular damage on the embryos were likely to happen upon introduction and positioning (Fig. 4.15). Especially, the palp cells, one of the most interesting cell candidates for chemosensation (cf. Chap. 3) are highly susceptible to breakage as they are on the anterior end of the animal.

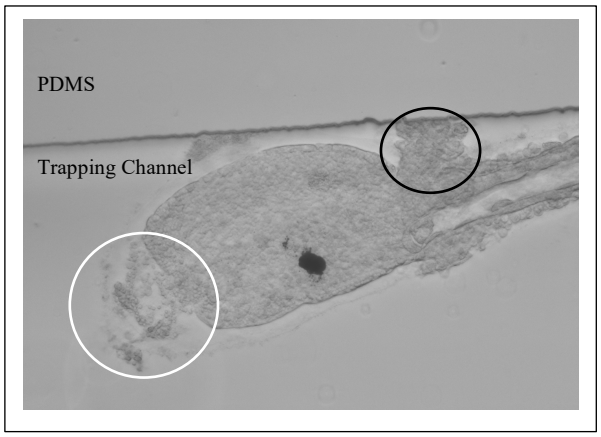


Figure 4.15. Due to the low height used in the chip, the appearance of damages in the form of leakage (black circle) are likely to happen in the chip if the Ciona larva is pushed too brutally in the trapping channel. Especially, the anterior end (white circle) containing the palps was the most fragile part of the larva. Note that this larva represents an extreme case of the possible damages.

This limitation made me unable to test accurately individual palp cells as initially intended. This design however worked well to measure chemosensory responses in territories close or within the brain vesicle. The risk of embryonic damage upon introduction of an invertebrate marine larva in a PDMS microfluidic chip was already raised by Thomas Chartier, which succeeded to test 6 dpf *Platynereis* larvae but could not apply their methodology for 3 dpf larvae which are much more fragile and required more care as damage in the form of leakage often appeared. The occurrence of embryonic damage is not ideal as physical stresses are likely to interfere with the recordings. Ideally, only a limited and temporary stress that impaired neither their health nor their aptitude to move and develop should be used. A good chip design should allow to release trapped larvae and assess for potential body damage introduced after trapping. My personal observation is that heights ≥60 μM were not damaging *Ciona* larvae and that after their release from a chip of that height in sea water the larvae would swim back normally and metamorphose. They would even start metamorphosis inside the PDMS chip if left long enough. This shows that it remains possible to introduce a *Ciona* larva inside a microfluidic without damage and deleterious interference with the larval condition if the height is appropriate.

Other options for immobilisation were prospected such as restriction with low-melting point agarose. Although agarose is an efficient way to immobilise *Ciona*, its use was precluded in a microfluidic device setup receiving repeated chemical stimuli. Microfluidic chips enabling the delivery of chemical stimuli, and which use an agar-embedding approach do exist. This type of immobilisation has been efficiently used for zebrafish (*Danio rerio*) larvae but requires manual removing of the jellified agar around the larva's

face to expose the olfactory organs (Candelier et al., 2015). This would be very difficult to achieve in *Ciona* larvae due to their much smaller size (~900 µm in total length as opposed to ~4 mm for zebrafish larvae). Besides, a gelatinous medium like agar is unnatural and it is certainly important for the animal's chemosensory organs to be exposed to chemicals freely as they would in a natural environment. For these reasons, the microfluidic trap used here relied solely on a mechanical method for immobilisation which even though not yet perfect in its current configuration is already a decent compromise between the necessity of immobilisation for stable recordings and the ecological relevance as the chemicals flow freely around the sensory organs, similar to what is expected in the sea. The use of methods involving chemicals such as the anaesthetic MS-222, which efficiently blocks muscle contraction and prevent swimming of *Ciona* larvae, were considered but abandoned as they might affect general cell physiology (Matsunobu and Sasakura, 2015). Also swimming movements are not problematic here as the larvae tend to stay still when inside a tight PDMS channel.

4.4.2 Stimulus delivery

The main advantage of the current setup is that chemical stimuli can be repeatability presented to the animal and this with a precise concentration. The possibility to test the reproducibility of individual cellular responses is fundamental to make any sort of physiological statements. The repetition of a given physiological response with the use of fast stimulus onsets/offset allows to separate genuine chemosensory responses from spontaneous activation of neurons. In this study, only one type of chemical cue (CO_2) was tested but, the chip design enables to test any type of water-soluble chemicals in a controlled fashion. For instance, chemical cues such as sugars, amino acids, proteins,

metabolic products, etc. are known to constitute much of the language of life in the sea (Hay, 2009). In particular, it would be very interesting to test exopolymers secreted by marine bacteria of the genus *Pseudomonas* that are known to increase the extent of larval settlement (Szewzyk et al., 1991). It is possible that *Ciona* larvae sense some of these excreted bacterial chemicals to find a good place for their lifelong settlement. It is also likely that some metabolic products excreted from adult Ciona come into play for larval population recruitment. In fact, there is evidence showing that larval ascidians preferably settle close to adult populations (Petersen and Svane, 1995; Bingham and Young, 1991). Although the current system is not flawless, the possibility to do functional imaging experiments is a great step forward and could soon enable to test a whole new repertoire of stimuli which have been so far out of reach experimentally.

4.4.3 Image quality

GCaMP6m expression was obtained through plasmid electroporation in *Ciona* zygotes, thus providing expression throughout embryogenesis when the chosen regulatory region is activated by TFs. This represents a net advantage upon other techniques such as microinjection which are difficult to perform and have generally lower level of expression due to mRNA degradation. Indeed, there is always a certain latency between mRNA injection into a zygote and the actual moment fluorescence recording are performed in a more developed animal. In this study, the upstream expression of GCaMP6m mRNA by the *Dmrt* regulatory region during development yielded a signal high enough to observe cellular activity. Nonetheless, the transgenic *Ciona* larvae showed variability in signal intensity and some mosaicism in expression territories which are inevitable occurrences of the transgenesis process. Mosaicism can be problematic as they may affect the given

frequency a certain response is observed in a certain cell/tissue or organ. These possible differences between embryos necessitate to do a screening of the transgenic animals before their introduction in the microfluidic chip. This extra step is time-consuming, especially if only one animal at a time can be introduced in the device. This problem was mitigated with the chip design used here as multiple animals could be introduced at once, which enabled directly to screen for the best larvae for recording in the chip. Ideally, what should be done in *Ciona* is the creation of permanent and stable adult transgenic lines containing in their genomes, the DNA for a calcium sensor such as GCaMP6m with either a ubiquitous or pan-neuronal promoter. This would greatly facilitate the process as the time needed to perform the experiments would be reduced. Indeed, the larvae from such transgenic lines could be used directly as the transgenesis protocol and screening step due to mosaicism would be suppressed.

An alternative could be to use chemical calcium dyes. I did not try to perform calcium imaging with these as they are thought to be less sensitive than genetically-encoded sensors and do not allow precise cellular targeting. Also, it was unclear whether they would pass through the cellulose tunic of late larva and penetrate all cell types, though this could be tested, and it could be interesting to examine whether variability in signal intensity between animals and the mosaicism effects (as well as the transgenesis protocol) could be avoided with such non-genetic dyes. Overall, however, the Dmrt>GCaMP6m signal was good enough to observe sensory activity with the current imaging settings. Responding cell position was identified based on location within the head and with anatomical landmarks such as the ocellus and otolith. This is enough to claim that a cell belongs to embryonic regions such as the palps or the brain vesicle, but some uncertainty remains about what are precisely the type of cells responding in those

regions. For example, in the palps, it is not possible to clearly differentiate between the three different palp cell types despite their distinct features. Further improvements are needed to identify individual cells as currently the borders between the different cells are ambiguous in the calcium recordings. Imaging of better quality could be achieved using a confocal microscope as the entire head depth could be recorded with the use of multiple focal plans. This might also reveal finer neuronal details such as axons and enable conclusions to be drawn on individual cell physiology and neuronal networks. What could also be useful to identify individual cells is to co-electroporate a plasmid coding for a cell nuclei fluorescent marker alongside the GCaMP plasmid. Such a marker could be the general chromatin marker "H2B-mcherry" that has been used previously in *Ciona* (Wagner and Levine, 2012). In fact, it is even possible to include such a marker in the same plasmid used in this study which already contains mCherry in addition of GCAMp6m both under the control of Dmrt regulatory region. The splitting into two independent proteins is done by the presence of a ribosomal skipping sequence P2A between mCherry and GCAMp6m. Lastly, it is important to state that even though high-quality images may be desired for presentations, they are not essential to make valid experimental conclusions. In fact, a cellular response can be quantified with only a low number of pixels if the signal-to-noise ratio is sufficient and if the ROI stays unchanged throughout the recording process. In consequence, any calcium activity traces that clearly display the presence of a sensory cellular response has to be considered of high enough quality, regardless of aesthetic standards.

4.4.4 Ecological relevance of CO_2 as a chemical cue for *Ciona*

What could be the main reason for *Ciona* larva to detect variation in external CO_2 levels? The main question would be to know if CO_2 acts as a repulsion or attraction cue. CO_2 is a complex chemical cue as different species have developed distinct behavioural and physiological responses to it. In fact, CO_2 can be a negative or positive valence depending on context, previous experience, and life stage of the animal (Banerjee and Hallem, 2020). In vertebrates, sudden environmental increase in CO_2 is mainly known to inform the animal for the proximity of a potential predator. For instance, in zebrafish larvae the GnRH3 neurons of the terminal nerve sense CO_2, and this detection triggers an escape motor response (Koide et al., 2018). Similarly, in rodents, CO_2 increase induced by predator respiration is detected by necklace OSNs and causes an innate avoidance behaviour even at near-atmospheric concentrations (Hu et al., 2007). Hence, we can assume that *Ciona* larva could also avoid CO_2-rich environment to escape predators or to find emptier settlement sites with less competition for resources. But the reverse is not impossible and maybe CO_2 acts as an attraction cue instead. For example, in the larvae of the rhinoceros beetle (*Trypoxylus dichotomus*), the CO_2 emitted by conspecifics act as an attractive source and leads to larval aggregation (Kojima, 2015). *Ciona* are gregarious ascidians living in close groups often attaching to one another. The larvae could perhaps be attracted to dense populations of kin through cues like CO_2, which would hint to a favourable site for survival. Behavioural assays could resolve this "attraction vs rejection" question for CO_2 in *Ciona* without the need to study the detailed physiological responses as it was attempted here. Nonetheless, the data presented here support a view where the ability to detect external CO_2 is a shared trait of Olfactores. In addition, *Ciona* may detect other dissolved gas in seawater such as O_2. For instance, nematodes such a *C. elegans* rely

on CO_2 and O_2 level detection to stay in a preferred environment (Carrillo et al., 2013). The mechanism of O_2 detection in insects and nematodes relies on soluble guanylate cyclases which are also present in the palp ACCs according to SC-RNAseq data (Gray et al., 2004; Vermehren et al., 20006; Sharma et al., 2019; Johnson et al., 2020).

Chapter 5 General Discussion

The work presented in the preceding chapters explored different unknowns in our understanding of the evolutionary origin of the olfactory neurons in the Olfactores clade. The developing lamprey was the first point of investigation in the search for the origin and specification mechanism of hypothalamic GnRH neurons. Then, the exploration of the genomic evolution, regulation, function, and expression of *GnRH* and *MS4A* offered two complementary channels for comparing putative homologies of vertebrate olfactory sensory and secretory neurons to *Ciona* palp cells. Finally, a first attempt was made to assess chemosensory responses in the nervous system of *Ciona* larva. Here, I will discuss how these data come together to advance our understanding of olfactory system evolution and development. I will be re-examining the main topics and provide additional information discussing how it complements or changes existing notions.

Work Declaration

The discussion and review in this Chapter were done by me, Guillaume Poncelet.

5.1 Conclusions and prospects

In the different chapters of this thesis, I explored the origin, specification, regulation, and functionality of emerging olfactory neurons in the Olfactores clade and used their conserved developmental signature to track the evolution of vertebrate olfactory sensory and secretory cells.

Chapter 2 explored the origin of hypothalamic GnRH neurons in lamprey through molecular, cell lineage tracing and drug test methodologies. The study revealed some key

element in the brook lamprey: (1) GnRH neurons appear *in situ* within the hypothalamus according to GnRH expression. (2) Cell lineage tracing data confirmed that GnRH neuron do not derive from NHP progenitors. (3) FGF and RA signalling influence GnRH-III neuron specification. (4) NELF likely co-localises with GnRH-II neurons, while ISLA is possibly co-localised with GnRH-I and -III neurons. Based on this collection of data, I proposed a model to explain the evolutionary origin of those cells in the Olfactores lineage (cf. Chapter 2). Key features of the model open questions: especially on the supposed heterotopic shifts in tissue origin as *Ciona* and jawed vertebrates (except some GnRH neuron populations in teleosts) all seem to have a 'placodal' origin for their anterior most GnRH neurons, but not lamprey. This brainteaser should be clarified further in future studies by fully evaluating GnRH cell developmental mechanisms and cell lineage in lamprey. Here, the study mainly focused on the expression of specific marker gene by *in situ* hybridisation which already provided plenty of information, but other techniques such as knockdown by morpholinos or CRISPR/Cas9 or even additional drug treatments that can be used in lampreys could be used to investigate the functional role and regulation of these gene in better details. Such additional techniques will certainly provide invaluable information to study the lamprey GnRH system from a functional perspective and help understand if there is an ancestral role in GnRH development specifically. For instance, two TFs appear particularly interesting in that *LpISlA* and *LpNELF* genes are likely co-localising with lamprey GnRH neurons. If *LpISlA* and *LpNELF* genuinely colocalise with GnRH neurons, which could be tested by double fluorescent *in situ* hybridisation or single cell sequencing, then it would offer a means to follow GnRH progenitors from earlier time in development, i.e., prior to the onset of GnRH expression. More advanced cell lineage tracing methodology such as lineage reconstruction with live cell tracking using confocal imaging could resolve this issue. For instance, the regulatory

regions of NELF and ISLA could be tagged to GFP and incorporated in transgenic lamprey zygotes then, these fluorescent cells could be followed up with automated live tracking algorithms such as in the Imaris software (Oxford instruments).

To know, whether the lack of a nasal placode origin is a specificity of lamprey, it would be interesting to test whether the pattern of expression of GnRH-related genes is similar between lamprey and hagfish embryos. Hagfishes are the other extant members of the cyclostome lineage and diverged from lampreys 470-390 million years ago (Kuraku and Kuratani, 2006). Previous chromatographical and immunocytochemical studies confirmed the presence of a GnRH system in the hypothalamus of two adult hagfish species: *Eptatretus stouti* (Braun et al. 1995) and *Myxine glutinosa* (Sower et al., 1995). Despite these findings, the exact hagfish GnRH gene(s) and primary structure(s) remain unresolved. Using the recently released genome of *Eptaterus burgeri*, I report the presence of a unique GnRH gene in that primitive vertebrate. The *E. burgeri GnRH* gene structure is of two exons and one intron and its decapeptide sequence is QHWSRKWQPG. In comparison to lampreys which have three different GnRH genes, the presence of a single gene in hagfish is therefore surprising but may facilitate the study of these neuroendocrine cells in that model system and comparison to jawed vertebrates like mice, which also possess a single hypothalamic GnRH gene. Cell lineage tracing methods in hagfish as used in this study could also be useful to confirm that hypothalamic GnRH neurons do not arise from the nasal placode, although given the extreme scarcity of hagfish embryos this will prove logistically challenging. In fact, if all cyclostomes show this pattern, then it would reinforce the view that the hypothalamic GnRH cell of the vertebrate common ancestor did not have a placodal origin. Taken together, there is still a wide area to be exploited by studying the GnRH cells in protochordates (tunicates and

amphioxus) and in lampreys to have a much better understanding of the putative ancestral state of the vertebrate hypothalamic GnRH system and/or ancient olfactory/GnRH developmental mechanisms.

Chapter 3 is a genomics contribution to the identification and functional aspects of the chemosensory system of ascidian. The study of the expression, regulation, and putative function of some *GnRH and Ms4A* genes in the palps informed on the evolution of olfactory neurons and lays the ground for further examination through experimental approaches. First and foremost, it provided a robust connection between ACCs and olfactory neurons, bringing existing data into a new perspective. In particular, the data highlight that the ACCs represent a unique opportunity to understand how segregation of function from a multifunctional GnRH/chemosensory precursor into distinct specialised cell types may have happened during evolution. At the regulatory level, functional segregation involves expression loss of cell type-specific transcription factors. The set of transcription factors that control cell type-specific gene expression are known as terminal selectors and cooperatively interact to form a CoRC (Hobert, 2016). The forced expression of one or a few of these terminal selectors is usually sufficient to change cell type identity. Such terminal selectors are identified in *Ciona* for instance, the bipolar tail neurons (BTNs) which have been recently suggested to be homologous to cranial sensory ganglia cell types depends on the TF Hmx for their cellular identity (Papadogiannis et al., 2020). Furthermore, the misexpression of terminal selectors like FoxC turns those BTNs into ACCs while its absence in the palps turn ACCs into aATENs (Horie et al., 2018). The identification of terminal selectors forming the CoRC of key olfactory effector genes such as GnRH and Ms4a, may have the potential to explain how this functional segregation occurred. In theory, the evolution of a new cell type involves the evolution of a new CoRC.

Deciphering in the details how these CoRCs in *Ciona* have evolved to lead to distinct cell types in vertebrates would be a challenging task but could reveal key molecular divergence in regulation responsible for the emergence of new sister cell types. My work on the GnRH1 regulatory region, expression data and knowledge-based interaction pathways suggest that a CoRC is conserved for GnRH cell expression across Olfactores. Indeed, it supports a preserved role of ATTA-binding homeobox genes such as *Pbx*, *Otx2, Nk4, Six3/6* and *Msx1/2* in the regulation of GnRH expression (Dardaillon et al., 2020). In vertebrates, the orthologues of those TFs bind to conserved TFBS on the GnRH regulatory region as molecular complexes (Clark and Mellon, 1995). A more rapid way to validate that such regulatory conservation exists could be to test if the *Ciona* regulatory regions of Ms4a and GnRH induce expression in the necklace OSNs and GnRH neurons of vertebrates, respectively. This type of cross-species activity approach has been undertaken for the regulatory region of ßγ-crystallin, a lens cell effector gene present in *Ciona* like in the ACCs. The latter was shown to be expressed in the lens of transgenic Xenopus showing that the ACCs likely have a preserved CoRC with placode-derived vertebrate lens cells even though *Ciona* predates the evolution of this structure (Shimeld et al., 2005). The close association between molecular markers of lens cell (ßγ-crystallin) and olfactory neurons (Ms4a and GnRH) in the ACCs is not that surprising if we account for the proximity of lens and olfactory placode territories in vertebrates. Interestingly, connection between olfactory, GnRH and the visual system exist in vertebrates such as in teleosts (Takashi et al., 2016), so we could assume that such connectivity between cell types might have been facilitated by the existence of an ancestral olfactory/lens precursor.

Chapter 3 also open the door for future research on *MS4A* genes in ascidian. First, these chemoreceptors should be tested for their ability to produce chemosensory responses in neurons *in vivo* or *in vitro*. Testing options that could easily be put in place exist, for instance specific interactions between odours and mouse Ms4a proteins was tested by heterologous expression of individual Ms4a proteins together with the genetically encoded fluorescent calcium indicator GCaMP6s in human embryonic kidney cells (HEK293) (Greer et al., 2016). This approach could be used to test specific cues recognized by *Ciona* MS4A for instance to know indirectly which chemical cues is triggering response in the palp ACCs. The method downside is that it does not enable to study the upstream circuitry happening *in vivo* inside the nervous system of *Ciona* as with the approach developed in Chapter 4. The main advantage is the fast lane it could provide to identify the chemical cues responsible for settlement which could quickly be applied for new control strategies to fight biofouling.

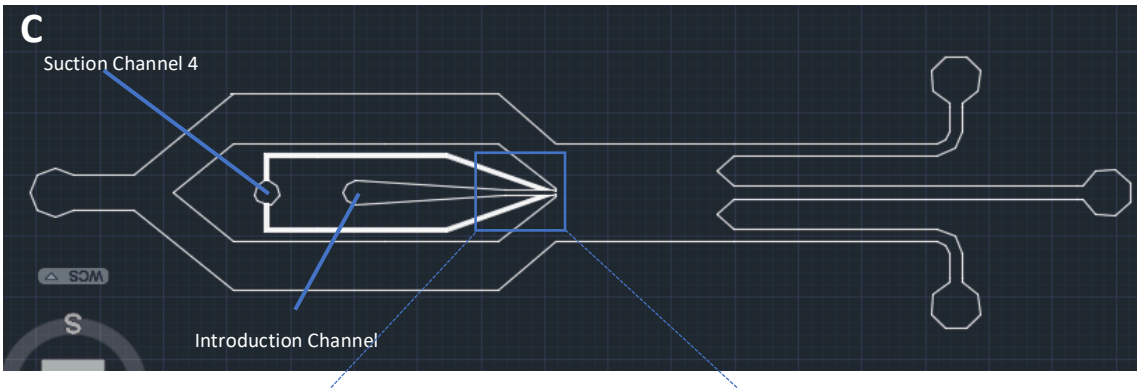
Chapter 4 represent the first successful step in the creation of microfluidic tool to measure chemosensory activities in the nervous system of *Ciona in vivo* and presents the first data of that kind in this species. As stated in Chapter 4, the microfluidic chip used in this study could be improved to load the *Ciona* larvae with more ease and without the prospect of cellular damage which could affect subsequent recordings. For this technology to be robust, the microfluidic system must be upgraded in the future. I will develop herein what should be the next step and possible solution to move it forward based on my personal experience and data gathered. The main solution I suggest is to implement a fourth channel with a negative pressure in the pressure controller MFCS™-EZ (Fig. 5.1A). This upgrade would allow to use suction and larvae could be immobilised easily in that way. Immobilising animals through suction in a PDMS microfluidic chip has

been succeeded with *Caenorhabditis elegans*, which has a similar size (about 1 mm) and a tubular shape like *Ciona* (Rohde *et al*., 2007; Zeng *et al*., 2008; McCormick *et al*., 2011; Zhao *et al*., 2013). Note that such upgrade would not complicate further the existing setup as it could be integrated into existing installations. The reason this strategy was not chosen immediately was due to the limited funding available which enabled the purchase of control devices for at most three pressure channels. Future funds are therefore necessary to make this technology more usable, but it is within technical reach. I estimated the cost of the upgrade by Fluigent to be about 1, 500 £ (Fig. 5.1B). Based on my experience, I suggest what that next microfluidic chip design could look like if a fourth negative pressure channel is implemented in the system (Fig. 5.1C). Overall, this technology if one decides to carry on with it, has the potential to be revolutionary in that field and not only for *Ciona* but for other protochordates such as amphioxus. Having such a tool developed would open a whole new area of research and bring answers to exciting questions that could not be tested so far.

B

PRICE SUMMARY

Description	Part Number	Unit price (EUR)	Quantity	Total price (EUR)		Final net price (EUR)
MFCS extension channel -345 mbar upgrade	R80345001	1940	1	1 940,00		1 940,00
discount on the extension						- 355,00
Bottle-CAP: Pressure CAP for lab bottle pressurization	Res-CAP	102	1	102,00		102,00
Bottle-CAP connector and tubing kit	CTQ-KIT-BC	39	1	39,00		39,00
Packing, shipping and insurance cost – Area 1	SHIP-1	100	1	100,00		100,00
FINAL PACKAGE PRICE (Euros)						1 826,00 €



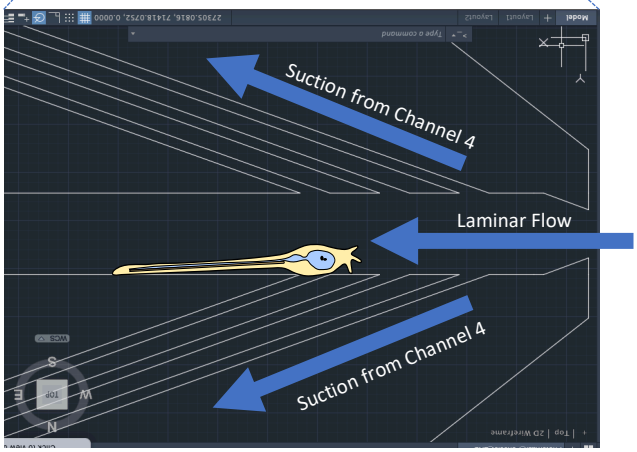


Figure 5.1. Suggestion of an upgraded design using suction. (A) The pressure controller MFCES-EZ can have up to four pressure channels, so a fourth one can be added (red arrow) to drive negative pressure. (B) Estimation of the fourth channel upgrade cost by Fluigent. (C) Suggestion of the new upgraded microfluidic design. After placement in the introduction channel the larva could be immobilised by the suction forces coming from the channel 4 and stay in position upon application of subsequent laminar flows.

5.2 Lamprey developmental GnRH-related gene markers reveal homologies and discrepancies at different levels

In Chapter 2, the different GnRH developmental marker genes characterised in lamprey were first identified as expressed in developing hypothalamic GnRH neurons of various vertebrates. The selected GnRH-related marker genes were candidates with regulatory potential, that is transcription factors and signalling molecules with specific developmental roles. The list mostly comprised early marker genes highly active at the placodal stages, although also persisting later in development (*FGF8*, *ISL1*, *COE2* and *NELF*), but also a marker of mature neuronal identity (*GnRH1*). In lamprey, the orthologues of these TFs and signalling molecule genes showed a great deal of variation in their expression specificity, presumably reflecting their various role in cell fate determination. In contrast, the GnRH genes presented highly restricted expression as they associate with a specific neuronal cell fate. Hence, these complementary sets of marker genes were valuable to decipher different degrees of homologies and hunt down the origin of different components of GnRH cell specification. Overall, the data collected on the expression patterns of these regulatory genes highlight several heterochronic and heterotopic changes in comparison to Gnathostomes. Whether such changes could explain the evolutionary reprogramming in the specification processes of hypothalamic GnRH cells is one of the key questions for evolutionary developmental biology. Further studies in the lamprey will be necessary to elucidate the function of the GnRH regulatory genes in this model system and will contribute to further advancements in this field.

5.3 The evolutionary link between the nose, hypothalamus, and adenohypophysis

The precise evolutionary history of chemoreception remains enigmatic as no unifying theory exists to explain how the different olfactory systems have evolved in metazoans. More mysterious in this story is the anatomical links between the nose, hypothalamus, and adenohypophysis in vertebrates. In this thesis, I focused on the Olfactores lineage with lamprey and the ascidian *C. intestinalis* as models. The latter taxon, the ascidian, is appealing as it presents an opportunity to better understand the evolutionary study of chemoreception despite the high variation in ascidian's morphologies and lifestyles. In fact, several territories in ascidians are thought to host chemosensory neurons. In particular, the larval palps have been suggested to host chemosensory and GnRH cells (cf. Chapter 1). This has been based on several characters: the expression of sensory and secretory cell markers in palp cells that are also expressed in olfactory neurons, the developmental origin of their progenitors, and their position anterior to the brain (Poncelet and Shimeld, 2020; this study). While one recent study points to a connection between the palp ACCs and myoepithelial cell types (Johnson et al., 2020), the results presented in Chapter 3 instead provide insight into an evolutionary connection between palp cells and olfactory neurons. The comparison of the expression and putative function of *CiMs4a and CiGnRH1* in *Ciona* and vertebrates argue that ACCs are not a myoepithelial homolog, instead highlighting the characters shared by ACCs and olfactory neurons. ACCs develop in a comparable fashion to olfactory progenitors from the edges of the neural plate, express *Ms4A* and *GnRH* and their specification is driven by TFs such as *ISL, COE, DLX and DMRT (*Wagner et al., 2012; Johnson et al., 2020), like OSNs and olfactory-derived

GnRH1/3 neurons (Duan and Allard, 2020). The ACCs are not the only cells which have been postulated to have combined chemosensory and GnRH neuroendocrine activities as the apical anterior trunk epidermal neurons (aATENs) have also been discussed to have such dual functions (Abitua et al., 2015). As mentioned in Chapter 3, we can argue that such anterior cells resemble the ancestral multifunctional precursors that have segregated into olfactory and hypothalamic sister cell types to evolve into the 'olfacto-hypothalamo-adenohypophyseal' circuit of vertebrates. The evolution of such complex neuronal circuits can be elegantly explained through the principle of cell type functional segregation. To explain the evolution of the olfactory nervous system, it requires us to accept the following assumptions: (1) the emergent functionally separated cell types moved away from each other due to their specialisation (2) the segregated sister cell types remain interconnected through axonal and dendritic contact to maintain cellular contact for functional coordination. Having in mind these ideas, we can envision an evolutionary scenario to explain the existing axon tracts that interconnect the vertebrate nose, hypothalamus, and adenohypophysis. This scenario presumes that in the vertebrate nose some of the olfactory sensory neurons and GnRH neuroendocrine neurons of the hypothalamus emerged by functional segregation of sister cell types, which evolved from a chemosensory–neurosecretory organ/cell type that was related to the palp ACCs in extant *Ciona* (cf. Chapter 3). Thus, these sister cells would have moved apart from each other according to their specific function but would have remained interconnected via axons from the nose to the hypothalamus and through the GnRH neuroendocrine process from the hypothalamus to the adenohypophysis. Consequently, from a vertebrate perspective, the *Ciona* ACCs seem to combine olfactory (chemosensory) and hypothalamic (GnRH-secreting) functions but not adenohypophyseal function (gonadotropin- and luteinizing hormone-secreting). In *Ciona*, the homolog structure to

the pituitary is thought to be the neurohypophyseal duct derived from the oral siphon primordium (OSP) in embryos (Boorman and Shimeld, 2002, Christiaen et al., 2002, Graham and Shimeld, 2013). The OSP and palps derive from the same ectodermal cell line (a10.79 and a10.79* lines at late neurula), with OSP progenitors inserted between the GnRH/chemosensory ACCs and aATENs. The close developmental connectivity between olfactory neurons and adenohypophysis homolog in *Ciona* is supported by the fact that in vertebrates these domains are initially congruent: while they detach into distinct olfactory, GnRH and adenophyposeal systems during development of jawed vertebrates, they remain tightly connected in the cyclostomes, consistent with a common developmental origin of the adenohypophyseal and olfactory cell types. In agreement with that theory is the fact that the hypothalamic GnRH-secreting cells in gnathostomes originate from the olfactory placodes, which reinforces the view that they emerged from a peripheral endocrine organ that was associated with the olfactory system. Although, in lamprey, the olfactory placodal developmental origin of GnRH neuroendocrine cells appears to have been lost, there is evidence of direct axonal projection pathways from olfactory neurons to the GnRH neurons in the hypothalamus, revealing that these systems are still connected as the chemical cues detected in the environment likely influence the lamprey reproductive GnRH systems functionally (Tobet et al., 1996).

5.4 Alternative evolutionary scenario for ACCs

While data in Chapter 3 suggest that *Ciona* ACCs represent and homolog to vertebrate olfactory neurons, it is important to note that the different olfactory neurons in the Olfactores lineage might have independent evolutionary origins and be non-homologous.

In fact, other processes can lead to similarities between structures that have quite different evolutionary origins due to common response to a similar environment and selection pressure. Thus, the simple evolutionary scenario proposed in Chapter 3 could be more complex. Indeed, increased number of cell types by functional segregation of an ancestral multifunctional cell type is not the only scenario possible to explain the apparent multifunctionality of ACCs. The opposite trend is also possible as in extreme case there can be a decrease in cell type number which can occur due to the loss of function but also through fusion of cellular identities. This process resulting in a functionally hybrid cell requires the co-option of CoRCs from different cell types. The ACCs may have evolved through the co-option of a myoepithelial gene regulatory network into olfactory cells resulting in a fusion of cell types, rather than by diversification of a pre-existing multifunctional cell type. Such examples exists, for instance the Merckel cells, a type of vertebrate mechanoreceptor cells are thought to have evolved by cell type fusion from the co-option of a mechanosensory gene regulatory network involving *Atoh1*, *Pou4f3* and *Gfi1* into epidermal cells that used Piezo channel to sense membrane (Arendt et al., 2016).

5.5 MS4A biological functions in ascidians

Nothing is known about the specific functions of ascidian MS4As although information gathered on mammalian MS4As can be extrapolated. Mouse Ms4a family members are odorant receptors causing calcium increase upon stimulation with some long chain fatty acids, steroids, and heterocyclic compounds (Greer et al., 2016). Ms4a ligands in mice have significance as they involve appetitive long-chain fatty acids enriched in seeds and nuts, as well as the female pheromone 2,5-dimethylpyrazine (Greer et al., 2016). What

biological processes might MS4A receptors mediate in ascidians? The most obvious role we can speculate is that ascidian's MS4As are required to choose an appropriate site for settlement. Two elements support that claim: (1) A *MS4A* was shown to be expressed in the most anterior end of *Ciona* swimming larvae in the external protrusions of the palp ACCs, which would be the first point of contact with a substrate to settle on; (2) Tunicate *MS4A* genes were only detected in ascidians and not in thaliacean and larvacean which have pelagic lifestyle and do not fix to a substrate to have a sessile existence. Therefore, we can infer that the lack of a sessile lifestyle made those genes obsolete in those latter tunicates and there was therefore no incentive to keep that genomic information. However, it remains possible that *MS4A* genes were not detected in those lineages due to the more limited amount of genomic data available for larvacean and thaliacean. Among the possible cues suggested to be recognized by ascidian larvae during settlement are exopolymers secreted from marine bacteria (Szewzyk et al., 1991; Holmström et al., 1992). Recognizing such chemical cues has evolutionary advantages on where or not to settle, as some marine bacteria like *Pseudoalteromonas tunicata* produce antifouling agents that kills the larvae of *Ciona* (Holmström et al., 1998). There are however other roles we can speculate on, as MS4A represents the only convincing chemoreceptor family described in ascidian genomes it is likely that they have a wider array of functions. In mammals, the *MS4As* are expressed in various immune cell types such as *MS4A1* in human B-cells or *MS4A4B* in mice naïve CD8+ and CD4+ T-cells and natural killer cells (NK) (Zuccolo et al., 2013; Xu et al., 2006). It is therefore conceivable that ascidian immune cells detect non-self material through their numerous MS4As. Interestingly, the ascidian MS4As are not the only kind of receptors in the palp cells whose homologs are expressed in the immune system of vertebrates. For instance, Ciona express in the palps a C-type lectin homolog to the CD94 vertebrate receptor expressed on NK cells (Zucchetti et al.,

2008). Another role we can hypothesise for *MS4As* in ascidians is that these genes are required for reproduction. In mouse, the MS4A13-2 protein localizes in the head of spermatozoa and is necessary for an adequate fertilisation of the oocyte via its interaction with the zona pellucida (Kaneko et al., 2017). We could thus hypothesise that the phenomenon of sperm chemotaxis in ascidian towards chemoattractants (such as sulphated steroid) release by the eggs also requires MS4As (Yoshida et al., 2002). Another described role of MS4A in mammals is in the intestinal tract. For example, human MS4A8B and MS4A12 are expressed in epithelial cells at the lumenal surface of the small and large intestine, which sense dietary lipids (Michel et al., 2013; Koslowski et al., 2008). The intestinal tract of adult ascidians may also sense various molecules in their filtered aliments via their MS4As. All these functions discussed above are of course hypothetical, but they provide further route for future research to better characterise the shared roles of those chemoreceptor genes in the Olfactores lineage.

5.6 A first step towards a new technology to detect chemosensory responses at single-cell level in *Ciona intestinalis*

Pioneering work of functional calcium imaging in *Ciona* using the genetically encoded calcium sensor GCaMP was first achieved by Christopher Hackley and colleagues to demonstrate the requirement of transiently expressed connexin for anterior neural plate development (Hackley et al., 2013). Since then, research on calcium imaging in *Ciona* embryos has been largely stagnant with very few studies. Especially, the calcium imaging of late *Ciona* larvae is only at its beginning mainly because the strong muscle cells present in larva's tail and resultant rapid movement have so far prohibited steady calcium imaging of individual cells. This blockage has started to be overcome here as the calcium

imaging presented in Chapter 4 has demonstrated the feasibility of systematic functional imaging of chemosensation in *Ciona* larvae. In fact, the calcium imaging experiments in this work constitute the first direct and comprehensive attempt to physiologically study the chemosensory systems of *C. intestinalis* larvae and more generally of any protochordate. The experiments have allowed me to physiologically test and confirm the existence of CO_2-evoked responses, a type of data that has been out of reach in the field, although the protocol developed comes with some limitations. Further systematic experiments targeted at additional tissues and cells would be necessary to understand the precise chemical repertoire recognised by the different organs of *Ciona* larvae but overall, it sheds light on the chemosensory capabilities of ascidians. The study also provides a template for the development of whole-brain imaging approaches, which strengthens the attractiveness of *Ciona* larvae as models for system-wide understanding of chordate nervous circuits at single-cell resolution. Nonetheless, there is room for improvement, for instance a more sophisticated microfluidic trap than the one used here could be developed to improve the fine-tuned immobilisation. If such devices are successfully created in the future, then it would allow the physiological mapping of chemosensory cells in the entire Ciona body, as well as their downstream circuits. Especially in the experimentally active field of chemosensation, we could expect this to be achieved within a few years.

Ciona larvae are particularly interesting for the easy identification of numerous gene regulatory regions enabling precise cellular targeting of transgenes. In this study, I used a regulatory region (*Dmrt*) with a wide expression territory because of the uncertainty of the responding cells, but nothing precludes the experimenter from choosing a more precise regulatory region that only targets a highly limited number of cells. This type of

approach is particularly important if more precise markers of the palp are to be used next. Some work has already been done in that direction, as I cloned the B-Y-crystallin (Fig. 5.2A) and pVglutT regulatory regions in front of GCaMP6m (Fig. 5.2B). These two regulatory regions distinguish between the palp ACCs and PSNs (Zeng et al., 2019). These molecular constructs could be used to differentiate the different sensory responses of those palp cells, which are the most interesting from their anatomy and the expression of olfactory neuronal markers such as GnRH and MS4A (cf. Chapter 3).

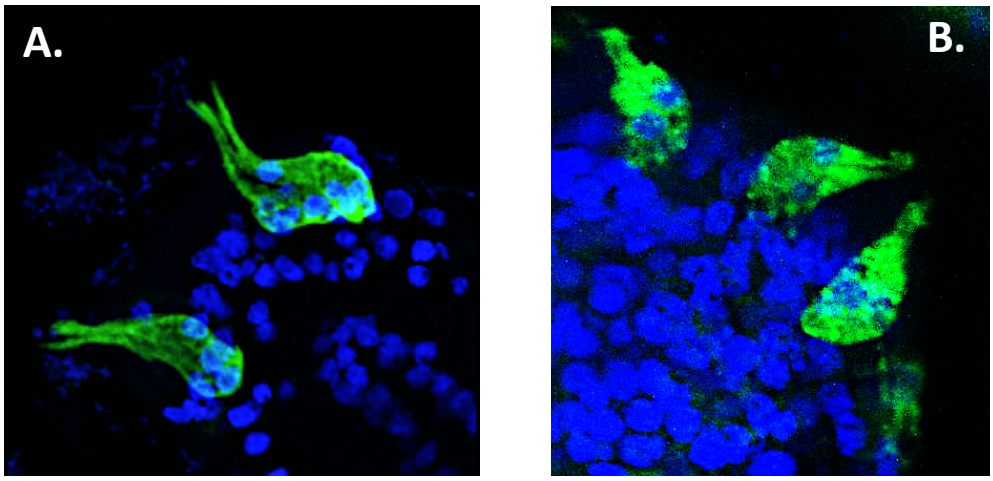


Figure 5.2. Additional molecular constructs for cell-type specific calcium imaging in the palp (A) B-Y-crystallin>GCaMP6m (green) marking the palp ACCs. (B) pVGLUT>GCaMP6m (green) marking the palp PSNs. Both images are in a transgenic 22hpf *Ciona intestinalis* larvae. Dapi staining in blue mark cell nuclei.

5.7 Other type of genetic tool applicable to Ciona

New molecular tools such as GCaMP6s, first described in 2013 (Chen *et al*., 2013) were rapidly translated into non-conventional model systems such as *Ciona* four years later (Akahoshi *et al*., 2017). Therefore, novel fast developing tools such as genetically encoded fluorescent voltage indicators (GEVIs) which allow to monitor neuronal activity at higher spatial and temporal resolution than GCaMP might reasonably soon be applied to *Ciona* (Abdelfattah et al., 2019). For instance, the GEVI ArcLight, a chimeric voltage indicator

made of a mutated version of super ecliptic pHluorin (a GFP) and of a voltage sensitive phosphatase derived from *Ciona intestinalis*, has recently shown its potential to do precise voltage imaging of olfactory cells in transgenic mice (Platisa *et al*., 2020). Also, photoconvertible probes such as CaMPARI (Fosque *et al*., 2015) which enable one to identify active neurons and circuits *in vivo* during a chosen period could well be applied in *Ciona* in the future and enlarge the toolbox of functional imaging methods to understand the neural basis of behaviour.

5.8 Physiological information joined to Single-cell RNA-seq and connectomic data

The experimental suitability and simple nervous system architecture of only 177 neurons makes *Ciona* a model organism of choice for comparative studies in evolutionary neurobiology. In addition, its phylogenetic position is a main advantage over main invertebrate models such as *Drosophila melanogaster* and *Caenorhabditis elegans* as it has conservation of chordate ancestral features in its body plan and homologies in cell types (Bezares-Calderon and Jekely, 2016). Furthermore, a key strength of *Ciona intestinalis* as a model organism is the existence of single-cell transcriptomic and connectomic data (Ryan, Lu & Meinertzhagen, 2016; Sharma *et al*., 2019; Cao *et al*., 2019). The possibility to add and map physiological data from single cells to such resources would open the door to yet more powerful investigations of cell types and represents another level in the system-wide understanding of organs and cell types. Functional imaging data in *Ciona* are scarce but such an approach has been recently taken to suggest a connection between the activity of GnRH2-expressing cells and the tail movement of *Ciona* larvae (Okawa *et al*., 2020).

5.9 Potential applications for biofouling control strategies

The study of *Ciona* chemosensory systems is interesting for the insights it brings into our understanding of vertebrate olfactory evolution. However, another reason to investigate *Ciona* olfaction is for the potential applications that may be derived to control this highly invasive species. In fact, *Ciona* invasions are a particular issue in aquaculture farms as they weigh down equipment and compete with farmed bivalves for resources, leading to substantial economic losses (Global Invasive Species Database, 2015). Another concerning fact of the negative impact on the fishing industry is that *Ciona* has been shown to be a carrier of *Neoparamoeba pemaquidensis,* a parasite causing amoebic gill disease (Tan *et al.,* 2002). In consequence, there is a genuine need for effective control of tunicate invasions, but management plans tried to date have been unsuccessful. Chemical methods could be used to control and prevent settlement as there is evidence to suggest that *Ciona* larvae use chemical cues to determine where to settle (Ryan, Lu & Meinertzhagen, 2016, this study). Chemical control strategies using calcium hydroxide or acetic acid have been shown effective against tunicates, but they are biocidal to a variety of non-target species and difficult to apply in natural settings (Locke *et al*., 2009). Therefore, the study of the olfactory system in *Ciona* using same or similar methods as described in Chapter 4, would help us identify the key chemical cues necessary to attract *Ciona* larva and lead to the development of effective chemical trap and management strategies. An effective control of *Ciona* invasions would reduce economic losses to aquaculture farms and protect biodiversity and ecosystem functions of native communities.

REFERENCES

Abdelfattah, S. A. K., T.; Singh, A.; Novak, O.; Liu, H.; Shuai, Y.; Huang, Y-C.; Campagnola, L.; Seeman, S. C.; Yu, J.; Zheng, J.; Grimm, J. B.; Patel, R.; Friedrich, J.; Mensh, B. D.; Paninski, L.; Macklin, J. J.; Murphy G. J.; and K. L. Podgorski, B-J.; Chen, T-W.; Turner, G. C.; Liu, Z.; Koyama, M.; Svoboda, K.; Ahrens, M. B.; Lavis, L. D.; Schreiter, E. R. (2019). "Bright and photostable chemigenetic indicators for extended in vivo voltage imaging." Science **364**: 699-704.

Abdul-Wajid, S., H. Morales-Diaz, S. M. Khairallah and W. C. Smith (2015). "T-type Calcium Channel Regulation of Neural Tube Closure and EphrinA/EPHA Expression." Cell Rep **13**(4): 829-839.

Abitua, P. B., T. B. Gainous, A. N. Kaczmarczyk, C. J. Winchell, C. Hudson, K. Kamata, M. Nakagawa, M. Tsuda, T. G. Kusakabe and M. Levine (2015). "The pre-vertebrate origins of neurogenic placodes." Nature **524**(7566): 462-465.

Abraham, E., O. Palevitch, S. Ijiri, S. J. Du, Y. Gothilf and Y. Zohar (2008). "Early development of forebrain gonadotrophin-releasing hormone (GnRH) neurones and the role of GnRH as an autocrine migration factor." J Neuroendocrinol **20**(3): 394-405.

Adams, B. A., J. A. Tello, J. Erchegyi, C. Warby, D. J. Hong, K. O. Akinsanya, G. O. Mackie, W. Vale, J. E. Rivier and N. M. Sherwood (2003). "Six novel gonadotropin-releasing hormones are encoded as triplets on each of two genes in the protochordate, Ciona intestinalis." Endocrinology **144**(5): 1907-1919.

Aguillon, R., J. Batut, A. Subramanian, R. Madelaine, P. Dufourcq, T. F. Schilling and P. Blader (2018). "Cell-type heterogeneity in the early zebrafish olfactory epithelium is generated from progenitors within preplacodal ectoderm." Elife **7**.

Aguillon, R., P. Blader and J. Batut (2016). "Patterning, morphogenesis, and neurogenesis of zebrafish cranial sensory placodes." Methods Cell Biol **134**: 33-67.

Ahuja, G., S. Bozorg Nia, V. Zapilko, V. Shiriagin, D. Kowatschew, Y. Oka and S. I. Korsching (2014). "Kappe neurons, a novel population of olfactory sensory neurons." Sci Rep **4**: 4037.

Akahoshi, T., K. Hotta and K. Oka (2017). "Characterization of calcium transients during early embryogenesis in ascidians Ciona robusta (Ciona intestinalis type A) and Ciona savignyi." Dev Biol **431**(2): 205-214.

Almeida, F. C., A. Sánchez-Gracia, K. K. O. Walden, H. M. Robertson and J. Rozas (2015). "Positive selection in extra cellular domains in the diversification of Strigamia maritima chemoreceptors." Frontiers in Ecology and Evolution **3**.

Appelgate, V. C. (1950). "Natural history of the sea lamprey (Petromyzon marinus) in Michigan. ." U.S. Fish and Wildlife Service Special Report: Fisheries **55**.

Arendt, D. (2008). "The evolution of cell types in animals: emerging principles from molecular studies." Nat Rev Genet **9**(11): 868-882.

Arendt, D. (2021). "Elementary nervous systems." Philos Trans R Soc Lond B Biol Sci **376**(1821): 20200347.

Arendt, D., J. M. Musser, C. V. H. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev, G. Schlosser, S. Widder, M. D. Laubichler and G. P. Wagner (2016). "The origin and evolution of cell types." Nat Rev Genet **17**(12): 744-757.

Baatrup, E. (1981). "Primary Sensory Cells in the Skin of Amphioxus (Branchiostorna lanceolaturn (P))." Acta Zoologica **62**(3): 147-157.

Baatrup, E. (1982). "On the Structure of the Corpuscles of de Quatrefages (Branchiostorna lanceolaturn (P))." Acta Zoologica **63**: 39-44.

Bahn, Y. S. and F. A. Muhlschlegel (2006). "CO2 sensing in fungi and beyond." Curr Opin Microbiol **9**(6): 572-578.

Bailey, A. P., S. Bhattacharyya, M. Bronner-Fraser and A. Streit (2006). "Lens specification is the ground state of all sensory placodes, from which FGF promotes olfactory identity." Dev Cell **11**(4): 505-517.

Baker, C. V. and M. Bronner-Fraser (1997). "The origins of the neural crest. Part II: an evolutionary perspective." Mechanisms of Development **69**: 13-29.

Baker, C. V. and M. Bronner-Fraser (2001). "Vertebrate cranial placodes I. Embryonic induction." Dev Biol **232**(1): 1-61.

Banerjee, N. and E. A. Hallem (2020). "The role of carbon dioxide in nematode behaviour and physiology." Parasitology **147**(8): 841-854.

Barnett, L. M., T. E. Hughes and M. Drobizhev (2017). "Deciphering the molecular mechanism responsible for GCaMP6m's Ca2+-dependent change in fluorescence." PLoS One **12**(2): e0170934.

Barraud, P., A. A. Seferiadis, L. D. Tyson, M. F. Zwart, H. L. Szabo-Rogers, C. Ruhrberg, K. J. Liu and C. V. Baker (2010). "Neural crest origin of olfactory ensheathing glia." Proc Natl Acad Sci U S A **107**(49): 21040-21045.

Bassham, S. and J. H. Postlethwait (2005). "The evolutionary history of placodes: a molecular genetic investigation of the larvacean urochordate Oikopleura dioica." Development **132**(19): 4259-4272.

Bear, D. M., J. M. Lassance, H. E. Hoekstra and S. R. Datta (2016). "The Evolving Neural and Genetic Architecture of Vertebrate Olfaction." Curr Biol **26**(20): R1039-R1049.

Bezares-Calderon, L. A. and G. Jekely (2016). "Think small." Elife **5**.

Biechl, D., K. Tietje, S. Ryu, B. Grothe, G. Gerlach and M. F. Wullimann (2017). "Identification of accessory olfactory system and medial amygdala in the zebrafish." Sci Rep **7**: 44295.

Bingham, B. L. Y., C. M. (1991). "Larval behavior of the ascidian Ecteinascidia turbinata Herdman; an in situ experimental study of the effects of swimming on dispersal." Zoolog Sci. **145**(2): 0-204.

Bjerselius, R., W. Li, J. H. Teeter, J. G. Seelye, P. B. Johnsen, P. J. Maniak, G. C. Grant, C. N. Polkinghorne and P. W. Sorensen (2000). "Direct behavioral evidence that unique bile acids released by larval sea lamprey (Petromyzon marinus) function as a migratory pheromone." Canadian Journal of Fisheries and Aquatic Sciences **57**(3): 557-569.

Blähser, S. K., J. A.; Kuenzel, W. J. (1989). "Testing of Arg-8-gonadotropin-releasing hormone-directed antisera by immunological and immunocytochemical methods for use in comparative studies. Histochemistry **93**: 39-48.
Bollner, T., K. Holmberg and R. Olsson (1986). "A Rostral Sensory Mechanism in Oikopleura dioica (Appendicularia)." Acta Zoologica **Vol. 67, No. 4, pp. 235-241**.

Bone, Q. (1998). "The Biology of Pelagic Tunicates." Oxford University Press, Oxford : 340.

Bone, Q. and A. C. G. Best (1978). "Ciliated Sensory Cells in Amphioxus (Branchiostoma)." Journal of the Marine Biological Association of the United Kingdom **58**(2): 479-486.

Boorman, C. S., 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**(5): 354-365.

Braun, C. B. (1998). "Schreiner Organs: A New Craniate Chemosensory Modality in Hagfishes." The Journal of Compartive Neurology **392**: 135-163.

Braun, C. B., H. Wicht and G. R. Northcutt (1995). "Distribution of Gonadotropin-Releasing Hormone Immunoreactivity in the Brain of the Pacific Hagfish, Eptatretus stouti (Craniata:Myxinoidea)." The journal of compartive neurology **353**: 464-476.

Buck, L. A., R. (1991). "A Novel Multigene Family May Encode Odorant Receptors: A Molecular Basis for Odor Recognition." Cell **65**: 175-187.

Burgus, R. B., M.; Amoss, M.; Ling, N.; Monahan, M.H.; Rivier, J.; Fellows, R.; Blackwell, R.; Vale, W.; and Guillemin, R. (1972). "Primary structure of the ovine hypothalamic luteinizing hormone releasing factor (LRF). Proc. Natl. Acad. Sci., 69, 278 (1972)." Proc. Natl. Acad. Sci. **69**: 278-282.

Caicci, F., G. Zaniolo, P. Burighel, V. Degasperi, F. Gasparini and L. Manni (2010). "Differentiation of papillae and rostral sensory neurons in the larva of the ascidian Botryllus schlosseri (Tunicata)." J Comp Neurol **518**(4): 547-566.

Candelier, R., M. S. Murmu, S. A. Romano, A. Jouary, G. Debregeas and G. Sumbre (2015). "A microfluidic device to study neuronal and motor responses to acute chemical stimuli in zebrafish." Sci Rep **5**: 12196.

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

Canestro, C., R. Albalat, M. Irimia and J. Garcia-Fernandez (2013). "Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates." Semin Cell Dev Biol **24**(2): 83-94.

Cao, C., L. A. Lemaire, W. Wang, P. H. Yoon, Y. A. Choi, L. R. Parsons, J. C. Matese, W. Wang, M. Levine and K. Chen (2019). "Comprehensive single-cell transcriptome lineages of a proto-vertebrate." Nature **571**(7765): 349-354.

Capella-Gutierrez, S., J. M. Silla-Martinez and T. Gabaldon (2009). "trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses." Bioinformatics **25**(15): 1972-1973.

Carrillo, M. A., M. L. Guillermin, S. Rengarajan, R. P. Okubo and E. A. Hallem (2013). "O2-sensing neurons control CO2 response in C. elegans." J Neurosci **33**(23): 9675-9683.
Chang, C. Y. L., Y. X.; Zhu, Y. T.; Zhu, H. H. (1985). "The reproductive endocrinology of Amphioxus." Frontiers in Physiological Research. Eds. by D. G. Carlick and P. I. Korner, Autstralian Academy of Science, Canberra: 79-86.

Chang, S., Y.-W. Chung-Davidson, S. V. Libants, K. G. Nanlohy, M. Kiupel, B. C. T. and W. Li (2013). "The sea lamprey has a primordial accessory olfactory system." BMC Evolutionary Biology **13**(172): 1-11.

Chartier, T. F., J. Deschamps, W. Durichen, G. Jekely and D. Arendt (2018). "Whole-head recording of chemosensory activity in the marine annelid Platynereis dumerilii." Open Biol **8**(10).

Chen, J. Y., D. Y. Huang, Q. Q. Peng, H. M. Chi, X. Q. Wang and M. Feng (2003). "The first tunicate from the Early Cambrian of South China." Proc Natl Acad Sci U S A **100**(14): 8314-8318.

Chen, M., Z. Peng and S. He (2010). "Olfactory receptor gene family evolution in stickleback and medaka fishes." Sci China Life Sci **53**(2): 257-266.

Chen, T. W., T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman, L. L. Looger, K. Svoboda and D. S. Kim (2013). "Ultrasensitive fluorescent proteins for imaging neuronal activity." Nature **499**(7458): 295-300.

Cho, H. J., Y. Shan, N. C. Whittington and S. Wray (2019). "Nasal Placode Development, GnRH Neuronal Migration and Kallmann Syndrome." Front Cell Dev Biol **7**: 121.

Christiaen, L., E. Wagner, W. Shi and M. Levine (2009). "X-gal staining of electroporated sea squirt (Ciona) embryos." Cold Spring Harb Protoc **2009**(12): pdb prot5346.

Chung, W. C., M. L. Linscott, K. M. Rodriguez and C. E. Stewart (2016). "The Regulation and Function of Fibroblast Growth Factor 8 and Its Function during Gonadotropin-Releasing Hormone Neuron Development." Front Endocrinol (Lausanne) **7**: 114.

Chung, W. C. J. and P. S. Tsai (2010). "Role of fibroblast growth factor signaling in gonadotropin-releasing hormone neuronal system development." Front Horm Res **39**: 37-50.

Chuong, E. B., N. C. Elde and C. Feschotte (2017). "Regulatory activities of transposable elements: from conflicts to benefits." Nat Rev Genet **18**(2): 71-86.

Churcher, A. M. and J. S. Taylor (2009). "Amphioxus (Branchiostoma floridae) has orthologs of vertebrate odorant receptors." BMC Evol Biol **9**: 242.

Churcher, A. M. and J. S. Taylor (2011). "The antiquity of chordate odorant receptors is revealed by the discovery of orthologs in the cnidarian Nematostella vectensis." Genome Biol Evol **3**: 36-43.

Clark, M. E. M., P. L. (1995). "The POU Homeodomain Transcription Factor Oct-1 Is Essential for Activity of the Gonadotropin-Releasing Hormone Neuron-Specific Enhancer." Molecular and Cellular Biology **15**(11): 6169-6177.

Cloney, R. (1982). "Ascidians larvae and the events of metamorphosis." Amer. Zool. **22**: 817-826.

Comai, G. E., M. Tesarova, V. Dupe, M. Rhinn, P. Vallecillo-Garcia, F. da Silva, B. Feret, K. Exelby, P. Dolle, L. Carlsson, B. Pryce, F. Spitz, S. Stricker, T. Zikmund, J. Kaiser, J. Briscoe, A. Schedl, N. B. Ghyselinck, R. Schweitzer and S. Tajbakhsh (2020). "Local retinoic acid signaling directs emergence of the extraocular muscle functional unit." PLoS Biol **18**(11): e3000902.

Conway Morris, S. (1982). "Atlas of the Burgess Shale." London, UK: Palaeontological Assoiciation.

Corradi, A., L. Croci, V. Broccoli, S. Zecchini, S. Previtali, W. Wurst, S. Amadio, R. Maggi, A. Quattrini and G. G. Consalez (2003). "Hypogonadotropic hypogonadism and peripheral neuropathy in Ebf2-null mice." Development **130**(2): 401-410.

Croset, V., R. Rytz, S. F. Cummins, A. Budd, D. Brawand, H. Kaessmann, T. J. Gibson and R. Benton (2010). "Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction." PLoS Genet **6**(8): e1001064.

Dardaillon, J., D. Dauga, P. Simion, E. Faure, T. A. Onuma, M. B. DeBiasse, A. Louis, K. R. Nitta, M. Naville, L. Besnardeau, W. Reeves, K. Wang, M. Fagotto, M. Gueroult-Bellone, S. Fujiwara, R. Dumollard, M. Veeman, J. N. Volff, H. Roest Crollius, E. Douzery, J. F. Ryan, B. Davidson, H. Nishida, C. Dantec and P. Lemaire (2020). "ANISEED 2019: 4D exploration of genetic data for an extended range of tunicates." Nucleic Acids Res **48**(D1): D668-D675.

Darwin, C. (1871). "The Descent of Man, and Selection in Relation to Sex." London John Murray(1-2): 107.

Dawson, H. A., B. R. Quintella, P. R. Almeida, A. J. Treble and J. C. Jolley (2015). The Ecology of Larval and Metamorphosing Lampreys. Lampreys: Biology, Conservation and Control**:** 75-137.

de Beer, G. R. (1924). "On a Problematical Organ of the Lamprey." J Anat **59**: 97-107.

de Quatrefages, M. A. (1845). "Sur le systeme nerveux et sur I'histologie du Branchiostome ou amphioxus." Anals. Sci. Nat. Zool. **4**: 197-248.

Dean, B. (1899). "On the embryology of Bdellostoma stouti. A genera account of myxinoid development from the egg and segmentation to hatching." Festschrift zum 70ten Geburststag Carl von Kupffer". Gustav Fischer Verlag. Jena.: 220-276.

Decatur, W. A., J. A. Hall, J. J. Smith, W. Li and S. A. Sower (2013). "Insight from the lamprey genome: glimpsing early vertebrate development via neuroendocrine-associated genes and shared synteny of gonadotropin-releasing hormone (GnRH)." Gen Comp Endocrinol **192**: 237-245.

Dehal, P., Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A. Di Gregorio, M. Gelpke, D. M. Goodstein, N. Harafuji, K. E. Hastings, I. Ho, K. Hotta, W. Huang, T. Kawashima, P. Lemaire, D. Martinez, I. A. Meinertzhagen, S. Necula, M. Nonaka, N. Putnam, S. Rash, H. Saiga, M. Satake, A. Terry, L. Yamada, H. G. Wang, S. Awazu, K. Azumi, J. Boore, M. Branno, S. Chin-Bow, R. DeSantis, S. Doyle, P. Francino, D. N. Keys, S. Haga, H. Hayashi, K. Hino, K. S. Imai, K. Inaba, S. Kano, K. Kobayashi, M. Kobayashi, B. I. Lee, K. W. Makabe, C. Manohar, G. Matassi, M. Medina, Y. Mochizuki, S. Mount, T. Morishita, S. Miura, A. Nakayama, S. Nishizaka, H. Nomoto, F. Ohta, K. Oishi, I. Rigoutsos, M. Sano, A. Sasaki, Y. Sasakura, E. Shoguchi, T. Shin-i, A. Spagnuolo, D. Stainier, M. M. Suzuki, O. Tassy, N. Takatori, M. Tokuoka, K. Yagi, F. Yoshizaki, S. Wada, C. Zhang, P. D. Hyatt, F. Larimer, C. Detter, N. Doggett, T. Glavina, T. Hawkins, P. Richardson, S. Lucas, Y. Kohara, M. Levine, N. Satoh

and D. S. Rokhsar (2002). "The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins." Science **298**(5601): 2157-2167.
Delarbre, C., C. Gallut, V. Barriel, P. Janvier and G. Gachelin (2002). "Complete mitochondrial DNA of the hagfish, Eptatretus burgeri: the comparative analysis of mitochondrial DNA sequences strongly supports the cyclostome monophyly." Mol Phylogenet Evol **22**(2): 184-192.

Dellovade, T. L. P., D.W.; Schwanzel-Fukuda, M. (1998). "The gonadotropin-releasing hormone system does not develop in Small-Eye (Sey) mouse phenotype. 107 (2), 233–240." Brain Res. Dev. **107**(2): 233-240.

Delsuc, F., H. Brinkmann, D. Chourrout and H. Philippe (2006). "Tunicates and not cephalochordates are the closest living relatives of vertebrates." Nature **439**(7079): 965-968.

Delsuc, F., H. Philippe, G. Tsagkogeorga, P. Simion, M. K. Tilak, X. Turon, S. Lopez-Legentil, J. Piette, P. Lemaire and E. J. P. Douzery (2018). "A phylogenomic framework and timescale for comparative studies of tunicates." BMC Biol **16**(1): 39.

Derjean, D., A. Moussaddy, E. Atallah, M. St-Pierre, F. Auclair, S. Chang, X. Ren, B. Zielinski and R. Dubuc (2010). "A novel neural substrate for the transformation of olfactory inputs into motor output." PLoS Biol **8**(12): e1000567.

Derobert, Y. B. B. L., M.; Mazan, S. (2002). "Pax6 expression patterns in Lampetra fluviatilis and Scyliorhinus canicula embryos suggest highly conserved roles in the early regionalization of the vertebrate brain." Brain Research Bulletin **57**: 277-280.

Dieterich, D. C., A. Karpova, M. Mikhaylova, I. Zdobnova, I. Konig, M. Landwehr, M. Kreutz, K. H. Smalla, K. Richter, P. Landgraf, C. Reissner, T. M. Boeckers, W. Zuschratter, C. Spilker, C. I. Seidenbecher, C. C. Garner, E. D. Gundelfinger and M. R. Kreutz (2008). "Caldendrin-Jacob: a protein liaison that couples NMDA receptor signalling to the nucleus." PLoS Biol **6**(2): e34.

Dolcemascolo, G. P., R.; De Bernardi, F.; Damiani, F.; Gianguzza, M. (2009). "Ultrastructural comparative analysis on the adhesive papillae of the swimming larvae of three ascidian species." ISJ **6**: S77-S86.

Døving, K. B. (1998). "The olfactory system of hagfishes."

Døving, K. B. and K. Holmberg (1974). "A Note on the Function of the Olfactory Organ of the Hagfish Myxine glutinosa." Acta physiol. scand. **91**: 430-432.

Duan, C. and J. Allard (2020). "Gonadotropin-releasing hormone neuron development in vertebrates." Gen Comp Endocrinol **292**: 113465.

Dubois, L. V., A. (2001). "The COE – Collier/Olf1/EBF – transcription factors: structural conservation and diversity of developmental functions." Mechanisms of Development **108**: 3-12.

Dulac, C. and A. T. Torello (2003). "Molecular detection of pheromone signals in mammals: from genes to behaviour." Nat Rev Neurosci **4**(7): 551-562.

Eisthen, H. L. (1992). "Phylogeny of the Vomeronasal System and of Receptor Cell Types in the Olfactory and Vomeronasal Epithelia of Vertebrates
." microscopy research and technique **23**: 1-21.

Eisthen, H. L. D., R. J.; Wirsig-Wiechmann, C. R.; Dionne, V. E. (2000). "Neuromodulatory Effects of Gonadotropin Releasing Hormone on Olfactory Receptor Neurons." The Journal of Neuroscience **20**(11): 3947-3955.

Ericson, J. T., S.; Edlund, T.; Jessell, T.M.; Yamada, T. (1992). "Early stages of motor neuron differentiation revealed by expression of homeobox gene Islet-1." Science **256**: 1555-1560.

Eyun, S. I., H. Y. Soh, M. Posavi, J. B. Munro, D. S. T. Hughes, S. C. Murali, J. Qu, S. Dugan, S. L. Lee, H. Chao, H. Dinh, Y. Han, H. Doddapaneni, K. C. Worley, D. M. Muzny, E. O. Park, J. C. Silva, R. A. Gibbs, S. Richards and C. E. Lee (2017). "Evolutionary History of Chemosensory-Related Gene Families across the Arthropoda." Mol Biol Evol **34**(8): 1838-1862.

Fabian, P. T., K-C.; Smeeton, J.; Lancman, J. J.; Si Dong, P. D.; Cerny, R.; Crump, J. G. (2020). "Lineage analysis reveals an endodermal contribution to the vertebrate pituitary." Science **370**: 463-467.

Falardeau, J., W. C. Chung, A. Beenken, T. Raivio, L. Plummer, Y. Sidis, E. E. Jacobson-Dickman, A. V. Eliseenkova, J. Ma, A. Dwyer, R. Quinton, S. Na, J. E. Hall, C. Huot, N. Alois, S. H. Pearce, L. W. Cole, V. Hughes, M. Mohammadi, P. Tsai and N. Pitteloud (2008). "Decreased FGF8 signaling causes deficiency of gonadotropin-releasing hormone in humans and mice." J Clin Invest **118**(8): 2822-2831.

Ferrandiz, C. and A. Sessions (2008). "Preparation and hydrolysis of digoxygenin-labeled probes for in situ hybridization of plant tissues." CSH Protoc **2008**: pdb prot4942.

Ferrando, S., M. Bottaro, L. Gallus, L. Girosi, M. Vacchi and G. Tagliafierro (2006). "Observations of crypt neuron-like cells in the olfactory epithelium of a cartilaginous fish." Neurosci Lett **403**(3): 280-282.

Ferrando, S., L. Gallus, C. Gambardella, M. Vacchi and G. Tagliafierro (2010). "G protein alpha subunits in the olfactory epithelium of the holocephalan fish Chimaera monstrosa." Neurosci Lett **472**(1): 65-67.

Flores, A. R. F., Z. (2008). "Texture preferences of ascidian tadpole larvae during settlement." Marine and Freshwater Behaviour and Physiology **41**(3): 155-159.

Forni, P. E., K. Bharti, E. M. Flannery, T. Shimogori and S. Wray (2013). "The indirect role of fibroblast growth factor-8 in defining neurogenic niches of the olfactory/GnRH systems." J Neurosci **33**(50): 19620-19634.

Forni, P. E., C. Taylor-Burds, V. S. Melvin, T. Williams and S. Wray (2011). "Neural crest and ectodermal cells intermix in the nasal placode to give rise to GnRH-1 neurons, sensory neurons, and olfactory ensheathing cells." J Neurosci **31**(18): 6915-6927.

Forni, P. E. and S. Wray (2012). "Neural crest and olfactory system: new prospective." Mol Neurobiol **46**(2): 349-360.

Fosque, B. F. S., Y.; Dana, H.; Yang, C.-T. Ohyama,T.; Tadross, M. R.; Patel, R.; Zlatic, M.; Kim, D. S.; Ahrens, M. B.; Jayaraman, V.; Looger, L. L.; Schreiter, E. R. (2015). "Labeling of active neural circuits in vivo with designed calcium integrators." Science **347**(6223): 755-760.

Freitag, J., A. Beck, L. Gunther, L. von Buchholtz and H. Breer (1999). "On the origin of the olfactory receptor family: receptor genes of the jawless fish (Lampetra fluviatilis)." Gene **226**: 165-174.

Frontini, A., A. U. Zaidi, H. Hua, T. P. Wolak, C. A. Greer, K. W. Kafitz, W. Li and B. S. Zielinski (2003). "Glomerular territories in the olfactory bulb from the larval stage of the sea lamprey Petromyzon marinus." J Comp Neurol **465**(1): 27-37.

Gai, Z., P. C. Donoghue, M. Zhu, P. Janvier and M. Stampanoni (2011). "Fossil jawless fish from China foreshadows early jawed vertebrate anatomy." Nature **476**(7360): 324-327.

Gaillard, A. L., B. H. Tay, D. I. Perez Sirkin, A. G. Lafont, C. De Flori, P. G. Vissio, S. Mazan, S. Dufour, B. Venkatesh and H. Tostivint (2018). "Characterization of Gonadotropin-Releasing Hormone (GnRH) Genes From Cartilaginous Fish: Evolutionary Perspectives." Front Neurosci **12**: 607.

Gainous, T. B., E. Wagner and M. Levine (2015). "Diverse ETS transcription factors mediate FGF signaling in the Ciona anterior neural plate." Dev Biol **399**(2): 218-225.

Gans, C. and R. G. Northcutt (1983). "Neural Crest and the Origin of Vertebrates: A New Head." Science **220**(4594): 268-274.

Gasparini, F. and L. Ballarin (2018). Reproduction in Tunicates. Encyclopedia of Reproduction**:** 546-553.
Gee, H. (2018). "Across the Bridge: Understanding the Origin of the Vertebrates. ." Chicago: University of Chicago Press.

Giuliano, P. M., R.; M.; Pinto, M. R.; De Santis, R. (1998). "Gene expression pattern Identification and developmental expression of Ci-isl, a homologue of vertebrate islet genes, in the ascidian Ciona intestinalis." Mech Dev **78**: 199-202.

Glover, C. N., D. Newton, J. Bajwa, G. G. Goss and T. J. Hamilton (2019). "Behavioural responses of the hagfish Eptatretus stoutii to nutrient and noxious stimuli." Sci Rep **9**(1): 13369.

Gonzalez, A., R. Morona, J. M. Lopez, N. Moreno and R. G. Northcutt (2010). "Lungfishes, like tetrapods, possess a vomeronasal system." Front Neuroanat **4**: 1-11.

Gorbman, A. T., A. (1985). "Early development of oral, olfactory and adenohypophyseal structures of Agnathans and its evolutionary implications." Evolutionary Biology of Primitive Fishes: 165-185.

Graham, A. and S. M. Shimeld (2013). "The origin and evolution of the ectodermal placodes." J Anat **222**(1): 32-40.

Gray, J. K., D.; Lu, H.; Chang, A.; Chang, J.; Ellis, R. E.; Marletta, M. A.; Bargmann, C. I. (2004). "Oxygen sensation and social feeding mediated by a C. elegans guanylate cyclase homologue. ." Nature **430**: 317-322.

Green, W. W., A. Basilious, R. Dubuc and B. S. Zielinski (2013). "The neuroanatomical organization of projection neurons associated with different olfactory bulb pathways in the sea lamprey, Petromyzon marinus." PLoS One **8**(7): e69525.

Greer, P. L., D. M. Bear, J. M. Lassance, M. L. Bloom, T. Tsukahara, S. L. Pashkovski, F. K. Masuda, A. C. Nowlan, R. Kirchner, H. E. Hoekstra and S. R. Datta (2016). "A Family of non-GPCR Chemosensors Defines an Alternative Logic for Mammalian Olfaction." Cell **165**(7): 1734-1748.

Grienberger, C. and A. Konnerth (2012). "Imaging calcium in neurons." Neuron **73**(5): 862-885.

Groppelli, S., R. Pennati, G. Scarì, C. Sotgia and F. De Bernardi (2003). "Observations on the settlement of phallusia mammillata larvae: Effects of different lithological substrata." Italian Journal of Zoology **70**(4): 321-326.

Grus, W. E. and J. Zhang (2006). "Origin and evolution of the vertebrate vomeronasal system viewed through system-specific genes." Bioessays **28**(7): 709-718.

Grus, W. E. and J. Zhang (2009). "Origin of the genetic components of the vomeronasal system in the common ancestor of all extant vertebrates." Mol Biol Evol **26**(2): 407-419.

Gwee, P. C., B. H. Tay, S. Brenner and B. Venkatesh (2009). "Characterization of the neurohypophysial hormone gene loci in elephant shark and the Japanese lamprey: origin of the vertebrate neurohypophysial hormone genes." BMC Evol Biol **9**: 47.

Hackley, C., E. Mulholland, G. J. Kim, E. Newman-Smith and W. C. Smith (2013). "A transiently expressed connexin is essential for anterior neural plate development in Ciona intestinalis." Development **140**(1): 147-155.

Haeckel, E. (1874). "Die Gastraea-Theorie, die phylogenetische Classification des Thierreichs und die Homologie der Keimblatter." Jenaische Zischr Naturw **8**: 1-55.

Hagelin, L. and A. G. Johnels (1955). "On the structure and function of the accessory olfactory organ in lampreys." Acta Zool **36**: 113–125.

Hammond, K. L. and T. T. Whitfield (2006). "The developing lamprey ear closely resembles the zebrafish otic vesicle: otx1 expression can account for all major patterning differences." Development **133**(7): 1347-1357.

Harafuji, N., D. N. Keys and M. Levine (2002). "Genome-wide identification of tissue-specific enhancers in the Ciona tadpole." Proc Natl Acad Sci U S A **99**(10): 6802-6805.

Hatschek, B. (1881). "Studien über die Entwicklung des Amphioxus." Arb Zool. InstWien. **4**: 1-99.

Hecht, S. (1918). "The physiology of Ascidia atra lesueur." Journal of Experimental Zoology **25**(1): 229-259.

Heimberg, A. M., R. Cowper-Sal-lari, M. Semon, P. C. Donoghue and K. J. Peterson (2010). "microRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate." Proc Natl Acad Sci U S A **107**(45): 19379-19383.

Hilal, E. M., J. H. Chen and A.-J. Silverman (1996). "Joint Migration of Gonadotropin-Releasing Hormone (GnRH) and NeuropeptideY (NPY) Neurons from Olfactory Placode to Central Nervous System." Journal of Neurobiology **31**(4): 487-502.

Hobert, O. (2016). "Terminal Selectors of Neuronal Identity." Curr Top Dev Biol **116**: 455-475.

Holland, L. Z. (2005). "Non-neural ectoderm is really neural: Evolution of developmental patterning mechanisms in the non-neural ectoderm of chordates and the problem of sensory cell homologies." Journal of Experimental Zoology(304B): 304-323.

Holland, L. Z. (2016). "Tunicates." Curr Biol **26**(4): R146-152.

Holland, L. Z., M. Schubert, N. D. Holland and T. Neuman (2000). "Evolutionary conservation of the presumptive neural plate markers AmphiSox1/2/3 and AmphiNeurogenin in the invertebrate chordate amphioxus." Dev Biol **226**(1): 18-33.

Holland, N. D. Y., J-K. (2002). "Epidermal receptor development and sensory pathways in vitally stained amphioxus (Branchiostoma floridae)." Acta Zoologica **83**: 309-319.

Holmes, W. M., R. Cotton, V. B. Xuan, A. D. Rygg, B. A. Craven, R. L. Abel, R. Slack and J. P. Cox (2011). "Three-dimensional structure of the nasal passageway of a hagfish and its implications for olfaction." Anat Rec (Hoboken) **294**(6): 1045-1056.

Holmström, C. J., S.; Neilan, B. A.; White, D. C.; Kjellebergl, S. (1998). "Pseudoalteromonastunicata sp. now, a bacterium that produces antifouling agents." International Journal of Systematic Bacteriology **48**: 1205-1212.

Holmstrom, C. R., D.; Kjelleberg, S. (1992). "Inhibition of Settlement by Larvae of Balanus amphitrite and Ciona intestinalis by a Surface-Colonizing Marine Bacterium." Appl and Environ Microbiol **58**(7): 2111-2115.
Hong, C. S., B. Y. Park and J. P. Saint-Jeannet (2007). "The function of Dmrt genes in vertebrate development: it is not just about sex." Dev Biol **310**(1): 1-9.

Horie, R., A. Hazbun, K. Chen, C. Cao, M. Levine and T. Horie (2018). "Shared evolutionary origin of vertebrate neural crest and cranial placodes." Nature **560**(7717): 228-232.

Hozumi, A., S. Matsunobu, K. Mita, N. Treen, T. Sugihara, T. Horie, T. Sakuma, T. Yamamoto, A. Shiraishi, M. Hamada, N. Satoh, K. Sakurai, H. Satake and Y. Sasakura (2020). "GABA-Induced GnRH Release Triggers Chordate Metamorphosis." Curr Biol.

Hu, J. Z., C.; Ding, C.; Chi, Q.; Walz, A.; Mombaerts, P.; Matsunami, H.; Luo, M. (2007). "Detection of near-atmospheric concentrations of CO2 by an olfactory subsystem in the mouse." Science **317**: 953–957.

Hu, L., J. Ye, H. Tan, A. Ge, L. Tang, X. Feng, W. Du and B. F. Liu (2015). "Quantitative analysis of Caenorhabditis elegans chemotaxis using a microfluidic device." Anal Chim Acta **887**: 155-162.

Huang, X., C. S. Hong, M. O'Donnell and J. P. Saint-Jeannet (2005). "The doublesex-related gene, XDmrt4, is required for neurogenesis in the olfactory system." Proc Natl Acad Sci U S A **102**(32): 11349-11354.

Hurst, L. D., C. Pal and M. J. Lercher (2004). "The evolutionary dynamics of eukaryotic gene order." Nat Rev Genet **5**(4): 299-310.

Hussain, A. (2011). "The Olfactory Nervous System Of Terrestrial And Aquatic Vertebrates." Nature Precedings.

Hussain, A., L. R. Saraiva and S. I. Korsching (2009). "Positive Darwinian selection and the birth of an olfactory receptor clade in teleosts." Proc Natl Acad Sci U S A **106**(11): 4313-4318.

Imai, J. H. and I. A. Meinertzhagen (2007a). "Neurons of the ascidian larval nervous system in Ciona intestinalis: I. Central nervous system." J Comp Neurol **501**(3): 316-334.

Imai, J. H. and I. A. Meinertzhagen (2007b). "Neurons of the ascidian larval nervous system in Ciona intestinalis: II. Peripheral nervous system." J Comp Neurol **501**(3): 335-352.

Ishibashi, K. S., M.; Sasaki, S.; Imai, M. (2001). "Identification of a new multigene four-transmembrane family (MS4A) related to CD20, HTm4 and b subunit of the high-affinity IgE receptor." Gene **264**: 87-93.

Jackman, W. R., J. A. Langeland and C. B. Kimmel (2000). "islet reveals segmentation in the Amphioxus hindbrain homolog." Dev Biol **220**(1): 16-26.

Jandzik, D., M. B. Hawkins, M. V. Cattell, R. Cerny, T. A. Square and D. M. Medeiros (2014). "Roles for FGF in lamprey pharyngeal pouch formation and skeletogenesis highlight ancestral functions in the vertebrate head." Development **141**(3): 629-638.

Janvier, P. (1996). "Early Vertebrates." Oxford Scientific Publications New York.

Jefferies, R. P. S. (1991). "Biological Asymmetry and Handedness " Wiley: 94-127.

Johanson, Z. B., C. A.; Trinajstic, K. (2019). Early Vertebrates and the Emergence of Jaws. Heads, Jaws, and Muscles**:** 23-44.

Johnson, C. J., F. Razy-Krajka and A. Stolfi (2020). "Expression of smooth muscle-like effectors and core cardiomyocyte regulators in the contractile papillae of Ciona." Evodevo **11**: 15.

Johnson, D. S., B. Davidson, C. D. Brown, W. C. Smith and A. Sidow (2004). "Noncoding regulatory sequences of Ciona exhibit strong correspondence between evolutionary constraint and functional importance." Genome Res **14**(12): 2448-2456.

Jones, W. (2013). "Olfactory carbon dioxide detection by insects and other animals." Mol Cells **35**(2): 87-92.
Kamesh, N., G. K. Aradhyam and N. Manoj (2008). "The repertoire of G protein-coupled receptors in the sea squirt Ciona intestinalis." BMC Evol Biol **8**: 129.

Kaneko, T., K. Toshimori and H. Iida (2017). "Subcellular localization of MS4A13 isoform 2 in mouse spermatozoa." Reproduction **154**(6): 843-857.

Kaplam, J. M. H., H.R. (1993). "A dual mechanosensory and chemosensory neuron in Caenorhabditis elegans ." Proc. Natl. Acad. Sci. USA **90**: 2227-2231.

Kari, W., F. Zeng, L. Zitzelsberger, J. Will and U. Rothbacher (2016). "Embryo Microinjection and Electroporation in the Chordate Ciona intestinalis." J Vis Exp(116).

Karpova, A., M. Mikhaylova, S. Bera, J. Bar, P. P. Reddy, T. Behnisch, V. Rankovic, C. Spilker, P. Bethge, J. Sahin, R. Kaushik, W. Zuschratter, T. Kahne, M. Naumann, E. D. Gundelfinger and M. R. Kreutz (2013). "Encoding and transducing the synaptic or extrasynaptic origin of NMDA receptor signals to the nucleus." Cell **152**(5): 1119-1133.

Katoh, H., S. Shibata, K. Fukuda, M. Sato, E. Satoh, N. Nagoshi, T. Minematsu, Y. Matsuzaki, C. Akazawa, Y. Toyama, M. Nakamura and H. Okano (2011). "The dual origin of the peripheral olfactory system: placode and neural crest." Mol Brain **4**: 34.

Katoh, K., J. Rozewicki and K. D. Yamada (2019). "MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization." Brief Bioinform **20**(4): 1160-1166.

Kauffman, A. S. (2004). "Emerging functions of gonadotropin-releasing hormone II in mammalian physiology and behaviour." J Neuroendocrinol **16**(9): 794-806.

Kavanaugh, S. I., M. Nozaki and S. A. Sower (2008). "Origins of gonadotropin-releasing hormone (GnRH) in vertebrates: identification of a novel GnRH in a basal vertebrate, the sea lamprey." Endocrinology **149**(8): 3860-3869.

Kawai, T., Y. Oka and H. Eisthen (2009). "The role of the terminal nerve and GnRH in olfactory system neuromodulation." Zoolog Sci **26**(10): 669-680.

Kawauchi, S., J. Shou, R. Santos, J. M. Hebert, S. K. McConnell, I. Mason and A. L. Calof (2005). "Fgf8 expression defines a morphogenetic center required for olfactory neurogenesis and nasal cavity development in the mouse." Development **132**(23): 5211-5223.

Khan, A., O. Fornes, A. Stigliani, M. Gheorghe, J. A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Cheneby, S. R. Kulkarni, G. Tan, D. Baranasic, D. J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W. W. Wasserman, F. Parcy and A. Mathelier (2018). "JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework." Nucleic Acids Res **46**(D1): D260-D266.

Kim, K., M. E. Colosimo, H. Yeung and P. Sengupta (2005). "The UNC-3 Olf/EBF protein represses alternate neuronal programs to specify chemosensory neuron identity." Dev Biol **286**(1): 136-148.

King, C. J. S., S. A.; Anthony, E. L. P. (1988). "Neuronal systems immunoreactive with antiserum to lamprey gonadotropin-releasing hormone in the brain of Petromyzon marinus." Cell Tissue Res **253**: 1-8.

Kleerekoper, H. and G. A. van Erkel (1960). "The olfactory apparatus of Petromyzon marinus L." Canadian Journal of Zoology **38**: 209-223.

Ko, E. K., L. P. Chorich, M. E. Sullivan, R. S. Cameron and L. C. Layman (2018). "JAK/STAT signaling pathway gene expression is reduced following Nelf knockdown in GnRH neurons." Mol Cell Endocrinol **470**: 151-159.
Koide, T., Y. Yabuki and Y. Yoshihara (2018). "Terminal Nerve GnRH3 Neurons Mediate Slow Avoidance of Carbon Dioxide in Larval Zebrafish." Cell Rep **22**(5): 1115-1123.

Kojima, W. (2015). "Attraction to Carbon Dioxide from Feeding Resources and Conspecific Neighbours in Larvae of the Rhinoceros Beetle Trypoxylus dichotomus." PLoS One **10**(11): e0141733.

Kosakovsky Pond, S. L. and S. D. Frost (2005). "Not so different after all: a comparison of methods for detecting amino acid sites under selection." Mol Biol Evol **22**(5): 1208-1222.

Koslowski, M., U. Sahin, K. Dhaene, C. Huber and O. Tureci (2008). "MS4A12 is a colon-selective store-operated calcium channel promoting malignant cell processes." Cancer Res **68**(9): 3458-3466.

Kozmik, Z., N. D. Holland, J. Kreslova, D. Oliveri, M. Schubert, K. Jonasova, L. Z. Holland, M. Pestarino, V. Benes and S. Candiani (2007). "Pax-Six-Eya-Dach network during amphioxus development: conservation in vitro but context specificity in vivo." Dev Biol **306**(1): 143-159.

Kramer, P. R. W., S. (2000). "Novel gene expressed in nasal region influences outgrowth of olfactory axons and migration of luteinizing hormone-releasing hormone (LHRH) neurons." Genes Dev **14**: 1824-2834.
Kramer, P. R. W., S. (2001). "Nasal embryonic LHRH factor (NELF) expression within the CNS and PNS of the rodent." Gene Expr Patterns **1**: 23-26.

Kuraku, S. (2013). "Impact of asymmetric gene repertoire between cyclostomes and gnathostomes." Semin Cell Dev Biol **24**(2): 119-127.

Kuraku, S. and S. Kuratani (2006). "Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences." Zoolog Sci **23**(12): 1053-1064.

Kuratani, S., Y. Nobusada, N. Horigome and Y. Shigetani (2001). "Embryology of the lamprey and evolution of the vertebrate jaw: insights from molecular and developmental perspectives." Philos Trans R Soc Lond B Biol Sci **356**(1414): 1615-1632.

Kuratani, S. U., T.; Aizawa, S.; Hirano, S. (1997). "Peripheral development of cranial nerves in a cyclostome, Lampetra japonica: morphological distribution of nerve branches and the vertebrate body plan." J Comp Neurol **384**(4): 483-500.

Kusakabe, T., R. Yoshida, I. Kawakami, R. Kusakabe, Y. Mochizuki, L. Yamada, T. Shin-i, Y. Kohara, N. Satoh, M. Tsuda and Y. Satou (2002). "Gene expression profiles in tadpole larvae of Ciona intestinalis." Dev Biol **242**(2): 188-203.

Kusakabe, T. G., T. Sakai, M. Aoyama, Y. Kitajima, Y. Miyamoto, T. Takigawa, Y. Daido, K. Fujiwara, Y. Terashima, Y. Sugiuchi, G. Matassi, H. Yagisawa, M. K. Park, H. Satake and M. Tsuda (2012). "A conserved non-reproductive GnRH system in chordates." PLoS One **7**(7): e41955.

Lacalli, T. C. (2004). "Sensory systems in amphioxus: a window on the ancestral chordate condition." Brain Behav Evol **64**(3): 148-162.

Lacalli, T. C., T. H. J. Gilmour and S. J. Kelly (1999). "The oral nerve plexus in amphioxus larvae: function, cell types and phylogenetic signifcance." Proc. R. Soc. Lond. B. **266**: 1461-1470.

Lacalli, T. C. H., S. (1999). "A reexamination of the epithelial sensory cells of amphioxus (Branchiostoma)." Acta Zoologica **80**: 125-134.

Laframboise, A. J., X. Ren, S. Chang, R. Dubuc and B. S. Zielinski (2007). "Olfactory sensory neurons in the sea lamprey display polymorphisms." Neurosci Lett **414**(3): 277-281.

Lamarck, J. B. (1816). "Histoire naturelle des animaux sans vertebres, vol. III: Tuniciers." Paris, France: Deterville.

Lara-Ramirez, R., C. Patthey and S. M. Shimeld (2015). "Characterization of two neurogenin genes from the brook lamprey lampetra planeri and their expression in the lamprey nervous system." Dev Dyn **244**(9): 1096-1108.

Lara-Ramirez, R., G. Poncelet, C. Patthey and S. M. Shimeld (2017). "The structure, splicing, synteny and expression of lamprey COE genes and the evolution of the COE gene family in chordates." Dev Genes Evol **227**(5): 319-338.

Larsson, A. (2014). "AliView: a fast and lightweight alignment viewer and editor for large datasets." Bioinformatics **30**(22): 3276-3278.

Leach, J. W. (1951). "The hypohysis of lampreys in relation to the nasal apparatus."

Li, K., C. O. Brant, M. Huertas, E. J. Hessler, G. Mezei, A. M. Scott, T. R. Hoye and W. Li (2018). "Fatty-acid derivative acts as a sea lamprey migratory pheromone." Proc Natl Acad Sci U S A **115**(34): 8603-8608.

Li, W., P. W. Sorensen and D. D. Gallaher (1995). "The Olfactory System of Migratory Adult Sea Lamprey (Petromyzon marinus) Is Specifically and Acutely Sensitive to Unique Bile Acids Released by Conspecific Larvae." J. Gen. Physiol. **105**: 569-587.

Libants, S., K. Carr, H. Wu, J. H. Teeter, Y. W. Chung-Davidson, Z. Zhang, C. Wilkerson and W. Li (2009). "The sea lamprey Petromyzon marinus genome reveals the early origin of several chemosensory receptor families in the vertebrate lineage." BMC Evol Biol **9**: 180.

Lien, S., B. F. Koop, S. R. Sandve, J. R. Miller, M. P. Kent, T. Nome, T. R. Hvidsten, J. S. Leong, D. R. Minkley, A. Zimin, F. Grammes, H. Grove, A. Gjuvsland, B. Walenz, R. A. Hermansen, K. von Schalburg, E. B. Rondeau, A. Di Genova, J. K. Samy, J. Olav Vik, M. D. Vigeland, L. Caler, U. Grimholt, S. Jentoft, D. I. Vage, P. de Jong, T. Moen, M. Baranski, Y. Palti, D. R. Smith, J. A. Yorke, A. J. Nederbragt, A. Tooming-Klunderud, K. S. Jakobsen, X. Jiang, D. Fan, Y. Hu, D. A. Liberles, R. Vidal, P. Iturra, S. J. Jones, I. Jonassen, A. Maass, S. W. Omholt and W. S. Davidson (2016). "The Atlantic salmon genome provides insights into rediploidization." Nature **533**(7602): 200-205.

Lionel Christiaen, L. B., P.; Smith, W. C.; Verniera, P.; Bourrata, F.; Joly, J-S. (2002). "Pitx genes in Tunicates provide new molecular insight into the evolutionary origin of pituitary." Gene **287**: 107-113.

Liu, B. and Y. Satou (2019). "Foxg specifies sensory neurons in the anterior neural plate border of the ascidian embryo." Nat Commun **10**(1): 4911.

Liu, B. and Y. Satou (2020). "The genetic program to specify ectodermal cells in ascidian embryos." Dev Growth Differ **62**(5): 301-310.

Locke, A. (2009). "Preliminary evaluation of effects of invasive tunicate management with acetic acid and calcium hydroxide on non-target marine organisms in Prince Edward Island, Canada." Aquatic Invasions **4**(1): 221-236.

Lu, T. M., Y. J. Luo and J. K. Yu (2012). "BMP and Delta/Notch signaling control the development of amphioxus epidermal sensory neurons: insights into the evolution of the peripheral sensory system." Development **139**(11): 2020-2030.

Lund, C., K. Pulli, V. Yellapragada, P. Giacobini, K. Lundin, S. Vuoristo, T. Tuuri, P. Noisa and T. Raivio (2016). "Development of Gonadotropin-Releasing Hormone-Secreting Neurons from Human Pluripotent Stem Cells." Stem Cell Reports **7**(2): 149-157.

Lund, C., V. Yellapragada, S. Vuoristo, D. Balboa, S. Trova, C. Allet, N. Eskici, K. Pulli, P. Giacobini, T. Tuuri and T. Raivio (2020). "Characterization of the human GnRH neuron developmental transcriptome using a GNRH1-TdTomato reporter line in human pluripotent stem cells." Dis Model Mech **13**(3).

Mackie, G. (1970). "Neuroid conduction and the evolution of conducting tissues." Q. Rev. Biol. **45**: 319-332.

Mackie, G. O. (1995). "On the “Visceral Nervous System” of Ciona." J. mar. bio. **75**: 141-151.

Madin, L. P. (1995). "Sensory ecology of salps (Tunicata, thaliacea): More questions than answers." Mar. Fresh. Behav. Physiol. **26**: 175-195.

Magdaleno, S., P. Jensen, C. L. Brumwell, A. Seal, K. Lehman, A. Asbury, T. Cheung, T. Cornelius, D. M. Batten, C. Eden, S. M. Norland, D. S. Rice, N. Dosooye, S. Shakya, P. Mehta and T. Curran (2006). "BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system." PLoS Biol **4**(4): e86.

Mallatt, J. and C. J. Winchell (2007). "Ribosomal RNA genes and deuterostome phylogeny revisited: more cyclostomes, elasmobranchs, reptiles, and a brittle star." Mol Phylogenet Evol **43**(3): 1005-1022.

Manni, L., A. Agnoletto, G. Zaniolo and P. Burighel (2005). "Stomodeal and neurohypophysial placodes in Ciona intestinalis: Insights into the origin of the pituitary gland." Journal of Experimental Zoology Part B-Molecular and Developmental Evolution **304b**(4): 324-339.

Marletaz, F., P. N. Firbas, I. Maeso, J. J. Tena, O. Bogdanovic, M. Perry, C. D. R. Wyatt, E. de la Calle-Mustienes, S. Bertrand, D. Burguera, R. D. Acemel, S. J. van Heeringen, S. Naranjo, C. Herrera-Ubeda, K. Skvortsova, S. Jimenez-Gancedo, D. Aldea, Y. Marquez, L. Buono, I. Kozmikova, J. Permanyer, A. Louis, B. Albuixech-Crespo, Y. Le Petillon, A. Leon, L. Subirana, P. J. Balwierz, P. E. Duckett, E. Farahani, J. M. Aury, S. Mangenot, P. Wincker, R. Albalat, E. Benito-Gutierrez, C. Canestro, F. Castro, S. D'Aniello, D. E. K. Ferrier, S. Huang, V. Laudet, G. A. B. Marais, P. Pontarotti, M. Schubert, H. Seitz, I. Somorjai, T. Takahashi, O. Mirabeau, A. Xu, J. K. Yu, P. Carninci, J. R. Martinez-Morales, H. R. Crollius, Z. Kozmik, M. T. Weirauch, J. Garcia-Fernandez, R. Lister, B. Lenhard, P. W. H. Holland, H. Escriva, J. L. Gomez-Skarmeta and M. Irimia (2018). "Amphioxus functional genomics and the origins of vertebrate gene regulation." Nature **564**(7734): 64-70.

Marquet, N., J. C. R. Cardoso, B. Louro, S. A. Fernandes, S. C. Silva and A. V. M. Canario (2020). "Holothurians have a reduced GPCR and odorant receptor-like repertoire compared to other echinoderms." Sci Rep **10**(1): 3348.

Martinez, I., E. G. Jones, S. L. Davie, F. C. Neat, B. D. Wigham and I. G. Priede (2011). "Variability in behaviour of four fish species attracted to baited underwater cameras in the North Sea." Hydrobiologia **670**(1): 23-34.

Matsunobu, S. and Y. Sasakura (2015). "Time course for tail regression during metamorphosis of the ascidian Ciona intestinalis." Dev Biol **405**(1): 71-81.

Matsuo, H., Balea, Y., Nair, R.M., Arimura, A. and Schally, A.V. (1971). "Structure of the porcine LH- and FSH-releasing hormone. I. The proposed amino acid sequence." Biochem. Biophys. Res. Commun. **43**: 1334-1339.

Mazet, F., J. A. Hutt, J. Milloz, J. Millard, A. Graham and S. M. Shimeld (2005). "Molecular evidence from Ciona intestinalis for the evolutionary origin of vertebrate sensory placodes." Dev Biol **282**(2): 494-508.

Mazet, F., S. Masood, G. N. Luke, N. D. Holland and S. M. Shimeld (2004). "Expression of AmphiCoe, an amphioxus COE/EBF gene, in the developing central nervous system and epidermal sensory neurons." Genesis **38**(2): 58-65.

McCormick, K. E., B. E. Gaertner, M. Sottile, P. C. Phillips and S. R. Lockery (2011). "Microfluidic devices for analysis of spatial orientation behaviors in semi-restrained Caenorhabditis elegans." PLoS One **6**(10): e25710.

McHenry, M. J. (2005). "The morphology, behavior, and biomechanics of swimming in ascidian larvae." Canadian Journal of Zoology **83**(1): 62-74.

Mehta, T. K., V. Ravi, S. Yamasaki, A. P. Lee, M. M. Lian, B. H. Tay, S. Tohari, S. Yanai, A. Tay, S. Brenner and B. Venkatesh (2013). "Evidence for at least six Hox clusters in the Japanese lamprey (Lethenteron japonicum)." Proc Natl Acad Sci U S A **110**(40): 16044-16049.

Meulemans, D. and M. Bronner-Fraser (2007). "The Amphioxus SoxB Family: Implications for the Evolution of Vertebrate Placodes." Int. J. Biol. Sci. **3**: 356-364.

Meyer, A. and Y. Van de Peer (2005). "From 2R to 3R: evidence for a fish-specific genome duplication (FSGD)." Bioessays **27**(9): 937-945.

Michel, J., K. Schonhaar, K. Schledzewski, C. Gkaniatsou, C. Sticht, B. Kellert, F. Lasitschka, C. Geraud, S. Goerdt and A. Schmieder (2013). "Identification of the novel differentiation marker MS4A8B and its murine homolog MS4A8A in colonic epithelial cells lost during neoplastic transformation in human colon." Cell Death Dis **4**: e469.

Miura, K., J. S. Acierno, Jr. and S. B. Seminara (2004). "Characterization of the human nasal embryonic LHRH factor gene, NELF, and a mutation screening among 65 patients with idiopathic hypogonadotropic hypogonadism (IHH)." J Hum Genet **49**(5): 265-268.

Miyashita, T., M. I. Coates, R. Farrar, P. Larson, P. L. Manning, R. A. Wogelius, N. P. Edwards, J. Anne, U. Bergmann, A. R. Palmer and P. J. Currie (2019). "Hagfish from the Cretaceous Tethys Sea and a reconciliation of the morphological-molecular conflict in early vertebrate phylogeny." Proc Natl Acad Sci U S A **116**(6): 2146-2151.

Modrell, M. S., D. Hockman, B. Uy, D. Buckley, T. Sauka-Spengler, M. E. Bronner and C. V. Baker (2014). "A fate-map for cranial sensory ganglia in the sea lamprey." Dev Biol **385**(2): 405-416.

Mohammadi, M. M., G.; Sun, L.; Tang, C.; Hirth, P.; Yeh, B. K.; Hubbard, S. R.; Schlessinger, J. (1997). "Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors." Science **276**(5314): 955-960.

Mojica Prieto, F. J. M., F. J. (2002). "The values of pK1 + pK2 for the dissociation of carbonic acid in seawater." Geochimica et Cosmochimica Acta **66**(14): 2529-2540.

Morris, S. C. and J. B. Caron (2014). "A primitive fish from the Cambrian of North America." Nature **512**(7515): 419-422.

Murakami, S. K., S.; Arai, Y. (1992). "The origin of the luteinizing hormone-releasing hormone (LHRH) neurons in newts(Cynopspyrrhogaster): the effect of olfactory placode ablation." Cell &Tissue Research **269**: 21-27.

Murakami, Y. O., M.; Sugahara, F.; Hirano, S.; Satoh, N. ; Kuratani, S. (2001). "Identification and expression of the lamprey Pax6 gene: evolutionary origin of the segmented brain of vertebrates." Development **128**: 3521-3531.

Muske, L. E. (1993). "Evolution of Gonadotropin-Releasing Hormone (GnRH) Neuronal Systems." Brain Behav Evol **42**: 215-230.

Myojin, M., T. Ueki, F. Sugahara, Y. Murakami, Y. Shigetani, S. Aizawa, S. Hirano and S. Kuratani (2001). "Isolation of Dlx and Emx Gene Cognates in an Agnathan Species, Lampetra japonica, and Their Expression

Patterns During Embryonic and Larval Development: Conserved and Diversified Regulatory Patterns of Homeobox Genes in Vertebrate Head Evolution." Journal of Experimental Zoology **291**: 68-84.

Nakai, U. O., M.; Imoto, K. (2001). "A high signal-to-noise Ca2+ probe composed of a single green fluorescent protein." Nature Biotechnology **19**: 137-141.

Nakashima, K., L. Yamada, Y. Satou, J. Azuma and N. Satoh (2004). "The evolutionary origin of animal cellulose synthase." Dev Genes Evol **214**(2): 81-88.

Nei, M., Y. Niimura and M. Nozawa (2008). "The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity." Nat Rev Genet **9**(12): 951-963.

Nei, M. T. G., T. (1986). "Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions." Mol. Biol. Evol. . **3**(5): 418-426.

Niimura, Y. (2009). "On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species." Genome Biol Evol **1**: 34-44.

Niimura, Y. (2009). "Evolutionary dynamics of olfactory receptor genes in chordates: Interaction between environments and genomic contents." Human Genomics **4**(2): 107-118.

Niimura, Y., A. Matsui and K. Touhara (2014). "Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals." Genome Res **24**(9): 1485-1496.

Nitta, K. R., R. Vincentelli, E. Jacox, A. Cimino, Y. Ohtsuka, D. Sobral, Y. Satou, C. Cambillau and P. Lemaire (2019). "High-Throughput Protein Production Combined with High- Throughput SELEX Identifies an Extensive Atlas of Ciona robusta Transcription Factor DNA-Binding Specificities." Methods Mol Biol **2025**: 487-517.

Nordstrom, K. J., R. Fredriksson and H. B. Schioth (2008). "The amphioxus (Branchiostoma floridae) genome contains a highly diversified set of G protein-coupled receptors." BMC Evol Biol **8**: 9.

Northcutt, L. G. M., L. E. (1994). "Multiple embryonic origins of gonadotropin-releasing hormone(GnRH) immunoreactive neurons." Developmental Brain Research Report **78**: 279-290.

Nozaki, M. and A. Gorbman (1992). "The Question of Functional Homology of Hatschek's Pit of Amphioxus (Branchiostoma helcheri) and the Vertebrate Adenohypophysis." Zoological Science **9**: 387-395.

Nozaki, M., K. Ominato, A. Gorbman and S. A. Sower (2000). "The distribution of lamprey GnRH-III in brains of adult sea lampreys (Petromyzon marinus)." Gen Comp Endocrinol **118**(1): 57-67.

Ohno, S. (1970). "Evolution by gene duplication." Springer Science & Business Media.
Oisi, Y., K. G. Ota, S. Kuraku, S. Fujimoto and S. Kuratani (2013). "Craniofacial development of hagfishes and the evolution of vertebrates." Nature **493**(7431): 175-180.

Oka, Y. and S. I. Korsching (2011). "Shared and unique G alpha proteins in the zebrafish versus mammalian senses of taste and smell." Chem Senses **36**(4): 357-365.

Okawa, N., K. Shimai, K. Ohnishi, M. Ohkura, J. Nakai, T. Horie, A. Kuhara and T. G. Kusakabe (2020). "Cellular identity and Ca(2+) signaling activity of the non-reproductive GnRH system in the Ciona intestinalis type A (Ciona robusta) larva." Sci Rep **10**(1): 18590.

Okubo, K., F. Sakai, E. L. Lau, G. Yoshizaki, Y. Takeuchi, K. Naruse, K. Aida and Y. Nagahama (2006). "Forebrain gonadotropin-releasing hormone neuronal development: insights from transgenic medaka and the relevance to X-linked Kallmann syndrome." Endocrinology **147**(3): 1076-1084.

Ota, K. G., S. Fujimoto, Y. Oisi and S. Kuratani (2011). "Identification of vertebra-like elements and their possible differentiation from sclerotomes in the hagfish." Nat Commun **2**: 373.

Palaniappan, T. K., L. Slekiene, L. Gunhaga and C. Patthey (2019). "Extensive apoptosis during the formation of the terminal nerve ganglion by olfactory placode-derived cells with distinct molecular markers." Differentiation **110**: 8-16.

Palevitch, O., E. Abraham, N. Borodovsky, G. Levkowitz, Y. Zohar and Y. Gothilf (2009). "Nasal embryonic LHRH factor plays a role in the developmental migration and projection of gonadotropin-releasing hormone 3 neurons in zebrafish." Dev Dyn **238**(1): 66-75.

Pandolfi, M., I. S. Parhar, M. A. Ravaglia, F. J. Meijide, M. C. Maggese and D. A. Paz (2002). "Ontogeny and distribution of gonadotropin-releasing hormone (GnRH) neuronal systems in the brain of the cichlid fish Cichlasoma dimerus." Anat Embryol (Berl) **205**(4): 271-281.

Papdogiannis, V., H. J. Parker, A. Pennati, C. Patthey, M. E. Bronner and S. M. Shimeld (2020). "Hmx gene conservation identifies the evolutionary origin of vertebrate cranial ganglia." Bioarchiv.

Paredes, R. M., J. C. Etzler, L. T. Watts, W. Zheng and J. D. Lechleiter (2008). "Chemical calcium indicators." Methods **46**(3): 143-151.

Parker, G. H. (1908). "The sensory reactions of amphioxus." Proc Am Acad Arts Sci **43**: 414-423.
Parker, H. J., M. E. Bronner and R. Krumlauf (2014). "A Hox regulatory network of hindbrain segmentation is conserved to the base of vertebrates." Nature **514**(7523): 490-493.

Parlier, D., V. Moers, C. Van Campenhout, J. Preillon, L. Leclere, A. Saulnier, M. Sirakov, H. Busengdal, S. Kricha, J. C. Marine, F. Rentzsch and E. J. Bellefroid (2013). "The Xenopus doublesex-related gene Dmrt5 is required for olfactory placode neurogenesis." Dev Biol **373**(1): 39-52.

Passamaneck, Y. J. and A. Di Gregorio (2005). "Ciona intestinalis: chordate development made simple." Dev Dyn **233**(1): 1-19.

Patthey, C., G. Schlosser and S. M. Shimeld (2014). "The evolutionary history of vertebrate cranial placodes--I: cell type evolution." Dev Biol **389**(1): 82-97.

Pennati, R., A. Dell'Anna, G. Zega and F. De Bernardi (2012). "Immunohistochemical study of the nervous system of the tunicate Thalia democratica (Forsskal, 1775)." European Journal of Histochemistry **56**: 96-101.

Pennati, R. G., S.; De Bernardi, F.; Mastrototaro,F.; Zega, G. (2009). "Immunohistochemical analysis of adhesive papillae of Clavelina lepadiformis (Müller, 1776) and Clavelina phlegraea (Salfi, 1929) (Tunicata, Ascidiacea)." European Journal of Histochemistry **53**: 25-34.

Petersen, J. K. S., T. (1995). "Larval dispersal in the ascidian Ciona intestinalis (L.). Evidence for a closed population." Journal of Experimental Marine Biology and Ecology **186**: 89-102.

Pitteloud, N., R. Quinton, S. Pearce, T. Raivio, J. Acierno, A. Dwyer, L. Plummer, V. Hughes, S. Seminara, Y. Z. Cheng, W. P. Li, G. Maccoll, A. V. Eliseenkova, S. K. Olsen, O. A. Ibrahimi, F. J. Hayes, P. Boepple, J. E. Hall, P. Bouloux, M. Mohammadi and W. Crowley (2007). "Digenic mutations account for variable phenotypes in idiopathic hypogonadotropic hypogonadism." J Clin Invest **117**(2): 457-463.

Platisa, J., H. Zeng, L. Madisen, L. B. Cohen, V. A. Pieribone and D. A. Storace (2020). "Voltage imaging using transgenic mouse lines expressing the GEVI ArcLight in two olfactory cell types. Bioarchiv.

Pombal, M. A. M., M. (2019). "Development and Functional Organization of the Cranial Nerves in Lampreys." The Anatomical Record **302**: 512-539.

Poncelet, G. and S. M. Shimeld (2020). "The evolutionary origins of the vertebrate olfactory system." Open Biol **10**(12): 200330.

Potter, I. C. (1980). "The Petromyzoniformes with particular reference to paired species." Can. J. Fish. Aquat. Sci. **37**: 1595-1615.

Powell, J. F. R., S.; Prakash, M.; Fischer, W.; Park, M.; Rivier, J.; Craig, G.; Mackie, G.; Sherwood, N. M. (1996). "Two new forms of gonadotropin-releasing hormone in a protochordate and the evolutionary implications." Proc. Natl. Acad. Sci. **93**: 10461-10464.

Prasad, B. C. Y., B.; Zackhary, R.; Schrader, K.; Seydoux, G.; Reed, R. R. (1998). "unc-3, a gene required for axonal guidance in Caenorhabditis elegans, encodes a member of the O/E family of transcription factors." Development **125**: 1561-1568.

Qiu, H. T., J. M. O. Fernandes, W. S. Hong, H. X. Wu, Y. T. Zhang, S. Huang, D. T. Liu, H. Yu, Q. Wang, X. X. You and S. X. Chen (2019). "Paralogues From the Expanded Tlr11 Gene Family in Mudskipper (Boleophthalmus pectinirostris) Are Under Positive Selection and Respond Differently to LPS/Poly(I:C) Challenge." Front Immunol **10**: 343.

Quaynor, S. D., E. K. Ko, L. P. Chorich, M. E. Sullivan, D. Demir, J. L. Waller, H. G. Kim, R. S. Cameron and L. C. Layman (2015). "NELF knockout is associated with impaired pubertal development and subfertility." Mol Cell Endocrinol **407**: 26-36.

Ramanathan, N., O. Simakov, C. A. Merten and D. Arendt (2015). "Quantifying Preferences and Responsiveness of Marine Zooplankton to Changing Environmental Conditions using Microfluidics." PLoS One **10**(10): e0140553.

Rast, J. P. and K. M. Buckley (2013). "Lamprey immunity is far from primitive." Proc Natl Acad Sci U S A **110**(15): 5746-5747.

Ratnere, I. and I. Dubchak (2009). "Obtaining comparative genomic data with the VISTA family of computational tools." Curr Protoc Bioinformatics **Chapter 10**: Unit 10 16.

Reed, K. L. M., J. K.; Tobet S. A.; Trudeauc, V. L.; MacEachernc, L.; Rubin, B. S.; Sower, S. A. (2002). "The Spatial Relationship of Gamma-Aminobutyric Acid (GABA) Neurons and Gonadotropin-Releasing Hormone (GnRH) Neurons in Larval and Adult Sea Lamprey, Petromyzon marinus." Brain Behav Evol **60**: 1-12.

Reilly, D. K., D. E. Lawler, D. R. Albrecht and J. Srinivasan (2017). "Using an Adapted Microfluidic Olfactory Chip for the Imaging of Neuronal Activity in Response to Pheromones in Male C. Elegans Head Neurons." J Vis Exp(127).

Ren, X., S. Chang, A. Laframboise, W. Green, R. Dubuc and B. Zielinski (2009). "Projections from the accessory olfactory organ into the medial region of the olfactory bulb in the sea lamprey (Petromyzon marinus): a novel vertebrate sensory structure?" J Comp Neurol **516**(2): 105-116.

Riviere, S., L. Challet, D. Fluegge, M. Spehr and I. Rodriguez (2009). "Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors." Nature **459**(7246): 574-577.

Robertson, F. M., M. K. Gundappa, F. Grammes, T. R. Hvidsten, A. K. Redmond, S. Lien, S. A. M. Martin, P. W. H. Holland, S. R. Sandve and D. J. Macqueen (2017). "Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification." Genome Biol **18**(1): 111.

Roch, G. J., J. A. Tello and N. M. Sherwood (2014). "At the transition from invertebrates to vertebrates, a novel GnRH-like peptide emerges in amphioxus." Mol Biol Evol **31**(4): 765-778.

Rohde, C. B., F. Zeng, R. Gonzalez-Rubio, M. Angel and M. F. Yanik (2007). "Microfluidic system for on-chip high-throughput whole-animal sorting and screening at subcellular resolution." Proc Natl Acad Sci U S A **104**(35): 13891-13895.

Root, A. R., N. V. Nucci, J. D. Sanford, B. S. Rubin, V. L. Trudeau and S. A. Sower (2005). "In situ characterization of gonadotropin- releasing hormone-I, -III, and glutamic acid decarboxylase expression in the brain of the sea lamprey, Petromyzon marinus." Brain Behav Evol **65**(1): 60-70.

Roure, A., P. Lemaire and S. Darras (2014). "An otx/nodal regulatory signature for posterior neural development in ascidians." PLoS Genet **10**(8): e1004548.

Ryan, K., Z. Lu and I. A. Meinertzhagen (2016). "The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling." Elife **5**.

Ryan, K. and I. A. Meinertzhagen (2019). "Neuronal identity: the neuron types of a simple chordate sibling, the tadpole larva of Ciona intestinalis." Curr Opin Neurobiol **56**: 47-60.

Rychel, A. L., S. E. Smith, H. T. Shimamoto and B. J. Swalla (2006). "Evolution and development of the chordates: collagen and pharyngeal cartilage." Mol Biol Evol **23**(3): 541-549.

Sabado, V., P. Barraud, C. V. Baker and A. Streit (2012). "Specification of GnRH-1 neurons by antagonistic FGF and retinoic acid signaling." Dev Biol **362**(2): 254-262.

Sabado, V., P. Barraud, C. V. Baker and A. Streit (2012). "Specification of GnRH-1 neurons by antagonistic FGF and retinoic acid signaling." Dev Biol **362**(2): 254-262.

Saint-Jeannet, J. P. and S. A. Moody (2014). "Establishing the pre-placodal region and breaking it into placodes with distinct identities." Dev Biol **389**(1): 13-27.

Sakurai, D., M. Goda, Y. Kohmura, T. Horie, H. Iwamoto, H. Ohtsuki and M. Tsuda (2004). "The role of pigment cells in the brain of ascidian larva." J Comp Neurol **475**(1): 70-82.

Saraiva, L. R. and S. I. Korsching (2007). "A novel olfactory receptor gene family in teleost fish." Genome Res **17**(10): 1448-1457.

Satoh, N. (1994). "Developmental Biology of Ascidians. ." Cambridge University Press, New York.

Satoh, G. (2005). "Characterization of novel GPCR gene coding locus in amphioxus genome: gene structure, expression, and phylogenetic analysis with implications for its involvement in chemoreception." Genesis **41**(2): 47-57.

Satoh, N., D. Rokhsar and T. Nishikawa (2014). "Chordate evolution and the three-phylum system." Proc Biol Sci **281**(1794): 20141729.

Scheffler, K. and C. Seoighe (2005). "A Bayesian model comparison approach to inferring positive selection." Mol Biol Evol **22**(12): 2531-2540.

Schlosser, G. (2005). "Evolutionary origins of vertebrate placodes: insights from developmental studies and from comparisons with other deuterostomes." J Exp Zool B Mol Dev Evol **304**(4): 347-399.

Schlosser, G. (2006). "Induction and specification of cranial placodes." Dev Biol **294**(2): 303-351.

Schlosser, G. (2017). "From so simple a beginning - what amphioxus can teach us about placode evolution." Int J Dev Biol **61**(10-11-12): 633-648.

Schlosser, G., C. Patthey and S. M. Shimeld (2014). "The evolutionary history of vertebrate cranial placodes II. Evolution of ectodermal patterning." Dev Biol **389**(1): 98-119.

Schwanzel-Fukuda, M., Pfaff, D (1989). "Origin of luteinizing hormone-releasing hormone neurons." Nature **338**: 161-164.

Scott, K. (2011). "Out of thin air: sensory detection of oxygen and carbon dioxide." Neuron **69**(2): 194-202.
Scott, W. B. (1887). "Notes on the development of Petromyzon." J Morphol **1**: 253-310.

Seo, H. C., M. Kube, R. B. Edvardsen, M. F. Jensen, A. Beck, E. Spriet, G. Gorsky, E. M. Thompson, H. Lehrach, R. Reinhardt and D. Chourrout (2001). "Miniature genome in the marine chordate Oikopleura dioica." Science **294**(5551): 2506.

Shan, Y., H. Saadi and S. Wray (2020). "Heterogeneous Origin of Gonadotropin Releasing Hormone-1 Neurons in Mouse Embryos Detected by Islet-1/2 Expression." Front Cell Dev Biol **8**: 35.

Sharma, K., A. S. Syed, S. Ferrando, S. Mazan and S. I. Korsching (2019). "The Chemosensory Receptor Repertoire of a True Shark Is Dominated by a Single Olfactory Receptor Family." Genome Biol Evol **11**(2): 398-405.

Sharma, S., W. Wang and A. Stolfi (2019). "Single-cell transcriptome profiling of the Ciona larval brain." Dev Biol **448**(2): 226-236.

Shigetani, Y. S., F.; Kawakami, Y.; Murakami, Y.; Hirano, S.; Kuratani, S. (2002). "Heterotopic Shift of Epithelial-Mesenchymal Interactions in Vertebrate Jaw Evolution." Science **96**: 13-17-1319.

Shimeld, S. M. and P. C. Donoghue (2012). "Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish)." Development **139**(12): 2091-2099.

Shimeld, S. M., A. G. Purkiss, R. P. Dirks, O. A. Bateman, C. Slingsby and N. H. Lubsen (2005). "Urochordate betagamma-crystallin and the evolutionary origin of the vertebrate eye lens." Curr Biol **15**(18): 1684-1689.
Shu, D. G. C. M., S.; Zhang , X. L. (1996). "A Pikaia-like chordate from the Lower Cambrian of China." Nature **384**: 157-158.

Shu, D. G. C. M., S.; Han, J.; Zhang, Z. F.; Yasui, K.; Janvier, P.; Chen, L.; Zhang, X. L.; Liu, J. N.; Li, Y.; Liu, H. Q. (2003). "Head and backbone of the Early Cambrian vertebrate Haikouichthys." Nature **421**(6922): 526-529.

Siefkes, M. J. and W. Li (2004). "Electrophysiological evidence for detection and discrimination of pheromonal bile acids by the olfactory epithelium of female sea lampreys (Petromyzon marinus)." J Comp Physiol A Neuroethol Sens Neural Behav Physiol **190**(3): 193-199.

Silver, M. R., H. Kawauchi, M. Nozaki and S. A. Sower (2004). "Cloning and analysis of the lamprey GnRH-III cDNA from eight species of lamprey representing the three families of Petromyzoniformes." Gen Comp Endocrinol **139**(1): 85-94.

Simakov, O., F. Marletaz, J. X. Yue, B. O'Connell, J. Jenkins, A. Brandt, R. Calef, C. H. Tung, T. K. Huang, J. Schmutz, N. Satoh, J. K. Yu, N. H. Putnam, R. E. Green and D. S. Rokhsar (2020). "Deeply conserved synteny resolves early events in vertebrate evolution." Nat Ecol Evol.

Simmen, M. W. B., A. (2000). "Sequence Analysis of Transposable Elements in the Sea Squirt, Ciona intestinalis." Mol Biol Evol **17**(11): 1685-1694.

Simossis, V. A. and J. Heringa (2005). "PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information." Nucleic Acids Res **33**(Web Server issue): W289-294.

Singh, P. P., J. Arora and H. Isambert (2015). "Identification of Ohnolog Genes Originating from Whole Genome Duplication in Early Vertebrates, Based on Synteny Comparison across Multiple Genomes." PLoS Comput Biol **11**(7): e1004394.

Sjodal, M., T. Edlund and L. Gunhaga (2007). "Time of exposure to BMP signals plays a key role in the specification of the olfactory and lens placodes ex vivo." Dev Cell **13**(1): 141-149.

Smadja, C., P. Shi, R. K. Butlin and H. M. Robertson (2009). "Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, Acyrthosiphon pisum." Mol Biol Evol **26**(9): 2073-2086.

Smith, J. J. K., M. C. (2015). "The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. ." Genome Res. **25**: 1081-1090.

Sotgia, C., U. Fascio, G. Melone and F. De Bernardi (1998). "Adhesive Papillae of Phallusia mamillata Larvae: Morphology and Innervation." Zoolog Sci **15**(3): 363-370.

Sower, S. A. (2015). The Reproductive Hypothalamic-Pituitary Axis in Lampreys. Lampreys: Biology, Conservation and Control**:** 305-373.

Sower, S. A. N., M.; Knox, C. J.; Gorbman, A. (1995). "The occurence and distribution of GnRH in the brain of Atlantic hagfish, an agnatha determined by chromatography and immunocytichemistry." Gen Comp Endocrinol **97**(3): 300-307.

Spilker, C. N., S.; Grochowska, K. M.; Schumacher, A.; Butnaru, I.; Macharadze, T.; Gomes, G. M.; Yuanxiang,P.; Bayraktar, G.; Rodenstein, C.; Geiseler, C.; Kolodziej, A.; Lopez-Rojas, J.; Montag, D.; Angenstein, F.; Bär, J.; D'Hanis, W.; Roskoden, T.; Mikhaylova, M; Budinger, E.; Ohl, F. W.; Stork, O.; Zenclussen, A. C.; Karpova, A.; Herbert Schwegler, H.; Kreutz, M. R. (2016). "A Jacob/Nsmf Gene Knockout Results in Hippocampal Dysplasia and Impaired BDNF Signaling in Dendritogenesis." PLoS Genet **12**(3): 1-32.

Stokes, M. D. and N. D. Holland (1995). "Embryos and Larvae of a Lancelet, Branchiostoma floridae,from Hatching through Metamorphosis: Growth in the Laboratory and External Morphology." Acta Zoologica **76**(2): 105-120.

Suarez, R., D. Garcia-Gonzalez and F. de Castro (2012). "Mutual influences between the main olfactory and vomeronasal systems in development and evolution." Front Neuroanat **6**: 50.

Sugahara, F., S. Aota, S. Kuraku, Y. Murakami, Y. Takio-Ogawa, S. Hirano and S. Kuratani (2011). "Involvement of Hedgehog and FGF signalling in the lamprey telencephalon: evolution of regionalization and dorsoventral patterning of the vertebrate forebrain." Development **138**(6): 1217-1226.

Sugahara, F., J. Pascual-Anaya, Y. Oisi, S. Kuraku, S. Aota, N. Adachi, W. Takagi, T. Hirai, N. Sato, Y. Murakami and S. Kuratani (2016). "Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain." Nature **531**(7592): 97-100.

Suyama, M., D. Torrents and P. Bork (2006). "PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments." Nucleic Acids Res **34**(Web Server issue): W609-612.

Suzuki, K. G., R. L.; Sower, S. A. (2000). "Multiple transcripts encoding lamprey gonadotropin-releasing hormone-I precursors." Journal of Molecular Endocrinology **24**: 365–376.

Swanson, W. J. N., R.; Yang, Q. (2003). "Pervasive Adaptive Evolution in Mammalian Fertilization Proteins." Mol. Biol. Evol. **20**(1): 18-20.

Szewzyk, U. H., C.; Wrangstadh, M.; Samuelsson, M.; Maki, J. S.; Kjellebergfm S. (1991). "Relevance of the exopolysaccharide of marine Pseudomonas sp. strain S9 for the attachment of Ciona intestinalis larvae."
Tahara, Y. (1988). "Normal stages of development in the lamprey, Lampetra reissneri (Dybowski)." Zool Sci **5**: 109-118.

Takahashi, A., M. S. Islam, H. Abe, K. Okubo, Y. Akazome, T. Kaneko, H. Hioki and Y. Oka (2016). "Morphological analysis of the early development of telencephalic and diencephalic gonadotropin-releasing hormone neuronal systems in enhanced green fluorescent protein-expressing transgenic medaka lines." J Comp Neurol **524**(4): 896-913.

Takezaki, N., F. Figueroa, Z. Zaleska-Rutczynska and J. Klein (2003). "Molecular phylogeny of early vertebrates: monophyly of the agnathans as revealed by sequences of 35 genes." Mol Biol Evol **20**(2): 287-292.

Tan, C. K. F. B. F. N., B. F.; Hodson, S. L. (2002). "Biofouling as a reservoir of Neoparamoeba pemaquidensis (Page, 1970), the causative agent of amoebic gill disease in Atlantic salmon." Aquaculture **210**: 49-58.

Tank, E. M., R. G. Dekker, K. Beauchamp, K. A. Wilson, A. E. Boehmke and J. A. Langeland (2009). "Patterns and consequences of vertebrate Emx gene duplications." Evol Dev **11**(4): 343-353.

Taroc, E. Z. M., R. R. Katreddi and P. E. Forni (2020). "Identifying Isl1 Genetic Lineage in the Developing Olfactory System and in GnRH-1 Neurons." Frontiers in Physiology **11**.

Terakado, K. (2001). "Induction of gamete release by gonadotropin-releasing hormone in a protochordate, Ciona intestinalis." Gen Comp Endocrinol **124**(3): 277-284.

Terakado, K. (2009). "Placode formation and generation of gonadotropin-releasing hormone (GnRH) neurons in ascidians." Zoolog Sci **26**(6): 398-405.

Theisen, B. (1973). "The olfactory system in the hagfish Myxine glutinosa." Acta Zoologica **54**: 271-284.
Theisen, B. (1976). "The Olfactory System in the Pacific Hagfishes Eptatretus stoutii, Eptatretus deani, and Myxine circifrons." Acta Zoologica **57**: 167-173.

Thevenaz, P. R., U. E.; Unser, M. (1998). "A Pyramid Approach to Subpixel Registration Based on Intensity." IEEE TRANSACTIONS ON IMAGE PROCESSING **7**(1): 27-41.

Thornhill, R. A. (1967). "The ultrastructure of the olfactory epithelium of the lamprey Lampetra fluviatilis." J. Cell Sci. **2**: 591-602.

Tobet, S. A., T. W. Chickering and S. A. Sower (1996). "Relationship of Gonadotropin-Releasing Hormone (GnRH) Neurons to the Olfactory System in Developing Lamprey (Petromyzon marinus)." The Jounral of Comparative Neurology **376**: 97-111.

Tobet, S. A., M. Nozaki, J. H. Youson and S. A. Sower (1995). "Distribution of lamprey gonadotropin-releasing hormone-III (GnRH-III) in brains of larval lampreys (Petromyzon marinus)." Cell & Tissue Research **279**(2): 261-270.

Toro, S. and Z. M. Varga (2007). "Equivalent progenitor cells in the zebrafish anterior preplacodal field give rise to adenohypophysis, lens, and olfactory placodes." Semin Cell Dev Biol **18**(4): 534-542.

Torrence, S. A. and R. A. Cloney (1982). "Nervous System of Ascidian Larvae : Caudally Primary Sensory Neurons." Zoomorphology **99**: 103-115.

Torrence, S. A. and R. A. Cloney (1983). "Ascidian Larval Nervous System: Primary Sensory Neurons in Adhesive Papillae." Zoomorphology **102**: 11-123.

Tresser, J., S. Chiba, M. Veeman, D. El-Nachef, E. Newman-Smith, T. Horie, M. Tsuda and W. C. Smith (2010). "doublesex/mab3 related-1 (dmrt1) is essential for development of anterior neural plate derivatives in Ciona." Development **137**(13): 2197-2203.

Tsirigos, K. D., C. Peters, N. Shu, L. Kall and A. Elofsson (2015). "The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides." Nucleic Acids Res **43**(W1): W401-407.

Tsuda, M., D. Sakurai and M. Goda (2003). "Direct evidence for the role of pigment cells in the brain of ascidian larvae by laser ablation." J Exp Biol **206**(Pt 8): 1409-1417.

Tsuda, M. O., M.; Nakagawa, M.; Katagirib, Y. (2001). "Light Regulated GnRH Neurons in Biological Clock for Reproduction in the Ascidian, Halocynthia roretzi." The Biology of Ascidians: 131-132.

Tsutsui, H. Y., N.; Ito, H.; Oka, Y. (1998). "GnRH-Immunoreactive Neuronal System in the Presumptive Ancestral Chordate, Ciona intestinalis (Ascidian)." General and Comparative Endocrinology **112**: 426-432.

Ubeda-Banon, I., P. Pro-Sistiaga, A. Mohedano-Moriano, D. Saiz-Sanchez, C. de la Rosa-Prieto, N. Gutierrez-Castellanos, E. Lanuza, F. Martinez-Garcia and A. Martinez-Marcos (2011). "Cladistic analysis of olfactory and vomeronasal systems." Front Neuroanat **5**: 3.

Uchida, K., Y. Murakami, K. S., S. Hirano and S. Kuratani (2003). "Development of the Adenohypophysis in the Lamprey: Evolution of Epigenetic Patterning Programs in Organogenesis." Journal of Experimental Zoology.
Umatani, C. and Y. Oka (2019). "Multiple functions of non-hypophysiotropic gonadotropin releasing hormone neurons in vertebrates." Zoological Lett **5**: 23.

Van Gulick, E. R., T. J. Marquis and S. A. Sower (2018). "Co-localization of three gonadotropin-releasing hormone transcripts in larval, parasitic, and adult sea lamprey brains." Gen Comp Endocrinol **264**: 84-93.

VanDenbosshe, J., J. G. Seelye and B. S. Zielinski (1995). "The morphology of the olfactory epithelium in larval, juvenile and upstream migrant stages of the sea lamprey, Petromyzon marinus." Brain Behav Evol **45**: 19-24.

VanDenbosshe, J., J. H. Youson, D. Pohlman, E. Wong and B. S. Zielinski (1997). "Metamorphosis of the Olfactory Organ of the Sea Lamprey (Petromyzon marinus L.): Morphological Changes and Morphometric Analysis." Journal of Morphology **231**: 41-52.

Vermehren, A., K. K. Langlais and D. B. Morton (2006). "Oxygen-sensitive guanylyl cyclases in insects and their potential roles in oxygen detection and in feeding behaviors." J Insect Physiol **52**(4): 340-348.

Wagner, E. and M. Levine (2012). "FGF signaling establishes the anterior border of the Ciona neural tube." Development **139**(13): 2351-2359.

Wagner, E., A. Stolfi, Y. Gi Choi and M. Levine (2014). "Islet is a key determinant of ascidian palp morphogenesis." Development **141**(15): 3084-3092.

Wakai, M. K., M. J. Nakamura, S. Sawai, K. Hotta and K. Oka (2021). "Two-Round Ca(2+) transient in papillae by mechanical stimulation induces metamorphosis in the ascidian Ciona intestinalis type A." Proc Biol Sci **288**(1945): 20203207.

Watanabe, M., A. Fukuda and J. Nabekura (2014). "The role of GABA in the regulation of GnRH neurons." Front Neurosci **8**: 387.

Weirauch, M. T., A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J. C. Lozano, M. Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F. Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker and T. R. Hughes (2014). "Determination and inference of eukaryotic transcription factor sequence specificity." Cell **158**(6): 1431-1443.

Whitlock, K. E. (2005). "Origin and development of GnRH neurons." Trends Endocrinol Metab **16**(4): 145-151.

Whitlock, K. E., C. D. Wolf and M. L. Boyce (2003). "Gonadotropin-releasing hormone (gnrh) cells arise from cranial neural crest and adenohypophyseal regions of the neural plate in the zebrafish, danio rerio." Developmental Biology **257**(1): 140-152.

Wierman, M. E., K. Kiseljak-Vassiliades and S. Tobet (2011). "Gonadotropin-releasing hormone (GnRH) neuron migration: initiation, maintenance and cessation as critical steps to ensure normal reproductive function." Front Neuroendocrinol **32**(1): 43-52.

Winkler, C., U. Hornung, M. Kondo, C. Neuner, J. Duschl, A. Shima and M. Schartl (2004). "Developmentally regulated and non-sex-specific expression of autosomal dmrt genes in embryos of the Medaka fish (Oryzias latipes)." Mech Dev **121**(7-8): 997-1005.

Wirsig-Wiechmann, C. R. (2001). "Function of gonadotropin-releasing hormone in olfaction " Keio J Med **50**(2): 81-85.

Wray, S. (2002). "Molecular Mechanisms for Migration of Placodally Derived GnRH Neurons." Chem. Senses **27**: 569-572.

Wray, S. G., P.; Gainer, H. (1989). "Evidence that cells expressing luteinizing hormone-releasing hormone mRNA in the mouse are derived from progenitor cells in the olfactory placode (prenatal development/in situ hybridization/histochenustry/immunocytochemistry/[3H]thymidine autoradiography)." Proc. Natl. Acad. Sci. USA **86**: 8132-8136.

Wright, G. M., K. M. McBurney, J. H. Youson and S. A. Sower (1994). "Distribution of lamprey gonadotropin-releasing hormone in the brain and pituitary gland of larval, metamorphic, and adult sea lampreys, Petromyzon marinus." Canadian Journal of Zoology **72**(1): 48-53.

Xia, W., O. Smith, N. Zmora, S. Xu and Y. Zohar (2014). "Comprehensive analysis of GnRH2 neuronal projections in zebrafish." Sci Rep **4**: 3676.

Xia, Y. W., G. M. (1998). "Soft Lithography." Angew. Chem. Int. Ed. **37**: 550-575.

Xu, H., M. S. Williams and L. M. Spain (2006). "Patterns of expression, membrane localization, and effects of ectopic expression suggest a function for MS4a4B, a CD20 homolog in Th1 T cells." Blood **107**(6): 2400-2408.

Xu, N., B. Bhagavath, H. G. Kim, L. Halvorson, R. S. Podolsky, L. P. Chorich, P. Prasad, W. C. Xiong, R. S. Cameron and L. C. Layman (2010). "NELF is a nuclear protein involved in hypothalamic GnRH neuronal migration." Mol Cell Endocrinol **319**(1-2): 47-55.

Yagi, K. and K. W. Makabe (2002). "Retinoic acid differently affects the formation of palps and surrounding neurons in the ascidian tadpole." Dev Genes Evol **212**(6): 288-292.

Yamamoto, N. U., H.; Ohki-Hamazakic, H.; Hideaki Tanaka, H.; Ito, H. (1996). "Migration of GnRH-immunoreactive neurons from the olfactory placode to the brain: a study using avian embryonic chimeras." Developmenta Brain Research **95**: 234-244.

Yang, L., H. Jiang, Y. Wang, Y. Lei, J. Chen, N. Sun, W. Lv, C. Wang, T. J. Near and S. He (2019). "Expansion of vomeronasal receptor genes (OlfC) in the evolution of fright reaction in Ostariophysan fishes." Commun Biol **2**: 235.

Yang, Z. a. S., W. J. (2002). "Codon-Substitution Models to Detect Adaptive Evolution that Account for Heterogeneous Selective Pressures Among Site Classes." Mol. Biol. Evol. **19**(1): 49-57.

Yi, G., S. H. Sze and M. R. Thon (2007). "Identifying clusters of functionally related genes in genomes." Bioinformatics **23**(9): 1053-1060.

Yokoyama, T. D., K. Hotta and K. Oka (2014). "Comprehensive morphological analysis of individual peripheral neuron dendritic arbors in ascidian larvae using the photoconvertible protein Kaede." Dev Dyn **243**(10): 1362-1373.

Yongqiang, F. W., H.; Lei, C. (1999). "Distribution of gonadotropin-releasing hormone in the brain and Hatschek's pit of amphioxus (Branchiostoma belcheri)." Acta Zool Sinica **45**: 106-111.

Yoshida, M., M. Murata, K. Inaba and M. Morisawa (2002). "A chemoattractant for ascidian spermatozoa is a sulfated steroid." Proc Natl Acad Sci U S A **99**(23): 14831-14836.

Youson, J. H., J. A. Heinig, S. F. Khanam, S. A. Sower, H. Kawauchi and F. W. Keeley (2006). "Patterns of proopiomelanotropin and proopiocortin gene expression and of immunohistochemistry for gonadotropin-releasing hormones (lGnRH-I and III) during the life cycle of a nonparasitic lamprey: relationship to this adult life history type." Gen Comp Endocrinol **148**(1): 54-71.

Zega, G., M. C. Thorndyke and E. R. Brown (2006). "Development of swimming behaviour in the larva of the ascidian Ciona intestinalis." J Exp Biol **209**(Pt 17): 3405-3412.

Zeng, F., C. B. Rohde and M. F. Yanik (2008). "Sub-cellular precision on-chip small-animal immobilization, multi-photon imaging and femtosecond-laser manipulation." Lab Chip **8**(5): 653-656.

Zeng, F., J. Wunderer, W. Salvenmoser, T. Ederth and U. Rothbacher (2019b). "Identifying adhesive components in a model tunicate." Philos Trans R Soc Lond B Biol Sci **374**(1784): 20190197.

Zeng, F., J. Wunderer, W. Salvenmoser, M. W. Hess, P. Ladurner and U. Rothbacher (2019a). "Papillae revisited and the nature of the adhesive secreting collocytes." Dev Biol **448**(2): 183-198.

Zhao, X., F. Xu, L. Tang, W. Du, X. Feng and B. F. Liu (2013). "Microfluidic chip-based C. elegans microinjection system for investigating cell-cell communication in vivo." Biosens Bioelectron **50**: 28-34.

Zhao, Y., M. C. Lin, M. Farajzadeh and N. L. Wayne (2013). "Early development of the gonadotropin-releasing hormone neuronal network in transgenic zebrafish." Front Endocrinol (Lausanne) **4**: 107.

Zhong, Y., C. Herrera-Ubeda, J. Garcia-Fernandez, G. Li and P. W. H. Holland (2020). "Mutation of amphioxus Pdx and Cdx demonstrates conserved roles for ParaHox genes in gut, anus and tail patterning." BMC Biol **18**(1): 68.

Zieger, E., G. Garbarino, N. S. M. Robert, J. K. Yu, J. C. Croce, S. Candiani and M. Schubert (2018). "Retinoic acid signaling and neurogenic niche regulation in the developing peripheral nervous system of the cephalochordate amphioxus." Cell Mol Life Sci **75**(13): 2407-2429.

Zielinski, B. S., J. K. Osahan, T. J. Hara, M. Hosseini and E. Wong (1996). "Nitric Oxide Synthase in the Olfactory Mucosa of the Larval Sea Lamprey (Petromyzon marinus)." The Jounral of Comparative Neurology **365**: 18-26.

Ziermann, J. M. (2019). Cranium, Cephalic Muscles, and Homologies in Cyclostomes. Heads, Jaws, and Muscles**:** 45-63.

Zucchetti, I., R. Marino, M. R. Pinto, J. D. Lambris, L. Du Pasquier and R. De Santis (2008). "ciCD94-1, an ascidian multipurpose C-type lectin-like receptor expressed in Ciona intestinalis hemocytes and larval neural structures." Differentiation **76**(3): 267-282.

Zuccolo, J., J. Bau, S. J. Childs, G. G. Goss, C. W. Sensen and J. P. Deans (2010). "Phylogenetic analysis of the MS4A and TMEM176 gene families." PLoS One **5**(2): e9369.

Zuccolo, J., L. Deng, T. L. Unruh, R. Sanyal, J. A. Bau, J. Storek, D. J. Demetrick, J. M. Luider, I. A. Auer-Grzesiak, A. Mansoor and J. P. Deans (2013). "Expression of MS4A and TMEM176 Genes in Human B Lymphocytes." Front Immunol **4**: 195.

The Evolution and Development of the Olfactory Neurons in the Olfactores clade

Guillaume Poncelet
Linacre College
Michaelmas 2021

Appendix

Chapter 2 Supplementary material

S2.1 Transcript sequences mapping to the identified lamprey loci

S2.1.1 GnRH-I

> ENSPMAT00000000280
ATGGCACTGCGCGGTCAGAGTCTGACACTCCTGCTGCTGGCGACGGCTCTGCTCGTCTCCTTGAATTAC
GCGCAGCACTACTCCCTGGAATGGAAACCCGGGGGCAAACGAGACCTGGAGGTCAGTCACACACGAGA
GCTGGAGCAGGAGCTTGAGCCGCCGAGCAACGCGTTCGAGTGCGACGGACCCGAATGCGCCTTCTCTC
GAGTGCCCAACACCAAGCTGATCAGGGAGCTCGCGAGCTACCTCTCGCAGAGGAATTATGATCGGAAA
GGAGCTCTGAAGTAAGACCCCACGCATCTCAGGGTGCAGATGAAGACGACCAAGGCTCCGACACTTCG
CAACATTCCTCTGAAGCACGATGTTTAGAAACCTTTCAATGAACGCCCTTTCACCTTCGTCTTAACCAC
GTGTTATTGCCCGTGACGTCTCTTCGACACGTTGTGTCCTTAATTACATTCATCGCCGCTCAGCGTATC
GTTAACATGATTGTGACCGTACACATTTTATTTGTGCAGTACATAGTACTATGTTTCAGAGTGGCTGT
CCTGTCTGTGGGTGGGATACACTGTGCTATGCGTGGCCATGTATAACAAGGAGGTGTTGTGCAT
AAATAAAATTGTT

S2.1.2 GNRH-II

> TRINITY_DN86187_c0_g1_i1
GGCTGCTGATGGCCCCGCCGGCCTCGCTGGGGCAGCATTGGTCGCACGGTTGGTTCCCCGGGGG
CAAACGGAGCGTTCAGGAGCCGCCCCGACCCTCCTACGAGAGCGTCAGTCCGTCGGATGCTCCCCCCTT
CACTCCTGTGCTCAGCAGAGCAGGCGCCGAATGCCCTGTCATCGCTGACACCGACTACCCACGCATACA
CCTTTTGCGCGACGTGCTGGTGGAGTTAATGGAAAGAAATTCACAGAAGAGAATATATCATCGGAAT
GAAATGCGAAGAAATGTTGGTTAGGAAACGGATATCCATGGCCAGGACGTATTGCGAGCTTGTATTCG
TGCATATAGTTGCTACTTAATTTGATTTCTCACAGCTTTATATGCTGTAATATATAACGTCTCACAAA
CGTACGCATAATTCTACGTAAAATTATACCTGATTAAACTTCGTTGTACCTGTATAGTTAAACTGTGT
AAGCCATATTCCACGTTGTATACTAAACATGATCCGTCACTCCCGT

> ENSPMAT00000011357
ATGAATGAACGAATCGATTGCTGTCTGTCCCTCGCGCTGTGGAGGATGGGGCGCGCGTCGCTGTCCCT
GGTGCTGATCCTGTGGCTGCTGACGGCCCCGCCAGCCTCGCTGGGGCAGCATTGGTCGCACGGTTGGTT
CCCTGGAGGCAAGCGAGGCGTTCAGGAGCCGCCCCGAGCCTCCTACGAGAACGTCAGCCCGTCGGACGG
TTCCCCTTTCACTCCTGTGAGCTCAGGTCTTCAGGTCGCGGACTGGCATGTTGTTTGTTCACGCTCTCC
AAATGGGTTTTCGGGTTGTGCCATGTGTTGTTCGCCAGGATGTCCGACGTTTCTCTCCCAAGTGCTGG
AGGCTTCTTTGGGAACTTAA

S2.1.3 GNRH-III

> ENSPMAT00000000282
CAACTACCGAAACAGATTCCTCTCCGAGCTCGTCTTCGCGCGGTGGTTTATTTTCTCAACAGACCGTCT
GGAATCATCACAGAAGCCACACTCGGCTGCTGTAGAGATGGCACTGCGCGGTCAAAGCCTGGTTCTGC
TGCTGCTGGCGTCGGCGCTGCTGGTGTCGCTGACGCACACACAGCACTGGTCCCACGACTGGAAACCCG
GAGGGAAACGCGACCTGGAGGCCATGAGACCACTGCTGGAGCAGGAGCTTGAGGCGCCGAACAGCGCG
TTCGAATGCGACGGACCCGAATGCGCCTTCGCTCGAGTGCCGACTGGTGAGCTCGTCAGGGAGATCGT
GAGTTACCTCTCGCAGAAGAATTATCAAAGGAAAGTTCTGAAGTAAAAGCCCCGCGTCTCAAGCTGCA
GATGAAGACAATCAACGCTCCCGACAATTCGAGAAGTTCCTCGAGAGCGCCACGTGACACGAACCCTG
TCAATGAAATGCCCTCGCTGTGGTCTGTGACGTCTCTTAGACCCTTTGTGTTTATTTAATTCTTCATC
GCCGCTCAGCGTATCGTCAACATGACCGTGGCCATATGCGTTTATGTGCACCATACATAGTACCGTGT
TCCAGAGTGGCTGTGTGTAGGTGGCGTACACTTTGTTATTATTGGTCACGAACAGGCCA
CGCGTAACAACAAGGCGTTGTCCGTGAATAAAGTTGTTATAAGCG

S2.1.4 ISLA

> TRINITY_DN41193_c0_g1_i2
GCCCCACCGCCGCCGCCGCTGCCGTCGCTGCCGCTGCTGCTGCTGCTGCTCAGGCCTCGATGGGGCTGG
AGCCGATGCTGGTGGGAGTGTCGGGCAGCTGCGAGAGCGACGTGATCTCACTGCCCGTCGAGTTGGAG
CCCGGCCGCCCCTCCGAGAAGTTCACCAGCTGCTGGAAGGGCCCCTGGTCGAGCTCGCTGCTGAGCGCG
AAGTCGCTGAGCGCCTTCCAGGGCGGCTGGTGGAAGGTCTGCACGTCGACGGCGCCCACGGGAAGGCC
CGGCCCGTCCTGCCGCACGGGGCTGCCGGCCACCAGCGGCGTGCCCGTCAGGCCTTGCAGGTTCGTCTT
GTCGCCTCCCTGCTGCTGGATCTGCTTCATGAGGATGCTGCGCTTCTTGTCCTTGCAGCGCTTGTTCTG
GAACCAGACGCGGATGACGCGGGGACTCAGGCCCGTCATCTCCACGAGCTGCTCCTTCATGAGCGCGTC
GGGCCGCGGGTTGGCCGCGTAGCAGGTTCGCAGCGTGTGCAGCTGCTTCTCGTTCAGCACCGTGCGCAC
GCGCGTCGTCTTCTCCGCCTGCTTGTGCACGTGCGGGCGCAGTGGGTGGTGGCGGCCAGCCGCGGACAT
GGGCTCGGCGGCGAGCTGCAGGGGCCGGCCGCCCCCCTGCCCCGGGCTGCCCATGCCGTCGGCGGTGCC
GGCCCTCTCCAGGAGCTCGTGATCGGCGCGGCAGAAGAGGCCGTCCTCCCGCAGCGCAAACTCGTCACC
CGGGATGAGCTGGCGGCTGCACGCCACGCAGCGGAAGCACTCGAGGTGGAACACCTGGGTCCGCACCC
TCATCACGAAGTCCGTCTTGCTGAAGCCCACGCCACACTTAGCGCACTTGATGCCAAACAGCCTGCGGG
CGCCCACAGGGAGAGGGCAGCGAGAGGGAGGGGGAGAGACGGAGGAGGGAGAGGGATGGAGAGAG

> TRINITY_DN41193_c0_g1_i1
GCCCCACCGCCGCCGCCGCTGCCGTCGCTGCCGCTGCTGCTGCTGCTGCTCAGGCCTCGATGGGGCTGG
AGCCGATGCTGGTGGGAGTGTCGGGCAGCTGCGAGAGCGACGTGATCTCACTGCCCGTCGAGTTGGAG
CCCGGCCGCCCCTCCGAGAAGTTCACCAGCTGCTGGAAGGGCCCCTGGTCGAGCTCGCTGCTGAGCGCG
AAGTCGCTGAGCGCCTTCCAGGGCGGCTGGTGGAAGGTCTGCACGTCGACGGCGCCCACGGGAAGGCC
CGGCCCGTCCTGCCGCACGGGGCTGCCGGCCACCAGCGGCGTGCCCGTCAGGCCTTGCAGGTTCGTCTT
GTCGCCTCCCTGCTGCTGGATCTGCTTCATGAGGATGCTGCGCTTCTTGTCCTTGCAGCGCTTGTTCTG
GAACCAGACGCGGATGACGCGGGGACTCAGGCCCGTCATCTCCACGAGCTGCTCCTTCATGAGCGCGTC
GGGCCGCGGGTTGGCCGCGTAGCAGGTTCGCAGCGTGTGCAGCTGCTTCTCGTTCAGCACCGTGCGCAC
GCGCGTCGTCTTCTCCGCCTGCTTGTGCACGTGCGGGCGCAGTGGGTGGTGGCGGCCAGCCGCGGACAT
GGGCTCGGCGGCGAGCTGCAGGGGCCGGCCGCCCCCCTGCCCCGGGCTGCCCATGCCGTCGGCGGTGCC
GGCCCTCTCCAGGAGCTCGTGATCGGCGCGGCAGAAGAGGCCGTCCTCCCGCAGCGCAAACTCGTCACC
CGGGATGAGCTGGCGGCTGCACGCCACGCAGCGGAAGCACTCGAGGTGGAACACCTGGGTCCGCACCC
TCATCACGAAGTCCGTCTTGCTGAAGCCCACGCCACACTTAGCGCACTTGATGCCAAACAGCCTGACGT
AGTCCCGCTTGCAGTAGGTCTTGCCGTCGCGCACGAAGCATGTGCACGTCTCGTCCAGGTACTGCGCGC
ACTCGGCGCACTTGAGGCACGAGGCGTGCCACTCGAGGTCGGGGGATACGCGCAGGATGAAGGGGTCT
CGGATCTGGGCGCCGCAGCCAACGCACACGGACGCCAGCCGCTTTTTTGGAGTTGTCTCCCATATCACC
CATGTCCTCGTCTTCTTCGCCGTGTTTCTTCTCGCCACGGCCAACGAGCTCTCCGCCTCGATCTCCTCT
GCCGAGGAGGAGGAGGAGGAAGAGGCGGCGGCGGCGGTGATGCTGGTGATGGAGAGTCCTCCT

> Locus_139360_Transcript_1/1_Confidence_1.000_Length_586
CGGGGGCGCTGCTCTGCCGCACCGCCGCTGCTGCTGCTGCTCAGGCCTCGATGGGGCTGGAGCCGA
TGCTGGTGGGAGTGTCGGGCAGCTGCGACAGCGACGTGATCTCACTGCCCGTCGAGTTGGAACCCGGC
CGTCCCTCCGAGAAGTTCACCAGCTGCTGGAAGGGCCCCTGGTCGAGCTCGCTGCTGAGCGCAAAGTCG
CTGAGAGCCTTCCAGGGCGGCTGGTGGAACGTCTGCACGTCGACAGCGCCCACCGGAAGTCCCGGCCCG
TCTTGCCGCACGGGGCTCCCGGCCACCAGCGGCGTCCCCGTCAGGCCCTGCAGGTTCGTTTTGTCTCCG
CCCTGCTGCTGGATCTGCTTCATGAGGATGCTCCGCTTCTTGTCCTTGCAGCGCTTGTTCTGGAACCAG
ACGCGGATGACGCGGGGGCTCAGGCCCGTCATCTCCACGAGCTGCTCCTTCATGAGCGCGTCGGGGCGC
GGGTTGGCCGCGTAGCAGGTGCGCAGCGTGTGCAGCTGCTTCTCGTTGAGCACCGTGCGCACGCGCGTC
GTCTTCTCCGCCTGCTTGTGCACGTGCGGGCGCAG

> comp151224_c2_seq2

CTGGAGTCGTGGAGGCTGGGCGCGCTGTCGCATTGTGGCTCAGCTCTCTCGCTGCTTGCCCCTGCCGCT
GCTTCTGCCTCTCGTCCCCCCTCTTCAGCCTCTCGCGCTGAGAGTCCCGTGCGGGCGCTCTTCTCTCTCG
CCTCTCGCTCGCTCTTTTCACAAGGCACACAGCGAGAGCGAGAGACACGCACGCGCGCGCGCACGCAGA
GAGATCGAGAGTACGAGAGAGAGAGTTAAAGAAGGGGGTTGTTGCATCTGTGGCTGCTGAGCGGAAC
GGCTGTGGTGGGCGGCGGGGGGGGTGGTGCCGAGACGGAGTCGATTCGTGAAGGGAAATGCGGCAGCG
GTGAGTCAGCCGGCTGCTCGCCTAACCACATACCCACACCTACAGACCATCCGACCCAGCCGCCTACCA
ACCAACGCACCTAACCTCCTGCATCCACCGACCCACCCAGCTACCTACCTACCGACCAACTCAACCCAT
CTTCCGGCAGATACAACATTCAACAACAACAGCCACTCACAACAACCTCCACCACCAGCAGCAACAAC
AGTCGCACCCCCGCCCGCCGCCACCCCGTGGCGCGGTGGCTCCAGAGATGCATGGCAGGGTGCACTGCG
CCTGGCACGCGGCTCGCCTCATCGTCGCGTGAAGGAGGACTCCCCATCACCAGCATCACCACCGCCGCC
TCTTCCTCCTCCTCCTCATCGGCCGAGGAGATCGAGGCGGAGGCCTCGTTGGCCGTGGCGAGAAGAAAC
ACGGCGAAGAAGACGAGGACATGGGTGATATGGGAGACAACTCCAAAAAGAAGCGGCTGGCGTCCGT
GTGCGTTGGCTGCGGCGCCCAGATCCGCGACCCCTTCATCCTGCGCGTGTCCCCCGACCTCGAGTGGCA
CGCATCGTGCCTCAAGTGCGCCGAGTGCGCGCAGTACCTGGACGAGACGTGCACATGCTTCGTGCGGG
ACGGCAAGACCTACTGCAAGCGCGACTACGTCAGGCTGTTCGGCATCAAGTGCGCTAAGTGTGGCGTG
GGCTTCAGCAAGACGGACTTCGTGATGAGGGTGCGAACCCAGGTGTTCCACCTCGAGTGCTTCCGCTG
CGTGGCGTGCAGCCGACAGCTCATCCCGGGGGACGAGTTTGCTCTGCGGGAGGACGGCCTCTTCTGCCG
CGCCGACCACGAGCTCCTGGAGAGGGCCGGCACCGCCGACGGCATGGGCAGCCCGGGACAGGGGGGCG
GCCGGCCCCTTCAGCTCGCCGCCGAGCCCATGTCTGCGGCGGGGCGCCACCACCCGCTGCGCCCGCACG
TGCACAAGCAGGCGGAGAAGACGACGCGCGTGCGCACGGTGCTCAACGAGAAGCAGCTGCACACGCTG
CGCACCTGCTACGCGGCCAACCCGCGCCCCGACGCGCTCATGAAGGAGCAGCTCGTGGAGATGACGGGC
CTGAGCCCCCGCGTCATCCGCGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGGAGCATCCT
CATGAAGCAGATCCAGCAGCAGGGCGGAGACAAAACGAACCTGCAGGGCCTGACGGGGACGCCGCTGG
TGGCCGGGAGCCCCGTGCGGCAAGACGGGCCGGGACTTCCGGTGGGCGCTGTCGACGTGCAGACGTTC
CACCAGCCGCCCTGGAAGGCTCTCAGCGACTTTGCGCTCAGCAGCGAGCTCGACCAGGGGCCCTTCCAG
CAGCTGCAGGTGAACTTCTCGGAGGGACGGCCGGGTTCCAACTCGACGGGCAGTGAGATCACGTCGCT
GTCGCAGCTGCCCGACACTCCCACCAGCATCGGCTCCAGCCCCATCGAGGCCTG

> comp166395_c3_seq4
ATCCCTGCCGTCTCTGCCGTCCCGGGCGGGGGGGGCCGGGCCGCTGGACGGGGCCCCTCTGGACGATCT
CGACGGGCTGCCCATGCAGGCCGGGGACGATATCGACGGGCTGCCCCTGGACGATGCTCTCCCCAAGAA
GCCCGTTCTGCCCATCGCGTCGTCCAAGTGGGAGCGCGTGGACGACCTGGACACGGACACCACCCGCCG
CTCCCCCGCTCCGTCGGGCAGCAAGAGGGACGACGACTCGGACGATGCCAGCGGCGGCAATGTTGCCGC
CGCGGGCGGCAAGTTGCCGCCGGGCTTTGCCGCCACGGCCGCCGCCTCTTCCTCCTCCTCCTCGGCAGA
GGAGATCGAGGCGGAGAGCTCGTTGGCCGTGGCGAGAAGAAACACGGCGAAGAAGACGAGGACATGG
GTGATATGGGAGACAACTCCAAAAAGAAGCGGCTGGCGTCCGTGTGCGTTGGCTGCGGCGCCCAGATC
CGAGACCCCTTCATCCTGCGCGTGTCGCCCGACCTCGAGTGGCACGCCTCGTGCCTCAAGTGCGCCGAG
TGCGCGCAGTACCTGGACGAGACGTGCACATGCTTCGTGCGCGACGGCAAGACCTACTGCAAGCGGGA
CTACGTCAGGCTGTTTGGCATCAAGTGCGCTAAGTGTGGCGTGGGCTTCAGCAAGACGGACTTCGTGA
TGAGGGTGCGGACCCAGGTGTTCCACCTCGAGTGCTTCCGCTGCGTGGCGTGCAGCCGCCAGCTCATCC
CGGGTGACGAGTTTGCGCTGCGGGAGGACGGCCTCTTCTGCCGCGCCGACCACGAGCTCCTGGAGAGG
GCCGGCACCGCCGACGGCATGGGCAGCCCGGGGCAGGGGGGCGGCCGGCCCCTGCAGCTCGCCGCCGAG
CCCATGTCCGCGGCTGGCCGCCACCACCCACTGCGCCCGCACGTGCACAAGCAGGCGGAGAAGACAACG
CGCGTGCGCACGGTGCTGAACGAGAAGCAGCTGCACACGCTGCGCACGTGCTACGCGGCCAACCCGCGG
CCCGACGCGCTCATGAAGGAACAGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCATCCGCGTCTGG
TTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCAGCATCCTCATGAAGCAGATCCAGCAGCAGGGCGG
TGACAAGACGAACCTGCAAGGCCTGACGGGCACGCCGCTGGTGGCCGGGAGCCCCGTGCGGCAGGACG
GGCCGGGCCTCCCCGTGGGCGCCGTCGACGTGCAGACCTTCCACCAGCCGCCCTGGAAGGCGCTCAGCG
ACTTCGCGCTCAGCAGCGAGCTCGACCAGGGGCCCTTCCAGCAGCTGCAGGTGAACTTCTCGGAGGGGC
GGCCGGGCTCCAACTCGACGGGCAGTGAGATCACGTCGCTCTCGCAGCTGCCCGACACTCCCACCAGCA
TCGGCTCCAGCCCCATCGAGGC

> comp166395_c3_seq14
CAAGAAGATTGTTGTTGTTGTTGTCTTTATTGTCGTTGTTGTTGTTGCTGTAGTGGACGATGCTCTCC
CCAAGAAGCCCGTTCTGCCCATCGCGTCGTCCAAGTGGGAGCGCGTGGACGACCTGGACACGGACACC
ACCCGCCGCTCCCCCGCTCCGTCGGGCAGCAAGAGGGACGACGACTCGGACGATGCCAGCGGCGGCAAT
GTTGCCGCCGCGGGCGGCAAGTTGCCGCCGGGCTTTGCCGCCACGGCCGCCGCCTCTTCCTCCTCCTCCT
CGGCAGAGGAGATCGAGGCGGAGAGCTCGTTGGCCGTGGCGAGAAGAAACACGGCGAAGAAGACGAG
GACATGGGTGATATGGGAGACAACTCCAAAAAGAAGCGGCTGGCGTCCGTGTGCGTTGGCTGCGGCGC
CCAGATCCGAGACCCCTTCATCCTGCGCGTGTCGCCCGACCTCGAGTGGCACGCCTCGTGCCTCAAGTG
CGCCGAGTGCGCGCAGTACCTGGACGAGACGTGCACATGCTTCGTGCGCGACGGCAAGACCTACTGCA
AGCGGGACTACGTCAGGCTGTTTGGCATCAAGTGCGCTAAGTGTGGCGTGGGCTTCAGCAAGACGGAC
TTCGTGATGAGGGTGCGGACCCAGGTGTTCCACCTCGAGTGCTTCCGCTGCGTGGCGTGCAGCCGCCAG
CTCATCCCGGGTGACGAGTTTGCGCTGCGGGAGGACGGCCTCTTCTGCCGCGCCGACCACGAGCTCCTG
GAGAGGGCCGGCACCGCCGACGGCATGGGCAGCCCGGGGCAGGGGGGCGGCCGGCCCCTGCAGCTCGCC
GCCGAGCCCATGTCCGCGGCTGGCCGCCACCACCCACTGCGCCCGCACGTGCACAAGCAGGCGGAGAAG
ACAACGCGCGTGCGCACGGTGCTGAACGAGAAGCAGCTGCACACGCTGCGCACGTGCTACGCGGCCAA
CCCGCGGCCCGACGCGCTCATGAAGGAACAGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCATCCG
CGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCAGCATCCTCATGAAGCAGATCCAGCAGC
AGGGCGGTGACAAGACGAACCTGCAAGGCCTGACGGGCACGCCGCTGGTGGCCGGGAGCCCCGTGCGG
CAGGACGGGCCGGGCCTCCCCGTGGGCGCCGTCGACGTGCAGACCTTCCACCAGCCGCCCTGGAAGGCG
CTCAGCGACTTCGCGCTCAGCAGCGAGCTCGACCAGGGGCCCTTCCAGCAGCTGCAGGTGAACTTCTCG
GAGGGGCGGCCGGGCTCCAACTCGACGGGCAGTGAGATCACGTCGCTCTCGCAGCTGCCCGACACTCCC
ACCAGCATCGGCTCCAGCCCCATCGAGGC

> comp151224_c2_seq1
TGATATGGGAGACAACTCCAAAAGCTGTTCGGCATCAAGTGCGCTAAGTGTGGCGTGGGCTTCAGCAA
GACGGACTTCGTGATGAGGGTGCGAACCCAGGTGTTCCACCTCGAGTGCTTCCGCTGCGTGGCGTGCA
GCCGACAGCTCATCCCGGGGGACGAGTTTGCTCTGCGGGAGGACGGCCTCTTCTGCCGCGCCGACCACG
AGCTCCTGGAGAGGGCCGGCACCGCCGACGGCATGGGCAGCCCGGGACAGGGGGGCGGCCGGCCCCTTC
AGCTCGCCGCCGAGCCCATGTCTGCGGCGGGGCGCCACCACCCGCTGCGCCCGCACGTGCACAAGCAGG
CGGAGAAGACGACGCGCGTGCGCACGGTGCTCAACGAGAAGCAGCTGCACACGCTGCGCACCTGCTAC
GCGGCCAACCCGCGCCCCGACGCGCTCATGAAGGAGCAGCTCGTGGAGATGACGGGCCTGAGCCCCCGC
GTCATCCGCGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGGAGCATCCTCATGAAGCAGAT
CCAGCAGCAGGGCGGAGACAAAACGAACCTGCAGGGCCTGACGGGGACGCCGCTGGTGGCCGGGAGCC
CCGTGCGGCAAGACGGGCCGGGACTTCCGGTGGGCGCTGTCGACGTGCAGACGTTCCACCAGCCGCCCT
GGAAGGCTCTCAGCGACTTTGCGCTCAGCAGCGAGCTCGACCAGGGGCCCTTCCAGCAGCTGGTGAAC
TTCTCGGAGGGGCGGCCGGGTTCCAACTCGACGGGCAGTGAGATCACGTCGCTGTCGCAGCTGCCCGAC
ACTCCCACCAGCATCGGCTCCAGCCCCATCGAGGCCTG

> comp166395_c3_seq1
CACCACCACCACCATCACCAGGAGCGCCCCGTTGGCGTGCCGTCGCTGAGGCCCGACAGGGCTCACAGG
GGCCCCGGGGAGAAGAGGGGACCTGTGCCACGAGCCTCGTGCGGGGACCTGCACATGTCCGGCTCGGG
CAAGGCGAGCCGGGTCCGCACCGTCCTGACGGAGCGACAGCTGCAGACGCTGCGCACGTGCTACGCGGC
CAACCCACGGCCCGACGCCCTCCTGAAGGAGCAGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCAT
CCGCGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCAGCATCCTCATGAAGCAGATCCAGC
AGCAGGGCGGTGACAAGACGAACCTGCAAGGCCTGACGGGCACGCCGCTGGTGGCCGGGAGCCCCGTG
CGGCAGGACGGGCCGGGCCTCCCCGTGGGCGCCGTCGACGTGCAGACCTTCCACCAGCCGCCCTGGAAG
GCGCTCAGCGACTTCGCGCTCAGCAGCGAGCTCGACCAGGGGCCCTTCCAGCAGCTGGTGAACTTCTCG
GAGGGGCGGCCGGGCTCCAACTCGACGGGCAGTGAGATCACGTCGCTCTCGCAGCTGCCCGACACTCCC
ACCAGCATCGGCTCCAGCCCCATCGAGGC

S2.1.5 ISLB

> TRINITY_DN41193_c2_g1_i1

GCGCAGCATCCTCATGAAGCAGATGCAGCAGCAGCATCACGCCGACAAGACGAACCTGCAGGGCCTGA
CGGGCACACCGCTGGTGGCCGGGAGCCCTGTGCCACACGAGGCGGCGATACCGGGGAACCCCGTCGAG
GTGCAGAGTTTCCAGCCAGCGCCGTGGAAAACACTCAGCGAGTTCGCGCTGAGCAGCGACCTCGACCA
GGGCCCGTTCCAGCACCTGATGGGGTTCTCTGACGGGGGACCCGGATCGAACTCCACCGCAGGGAGCG
ACCTGGGCTCCATTCCCTCCCAGCTCCCGGACACCCCAAGCAGCATGGTGGCCAGCCCCATCGAAGCCT
AAAACAACAACCCAACATGGCCGCGAGAACCATGACCACCTCTACGTTGACTCCGAGCCTCTTCCGCTC
CAAGCCCAGCAGCTAAGCGAGCCTCCTTTAAAGCGCTTCTGCTTAACG

> ENSPMAT00000002189
AGGCTGTTCGGGATCAAGTGCGCCAAGTGCAACGCGTCGTTCAGCAAGACGGACTTTGTGATGCGCGC
ACGCTCCCACATCTACCACATGGAGTGCTTCCGCTGCGAGACCTGCTCGCGCCAGCTGCTGCCCGGCGA
CGAGTTCGCGCTGCGCGACGGCGGCGGACTCCTGTGTCGAGCCGACCAGCACCACGGCGGCGGTGGCGG
CGGTGGCGCCGGTGGCGCCGGAGGCGGCGCCGGAGGCGGCGGCGGCGGGCTGCTCGGAACGGCGCTGA
CGCGACCGGGCAGTCCGGTGCACTCGTCGAGAAGTCTGCACCTGTCGGGGGGGATTCCGAGCGGGCCG
GAGAAGGCGACGCGGGTTCGCACGGTGCTCAACGAGAAGCAGCTGCAGACGCTGCGCACGTGCTACGC
GGCAAACCCGCGGCCCGACGCGCTCATGAAGGAGCAGCTCGTGGAGATGACGGGCCTGAGCCCGCGCG
TCATCCGCGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCAGCATCCTCATGAAGCAGATG
CAGCAGCAGCATCACGCCGACAAGACGAACCTGCAGGGCCTGACGGGGACGCCGCTGGTGGCCGGGAG
CCCCGTGCCACACGAGGCGGCGATACCGGGGAACCCCGTCGAGGTGCAGAGTTTCCAGCCAGCGCCGTG
GAAAACGCTCAGCGAGTTCGCGCTGAGCAGCGACCTCGACCAGGGCCCGTTCCAGCACCTGATGGGGT
TCTCTGACGGGGGACCCGGATCGAACTCCACCGCAGGGAGCGACCTTGGCTCCATTCCCTCCCAGCTCC
CGGACACCCCAAGCAGCATGGTGGCCAGCCCCATCGAGGCCTAA

> comp166395_c3_seq2
CACCACCACCACCATCACCAGGAGCGCCCCGTTGGCGTGCCGTCGCTGAGGCCCGACAGGGCTCACAGG
GGCCCCGGGGAGAAGAGGGGACCTGTGCCACGAGCCTCGTGCGGGGACCTGCACATGTCCGGCTCGGG
CAAGGCGAGCCGGGTCCGCACCGTCCTGACGGAGCGACAGCTGCAGACGCTGCGCACGTGCTACGCGGC
CAACCCACGGCCCGACGCCCTCCTGAAGGAGCAGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCAT
CCGCGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCAGCATCCTCATGAAGCAGATCCAGC
AGCAGGGCGGTGACAAGACGAACCTGCAAGGCCTGACGGGCACGCCGCTGGTGGCCGGGAGCCCCGTG
CCACACGAGTCGGCGATACCGGGGAACCCCGTCGAGGTGCAGAGTTTCCAGCCAGCGCCGTGGAAAAC
GCTCAGCGAGTTCGCGCTGAGCAGCGACCTCGACCAGGGCCCGTTCCAGCACCTGATGGGGTTCTCTGA
TGGGGGACCCGGATCGAACTCCACCGCAGGGAGTGACCTGGGCTCCATTCCCTCCCAGCTGCCGGACAC
CCCAAGCAGCATGGTGGCCAGCCCCATCGAAGCCTAAAACAACAACCCAACGTGGCCGCGAGAACCAT
GACCACCTCTACGGTGACTCCGAGCCTCTTCCGCTCCAAGCCCAGCAGCTAAGCGAGCCTCCTTTAAAA
CGCTTCTGCTCAACGCCATCAGCCCAGTTTTTTTTACGTAGCTTACGGAGGAGGCATTGAAACTAAAT
TATATTACCACGGGGGTGGAAAATCGGCTGTTGCGTCATGGACAGCGACAACTGTTGAGTTTGGGTGG
TGAAAGAATTAATTTGTTTCGTCAGAAGAATGAACGTAGCTTCGTTCTTCCGCCGTTCGGGAATATAA
GAGCCACTTGAGGTCATCTTTTATGTTTATTTACATCTCAAACGAAGAGGTCATCGATTAATCCACAA
TTTTATTCAGGAATCGTTCGACTGTCAAGCCAGTCAGTGAGCATCCTTTATTTTAACCAGGGGACGAA
TTCAACCTGTTGCGACACTTGGCACGTTACCTCTTGTGCAAGGGTGACCTGGGTATAGTCACGGCTGC
GAATAGCCGTGGCTGTGGCACGAAACAATTATTTTTTTTTGTTTTATCAGAAGACACGGCGAGCTGGG
GAAAGTGCAAGAATGAATTACAAACGACAAAAAAATTGTAAAAAAAATAGAAAG

> comp166395_c3_seq12
GCAGCAGCAGCAACAACAACAACATTTGCACCCCTGCCCGCAGCCACCCAGTGGCGCGGTAGCTCCAGA
GATGCATGGCAGGGTGCACTGTGCCTGGCACGCGGCTCGCATCATCGTCGCGTGAAGGAGGACTTCCC
ATCACCAGCATCACCGCCGCCGCCTCTTCCTCCTCCTCCTCGGCAGAGGAGATCGAGGCGGAGAGCTCG
TTGGCCGTGGCGAGAAGAAACACGGCGAAGAAGACGAGGACATGGGTGATATGGGAGACAACTCCAA
AAAGAAGCGGCTGGCGTCCGTGTGCGTTGGCTGCGGCGCCCAGATCCGAGACCCCTTCATCCTGCGCGT
GTCGCCCGACCTCGAGTGGCACGCCTCGTGCCTCAAGTGCGCCGAGTGCGCGCAGTACCTGGACGAGAC
GTGCACATGCTTCGTGCGCGACGGCAAGACCTACTGCAAGCGGGACTACGTCAGGCTGTTTGGCATCA
AGTGCGCTAAGTGTGGCGTGGGCTTCAGCAAGACGGACTTCGTGATGAGGGTGCGGACCCAGGTGTTC
CACCTCGAGTGCTTCCGCTGCGTGGCGTGCAGCCGCCAGCTCATCCCGGGTGACGAGTTTGCGCTGCGG

GAGGACGGCCTCTTCTGCCGCGCCGACCACGAGCTCCTGGAGAGGGCCGGCACCGCCGACGGCATGGGC
AGCCCGGGGCAGGGGGGCGGCCGGCCCCTGCAGCTCGCCGCCGAGCCCATGTCCGCGGCTGGCCGCCAC
CACCCACTGCGCCCGCACGTGCACAAGCAGGCGGAGAAGACAACGCGCGTGCGCACGGTGCTGAACGA
GAAGCAGCTGCACACGCTGAGGACCTGCTACGCGGCCAACCCGCGGCCCGATGCGCTCATGAAGGAGC
AGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCATCCGCGTCTGGTTCCAGAACAAGCGCTGCAAG
GACAAGAAGCGCAGCATCCTCATGAAGCAGATCCAGCAGCAGGGCGGTGACAAGACGAACCTGCAAGG
CCTGACGGGCACGCCGCTGGTGGCCGGGAGCCCCGTGCCACACGAGTCGGCGATACCGGGGAACCCCGT
CGAGGTGCAGAGTTTCCAGCCAGCGCCGTGGAAAACGCTCAGCGAGTTCGCGCTGAGCAGCGACCTCG
ACCAGGGCCCGTTCCAGCACCTGATGGGGTTCTCTGATGGGGGACCCGGATCGAACTCCACCGCAGGG
AGTGACCTGGGCTCCATTCCCTCCCAGCTGCCGGACACCCCAAGCAGCATGGTGGCCAGCCCCATCGAA
GCCTAAAACAACAACCCAACGTGGCCGCGAGAACCATGACCACCTCTACGGTGACTCCGAGCCTCTTCC
GCTCCAAGCCCAGCAGCTAAGCGAGCCTCCTTTAAAACGCTTCTGCTCAACGCCATCAGCCCAGTTTTT
TTTACGTAGCTTACGGAGGAGGCATTGAAACTAAATTATATTACCACGGGGGTGGAAAATCGGCTGTT
GCGTCATGGACAGCGACAACTGTTGAGTTTGGGTGGTGAAAGAATTAATTTGTTTCGTCAGAAGAAT
GAACGTAGCTTCGTTCTTCCGCCGTTCGGGAATATAAGAGCCACTTGAGGTCATCTTTTATGTTTATT
TACATCTCAAACGAAGAGGTCATCGATGAATCCACAAATTTCTTAAAGCTCTGGTCAGCCCGTACATT
CGGGAATCGTTTGACTGTCAAGCCAGTCAGCGAGCATCCTTTATTTTAACCAGGGGACGAATTCAACC
TGTTGCGACACTTGGCACGTTACCTCTTGTGCAAGGGTGACCTGGGTATAGTCACGGCTGCGAATAGC
CGTGGCTGTGGCACGAAACAATTATTTTTTTTTGTTTTATCAGAAGACACGGCGAGCTGGGGAAAGTG
CAAGAATGAATTACAAACGACAAAAAAATTGTAAAAAAAATAGAAAG
> comp23347_c0_seq1
TTTGTTAGGTTATATTTATCACAGCATACATTGTTGTTATTATATTTCTTTTGTGGATACATCGAGTT
AGAAAGTCGAAGTGGATTTTTCGACCTTCATGGGCGATATGGGAGACTCGGCCAAGAAGCGGAGGAG
AGCTGCGCTGTGCGTCGGCTGTGGCACGCACATCCAGGACCCGTTCATCCTGCGCGTGTCTCCGGACCT
CGAGTGGCACGCCGCCTGCCTCAAGTGCGCCGAATGCGGCCAGAGCCTCGACGAGACGTGCACCTGCTT
CGTGCGCGATGGCAAAACCTTCTGCAAGCGCGACTACAGCAGGCTGTTCGGGATCAAGTGCGCCAAGT
GCAACGCGTCGTTCAGCAAGACGGACTTTGTGATGCGCGCGCGCTCCCACATCTACCACATGGAGTGC
TTCCGCTGCGAGACCTGCTCGCGCCAGCTGCTGCCCGGCGACGAGTTCGCGCTGCGCGACGGCGGCGGA
CTCCTGTGTCGAGCCGACCAGCACCACGGCG

> comp166395_c3_seq15
CACCACCACCACCATCACCAGGAGCGCCCCGTTGGCGTGCCGTCGCTGAGGCCCGACAGGGCTCACAGG
GGCCCCGGGGAGAAGAGGGGACCTGTGCCACGAGCCTCGTGCGGGGACCTGCACATGTCCGGCTCGGG
CAAGGCGAGCCGGGTCCGCACCGTCCTGACGGAGCGACAGCTGCAGACGCTGCGCACGTGCTACGCGGC
CAACCCACGGCCCGACGCCCTCCTGAAGGAGCAGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCAT
CCGCGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCAGCATCCTCATGAAGCAGATCCAGC
AGCAGGGCGGTGACAAGACGAACCTGCAAGGCCTGACGGGCACGCCGCTGGTGGCCGGGAGCCCCGTG
CCACACGAGTCGGCGATACCGGGGAACCCCGTCGAGGTGCAGAGTTTCCAGCCAGCGCCGTGGAAAAC
GCTCAGCGAGTTCGCGCTGAGCAGCGACCTCGACCAGGGCCCGTTCCAGCACCTGATGGGGTTCTCTGA
TGGGGGACCCGGATCGAACTCCACCGCAGGGAGTGACCTGGGCTCCATTCCCTCCCAGCTGCCGGACAC
CCCAAGCAGCATGGTGGCCAGCCCCATCGAAGCCTAAAACAACAACCCAACGTGGCCGCGAGAACCAT
GACCACCTCTACGGTGACTCCGAGCCTCTTCCGCTCCAAGCCCAGCAGCTAAGCGAGCCTCCTTTAAAA
CGCTTCTGCTCAACGCCATCAGCCCAGTTTTTTTTACGTAGCTTACGGAGGAGGCATTGAAACTAAAT
TATATTACCACGGGGGTGGAAAATCGGCTGTTGCGTCATGGACAGCGACAACTGTTGAGTTTGGGTGG
TGAAAGAATTAATTTGTTTCGTCAGAAGAATGAACGTAGCTTCGTTCTTCCGCCGTTCGGGAATATAA
GAGCCACTTGAGGTCATCTTTTATGTTTATTTACATCTCAAACGAAGAGGTCATCGATGAATCCACAA
ATTTCTTAAAGCTCTGGTCAGCCCGTACATTCGGGAATCGTTTGACTGTCAAGCCAGTCAGCGAGCAT
CCTTTATTTTAACCAGGGGACGAATTCAACCTGTTGCGACACTTGGCACGTTACCTCTTGTGCAAGGG
TGACCTGGGTATAGTCACGGCTGCGAATAGCCGTGGCTGTGGCACGAAACAATTATTTTTTTTTGTTT
TATCAGAAGACACGGCGAGCTGGGGAAAGTGCAAGAATGAATTACAAACGACAAAAAAATTGTAAAA
AAAATAGAAAG

> comp166395_c3_seq17

GCAGCAGCAGCAACAACAACAACATTTGCACCCCTGCCCGCAGCCACCCAGTGGCGCGGTAGCTCCAGA
GATGCATGGCAGGGTGCACTGTGCCTGGCACGCGGCTCGCATCATCGTCGCGTGAAGGAGGACTTCCC
ATCACCAGCATCACCGCCGCCGCCTCTTCCTCCTCCTCCTCGGCAGAGGAGATCGAGGCGGAGAGCTCG
TTGGCCGTGGCGAGAAGAAACACGGCGAAGAAGACGAGGACATGGGTGATATGGGAGACAACTCCAA
AAAGAAGCGGCTGGCGTCCGTGTGCGTTGGCTGCGGCGCCCAGATCCGAGACCCCTTCATCCTGCGCGT
GTCGCCCGACCTCGAGTGGCACGCCTCGTGCCTCAAGTGCGCCGAGTGCGCGCAGTACCTGGACGAGAC
GTGCACATGCTTCGTGCGCGACGGCAAGACCTACTGCAAGCGGGACTACGTCAGGCTGTTTGGCATCA
AGTGCGCTAAGTGTGGCGTGGGCTTCAGCAAGACGGACTTCGTGATGAGGGTGCGGACCCAGGTGTTC
CACCTCGAGTGCTTCCGCTGCGTGGCGTGCAGCCGCCAGCTCATCCCGGGTGACGAGTTTGCGCTGCGG
GAGGACGGCCTCTTCTGCCGCGCCGACCACGAGCTCCTGGAGAGGGCCGGCACCGCCGACGGCATGGGC
AGCCCGGGGCAGGGGGGCGGCCGGCCCCTGCAGCTCGCCGCCGAGCCCATGTCCGCGGCTGGCCGCCAC
CACCCACTGCGCCCGCACGTGCACAAGCAGGCGGAGAAGACAACGCGCGTGCGCACGGTGCTGAACGA
GAAGCAGCTGCACACGCTGAGGACCTGCTACGCGGCCAACCCGCGGCCCGATGCGCTCATGAAGGAGC
AGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCATCCGCGTCTGGTTCCAGAACAAGCGCTGCAAG
GACAAGAAGCGCAGCATCCTCATGAAGCAGATCCAGCAGCAGGGCGGTGACAAGACGAACCTGCAAGG
CCTGACGGGCACGCCGCTGGTGGCCGGGAGCCCCGTGCCACACGAGTCGGCGATACCGGGGAACCCCGT
CGAGGTGCAGAGTTTCCAGCCAGCGCCGTGGAAAACGCTCAGCGAGTTCGCGCTGAGCAGCGACCTCG
ACCAGGGCCCGTTCCAGCACCTGATGGGGTTCTCTGATGGGGGACCCGGATCGAACTCCACCGCAGGG
AGTGACCTGGGCTCCATTCCCTCCCAGCTGCCGGACACCCCAAGCAGCATGGTGGCCAGCCCCATCGAA
GCCTAAAACAACAACCCAACGTGGCCGCGAGAACCATGACCACCTCTACGGTGACTCCGAGCCTCTTCC
GCTCCAAGCCCAGCAGCTAAGCGAGCCTCCTTTAAAACGCTTCTGCTCAACGCCATCAGCCCAGTTTTT
TTTACGTAGCTTACGGAGGAGGCATTGAAACTAAATTATATTACCACGGGGGTGGAAAATCGGCTGTT
GCGTCATGGACAGCGACAACTGTTGAGTTTGGGTGGTGAAAGAATTAATTTGTTTCGTCAGAAGAAT
GAACGTAGCTTCGTTCTTCCGCCGTTCGGGAATATAAGAGCCACTTGAGGTCATCTTTTATGTTTATT
TACATCTCAAACGAAGAGGTCATCGATTAATCCACAATTTTATTCAGGAATCGTTCGACTGTCAAGCC
AGTCAGTGAGCATCCTTTATTTTAACCAGGGGACGAATTCAACCTGTTGCGACACTTGGCACGTTACC
TCTTGTGCAAGGGTGACCTGGGTATAGTCACGGCTGCGAATAGCCGTGGCTGTGGCACGAAACAATTA
TTTTTTTTTGTTTTATCAGAAGACACGGCGAGCTGGGGAAAGTGCAAGAATGAATTACAAACGACAA
AAAAATTGTAAAAAAAATAGAAAG

> comp106466_c0_seq1
GGTGTCCGGGAGCTGGGAGGGAATGGAGCCAAGGTCGCTCCCTGCGGTGGAGTTCGATCCGGGTCCCC
CGTCAGAGAACCCCATCAGGTGCTGGAACGGGCCCTGGTCGAGGTCGCTGCTCAGCGCGAACTCGCTG
AGCGTTTTCCACGGCGCTGGCTGGAAACTCTGCACCTCGACGGGGTTCCCCGGTATCGCCGCCTCGTGT
GGCACGGGGCTCCCGGCCACCAGC

S2.1.6 ISLC

> TRINITY_DN77023_c0_g1_i1
GCGGCGACGATGGGAACCTACGAGGAGGAGGCGGCGGCGGTGGTGCCGACGGCGCCGAGGGCGTCTCC
GCGGGCTGCGCGGGGTGCGGGAGGCCCATCCGGGACGCGTTCCTGCTGCGCGTGTGGCCCGACCTGCGG
TTCCACGCCGCCTGCCTGCGCTGCGCCGAGTGCCGCGCGCAACTGCACGAGGCCCGCTCCTGCTTCGTG
CGCGCAGGCCGCACCTTCTGCCAGCGGGACTACAACCGTCTGTTCGGGGTCAAGTGCTCGCGCTGCTCG
CTGGGTCTGAGCCGCACCGAGCTCGTGATGCGAGTGCGCGGCCGCGTCTACCACCTCGCGTGC

> TRINITY_DN50605_c0_g1_i1
GCTCAGGCCCGTCATCTCCACGAGCTGCTCCTTCAGGAGGGCGTCGGGCCGCGGGTTGGCCGCGTAGCA
CGTGCGCAGCGTCTGCAGCTGTCGCTCCGTCAGGACGGTGCGGACCCGGCTCGCCTTGCCCGAGCCGGA
CATGTGCAGGTCCCCGCACGAGGCTCGTGGCACTGGTCCCCTCTTCTCCCCGGGGCCCCTGTGAGCCCT
GTCGGGCCTCAGCGACGGCACGCCAACGGGGCGCTCCTGGTGATGGTGGTGGTGGTGATGGTTGTGAT
GGTTGTGATGGTTGTGGCCCTGGTTATGGACGTGAGGATGATGATGGTGGTGGTGATGGTTGTG

> comp128550_c0_seq2

GAGGAGGAGGAGGCGGCGGTGCCGACGGCGCCGAGGGCGTCTCCGCGGGCTGCGCGGGGTGCGGGAGG
CCCATCCGGGACGCGTTCCTCCTGCGCGTGTGGCCCAACCTGCGGTTCCACGCCGCCTGCCTGCGCTGC
GCCGAGTGCCGTGCGCAGCTGCACGAGGCCCGCTCCTGCTTCGTGCGCGCAGGCCGCACCTTCTGCCAG
CGGGACTACAACCGTCTGTTCGGGGTCAAGTGCTCGCGCTGCTCGCTGGGTCTGAGCCGCACCGAGCTC
GTGATGCGAGTGCGCGGCCGCGTCTACCACCTCGCGTGCTTCCGCTGCTGGGCGTGCGCCCGGCGCCTG
CTGCCGGGAGACGAGGTCTCCCTGCGCCCCGCCGGAGAGCTCCTGTGCAGGGCGCACAGCCTCCCGCCC
CCACCG

> comp166395_c3_seq11
CAAGAAGATTGTTGTTGTTGTTGTCTTTATTGTCGTTGTTGTTGTTGCTGTAGTGGACGATGCTCTCC
CCAAGAAGCCCGTTCTGCCCATCGCGTCGTCCAAGTGGGAGCGCGTGGACGACCTGGACACGGACACC
ACCCGCCGCTCCCCCGCTCCGTCGGGCAGCAAGAGGGACGACGACTCGGACGATGCCAGCGGCGGCAAT
GTTGCCGCCGCGGGCGGCAAGTTGCCGCCGGGCTTTGCCGCCACGGCCGCCGCCTCTTCCTCCTCCTCCT
CGGCAGAGGAGATCGAGGCGGAGAGCTCGTTGGCCGTGGCGAGAAGAAACACGGCGAAGAAGACGAG
GACATGGGTGATATGGGAGACAACTCCAAAAAGAAGCGGCTGGCGTCCGTGTGCGTTGGCTGCGGCGC
CCAGATCCGAGACCCCTTCATCCTGCGCGTGTCGCCCGACCTCGAGTGGCACGCCTCGTGCCTCAAGTG
CGCCGAGTGCGCGCAGTACCTGGACGAGACGTGCACATGCTTCGTGCGCGACGGCAAGACCTACTGCA
AGCGGGACTACGTCAGGCTGTTTGGCATCAAGTGCGCTAAGTGTGGCGTGGGCTTCAGCAAGACGGAC
TTCGTGATGAGGGTGCGGACCCAGGTGTTCCACCTCGAGTGCTTCCGCTGCGTGGCGTGCAGCCGCCAG
CTCATCCCGGGTGACGAGTTTGCGCTGCGGGAGGACGGCCTCTTCTGCCGCGCCGACCACGAGCTCCTG
GAGAGGGCCGGCACCGCCGACGGCATGGGCAGCCCGGGGCAGGGGGGCGGCCGGCCCCTGCAGCTCGCC
GCCGAGCCCATGTCCGCGGCTGGCCGCCACCACCCACTGCGCCCGCACGTGCACAAGCAGGCGGAGAAG
ACAACGCGCGTGCGCACGGTGCTGAACGAGAAGCAGCTGCACACGCTGAGGACCTGCTACGCGGCCAA
CCCGCGGCCCGATGCGCTCATGAAGGAGCAGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCATCCG
CGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCACCCTGGCGGGGCGGCACGACGGAGAGC
CGCCGTCGCCGGACGCCAAGACCTGCCTGCAGGGCCTGGAGGGCAAAGCCCTGGTGGCGCGGAGCCCCG
TGACACAGGACAGCGAGGCGGAGGCCGAAGCTGCCGGAGTCGGCGGTGGGTGTGGCGGCGGCGGCGCC
GTGGCGG

> comp166395_c3_seq16
CACCACCACCACCATCACCAGGAGCGCCCCGTTGGCGTGCCGTCGCTGAGGCCCGACAGGGCTCACAGG
GGCCCCGGGGAGAAGAGGGGACCTGTGCCACGAGCCTCGTGCGGGGACCTGCACATGTCCGGCTCGGG
CAAGGCGAGCCGGGTCCGCACCGTCCTGACGGAGCGACAGCTGCAGACGCTGCGCACGTGCTACGCGGC
CAACCCACGGCCCGACGCCCTCCTGAAGGAGCAGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCAT
CCGCGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCACCCTGGCGGGGCGGCACGACGGAG
AGCCGCCGTCGCCGGACGCCAAGACCTGCCTGCAGGGCCTGGAGGGCAAAGCCCTGGTGGCGCGGAGCC
CCGTGACACAGGACAGCGAGGCGGAGGCCGAAGCTGCCGGAGTCGGCGGTGGGTGTGGCGGCGGCGGC
GCCGTGGCGG

> comp166395_c3_seq2
CACCACCACCACCATCACCAGGAGCGCCCCGTTGGCGTGCCGTCGCTGAGGCCCGACAGGGCTCACAGG
GGCCCCGGGGAGAAGAGGGGACCTGTGCCACGAGCCTCGTGCGGGGACCTGCACATGTCCGGCTCGGG
CAAGGCGAGCCGGGTCCGCACCGTCCTGACGGAGCGACAGCTGCAGACGCTGCGCACGTGCTACGCGGC
CAACCCACGGCCCGACGCCCTCCTGAAGGAGCAGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCAT
CCGCGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCAGCATCCTCATGAAGCAGATCCAGC
AGCAGGGCGGTGACAAGACGAACCTGCAAGGCCTGACGGGCACGCCGCTGGTGGCCGGGAGCCCCGTG
CCACACGAGTCGGCGATACCGGGGAACCCCGTCGAGGTGCAGAGTTTCCAGCCAGCGCCGTGGAAAAC
GCTCAGCGAGTTCGCGCTGAGCAGCGACCTCGACCAGGGCCCGTTCCAGCACCTGATGGGGTTCTCTGA
TGGGGGACCCGGATCGAACTCCACCGCAGGGAGTGACCTGGGCTCCATTCCCTCCCAGCTGCCGGACAC
CCCAAGCAGCATGGTGGCCAGCCCCATCGAAGCCTAAAACAACAACCCAACGTGGCCGCGAGAACCAT
GACCACCTCTACGGTGACTCCGAGCCTCTTCCGCTCCAAGCCCAGCAGCTAAGCGAGCCTCCTTTAAAA
CGCTTCTGCTCAACGCCATCAGCCCAGTTTTTTTTACGTAGCTTACGGAGGAGGCATTGAAACTAAAT

TATATTACCACGGGGGTGGAAAATCGGCTGTTGCGTCATGGACAGCGACAACTGTTGAGTTTGGGTGG
TGAAAGAATTAATTTGTTTCGTCAGAAGAATGAACGTAGCTTCGTTCTTCCGCCGTTCGGGAATATAA
GAGCCACTTGAGGTCATCTTTTATGTTTATTTACATCTCAAACGAAGAGGTCATCGATTAATCCACAA
TTTTATTCAGGAATCGTTCGACTGTCAAGCCAGTCAGTGAGCATCCTTTATTTTAACCAGGGGACGAA
TTCAACCTGTTGCGACACTTGGCACGTTACCTCTTGTGCAAGGGTGACCTGGGTATAGTCACGGCTGC
GAATAGCCGTGGCTGTGGCACGAAACAATTATTTTTTTTTGTTTTATCAGAAGACACGGCGAGCTGGG
GAAAGTGCAAGAATGAATTACAAACGACAAAAAAATTGTAAAAAAAATAGAAAG

> comp166395_c3_seq15
CACCACCACCACCATCACCAGGAGCGCCCCGTTGGCGTGCCGTCGCTGAGGCCCGACAGGGCTCACAGG
GGCCCCGGGGAGAAGAGGGGACCTGTGCCACGAGCCTCGTGCGGGGACCTGCACATGTCCGGCTCGGG
CAAGGCGAGCCGGGTCCGCACCGTCCTGACGGAGCGACAGCTGCAGACGCTGCGCACGTGCTACGCGGC
CAACCCACGGCCCGACGCCCTCCTGAAGGAGCAGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCAT
CCGCGTCTGGTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCAGCATCCTCATGAAGCAGATCCAGC
AGCAGGGCGGTGACAAGACGAACCTGCAAGGCCTGACGGGCACGCCGCTGGTGGCCGGGAGCCCCGTG
CCACACGAGTCGGCGATACCGGGGAACCCCGTCGAGGTGCAGAGTTTCCAGCCAGCGCCGTGGAAAAC
GCTCAGCGAGTTCGCGCTGAGCAGCGACCTCGACCAGGGCCCGTTCCAGCACCTGATGGGGTTCTCTGA
TGGGGGACCCGGATCGAACTCCACCGCAGGGAGTGACCTGGGCTCCATTCCCTCCCAGCTGCCGGACAC
CCCAAGCAGCATGGTGGCCAGCCCCATCGAAGCCTAAAACAACAACCCAACGTGGCCGCGAGAACCAT
GACCACCTCTACGGTGACTCCGAGCCTCTTCCGCTCCAAGCCCAGCAGCTAAGCGAGCCTCCTTTAAAA
CGCTTCTGCTCAACGCCATCAGCCCAGTTTTTTTTACGTAGCTTACGGAGGAGGCATTGAAACTAAAT
TATATTACCACGGGGGTGGAAAATCGGCTGTTGCGTCATGGACAGCGACAACTGTTGAGTTTGGGTGG
TGAAAGAATTAATTTGTTTCGTCAGAAGAATGAACGTAGCTTCGTTCTTCCGCCGTTCGGGAATATAA
GAGCCACTTGAGGTCATCTTTTATGTTTATTTACATCTCAAACGAAGAGGTCATCGATGAATCCACAA
ATTTCTTAAAGCTCTGGTCAGCCCGTACATTCGGGAATCGTTTGACTGTCAAGCCAGTCAGCGAGCAT
CCTTTATTTTAACCAGGGGACGAATTCAACCTGTTGCGACACTTGGCACGTTACCTCTTGTGCAAGGG
TGACCTGGGTATAGTCACGGCTGCGAATAGCCGTGGCTGTGGCACGAAACAATTATTTTTTTTTGTTT
TATCAGAAGACACGGCGAGCTGGGGAAAGTGCAAGAATGAATTACAAACGACAAAAAAATTGTAAAA
AAAATAGAAAG

> comp166395_c3_seq5
ATCCCTGCCGTCTCTGCCGTCCCGGGCGGGGGGGGCCGGGCCGCTGGACGGGGCCCCTCTGGACGATCT
CGACGGGCTGCCCATGCAGGCCGGGGACGATATCGACGGGCTGCCCCTGGACGATGCTCTCCCCAAGAA
GCCCGTTCTGCCCATCGCGTCGTCCAAGTGGGAGCGCGTGGACGACCTGGACACGGACACCACCCGCCG
CTCCCCCGCTCCGTCGGGCAGCAAGAGGGACGACGACTCGGACGATGCCAGCGGCGGCAATGTTGCCGC
CGCGGGCGGCAAGTTGCCGCCGGGCTTTGCCGCCACGGCCGCCGCCTCTTCCTCCTCCTCCTCGGCAGA
GGAGATCGAGGCGGAGAGCTCGTTGGCCGTGGCGAGAAGAAACACGGCGAAGAAGACGAGGACATGG
GTGATATGGGAGACAACTCCAAAAAGAAGCGGCTGGCGTCCGTGTGCGTTGGCTGCGGCGCCCAGATC
CGAGACCCCTTCATCCTGCGCGTGTCGCCCGACCTCGAGTGGCACGCCTCGTGCCTCAAGTGCGCCGAG
TGCGCGCAGTACCTGGACGAGACGTGCACATGCTTCGTGCGCGACGGCAAGACCTACTGCAAGCGGGA
CTACGTCAGGCTGTTTGGCATCAAGTGCGCTAAGTGTGGCGTGGGCTTCAGCAAGACGGACTTCGTGA
TGAGGGTGCGGACCCAGGTGTTCCACCTCGAGTGCTTCCGCTGCGTGGCGTGCAGCCGCCAGCTCATCC
CGGGTGACGAGTTTGCGCTGCGGGAGGACGGCCTCTTCTGCCGCGCCGACCACGAGCTCCTGGAGAGG
GCCGGCACCGCCGACGGCATGGGCAGCCCGGGGCAGGGGGGCGGCCGGCCCCTGCAGCTCGCCGCCGAG
CCCATGTCCGCGGCTGGCCGCCACCACCCACTGCGCCCGCACGTGCACAAGCAGGCGGAGAAGACAACG
CGCGTGCGCACGGTGCTGAACGAGAAGCAGCTGCACACGCTGAGGACCTGCTACGCGGCCAACCCGCG
GCCCGATGCGCTCATGAAGGAGCAGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCATCCGCGTCTG
GTTCCAGAACAAGCGCTGCAAGGACAAGAAGCGCACCCTGGCGGGGCGGCACGACGGAGAGCCGCCGT
CGCCGGACGCCAAGACCTGCCTGCAGGGCCTGGAGGGCAAAGCCCTGGTGGCGCGGAGCCCCGTGACAC
AGGACAGCGAGGCGGAGGCCGAAGCTGCCGGAGTCGGCGGTGGGTGTGGCGGCGGCGGCGCCGTGGCG
G

> comp171797_c0_seq1

AGGAGGAGGAGGCGGCGGCGGTGGTGCCGACGGCGCCGAGGGCGTCTCCGCGGGCTGCGCGGGGTGCG
GGAGGCCCATCCGGGACGCGTTCCTGCTGCGCGTGTGGCCCGACCTGCGGTTCCACGCCGCCTGCCTGC
GCTGTGCCGAGTGCCGCGCGCAGCTGCACGAGGCCCGCTCCTGCTTCGTGCGCGCAGGCCGCACCTTCT
GCCAGCGGGACTACAACCGTCTGTTCGGGGTCAAGTGCTCGCGCTGCTCGCTGGGTCTGAGCCGCACCG
AGCTCGTGATGCGAGTGCGCGGCCGCGTCTACCACCTCGCGTGCTTCCGCTGCTGGGCGTGCGCCCGGC
GCCTGCTGCCGGGAGACGAGGTCTCCCTGCGCCCCGCCGGAGAGCTCTTGTGCAGGGCGCACAGCCTCC
CGCCCCCACCGCACCATCACCACCACCACCATCAT

> comp128550_c0_seq1
ACACGGGGCTGGTGGCTGCGATACTCTGTTCGGGGTCAAGTGCTCGCGCTGCTCGCTGGGTCTGAGCC
GCACCGAGCTCGTGATGCGAGTGCGCGGCCGCGTCTACCACCTCGCGTGCTTCCGCTGCTGGGCGTGCG
CCCGGCGCCTGCTGCCGGGAGACGAGGTCTCCCTGCGCCCCGCCGGAGAGCTCCTGTGCAGGGCGCACA
GCCTCCCGCCCCCACCG

> comp166395_c3_seq3
GCAGCAGCAGCAACAACAACAACATTTGCACCCCTGCCCGCAGCCACCCAGTGGCGCGGTAGCTCCAGA
GATGCATGGCAGGGTGCACTGTGCCTGGCACGCGGCTCGCATCATCGTCGCGTGAAGGAGGACTTCCC
ATCACCAGCATCACCGCCGCCGCCTCTTCCTCCTCCTCCTCGGCAGAGGAGATCGAGGCGGAGAGCTCG
TTGGCCGTGGCGAGAAGAAACACGGCGAAGAAGACGAGGACATGGGTGATATGGGAGACAACTCCAA
AAAGAAGCGGCTGGCGTCCGTGTGCGTTGGCTGCGGCGCCCAGATCCGAGACCCCTTCATCCTGCGCGT
GTCGCCCGACCTCGAGTGGCACGCCTCGTGCCTCAAGTGCGCCGAGTGCGCGCAGTACCTGGACGAGAC
GTGCACATGCTTCGTGCGCGACGGCAAGACCTACTGCAAGCGGGACTACGTCAGGCTGTTTGGCATCA
AGTGCGCTAAGTGTGGCGTGGGCTTCAGCAAGACGGACTTCGTGATGAGGGTGCGGACCCAGGTGTTC
CACCTCGAGTGCTTCCGCTGCGTGGCGTGCAGCCGCCAGCTCATCCCGGGTGACGAGTTTGCGCTGCGG
GAGGACGGCCTCTTCTGCCGCGCCGACCACGAGCTCCTGGAGAGGGCCGGCACCGCCGACGGCATGGGC
AGCCCGGGGCAGGGGGGCGGCCGGCCCCTGCAGCTCGCCGCCGAGCCCATGTCCGCGGCTGGCCGCCAC
CACCCACTGCGCCCGCACGTGCACAAGCAGGCGGAGAAGACAACGCGCGTGCGCACGGTGCTGAACGA
GAAGCAGCTGCACACGCTGAGGACCTGCTACGCGGCCAACCCGCGGCCCGATGCGCTCATGAAGGAGC
AGCTCGTGGAGATGACGGGCCTGAGCCCCCGCGTCATCCGCGTCTGGTTCCAGAACAAGCGCTGCAAG
GACAAGAAGCGCACCCTGGCGGGGCGGCACGACGGAGAGCCGCCGTCGCCGGACGCCAAGACCTGCCT
GCAGGGCCTGGAGGGCAAAGCCCTGGTGGCGCGGAGCCCCGTGACACAGGACAGCGAGGCGGAGGCCG
AAGCTGCCGGAGTCGGCGGTGGGTGTGGCGGCGGCGGCGCCGTGGCGG

S2.1.7 NELF

> comp47263_c0_seq1
GCATTCCTTCGACCTGGAGCAGGAGAAGGCCGAGCAGGAGGCGCTGAACAGCAAGAAGAAGCTGGAGC
GGATGTACAGCATGGATCGCATCTCCGACGACATCGAGTGCCGGAGCTGGTTCCCCAGCGAGAGCATG
CACATGCTGCACAGCACCGCGTCGACGCTGCAGGCCATCGCTGCCTTCCGTGGATACGCGGAGCGGAAG
AGACGCAAGCGGGAACTGGATCCTTCAGCGATGGA

> comp158004_c0_seq1
CTGCATCGGGAGCCCAACAGGGTTGCACAGGGAGCCCAACAGCGCCCGCTGGGAGGAGGAGAGGGAGA
TCGAGCTCGAATCTTGCAGCCGAACGGACTTCATCCCGCCTAAAGTTATGCTGATTTCATCAAAGGTC
CCAAAGGCCGAGTACATCCCCAACATCATCCGCAAAGATGATCCGTCCATCATCGCTATCCTCTACGAC
CATGAGCATGCAACCTTTTCTGACATTCTAGATGAGATAGAGAAGAAGCTGAATGAATATCGAAGAG
GCTGCAAAATCTGGAAGCTTCTTATTTACTGCCAGGGAGGGCCCGGGTACTTGTATCTGTTGAAGAAT
AAGGTGGCAACATTCGCCAAGCTGGAGAAAGAAGATGATTTAATTATGTTCTGGAAGTGCCTCGGAA
GCTTGATGAGCAAGATCAACACCGAGCTCAACGTCGTCCACATTATGGGATGCTACGTGCTGGGCAAT
CACAATGGAGAGAAGCTGTTGGACAGCCTCAAGTCTATAATGGGGCCCTATCGGGTGTCTTTTGAGTC
ACCACTGGAGTTGTCTGCTCAAGGCAAGAAGATGATAGAAACTTACTTTAGCTTTCGTCTTTATAAGT
TATGGAAGGCTCGGCAGCATTCCAAACTTGTGGATGATTTTGACGAAATCTTATGAGTAAACGTCCGG

TGTCGAAGCAAGAGATGGAACCACTGAAGGGGTTTGAAGTTTGAAAAGGAGGGAAAAGGAGAGCGCT
TTTGCTCTGTATCCGTCCTCAATGAAGGCAGTTTTGTGGCCATGGCGGCTGTTATATGTGCTCTGATG
AAATTCACACCCAGGCACAAGAGAAAGCCAAGCCTTAGGCCTACGTTTCTCTGCCTCGTGTCTCTTTTT
AAGGTAAAATGCACATCGAATGCCCCGGGAGGCATGTGCAATTTGTATTGAACATCATTGCACCAAAG
CAATTGGCGGCTGGGATTTTTAGTGTGAAAACATTTAAGGAGCTTCCTCGGCCATTTGTTTCTGTGCC
ACTCAGAGCATCGTCCGGAAAGTGGGGGCCTTTAAAGGCTACGAAGTATATTTGGAAGGTCGTTGTGC
ATACCAGCACACAGTTGGTGAGCAGTAGATTGCTTTTGATGGTTTTAAAGACGATGGTACTGGTTGAC
TTTCACGTGGCATATAAGGCATGGTGTTGAAACGTGCTGTTTAATGTACATTTTGCTACACCGTTTTT
AAATACATGTTACACTACACTTAAGACACTGTCACTAACACAGACAAAAACTGTTGTTGCTGCAGACA
TAAACATTTTAATTTTGTGTCGCATTTACAAAACAATACATTGGTATTTTACGTACTAAAAATACTTT
AACGGAAAATGTAGGAGTATTTGTATACCGCACATTTTAGTGAATGGATATTGTACAAATACGGTTG
TTTGTAAGAACAAAAAAGATCTAAGCCCAGTTACAGACAATTTAGAATTTTCACTTTTCAAAATTGTT
GATGTGCATCTAAAATGGCACCACCCCAATGACATATCTCATGGCAATAGTTCTGTTGACTTGTAAAA
AAAAAACCTCTTATGAAATGGATGTGCCTTTATTTAGTTTATTACTCTGGAAAATAAATATATATTAC
GCTTTAGCGATATTTACCAGCAAAACTAAGGCATGTTGTCCTCCCAGTTGAGCCGATAAAACAGATGC
AAATCATTTGATTGTTACGTTTGTTTCATTCACCTGCAGTTGATGTGCAACTCCGTTTAAACATTTAA
TATCTTTATAGGCGTGCATTGTAAACTATATTGCCTATCGTTGTAAACATGTACACGCAGCTACTATC
CTTTATCACACTTGTCTCTCATTAGACACAGACTAAGCATAACATAATATTGCTTGGCATGGTTATTT
ATTGAGATCACTTAACTGGTAGGCAATATTGATATACTGTATCGTACACACACATTTCACTTCGACTT
CCTAATGGGTCGAGTTTGTCATTTAGCAATTCTTAATGTCATGAAGTCTGCTGCTAATGTAATGCAGC
AGTACAGTTTATATTGTTATTTAAACATCATCAGAAGTGTGTCATTGTCAGGATAACTTGTACAAGTC
TCCTGAGATTAATGCCTGATCTTTTGTAAAAGTATATAACTTAGCGAGATGATTTTTGCCTTTGAAGG
AGCTGTGATGTATAATGATCTATTTTGTGCTTTAATTGCCTACTCGAGTTTATGACAGAGTGCTTTCT
TGTATACAGATCCTCCGTAATCACACACTTTCTGTGGCCGCATATTTTTACGATTGCTGGCAGAAGTC
ATCATGCGCAAGTTTTGTGTTATATATTCTATTGAAACATGGTTAATGAAAAGCACAAAATACTAAT
TATGGTAATAAATACAGTACTATATATACACCTGTACCTATACCTGTTTTTCGTAATTTTATCCATAT
TAGGCAAGCTAGCATTCCCATGATGCTTATGAACGTAACACGGTTATTGTTAAGTAATACTTCAGAAT
TTAACGTTTTATTGCATAGTCCCGAAAAGTGTGTGTTAGGTGCTGAGCACATAACAGCGTTGCTTTTT
AAAGAAATGCTGTCGTTAGGTCTGAATATGCGTTGTGGTTACAGTTGGTTAACTTTGAGTGTAGCGTC
ATGAATTGGAAACCATGTCTGCAAAGATAATTCAATTGAACTACACCCGTGAAGCATTACAGTTGTTA
AAGCAGATCTCGAATGTGCAGGATGTTACAGTGGCTGGCCCTAGCGTTGCCACCAATTTTCACAGACA
GTGTTTCTCAGCCAGCCTAGATCTGTCAAAACCGTCAGTTTAATACATCGGCTTTCGAGACCAATAAT
ACCCACAATATCGGTTTTACTCTCCAACACACCAGTGTG

> comp158004_c0_seq2
CTGCTGCCGCCGCTGCTGCCGCTGAAGTGCGATCCCCGGGAGATTGCGTGAAGCCAGCGCAGGTGGCGC
CTGCCCTTGTTCTCCTTGCTGTGGCCTCCGCCGCCGTGGCTGCAGCCGGCGGCCGGAGAGGTCCGGGCG
GACGCGGAGCGGGAGGGCGGCGGCTCCAGCGTGAGGGAGCGCGACTCCCTGGAGTGCCCGGCTTGCGA
GAACCTCGGCGTCTTCAAGCATTCCTTCGACCTGGAGCAAGAGAAGGCCGAGCAGGAGGCGCTGAGCA
GCAAGAAGAAGCTGGAGCGCATGTACAGCATGGATCGCATCTCCGACGACATCGAGTGCCGGAGTTGG
TTCCCCAGCGAGAGCATGCACATGCTGCACAGCACTGCGTCGACGCTGCAAGCCATCGCTGCCTTCCGT
GGATACGCGGAGCGGAAGAGACGCAAGCGGGAACTGGATCCTTCAGCGATGGTGCAGAGGAACTTCCA
CAAACATCTGCGGATGGTGGGACGGCACCGCGGGAGGGCTGCAGTGTCCCTCGACCGTCAAGAAAGCA
GAGAGTGTGGCCGGGCCCTGCCTCTGAGCGACGTGTCGCAGTCCAAGTCCGACCCAGTCCTCGATGCTC
GATTCAGTCACGACATGCAGAGCGCCTACGAGAGGTTGCACAGGGAGCCCAACAGCGCCCGCTGGGAG
GAGGAGAGGGAGATCGAGCTCGAATCTTGCAGCCGAACGGACTTCATCCCGCCTAAAGTTATGCTGAT
TTCATCAAAGGTCCCAAAGGCCGAGTACATCCCCAACATCATCCGCAAAGATGATCCGTCCATCATCG
CTATCCTCTACGACCATGAGCATGCAACCTTTTCTGACATTCTAGATGAGATAGAGAAGAAGCTGAAT
GAATATCGAAGAGGCTGCAAAATCTGGAAGCTTCTTATTTACTGCCAGGGAGGGCCCGGGTACTTGTA
TCTGTTGAAGAATAAGGTGGCAACATTCGCCAAGCTGGAGAAAGAAGATGATTTAATTATGTTCTGG
AAGTGCCTCGGAAGCTTGATGAGCAAGATCAACACCGAGCTCAACGTCGTCCACATTATGGGATGCTA
CGTGCTGGGCAATCACAATGGAGAGAAGCTGTTGGACAGCCTCAAGTCTATAATGGGGCCCTATCGGG
TGTCTTTTGAGTCACCACTGGAGTTGTCTGCTCAAGGCAAGAAGATGATAGAAACTTACTTTAGCTTT
CGTCTTTATAAGTTATGGAAGGCTCGGCAGCATTCCAAACTTGTGGATGATTTTGACGAAATCTTATG

AGTAAACGTCCGGTGTCGAAGCAAGAGATGGAACCACTGAAGGGGTTTGAAGTTTGAAAAGGAGGGA
AAAGGAGAGCGCTTTTGCTCTGTATCCGTCCTCAATGAAGGCAGTTTTGTGGCCATGGCGGCTGTTAT
ATGTGCTCTGATGAAATTCACACCCAGGCACAAGAGAAAGCCAAGCCTTAGGCCTACGTTTCTCTGCC
TCGTGTCTCTTTTTAAGGTAAAATGCACATCGAATGCCCCGGGAGGCATGTGCAATTTGTATTGAACA
TCATTGCACCAAAGCAATTGGCGGCTGGGATTTTTAGTGTGAAAACATTTAAGGAGCTTCCTCGGCCA
TTTGTTTCTGTGCCACTCAGAGCATCGTCCGGAAAGTGGGGGCCTTTAAAGGCTACGAAGTATATTTG
GAAGGTCGTTGTGCATACCAGCACACAGTTGGTGAGCAGTAGATTGCTTTTGATGGTTTTAAAGACGA
TGGTACTGGTTGACTTTCACGTGGCATATAAGGCATGGTGTTGAAACGTGCTGTTTAATGTACATTTT
GCTACACCGTTTTTAAATACATGTTACACTACACTTAAGACACTGTCACTAACACAGACAAAAACTGT
TGTTGCTGCAGACATAAACATTTTAATTTTGTGTCGCATTTACAAAACAATACATTGGTATTTTACGT
ACTAAAAATACTTTAACGGAAAATGTAGGAGTATTTGTATACCGCACATTTTAGTGAATGGATATTG
TACAAATACGGTTGTTTGTAAGAACAAAAAAGATCTAAGCCCAGTTACAGACAATTTAGAATTTTCA
CTTTTCAAAATTGTTGATGTGCATCTAAAATGGCACCACCCCAATGACATATCTCATGGCAATAGTTC
TGTTGACTTGTAAAAAAAAAACCTCTTATGAAATGGATGTGCCTTTATTTAGTTTATTACTCTGGAAA
ATAAATATATATTACGCTTTAGCGATATTTACCAGCAAAACTAAGGCATGTTGTCCTCCCAGTTGAGC
CGATAAAACAGATGCAAATCATTTGATTGTTACGTTTGTTTCATTCACCTGCAGTTGATGTGCAACTC
CGTTTAAACATTTAATATCTTTATAGGCGTGCATTGTAAACTATATTGCCTATCGTTGTAAACATGTA
CACGCAGCTACTATCCTTTATCACACTTGTCTCTCATTAGACACAGACTAAGCATAACATAATATTGC
TTGGCATGGTTATTTATTGAGATCACTTAACTGGTAGGCAATATTGATATACTGTATCGTACACACAC
ATTTCACTTCGACTTCCTAATGGGTCGAGTTTGTCATTTAGCAATTCTTAATGTCATGAAGTCTGCTG
CTAATGTAATGCAGCAGTACAGTTTATATTGTTATTTAAACATCATCAGAAGTGTGTCATTGTCAGGA
TAACTTGTACAAGTCTCCTGAGATTAATGCCTGATCTTTTGTAAAAGTATATAACTTAGCGAGATGAT
TTTTGCCTTTGAAGGAGCTGTGATGTATAATGATCTATTTTGTGCTTTAATTGCCTACTCGAGTTTAT
GACAGAGTGCTTTCTTGTATACAGATCCTCCGTAATCACACACTTTCTGTGGCCGCATATTTTTACGA
TTGCTGGCAGAAGTCATCATGCGCAAGTTTTGTGTTATATATTCTATTGAAACATGGTTAATGAAAAG
CACAAAATACTAATTATGGTAATAAATACAGTACTATATATACACCTGTACCTATACCTGTTTTTCGT
AATTTTATCCATATTAGGCAAGCTAGCATTCCCATGATGCTTATGAACGTAACACGGTTATTGTTAAG
TAATACTTCAGAATTTAACGTTTTATTGCATAGTCCCGAAAAGTGTGTGTTAGGTGCTGAGCACATAA
CAGCGTTGCTTTTTAAAGAAATGCTGTCGTTAGGTCTGAATATGCGTTGTGGTTACAGTTGGTTAACT
TTGAGTGTAGCGTCATGAATTGGAAACCATGTCTGCAAAGATAATTCAATTGAACTACACCCGTGAAG
CATTACAGTTGTTAAAGCAGATCTCGAATGTGCAGGATGTTACAGTGGCTGGCCCTAGCGTTGCCACC
AATTTTCACAGACAGTGTTTCTCAGCCAGCCTAGATCTGTCAAAACCGTCAGTTTAATACATCGGCTT
TCGAGACCAATAATACCCACAATATCGGTTTTACTCTCCAACACACCAGTGTG

> comp158004_c0_seq3
ATCGCACTTCAGCGGCAGCGGCGGCGGCTCCAGCGTGAGGGAGCGCGACTCCCTGGAGTGCCCGGCTTG
CGAGAACCTCGGCGTCTTCAAGCATTCCTTCGACCTGGAGCAAGAGAAGGCCGAGCAGGAGGCGCTGA
GCAGCAAGAAGAAGCTGGAGCGCATGTACAGCATGGATCGCATCTCCGACGACATCGAGTGCCGGAGT
TGGTTCCCCAGCGAGAGCATGCACATGCTGCACAGCACTGCGTCGACGCTGCAAGCCATCGCTGCCTTC
CGTGGATACGCGGAGCGGAAGAGACGCAAGCGGGAACTGGATCCTTCAGCGATGGTGCAGAGGAACTT
CCACAAACATCTGCGGATGGTGGGACGGCACCGCGGGAGGGCTGCAGTGTCCCTCGACCGTCAAGAAA
GCAGAGAGTGTGGCCGGGCCCTGCCTCTGAGCGACGTGTCGCAGTCCAAGTCCGACCCAGTCCTCGATG
CTCGATTCAGTCACGACATGCAGAGCGCCTACGAGAGGTTGCACAGGGAGCCCAACAGCGCCCGCTGG
GAGGAGGAGAGGGAGATCGAGCTCGAATCTTGCAGCCGAACGGACTTCATCCCGCCTAAAGTTATGCT
GATTTCATCAAAGGTCCCAAAGGCCGAGTACATCCCCAACATCATCCGCAAAGATGATCCGTCCATCA
TCGCTATCCTCTACGACCATGAGCATGCAACCTTTTCTGACATTCTAGATGAGATAGAGAAGAAGCTG
AATGAATATCGAAGAGGCTGCAAAATCTGGAAGCTTCTTATTTACTGCCAGGGAGGGCCCGGGTACTT
GTATCTGTTGAAGAATAAGGTGGCAACATTCGCCAAGCTGGAGAAAGAAGATGATTTAATTATGTTC
TGGAAGTGCCTCGGAAGCTTGATGAGCAAGATCAACACCGAGCTCAACGTCGTCCACATTATGGGATG
CTACGTGCTGGGCAATCACAATGGAGAGAAGCTGTTGGACAGCCTCAAGTCTATAATGGGGCCCTATC
GGGTGTCTTTTGAGTCACCACTGGAGTTGTCTGCTCAAGGCAAGAAGATGATAGAAACTTACTTTAGC
TTTCGTCTTTATAAGTTATGGAAGGCTCGGCAGCATTCCAAACTTGTGGATGATTTTGACGAAATCTT
ATGAGTAAACGTCCGGTGTCGAAGCAAGAGATGGAACCACTGAAGGGGTTTGAAGTTTGAAAAGGAG
GGAAAAGGAGAGCGCTTTTGCTCTGTATCCGTCCTCAATGAAGGCAGTTTTGTGGCCATGGCGGCTGT

TATATGTGCTCTGATGAAATTCACACCCAGGCACAAGAGAAAGCCAAGCCTTAGGCCTACGTTTCTCT
GCCTCGTGTCTCTTTTTAAGGTAAAATGCACATCGAATGCCCCGGGAGGCATGTGCAATTTGTATTGA
ACATCATTGCACCAAAGCAATTGGCGGCTGGGATTTTTAGTGTGAAAACATTTAAGGAGCTTCCTCGG
CCATTTGTTTCTGTGCCACTCAGAGCATCGTCCGGAAAGTGGGGGCCTTTAAAGGCTACGAAGTATAT
TTGGAAGGTCGTTGTGCATACCAGCACACAGTTGGTGAGCAGTAGATTGCTTTTGATGGTTTTAAAGA
CGATGGTACTGGTTGACTTTCACGTGGCATATAAGGCATGGTGTTGAAACGTGCTGTTTAATGTACAT
TTTGCTACACCGTTTTTAAATACATGTTACACTACACTTAAGACACTGTCACTAACACAGACAAAAAC
TGTTGTTGCTGCAGACATAAACATTTTAATTTTGTGTCGCATTTACAAAACAATACATTGGTATTTTA
CGTACTAAAAATACTTTAACGGAAAATGTAGGAGTATTTGTATACCGCACATTTTAGTGAATGGATA
TTGTACAAATACGGTTGTTTGTAAGAACAAAAAAGATCTAAGCCCAGTTACAGACAATTTAGAATTT
TCACTTTTCAAAATTGTTGATGTGCATCTAAAATGGCACCACCCCAATGACATATCTCATGGCAATAG
TTCTGTTGACTTGTAAAAAAAAAACCTCTTATGAAATGGATGTGCCTTTATTTAGTTTATTACTCTGG
AAAATAAATATATATTACGCTTTAGCGATATTTACCAGCAAAACTAAGGCATGTTGTCCTCCCAGTTG
AGCCGATAAAACAGATGCAAATCATTTGATTGTTACGTTTGTTTCATTCACCTGCAGTTGATGTGCAA
CTCCGTTTAAACATTTAATATCTTTATAGGCGTGCATTGTAAACTATATTGCCTATCGTTGTAAACAT
GTACACGCAGCTACTATCCTTTATCACACTTGTCTCTCATTAGACACAGACTAAGCATAACATAATAT
TGCTTGGCATGGTTATTTATTGAGATCACTTAACTGGTAGGCAATATTGATATACTGTATCGTACACA
CACATTTCACTTCGACTTCCTAATGGGTCGAGTTTGTCATTTAGCAATTCTTAATGTCATGAAGTCTG
CTGCTAATGTAATGCAGCAGTACAGTTTATATTGTTATTTAAACATCATCAGAAGTGTGTCATTGTCA
GGATAACTTGTACAAGTCTCCTGAGATTAATGCCTGATCTTTTGTAAAAGTATATAACTTAGCGAGAT
GATTTTTGCCTTTGAAGGAGCTGTGATGTATAATGATCTATTTTGTGCTTTAATTGCCTACTCGAGTT
TATGACAGAGTGCTTTCTTGTATACAGATCCTCCGTAATCACACACTTTCTGTGGCCGCATATTTTTA
CGATTGCTGGCAGAAGTCATCATGCGCAAGTTTTGTGTTATATATTCTATTGAAACATGGTTAATGAA
AAGCACAAAATACTAATTATGGTAATAAATACAGTACTATATATACACCTGTACCTATACCTGTTTTT
CGTAATTTTATCCATATTAGGCAAGCTAGCATTCCCATGATGCTTATGAACGTAACACGGTTATTGTT
AAGTAATACTTCAGAATTTAACGTTTTATTGCATAGTCCCGAAAAGTGTGTGTTAGGTGCTGAGCACA
TAACAGCGTTGCTTTTTAAAGAAATGCTGTCGTTAGGTCTGAATATGCGTTGTGGTTACAGTTGGTTA
ACTTTGAGTGTAGCGTCATGAATTGGAAACCATGTCTGCAAAGATAATTCAATTGAACTACACCCGTG
AAGCATTACAGTTGTTAAAGCAGATCTCGAATGTGCAGGATGTTACAGTGGCTGGCCCTAGCGTTGCC
ACCAATTTTCACAGACAGTGTTTCTCAGCCAGCCTAGATCTGTCAAAACCGTCAGTTTAATACATCGG
CTTTCGAGACCAATAATACCCACAATATCGGTTTTACTCTCCAACACACCAGTGTG

> ENSPMAT00000004778
ATGGGGACCTCTGTGTCCAAGAGAAGACAAATGCGAGCAGAAGCTATTTCCTCAGCGGCAATTAAATT
CAGGGCAGTGAAGTCTTTTGGGGACTACGTTCAGCTCGGTCACATGCAGGGCCACGAAAGAGGAGGCA
CAGTGGTGGTGGTGTCGGCGTCAGCCGCGACGACGCCCTCCCCGTGCCTACAGCCCCCCTCGCTGCTCC
AGCAGAAGCGTCGCCTCTCGCTGAGCCGCAGCTTCTCGGACGAGGCGGCGGAGCGTACCCCCGCGACGC
CGGCTCCCTCCCCCTCCGTCCCCTCCATCCCCTCCTCGTCTCCGTCTGCCGTTCCCGCGCCCCACGGCGG
CGGAGGCCACAGCAAGGAGAACAAGGGCAGGCGCCACCTCCGCTGGCTTCACGCCATCTCCCGGGGATC
GCACTTCAGCGGCAGCGGCGGCAGCAGCAGCAGCGCCGGGAGCAGCCGGGATGGTGGTGGCGGCGGCT
CCAGCGTGAGGGAGCGCGACTCCCTGGAGTGTCCGGCTTGCGAGAGCCTCGGCGTCTTCAAGCATTCCT
TCGACCTGGAGCAGGAGAAGGCCGAGCAGGAGGCACTGAACAGCAAGAAGAAGCTGGAGCGGATGTA
CAGCATGGATCGCATCTCCGACGACATCGAGTGCCGGAGCTGGTTCCCCAGCGAGAGCATGCACATGC
TGCACAGCACCGCGTCGACGCTGCAGGCGTACGTCTGCAGTGCCTTCCGTGGATACGCGGAGCGGAAG
AGACGCAAGCGGGAACTGGATCCTTCAGCGATGGTGCAGAGGAACTTCCACAAACATCTGCGGATGGT
GGGACGGCACCGCGGGAGGGCTGCAGTAAAAATGTCGCACCGCGAGAAAGCTTTTAGTGTTGCAATAT
CCTCCCCAACTGAAATGCGGTGGTCTCGGAACCTGGTGTCAGGTCACGACATGCAGAGCACCTACGAG
AGACTGCATAGGGAGCCCAACAGCGCCCGCTGGGAGGAGGAGAGGGAGATCGAGCTGGAATCTTGCAG
CCGAACGGACTTCATCCCGCCTAAAGTTATGCTGATTTCATCAAAGGTCCCAAAGGCCGAGTACATCC
CCAACATCATCCGCAAAGATGATCCGTCCATCATCGCTATCCTCTACGACCATGAGCATGCAACCTTTT
CTGACATTTTAGATGAGATAGAGAAGAAGCTGAATGAATATCGAAGAGGCTGCAAAATCTGGAAGCT
TCTGATTTACTGCCAGGGAGGGCCCGGATACTTGTATCTGTTGAAGAATAAGGTGGCAACATTCGCCA
AGCTGGAGAAAGAAGATGATTTAATTATGTTCTGGAAGTGCCTCGGAAGCTTGATGAGCAAGATCAA
CACCGAGCTCAACGTCGTCCACATTATGGGATGCTACGTGCTGGGCAATCACAATGGAGAGAAGCTGT

TGGACAGCCTCAAGTCTATAATGGGGCCCTATCGGGTGTCTTTTGAGTCACCACTGGAGTTGTCTGCT
CAAGGCAAGAAGATGATAGAAACTTACTTTAGCTTTCGTCTTTATAAGTTATGGAAGGCTCGGCAGCA
TTCCAAACTTGTGGATGATTTTGATGAAATCTTATGA

> Locus_109486_Transcript_1/1_Confidence_1.000_Length_421
GCATCCCATAATGTGGACGACGTTGAGCTCGGTGTTGATCTTGCTCATCAAGCTTCCGAGGCACTTCC
AGAACATAATTAAATCATCTTCTTTCTCCAGCTTGGCGAATGTTGCCACCTTATTCTTCAACAGATAC
AAGTATCCGGGCCCTCCCTGGCAGTAAATCAGAAGCTTCCAGATTTTGCAGCCTCTTCGATATTCATT
CAGCTTCTTCTCTATCTCATCTAAAATGTCAGAAAAGGTTGCATGCTCATGGTCGTAGAGGATAGCGA
TGATGGACGGATCATCTTTGCGGATGATGTTGGGGATGTACTCGGCCTTTGGGACCTTTGATGAAATC
AGCATAACTTTAGGCGGGATGAAGTCCGTTCGGCTGCAAGATTCCAGCTCGATCTCCCTCTCCTCTACC
GCCTGCTCGCCG

> Locus_178211_Transcript_1/1_Confidence_1.000_Length_332
TAGAAACTTACTTTAGCTTTCGTCTTTATAAGTTATGGAAGGCTCGGCAGCATTCCAAACTTGTGGAT
GATTTTGATGAAATCTTATGAGTAAACGTCTGGTGTCGAAGTAAGAGATGGAACCACTGAAGGGGTT
TGAAGTTTGAAAAGGAGGGAAAAGGAGAGCGATTTTGCTCTTTATCCGTCCTCAATGAAGGCAGTTTT
GTGGTCATGGCGGCTGCTGTATGTGCTCTGATGAAATTCACACCCAGGCACAAGAGAAAGCCAAGCCT
CGGGCCTACGTTTCTCTGCCTCATGTCTCTTTTTAAGGTAAAATGCACATAGAATGCCCCG

> TRINITY_DN44340_c5_g1_i1
AATGTATCGTTAAACCCGCCACCACCCGACGCAGCCAATTAGTAAGATTTGTCACGTAAATAAAACGT
GTAAATCTGTTACTAGTACACAACAGCACACTGGTGTGTTGGAGAGTAAACCCGATATTGTGGGTATT
ATTGGTCTCGAAAGCCGAGGTATTAAACTGACGGTTTTGACAGATCTAGGCTGGCTGAGAAACACTGT
CTGTGAAAATTGGTGGCAACGCTAGGGCCAGCCACTGTTACATCCTGCACATTCGAGATCTGCTTTAA
CAATTGTAATGCTTCACGGGTGTAGTTCAATTGAATTATCTTTGCAGACATGGTTTCCAATTCATGAC
GCTACACTCAAAGTTAACCAACTGTAACCACAACGCATATTCAGACCTAACGACAGCATTTCTTTAAA
AAGCAACGCTGTAATGTGCTCAGCACCTAACACACACTTTTCGTGACTATGCAATAAAACGTTAAATT
CTGAAGTATTACTTAACAATAACCGTGTTACGTTCATAAGCATCATGGGAATGCTAGCTTGCCTAATA
TGGATAAAATTACGAAAAACAGGTATAGGTACAGGTGTATATATAGTACTGTATTTATTACCATAAT
TAGTATTTTGTGCTTTTCATTAACCATGTTTCAATAGAATATATAACACAAAACTTGCGCATGATGAC
TTCTGCCAGCAATCGTAAAAATATGCGGCCACAGAAAGTGTGTGATTACGGAGGATCTGTATACAAGA
AAGCACTCTGTCATAAACTCGAGTAGGCAATTAAAGCACAAAATAGATCATTATACATCACAGCTCCT
TCAAAGGCAAAAATCATCTCGCTAAGTTATGTACTTTTACAAAAGATCAGGCATTAATCTCAGGAGAC
TTGTACAAGTTATCCTGACAATGACACACTTCTGATGATGTTTAAATAACAATATAAACTGTTCTGCT
GCATTACATCAGCAGCAGACTTCATGACATTAAGAATTGCTAAATGACAAACTCGACCCATTAGGAAG
TCGAAGTGAAATGTGTGTACGATACAGTATATCAATATTGCCTACCAGTTAAGTGATCTCAATAAATA
ACCATGCCAAGCAATATTATGTTATGCTTAGTCTGTGTCTAATGAGAGACAAGTGTGATAAAGGAGA
GTATCTGCGTGTACGTGTTTACAACGATAGGCAATATAGTTTACAATGCACGTCTATAAAGATATTAA
ATGTTTAAACGGAGTTGCACATCAACTGCAGGTGAATGAAACAAACGTAACAATCAAATGATTTGCA
TCTGTTTTGTCGGCTCAACTGGGAGGACAACATGCCTTAGTTTTGCTGGTAAATATCGCTAAAGCGTA
ATATATATTTATTTTCCAGAGTAATAAACTAAATAAAGGCACATCCATTTCATAAGAGGTTTTTTTTT
TTTACAAGTCAACAGAACTATTGCCATGAGATATGTCATTGGGGCGGTGCCATTTTAGATGCACATCA
ACAATTTTGAAAAGTGAAAATTCTTAATTGCGCAAGACGTCTGTAACGGGGCTTAGATCGTTTTCGTT
CTTACAAACAACCGTATTTGTGCAATATCCATTCACTAAAATGTGCGGTATACAAATACTCCTACGTT
TTCCGTTAAAGTATTTTTAGTACGTAAAATACCAATGTATTTTTTTGTAAATGCGACACAAAATTAAA
ATGTTTCTGTCTGCAGCAACAACAGTAGTTTTTGTCTGTGTTAACGACAGTGTCTTAAGTGTAGTGTA
ACGTGTATTTAAAAACGGTGTAGCAAAATGTACATTAAACAGCACGTTTCAACACCATGCCTTATATG
CCACGTGAAAGTCAACCAGTGCCATCGTCTTTAAAACCATCAAAAGCAATCTACTGCTCACCAACTGT
GTGCTGGTATGCACAACGACCTTCCAAATATACTTCGTAGCCTTTAAAGGCCCCACTTTCCGGACGAT
GCTCTGAGTGGCACACAAACAAATGGCCGAGGAAGCTCCTTAAATGTTTTCACACTAAAAATCCCAGC
CGCCAATTGCTTTGGTGCAATGATGTTCAATACAAATTGCACATGCCTCCGGGGCATTCTATGTGCAT
TTTACCTTAAAAAGAGACACGAGGCAGAGAAACGTAGGCCTAAGGCTTGGCTTTCTCTTGTGCCTGGG

TGTGAATTTCATCAGAGCACATACAACAGCCGCCATGGCCACAAAACTGCCTTCATTGAGGACGGATA
CAGAGCAAAAGCGCTCTCCTTTTCCCTCCTTTTCAAACTTCAAACCCCTTCAGTGGTTCCATCTCTTGC
TTCGACACTGGACGTTTACTCATAAGATTTCGTCAAAATCATCCACAAGTTTGGAATGCTGCCGAGCC
TTCCATAACTTATAAAGACGAAAGCTAAAGTAAGTTTCTATCATCTTCTTGCCTTGAGCAGACAACTC
CAGTGGTGACTCAAAAGACACCCGATAGGGCCCCATTATAGACTTGAGGCTGTCCAACAGCTTCTCTC
CATTGTGATTGCCCAGCACGTAGCATCCCATAATGTGGACGACGTTGAGCTCGGTGTTGATCTTGCTC
ATCAAGCTTCCGAGGCACTTCCAGAACATAATTAAATCATCTTCTTTCTCCAGCTTGGCGAATGTTGC
CACCTTATTCTTCAACAGATACAAGTATCCGGGCCCTCCCTGGCAGTAAATCAGAAGCTTCCAGATTT
TGCAGCCTCTTCGATATTCATTCAGCTTCTTCTCTATCTCATCTAAAATGTCAGAAAAGGTTGCATGC
TCATGGTCGTAGAGGATAGCGATGATGGACGGATCATCTTTGCGGATGATGTTGGGGATGTACTCGGC
CTTTGGGACCTTTGATGAAATCAGCATAACTTTAGGCGGGATGAAGTCCGTTCGGCTGCAAGATTCGA
GCTCGATCTCCCTCTCCTCCTCCCAGCGGGCGCTGTTGGGCTCCCTATGCAGCCTCTCGTAGGCGCTCT
GCATGTCGTGACTGAATCGAGCATCGAGGACTGGGTCGGACTTGGACTGCGACACGTCGCTCAGAGGC
AGGGCCCGGCCACACTCTCTGCTTTCTTGACGGTCGAGGGACACTGCAGCTCTCCCGCGGTGGCGTCCC
ACCATCCGCAGATGTTTGTGGAAGTTCCTCTGCACCATCGCTGAAGGATCCAGTTCCCGCTTGCGTCTC
TTCCGCTCCGCGTATCCACGGAAGGCAGCGATGGCCTGCAACGTCGACGCAGTGCTGTGCAGCATGTGC
ATGCTCTCGCTGGGGAACCAGCTCCGGCACTCGATGTCGTCGGAGATGCGATCCATGCTGTACATGCG
CTCCAGCTTCTTCTTGCTGCTCAGCGCCTCCTGCTCGGCCTTCTCCTGCTCCAGGTCGAAGGAATGCTT
GAAGACGCCGAGGTTCTCGCAAGCCGGGCACTCCAGGGAGTCGCGCTCCCTCACGCTGGAGCCGCCGCC
GCCGCCGCCGTCCCGGCTGCTCCCGGCGCTGCTGCCGCCGCTGCTGCCGCTAAAGTGCG

S2.1.8 FGF8/17

> TRINITY_DN94963_c0_g1_i1
GTCTCCTGGTCCCAAGGCCCCCCGAAACCTCGCGACCCCCGTTCACGACGCCCAAGGTACCCGTTGGCA
CAAGGTTTTCACGTCGTCCGGGCCTCGCGTCCCCGTGGCCACATGGCCCCGCGGCCGTGTGGCGGCGTT
GGAGGCCTCATCCCCGCCGCTGGGCCCCGTGCCTCGACTGCCGGGTGCGCCTACTCCTGCGACCGCCTC
GCCCCGGGCCGCCATTACCGCCGCCGCCACCGCCGGGCCCGCTGCCCGGCCGCACCACCCAGAAGATGG
GCGTGACGGGCTCTTGCGACCGGCCCCGCGTCAGCCGCTTGATGAAGTGCACCTCCTGCTGGCCACGCA
TCGTGCGCGTGCCGTTCTTGGGACGCCCGTGCCGGTTGAAGGCCATGTACCACTCGCGGTGCTTGGCGC
TCTGCAGCGCCGTGTAGTTGTTCTCCAGGTAGATCTCGATGAAGACGCAGTCTCGGTTGCGCCCGTTG
AGCTTGCCCACCAGCTTGCCGCGCTTGTTCATGCAGATGTAGTGCCCCGTCTCAACCCCACGGAGCCGA
ATGTTCCCACCAAACGAGTCCGTCTCCGCCTGGAGCATGGCGTGCGGGTCTCCGTCGTCTCCGCGTGCC
TCGATGTTGCGGCCCAGCACCTGCACGTGCTTGCCGCTGGTGCGCGAGTAGAGTTGGTAGGTGCGCAC
GTGCCGCCGAGACGCCTCGTCCGTGTGCTGCGCCTCCTTCTGCACGTGGCGCCCGAAGTCCGGCGACGG
CGTGGGCGACACTAACTGAACTTGGAGGTAGAGCGTGAGTAACCGCGAACACAGCTGCAGGAGCCTCG
CCATGGCTTCGGGTTCCGTGCCGACGAGTTAAACTCCGACCCACAGGGCGCCAGGGAGACACGCGGCG
GAGACACGCGGTGGAGACGACACGCGGAGACGATCCGCCGACTCACAACCCCCGGACCCCCGCGTACTG
CACGCGGGAGC

> TRINITY_DN6733_c0_g1_i1
TCACATCATCCTCCGCTTGCCCGGCCGGTCGCATTGCAGCAAGAGGTTGCGCCGGCGAGGATTAACTAC
CCCGCTGTCTACATCGACTGCGTCCTGCGCCGCGGCCAGAGAGCGCCGCGCGTCGTTCAAGATGCGACT
CTCCCAGTACCGGCAGCCACTCGCGTCTCTATGCCTACACTTCCTGGTGCTCTGCTTTCAAGTGCAGGC
GGCGCCGGATTTTACGCAGCATGTGGAGGCGCAGCTCCAGCGACCGGACGCCGTGAGCCGCAAGCACA
TCCGCTCCTACCAGCTCTACAGCCGCACGAGCGGGAAACACGTGCAGATCGTAAACAAGCGCATCAAC
GCGCGCGCCGACGACGGCAATAAGTTCGCCAAACTGACCGTGGAAACGGACACGTTCGGCAGCCGCGT
ACGGATAAAAGGAGCAGAATCCGGCTACTACATCTGCATGAACAACAAGGGGAAGCTGGTGGGCAAG
AAGGAGGGCAAAGACACCGACTGCGTCTTCAAGGAGATCGTGCTGGAGAACAACTACACCGCGCTCGA
GTCGGTCATGTACGGCGGCTGGTACATGGGCTTCACGCGCAAGGGCCGGCCGCGCAAGGGCTCGCAGA
CGAGCCAGCACCAGCGCGAGGTCCACTTCATGAAACGCTTCCAGCGCACGTCGGAGGAACGCGAGCGC
AAGTTCATCCAGGTCGCCACGGGGGCCGCACGCCGCAGCAAGCGCATGCACGGGGGGGCCGCCGCCGCA
GCCGCAGCCGCCGGTGCCGACTCGCCCCGCCTGCGCCGAGCCTTCTACTACAATAAGTATGGTGGTGAC

GGCTCCCACATGGGCTCTGTGAGTGAGCACGGGGGGCACGGATACCCTGCCGGGTCAGAGCTCTTCCA
CGCCCCCAACTTGGGCGACGAGGAGTTTGAACTCCCGGCCATCGGCCCTCCGGGCGACTACGACGACTC
G

> Locus_48593_Transcript_1/1_Confidence_1.000_Length_612
CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGCGCTTCATGAAGTGGACCTCACGCTGGTGCTGGCT
CGTCTGCGAGCCCTTGCGCGGCCGGCCCTTGCGCGTGAAGCCCATATACCAGCCGCCGTACATGACCGA
CTCGAGCGCGGTGTAGTTGTTCTCCAGCACGATCTCCTTGAAGACGCAGTCCGTGTCTTTACCCTCCTT
CTTGCCCACCAGCTTCCCCTTTTTGTTCATGCAGATGTAGTAGCCGGATTCTGCTCCTTTTATCCGCAC
GCGGCTGCCGAACGTGTCCGTTTCCACGGTCAGTTTGGCGAACTTATTTCCGTCGTCGGCGCGCGCGTT
GATGCGCTTGCCCACGATCTGCACGTGCTTCCCGCTCGTGCGGCTGTAGAGCTGATAGGAGCGGATGT
GCTTGCGGCTCACGGCGTCCGGTCGCTGGAGCTGCGCCTCCACGTGCTGCGTAAAATCCGGCGCCGCCT
GCACTTGAAAGCAGAGCACCAGGAAGTGTAGGCATAGGGACGCGAGTGGCCGCCGGCAGTGCGAGGGT
CGCATCTTGGACGGCGCGCGGCGCTCTTCGCTGCGGCGCAGGACGCGGTCGATGGAGACAGC

> comp168772_c1_seq1
GTTGTTGCTGCTGCTGCTGCTGGTAGTAGGAGGTGGCGGCCGAGTACACCGCGGCCTCGTAGCCCGAG
TACGCGCCGCCCGTGTAGGTGACGGGCGTGGTGCTGTAGCTGGTGCTCTGCGTGTAGGAGGGCACCAC
GGCGGGCGCCGCCGCCACCACGGGCTGCACGGCGGAGCTCACGGGGTACACGGTGTAGGTGCTCGTCGC
CGTGCTCGGCGTCGCCTGCTTGATGGCCGTCACCTGGCGCGGCTGCTGCGTCTGGGTGAACTGCGGCGC
CGCCTGGCTGAAGCCGACCTTCTGCGCCGCCGACACGGACAGGCTCGATGTGATCGGCGGGATCACGCT
CTGATAGTACGTCTCGGCGCCGGTGTGCTGCGGCTGAGACACGGCCGGCTGTGGGTAGTATTGCTTGT
CATATCCGACGACCCCAGCCGACGAGCGCACGTAGGTGTAGGTCTCCTGGTTTCCCTGCCCGGGTCCTT
GCATCTCCCAAGGCAAAGGGAGGTGTGTGACATCATGCATGTGGTTCCTCCTATGGCCGATGACGGCA
GCCATGGCCTGCGGAATTAGCTGGTAGTTCTGCGTGGTGGTGGGCGGCGGCGGGGGCGGCGGAGCTTC
CTGCTGGCGCTGTGCGTACCCGTAGTCGGCCCCGTGGCTTGCTGCGGCGGCGGCAGCGGCGGCGGCCCC
CCCGTGCATGCGCTTGCTGCGGCGCGCGGCCCCCGTGGCGACCTGGATGAACTTGCGCTCGCGTTCCTC
CGACGTGCGCTGGAAGCGTTTCATGAAGTGGACCTCGCGCTGGTGCTGGCTCGTCTGCGAGCCCTTGCG
CGGCCGGCCCTTGCGCGTGAAGCCCATGTACCAGCCGCCGTACATGACCGACTCGAGCGCGGTGTAGTT
GTTCTCCAGCACGATCTCCTTGAAGACGCAGTCGGTGTCTTTGCCCTCCTTCTTGCCCACCAGCTTCCC
CTTGTTGTTCATGCAGATGTAGTAGCCGGATTCTGCTCCTTTTATCCGCACGCGGCTGCCGAACGTGTC
CGTTTCCACGGTCAGTTTGGCGAACTTATTGCCGTCGTCGGCGCGCGCGTTGATGCGCTTGTTTACGAT
CTGCACGTGCTTGCCGCTCGTGCGGCTGTAGAGCTGGTAGGAGCGGATGTGCTTGCGGCTCACGGCGT
CCGGTCGCTGGAGCTGCGCCTCCACGTGCTGCGTAAAATCCGGCGCCGCCTGCACTTGAAAGCAGAGCA
CCAGGAAGTGTAGGCATAGGGACGCGAGTGGCTGCCGGTACTGGGAGAGTCGCATCTTGGACGACGCG
CGGCGCTCTCTCGCCGCGGCGCAGGACGCAGTCGATGTAGACAGCGGGGTAGTTAATCCTCGCCGGCGC
AACCTCTTGCTGCAATGCGACCGGCCGGGCAAGCGGAGGATGATGTGATGATGAGGATGGGGGGTAGG
GTAAGGG

> comp168772_c1_seq2
CTCTCCCCCTCTCCACTCTCTCCTCCCCTCCCTCCTGCCGCGTCCGTCCTTCTTTCGCGCCGCGTGCCAC
GTGCAGCTCGCGCGCGCTCCCTTCCTCGCGCATCTACCGCGGGCGGGAAGCCAACGAAGACGACGACTC
GGACTTAGACGACGCAGACGACGATGACGACGATCCTGCGGCGGCGGCAGCGGCGGCGGCCCCCCCGT
GCATGCGCTTGCTGCGGCGCGCGGCCCCCGTGGCGACCTGGATGAACTTGCGCTCGCGTTCCTCCGACG
TGCGCTGGAAGCGTTTCATGAAGTGGACCTCGCGCTGGTGCTGGCTCGTCTGCGAGCCCTTGCGCGGCC
GGCCCTTGCGCGTGAAGCCCATGTACCAGCCGCCGTACATGACCGACTCGAGCGCGGTGTAGTTGTTCT
CCAGCACGATCTCCTTGAAGACGCAGTCGGTGTCTTTGCCCTCCTTCTTGCCCACCAGCTTCCCCTTGT
TGTTCATGCAGATGTAGTAGCCGGATTCTGCTCCTTTTATCCGCACGCGGCTGCCGAACGTGTCCGTTT
CCACGGTCAGTTTGGCGAACTTATTGCCGTCGTCGGCGCGCGCGTTGATGCGCTTGTTTACGATCTGCA
CGTGCTTGCCGCTCGTGCGGCTGTAGAGCTGGTAGGAGCGGATGTGCTTGCGGCTCACGGCGTCCGGT
CGCTGGAGCTGCGCCTCCACGTGCTGCGTAAAATCCGGCGCCGCCTGGTCAGTGAGTGGGGGAGGCAA
GGTCTCCATTACTTACACTCGCGCATGCACTCACACACGCACGCG

> comp168772_c1_seq4
GTTGTTGCTGCTGCTGCTGCTGGTAGTAGGAGGTGGCGGCCGAGTACACCGCGGCCTCGTAGCCCGAG
TACGCGCCGCCCCGTGTAGGTGACGGGCGTGGTGCTGTAGCTGGTGCTCTGCGTGTAGGAGGGCACCAC
GGCGGGCGCCGCCGCCACCACGGGCTGCACGGCGGAGCTCACGGGGTACACGGTGTAGGTGCTCGTCGC
CGTGCTCGGCGTCGCCTGCTTGATGGCCGTCACCTGGCGCGGCTGCTGCGTCTGGGTGAACTGCGGCGC
CGCCTGGCTGAAGCCGACCTTCTGCGCCGCCGACACGGACAGGCTCGATGTGATCGGCGGGATCACGCT
CTGATAGTACGTCTCGGCGCCGGTGTGCTGCGGCTGAGACACGGCCGGCTGTGGGTAGTATTGCTTGT
CATATCCGACGACCCCAGCCGACGAGCGCACGTAGGTGTAGGTCTCCTGGTAGTTCTGCGTGGTGGTG
GGCGGCGGCGGGGGCGGCGGAGCTTCCTGCTGGCGCTGTGCGTACCCGTAGTCGGCCCCGTGGCTTGCT
GCGGCGGCGGCAGCGGCGGCGGCCCCCCCGTGCATGCGCTTGCTGCGGCGCGCGGCCCCCGTGGCGACC
TGGATGAACTTGCGCTCGCGTTCCTCCGACGTGCGCTGGAAGCGTTTCATGAAGTGGACCTCGCGCTG
GTGCTGGCTCGTCTGCGAGCCCTTGCGCGGCCGGCCCTTGCGCGTGAAGCCCATGTACCAGCCGCCGTA
CATGACCGACTCGAGCGCGGTGTAGTTGTTCTCCAGCACGATCTCCTTGAAGACGCAGTCGGTGTCTT
TGCCCTCCTTCTTGCCCACCAGCTTCCCCTTGTTGTTCATGCAGATGTAGTAGCCGGATTCTGCTCCTT
TTATCCGCACGCGGCTGCCGAACGTGTCCGTTTCCACGGTCAGTTTGGCGAACTTATTGCCGTCGTCGG
CGCGCGCGTTGATGCGCTTGTTTACGATCTGCACGTGCTTGCCGCTCGTGCGGCTGTAGAGCTGGTAG
GAGCGGATGTGCTTGCGGCTCACGGCGTCCGGTCGCTGGAGCTGCGCCTCCACGTGCTGCGTAAAATCC
GGCGCCGCCTGGTCAGTGAGTGGGGGAGGCAAGGTCTCCATTACTTACACTCGCGCATGCACTCACAC
ACGCACGCG

> comp168772_c1_seq6
GTTGTTGCTGCTGCTGCTGCTGGTAGTAGGAGGTGGCGGCCGAGTACACCGCGGCCTCGTAGCCCGAG
TACGCGCCGCCCCGTGTAGGTGACGGGCGTGGTGCTGTAGCTGGTGCTCTGCGTGTAGGAGGGCACCAC
GGCGGGCGCCGCCGCCACCACGGGCTGCACGGCGGAGCTCACGGGGTACACGGTGTAGGTGCTCGTCGC
CGTGCTCGGCGTCGCCTGCTTGATGGCCGTCACCTGGCGCGGCTGCTGCGTCTGGGTGAACTGCGGCGC
CGCCTGGCTGAAGCCGACCTTCTGCGCCGCCGACACGGACAGGCTCGATGTGATCGGCGGGATCACGCT
CTGATAGTACGTCTCGGCGCCGGTGTGCTGCGGCTGAGACACGGCCGGCTGTGGGTAGTATTGCTTGT
CATATCCGACGACCCCAGCCGACGAGCGCACGTAGGTGTAGGTCTCCTGGTAGTTCTGCGTGGTAGTG
GGCGGCGGCGGGGGCGGCGGAGCTTCCTGCTGGCGCTGTGCGTACCCGTAGTCGGCCCCGTGGCTTGCT
GCGGCGGCGGCAGCGGCGGCGGCCCCCCCGTGCATGCGCTTGCTGCGGCGCGCGGCCCCCGTGGCGACC
TGGATGAACTTGCGCTCGCGTTCCTCCGACGTGCGCTGGAAGCGTTTCATGAAGTGGACCTCGCGCTG
GTGCTGGCTCGTCTGCGAGCCCTTGCGCGGCCGGCCCTTGCGCGTGAAGCCCATGTACCAGCCGCCGTA
CATGACCGACTCGAGCGCGGTGTAGTTGTTCTCCAGCACGATCTCCTTGAAGACGCAGTCGGTGTCTT
TGCCCTCCTTCTTGCCCACCAGCTTCCCCTTGTTGTTCATGCAGATGTAGTAGCCGGATTCTGCTCCTT
TTATCCGCACGCGGCTGCCGAACGTGTCCGTTTCCACGGTCAGTTTGGCGAACTTATTGCCGTCGTCGG
CGCGCGCGTTGATGCGCTTGTTTACGATCTGCACGTGCTTGCCGCTCGTGCGGCTGTAGAGCTGGTAG
GAGCGGATGTGCTTGCGGCTCACGGCGTCCGGTCGCTGGAGCTGCGCCTCCACGTGCTGCGTAAAATCC
GGCGCCGCCTGCACTTGAAAGCAGAGCACCAGGAAGTGTAGGCATAGGGACGCGAGTGGCTGCCGGTA
CTGGGAGAGTCGCATCTTGGACGACGCGCGGCGCTCTCTCGCCGCGGCGCAGGACGCAGTCGATGTAG
ACAGCGGGGTAGTTAATCCTCGCCGGCGCAACCTCTTGCTGCAATGCGACCGGCCGGGCAAGCGGAGG
ATGATGTGATGATGAGGATGGGGGGTAGGGTAAGGG

> comp168772_c1_seq1
GTTGTTGCTGCTGCTGCTGCTGGTAGTAGGAGGTGGCGGCCGAGTACACCGCGGCCTCGTAGCCCGAG
TACGCGCCGCCCCGTGTAGGTGACGGGCGTGGTGCTGTAGCTGGTGCTCTGCGTGTAGGAGGGCACCAC
GGCGGGCGCCGCCGCCACCACGGGCTGCACGGCGGAGCTCACGGGGTACACGGTGTAGGTGCTCGTCGC
CGTGCTCGGCGTCGCCTGCTTGATGGCCGTCACCTGGCGCGGCTGCTGCGTCTGGGTGAACTGCGGCGC
CGCCTGGCTGAAGCCGACCTTCTGCGCCGCCGACACGGACAGGCTCGATGTGATCGGCGGGATCACGCT
CTGATAGTACGTCTCGGCGCCGGTGTGCTGCGGCTGAGACACGGCCGGCTGTGGGTAGTATTGCTTGT
CATATCCGACGACCCCAGCCGACGAGCGCACGTAGGTGTAGGTCTCCTGGTTTCCCTGCCCGGGTCCTT
GCATCTCCCAAGGCAAAGGGAGGTGTGTGACATCATGCATGTGGTTCCTCCTATGGCCGATGACGGCA
GCCATGGCCTGCGGAATTAGCTGGTAGTTCTGCGTGGTGGTGGGCGGCGGCGGGGGCGGCGGAGCTTC
CTGCTGGCGCTGTGCGTACCCGTAGTCGGCCCCGTGGCTTGCTGCGGCGGCGGCAGCGGCGGCGGCCCC

CCCGTGCATGCGCTTGCTGCGGCGCGCGGCCCCCGTGGCGACCTGGATGAACTTGCGCTCGCGTTCCTC
CGACGTGCGCTGGAAGCGTTTCATGAAGTGGACCTCGCGCTGGTGCTGGCTCGTCTGCGAGCCCTTGCG
CGGCCGGCCCTTGCGCGTGAAGCCCATGTACCAGCCGCCGTACATGACCGACTCGAGCGCGGTGTAGTT
GTTCTCCAGCACGATCTCCTTGAAGACGCAGTCGGTGTCTTTGCCCTCCTTCTTGCCCACCAGCTTCCC
CTTGTTGTTCATGCAGATGTAGTAGCCGGATTCTGCTCCTTTTATCCGCACGCGGCTGCCGAACGTGTC
CGTTTCCACGGTCAGTTTGGCGAACTTATTGCCGTCGTCGGCGCGCGCGTTGATGCGCTTGTTTACGAT
CTGCACGTGCTTGCCGCTCGTGCGGCTGTAGAGCTGGTAGGAGCGGATGTGCTTGCGGCTCACGGCGT
CCGGTCGCTGGAGCTGCGCCTCCACGTGCTGCGTAAAATCCGGCGCCGCCTGCACTTGAAAGCAGAGCA
CCAGGAAGTGTAGGCATAGGGACGCGAGTGGCTGCCGGTACTGGGAGAGTCGCATCTTGGACGACGCG
CGGCGCTCTCTCGCCGCGGCGCAGGACGCAGTCGATGTAGACAGCGGGGTAGTTAATCCTCGCCGGCGC
AACCTCTTGCTGCAATGCGACCGGCCGGGCAAGCGGAGGATGATGTGATGATGAGGATGGGGGGTAGG
GTAAGGG

> comp168772_c1_seq7
GTTGTTGCTGCTGCTGCTGCTGGTAGTAGGAGGTGGCGGCCGAGTACACCGCGGCCTCGTAGCCCGAG
TACGCGCCGCCCGTGTAGGTGACGGGCGTGGTGCTGTAGCTGGTGCTCTGCGTGTAGGAGGGCACCAC
GGCGGGCGCCGCCGCCACCACGGGCTGCACGGCGGAGCTCACGGGGTACACGGTGTAGGTGCTCGTCGC
CGTGCTCGGCGTCGCCTGCTTGATGGCCGTCACCTGGCGCGGCTGCTGCGTCTGGGTGAACTGCGGCGC
CGCCTGGCTGAAGCCGACCTTCTGCGCCGCCGACACGGACAGGCTCGATGTGATCGGCGGGATCACGCT
CTGATAGTACGTCTCGGCGCCGGTGTGCTGCGGCTGAGACACGGCCGGCTGTGGGTAGTATTGCTTGT
CATATCCGACGACCCCAGCCGACGAGCGCACGTAGGTGTAGGTCTCCTGGTTTCCCTGCCCGGGTCCTT
GCATCTCCCAAGGCAAAGGGAGGTGTGTGACATCATGCATGTGGTTCCTCCTATGGCCGATGACGGCA
GCCATGGCCTGCGGAATTAGCTGGTAGTTCTGCGTGGTGGTGGGCGGCGGCGGGGGCGGCGGAGCTTC
CTGCTGGCGCTGTGCGTACCCGTAGTCGGCCCCGTGGCTTGCTGCGGCGGCGGCAGCGGCGGCGGCCCC
CCCGTGCATGCGCTTGCTGCGGCGCGCGGCCCCCGTGGCGACCTGGATGAACTTGCGCTCGCGTTCCTC
CGACGTGCGCTGGAAGCGTTTCATGAAGTGGACCTCGCGCTGGTGCTGGCTCGTCTGCGAGCCCTTGCG
CGGCCGGCCCTTGCGCGTGAAGCCCATGTACCAGCCGCCGTACATGACCGACTCGAGCGCGGTGTAGTT
GTTCTCCAGCACGATCTCCTTGAAGACGCAGTCGGTGTCTTTGCCCTCCTTCTTGCCCACCAGCTTCCC
CTTGTTGTTCATGCAGATGTAGTAGCCGGATTCTGCTCCTTTTATCCGCACGCGGCTGCCGAACGTGTC
CGTTTCCACGGTCAGTTTGGCGAACTTATTGCCGTCGTCGGCGCGCGCGTTGATGCGCTTGTTTACGAT
CTGCACGTGCTTGCCGCTCGTGCGGCTGTAGAGCTGGTAGGAGCGGATGTGCTTGCGGCTCACGGCGT
CCGGTCGCTGGAGCTGCGCCTCCACGTGCTGCGTAAAATCCGGCGCCGCCTGGTCAGTGAGTGGGGGA
GGCAAGGTCTCCATTACTTACACTCGCGCATGCACTCACACACGCACGCG

> comp168772_c1_seq8
CTCTCCCCCTCTCCACTCTCTCCTCCCCTCCCTCCTGCCGCGTCCGTCCTTCTTTCGCGCCGCGTGCCAC
GTGCAGCTCGCGCGCGCTCCCTTCCTCGCGCATCTACCGCGGGCGGGAAGCCAACGAAGACGACGACTC
GGACTTAGACGACGCAGACGACGATGACGACGATCCTGCGGCGGCGGCAGCGGCGGCGGCCCCCCCGT
GCATGCGCTTGCTGCGGCGCGCGGCCCCCGTGGCGACCTGGATGAACTTGCGCTCGCGTTCCTCCGACG
TGCGCTGGAAGCGTTTCATGAAGTGGACCTCGCGCTGGTGCTGGCTCGTCTGCGAGCCCTTGCGCGGCC
GGCCCTTGCGCGTGAAGCCCATGTACCAGCCGCCGTACATGACCGACTCGAGCGCGGTGTAGTTGTTCT
CCAGCACGATCTCCTTGAAGACGCAGTCGGTGTCTTTGCCCTCCTTCTTGCCCACCAGCTTCCCCTTGT
TGTTCATGCAGATGTAGTAGCCGGATTCTGCTCCTTTTATCCGCACGCGGCTGCCGAACGTGTCCGTTT
CCACGGTCAGTTTGGCGAACTTATTGCCGTCGTCGGCGCGCGCGTTGATGCGCTTGTTTACGATCTGCA
CGTGCTTGCCGCTCGTGCGGCTGTAGAGCTGGTAGGAGCGGATGTGCTTGCGGCTCACGGCGTCCGGT
CGCTGGAGCTGCGCCTCCACGTGCTGCGTAAAATCCGGCGCCGCCTGCACTTGAAAGCAGAGCACCAGG
AAGTGTAGGCATAGGGACGCGAGTGGCTGCCGGTACTGGGAGAGTCGCATCTTGGACGACGCGCGGCG
CTCTCTCGCCGCGGCGCAGGACGCAGTCGATGTAGACAGCGGGGTAGTTAATCCTCGCCGGCGCAACCT
CTTGCTGCAATGCGACCGGCCGGGCAAGCGGAGGATGATGTGATGATGAGGATGGGGGGTAGGGTAA
GGG

> comp137350_c0_seq3

CGCAGAAGCACAGCAGCCAACACGCAGGGTGCAGCTGCCCCTTGTGCGAGCGTGCAGTGCGTGCGCCT
TCCTGCTCTGGCACTATTTCCCCCCACCCCTTCTCCTCCTCCTCCTCCTCCACTCTCTCTCTCTCTGCAC
TCTCTCCCTCTCTCTCCTCCCCTCCCTCCTGCCGCGTTCGCCCCCTTCTTCCGCGCCGCGCGCCCGCGCA
CGCAGCGCGCGCGCTCCCTTCCTCGCGCGTCTACCGCGGGCGGGAAGCCAACGAAGACGACGACTCAGA
CGCAGACGACGACGCAGACGACGAGGACGACGATCCGGCGGCGGCGGCGGCCCCGTGCATGCGCTTGC
TGCGGCGCGCGGCCCCCGTGGCCACCTGGATGAACTTGCGCTCACGCTCCTCCGAGGTGCGCTGGAAGC
GCTTCATGAAGTGGACCTCACGCTGGTGCTGGCTCGTCTGCGAGCCCTTGCGCGGCCGGCCCTTGCGCG
TGAAGCCCATATACCAGCCGCCGTACATGACCGACTCGAGCGCGGTGTAGTTGTTCTCCAGCACGATCT
CCTTGAAGACGCAGTCCGTGTCTTTACCCTCCTTCTTGCCCACCAGCTTCCCCTTTTTGTTCATGCAGA
TGTAGTAGCCGGATTCTGCTCCTTTTATCCGCACGCGGCTGCCGAACGTGTCCGTTTCCACGGTCAGTT
TGGCGAACTTATTTCCGTCGTCGGCGCGCGCGTTGATGCGCTTGCCCACGATCTGCACGTGCTTCCCGC
TCGTGCGGCTGTAGAGCTGATAGGAGCGGATGTGCTTGCGGCTCACGGCGTCCGGTCGCTGGAGCTGC
GCCTCCACGTGCTGCACTTGAAAGCAGAGCACCAGGAAGTGTAGGCATAGGGACGCGAGTGGCCGCCG
GCAGTGCGAGGGTCGCATCTTGGACGGCGCGCGGCGCTCTTCGCTGCGGCGCAGGACGCGGTCGATGG
AGACAGCGGGGTCGTTAATCCTCGCCGGCGCAACCTCTTGCTGCAATGCAACCGGCCGGGCAAGCGGA
GGATGATGTGATGATGAGGATGGGGGGTAGG

> comp137350_c0_seq2
CGCAGAAGCACAGCAGCCAACACGCAGGGTGCAGCTGCCCCTTGTGCGAGCGTGCAGTGCGTGCGCCT
TCCTGCTCTGGCACTATTTCCCCCCACCCCTTCTCCTCCTCCTCCTCCTCCACTCTCTCTCTCTCTGCAC
TCTCTCCCTCTCTCTCCTCCCCTCCCTCCTGCCGCGTTCGCCCCCTTCTTCCGCGCCGCGCGCCCGCGCA
CGCAGCGCGCGCGCTCCCTTCCTCGCGCGTCTACCGCGGGCGGGAAGCCAACGAAGACGACGACTCAGA
CGCAGACGACGACGCAGACGACGAGGACGACGATCCGGCGGCGGCGGCGGCCCCGTGCATGCGCTTGC
TGCGGCGCGCGGCCCCCGTGGCCACCTGGATGAACTTGCGCTCACGCTCCTCCGAGGTGCGCTGGAAGC
GCTTCATGAAGTGGACCTCACGCTGGTGCTGGCTCGTCTGCGAGCCCTTGCGCGGCCGGCCCTTGCGCG
TGAAGCCCATATACCAGCCGCCGTACATGACCGACTCGAGCGCGGTGTAGTTGTTCTCCAGCACGATCT
CCTTGAAGACGCAGTCCGTGTCTTTACCCTCCTTCTTGCCCACCAGCTTCCCCTTTTTGTTCATGCAGA
TGTAGTAGCCGGATTCTGCTCCTTTTATCCGCACGCGGCTGCCGAACGTGTCCGTTTCCACGGTCAGTT
TGGCGAACTTATTTCCGTCGTCGGCGCGCGCGTTGATGCGCTTGCCCACGATCTGCACGTGCTTCCCGC
TCGTGCGGCTGTAGAGCTGATAGGAGCGGATGTGCTTGCGGCTCACGGCGTCCGGTCGCTGGAGCTGC
GCCTCCACGTGCTGCGTAAAATCCGGCGCCGCCTGCACTTGAAAGCAGAGCACCAGGAAGTGTAGGCA
TAGGGACGCGAGTGGCCGCCGGCAGTGCGAGGGTCGCATCTTGGACGGCGCGCGGCGCTCTTCGCTGC
GGCGCAGGACGCGGTCGATGGAGACAGCGGGGTCGTTAATCCTCGCCGGCGCAACCTCTTGCTGCAAT
GCAACCGGCCGGGCAAGCGGAGGATGATGTGATGATGAGGATGGGGGGTAGG

> comp163850_c0_seq1
CGACGGGCTCCAGCGACCGGCCCCGCGTCAGCCGCTTGATGAAGTGCACCTCCTGCTGGCCGCGCATCG
TGCGCGTGCCGTTCTTGGCACGCCCGTGCCGGTTGAAGGCCATGTACCACTCGCGGTGCTTGGCGCTCT
GCAGCGCCGTGTAGTTGTTCTCCAGGTAGATCTCGATGAAGACGCAGTCTCGGTTGCGCCCGTTGAGC
TGCCGGGAAGCGGG

S2.2 Plasmid maps for lamprey *in situ* probes

The electronic files are attached in the Additional Material and can be visualized with SnapGene viewer
Files names:

- GNRH-I_lamprey_in_situ.dna
- GnRH-II_lamprey_in_situ.dna
- GnRH-III_lamprey_in_situ.dna
- ISLA_lampey_in_situ.dna
- ISLB_lamprey_in_situ.dna
- ISLC_lamprey_in_situ.dna
- NELF_lamprey_in_situ.dna
- FGF8_17_lamprey_in_situ.dna

S2.3 Sequence references and alignments used for maximum-likelihood phylogenetic tree

S2.3.1 GnRH

>GnRH1_Human_(Homo sapiens)_(NP_000816.4)
KPIQKLLAGLILLTWCVEGCSSQHWSYGLRPGGKRDAENLIDSFQEKEVG-
QRFECTTHQPSPLRDLKGAESLIEEETGQKK
>GnRH1_Panda_(Ailuropoda melanoleuca)_(XP_002914474.2)
EPIPKLIAGLLLLTLCVVGCSGQHWSYGLRPGGKRNAEKLIDSFQEKELD-
QHLECTIHQPTPLRDLKGAESLIEEENGQKR
>GnRH1_Rat_(Rattus norvegicus)_(NP_036899.1)
ETIPKLMAAVVLLTVCLEGCSSQHWSYGLRPGGKRNTEHLVDSFQEKEED-
QNFECTVHWPSPLRDLRGAERLIEEEAGQKK
>GnRH1_Koala_(Phascolarctos cinereus)_(AJT59745.1)
ELTQKLVAGCLLLTVCVTISSGQHWSYGLRPGGKRDADNLIDSFQEDEGN-
QRFECTIHQPSPLRDLKGVASLIEGEAGRKK
>GnRH1_Gallus_gallus_(NP_001074346.1)
EKSRKILVGVLLETASVAICLAQHWSYGLQPGGKRNAENLVESFQENEME-
QKAECPGSYQPRLSDLKETASLIEGEARRKE
>GnRH1_Pelican_(Pelecanus crispus)_(XP_009482176.1)
EKSRKIFVSILLEVMSVEICLAQHWSYGLQPGGKRNAENLVESFQENEME-
QKTECPGLRQSRFSDLKEAESLIEGEARRKK
>GnRH1_Duck_(Anas platyrhynchos)_(XP_012959602.1)
QKSRKAFVGILLFIVSVEICLAQHWSYGLQPGGKRNVDNLGELFQENDME-
QKTECPGSYQPQFTDLKEAASLIEGEARRKK
> >GnRH1_Alligator_(Alligator mississippiensis)_(KQL69429.1)
QKTRKVFVSLLLLILSVDICLAQHWSYGLQPGGKRNAENVVESFQQSDME-
QQFECSGPHQSKLSDLKKAASLIEGEAGRKK
>GnRH1_Turtle_(Chrysemys picta bellii)_(XP_005285222.1)
EKTRKLEVRELMFILSVEICLAQHWSYGLQPGGKRDAENLVESFQESEME-
QHFECTGPHQSMLSGLKGAASLIEGDAGRKK
>GnRH1_Gecko_(Eublepharis macularius)_(ABB89899.1)
-------GSLLFLLLCVAIGSAQHWSYGLQPGGKRDAENLIESFQENEVD-
QHLECTASQQPTLQGLKGAASLIDRETGQKK
>GnRH1_Sturgeon_(Acipenser sinensis)_(AGK30598.1)
AVSRGAFVWLLSLMAVSEVCYGQHWSYGLRPGGKRETETLLDTLQE-DIE-
DHSECALSSQSQLSDLKGVARLVGGESARKK
>GnRH1_Gar_(Lepisosteus oculatus)_(XP_006625369.1)
KAQKSSLFWLVVAMTLVTQACSQHWSYGLRPGGKREVESLQDTLQDEEVR-
RQPGCADVSPSRLSSLRELASLAEEERGRKN
>GnRH1_Eel_(Anguilla anguilla)_(ADD92012.1)

MADKSALLWLGLAVALVCQGCCQHWSYGLRPGGKRGADSLQDTLQDEELQ-
SLPSCNDLSPITLSSLKEIANLADRETGRKN
>GnRH1_Whitefish_(Coregonus clupeaformis)_(AAP57221.1)
EEKKVLLLLLLLVVALVSQGCCQHWSYGMNPGGKRATGSLSDTQDNEDLLCSLFGCADVSPAKMYRLRALASLAD
RQSGLNN
>GnRH1_Medaka_(Oryzias latipes)_(BAB16303.1)
MVVKTWMPW-LLVSSVLSQGCCQHWSFGLSPGGKRELKYFPNTLENR-
LLCSDLSHLEESSAKIYRIKGLGSVTEAKNGYRT
>GnRH1_Seabream_(Dicentrarchus labrax)_(AAF62898.1)
MAAQTFALRLLLLGTLLGQGCCQHWSYGLSPGGKRELDGLSETLGNGSFPCRVLGCAEESPPKIYRMKGFDAVTDR
ENGNRT
>GnRH1_Chanchita_(Cichlasoma dimerus)_(AKG03158.1)
MAAKILALWLLLAGTAFPQGCCQHWSYGLSPGGKRDLDTESDALGNEEFPCSVEGCAEESPAKMYRVKGLGSVTE
RENGHRT
>GnRH1_Tilapia_(Oreochromis niloticus)_(BAC56849.1)
MAAKILALWLLLAGTVFPQGCCQHWSYGLSPGGKRDLDNESDTLGNEEFPCSVEGCAEESPAKMYRVKGLASLAE
TDTGHSR
>GnRH1_Mackerel_(Scomber japonicus)_(ADP89591.1)
MAMQTLALWLLLLGSVVPQVCCQHWSYGLSPGGKRELDSLSDTMDDEGFPCSFLGCAEESPAKIYRMKGLGSVTN
RENGHRN
>GnRH1_Pejerrey_(Odontesthes bonariensis)_(AAU94309.2)
MAVRTWALWPLLVGSVLLQVTCQHWSFGLSPGGKRDLDTFSDTLDNEGFPCRVVGCADESPAKIYRMKGFGGVT
DRENGRRV
>GnRH1_Xenopus_(Xenopus tropicalis)_(NP_001107165.1)
KAISTYALLLLVLLFSAHVGHAQHWSYGLRPGGKRDAESLQDMYPENEVP-ERLECSV--
PSRLNVLRGAMSWLDGEN-RKK
>GnRH1_Bullfrog_(Rana catesbeiana)_(090V63.1)
RHVTVVLLLAIVLLLSSHMIHGQHWSYGLRPGGKREVESLQESYAENEVS-QHLECSI--
PNRISLVRDAMNWLEGENARKK
>GnRH1_Catshark_(Scyliorhinus canicula)_(MH468810)
---KLLVCFALGSAIFVNELSAQHWSFDLRPGGKREADDLVESFQEGNVDCPFPDC----------LRGTAKFTP----RRK
>GnRH1_Anchovy_(Engraulis japonicas)_(AF070217.1)
-RSKGALVCLLLVTAAVLQCSSQHWSHGLSPGGKREADSPSESQVMEGLPRGGARCGSDTRERPSTLEQ-ISLMS----
-RE
>GnRH1a_Elephant shark_(Callorhinchus milli)_(SRA054255)
ALGKRLLWLSLTLAVLTALTSAQHWSIDNRPGKKRGTEHMIEFLQGGEVEVELPECSGDNP----------
GKMVRKN
>GnRH1b_Elephant shark_(Callorhinchus milli)_(ND)
VLGKRLLWLVLILAVLTALTSAQHWSIDNRPGRKRGTEHMIEFLQGGEVEVELPECPGDKP----------
RKMVRKN
>GnRH2_Medaka_(Oryzias latipes)_(BAB16300.1)
-MSRLVLLLGVLLYVGAQLSQAQHWSHGWYPGGKRELDSF---
EVSEEMKCETGECSYMRPQRRSFLRNILDALARELQKRK
>GnRH2_Chanchita_(Cichlasoma dimerus)_(ADV31310.1)
CVSRLVLLLGLLLCVGAQLSFAQHWSHGWYPGGKRELDSFGTSEISEEIKCEAGECSYLRPQRRGILRNILDALAREL
QKRK
>GnRH2_Tilapia_(Oreochromis niloticus)_(BAC56850.1)
CVSRLALLLGLLLCVGAQLSFAQHWSHGWYPGGKRELDSEGTSEISEEIKCEAGECSYLRPQRRSILRNILDALAREL
QKRK
>GnRH2_Seabream_(Dicentrarchus labrax)_(Q9IA08.1)
CVSRLVLLFGLLLCVGAQLSNAQHWSHGWYPGGKRELDSEGTSEISEEIKCEAGECSYLRPQRRSVLRNILDALARE
LQKRK
>GnRH2_Mackerel_(Scomber japonicus)_(ADP89592.1)
CVSRLVLLLGLLLCVGAQLSNAQHWSHGWYPGGKRELDSFGTPEISEEIKCEAGECSYLRPQRRSFLRNILDALARE
LQKRK
>GnRH2_Pejerrey_(Odontesthes bonariensis)_(AAU94307.1)
-
MSRLVLLLVLLLYVGAQLSYAQHWSHGWYPGGKRELDSFSTSEISEENKCEAGECSYLRPQRQNVLRNILDALARE
LQKRK

>GnRH2_Whitefish_(Coregonus clupeaformis)_(AAP57219.1)
SVARLVFMLGLLLCLGAQLSSSQHWSHGWYPGGKRELDSFTTSEISEEIKCEAGECSYLRPQRRNILKNILDALARE
FEKRK
>GnRH2a_Salmon_(Salmo salar)_(XP_013987737.1)
SVARLVFMLGLLLCLGAQLSSSQHWSHGWYPGGKRELDSFTTSEISEEIKCEAGECSYLRPQRRNILKNILDVLARE
FQKRK
>GnRH2b_Salmon_(Salmo salar)_(NP_001134611.1)
SVARLVLMLGLLLCLGAQLSSSQHWSHGWYPGGKRELDSFTTSEISEEIKCEAGECSYLRPQRRNILRNILDALARE
FEKIK
>GnRH2_Eel_(Anguilla anguilla)_(ADD92005.1)
NTGRLVLILGVLLCLGAQLSLCQHWSHGWYPGGKRELDSLTTAEVLDEIKCDGGECSYLRPQRKSLLKNILDALAR
EFQRKR
>GnRH2_Goldfish_(Carassius auratus)_(042471.1)
HICRLFVVMGMLMFLSVQFASSQHWSHGWYPGGKREIDVYDPSEVSEEIKCNAGKCSFLIPQGRNILKTILDALTR
DFQKRK
>GnRH2_Zebrafish_(Danio rerio)_(AAU43784.1)
LVCRLLLVMGLMLCLSAQLSSAQHWSHGWYPGGKREIDLYDTSEVSEEVKCEAGKCSYLRPQGRNILKTILDALIR
DFQKRK
>GnRH2_Anchovy_(Engraulis japonicus)_(AF070218.1 GnRH2)
CGYRWVLLAAVLLFLGVELSGAQHWSHGWYPGGKRDVDTENSAQVSEEIKCEAGECSYLRPQRRNLLKSILEALT
REFQRRK
>GnRH2_Arowana_(Scleropages jardinii)_(BAB72183.1)
CVGRLTLLLGILLCSGAQLSCSQHWSHGWYPGGKRELNSLTASEVSGKIKCEDRKCSYLRPQQKNIL-
TIVDASTREFRGKR
>GnRH2_Sturgeon_(Acipenser sinensis)_(AGK30597.1)
CQGKLLVLLAVLLALSAQLSSGQHWSHGWYPGGKRELEGLQSPEDSDEVKCDGDECSYLRHPRKNILRSIADMLT
RQMQRKK
>GnRH2_Coelacanth_(Latimeria menadoensis)_(ABZ04537.1)
CQRSLVILLEVLLAVSIQLCSTQHWSHGWYPGGKRELAIPQTPEVSEEIKCDGEECTYLRSPRKSILKEIADIIAWQIQ
KKK
>GnRH2_Bullfrog_(Rana catesbeiana)_(AAL05971.1)
CQRHLLFLLLVLFAVSTQLSHGQHWSHGWYPGGKRELDMPASPEVSEEIKCEGEECAYLRNPRKNLLKNIADVLA
RQLQKK-
>GnRH2_Turtle_(Pelodiscus sinensis)_(XP_006139278.1)
CQRPFLLLLLVVLAVSTHLSRAQHWSHGWYPGGKRELDLSQAPEASEEIKCDGEACAYLRSPRKTIVNTLADLLAR
QLQKKK
>GnRH2_Alligator_(Alligator sinensis)_(XP_006020952.1)
CPRSLLLLLVLLLAIGVPLARAQHWSHGWYPGGKRELDLSQAPQASEEIKCGGEECAYVRSPRMNVVKTLADMLA
RQLQKKK
>GnRH2_Gegko_(Eublepharis macularius)_(BAC99084.1)
CHRPLLLFLCIMIIATIHLSKAQHWSHGWYPGGKREVDLSQSPEVSEDIKCDGDDCTYLKIPREKIVTSLADLLAKH
LQKKK
>GnRH2_Elephant shark_(Callorhinchus milii)_(XP_007896209.1)
LQRNLLLLLLVLLAINTQVSRAQHWSHGWYPGGKRELGQAQTPEVSEVFQCEGDDCAFVRSPRTNLFRSIADLVA
GRFQKKK
>GnRH2_Catshark_(Scyliorhinus canicula)_(MH468811)
FQRNALFLIFLLLIVNTQFSRAQHWSHGWYPGGKRELSLSQSPEVSEEIKCRGDGCLFLGSPRKDVIRSITDMLMQQ
IQKKK
>GnRH2_Whale shark_(Rhincodon typus)_(ND)
FQRNLHFLVFLLLIVNTEFSTAQHWSHGWYPGGKREVSLSQSPDASEEIKCQGEGCLLLRSPRRGIIRSIMDMLVQQ
IQKKK
>GnRH2_Xenopus_(Xenopus tropicalis)_(NP_001107550.1)
CQGHLVLLLIVLFAFSTHLSNAQHWSHGWYPGGKRQLDTRSIPEISDELKCEGESCDYPMN-
EMSILKGLTRFLFPRERQRK
>GnRH2_Koala_(Phascolarctos cinereus)_(AJT59746.1)
CLRP--LLLLGLLVLWTQISYAQHWSHGWYPGGKRALDEIPGLEASEEGKWDGGE--------
RSLLKTLADVLAQQQQK--
>GnRH-I_Lamprey_Petromyzon_marinus_ (AAF78456.1)

LRGQSLTLLLLATALLVSLNYAQHYSLEWKPGGKRDLEVSHTRELEQELECDGPECAFSRVPNTKLIRELSYLSQRN
YDRKK
>GnRH-I_Lamprey_Lethenteron camtschaticum
LRGQSLTLLLLATALLVSFNDAQHYSLEWKPGGKRDLEVSHTRELEQELECDGPECAFSRVPNTKLIRELSYLSQR
NYDRKK
>GnRH-I_Lamprey_Lampetra_planeri
----------HYSLEWKPGGKRDLEASRTREPEQELECDGPECAFSRVPNTKLIRELNYLSQRNYERKK
>GnRH-III_Lamprey_Petromyzon_marinus_(AAL12249.1)
LRGQSLVLLLLASALLVSLTHTQHWSHDWKPGGKRDLEAMRPLLE-
EELECDGPECAFARVPTGELVREISYLSQKNYQRKK
>GnRH-III_Lamprey_Lampetra_planeri
LRGQSLALLLLASALPVSLTHTQHWSHDWKPGGKRDLEAMRPLL--
EELECDGPECAFARVPSSELVREISYLSQKNYQRKK
>GnRH-III_Lamprey_Lethenteron camtschaticum
LRGQSLALLLLASALLVSLTHTQHWSHDWKPGGKRDLEAMRPLLESRSLRPNAPSLECRAVSSSGR----
SYLSQKNYQRKK
>GnRH2_Chicken_(Gallus gallus)_(BAE80719.1)
-------CLLLALLLLAGTAQGQHWSHGWYPGGKRDLSAPQVPAAL----CPTPPCRPLPP-
MPSTLRAAWRPLEAALRQH-
>GnRH2_Human_(Homo sapiens)_(NP_847901.1)
ASSRRGLLLLLLLTAHLGPSEAQHWSHGWYPGGKRALSSAQDPQNAAGSPSDALAPDDSMPRKRHLARTLTAAR
ERPAPSSK
>GnRH2_Panda_(Ailuropoda melanoleuca)_(XP 011222695.1)
ASCRLG--
FLILLTVHPGSLKAQHWSHGWYPGGKRASSSAQHPQRAASSPSNALAPENSVPQKQHLVKTLTGRRARPVAQ--
>GnRH2_Sheep_(Ovis aries)_(XP_012044160.1)
ASFGLGLLLLLLLTTHPGPSKAQHWSHSXYPGGKRASSLPRDPQHPAQSPSDALAWEDSVPGKQHLVQTLVSKVE
HPWPQRE
>GnRH-II_Lamprey_(Petromyzon_marinus)_ (ABE66462.1)
GRASLSLVLILWLLTAPPASLGQHWSHGWFPGGKRGVQEPPRASYESDGSCCSPGCPTF----------SQVLE------S
>GnRH-II_Lamprey_Lethenteron camtschaticum
GRVSLSLVLILWLLTAPPASLGQHWSHGWFPGGKRGVQEPPRPSYESDAS--------------------------------
>GnRH-II_Lamprey_Lampetra_planeri

LLMAPPASLGQHWSHGWFPGGKRSVQEPPRPSYESDAPRAGAECPVIADPRIHLLRDVVELMERNSQKRE
>GnRH3_Goldfish_(Carassius auratus)_(BAB18904.1)
EWNGRLLVQLLMLVCVLEVSLCQHWSYGWLPGGKRSVGEVEATFKM----MDAGDAVLSIPSP---
MEQLNEVDADGLPKER
>GnRH3_Zebrafish_(Danio rerio)_(NP_878307.2)
EWKGRLLVQLLLLVCVLEVSLCQHWSYGWLPGGKRSVGEMEATFRM----LDPGDTVLSIPSP---
MEQLNEVDAEGLPKGR
>GnRH3_Medaka_(Oryzias latipes)_(NP_001098142.1)
DVSSKVVVQVLLLALVVQVTLCQHWSYGWLPGGKRSVGELEATIRM----
MGTGRVVSLPEASAQTQERLNDGSTYFD-RKR
>GnRH3_Seabream_(Dicentrarchus labrax)_(Q9IA09.1)
EANSRVMVRVLLLALVVQVTLSQHWSYGWLPGGKRSVGELEATIRM----
MGTGEVVSLPEASAQTQERLNDDSSHFD-RKR
>GnRH3_Tilapia_(Oreochromis niloticus)_(XP_013126782.1)
EAGSRVIMQVLLLALVVQVTLSQHWSYGWLPGGKRSVGELEATIRM----
MGTGEVVSLPDANAQTQERLNDDSSHFD-RKK
>GnRH3_Mackerel_(Scomber japonicus)_(ADP89593.1)
EASSRVTVQVLLLALVVQVTLSQHWSYGWLPGGKRSVGELEATIRM----
MGTGGVVSLPEASAQTQEGLNDDSSHFD-RKR
>GnRH3_Chanchita_(Cichlasoma dimerus)_(ADV31311.2)
EASTRVAMQVLLLALVVQVTLSQHWSYGWLPGGKRSVGELEATIRM----
MGTGGVVSLPEASAQTQERLNDDSSHFD-GKR
>GnRH3_Pejerrey_(Odontesthes bonariensis_(AAU94308.1)
EASSRVMVQVLLLALVVQVSLCQHWSYGWLPGGKRSVGELEATIRM----
MGTGGVVSLPEASAQIQERFNDDSSHLDTRKK

>GnRH3_Whitefish_(Coregonus clupeaformis)_(AAP57220.1)
DLSSRTVVQVVVLVLVAQVTLSQHWSYGWLPGGKRSVGELEATIRM----MDTGGEVALPETSAHVSERLS---------
KW
>GnRH3a_Salmon_(Salmo salar)_(XP_014009933.1)
DLSNRTVVQVVVLALVAQVTLSQHWSYGWLPGGKRSVGELEASIKM----MDTGGVVALPETSAHVSERLL---------
KW
>GNRH3b_Salmon_(Salmo salar)_(XP_014062302.1)
DLSSKTVVQVVMLALIAQVTFSQHWSYGWLPGGKRSVGELEATIRM----MDTGGVMVLPETGAHVPERLS---------
KR
>GnRH3_Arowana_(Scleropages jardinii)_(BAB72182.1)
ELTGKSVLHLVVLAHVAQIGFSQHWSYGWLPGGKRSTGDTEAKVKM----
MDSGDLVTFFEASPFVPESLSEDGGEFTRKRW
>GnRH3_Anchovy_(Engraulis japonicus)_(AF070219.1)
EQGR--LVLLLVLACACKECVCQHWSYGWLPGGKRSIGELEATFRM----MDAGDTLIPLT-----
AEKLIDDIEENAVRRR
>GnRH3_Catshark_(Scyliorhinus canicula)_(MH468812)
EVTKIVVHFLIAIVFTAHGCISQHWSHGWLPGGKRNAVSMDAYLEMEDIIFEIPKY-----
QKMNSPPAYPDISDRKFQEKK
>GnRH3_Whaleshark_Rhincodon_typus
EVTKTIIHFLIAVMFIAHGCISQHWSHGWLPGGKRSAVSMDAYLEMEDVIFEIPRY-----
QRANNPQAIPDLNDRKIPKKK
>GnRH-like_Amphioxus_(Branchiostoma floridae)_(AHE40598.1)
-AARLPALLAV--
LLLAQILCARAFTYTHTWGRKRADSSELLTPHADSVSYDASEGSEVTKMAVRTLFRIGDYLQKRTNQN-

S.2.3.2 ISL

>ISL1_Human_(Homo sapiens)_(NP_002193.2)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Sheep_(Ovis aries)_(XP_004017051.2)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Mouse_(Mus musculus)_(NP_067434.3)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Killer_Whale_(Orcinus orca)_(XP_004265987.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Tasmanian_Devil_(Sarcophilus harrisii)_(XP_003759399.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Alligator_(Alligator mississippiensis)_(XP_006262621.2)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Green_Sea_Turtle_(Chelonia mydas)_(XP_007052879.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Painted_Turtle_(Chrysemys picta bellii)_(XP_008166937.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Common_Lizard_(Zootoca vivipara)_(XP_034956755.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCNIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Gekko_(Gekko japonicus)_(XP_015263958.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCNIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Wall_Lizard_(Podarcis muralis)_(XP_028604472.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCNIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Platypus_(Ornithorhynchus anatinus)_ (XP_028913171.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV

RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMAGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Panda_(Ailuropoda melanoleuca)_(XP_002919684.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Coelacanth_(Latimeria chalumnae)_(XP_005999085.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCNIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASIGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVSSPIEA

>ISL1_Great_Tit_(Parus major)_ (XP_033367233.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARAKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Starling_(Sturnus vulgaris)_(XP_014739774.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARAKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Spotted_Gar_(Lepisosteus oculatus)_(XP_015219661.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFV
RDGKTYCKRDYIRLYGIKCAKCNIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERATMGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSILM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVSFSEGGPGSNSTGSEVASMSSQLPDTPNSMVSSPIEA

>ISL1A_Zebrafish_(Danio rerio)_(NP_571037.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFV
RDGKTYCKRDYIRLYGIKCAKCNIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERATMGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSILM
KQLQQQQPNDKTNIQGMTGTPMVATSPERHDGGLQANQVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Medaka_(Oryzias latipes)_(NP_001295943.1)
MGDMGDPPKKKRLVSLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECSQYLDESCTCFV
RDGKTYCKRDYIRLYGIKCAKCNIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYNAKPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSLLM

KQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGIQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVSFSEGGPGSNSTGSEVASMSSQLPDTPNSMVSSPIEA

>ISL1_Tilapia_(Oreochromis niloticus)_(XP_003457079.1)
MGDMGDPPKKKRLVSLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECSQYLDESCTCFV
RDGKTYCKRDYIRLYGIKCAKCNIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGGDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSQLM
KQLQQQQSSDKTNIQGMTGTPMVAASPERHDGGIQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVSFSEGGPGSNSTGSEVASMSSQLPDTPNSMVSSPIEA

>ISL1_Toad_(Bufo bufo)_(XP_040276102.1)
MGDMGDPPKKKRLMSLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGIKCAKCSLGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASLGSDPLSPLHARPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSILM
KQLSQQQPNDKTNIQGMTGTPMVASSPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDI
DQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA

>ISL1_Thorny_Skate_(Amblyraja radiata)_(XP_032886821.1)
MGDMGDTPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGTKCAKCNIGFSKNDFVMRARNKVYHIDCFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASVGPDPLSPMHNRPLQMAEPISVRQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTSLSPRVIRVWFQNKRCKDKKRSVLM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDSSLQANPVEVQSYQPPWKVLSDFALQSDI
EQPAFQQLVSAS------NMQNTITTTTIQTLPDTHQSIYRKIAES

>ISL1_Whale_Shark_(Rhincodon typus)_(XP_020375890.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGTKCAKCNIGFSKNDFVMRARSKVYHIDCFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASVGPDPLSPMHNRPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDSSLQANPVEVQSYQPPWKVLSDFALQSDI
EQPAFQQLVSAS------NMQNTITTTTIQPAPETHQSIYRKIAES

>ISL1_Ghostshark_(Callorhinchus milii)_(XP_007890395.1)
MGDMGDPPKKKRLISLCVGCGNQIHDQFILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLYGTKCAKCNIGFSKNDFVMRARSKVYHIDCFRCVACSRQLIPGDEF
ALREDGLFCRADHDVVERASVGPDPLSPMHNRPLQMAEPISARQPALRPHVHKQPEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKTILM
KQLQQQQPNDKTNIQGMTGTPMVAASPERHDSSLQANPVEVQSYQPPWKVLSDFALQSDI
EQPAFQQLVSAS------NMQNTITITT-QTAPETHQSLYRKVAES

>ISL2_Human_(Homo sapiens)_(NP_665804.1)
LGAMGDHSKKKPGTAMCVGCGSQIHDQFILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAKCQVGFSSSDLVMRARDSVYHIECFRCSVCSRQLLPGDEF
SLREHELLCRADHGLLLERAAAGSPRSPLPARGLHLPDAGSGRQPALRPHVHKQTEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHSDKTSLQGLTGTPLVAGSPIRHENAVQGSAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Sheep_(Ovis aries)_(XP_004017877.2)
LGTMGDHSKKKPGTAMCVGCGSQIHDQFILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAKCQVGFSSSDLVMRARDSVYHIECFRCSVCSRQLLPGDEF
SLREHELLCRADHGLLLERAAAGSPRSPLPARGLHLPDPGSGRQPSLRPHVHKQTEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHNDKTSLQGLTGTPLVAGSPIRHESAVQGSAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Killer_Whale_(Orcinus orca)_(XP_004276396.1)
LGTMGDHSKKKPGTAMCVGCGSQIHDQFILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAKCQVGFSSSDLVMRARDSVYHIECFRCSVCSRQLLPGDEF
SLREHELLCRADHGLLLERAAAGSPRSPLPARGLHLPDPGSGRQPSLRPHVHKQTEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHNDKTSLQGLTGTPLVAGSPIRHESAVQGSAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Panda_(Ailuropoda melanoleuca)_(XP_034524749.1)
LGAMGDHSKKKPGTAMCVGCGSQIHDQFILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAKCQVGFSSSDLVMRARDSVYHIECFRCSVCSRQLLPGDEF
SLREHELLCRADHGLLLERAAAGSPRSPLPARGLHLPDPGSGRQPSLRPHVHKQTEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHNDKTSLQGLTGTPLVAGSPIRHESAVQGSAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Mouse_(Mus musculus)_(NP_081673.2)
LGAMGDHSKKKPGTAMCVGCGSQIHDQFILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAQCQVGFSSSDLVMRARDSVYHIECFRCSVCSRQLLPGDEF
SLREHELLCRADHGLLLERAAAGSPRSPLPARGLHLPDAGSGRQPSLRTHVHKQAEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHSDKASLQGLTGTPLVAGSPIRHENAVQGSAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Platypus_(Ornithorhynchus anatinus)_(XP_028922094.1)
LGAMGDHSKKKPGIAMCVGCGSQIHDQYILKVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYIRLFGIKCAKCKVGFSSSDLVMRARENVYHIECFRCSVCSRQLLPGDEF
SLRDHDLLCRADHSLLLDRGSADSPRSPLPSRGLHLTDPGSGRQPSLRPHVHKQAEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHNDKTSLQGLTGTPLVAGSPIRHDSAVQGSAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Tasmanian_Devil_(Sarcophilus harrisii)_(XP_031812658.1)
LGAMGDHSKKKPGIAMCVGCGTQIHDQYILKVSPDLEWHAACLKCAECSQYLDETCTCFV
RNGKTYCKRDYIRLFGIKCAKCKVGFSSSDLVMRARENVYHIECFRCSVCSRQLLPGDEF
SLREHELLCRADHSLLLDRASAESPRSPLASRSLHLSEPGSGRQPSLRPHVHKQTEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHSDKTSLQGLTGTPLVAGSPIRHESAVQGSAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Alligator_(Alligator mississippiensis)_(XP_006261413.1)
LGAMGDHSKKKPGIAMCVGCGSQIHDQYILKVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYIRLFGIKCAKCKVGFSSSDLVMRARDNVYHIECFRCSVCSRQLLPGDEF
SLRDHELLCRADHSLLLDRSAADSPRSPLQTRALHLAEPGPGRQPALRPHVHKASEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSLLM
KQLQQQQHSDKASLQGLTGTPLVAGSPIRHESAVQGSAVEVQTYQPPWKALSDFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Green_Sea_Turtle_(Chelonia mydas)_(XP_007068425.1)
LGAMGDHSKKKPGIAMCVGCGSQIHDQYILKVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYIRLFGIKCAKCKVGFSSSDLVMRARENVYHIECFRCSVCSRQLLPGDEF
SLRDHELLCRADHSLLLDRSSADSPRSPLQTRGLHLSDSVPGRQPSLRPHVHKQTEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHGDKTSLQGLTGTPLVAGSPIRHESAVQGNAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Painted_Turtle_(Chrysemys picta bellii)_(XP_005311676.1)
LGAMGDHSKKKPGIAMCVGCGSQIHDQYILKVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYIRLFGIKCAKCKVGFSSSDLVMRARENVYHIECFRCSVCSRQLLPGDEF

SLRDHELLCRADHSLLLDRSSADSPRSPLQTRGLHLSDSVPGRQPSLRPHVHKQTEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHSDKTSLQGLTGTPLVAGSPIRHESAVQGNAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Coelacanth_(Latimeria chalumnae)_(XP_005995696.1)
VGAMGDHSKKKPGIAMCVGCGSQIHDQYILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYIRLFGIKCAKCNSGFSSSDLVMRARDNVYHIECFRCSVCSRQLLPGDEF
SLREHQLLCRADHSLLLERSSTESPLSPLQSRSLHLVDPVSARQPSLRAHVHKQTEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHSDKTSLQGLTGTPLVAGSPIRHENAVQGNAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Spotted_Gar_(Lepisosteus oculatus)_(XP_006628961.1)
LGDMGDHSKKKSGIAMCVGCGSQIHDQYILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAKCNVGFSSSDLVMRARDTVYHIECFRCSVCSRQLLPGDEF
SLRDEELLCRADHSLLVERSSAGSPVSPIHTRSLHMADPVSVRQPQLRTHVHKQSEKTTR
VRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSILM
KQLQQQQHNDKTNLQGLTGTPLVAGSPIRHDNAVQGNPVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2B_Medaka_(Oryzias latipes)_(XP_023809342.1)
LDDMGDHSKKKPGFAMCVGCGSQIHDQYILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAKCNLGFSSSDLVMRARDNVYHIECFRCSMCSRQLLPGDEF
SLQEGDLLCRADHSMLLERTSAGSPISPIHNRPLHMADPVTVRQAPHRNHVHKQSEKTTR
VRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHSDKTNLQGLTGTPLVAGSPIRHESTVQGNPVEVQTYQPPWKALSDFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2B_Tilapia_(Oreochromis niloticus)_(XP_013124225.1)
LGDMGDHSKKKPGFAMCVGCGSQIHDQYILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAKCNLGFSSSDLVMRARDNVYHIECFRCSVCSRQLLPGDEF
SLREEELLCRADHSLLLERSSAGSPVSPIHNRPLHLADPVTVRQAPHRNHVHKQSEKTTR
VRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQQSDKTSLQGLTGTPLVAGSPIRHEGNVQGNPVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Common_Lizard_(Zootoca vivipara)_(XP_034969087.1)
LGAMGDHSKKKPGIAMCVGCGSEIHDQYILKVSPDLEWHAACLKCADCSQYLDETCTCFV
RDGKTYCKRDYIRLFGIKCAKCAAGFSSSDLVMRARDNVYHLECFRCSVCSRQLLPGDEF
SLRDHELLCRADHSLLIDRAGSSPPRSPLQAR---LADSLPGRQPSLRPHVHKQAEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHSDKASLQGLTGTPLVAGSPIRHDSTVQGNAVEVQTYQPPWKALSDFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Wall_Lizard_(Podarcis muralis)_(XP_028600481.1)
LGAMGDHSKKKPGIAMCVGCGSEIHDQYILKVSPDLEWHAACLKCADCSQYLDETCTCFV
RDGKTYCKRDYIRLFGIKCAKCASGFSSSDLVMRARDNVYHLECFRCSVCSRQLLPGDEF
SLRDHELLCRADHSLLVDRAGSSPPRSPLQARGLQLADSVPGRQPSLRPHVHKQAEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHSDKASLQGLTGTPLVAGSPIRHDSAVQGNAVEVQTYQPPWKALSDFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Gekko_(Gekko japonicus)_(XP_015279053.1)
LGAMGDHSKKKPGIAMCVGCGSEIHDQYILKVSPDLEWHAACLKCADCSQYLDENCTCFV
RDGKTYCKRDYIRLFGVKCAKCAAGFSSSDLVMRARDHVYHLECFRCSVCSRQLLPGDEF
SLRDHELLCRADHSLLLDRAAASPPRSPLPARGLPLTDSVPGRQSSLRPHVHKTAEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHNDKASLQGLTGTPLVAGSPIRHDSAVQGNAVEVQTYQPPWKALSDFALQSEL

DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2A_Zebrafish_(Danio_rerio)_(NP_571045.1)
LGDMGDHSKKKSGIAMCVGCGSQIHDQYILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAKCNIGFCSSDLVMRARDNVYHMECFRCSVCSRHLLPGDEF
SLRDEELLCRADHGLLMERASAGSPISPIHSRPLHIPEPVPVRQPPHRNHVHKQSEKTTR
VRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHNDKTNLQGLTGTPLVAGSPIRHDTTVQGNPVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2A_Medaka_(Oryzias latipes)_(XP_004069695.1)
LDDMGDHSKKKSGIAMCVGCGSQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYARLFGIKCAKCNTGFCSSDLVMRAREKVYHMECFRCSVCSRHLLPGDEF
SLREDELLCRANHDL-LERASAGSPLSPLHKRTLHISDPISVRHPSHRNHVHKQSEKTTR
VRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQIQQQQHNDKTNLQGLTGTPLVAGSPIRHDTTVQGNPVEVQTYQTPWKTLSDLALETDL
DQPAFQHLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVDT

>ISL2A_Tilapia_(Oreochromis niloticus)_(XP_003440468.1)
LDDMGDHSKKKSGIAMCVGCGSQIHDQYILRVSPDLEWHAACLKCAECNQHLDETCTCFV
RDGKTYCKRDYARLFGIKCAKCNMGFCSSDLVMRARDNVYHMECFRCSVCSRHLLPGDEF
SLRDDELLCRADHGLMMERASAGSPLSPIHNRPLHISDPVSVRHPPHRNHVHKPSEKTTR
VRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHNDKTNLQGLTGTPLVAGSPIRHDNTVQGNPVEVQSYQPPWKTLSDFALQTDL
DQPAFQQLVSFSEAGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVDT

>ISL2B_Zebrafish_(Danio_rerio)_(NP_571039.1)
LDDMGDHSKKKSGLAMCVGCGSQIHDQYILRVSPDLEWHAACLKCVECNQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAKCTLGFSSSDLVMRARDSVYHIECFRCSVCSRQLLPGDEF
SVRDEELLCRADHGLALERGPGGSPLSPIHTRGLHMADPVSVRQTPHRNHVHKQSEKTTR
VRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSILM
KQL-QQQHGDKTNLQGMTGTALVAGSPIRHNPSVPGHPVDVQAYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLGNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Thorny_Skate_(Amblyraja radiata)_(XP_032870769.1)
LGSMGDHCKRKHGVALCVGCGSQIHDQYILRVSPDLEWHAACLKCAECSQYLDESCTCFV
KDGKTYCKRDYIRLFGTKCAKCNLTFSKNDLVMRARNRVYHIECFRCVACSRQLIPGDEF
ALRDDELFCRADHDVVERGSVGGDSLSPEPSRSLQMAEPIAVRQPPLRSHVHKQQEKTTR
VRTVLNEKQLHTLRTCYGANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRTIFM
KQMQQQQNSDKTSLQGLTGTPMVAASPIRHDGSMQGNAVEVQSYQPPWKALSEFALQSDL
DQPAFQQLVSFSESGSLPNSSGSDVTSLSSQLPDTPNSMVPSPVDT

>ISL2A_Ghostshark_(Callorhinchus milii)_(XP_007906404.1)
LSAMGDHSKKKHGIAMCVGCGSQIHDQYILRVSPDLEWHAACLKCAECNQYLDETCTCFV
RDGKTYCKRDYIRLFGTKCAKCNLSFGKSDFVMRARSSVYHIECFRCVACSRQLIPGDEF
ALRDNELFCRADHDVVERASAGGESLSPVPGRPLQMAEPISARQPPLRP--HKQQEKTTR
VRTVLNEKQLHTLRTCYGANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKTILM
KQIQQQQHSDKTSLQGLTGTPMVAGSPIRHDSSVQGNAVEVQTYQPPWKALSEFALQSDL
DQPAFQQLVTFSESSSLPNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2-like_Whale_Shark_(Rhincodon typus)_(XP_020392977.1)
LGAMGDHCKKKHGVAMCVGCGSQIHDQFILRVSPDLEWHAACLKCAECSQYLDETCTCFV
RDGKTYCKRDYIRLFGTKCAKCNLTFSKSDFVMKARSRVYHIDCFRCVACSRQLIPGDEF
ALRDDELFCRADHDVVERASAGGDPLSPVPSRSLQMAEPISARQPSLRPHVHKQQEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKITLM
KQLQQQQHSDKTSLQGLTGTPMVAGSPIRHD--VQGNAVEVQSYQAPWKALSEFALQSEL
DQPAFQQLVTFSESGSLPNSSGSDVTSLSSQLPDTPNSMVPSPVET

>ISL2_Toad_(Bufo bufo)_(XP_040271691.1)

LAAMGDQPKKKPGLAVCVGCGSHILDQYILRVSPDLEWHAACLKCAECSQYLDENCTCFV
RDGKTYCKRDYIRLFSTRCPRCQGTLPRSELVMRVGERVYHTDCFRCSICSRRLLPGEEI
SLRDQELLCAADHNI----------PDTGRSSSLRSHIHKQTEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQQNDKTSLQGLTGTPLIAGSPIRHDSAVQGAAVEVQTYQPPWKALSDFALQSDL
EQPAFQQLVSFSESGSLGNSSGSDITSLSSQLPDTPNSMVPSPVET

>ISL2_Starling_(Sturnus vulgaris)_(XP_014734647.1)
PGAMGEPSKRRPGLALCAGCGGRIQDPFLLRVSPDLEWHVACLKCAECGQPLDETCTCFL
RDGKAYCKRDYSRLFGIKCAQCRAAFSSSDLVMRARDHVYHLECFRCAACGRQLLPGDQF
CLRERDLLCRADHGPVPDGAAARGARSPLPAAAAHLAEPVPGRPPAPRPPAHKAAEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHSDKTSLQGLTGTPLVAGSPIRHESAVQGSAVEVQTYQPPWKALSEFALQSDL
EQPAFQQLVSFSESGSLGTSSGSDVTSLSSQLPDTPNSMVPSPAET

>ISL2_Great_Tit_(Parus major)_(XP_015494171.1)
QGAMGEPSKRRPGLALCAGCGGRIQDPFLLRVSPDLEWHVACLKCAECGQPLDETCTCFL
RDGKAYCKRDYSRLFGIKCAQCRAAFSSSDLVMRARDHVYHLECFRCAACGRQLLPGDQF
CLRERDLLCRADHGPPPDGTAARGPRSPLTAAAAHLAEPVPGRPPAPRPPAHKAAEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQLQQQQHSDKTSLQGLTGTPLVAGSPIRHESAVQGSAVEVQTYQPPWKALSEFALQSDL
EQPAFQQLVSFSESGSLGTSSGSDVTSLSSQLPDTPNSMVPSPAET

>ISLA_Sea_Lamprey_(Petromyzon marinus)_ (XP_032829736.1)
LASIWETTPKKRLASVCVGCGAQIRDPFILRVSPDLEWHASCLKCAECAQYLDETCTCFV
RDGKTYCKRDYVRLFGIKCAKCGVGFSKTDFVMRVRTQVFHLECFRCVACSRQLIPGDEF
ALREDGLFCRADHELLERAGTADGMGSPQGGRPLQLAEPMSARHHPLRPHVHKQAEKTTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSILM
KQIQQQG-GDKTNLQGLTGTPLVAGSPVRQDGGLPVGAVDVQTFQPPWKALSDFALSSEL
DQGPFQQLVNFSEGRPGSNSTGSEITSL-SQLPDTPTSIGSSPIEA

>ISLB_Sea_Lamprey_(Petromyzon marinus)_(XP_032824418.1)
MGDMGDSAKKRRRAALCVGCGTHIQDPFILRVSPDLEWHAACLKCAECGQSLDETCTCFV
RDGKTFCKRDYSRLFGIKCAKCNASFSKTDFVMRARSHIYHMECFRCETCSRQLLPGDEF
ALRDGGLLCRADQHHGGGGGGATRPGSPVHSRSLHLAGGGGGGGGGVPGGIPSGPEKATR
VRTVLNEKQLQTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSILM
KQMQQQHHADKTNLQGLTGTPLVAGSPVPHEAAIPGNPVEVQSFQPPWKTLSEFALSSDL
DQGPFQHLMGFSDGGPGSNSTGSDLGSIPSQLPDTPSSMVASPIEA

>ISL_Amphioxus_(Branchiostoma floridae)_(XP_035666393.1)
IESPGDPPKKNRRNAMCVGCGSHIHDQYILRVAPDLEWHAACLKCSDCNQYLDETCTCFV
REGKTYCKRCYVRLFGTKCAKCSLGFTKNDFVMRARNKIYHIDCFRCVACSRQLIPGDEF
ALREDGLFCKADHEVLERASNNVDSNGRSLSTDLEMTRPESHRSDQRRPQVHKQDHKPTR
VRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILM
KQMQEQASKQDLGIGRLNGVPMVAQEPVRHESQMQANPVEVQSYQPPWKALSDFALQSDI
EQPAFQQLKSGDVHAQLPCGGGMMVLTLH------PHSHTPGPGEA

>ISL1B_Zebrafish_(Danio rerio)_(NP_001002043.1)
LSDMGEQQNRTKLSSFCVGCGLEILDRFIVRVSPDLEWHARCLKCAECHQFLDESCTCFI
RDGKTFCREHYSRLSTSKCAKCDKAFISKEFVMRSQVNIYHVQCFRCEGCNRPLLSGDEY
VLQDGQLLCTDHHNKLMSASINQQKEA---------GDPSEEKSTLSWSSMQRRSERATR
VRTVLSETQLCMLQTCYTANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSLTM
RHTQKQ-------LEGVSEEPLLLVSTESQDSDVMIS--------PPWKLLTDFILQNEA
EHRSFPQMLSLPTEGPC--SAGSEVASVS----DTANSLTASPTDI

>ISL1-Like_Thorny_Skate_(Amblyraja radiata)_(XP_032902406.1)
MAI-----SKEVPLSTCRGCSKHITDPYILRVYPDLEWHAACLKCVECNQNLDESCTCFI
KEGKIYCKTDYFSKFSVRCAQCQAGLSSLDLVLRAGDLTYHQRCFCCVACSRQLMPGDEF
TLRFDGPYCIQDGGLPDSSLNHKGVFSPLATSNLYLEKLKAHR---SAAHAAKHSDKITR

VRTVLSEQQLLTLRTCYAANPRPDALMKQQLIEMTGLNSRVIRVWFQNKRCKDKKKNVLQ
KHIDQRDR-DKADIRGLIGTLMVAMSPLTQKADLYCSPIEVHKCYTPWEDLKYFTQQ---
--------VAFSGGRCHSNCSSSEVSAPSSQLPDTLSSEGSQAADL

>ISL1-like_Whale_Shark_(Rhincodon typus)_(XP_020375499.1)
MGKKEERSMKEDSTSVCFGCSKHITDPYILRVYPNLEWHAACLKCVVCNQYLNETCTCFL
REGKTYCKTDYFKKFSVRCAQCQAGLLSSDLVFRARGLIYHQQCFRCVACNRRLLPGEEC
RLRFDGPYCIEDAWLPDPSFIQQDLLLRGQDRSLYLGQLMFARQPALRA--PNHSDKITR
MRTVLNEQQLLTLRTCYAANPRPDALMKQQLMEMTGLNSRVIRVWFQNKRCKDKKKSIMP
KHMDPCDR-NKADIQGLVGTLMIATSPLPQKVDSQCNPVEVQRYWPPWEDLSDLPLQVFL
EEDHTSLVVVLKCHLLVLNLLQQTVMNLNHQTSDYDNSLLKDQLDG

>ISLC_Sea_Lamprey_(Petromyzon marinus)_(XP_032806737.1)
MSDGGGADGAEGVSAGCAGCGRPIRDAFLLRVWPNLRFHAACLRCAECRAQLHEARSCFV
RAGRTFCQRDYNRLFGVKCSRCSLGLSRTELVMRVRGRIYHLACFRCWACARRLLPGDEV
SLRPGELLCRA-HSLPPPPHHHQQQLNDLENHPAGVRGPGEKRGPVPRPSCMSGSGKASR
VRTVLTERQLQTLRTCYAANPRPDALLKEQLVEMTGLSPRVIRVWFQNKRCKDKKRTLAG
RHDGEPPSDAKTCLQGLEGKALVARSPVTQDSEAEAEAVRASPQPPSWRALAAFALRAQL
QPLSTRQAASLSDSGSCSDSSASDVTLLPSHLPDTPTSLASS----

S2.3.3 NELF

>NSMF_Human_(Homo sapiens)_(NP_001124441.1)
MGAAASRRRALRSEAMSSVAAKVRAARAFGEYLSQSHPENRNGADHLLADAYSGHDGSPE
MQPAPQNKRRLSLVSNGCYEGSLSEEPSIKPAGEGPQPRVYTISGEPALLPSPEAEAIEL
AVVKGRRQ-PHHHSQPLRASGSREDVSRPCQSWAGSRQGSKECPGCAQLAPPTPRAFGLD
QPPLPETSGRRKKLERMYSVDRVSDDIPIRTWFPKENLFSFQTATTTMQAISVFRGYAER
KRRKRENDSASVIQRNFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPMKADTSHDS
RDSSDLQSSHCTLDEAFDLDWDTEKGLEAVACDTEGFVPPKVMLISSKVPKAEYIPTIIR
RDDPSIIPILYDHEHATFEDILEEIERKLNVYHKGAKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIHFWKRLSRLMSKVNPEPNVIHIMGCYILGNPNGEKLFQNLRTLMTPYR
VTFESPLELSAQGKQMIETYFDFRLYRLWKSRQHSKLLDFDDVL

>NSMF_Panda_(Ailuropoda melanoleuca)_(XP_034520793.1)
MGAAASRRRALRSEAMSSVAAKVRAARAFGEYLSQSHPENRNGADHLLADAYSGHDGSPE
MQPAPQNKRRLSLVSNGRYEGSLSEEAVSKPAGEGPQPRVYTISGEPALLPSPEAEAIEL
AVVKGRRQRPHHHSQPLRASGSREDVSRPCQSWAGSRQGSKECPGCAQLAPPSPRAFGLD
QPPLPEATSRRKKLERMYSVDRVSDDTPIRTWFPKENLFSFQTATTTMQAISVFRGYAER
KRRKRENDSASVIQRNFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPMKADTSHDS
RDSSDLQSSHCTLGEAFDLDWETEKGLEAVACDTVGFVPPKVMLISSKVPKAEYIPTIIR
RDDPSIIPILYDHEHATFEDILEEIEKKLNIYHKGAKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIHFWKRLSRLMSKVNPEPNVIHVMGCYILGNPNGEKLFQNLRTLMTPYR
VTFESPLELSAQGKQMIETYFDFRLYRLWKSRQHSKLLDFEDVL

>NSMF_Mouse_(Mus musculus)_(NP_001034475.1)
MGAAASRRRALRSEAMSSVAAKVRAARAFGEYLSQSHPENRNGADHLLADAYSGHDGSPE
MQPAPQNKRRLSLVSNGRYEGSISDEAVSKPAIEGPQPHVYTISREPALLPGSEAEAIEL
AVVKGRRQRPHHHSQPLRASSSREDISRPCQSWAGSRQGSKECPGCAQLVPPSSRAFGLE
QPPLPEAPGRHKKLERMYSVDGVSDDVPIRTWFPKENLFSFQTATTTMQAISVFRGYAER
KRRKRENDSASVIQRNFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPMKADTSHDS
RDSSDLQSSHCTLDEACDLDWDTEKGLEAMACNTEGFLPPKVMLISSKVPKAEYIPTIIR
RDDPSIIPILYDHEHATFEDILEEIEKKLNIYHKGAKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIHFWKRLSRLMSKVNPEPNVIHIMGCYILGNPNGEKLFQNLRTLMTPYK
VTFESPLELSAQGKQMIETYFDFRLYRLWKSRQHSKLLDFDDVL

>NSMF_Killer_Whale_(Orcinus orca)_(XP_033271119.1)
MGAVASRRRALRSEAMSSVAAKVRAARAFGEYLSQSHPENRNGADHLLADAYSGHDGSPE
MQPAPQNKRRLSLISNGRYEDSLPEEAVSKPAGEGPQPRVYTISGEPALLPGPEAEAIEL
AVVKGRQQRPRHHSQPLRTSGSREDVSRPCQSWAGSRQGSRECPGCAQLAPPSPQAFGLD

QPPLPEAASRRKKLERMYSVDRVSDDIPIRTWFPKENLFSFQTATTTMQA--VFRGYAER
KRRKRENDSASVIQRNFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPMKADTSHDS
RDSSDLQSSHCTLGEAFDLDWETERGLEAVACDTEGFVPPKVMLISSKVPKAEYIPTILR
RDDPSIIPILYDHEHATFEDILEEIEKKLNVYHKGAKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIHFWKRLSHLMSKVNPEPNVIHVMGCYVLGNPNGEKLFQNLRTLMTPYR
VTFESPLELSAQGKRFEEA--SRAAYRLWNCKLHSTV-----VL

>NSMF_Sheep_(Ovis aries)_(XP_027821839.1)
GGQSASSPSVWRVPVPES-PRKPEWRRAPGDGRQDQQAASSPGADHLLADAYSGHDGSPE
MQPAPQNKRRLSLISNGRYEGSLPEEAASKPAGEGPQPRVYTISGEPALLPGPEAEAIEL
AVVKGRQQRPHHHSQTLRASGSHEDVSRPCQSWTGSRHDSKECPGCAQLAPPAPQAFGVD
QPPLPEAPGRRKKLERMYSVDRVSDDVPIHTWFPKENLFSFQTATTTMQAISVFRGYAER
KRRKRENDSASVIQRNFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPMKADTSHDS
RDSSDLQSSHCTLGEAFDLDWETEKGLEAVACDTEGFVPPKVMLISSKVPKAEYIPTIIR
RDDPAIIPILYDHEHATFEDILEEIEKKLNIYHKGAKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIHFWKRLSRLMSKVNPEPNVIHVMGCYVLGNPNGEKLFHNLRTLMTPYR
VIFESPLELSAQGKQMIETYFDFRLYRLWKSRQHSKLLDFEDVL

>NSMF_Chicken_(Gallus gallus)_(XP_015135145.1)
MGTAVSKRKTLRNEAMSSVAAKVRAARAFGEYLSQNHPEGRNGSDHLLADSYIGQEDSPE
MQQAAQNKRRLSVISDGKFERSFSEEQTEKMPSEGPKPRVYTISGERPMLSDHENESMEL
VVMKGAAQEECHHGHPVHGAGGSHGVSRHCKGWPGSRQGSKECPNCTRLAAPSHHSFDLE
PHQSGETGWHRKRLERMYSVDRVSDDVPIRTWFPKENLFSFQTATTTMQAISAFRGYAER
KRRKRENDSAAVIQRNFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPIKADSFHDS
RESHDLQDSCGTLDGDFDLNWEAEKELEAMACDGEDFIPPKIMLISSKVPKAEYVPTIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMILFWKRLSRLMSKINPEPNIIHIMGCYVLGNPNGEKLFQNLKNLMNPYR
VAFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Turkey_(Meleagris gallopavo)_(XP_010719059.1)
----------RAARAFGEYLSQNHPEGRNGSDHLLADSYIGQEDSPE
MQQAAQNKRRLSVISDGKFERSFSEEQTEKMPSEGPKPRVYTISGERPMLSDHENESMEL
VVMKGAAQEECHHGHPVHGAGGSHGVSRHCKGWPGSRQGSKECPNCTRLAAPSHHSFDLE
PHQSGETGWHRKRLERMYSVDRVSDDVPIRTWFPKENLFSFQTATTTMQA--AFRGYAER
KRRKRENDSAAVIQRNFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPIKADSFHDS
RESHDLQDSCGTLDGDFDLNWEAEKELEAMACDGEDFIPPKIMLISSKVPKAEYVPTIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMILFWKRLSRLMSKINPEPNIIHIMGCYVLGNPNGEKLFQNLKNLMNPYR
VAFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Pelgrine_falcon_(Falco peregrinus)_ (XP_027634050.1)
----------LSFAAFCG-AWICRAARAFGEYLSQNHPEGRNGSDHLLADSYIGQEDSPE
MQQAAQNKRRLSVISDGKFERSFSEEQAEKMPSEGPKPRVYTISGERPMLSDHENESMEL
VVMKGAAQEECHHGHQVHSAGSSHGISRHCKGWPGSRQGSKECPNCTRLAAPSQHSFDLE
QQQSGETGWHRKRLERMYSVDRVSDDVPIRTWFPKENLFSFQTATTTMQAISAFRGYAER
KRRKRENDSAAVIQRNFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPIKADSFHDS
RESHDLQDSCGTLDGDFDLNWEAEKELEAMACDGEDFIPPKIMLISSKVPKAEYVPTIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMILFWKRLSRLMSKVNPEPNIIHIMGCYVLGNPNGEKLFQNLKNLMNPYR
VAFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Emperor_Penguin_(Aptenodytes forsteri)_(XP_019328757.1)

MQQAAQNKRRLSVISDGKFERSFSEEQAEKMPSEGPKPRVYTISGERPMLSDHENESMEL
VVMKGAAQEECHHGHQVHSAGGSHGISRHCKGWPGSRQGSKECPNCTRLAAPSQHSFDLE
QHQSSETGWHRKRLERMYSVDRVSDDVPIRTWFPKENLFSFQTATTTMQA----------
----------NFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPIKADSFHDS
RESHDLQDSCGTLDGDFDLNWEAEKELEAMACDGEDFIPPKIMLISSKVPKAEYVPTIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA

TFAKVEKEEDMILFWKRLSRLMSKVNPEPNIIHIMGCYVLGNPNGEKLFQNLKNLMNPYR
VAFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Great_tit_(Parus major)_(XP_015500231.1)
MGTAVSKRKTLRNEAMSSVAAKVRAARAFGEYLSQNHPEGRNGSDHLLADSYIGQEDSPE
MQQVAQNKRRLSVISDGKFERSFTEEQAEKMPSEGPKPRVYTISGERPMLSEHESDSMEL
VVMKGAAHEECHHGHHAHGAGGSHGV-RHCKAWPGGRQGSKECPNCTRLAAPSQQSIELE
QHPPGEAGWHRKRLERMYSVDRVSDDVPIRTWFPKENLFSFHTATTTMQAISAFRGYAER
KRRKRENDSAALIQRNFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPIKADSFHDS
RESHDLQDSCGTLDGDFDLNWEAEKELEAMACDGEDFIPPKIMLISSKVPKAEYVPTIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDLILFWKRLSRLMSKVNPEPNIIHIMGCYVLGNPNGEKLFQNLKNLMNPYR
VAFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Alligator_(Alligator mississippiensis)_(XP_006268359.2)
MGTAVSKRKSLRNDAMSSVAAKVRAARAFGEYLSQNHPESRNGSDHLLADSYIGQEDSPE
MQQAAQNKRRLSVISDGKFERSFSEERAEKIPSEGPKPRVYTISGERPMLSDHENDSMEL
VVMKGTGQEECHHGQPLQSACSSHNISRHCKGWAGSRQGSKECPSCARLAASSQHSFDLE
QHQSNETGWRRKKLERMYSVDRVSDDVPIRTWFPKENLFSFQTATTTMQAISAFRGYAER
KRRKRENDSAAVIQRNFRKHLRMVGSRRVKAQTFAERRERSFSRSWSDPTPVKADSFHDS
RESHDLQDSCCALEDEFDLNWEAEKELEAAACDGEDFVPPKIMLISSKVPKAEYIPTIIR
RDDPSIIPILYDHEHATFEDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMILFWKKLSRLMSKINPEPNLIHIMGCYVLGNPNGEKLFQNLKNLMNPYR
VAFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Painted_Turtle_(Chrysemys picta bellii)_(XP_005292712.1)
MGTAVSKRKTLRNDAMSSVAAKVRAARAFGEYLSQNHPEGRNGSDHLLADSYIGQEDSPE
MQQAAQNKRRLSVISDGKFERSFSEEQAEKIPSEGPNARVYTISGERPMLSDHENESMEL
VVMKGTGREDCHHGQPLQSAGSSHNISRHCKGWAGSRQGSKECPTCAQLAAHSQHSFDLE
QHQSSETGWRRKKLERMYSIDRVSDDVPIRTWFPKENLFSFQTATTTMQAISAFRGYAER
KRRKRENDSAAVIQRNFRKHLRMVGSRRVKAQPFGERRERSFSRSWSDPTPIKADSFHDS
RESHDLQDYCCALDDDFDLNWETEKELEAMACDGDDFIPPKIMLISSKVPKAEYIPTIIR
RDDPSIIPILYDHEHATFEDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMILFWKRLSRLMSKINPEPNVIHIMGCYVLGNPNGEKLFQNLKNLMNPYR
VVFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDVL

>NSMF_Green_Sea_Turtle_(Chelonia mydas)_(XP_007067723.2)
MGTAVSKRKTLRNDAMSSVAAKVRAARAFGEYLSQNHPEGRNGSDHLLADSY---EDSPE
MQQAAQNKRRLSVISDGKFERSFSEEQAEKIPSEGPKARVYTISGERPMLSDPENESMEL
MVMKGTGREDCHHGQPLQSAGSSHNISRHCKGWAGSRQGSKDCPACARLAAHSQHSFDLE
QHQPSETGWRRKKLERMYSIDRVSGDVPIRTWFPKENLFSFQTATTTVQAISAFRGYAER
KRRKRENDSAAVIQRNFRKHLRMVGSRRVKAQTFGERRERSFSRSWSDPTPIKADSFHDS
RESHDLQDYCCALDDDFDLNWETEKELEAMACDGDDFIPPKIMLISSKVPKAEYIPTIIR
RDDPSIIPILYDHEHATFEDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMILFWKRLSRLMSKINPEPNVIHIMGCYVLGNPNGEKLFQNLKNLMTPYR
VVFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Coelacanth_(Latimeria chalumnae)_(XP_005994509.1)
MGTAVSKRRNLRNDAISSVAAKVRAARAFGEYLSQNHPEGRNGSDHLLADTFIGQEDSPE
TQLASQNKRRLSLISNSKFDRSFSEERGENIAGEGPKSRVYTISGESQMFSDRENESVEL
VAMKNGQE-VRHHTHPLQNSGSSHNISRHCKSWGGSRQGSRECPSCARMAVSSQNSFDLE
HHHSSEAPWRRKKLERMYSVDRVSDDVPIRTWFPKENLFSFHTATTTMQAISAFRGYAER
KRRKRENESAAVIQRNFRKHLRMVGSRRMRAQTFAERREKSFSRSWSDPTPIKTDSLNDS
RESHDLQTSYSALDDEFDLNWEEEKELEAMSCDGDDFIPPKIMLISSKVPKAEYIPTIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDLILFWKRLSRLMSKINPEPNVIHIMGCYILGNPNGEKLFQNLKNLMQPHR
ISFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Gekko_(Gekko japonicus)_(XP_015278597.1)
MGTAVSKRRSLRNEAVASVAAKVRAARAFGDYLSQNHPEGRNGSDHLLAESYVGPEDSPE
MHQAAQNKRRLSLISDGKFERSFSEERGEKLTGEGPKPRVYTISGERPMLSQHQSESMEL
VVIKGPEAEECSQGQPLQNAGSSHNISRHCKGWTSSRQVSKEFPGCAQLAAPSQHTFDLE
HHQAGETGWRRKKLERMYSIDRVSDDVPIRTWFPKENLFTFQTATTTMQAISAFRGYAER
KRRKRENDSAALIQRNFRKHLRMVGSRRMKAQTFAERRERSFSRSWSDPTPIKPDNLNDS
RESHEFQDSCSALDKSVDLNWEAEKELEAAACSGEDFIPPKIMLISSKVPKAEYVPTIIR
RDDSSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIYFWKRLGHLMSKLNPEPNVIHVMGCYVLGNHNGEKLFQNLKNLMSPCR
VNFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Wall_Lizard_(Podarcis muralis)_(XP_028570590.1)
MGTAVSKRRSLRNEAVASVAAKVRAARAFGDYLSQNHPEGQNGSDHLLAESYVGQEDSPD
MHQGAQNKRRLSLISDGKFERSFSEERGEKLAGEGPKPRVYTISGERPMLSQHQSESMEL
VVIKGAEAEDCSQGQPLQNAVSSHNIGRHCKSWASSRQVSKECPGCTHLAAPSQHAFDIE
HHQAGETGWRRKKLERMYSIDRVSDDVPIRTWFPKENLFTFQTATTTMQAISAFRGYAER
KRRKRENDSAALIQRNFRKHLRMVGSRRMKAQTFAERRERSFSRSWSDPTPIKPENLNDS
RESHELQDSCCALDKGVDLNWEAEKELEAAACSGEDFIPPKIMLISSKVPKAEYVPTIIR
RDDQSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIYFWKRLGHLMSKLNPEPNVIHIMGCYVLGNHNGEKLFQNLKNLMNPCR
VTFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Common_Lizard_(Zootoca vivipara)_(XP_034969860.1)
MGTAVSKRRSLRNEAVASVAAKVRAARAFGDYLSQNHPEGQNGSDHLLAESYVGQEDSPD
MHQGAQNKRRLSLISDGKFERSFSEERGEKLAGEGPKPRVYTISGERPMLSQHQSESMEL
VVIKGAEAEDCSQGQPLQNAVSSHNIGRHCKSWASSRQVSKECPSCNHLAAPSQHAFDIE
HHQAGETGWRRKKLERMYSIDRVSDDVPIRTWFPKENLFTFQTATTTMQAISAFRGYAER
KRRKRENDSAALIQRNFRKHLRMVGSRRMKAQTFAERRERSFSRSWSDPTPIKPENLNDS
RESHELQDSCCALDKGVDLNWEAEKELEAAACSGEDFIPPKIMLISSKVPKAEYVPTIIR
RDDQSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIYFWKRLGHLMSKLNPEPNIIHIMGCYVLGNHNGEKLFQNLKNLMNPCR
VTFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Common_Frog_(Rana temporaria)_(XP_040180442.1)
MGTAVSKKKALRNDTISHVAAKVRAARAFGEYLSQTHPDSRNGSDHLLV-----QEDSPE
MLQGSQNKRRLSAISDSKLERSFSEEHTEKVPSEGPKPRVYTISGERPMLSDHESDSMEL
VVMKPDQDEYSHHHQPLQNSGSAHNITRPSKSWSGSRQNSKECPNCAKLTVPAQHSFDLE
QHQANESGWRRKKLERMYSIDRVSDDVPIRTWFPKENLFTFQTATTTMQAISAFRGYAER
KRRKRENDSAAVIQRNFRKHLRMVGSRRMKAQSFADRRERSFSRSWSDPTPVKADSIHDS
KESNDLQNSCTALNEESDLNWEAEREMEITSCEGEDFIPPKIMLISSKVPKAEYIPTIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLNVYRKGCKIRQMLIFCQGGPGYLYLLKNKVA
TFAKVEKEEDLIQFWRKLSRLMSKTNPEPNVIHIMGCYVLGNPNGEKLFQNLKNLMSPHR
IDFKSPLELSAQGKQLIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Xenopus_(Xenopus tropicalis)_(NP_001116909.1)
MGTAVSKRKALRNDPISNVAAKVRAARAFGEYLSQNHPDSRNGSDHLLV-----QEDSPE
MLPASQHKRRLSSLSDSKLDRSFSEERGEKLPGEGSKPLVYTISGERPMLSDHESDSMEL
VVMKDGREEHHHHHHPLQSSGSAHNITRPCKSWSGSRQNSKECPTCTRLTVPTQHSFDLE
QHQASESGWRRKKLERMYSIDRVSDDVPVKTWFPKENLFTFQTATTTMQAISAFRGYAER
KRRKRENDSAAVIQRHFRKHLRMVGSRRVKAQTFADRRERSFSRSWSDPTPVKIDSIHDS
RESHELQDSCTALNEECDFNWEAEKEMEIQSCEGDDFIPPKIMLISSKVPKAEYIPTIIR
RDDSSIIPILYDHEHATFEDILEEIEKKLNIYRKGCKIWEMLIFCQGGPGYLYLLKNKVA
TFAKVEKEEDLILFWKRLGRLMSKVNPEANVINIMGCYVLGNPNGEKLFQNLKNLMSPHR
MEFKSPLELSAQGKQLIERYFDFRMYRLWKTRQHSKLLDYDDIL

>NSMF_Toad_(Bufo bufo)_(XP_040261778.1)
MGTAVSKKKALRNDTISHVAAKVRAARAFGEYLSQTHPDSQNGSDHLLV-----QEDSPE
MLQASQNKRRLSAISDSKLERSFSEERGEKLTSDGPKPHVYTISGERPMLSDHESDSMEL
VVMKTEQGEHHHHHHPLQNSGSAHNITRPCKSWSGSRQNSKECPNCAKLTVPAQHSFDLE

QHQSNESGWRKKKLERMYSIDRVSDDVPARTWLPKENLFTFQTATTTMQA----------
----------NFRKHLRMVGSRRMKAQTFADRRERSFSRSWSDPTPMKADSIHDS
RESHDLQNSCTALDEECDLNWESEREMEVLSCEGEEFIPPKIMLISSKVPKAEYIPTIIR
RDDPSIIPILYDHEHATFDDILEEVEKKLNVYRKGCKIWEMLIFCQGGPGYLYLLKNKVA
TFAKVEKEEDLILFWKRLSKLMSKINPEPNVIHIMGCYVLGNPNGEKLFRNLKNLMSPHQ
IEFKSPLELSAQGKQLIERYFDFRLYRLWKTRQHSSCWITRHFM

>NSMF_Pyhton_(Python bivittatus)_(XP_025022054.1)
QTLAISFLGSVYSFPSSPAANYPGAARAFGDYLSQNHPEGKNGSDHLLAESYVGQEDSPE
IHVGVQNKRRLSLISDGKFQPSFPEEAGEKLAGEGPKPRVYTISSEGPMLSQHQSESMEL
VVIKGTEEEDRSLGQPLQNAGSSHNISRQCKGWPSHQQGSKEYPSCAQLTMSSQHPFDLE
QNQAGETGWRRKKLERMYSIDRVSDEVPIRTWFPKENLFTFQTATTTMQA----------
----------NFRKHLRMVGSRRMKAQRFSER--RSFSRSWSDPTPIKPENLNDS
RESHEFQDSFCALDKGVDLNWEAEKELEAAACSGEDFLPPKIMLISSKVPKAEYVPTIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIYFWKRLGHLMSKLNPGPNVIHIMGCYVLGNHNGEKLFQNLKNLMSPCR
VTFESPLELSAQGKEMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Anolis_(Anolis carolinensis)_(XP_008123249.2)
MASGLESNCVLPPTLVESAFALLRAARAFGDYLSQNHPEGKNGSDHLLAESYLGPEDSPE
MLPAPQNKRRLSLISDGKFERSFSEEGGEKLPGEGPKPRVYTISGERPMLSQHQSESMEL
VVLKGAAEARGERGSQSQNGGSSHDVGRSCKDWAG----PKECPSCVPLAGTSHHHHHHH
HHQAGETSWRRKKLERMYSIDRGPDEVPIRTWFPKENLFTFQTATTTMQAIS--------
----------NFRKHLRMVGSRRMKAQTFAERRERSFSRSWSDPTPTKPENLNDS
RESHEPPDPCCVPDKGVDLNWEAEKEAEATACNGEDFVPPKIMLISSKVPKAEYVPTIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLNIYRKGCKIWKLLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIYFWKRLGHLMSKLNPDPNVVHIMGCYVLGNHNGEKLFQNLKNLMSPCR
VTFESPLELSAQGKQMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Spotted_Gar_(Lepisosteus oculatus)_(XP_015222022.1)
MGTAVSKRKHLRNDAISSVAAKVRAARAFGEYLSHAHPENRNGS----------
--------------------------GEPVSGAGSRP----------
----------GSRDCPQDCAGSHPAPVGAQHSLDL-
-----ESARQRKKLERMYSVDRVSEDVPIHSWFPKENMFSFQTATTTMQAISAFRGFAER
KRRKREHESAAVIQRNFRKHLRMVGSRRMRAQTFADRREKSFSRSWSDPTPIKTDPLHDS
RDSGDLQESCGTLDEGSDQIWEEEREEERLACQGEDFIPPKIMLISSKVPKAEYVPNIIR
RDDHSIIPILYDHEHATFDDILEEIEKKLNAYRRGCKIWRMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMILFWKRLSRLMTKLNPEPNVIHVMGCYVLGNPNGEKLFQNLKNLMKPQS
ITFESPLELSAQGKEMIETYFDFRLYRLWKTRQHSKLLDYDDIL

>NSMF_Ghostshark_(Callorhinchus milii)_(XP_007898540.1)
---------MKLHEVVATVIFLVGAARAFGEYLSQSQADGDPGSKHLLAEAFLEPDDTSQ
LQPVSQNKRRLSVTSHGKFERSFSDERSER-SSEGQKPRVYTISGESQLLSESENEATGM
---------SGKSGHPLHSSGSSHSVNKHCKSWAGSRQSSKDCSSCRKAAVSSNHSFDLD
HLQSREAARHRKKLERMYSVDKVSDDVPIRTWFPKENLFTFPSPTTTMQAISAFRGYAER
KRRKRENESAAVIQRNFRKHLRMVGSRRIRLQGFAERREKSFSRSWSDPTPIKSDLLCGS
RESQELQSSCSALDDDFDLTWEAEKELESLSCDGEDFVSPKIMLISSKVPKAEYIPTIIR
RDNPSIIPILYDHEHATFDDILEEIEKKLNAYRRGCKIWKMLFFCQGGPGHLYLLKNKVA
TFAKVGKEEDLILFWKRLSKLMSKINPEPNVIHIMGCYILGNPNGEKLFRSLKTLMAPHR
ISFESPLELSAQGKRMIETYFDFRLYRLWKTKQHSKLLGYDDVL

>NSMF_Whale_Shark_(Rhincodon typus)_(XP_020386609.1)
MGSAVSKRKNPRSNAICSAESKVRAAQAFREYLSQTEPSSTKASDHLVAETLMDPEHSSE
IQQPAQNKRRLSMISQGRFDRSFSEEGAEK-PADGQKPRVYTISGESQLLSERETESEGT
GTVTAATSTPGKEARPSQAGGGGGG---RCRNWAGSRQSSRECPGCAKAAGSSQLTVDGE
HVQAREAARHRKKLERMYSMDRVSDDVPIRTWFPKENLFNFQTATTTMQAXVXYRGYIDR
KRRKRENECAAVIQRNFRKHLRMVGSRRIRVQSLSERQEKSFGRSWSDPTPVKSDVIGEP
RESQELQNSCCALDDDFDLNWEAEKELELLSCDGEDFVPPKIMLISSKVPKAEYIPTIVR
RNDPSVIPILYDHEHAAFDDILEEIEKKLNAYRKGCKIWKLLIFCQGGPGHLYLLKNKVA

TFAKVEKEEDMILFWRRLSRLMSKINPDPNVIHIMGCYILGNPNGEKLFRSLSNLMTPHR
ISFESPLELSAQGKRMIETYFDFRLYRLWKSRQHSKLLDYDDIL

>NSMF_Thorny_Skate_(Amblyraja radiata)_(XP_032904506.1)
MGSAVSKRRPPRSGAVSPPEPSLRAAQVFGEYLSQSEAASSSAKEPLVVESPMGLEDV--
LQVPVQNKRRLSTMSQSRLERSFSEEGADK-PAEAQKARVYTISGESQLVSERESEGAR-
---------QAKAGS--QGPCGTHTIA--ARGWAAARAGSKDCPNCARAPIVSQQNFDLE
QAPSREPARHRKKLERMYSVDRVSDDVPIRTWFPKENLFNFQTATTAMQAISAFRGYTER
KRRKRENECAAVIQRNFRKHLRMVGSRRIRVQGFCERQEKNLGRSWSDPTPVKPDVMCET
RESQELQSVCAASDDDFDLSWEAEKELELLSCDGEDFVPPKIMLISSKVPKAEYIPTIIR
RDDSSIIPILYDHEHATFDDILEEIEKKLTTYRRGCKISKLLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDLILFWKRLSRMMSKINTEANVIHVMGCYILGNPNGEKLFQSLKDLMRPHR
ISFESPLELSAEGKRMIETYFDFRLYRLWKSRQRSKLLDYDDIL

>NSMF_Zebrafish_(Danio rerio)_(XP_021331973.1)
MGTAVSKKKNLRNDAISSVAAKVRAARAFGEYLSHTHPENRNRSDHLLSDTFIGQEDSPD
IQQGSQSKRRLSV------ERSLSSEDQQQRRTESSKPRVYTITRERDMLGQGSEESLEL
EVLKRTSEPHAPMAQPLQSSGSTHNI----RDWGSRRSRSRECVACIRPHCQSQRSLDLD
TSP-HGGGKQHKKLERMYSEDRVSSEDRTNSWFPKENMFSFQTATTTMQAISAFRGIAER
KRRKREQEATALMERNFRKHLRMVGSRRVKAQTFVDRKAKSFSRSWSDPTPVKPDSLHDS
RDSGDLQASSGNLDEEDDVDWEEERELERAACEGDDFIPPKLMLISSKVPKAEYVPNIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLTAYRKGCKIWNMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEGMMQFWKKLGRFMSLLNPEPNLIHIMGCYVLGNANGEKLFQNLKRLMKPHG
IEFKSPLELSAQGKEMIEMYFDFRLYRLWKTRQHSKLHDYDDLL

>NSMF_Carp_(Cyprinus carpio)_(XP_018953227.1)
MGTAVSKKKNLRNDSISSVAAKVRAARAFGEYLSHTHPENRNGSDHLLSDTFIGQDDSPD
IQPGSQSKRRLSV------ERSLSSEDQHQRCAESSKPRVYTITRERNMLSQGSKESLEL
EVLKRTTDP------PLQSSGSAHNI----RDWGSRRSRSRECVDCIRPHCQSQRSLDLE
TSP------------RTFLV--------TLDFFLLHCSLCTHTPTPTACNITAFRGIAER
KRRKREQEATALMERNFRKHLRMVGSRRVKVQTFADRKAKSFSRSWSDPTPIKPDSLHDS
RDSADLQASSGNLEEEDDVDWEEERELERAACEGDDFIPPKIMLISSKVPKAEYVPNIII
RDDPSIIPILYDHEHATFDDILEEIEKKLTAYRKGCKIWNMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIQFWKRLGRFMSLLNPEPNLIHIMGCYVLGNANGEKLFQNLKRLMKPHV
IEFKSPLELSAQGKEMIEMYFDFRLYRLWKSRQHSKLLDYDDLL

>NSMF_Tilapia_(Oreochromis niloticus)_(XP_019221155.1)
MGTAVSKRKNLRSDAISSVAAKVRAARAFGEYLSNTNPENRNGADHLLSDTFSGQ-DSPD
VNPQSQSKRRLSA------ERSLSIEEPPSRGSQGPKPRVYTITREGGMTGRGSEESLEL
EVLKGSREQPMSSSQPLQSSGSAGNI----RDWGMRRGGSRDCVACIRAPCQSQRSLDLD
TSP-RDGGKQRKKLERMYSEDRASTDDRPNSWFPKENMFSFQTATTTMQAISAFRGIAER
KRRKREQEAATMMERNFRKHLRMVGSRRMKAQTFVDRRAKSFSRSWSDPTPVKPESLHES
RDSGELQTSSGTLDEGLDADWEEEKDMERVACDGEDFIPPKIMLISSKVPKAEYVPNIIR
RDDPSIIPILYDHEHATFDDILEEIEKKLTAYRKGCKIWNMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMIQFWRRLSRLMSKLNPEPNLIHIMGCYVLGSANGEKLIQTLKRLMRPTS
VEFKSPLELSAQGKEMIEMYFDFRLFRLWKSRQHSKLLDYDDLL

>NSMF_Medaka_(Oryzias latipes)_(XP_023816497.1)
MGTAASRRRNLRNDAISSVAAKVRAARAFGEYLSHTHPENRNGSAHLLCDTLVGSDEAPS
DQPPPPKPERLSL------ERSFSTEEDQQKSVECTQPRVYTITTQHGMLGRGSKESLEL
DVLKDKSASHTHHHHPLQTSISAHNI----RGWGESGKGEAECEACSSAPSRSQGSLDLE
-NTSREAGKQHRRLERMWSVDRVTGEREDSNWFPKENMFTFQTATTTMQAIS--------
----------NFRKHLRMVGSRRIKAQTFAERRSKSFNRSWSDPTPIKTESPHEP
RDSGDLQTTSGTAEEGGNRGWEEEQELERLACEGDDFVPPKIMLISSKVPKAEYIPTIIR
RDDPSIIPILYDHEHATFDDILEEIDKKLTAYRKGSKFWRMLIFCQGGPGHLYLLKNKVA
TFAKVEKEEDMSQFWRRLSRFMSKLNPEPNKIHIMGCYVLGNPNGEKLFQKLKNLMRPYS
VEFESPLELSAQGKEMIEMYFDFRLYRLWKTRQHSKLLDYEDFL

>NSMF_Sea_Lamprey_(Petromyzon marinus)_(XP_032814793.1)

MGTSVSKRRQMRAEAISSAAIKFRAVKSFGDYVQLGHMQGRGGTERCLSEPGVEASPSPC
LQPPSQQKRRLSL------SRSFSDEAAETPATPAPSPSVPSIPSSSPPPSRSASARTSP
AA----------HFSGSSSSAGSSRD--GGGGSERDSLECPACESLGV--KHSFDLE
QEKAQEALNSKKKLERMYSMDRISDDIECRSWFPSESMHMLHSTASTLQAIAAFRGYAER
KRRKRELDPSAMVQRNFHKHLRMVGRHRGRAAVSLDRQERECGRALSDVSQSKSDPVLDA
RFSHDMQSTYERLHREPSARWEEEREIELESCSRTDFIPPKVMLISSKVPKAEYIPNIIR
KDDPSIIAILYDHEHATFSDILDEIEKKLNEYRRGCKIWKLLIYCQGGPGYLYLLKNKVA
TFAKLEKEDDLIMFWKCLGSLMSKINTELNVVHIMGCYVLGNHNGEKLLDSLKSIMGPYR
VSFESPLELSAQGKKMIETYFSFRLYKLWKARQHSKLVDFDEIL

>NSMF_Brook_Lamprey_(Lampetra planeri)

----------HFSGSGSSAGSSRD-GGGGGSERDSLECPACENLGV--KHSFDLE
QEKAQEALSSKKKLERMYSMDRISDDIECRSWFPSESMHMLHSTASTLQAIAAFRGYAER
KRRKRELDPSAMVQRNFHKHLRMVGRHRGRAAVSLDRQERECGRALSDVSQSKSDPVLDA
RFSHDMQSAYERLHREPSARWEEEREIELESCSRTDFIPPKVMLISSKVPKAEYIPNIIR
KDDPSIIAILYDHEHATFSDILDEIEKKLNEYRRGCKIWKLLIYCQGGPGYLYLLKNKVA
TFAKLEKEDDLIMFWKCLGSLMSKINTELNVVHIMGCYVLGNHNGEKLLDSLKSIMGPYR
VSFESPLELSAQGKKMIETYFSFRLYKLWKARQHSKLVDFDEIL

>NSMF_Japanese_Lamprey_(Lethenteron camtschaticum)
MGTSVSKRRQMRAEAISSAAIKFRAVKSFGDYVQLGHMQGRGGTERCLSEPGVEASPSPC
LQPPSQQKRRLSL------SRSFSDEAAETPATPAPSPSVPSIPSSSPPPSRSASARTSP
AA----------HFSGSSSSAGSSRDGGGGGGSERDSLECPACENLGV--KHSFDLE
QEKAQEALSSKKKLERMYSMDRISDDIECRSWFPSESMHMLHSTASTLQACSAFRGYAER
KRRKRELDPSAMVQRNFHKHLRMVGRHRGRAAVSLDRQERECGRALSDVSQSKSDPVLDA
RFSHDMQSAYERLHREPSARWEEEREIELESCSRTDFIPPKVMLISSKVPKAEYIPNIIR
KDDPSIIAILYDHEHATFSDILDEIEKKLNEYRRGCKIWKLLIYCQ----------
----------FWKCLGSLMSKINTELNVVHIMGCYVLGNHNGEKLLDSLKSIMGPYR
VSFESPLELSAQGKKMIETYFSFRLYKLWKARQHSKLVDFDEIL

>NSMF_Amphioxus_(Branchiostoma_floridae)_(XP_035685233.1)

MGSAASRRKKNAATKLSAISAVTSATLQREQVLQQKNPENGQSSDTTNGETIATTLEQPQ
RMAAVQTEQRMARVIQRQFKKAVLHERPANTASPIPSSRMGQRSPRTAKLADPVGERSPL
LIRKFGQAKQTQDDQDGKGQGSSEDSTNNFMGQGSSDSSSKNTTGQGSAEGSSADSMTAD
QQRAKVKGQSSPSPSK---SGKISDTNSLTAVSPRDSKDKEQTPEGSEVSIS---GEKSE
QLTPEDENSLQSSQRTYCTVCCCYGDSLVSLQHRKERREHR-EQPWTEDTPQHEDL--QW
KIGNELQLNIDGYQDDISLRWKGEQLEEKAALAAHGFLPPRVVLVSSTVPKAKFLPLTV-
KEDPSVLVVSYDHQSDSFEKILSAIQSKLTSYCPGCKAKSLLMYMQGGPAHLYLKKNKVT
TVAKLKKHGDMVCFWKALGNMMSKLEPDSTVIHIMGCNVMGYPKGIDLFECLQEMMKPSI
VRFEAPLEMSIEGNAVINSYFDTNKYKLWKSQRYSNL---DRVF

S2.3 Supplementary Table

Table 2.S1 loci positions in the Japanese lamprey genome (*L. camtschaticum)* and the set of mapped lamprey transcripts.

GENE	Japanese Lamprey Scaffold	Predicted Exon Start-End (exons separated by comma)	Lampetra Transcripts	Petromyzon Transcripts	Lethenteron Transcripts
GnRH-I	KE993747.1	906761-906834;908511-908679;911990-912084;913397-913772	-	ENSPMAT00000000280	-
GNRH-II	KE993680.1	1936923-1937221;1938046-1938133;1938812-1939172	TRINITY_DN86187_c0_g1_i1	ENSPMAT00000011357	
GNRH-III	KE993747.1	897624-897770;899158-899255;901016-901389	-	ENSPMAT00000000282	-
ISLA	KE994173.1	6281-6575;7671-7849;8358-8516	TRINITY_DN41193_c0_g1_i2; TRINITY_DN41193_c0_g1_i1	Locus_139360_Transcript_1_1_Confidence_1	comp151224_c2_seq2;comp166395_c3_seq4;comp166395_c3_seq14;comp151224_c2_seq1;comp166395_c3_seq1
ISLB	KE993743.1	1385351-1385475; 1419805-1419995; 1422624-1422957; 1424324-1424575;1425281-1425452; 1426537-1427296	TRINITY_DN41193_c2_g1_i1	ENSPMAT00000002189	comp166395_c3_seq2;comp166395_c3_seq15;comp166395_c3_seq17;comp106466_c0_seq1;comp23347_c0_seq1;comp166395_c3_seq12
ISLC	KE993689.1	2879031-2879280; 2879907-2880113 ; 2880300-2880430; 2881373-2881612; 2882173-2882293	TRINITY_DN77023_c0_g1_i1; TRINITY_DN50605_c0_g1_i1	-	comp128550_c0_seq2; comp166395_c3_seq16; comp166395_c3_seq2; comp166395_c3_seq15; comp166395_c3_seq11; comp166395_c3_seq5; comp171797_c0_seq1; comp128550_c0_seq1; comp166395_c3_seq3
NELF	KE993681.1	5204239-5204322; 5205218-5205391; 5205551-5205858; 5205628-5205858; 5210104-5210186; 5210906-5210985; 5213783-5213837; 5215145-5215247; 5218356-5218539; 5225578-5225663; 5231951-5231984; 5234702-5234780; 5237795-5237897; 5239070-5239148; 5240373-5242758;	TRINITY_DN44340_c5_g1_i1 5204322	Locus_109486_Transcript_1_1_Confidence_1; ENSPMAT00000004778	comp158004_c0_seq1; comp158004_c0_seq2; comp158004_c0_seq3; comp47263_c0_seq1
FGF8/17	KE994328.1	16085-16199; 16319-16442; 17137-17399; 18330-18412;	TRINITY_DN6733_c0_g1_i1 TRINITY_DN94963_c0_g1_i1	Locus_48593_Transcript_1_1_Confidence_1	comp168772_c1_seq1;comp168772_c1_seq2;comp168772_c1_seq4;comp168772_c1_seq6;comp168772_c1_seq1;comp168772_c1_seq7;comp168772_c1_seq8;comp137350_c0_seq3;comp137350_c0_seq2;comp163850_c0_seq1
	KE993680.1	5383145-5383332; 5384580-5384623; 5390091-5390325; 5401265-5401372; 5407350-5407856;			

Chapter 3 Supplementary

S3.1 Sequence alignments used for maximum-likelihood phylogenetic tree

>Boleac.g00005463_1
LGIMNCVMGFFSIVLSAVALNFVGTEGIWCGICLLLTGTTGVLTGVKASDAMYIVSGVVGIISGL
VNAGFGLCVVAFSLYVVIGSFCFAGIPMCIVQAVY
>Boleac.g00005463_2
LGITELVIGFLIFCLDIASLAFGYNNGIWGGLYICLSGALGVVYSYKPSPTMINTNMSLSLLAAAT
SIMLVVHGFAAGLHITIAILATCTTEIRLIISHH
>Boleac.g00005468_1
LGVAEIAVGVVAIMLGVGAETFSFTHGIWIGLFMIVTGTLGIVISKKSTKAVYIANMVFAIFTAT-
-IGLILSGLAAGLHIVLTILFVVMITFAMIHANI
>Boleac.g00005468_2
LAITEIGIGVILIILGIGAVXFXYTQGVWCGMFIIVTGVLGVLHVRNATRRSYIASLTVSIISIFVVIG
LILSGVAAGIHIAIAVLCSIIILINIAHAVI
>Boleac.g00005475_1
LGFAETVIGGLSILLSVVVYGYNYIQGIWGGVLIVITGGLGIRMNGRPTKTWYIVNMVMAIITSN
VAAVVIMSGIFAGLHAIITILCFVAFILMVIHAAF
>Boleac.g00005475_2
LGITEVVFGALTIILGIVTYYSLSLQGIWGGLFVLITGVLGIRVKSNPTQHNYIANMALAIITAILS
AALILSALAAALHALVTLLFLVSFILTITHAAF
>Boleac.g00005475_3
VGWIEVLVGGLSIILAIISLTLTYSQGIWSGFFILIAGVLGISLYMNSSKCYYIFNVIIACATGVIGVA
AFLSGWAAVLHVTIGSLSALGVILAIVQVAY
>Boleac.g00005475_4
LGVAEILLGGISIILAIVAVIITYVQGVWGGLCVVITGVLAMRVRSNPTKCQFITSMTMAIITVVT
ATAAILSIIASSLHQTIAAINFVCFILTITQAGL
>Boleac.g00005475_5
LGVAEILLGGLYIILAIVFEWE----
RIPIGSIIFITGVLGTRLKTNQXRCLYIANMTMXIITAYFTGALILSFSFGALHILIAVIAFICVILSIC
HAAF
>Boleac.g00005475_6
LGVAEDLLGVLLIILAIVASYAAST-
GIVSGLLVLITGRLGIRVKTSPTRCMYIANLTMAITTAYVTVVVILCSVYGALHITIAVLGTFCVII
TIIHAAF
>Boleac.g00005475_7
LGIAEVILGGVALISGIVVFGFIRAQGIWGGSLIIITGILGILARKWPTKCLFILNTLMAAITTKLLA
VIGMSIAAAALHGTIALAAAISCLLTLINAVY
>Boleac.g00005475_8
FGNN----
GIIAMAIGHGELTWTYSQGIWGGFLILITGVLGIRAKSNPTKCLFIANITMATLTANVAAAFILSV
VAAGLHSMISLFTFVSIILTIIHASF
>Boleac_g00013053
LGITLVILGLLCIIVNLILTAATYRTGFWAGPYFIATGVLALICGKTLKQWMIKLSMSAAIVCANT
AGAIVLDVIAVALHATLAALSLLSGIASIIYAAY
>Boleac_g00012995

IGIAECILGGLSVIFAIVTVALTYIQGIWAGAVIVITGILGVCIRAKPTVEMYNANMSLAIITAVAS
VGIILSGLATALHLLITLCLFGGFISSIVHSAF
>Boleac_g00012484
LQITLVILGLLCIIVNLILTAATYRTGFWAGPYFIATGVLALICGKTLKQWMIKLSMSAAIVCANT
AGAIVLDVIAVALHATLAALSLLSGIASIIYAAY
>Boleac_g00005764
LGITEIIXGLTSCALGVVTPHFFHIQGIWCGFILIISGILGVVACNHPTKAMYRANLALSMLGVIV
MILLFTSAMAAGYFIALSVEAFVGIIITSAHCGY
>Boleac_g00005763
LGIAEIVLGGIIFILDIAI------
QGIWCGSFLIASGIVAVLTRSQPSYCMFSANIALSILAAIFMILVILSGIYAGLQSTNAVLAFSGLII
AIVHSAY
>Boleac_g00005762
LSITEISLGVIILILDIAT------
QGIWCGAFLIASGIVGVVTRSQPSYCMFNTNIALSILAAAFMTLVILSSIFAGLQVTIVVLAFTGL
VIAIIHCVY
>Boleac_g00005474
LGITEIVLGVLMLAFGIAAHPILNTPGIWCGIFLITTGVLGILIDVSASKNMYIVNMAFAIFSSLLT
ACMILAGFAMAMNGITIPLAFVAMIISIFHASA
>Boleac_g00005473
LGATEIVFGVLILVAGIAAYQFMNSPGIWCGIFLIITGVLGVLIDVSSSKNMYIVNMVFAILSSLLT
ACMILAGVAIALHGITIALAFVAMIISMVHACA
>Boleac_g00005467
LGIVEIVFGSIALILGIAAASWSYIQGIWCGLFLLATGLLGILLAKNLSRHMYIANMAMAIISAVIV
TGLVLSALAAGMHVVIAIVCFAAMVIAILHASF
>Boleac_g00005466
LGIAEVVVGVISIAFAVAARYWSLIQGIWCGLLTLTTGILGIVLVKKTSRNLYIANMVAAIITACT
VAGIVLSGLTAGMQIVVAILCLIAMILTIXHATF
>Boleac_g00008864
VGILEIASGLLMATLEIVSYYIYLTPAIWCGIIFCVSGAIAFCSWRKATKCNVIAGIVLSVISILACW
LFWYEGEIVAFHVIILLCSIGQFLTSCGHLFF
>Boleac_g00007139
FGFIEILLGIAAFGTCVVGGSMLYASGLWCGIVMIVSGILGVKVSKQSSLNVIRANMAASIVSSIF
MIMFMMSMFASAIYVFLSIIASAGFILNIVHAAF
>Boschl.g00029432
LGITEIVLGALSILFTIPTNVFTYAAGIWGGICMVAMGWLGIRVKANPSKGSYIGNMIMAIVTAN
ITSAMFIAVLAAALHIAVAFLNAAATIATIVHSAF
>Boschl.g00060878
LGIAEVANGSLSILLGLAALYTSFVPGVWCGLLLVATGSLGIRVRNKPSRGMYIANMVMTIITTA
CLVGFLLSMIQGGIDGCSWHQSEAEAIKMHVYCKY
>Boschl.g00064236_1
LGITNCVLGVIGIVVSSVAYNIVAAEGIWCGIFLILTGILGILTKVKPSEPLYIASGVVGILAAIISVG
FGLSVLGVSLYAVVGTMCLSAIPICIVHTVF
>Boschl.g00064236_2
LGVAEIALGFVLLCLEAAALAFGYSIGMWGGLYICASGALGVLYKFRPTMTVINANLAMCMLSA
TTSLIVVISGLAAELRIALAALGFIAVIVNVVHSAF
>Boschl.g00001820_1
FGVTEIVLGCILTIFGIAVLVLSVSQGIWCGVFLVVTGVFGTVLKRKCTRKILTINLAMTFIACVL
VACFLMSVIAAFLHIIIAVVCIIAFGILLAHAIL

>Boschl.g00001820_2
FGVTEIVLGCILIIFGIAVSVLSESQGIWCGVFVIVTGICGTVLRRNVTRKILTINLAMTFIACVLVA
CFLMSVIAASLHVVVAVASVIAFGILLAHAIL
>Boschl.g00001820_3
FGVTEIVLGCIMAIFGIAVEILSVTQGIWCGVFVVVTGVLGTVLRRRFTSKILTNNLAMTVIACVL
VGCFIMSVIAASLHIVVAVACVIAFGILIAHAIL
>Boschl.g00001820_4
FGVTEIVLGCIMTIFGIAVFVLSVTQGIWCGVFLVATGAYGTVLRKKITKKMLTVNLVMTFIACV
LVGCFLMSVIAASLHIVVAVACAIVLGILIVHAIL
>Boschl_g00069488
VGIMEIILGSISIILGIVTVILLFVLGLWCGVMVLITGIIGMTLKQHATRCLHIASMLFSMTCFVIA
AALISSIIMLGIHATLFVMNIVILIVTIVHVCF
>Boschl_g00057620
MGIMEVVLXLISIILGIVTSDAELMQGLWCGVMILVAGVMGITLKKHATRCLHIANMIFSIMCSF
IAAALVISIINVWIHATLSVINFVLLILTIVHACF
>Boschl_g00042820
MGITEVVLGSISFTLGIVMLALFAIVGLWCGVMALITGIIGIKVKQHGTRCLHITSMLFTMICSFI
AAALIISSMMMGIHATLSVINIVTLIVTVVHACF
>Boschl_g00014523
FGFIEILLGIAAFGTCVVGESMIYASGLWCGTVMMVSGILGVKVSKQSSLYIIRANMTASILSSIF
MIMFMMSMFAAAIYVFLSIIASAGFILNIVHAAF
>Boschl_g00001962
FGIAEILIGGLSVLLTIITNKFVYAPGIWGGIFFITTGILAVRIKANHSRYMYIASMTMAIITACVA
SSTIISGYAAAIHVIITLLQIASWIIAIIHAAF
>Boschl_g00058817
LGIGEIVMGVLSVLMTIISPGFTYSPGIWGGISVIITGMLAVCIKSNPSRCLYIANMTMAIVTACIT
SSTIISGFAAFLHSIIMLLNFVSWIITIIHAAF
>Boschl_g00067862
FGIGEIVMGGLSVLLTIIPGRYTYTPGIWGGIFIIVTGILAVRIKSNPSRCMYIANMIMAIVTACVT
SSTIVSGFAAAVHVLITLLNFVSWIITIIHAAF
>Boschl_g00049784
MGILEIVLGSISIILGIVTPPIAASQGLWCGIMVLIIGIMGIALKKHATRCLHIANMIFSFICSCIAAA
LINSMIMTGMHATLSVINFMILILTIVHSCF
>Boschl_g00000976
IGFLHIILGTICIVIESGANHYPSISGIWAGLFFLLAGVFSIVSGAGKYIYVLISGVALSIIGSVISSYF
AQKLIVEPPLLILIAASFFELIVSLFHCAI
>Boschl_g00022634
LGVAEIITGAVALAVEAAICSYINAQGIWCGLVAFILGMLGTAFYEKASRCMYRANIAMAIFSVI
VFVGFILSAITADLHALVAALCFLELIFAVIHSTI
>Cirob_g00000727
WGIAHICFGAAIFLEEVFGGSDIYM-
GIWAGAIFMGTGALGVQASRFPAKGVIIAGLVMSVISGLLSGVLILAIRLIVLSVLNIIIGLGETIVA
TVFTGY
>Cirobu.g00003356
IGITEIVFGGLTVILCIISTYATHSAGIWCSAVAIAAGALNITVVKKQAFSLIYASMICSIVGACFMI
LIIIETISSIIHVVMLLFTVVQMVLLINSVIF
>Cirobu.g00010723
VGAMQIMLGTLSITTWIYADNFAYIAGVWCGIFFLIAGILATVSSVKPNKCKVVAGMVMAIFAA
IFATMFGIETAGATSGPLSSVYGYYGY----YSTYY

>Cirob_g00015934
IGIIEIVFGISCIILDLVVSVYSLSPGIWGGIFVIIAGSLGIAFSSEADTCANAAIVVMNVIAALASSTL
SLETLATSLHGTTAGLGSICFCLGITQLIL
>Cirob_g00014126
LAVSETILGVLSIVLGSEISFSHYFEGLLPGIWVLIAGILGIFASRRPSSSVINAHMSFSVTAAIFLG
QVAAGIVLGYLCFALAVIGFLSFILCIMSASF
>Cirob_g00014577
TGWSQILIGFLSILFASIAFNILDTSGIWCSVFFIGSGVLGILAGGQPEFTQINRAMFGAMASTIFG
MMFAIEIASAALHSILVLLAFSEVVVALSQLYY
>Cirob_g00014575
VGVVKLVVGCVSVLLGVWAYFEVYS-
PLWCAPFFIVAGGKGVNLSCKPSDKKMFTTMILAIVSGILGVMYGFELVDSILHGILTALGFLNF
LFAKVLVAF
>Cirob_g00010730
LGVIQIMLGTLSIITWTYADVFVYISGIWCGIFFLIAGILATVSSFNPNRCKVVAGMVMAIFAAIF
ATMFGIEVAGSTLHGFMAFFAIAELVVAIVHSVY
>Cirob_g00013381
IGIAELVFGSICFVLGIIGEELLLVAGVWGGVFIAVSGLLAFLHSKNPTSCLLKGNLTVSILSAVMS
AIVIVGAFSAILHGITAFFGFVSLCLSITQFAY
>Cirob_g00013319
LGITHIVFGVLSLCLGISAFHVAYVSGVWCGLMSITVGIISIFAARRKTTCMIVCLMAIATVATIFA
AQVTLAVFGAIIHCLIGLLGLVEFVLFIVSAAL
>Cirob_g00012729
IGITHTVLGIISIVSGILANPRAYFSGIWAGTVFTALGIVLIHASKFKTICPFVAALVLSIIGVPVAL
MLVMSGPLEILYSVITLTGLVEGILIIISIVV
>Cirob_g00012445
LGVTQLVMGVLSLGIAITANYYFDTAGIWCGVFFLVCGILGCISGQKPGKCIIVASMVMTIISAVI
ATMLSLETASAALHSSNVLVAFIELVICIIHSAY
>Cirob_g00012427
LVTIQFIAGMSSILLGLAAQSFFDTAGIWCGCFFLATGVLGYLADKQSKQCLIIGTMAMSILSAGL
AVMFALDFLSMAIHFVLDIIAFGVLLVSVNIAAY
>Cirob_g00012421
IGITNIVLGCIAAVLAVTGCCNRDIAGIWGGVMLIISGTIAVNTARMMTKCMLVLTMLASAISAG
VGAMFGVELAFAVIHGVLSFIGFTGMILSIVNTVY
>Cirob_g00012378
LGITQVVIGALSVVTGVGSYYYYEIEGIWCGAILIVAGVLSIISSGRPTVCLISLSLAAAIIATAIGAL
FGIELSTVIIHGVMGLLGFIAMIIGIVMVVY
>Cirob_g00012377
LGIVQVVIAVLSITLSTTTGYSINMSGIWCGALFLVSGVFCIVSGKKPSHCWITTGLVLSILSAVFA
AMASFEVTCAVLHAVLGCLGTVDIIISCIYAGY
>Cirob_g00011972
LGVVQLVLGLLILFIGILDFYLAKVAGIWIGIAVIATGCVGVTSRSLRRHTTLILFLCLSAISLVFCV
LISIAISEMTMEQQQKDQGITDMSACFNQTST
>Cirob_g00011373
LAITETILGVLSIVLGIVLLFTIYSKGLWTGIWVLIAGILGIFASRRPSSSVINAHMGLSITAAVFSA
GQVAAGITLRTSIALAAFGFISFVLCIVSASF
>Cirob_g00008774

LGVFEMSFGFISILMGILQSPQSY-
EGIWAGIWVLVAGVLCIAAAKKTKLWLIQTHLGFAIAGAIFAIAIAAVLLSLGAFGLLSCIASASV
TTPFYNALT
>Cirob_g00006128
LGITETLLGCVSIVVGIVLLPLTFVEGVWCGVWILVAGVLGIISRKSLNSCMINCHMGFAITAAVF
AVQIALASLLTNLTMVLAIVGFVSFIVCIVSASF
>Cirob_g00006124
LGVAETVLGFLSMILGSVQIPYSY-
EGIWSGIWVLTAGILGIASKKRLTSCLINCHMGIAITASVFAFQIGLSSVSANIAVVLCIVGFLSFVI
CIASASL
>Cirob_g00005790
LSITETIIGSISIVIGIVLYYASPVEGIWCGIWILVAGCLGIAATRRATSCMINCHMGFAITASVFAQ
TAGSGILASGFSILLSILGFIAFIICIVSASY
>Cirob_g00005375
LCISETVLGSISILLGIVL------
AGIWCGIWILVAGILGIVASRRTTPCLINCHMGFAITASVFSIQVTVASVQSAFSIVLVIIGLAAFII
CIVSASY
>Cirob_g00004442
FGMAQVITGALCILLTGVVGTDHYKPGLWLGVLFIVTGALGVLSGKKPKNITINASMLFTILSAI
GAVMVILQVTASRLHITLTVLALLQLFATTAHSGY
>Cirob_g00004593
LGCLQIFIGVLSIILSSVVAYYIYS-
GIGSGLWYIVAGILGVVSGRRPSSCPITGALVITIFAAITAMAILQSLTSAIIYAVLAVLAFMEFIITI
VHSGF
>Cirob_g00001167
LGISQLILGTLSVFFAIAASYMVEVSGIWCGFFFLLAGGISMHAATKPYSCKIIAGMVLCIFAALF
AVLTIISIVYAAFHALLAITGFIELVLSIVHSAF
>Cirob_g00000918
FSVMIVVMGFVSVVLGITVPRFPITAGIWAGLVTIGVGAVGVISGHNANADQLKTLMVFTVIAV
PCLVQAVFAGMWAGLLG-LAAFSLFCFIFVIYGSSL
>Cirob_g00000884
LGILQMTIGVLCIGLAIVSYYPSSIKGIWCGIFFLVAGILGFLASKKNSRCLVIATMVLSIMSSVMA
NLIIISITTAILYSVLAFFALFEFVLSIVHSAV
>Cirob_g00000825
-
GIAQIVFGIISIVGEMGFALGASTGGMWAGALYCVAGVLGIHAGKGPTRRLVIAGMVVSIVSCIL
AMCTGISAMIGIFRLIVAIAGLGEFITSIVFSAF
>Cisavi_g00002069
LGVLQILIGILSVILCGVSVGYIYY-
SIGSSIWYLVAGSLGVASGRNPSNCPITGALVVTIFAAIASMCLLESLTTVAIHAVLAFLAGAQFII
TIVHSGY
>Cisavi.g00003309
LGIAAIIY----
AVLASRYPHPAYLYGIVSGLLFMTVGIISILASKRQTTCFIVCLMALSTMALLVATGIGIYGIICIL
HALLTAIGLVEFVLLIITTVF
>Cisavi_g00001274
LGVLQILIGVLSVILCGVSVGYIYY-
SIGSSIWYLVAGSLGVASGRNPSNCPITGALVVTIFAAIASMCLLESLTTVAIHAVLAFLAGAQFII
TIVHSGY

>Cisavi_g00004900
LAVTEIVLGSLSMLLGVVIYRYTYREGLWCGIWILVAGSIGVPASRQDKACLINCHMGFAITAAI
FAQGITSSTMAAGLSSFLAFVGFASFVICIVSACY
>Cisavi_g00004897
LSVTETVLGALSILFGIVVYSYSYTEGIWCGIWILIAGILGIAASRKTKSCLIHCHMGFAITAAVFA
SQVIASSILAGVSIVIAIAGFVAFIICIVSASY
>Cisavi_g00004896
LSVTETVLGSISVLLGIVISSITSVEGMWCGVWVLVAGILGIAASRTRKESLIHWHMGFAITGAV
FSAQVTASTILAGLSIIVAVIGFIAFIICIVSASY
>Cisavi_g00011498
IGISQMLIAITCVAISIPL------
SGIWCSIFFATMGGLNFFADKKSFKAMLTFSATSLYFGVLMLELIEFSIPATVLHVVLCSLALSEI
LLTILAIII
>Cisavi_g00010447
LGTFQILLGTLSIFTWSYADIFAFISGIWCGIFFLVAGILAAVSSVEPNNCKIVSGLVMSIFAAIFA
TMFGIEITAATLHGFMAFFAFAELIVAITHSVY
>Cisavi_g00010445
LGIIQILLGTLSIITWSYVDVFAFICGIWCGIFFLVAGILATVSSEEPNNCKIIAGMVMSILAAIFAV
MLAIESIGATLHGFMAFFAFAELVIAITHSVF
>Cisavi_g00010390
LGILQILLGSLSIFTWSYTDAFAYVAGIWCGIFFLVAGILAFVSSVDPNNCKIVSGMVMSIFAAIF
ATMFAIEVAGATLHGFMAFFAFAELIIAITHAVY
>Cisavi_g00009323
LSITEIVLGVISILLGIVVMPFNSNAGIWCGIWTSVAGILGIVASRSRQPCVINCHMGLAITAAVFA
QVFASSVLAISLAILIALVGFSALIICIVSACY
>Cisavi_g00000827
LGITMIIFGSLSAVLGVGTYYSFDIEGIWCGALVVLSGSLSIVSSRNPTVCVIALSTVVAVIAGAMG
TLFGLDLTIVSLHGVLSLLGFIEMLLAITLVSY
>Cisavi_g00000826
IGITLIVLGCVAAVIAVIAYHHGDVSGIWGGIIFIVTGTISVAARK--
QTCALVLIGVTTVASGIGVAMFGIELANTVAHGILAFIGFSGMIISIVNTVY
>Cisavi_g00000113
TGITEIVLGSVTFIAGIVSDELSNVAGIWCSVFAITAGAMNIAFMKNPIICLALSNMITSIVGASM
TILLILELACSIIHIIMLLMTIAQLVLLINSIVF
>Haaura_g00006236
FGITELVLGCSACLLCVISNDLAYVSGIWCGVFLIVSGILGVLTVKRPTVCMYNANMVMTIIGAM
MMTMFGISLVAAFIHALQTIISFAGLIICIIHSAY
>Haaura_g00002003
ISISLIVLGSLCICLSTLAYFFVFLPGIWSGLMVITTGIIGVIAAHRKTACFIVTMMVTSLVATLMT
AGLYIAANFQVMNSCGSLLSLASTVLLLMFSLT
>Haaura_g00001705
FGITELVLGCSAFLLCVISGAITYIPGIWGGIFLIVIGILGVLTRKRPTVCMYNANMVMAIICALM
MTMFGISVPAAFIHTIQSIISFAGLIICIIHSAY
>Haaura.g00001706

AGVTIIQPSPALIFTFTQGIWCGIFPIISGILGIVARKKPTNCMYNANMAMTIISAVMLLLFSLSIL
ATACHIIISILSFACLIITIVHSAY
>Haaura.g00004400_1

LGIAEVVLGAISMLFALTSYVPSLIAGIWCGVFVVISGVFAVLLDKNSEIRRIKLNMILCIVASIFQ
FQIAVTTVAAMMSQVIKASGYASIALCVIHAAV
>Haaura.g00004400_2
FGFAKILLGVFACLIGIIGFDISSTKGLWCGLIVITSGVLGILTNKNNIRTMLGKDRVMSIVAIFFV
VQFFMSIVTFHINKCFEFLSMGTLVVCLLYSIH
>Harore_g00005028_1
LGIAEVVLGSISMLFALTSYVPSLIAGIWCGVFVVISGVFAVLLDKNSEIRRIKLNMILCIVASIFQF
QIAVATVAAMMSQVIKASGFASIALTVIHAAV
>Harore_g00005028_2
FGFSKILLGVFACLIGIISFDMSSTKGLWCGLIVITSGVLGILTNKNNIRTMLGKDRVMSIVAIFFV
VQFFMSIVTFHINSCLEFLSMGTLVVCLLYSIH
>Harore_g00012923
FGITELVLGSTACLLAVISDGLTYVSGIWCGVFLIVSGILGVLTVKRPTVCMYNANMVMTIIGAI
MMTMFGISVTAAFIHALQTIISFAGLIICIIHSAY
>Harore_g00003599
FGITELVLGCTAFLLCVISGVITYISGIWGGIFLIISGILGVLTRKKPTVCMYNANMVMAIICAVM
MTMFGISVPAAFIHSLQSLISFAGMIICIIHSAY
>Harore.g00005144
-
GATVPPQAGVTIIQPTPAVVFTYMQGIWCGIFPIISGILGIVAKKKPTNCMYNANMAMTIVSAV
MLLLFSLSILAMACHIIISIISFVCLIITIVHSAY
>Harore_g00012525
IGMAQIVFGCISVISCAATFSLDRSAGIWCGVLAIIAGSLSISARYRP-
YTLILSNMVVSIVTGLFGAYFIFSLRNLSLYLFQTSIALCWTTLAINAVTF
>Harore_g00015316
ISISLIVLGSLCICLSTLAYYFVFLPGIWSGLMVITSGIIGVIAARRKTACFIVSLMVTSLVATLMTS
SVGIYIAANVINSCGGLLSLASTVLLLVKFSL
>Moocci_g00010729
ISVAILIIGVICVVLGSGATTEHPTAGIWSGLMFIAMAIVGIIAARRKTTCMILSLLVLSVPCLIIAT
GTGFAIYSEIINAVLAFLIFVVIVLDIILFAF
>Moocci_g00008258
ISVTILLIGILAITSGIVTTIPTSVAGLWVGLMFVAVGTVGIIAAQRKSTCMILSLLVTSVPCFLFGA
GVGLGIYSEISNGILAFISLIATILDIILFAF
>Moocci.g00017042
I----VIAGMVLCIFSAIA-------RFFLGLGIIIVDSFS-----
DGDKTMHYVSLAASVLMFITASAFCCGAVCCTMHYVSLAASVLMFITAIVHSAF
>Moocci.g00017046
LGVLEILTGLVVLILGIVSSLYYYT-GIWTGLFILVAGIFGILNGRSPTDCNIIGGVVVTVLCAI--
GSLGGGITAAAASSSSECAGFYEF---------
>Moocci.g00017878
-MVLAIVT------
LRLPYVELSYVSGVWCGAFLITIGILGIKAKENPSQCLIVANMVMSIIASIFMIEFGFSVVSSALHA
ILAVISIVSFVLLVVHSAF
>Moocci.g00024436
A--
IVIATGASSWXXXXREYEYYVIFSILTGIFFFISGFFGILSGRNPSDCNVITGLVMTIFTALSALVFY
QDVELMAMHIILMVTGGIIFIVSIVHSAY
>Moocci.g00027903

LGITQCGLGMFVIVLDILMPSPSLLSPLWVSIVFVGAGVSTVVAYKKKGRFMTAIGFACSILSVLF
SIAISMMGMAQIIGSVLLLTVVAEVVISATNTLI
>Moocul_g00012680
LGMIEVSIATVCVVFGVVLGFISYLSAIWCPIPLFVTGGLGICARKNPSRCLIIAGMVMAIITSIFM
TMLIISISGATFHSINSILSLSAFVIVIVHSVY
>Moocul.g00014061
LGIIEIIIGLISLVMALLLNYENFYSEFVCITPLILSGCLGICASRKR----
LVSNFFSTILATAATTFAVFNTMASAYHLMIAALNSAAFFIVLLHFGF
>Moocul_g00012612
FGIIQLISGLVVIGFAGYGYWMLCI-
GFFTGIMFLISGILGLVSGCKPTNCNVIGGLVMTVLSAIAADLVIYQVVSVAVQYSAMAFGILTFI
VAIVHSAF
>Moocul.g00008648
IAIIEVVTALTVIIIGAISNFLLTI-
GIGNGIFFLMIASMAIASASCPNDCNVVAGMILCLFSAIITGFLLIGSLLAILHMIAIACSFIMFLTA
IVHSAL
>Moocul_g00008014
LGIIEIVIASICLISGIILGGISYVTAIWGTLPLYVTGGLGICSKNSKSRCLIIAGMVMAIITSLFMVLL
GFSITGAAFHSINALLSFAAFVIVIVHAAF
>Moocul_g00004119
IGIIELSIGGLLIVLGVVNLSNAYS-
GFWNGMFVCVAGSLGVAAGKSPNQCTVTGGLVMTIISAIFGIAEGLQMSVTAVHGVLLFLCIAG
FVLSIVHSGF
>Moocul_g00003041
LGITETTLGLIVFILGIVAHGYTYIGGIWGGLVVLTLGVCGVFVGYSKKNDLLTAQIVIGIIVAIITV
SMCIESASAYCHAVNASLLFISFVVAIVHASM
>Moocul_g00001460
IGVGQLLIGIFCVMAGVFSWDVFYLEGIWIGVLAFITGLFSVSSRSQRKKFMMRGFTVMTVLTL
VLVGLVIADLVTVNLTENETIISTTEQTTTINTTDN
>Phfumi_g00005311
LGFIHLLLGIFTLCIGLFDFSQAKATGIWVGTVIIITGMIGLVTSRSKRRTTILLFGGFSVACVLFS
VLMGISISEMGTNCTIQITNLTTPALRCAQTTF
>Phfumi_g00005291
LGITEAIIGLICCVMGILLVFLFNVEGLWAGVWTFVAGVLGAACGKTVTPCLINAYLAFSIVASLF
ATMFGINVTFAGVEATLAILSFAAFVICIVSACY
>Phfumi.g00002448
LGAVEAVLGLIGLVLGIVAQPYFRVQGIWGSVWVYV----------
HMAFSILAAVISAMLISSFVFCGLEIGIAVASLCTLVLCVVSAVL
>Phmamm.g00018199
LGIVEIAIGAISIICGIIAFTFALLTGIWCGVFIIGGMVMSIVSAL--AGGVLL---GMSVTAAI--
LSFGEEEIVFAHHGKLEVNTRSATDDIGSHTSS
>Phmamm.g00003413
LGIVEVVLGIIGFVLGIVASPYFRVQGVWGNIWVYV----------
HMALSIIAAAISGAMLLSSFVFGLEIGIAIASLFTLVLCVVSAVL
>Phmamm.g00006033
LGATETVLGLICSLLGTVAYYSTYLEGLWGGVWVIIGGSLGIACGKTVSSCLQLSGGNVMYVAT
N--QAFGQRPQQTTVPYVVQGVPDTAIPQTVLQAAP
>Phmamm.g00017658

IAIFQLLLSLLCFVFGVLTYLMDTVPGIWCSVFALAAAWKVLSTINKPTYTKLFWSNVTNWTA
SFFQFVVAIQVYSLAFISVGLVTSFILLLICVNAKKF
>Phmamm_g00018375
LGITEAVIGMICCIMGILLLYVFNLEGLWSGIWILIAGCLGAACGKTVTPCLLNSYLAFSIVASIFA
NMFGINVPFAGVEATLAFLSFTAFVICIVSACY
>Phmamm_g00016143
LGIAQVTIGIMSCTMGLLLVTYTHHEGLWCGFWILASGILGVICGKTINPCLVNNLLGFSIATAIF
SVMLCLGGLYAVIEIFLAIFGFFGFFLSYASAKY
>Phmamm_g00013947
IGATLVKISVLAMALSIPAYDYTYLRGIWCSVMSLVAGALCYKSVPTKNNAQITATFAFGIIASV
FALGVEVGAAVQQVHMIVSACAMYQLALSIIAVTY
>Phmamm_g00013192
LGAIEALLGTICSLLGLAVNKYTTTEGLWGGIWVIIGGALGIACEKKPTPCLVNCHMAFSIIGAV
ASSMFSTSVPFAACDAAIAVLSFITLALCITSACL
>Phmamm_g00010988
LGFIHLLLGVFTLCIGIFDFSQAKATGIWVGAVIIITGMIGLVTSRSKRRTTILLFGGFSVACVLVC
ILIGISISEMGTNCTIDVTSVTTPTVRCAQTTF
>Phmamm_g00009351
IGITQLVIALLCIVISIPLIHRSSASGIWCSAFFILMSGLNFSSNMRPFRVMVAFAATSVYFGVLML
ESVELAMLISCLHIVLCMFALTEIVLTIMGAVL
>Phmamm_g00005433
LGIFQILLGSVCIMMSIPYQTLTFISGIWGGSCSLIAGVLGVILSTKSSKCYIVTQLIASILAFLATA
MLGIEVAQTMYYLTMCLLALSSVIVAAIQSGY
>Phmamm_g00002791
LSIAELVLACISFVAGIGFINPAFMPGIWCGLLYGAAAIVGLKAAKHRSTCLIVTLMVLAITGAVF
ALLVGILGAAGAFNILLAVIGLVQMILLISSTLT
>Brflor_XP_035666292.1
LSITQIVCGSLAILLGIAISQFSWTYPIWTGAFFLSTGIIGIFAAKRKTNCMIIPLMVLSVLSSILAA
TLVLAAIALTFDVLLILLALLELGLAIATSVM
>Brflor_XP_035666291.1
LSITQIVCGSVAILLGIAISELFQTYPIWTGALFLSTGIIGIFAAKRKTKCMIIPLMILSILSSILAVTL
VLAAIALTFDVLLILLALLELGLAIATSVM

S3.2 Codon alignments used in the codon substitution models

S3.2.1 Clade 1

4 372
Boschl_g00049784

ATGGGAATTTTGGAGATCGTGTTAGGATCGATTTCCATTATTCTTGGAATCGTCACGGTGACAATTGTATCTAAC
GAACTTCTAAATCCATCCATCGCATTACCTCCTATCGCGGCAAGTTCGCAAGGGCTGTGGTGCGGTATCATGGTC
CTTATCATTGGCATTATGGGGATAGCATTAAAAAAACACGCTACCAGATGTCTTCATATCGCGAACATGATCTTC
AGCTTCATATGTTCTTGTATCGCCTGCGCAGCACTAATCAACTCCATCATGACATTTTTATCTCTTCTCATATCA
GTTGCTGTAGTCGGGATGCACGCTACGCTGTCAGTTATCAATTTCATGATACTGATATTGACAATTGTACAC
Boschl_g00057620

ATGGGAATTATGGAAGTTGTGTTAGGATTAATTTCTATTATTCTTGGAATCGTCACGGTGATAATTGTATCCAA
TCAAATCGACACAGACCCATTCATACCTTCTGATGCAGAGCTAATGTCGCAAGGGCTATGGTGCGGCGTCATGAT
CCTTGTCGCTGGCGTTATGGGGATTACACTGAAAAAACACGCAACCCGATGTCTTCATATCGCGAACATGATCTT
TAGCATCATGTGTTCTTTCATCGCGTTAGCAGCACTCGTCATCTCCATTAATGTATTTTTATCATTATCCACATC
GCACACTGCAGTCTGGATCCACGCAACGCTGTCTGTCATCAATTTCGTGCTACTGATATTGACAATTGTACAC
Boschl_g00069488

GTGGGAATCATGGAGATCATCTTAGGATCGATTTCCATTATTCTTGGAATCGTCACGGTGACAATTATATCCAAG
AAAGTCGACGAAGTTGAAGTCATATTACTGTTTGTCGGGTTAAGTTCGCAAGGACTTTGGTGTGGGGTCATGGT
ACTTATCACAGGCATTATAGGAATGACGTTGAAACAACACGCTACAAGATGTCTTCATATCGCGAGCATGCTCTT
TAGCATGACCTGTTTTGTCATCGCGTTAGCGGCACTAATCAGCTCAATAATGTTATTTATATCAATTATAATCGC
AACGCTGTTGGTTGGGATTCACGCAACTCTGTTTGTTATGAATATCGTGATACTGATTGTGACAATTGTACAC
Boschl_g00042820

ATGGGAATTACTGAGGTCGTGTTAGGATCGATCTCCTTTACTCTTGGAATCGTCATGGTGACAATTGAATCCAAT
CAAATCGGCGAACTTCGTCTCGCATTATTTGCTATCGGAGTAAGTTCGCAAGGACTTTGGTGTGGGGTCATGGCA
CTTATCACAGGTATTATAGGGATTAAGGTAAAACAACACGGTACCAGATGTCTTCATATCACGAGCATGCTCTTC
ACCATGATATGTTCTTTCATCGCGTTAGCAGCTTTAATCATCTCAATGATGATATTTGTATCATTTTCATTCTCG
GTTACTGCAGTCGGGATTCACGCAACGCTGTCTGTTATCAATATCGTGACACTAATTGTGACAGTAGTACAC

S3.2.2 Clade 2

4 396
Boschl.g00029432

CTTGGAATCACTGAAATAGTACTGGGGGCGTTATCTATCCTATTCACGATACCCACACTCGCGATTGTGTCTCGA
GAAACTAAACGATTTGAGTACACATATTCGTACTACATCAACAACGTATTTACATACGCCTCCGCTGGAATATGG
GGTGGAATTTGCATGGTGGCTATGGGCTGGCTTGGAATACGCGTCAAAGCCAATCCGTCGAAGGGCAGCTACATT
GGCAACATGATTATGGCGATTGTGACTGCCAACATCACATTTTCTGCTATGTTCATTGCTGTATTGGCGGCCGGC
CTATCTTACTATTCTAGCGCGCTGAGAGCGTTGCATATCGCTGTGGCGTTTCTAAACGCCGCGGCCACAATCGCC
ACCATCGTGCACTCGGCGTTT
Boschl_g00001962

TTTGGCATCGCAGAGATATTGATTGGTGGATTGTCGGTTCTACTGACTATAATAACTTTGTCTATAAATAACCGT
GCAGAAAGAAACTATTCGATACGACGCTACTACTACCCCAACAATAAGTTCGTGTACGCAGCGCCGGGAATATGG
GGAGGCATATTTTTTATTACTACCGGCATTTTGGCTGTGCGTATAAAGGCTAACCATTCCCGATATATGTACATT
GCAAGTATGACCATGGCCATTATTACCGCATGCGTTGCATTTTCTTCCACCATCATATCCGGCTATGCAGCCGCA
TTTTCCGATTATTCACCTGAATTAATCGCCATTCATGTGATCATAACATTGCTACAGATAGCATCATGGATTATA
GCGATCATCCACGCCGCCTTC
Boschl_g00058817

CTCGGCATCGGAGAGATAGTGATGGGTGTATTGTCGGTTCTAATGACCATAATTTCTCTGGTTATCAATCAGCAT
GCAATAAAATCGTATTGGAAGGAAAATAGACCATCCATCTTCCCTGGTTTCACGTACTCTTCACCTGGGATATGG
GGTGGTATATCTGTTATTATAACAGGGATGTTGGCAGTGTGCATCAAGTCTAATCCTTCCAGGTGCCTGTACATC

GCAAATATGACCATGGCGATTGTTACCGCTTGCATCACATTTTCGTCGACCATCATATCCGGATTTGCAGCCGTA
TGGTCTTTTTATTCATCTGCATTAATCTTCCTACATTCTATAATAATGCTGCTGAATTTCGTGTCATGGATCATA
ACGATTATTCACGCCGCGTTC
Boschl_g00067862

TTTGGCATCGGGGAGATAGTGATGGGTGGATTGTCCGTTCTATTGACCATAATACCACTAGCTACCAGCTCGCAG
GCAGCTTTTGCGTACCGGAATCAAAGCGGACGATATACCTATACAGGTTTCACGGACTCATCACCTGGGATATGG
GGAGGCATATTCATTATTGTAACAGGGATTCTGGCAGTGCGCATCAAATCCAATCCTTCCAGATGCATGTACATC
GCTAATATGATCATGGCGATTGTCACCGCTTGTGTCACATTTTCGTCGACCATCGTATCTGGCTTTGCAGCCAGT
GCATCTTTTTATTCGTCGGAATTAATCGCCGTACATGTCCTCATAACGCTGCTGAATTTCGTCTCATGGATCATA
ACAATCATTCACGCCGCGTTC

S3.2.2 Clade 3

4 372
Boschl.g00001820_1

ATCTTCCGGTTCGGCGTGACTGAAATAGTTTTAGGATGCATTTTGACTATATTCGGAATCGCGGTGGTCTTGATC
GGTCCTACTAACGGCGTCGACACCATTTGCGGAGTGGCCTTGGTCCTGTCAGTGTCTGCACAAGGAATATGGTGC
GGCGTGTTCCTGGTGGTTACTGGTGTGTTCGGAACGGTGTTGAAGAGGAAATGTACGCGAAAGATCTTGACTATC
AACTTGGCAATGACATTTATAGCCTGTGTTTTGGTGTTCGCTTGCTTCCTCATGTCGGTAATCGCAGCCGTAGCA
GAGAGGGGTTTGATTATGTTCCTGCACATTATTATTGCGGTGGTGTGCATTATCGCATTTGGCATCCTACTC
Boschl.g00001820_4

ATCTTCAGGTTTGGCGTGACTGAAATAGTTCTAGGGTGCATTATGACTATATTTGGGATCGCCGTTGCCGTTGTC
GGTGCTACCGTCATTCCAACCAGTATTTACAGTGTCGATTTCGTCCTGTCTGTAACCCCCCAAGGAATATGGTGC
GGCGTATTTCTAGTAGCTACTGGTGCTTATGGGACAGTTTTGAGGAAAAAAATTACGAAAAAGATGCTGACTGT
TAACTTGGTGATGACATTTATTGCCTGTGTTTTGGTGTTCGGTTGCTTCCTTATGTCGGTGATAGCAGCCGTTAT
TTCCATTAACGCTATAATGTCCTTGCACATCGTTGTTGCGGTGGCGTGCGCTATCGTGCTTGGTATCTTAATC
Boschl.g00001820_3

ATCTTCCGTTTCGGCGTAACTGAAATAGTTTTAGGCTGCATTATGGCAATATTCGGGATCGCCGTGACTGTTGTG
GTTGTCTTGAGCGGTGTGAACAATACTTGCGAAGTCGCTGAGATTCTGTCAGTGACTTCCCAAGGCATATGGTGC
GGCGTATTTGTGGTGGTTACTGGTGTGTTAGGAACGGTTTTGAGGAGGAGATTTACGAGCAAAATCTTGACTAA
CAACTTGGCGATGACAGTTATAGCCTGTGTTTTGGTATTCGGTTGCTTCATTATGTCTGTAATCGCGGCTGTAGC
AGGTCTGGGCTCGATTATGTCCCTGCACATCGTTGTTGCGGTGGCGTGCGTTATCGCGTTTGGCATACTAATC
Boschl.g00001820_2

ATCTTCCGGTTCGGCGTGACTGAAATAGTTTTAGGATGCATTTTGATAATATTCGGGATTGCAGTGGTCTTGTTT
CGTGAAACTCCTCACGGAGACAGTATTTGCACAGTCGCCTCAGTATTGTCAGAGTCTTCACAGGGCATATGGTGC
GGCGTATTTGTGATCGTTACTGGTATCTGCGGAACGGTATTGAGGAGGAACGTTACAAGAAAGATCTTGACTAT
CAACCTGGCGATGACATTTATAGCCTGTGTTTTGGTGTTCGCTTGCTTCCTTATGTCTGTAATCGCGGCAGTAGC
AGATGCCAGCGTGATTATGTCCCTGCACGTTGTTGTTGCGGTGGCGTCCGTTATCGCGTTTGGCATCCTACTC

S3.2.2 Clade 4

8 378
Boschl_g00049784

ATGGGAATTTTGGAGATCGTGTTAGGATCGATTTCCATTATTCTTGGAATCGTCACGGTGACAATTGTATCTAAC
GAACTTCTAAATCCATCCATCGCATTACCTCCTATCGCGGCAAGTTCGCAAGGGCTGTGGTGCGGTATCATGGTC
CTTATCATTGGCATTATGGGGATAGCATTAAAAAAACACGCTACCAGATGTCTTCATATCGCGAACATGATCTTC
AGCTTCATATGTTCTTGTATCGCCTGCGCAGCACTAATCAACTCCATCATGACATTTTTATCTTCAGTTGCTGTA
GTCGGGATGCACGCTACGCTGTCAGTTATCAATTTCATGATACTGATATTGACAATTGTACACTCGTGTTTCACC
TGT
Boschl_g00057620

ATGGGAATTATGGAAGTTGTGTTAGNATTAATTTCTATTATTCTTGGAATCGTCACGGTGATAATTGTATCCAA
TCAAATCGACACAGACCCATTCATACCTTCTGATGCAGAGCTAATGTCGCAAGGGCTATGGTGCGGCGTCATGAT

CCTTGTCGCTGGCGTTATGGGGATTACACTGAAAAAACACGCAACCCGATGTCTTCATATCGCGAACATGATCTT
TAGCATCATGTGTTCTTTCATCGCGTTAGCAGCACTCGTCATCTCCATTAATGTATTTTTATCATCGCACACTGC
AGTCTGGATCCACGCAACGCTGTCTGTCATCAATTTCGTGCTACTGATATTGACAATTGTACACGCGTGTTTTTG
CTGC
Boschl_g00069488

GTGGGAATCATGGAGATCATCTTAGGATCGATTTCCATTATTCTTGGAATCGTCACGGTGACAATTATATCCAAG
AAAGTCGACGAAGTTGAAGTCATATTACTGTTTGTCGGGTTAAGTTCGCAAGGACTTTGGTGTGGGGTCATGGT
ACTTATCACAGGCATTATAGGAATGACGTTGAAACAACACGCTACAAGATGTCTTCATATCGCGAGCATGCTCTT
TAGCATGACCTGTTTTGTCATCGCGTTAGCGGCACTAATCAGCTCAATAATGTTATTTATATCAGCAACGCTGTT
GGTTGGGATTCACGCAACTCTGTTTGTTATGAATATCGTGATACTGATTGTGACAATTGTACACGTTTGTTTTTG
TTGT
Boschl_g00042820

ATGGGAATTACTGAGGTCGTGTTAGGATCGATCTCCTTTACTCTTGGAATCGTCATGGTGACAATTGAATCCAAT
CAAATCGGCGAACTTCGTCTCGCATTATTTGCTATCGGAGTAAGTTCGCAAGGACTTTGGTGTGGGGTCATGGCA
CTTATCACAGGTATTATAGGGATTAAGGTAAAACAACACGGTACCAGATGTCTTCATATCACGAGCATGCTCTTC
ACCATGATATGTTCTTTCATCGCGTTAGCAGCTTTAATCATCTCAATGATGATATTTGTATCATCGGTTACTGCA
GTCGGGATTCACGCAACGCTGTCTGTTATCAATATCGTGACACTAATTGTGACAGTAGTACACGCTTGTTTTTGC
TGT
Boschl.g00029432

CTTGGAATCACTGAAATAGTACTGGGGGCGTTATCTATCCTATTCACGATACCCACACTCGCGATTGTGTCTCGA
GAAACTAAACGATTTGAGTACATCAACAACGTATTTACATACGCCTCCGCTGGAATATGGGGTGGAATTTGCATG
GTGGCTATGGGCTGGCTTGGAATACGCGTCAAAGCCAATCCGTCGAAGGGCAGCTACATTGGCAACATGATTATG
GCGATTGTGACTGCCAACATCACATTTTCTGCTATGTTCATTGCTGCGGCCGGCCTATCTTACTCTAGCGCGCTGA
GAGCGTTGCATATCGCTGTGGCGTTTCTAAACGCCGCGGCCACAATCGCCACCATCGTGCACTCGGCGTTTGGAT
GT
Boschl_g00001962

TTTGGCATCGCAGAGATATTGATTGGTGGATTGTCGGTTCTACTGACTATAATAACTTTGTCTATAAATAACCGT
GCAGAAAGAAACTATTCGTACCCCAACAATAAGTTCGTGTACGCAGCGCCGGGAATATGGGGAGGCATATTTTTT
ATTACTACCGGCATTTTGGCTGTGCGTATAAAGGCTAACCATTCCCGATATATGTACATTGCAAGTATGACCATG
GCCATTATTACCGCATGCGTTGCATTTTCTTCCACCATCATATCCGCAGCCGCATTTTCCGATTCACCTGAATTAA
TCGCCATTCATGTGATCATAACATTGCTACAGATAGCATCATGGATTATAGCGATCATCCACGCCGCCTTCTGCT
GT
Boschl_g00058817

CTCGGCATCGGAGAGATAGTGATGGGTGTATTGTCGGTTCTAATGACCATAATTTCTCTGGTTATCAATCAGCAT
GCAATAAAATCGTATTGGTCCATCTTCCCTGGTTTCACGTACTCTTCACCTGGGATATGGGGTGGTATATCTGTT
ATTATAACAGGGATGTTGGCAGTGTGCATCAAGTCTAATCCTTCCAGGTGCCTGTACATCGCAAATATGACCATG
GCGATTGTTACCGCTTGCATCACATTTTCGTCGACCATCATATCCGCAGCCGTATGGTCTTTTTCATCTGCATTA
ATCTTCCTACATTCTATAATAATGCTGCTGAATTTCGTGTCATGGATCATAACGATTATTCACGCCGCGTTCTGT
TGC
Boschl_g00067862

TTTGGCATCGGGGAGATAGTGATGGGTGGATTGTCCGTTCTATTGACCATAATACCACTAGCTACCAGCTCGCAG
GCAGCTTTTGCGTACCGGTATACCTATACAGGTTTCACGGACTCATCACCTGGGATATGGGGAGGCATATTCATT
ATTGTAACAGGGATTCTGGCAGTGCGCATCAAATCCAATCCTTCCAGATGCATGTACATCGCTAATATGATCATG
GCGATTGTCACCGCTTGTGTCACATTTTCGTCGACCATCGTATCTGCAGCCAGTGCATCTTTTTCGTCGGAATTA
ATCGCCGTACATGTCCTCATAACGCTGCTGAATTTCGTCTCATGGATCATAACAATCATTCACGCCGCGTTCTGT
TGC

S3.2.2 Clade 5

12 363
Boschl_g00049784

ATGGGAATTTTGGAGATCGTGTTAGGATCGATTTCCATTATTCTTGGAATCGTCACGGTGACAATTGTATCTAAC
GAAAATCCATCCATCGCATTACCTCCTATCGCGGCAAGTTCGCAAGGGCTGTGGTGCGGTATCATGGTCCTTATC
ATTGGCATTATGGGGATAGCATTAAAAAAACACGCTACCAGATGTCTTCATATCGCGAACATGATCTTCAGCTTC
ATATGTTCTTGTATCGCCTGCGCAGCACTAATCAACTCCATCATGACATTTTCAGTTGTAGTCGGGATGCACGCT
ACGCTGTCAGTTATCAATTTCATGATACTGATATTGACAATTGTACACTCGTGTTTCACCTGT
Boschl.g00001820_4

TTTGGCGTGACTGAAATAGTTCTAGGGTGCATTATGACTATATTTGGGATCGCCGTTGCCGTTGTCGGTGCTACC
GTCATTCCAACCAGTGTCGATTTCGTCCTGTCTGTAACCCCCCAAGGAATATGGTGCGGCGTATTTCTAGTAGCT
ACTGGTGCTTATGGGACAGTTTTGAGGAAAAAAATTACGAAAAAGATGCTGACTGTTAACTTGGTGATGACATT
TATTGCCTGTGTTTTGGTGTTCGGTTGCTTCCTTATGTCGGCAGCCGTTATTAACGCTATAATGTCCTTGCACAT
CGTTGTTGCGGTGGCGTGCGCTATCGTGCTTGGTATCTTAATCGTGCACGCGATACTGTCCCGG
Boschl.g00001820_3

TTCGGCGTAACTGAAATAGTTTTAGGCTGCATTATGGCAATATTCGGGATCGCCGTGACTGTTGTGGTTGTCTTG
AGCGGTGTGAACGAAGTCGCTGAGATTCTGTCAGTGACTTCCCAAGGCATATGGTGCGGCGTATTTGTGGTGGTT
ACTGGTGTGTTAGGAACGGTTTTGAGGAGGAGATTTACGAGCAAAATCTTGACTAACAACTTGGCGATGACAGT
TATAGCCTGTGTTTTGGTATTCGGTTGCTTCATTATGTCTGCGGCTGTAGCAGGCTCGATTATGTCCCTGCACAT
CGTTGTTGCGGTGGCGTGCGTTATCGCGTTTGGCATACTAATCGCCCACGCTATACTATCCCGA
Boschl.g00001820_1

TTCGGCGTGACTGAAATAGTTTTAGGATGCATTTTGACTATATTCGGAATCGCGGTGGTCTTGATCGGTCCTACT
AACGGCGTCGACGGAGTGGCCTTGGTCCTGTCAGTGTCTGCACAAGGAATATGGTGCGGCGTGTTCCTGGTGGTT
ACTGGTGTGTTCGGAACGGTGTTGAAGAGGAAATGTACGCGAAAGATCTTGACTATCAACTTGGCAATGACATT
TATAGCCTGTGTTTTGGTGTTCGCTTGCTTCCTCATGTCGGCAGCCGTAGCAGGTTTGATTATGTTCCTGCACAT
TATTATTGCGGTGGTGTGCATTATCGCATTTGGCATCCTACTCGCCCACGCGATACTATCGCGG
Boschl.g00001820_2

TTCGGCGTGACTGAAATAGTTTTAGGATGCATTTTGATAATATTCGGGATTGCAGTGGTCTTGTTTCGTGAAACT
CCTCACGGAGACACAGTCGCCTCAGTATTGTCAGAGTCTTCACAGGGCATATGGTGCGGCGTATTTGTGATCGTT
ACTGGTATCTGCGGAACGGTATTGAGGAGGAACGTTACAAGAAAGATCTTGACTATCAACCTGGCGATGACATT
TATAGCCTGTGTTTTGGTGTTCGCTTGCTTCCTTATGTCTGCGGCAGTAGCAAGCGTGATTATGTCCCTGCACGT
TGTTGTTGCGGTGGCGTCCGTTATCGCGTTTGGCATCCTACTCGCCCACGCGATACTATCTCGG
Boschl_g00057620

ATGGGAATTATGGAAGTTGTGTTAGNATTAATTTCTATTATTCTTGGAATCGTCACGGTGATAATTGTATCCAA
TCAAACAGACCCATTCATACCTTCTGATGCAGAGCTAATGTCGCAAGGGCTATGGTGCGGCGTCATGATCCTTGT
CGCTGGCGTTATGGGGATTACACTGAAAAAACACGCAACCCGATGTCTTCATATCGCGAACATGATCTTTAGCAT
CATGTGTTCTTTCATCGCGTTAGCAGCACTCGTCATCTCCATTAATGTATTTTCGCACGCAGTCTGGATCCACGCA
ACGCTGTCTGTCATCAATTTCGTGCTACTGATATTGACAATTGTACACGCGTGTTTTTGCTGC
Boschl_g00069488

GTGGGAATCATGGAGATCATCTTAGGATCGATTTCCATTATTCTTGGAATCGTCACGGTGACAATTATATCCAAG
AAAGAAGTTGAAGTCATATTACTGTTTGTCGGGTTAAGTTCGCAAGGACTTTGGTGTGGGGTCATGGTACTTAT
CACAGGCATTATAGGAATGACGTTGAAACAACACGCTACAAGATGTCTTCATATCGCGAGCATGCTCTTTAGCAT
GACCTGTTTTGTCATCGCGTTAGCGGCACTAATCAGCTCAATAATGTTATTTGCAACGTTGGTTGGGATTCACGC
AACTCTGTTTGTTATGAATATCGTGATACTGATTGTGACAATTGTACACGTTTGTTTTTGTTGT
Boschl_g00042820

ATGGGAATTACTGAGGTCGTGTTAGGATCGATCTCCTTTACTCTTGGAATCGTCATGGTGACAATTGAATCCAAT
CAAGAACTTCGTCTCGCATTATTTGCTATCGGAGTAAGTTCGCAAGGACTTTGGTGTGGGGTCATGGCACTTATC
ACAGGTATTATAGGGATTAAGGTAAAACAACACGGTACCAGATGTCTTCATATCACGAGCATGCTCTTCACCATG
ATATGTTCTTTCATCGCGTTAGCAGCTTTAATCATCTCAATGATGATATTTTCGGTTGCAGTCGGGATTCACGCA
ACGCTGTCTGTTATCAATATCGTGACACTAATTGTGACAGTAGTACACGCTTGTTTTTGCTGT
Boschl.g00029432

CTTGGAATCACTGAAATAGTACTGGGGGCGTTATCTATCCTATTCACGATACCCACACTCGCGATTGTGTCTCGA
GAACGATTTGAGTACATCAACAACGTATTTACATACGCCTCCGCTGGAATATGGGGTGGAATTTGCATGGTGGCT

ATGGGCTGGCTTGGAATACGCGTCAAAGCCAATCCGTCGAAGGGCAGCTACATTGGCAACATGATTATGGCGATT
GTGACTGCCAACATCACATTTTCTGCTATGTTCATTGCTGCGGCCGGCCTATCTAGCCTGAGAGCGTTGCATATC
GCTGTGGCGTTTCTAAACGCCGCGGCCACAATCGCCACCATCGTGCACTCGGCGTTTGGATGT
Boschl_g00001962

TTTGGCATCGCAGAGATATTGATTGGTGGATTGTCGGTTCTACTGACTATAATAACTTTGTCTATAAATAACCGT
GCAAACTATTCGTACCCCAACAATAAGTTCGTGTACGCAGCGCCGGGAATATGGGGAGGCATATTTTTTATTACT
ACCGGCATTTTGGCTGTGCGTATAAAGGCTAACCATTCCCGATATATGTACATTGCAAGTATGACCATGGCCATT
ATTACCGCATGCGTTGCATTTTCTTCCACCATCATATCCGCAGCCGCATTTTCACCTTTAATCGCCATTCATGTGA
TCATAACATTGCTACAGATAGCATCATGGATTATAGCGATCATCCACGCCGCCTTCTGCTGT
Boschl_g00058817

CTCGGCATCGGAGAGATAGTGATGGGTGTATTGTCGGTTCTAATGACCATAATTTCTCTGGTTATCAATCAGCAT
GCATCGTATTGGTCCATCTTCCCTGGTTTCACGTACTCTTCACCTGGGATATGGGGTGGTATATCTGTTATTATA
ACAGGGATGTTGGCAGTGTGCATCAAGTCTAATCCTTCCAGGTGCCTGTACATCGCAAATATGACCATGGCGATT
GTTACCGCTTGCATCACATTTTCGTCGACCATCATATCCGCAGCCGTATGGTCATCTTTAATCTTCCTACATTCTA
TAATAATGCTGCTGAATTTCGTGTCATGGATCATAACGATTATTCACGCCGCGTTCTGTTGC
Boschl_g00067862

TTTGGCATCGGGGAGATAGTGATGGGTGGATTGTCCGTTCTATTGACCATAATACCACTAGCTACCAGCTCGCAG
GCAGCGTACCGGTATACCTATACAGGTTTCACGGACTCATCACCTGGGATATGGGGAGGCATATTCATTATTGTA
ACAGGGATTCTGGCAGTGCGCATCAAATCCAATCCTTCCAGATGCATGTACATCGCTAATATGATCATGGCGATT
GTCACCGCTTGTGTCACATTTTCGTCGACCATCGTATCTGCAGCCAGTGCATCGTCGTTAATCGCCGTACATGTCC
TCATAACGCTGCTGAATTTCGTCTCATGGATCATAACAATCATTCACGCCGCGTTCTGTTGC

S3.2.2 Clade 6

18 291
Boschl_g00058817

CTCGGCATCGGAGAGATAGTGATGGGTGTATTGTCGGTTCTAATGACCATAATTTCTCTGGTTATCTATCCTGGT
TTCACGTACTCTTCACCTGGGATATGGGGTGGTATATCTGTTATTATAACAGGGATGTTGGCAGTGTGCATCAAG
TCTAATCCTTCCAGGTGCCTGTACATCGCAAATATGACCATGGCGATTGTTACCGCTTGCATCACATTTTCGTCG
ACCATCATATCCGCAGCCGTATGGCTGCTGAATTTCGTGTCATGGATCATAACGATTATTCACGCC
Boschl_g00000976

ATCGGGGTCTTGCATATAATCTTGGGAACTATATGCATCGTGATTGAATCAGGAGCAATCGTATTCTACAATCAT
TACCCGAGTATCGCGTCCGGAATATGGGCTGGATTATTTGTTTTATTGGCGGGAGTGTTTTCAATCGTATCTGGA
GCAACTGGAAAATATATATACGTGTTGATTAGCGGGGTGCTGTCAATCATCAGCTGCAATTCGTATTTCGCACAG
AAGCTTATAGTGGAACCAATCACCGCGGCGTCGTTCTTTGAGCTCATTGTGGCAGTCGTAGAACTG
Boschl_g00049784

ATGGGAATTTTGGAGATCGTGTTAGGATCGATTTCCATTATTCTTGGAATCGTCACGGTGACAATTCCACCTCCT
ATCGCGGCAAGTTCGCAAGGGCTGTGGTGCGGTATCATGGTCCTTATCATTGGCATTATGGGGATAGCATTAAAA
AAACACGCTACCAGATGTCTTCATATCGCGAACATGATCTTCAGCTTCATATGTTCTTGTATCGCCTGCGCAGCA
CTAATCAACTCCATCATGACATTTGTTATCAATTTCATGATACTGATATTGACAATTGTACACTCG
Boschl_g00057620

ATGGGAATTATGGAAGTTGTGTTAGNATTAATTTCTATTATTCTTGGAATCGTCACGGTGATAATTGACTCTGA
TGCAGAGCTAATGTCGCAAGGGCTATGGTGCGGCGTCATGATCCTTGTCGCTGGCGTTATGGGGATTACACTGAA
AAAACACGCAACCCGATGTCTTCATATCGCGAACATGATCTTTAGCATCATGTGTTCTTTCATCGCGTTAGCAGC
ACTCGTCATCTCCATTAATGTATTTGTCATCAATTTCGTGCTACTGATATTGACAATTGTACACGCG
Boschl_g00069488

GTGGGAATCATGGAGATCATCTTAGGATCGATTTCCATTATTCTTGGAATCGTCACGGTGACAATTGTTCTGTTT
GTCGGGTTAAGTTCGCAAGGACTTTGGTGTGGGGTCATGGTACTTATCACAGGCATTATAGGAATGACGTTGAA
ACAACACGCTACAAGATGTCTTCATATCGCGAGCATGCTCTTTAGCATGACCTGTTTTGTCATCGCGTTAGCGGC
ACTAATCAGCTCAATAATGTTATTTGTTATGAATATCGTGATACTGATTGTGACAATTGTACACGTT
Boschl_g00042820

ATGGGAATTACTGAGGTCGTGTTAGGATCGATCTCCTTTACTCTTGGAATCGTCATGGTGACAATTCTTTTTGCT
ATCGGAGTAAGTTCGCAAGGACTTTGGTGTGGGGTCATGGCACTTATCACAGGTATTATAGGGATTAAGGTAAA
ACAACACGGTACCAGATGTCTTCATATCACGAGCATGCTCTTCACCATGATATGTTCTTTCATCGCGTTAGCAGC
TTTAATCATCTCAATGATGATATTTGTTATCAATATCGTGACACTAATTGTGACAGTAGTACACGCT
Boschl_g00022634

CTGGGTGTAGCAGAGATCATCACTGGAGCAGTTGCACTCGCGGTTGAAGCCGCAATAGTGCGAGTGCCTTGCTCC
TACATCAACGCTGGTCAAGGAATATGGTGCGGACTGGTCGCTTTCATACTGGGTATGCTCGGAACTGCATTTTAC
GAGAAGGCGTCAAGATGTATGTATCGCGCAAATATAGCCATGGCAATATTCAGTGTCATCGTGTTCGAGGTCGGA
TTCATATTGTCCACGGCAGTATATGCGCTATGTTTTCTGGAACTCATTTTCGCGGTCATCCATTCC
Boschl.g00001820_4

TTTGGCGTGACTGAAATAGTTCTAGGGTGCATTATGACTATATTTGGGATCGCCGTTGCCGTTGTCCCATTCGTC
CTGTCTGTAACCCCCCAAGGAATATGGTGCGGCGTATTTCTAGTAGCTACTGGTGCTTATGGGACAGTTTTGAGG
AAAAAAATTACGAAAAAGATGCTGACTGTTAACTTGGTGATGACATTTATTGCCTGTGTTTTGGTGTTCGGTTG
CTTCCTTATGTCGGCAGCCGTTATTGTGGCGTGCGCTATCGTGCTTGGTATCTTAATCGTGCACGCG
Boschl.g00001820_3

TTCGGCGTAACTGAAATAGTTTTAGGCTGCATTATGGCAATATTCGGGATCGCCGTGACTGTTGTGGTGGAGATT
CTGTCAGTGACTTCCCAAGGCATATGGTGCGGCGTATTTGTGGTGGTTACTGGTGTGTTAGGAACGGTTTTGAGG
AGGAGATTTACGAGCAAAATCTTGACTAACAACTTGGCGATGACAGTTATAGCCTGTGTTTTGGTATTCGGTTGC
TTCATTATGTCTGCGGCTGTAGCAGTGGCGTGCGTTATCGCGTTTGGCATACTAATCGCCCACGCT
Boschl.g00001820_1

TTCGGCGTGACTGAAATAGTTTTAGGATGCATTTTGACTATATTCGGAATCGCGGTGGTCTTGATCGTCTTGGTC
CTGTCAGTGTCTGCACAAGGAATATGGTGCGGCGTGTTCCTGGTGGTTACTGGTGTGTTCGGAACGGTGTTGAAG
AGGAAATGTACGCGAAAGATCTTGACTATCAACTTGGCAATGACATTTATAGCCTGTGTTTTGGTGTTCGCTTGC
TTCCTCATGTCGGCAGCCGTAGCAGTGGTGTGCATTATCGCATTTGGCATCCTACTCGCCCACGCG
Boschl.g00001820_2

TTCGGCGTGACTGAAATAGTTTTAGGATGCATTTTGATAATATTCGGGATTGCAGTGGTCTTGTTTGGATCAGT
ATTGTCAGAGTCTTCACAGGGCATATGGTGCGGCGTATTTGTGATCGTTACTGGTATCTGCGGAACGGTATTGAG
GAGGAACGTTACAAGAAAGATCTTGACTATCAACCTGGCGATGACATTTATAGCCTGTGTTTTGGTGTTCGCTTG
CTTCCTTATGTCTGCGGCAGTAGCAGTGGCGTCCGTTATCGCGTTTGGCATCCTACTCGCCCACGCG
Boschl.g00060878

CTTGGAATTGCGGAGGTTGCCAACGGATCTTTGTCAATTCTGCTTGGGTTAGCTGCCATAGTATTTCTGTTATAT
ACATCTTTCGTTGCTCCGGGAGTATGGTGCGGATTACTCCTCGTAGCAACTGGTTCTCTTGGCATCAGAGTCAGG
AACAAACCATCGAGAGGAATGTACATCGCCAATATGGTGATGACTATTATAACCACCGCTTGCCTGTGCGTGGGA
TTTCTCCTTTCTCAAGGGAATATGGGTATCGACGGGTGCTCTTGGCATCAGGCCATCAAGATGCAT
Boschl_g00014523

TTCGGATTTATCGAGATTTTACTCGGAATAGCTGCATTCGGGACTTGCGTCGTGGGCGCCGTGATCTCCGAATCC
ATGATATACGCCACGTCTGGACTGTGGTGCGGGACCGTTATGATGGTGTCTGGAATATTAGGAGTAAAAGTATC
GAAACAGTCTTCGTTGTATATTATCAGAGCAAATATGACTGCCAGTATATTGTCAAGCATCTTCATGGCGATCAT
GTTTATGATGTCAGCCGCTCTGCAAATTATCGCGTCTGCCGGATTTATTTTGAATATCGTACACGCC
Boschl.g00029432

CTTGGAATCACTGAAATAGTACTGGGGGCGTTATCTATCCTATTCACGATACCCACACTCGCGATTTTTAACGTA
TTTACATACGCCTCCGCTGGAATATGGGGTGGAATTTGCATGGTGGCTATGGGCTGGCTTGGAATACGCGTCAAA
GCCAATCCGTCGAAGGGCAGCTACATTGGCAACATGATTATGGCGATTGTGACTGCCAACATCACATTTTCTGCT
ATGTTCATTGCTGCGGCCGGCCTATTTCTAAACGCCGCGGCCACAATCGCCACCATCGTGCACTCG
Boschl_g00001962

TTTGGCATCGCAGAGATATTGATTGGTGGATTGTCGGTTCTACTGACTATAATAACTTTGTCTATATATAATAA
GTTCGTGTACGCAGCGCCGGGAATATGGGGAGGCATATTTTTTATTACTACCGGCATTTTGGCTGTGCGTATAAA
GGCTAACCATTCCCGATATATGTACATTGCAAGTATGACCATGGCCATTATTACCGCATGCGTTGCATTTTCTTC
CACCATCATATCCGCAGCCGCATTTTTGCTACAGATAGCATCATGGATTATAGCGATCATCCACGCC

Boschl_g00067862

TTTGGCATCGGGGAGATAGTGATGGGTGGATTGTCCGTTCTATTGACCATAATACCACTAGCTACCTACACAGGT
TTCACGGACTCATCACCTGGGATATGGGGAGGCATATTCATTATTGTAACAGGGATTCTGGCAGTGCGCATCAAA
TCCAATCCTTCCAGATGCATGTACATCGCTAATATGATCATGGCGATTGTCACCGCTTGTGTCACATTTTCGTCG
ACCATCGTATCTGCAGCCAGTGCACTGCTGAATTTCGTCTCATGGATCATAACAATCATTCACGCC
Boschl.g00064236_1

ttgggaatcacgaattgcgtgctaggagtgatcggtattgtcgtatcgtcggtggccatgtccttgttttacaacatcgtggcggccggggagggaatctgg
tgcggaatatttctgattcttactggcattctaggaatattaacaaaagtaaagccgtccgagcctttgtacatagccagcggggtggtggggatattggcg
gcgataatcagttttgtcggatttggactgtcaggagtcctggcgacgatgtgcttgtccgctataccgatatgcattgtgcacact
Boschl.g00064236_2

cttggtgttgctgaaatagcgcttggattcgtcctactgtgtctagaagcagcggcgctgtcacacgttttagccttcggctactcaggaattggaatgtggg
gcgggttatatatttgcgcgtctggtgcgttaggagttctctacaaattcagacctaccatgactgttatcaacgcgaatttggcgatgtgcatgctgtcagc
gactacttccattctcatcgttgtgatttccgctgctatcacggcgctgggatttatagccgtgattgtcaatgtggtgcacagc

S3.3 Plasmid maps of regulatory region constructs

The electronic files are attached in the Additional Material and can be visualized with SnapGene viewer
Files names:

- GNRH1-A_LacZ
- GnRH1-B_LacZ
- GnRH1-C_LacZ
- GnRH1-D_LacZ
- GnRH1-E_LacZ
- GnRH1-F_LacZ
- GnRH1-G_LacZ
- GnRH1-H_LacZ
- GnRH1-I_LacZ
- GnRH1-J_LacZ
- GnRH1-J_GFP
- MS4A_mCherry
- Delta_like_mCherry

S3.3 Supplementary tables

Table 3.S1. Transcription factor binding sites detected in GnRH1-J with CIS-BP.

TF ID	KH ID	CIS-BP Name	Motif ID	GeneID	Family	Sequence	From	To	Direction	Score\
T207847_2.00	KH.S215.9	pbx	M06244_2.00	ENSCING00000009075	Homeodomain	CAACATAAAA	21	30	F	8.549\
T207847_2.00	KH.S215.9	pbx	M00276_2.00	ENSCING00000009075	Homeodomain	GTTAATTATGC	37	47	F	10.092\
T207847_2.00	KH.S215.9	pbx	M00276_2.00	ENSCING00000009075	Homeodomain	TTATAATTGGA	71	81	R	8.143\
T207847_2.00	KH.S215.9	pbx	M06244_2.00	ENSCING00000009075	Homeodomain	AAAATAGAAA	89	98	F	9.134\
T207875_2.00	KH.C4.84	otx2	M05389_2.00	ENSCING00000019349	Homeodomain	GGTATTATCT	51	60	R	12.799\
T207875_2.00	KH.C4.84	otx2	M03819_2.00	ENSCING00000019349	Homeodomain	AAATCTTATAATTGG	66	80	F	11.523\
T207875_2.00	KH.C4.84	otx2	M03844_2.00	ENSCING00000019349	Homeodomain	AGAATTATCGA	120	130	R	8.66\
T207848_2.00	KH.C2.957	msxb	M03105_2.00	ENSCING00000009129	Homeodomain	TTAATTATGCTTAG	38	51	F	13.948\
T207848_2.00	KH.C2.957	msxb	M08572_2.00	ENSCING00000009129	Homeodomain	TATTAT	53	58	R	8.987\
T207848_2.00	KH.C2.957	msxb	M06482_2.00	ENSCING00000009129	Homeodomain	TAATTGG	74	80	F	12.916\
T207848_2.00	KH.C2.957	msxb	M06325_2.00	ENSCING00000009129	Homeodomain	AATTAT	122	127	R	9.611\
T207841_2.00	KH.C8.482	nk4	M10743_2.00	ENSCING00000007218	Homeodomain	TAATTATG	39	46	R	11.712\
T207841_2.00	KH.C8.482	nk4	M02686_2.00	ENSCING00000007218	Homeodomain	ATAATTG	73	79	F	10.311\
T207841_2.00	KH.C8.482	nk4	M10743_2.00	ENSCING00000007218	Homeodomain	GAATTATC	121	128	R	9.712\
T207841_2.00	KH.C8.482	nk4	M08210_2.00	ENSCING00000007218	Homeodomain	AATCGATT	139	146	R	12.644\
T207841_2.00	KH.C8.482	nk4	M09146_2.00	ENSCING00000007218	Homeodomain	TATGAGTGCT	149	158	F	10.284\
T207841_2.00	KH.C8.482	nk4	M05044_2.00	ENSCING00000007218	Homeodomain	ACCGTTGAG	174	182	R	9.778\
T207879_2.00	KH.C3.553	ENSCING00000022188	M09157_2.00	ENSCING00000022188	Homeodomain	GGAAATCTTATAA	64	76	F	10.932\
T207879_2.00	KH.C3.553	ENSCING00000022188	M09157_2.00	ENSCING00000022188	Homeodomain	AGAAATGTTATAC	94	106	F	9.134\
T302687_2.00	KH.C9.580	rar	M09297_2.00	ENSCING00000006945	Nuclear receptor	TATGAACTGA	2	11	R	9.107\
T302687_2.00	KH.C9.580	rar	M00367_2.00	ENSCING00000006945	Nuclear receptor	CAACATAAAATAT	21	33	R	8.472\
T302687_2.00	KH.C9.580	rar	M00367_2.00	ENSCING00000006945	Nuclear receptor	AGGTATTATCTCA	50	62	F	8.384\
T302687_2.00	KH.C9.580	rar	M00367_2.00	ENSCING00000006945	Nuclear receptor	AAATCTTATAATT	66	78	F	9.093\
T302687_2.00	KH.C9.580	rar	M09608_2.00	ENSCING00000006945	Nuclear receptor	ATAGAAATGTTA	92	103	F	10.198\
T302681_2.00	KH.L17.15	coup	M09297_2.00	ENSCING00000002023	Nuclear receptor	TATGAACTGA	2	11	R	9.107\
T302687_2.00	KH.L17.15	coup	M00367_2.00	ENSCING00000006945	Nuclear receptor	CAACATAAAATAT	21	33	R	8.472\
T302687_2.00	KH.L17.15	coup	M00367_2.00	ENSCING00000006945	Nuclear receptor	AGGTATTATCTCA	50	62	F	8.384\
T302687_2.00	KH.L17.15	coup	M00367_2.00	ENSCING00000006945	Nuclear receptor	AAATCTTATAATT	66	78	F	9.093\
T302687_2.00	KH.L17.15	coup	M09608_2.00	ENSCING00000006945	Nuclear receptor	ATAGAAATGTTA	92	103	F	10.198\
T327695_2.00	KH.C7.523	ENSCING00000007901	M05769_2.00	ENSCING00000007901	Sox	CTAAAAATAG	86	95	R	9.314\

Table 3.S2. MS4A genes identified in selected chordates for synteny analysis.

Org_name	Symbol	description	genomic_nucleotide_accession	chromosome	start position	end_position
Homo sapiens	MS4A3	membrane spanning 4-domains A3	NC_000011.10	11	60056665	60071116
Homo sapiens	MS4A2	membrane spanning 4-domains A2	NC_000011.10	11	60088261	60098467
Homo sapiens	MS4A6A	membrane spanning 4-domains A6A	NC_000011.10	11	60171607	60184666
Homo sapiens	MS4A4E	membrane spanning 4-domains A4E	NC_000011.10	11	60201253	60243137
Homo sapiens	MS4A4A	membrane spanning 4-domains A4A	NC_000011.10	11	60280666	60308970
Homo sapiens	MS4A6E	membrane spanning 4-domains A6E	NC_000011.10	11	60327255	60341341
Homo sapiens	MS4A7	membrane spanning 4-domains A7	NC_000011.10	11	60378532	60395948
Homo sapiens	MS4A14	membrane spanning 4-domains A14	NC_000011.10	11	60396459	60417756
Homo sapiens	MS4A5	membrane spanning 4-domains A5	NC_000011.10	11	60429572	60447792
Homo sapiens	MS4A1	membrane spanning 4-domains A1	NC_000011.10	11	60455847	60470752
Homo sapiens	MS4A12	membrane spanning 4-domains A12	NC_000011.10	11	60492743	60507430
Homo sapiens	MS4A13	membrane spanning 4-domains A13	NC_000011.10	11	60515382	60543424
Homo sapiens	MS4A19P	membrane spanning 4-domains A19, pseudogene	NC_000011.10	11	60577856	60608418
Homo sapiens	MS4A8	membrane spanning 4-domains A8	NC_000011.10	11	60699612	60715807
Homo sapiens	MS4A18	membrane spanning 4-domains A18	NC_000011.10	11	60722909	60744213
Homo sapiens	MS4A15	membrane spanning 4-domains A15	NC_000011.10	11	60756867	60776733
Homo sapiens	MS4A10	membrane spanning 4-domains A10	NC_000011.10	11	60785332	60801305
Acipenser ruthenus	LOC117404752	membrane-spanning 4-domains subfamily A member 4A-like	NC_048374.1	52	45780	62457
Acipenser ruthenus	LOC117407886	membrane-spanning 4-domains subfamily A member 4A-like	NC_048374.1	52	650311	757324
Acipenser ruthenus	LOC117967640	membrane-spanning 4-domains subfamily A member 15-like	NC_048374.1	52	817684	858213
Acipenser ruthenus	LOC117407887	membrane-spanning 4-domains subfamily A member 15-like	NC_048374.1	52	884540	901811
Acipenser ruthenus	LOC117404722	membrane-spanning 4-domains subfamily A member 4A-like	NC_048374.1	52	1009751	1034304
Acipenser ruthenus	LOC117407783	membrane-spanning 4-domains subfamily A member 15-like	NC_048374.1	52	1191308	1210854
Acipenser ruthenus	LOC117404721	membrane-spanning 4-domains subfamily A member 4A-like	NC_048374.1	52	1299893	1313347
Acipenser ruthenus	LOC117967634	membrane-spanning 4-domains subfamily A member 4A-like	NC_048374.1	52	4066608	4073426
Amblyraja radiata	LOC116979486	membrane-spanning 4-domains subfamily A member 15-like	NC_045968.1	13	36040072	36057480
Petromyzon marinus	LOC116955408	uncharacterized LOC116955408	NC_046127.1	59	8076040	8110102
Branchiostoma floridae	LOC118419475	uncharacterized LOC118419475	NC_049985.1	7	708276	713314
Branchiostoma floridae	LOC118419911	uncharacterized LOC118419911	NC_049985.1	7	10878399	10884335
Branchiostoma floridae	LOC118419480	uncharacterized LOC118419480	NC_049985.1	7	17174075	17186441
Branchiostoma floridae	LOC118419477	uncharacterized LOC118419477	NC_049985.1	7	17266568	17286895
Branchiostoma floridae	LOC118409390	membrane-spanning 4-domains subfamily A member 5-like	NC_049980.1	2	8246406	8252688
Branchiostoma floridae	LOC118409393	mediator of RNA polymerase II transcription subunit 15-like	NC_049980.1	2	8256014	8260406
Branchiostoma floridae	LOC118409396	uncharacterized LOC118409396	NC_049980.1	2	8261096	8271352
Branchiostoma floridae	LOC118409394	mediator of RNA polymerase II transcription subunit 25-like	NC_049980.1	2	8266189	8322961
Branchiostoma floridae	LOC118409433	uncharacterized LOC118409433	NC_049980.1	2	18746526	18753300
Branchiostoma floridae	LOC118409978	uncharacterized LOC118409978	NC_049980.1	2	28229118	28235605
Branchiostoma floridae	LOC118406600	uncharacterized LOC118406600	NC_049997.1	19	14994292	15003943
Branchiostoma floridae	LOC118406602	uncharacterized LOC118406602	NC_049997.1	19	15004672	15008633
Branchiostoma floridae	LOC118406601	uncharacterized LOC118406601	NC_049997.1	19	15010728	15014266

Table 3.S3. Ascidian MS4a genes in Aniseed

Species	Unique Gene ID	Gene Model ID	Location
Botryllus leachi	Boleac.g00005463_1*	Boleac.CG.SB_v3.S230.g05463	S230:26057..27501
Botryllus leachi	Boleac.g00005463_2*	Boleac.CG.SB_v3.S230.g05463	S230:26057..27501
Botryllus leachi	Boleac.g00005468_1*	Boleac.CG.SB_v3.S230.g05468	S230:75980..77327
Botryllus leachi	Boleac.g00005468_2*	Boleac.CG.SB_v3.S230.g05468	S230:75980..77327
Botryllus leachi	Boleac.g00005475_1*	Boleac.CG.SB_v3.S230.g05475	S230:178716..198274
Botryllus leachi	Boleac.g00005475_2*	Boleac.CG.SB_v3.S230.g05475	S230:178716..198274
Botryllus leachi	Boleac.g00005475_3*	Boleac.CG.SB_v3.S230.g05475	S230:178716..198274
Botryllus leachi	Boleac.g00005475_4*	Boleac.CG.SB_v3.S230.g05475	S230:178716..198274
Botryllus leachi	Boleac.g00005475_5*	Boleac.CG.SB_v3.S230.g05475	S230:178716..198274
Botryllus leachi	Boleac.g00005475_6*	Boleac.CG.SB_v3.S230.g05475	S230:178716..198274
Botryllus leachi	Boleac.g00005475_7*	Boleac.CG.SB_v3.S230.g05475	S230:178716..198274
Botryllus leachi	Boleac.g00005475_8*	Boleac.CG.SB_v3.S230.g05475	S230:178716..198274
Botryllus leachi	Boleac.g00013053	Boleac.CG.SB_v3.S678.g13053	S678:66977..69150
Botryllus leachi	Boleac.g00012995	Boleac.CG.SB_v3.S670.g12995	S670:3015..3897
Botryllus leachi	Boleac.g00012484	Boleac.CG.SB_v3.S620.g12484	S620:70811..73735
Botryllus leachi	Boleac.g00005764	Boleac.CG.SB_v3.S241.g05764	S241:22666..23719
Botryllus leachi	Boleac.g00005763	Boleac.CG.SB_v3.S241.g05763	S241:11314..12085
Botryllus leachi	Boleac.g00005762	Boleac.CG.SB_v3.S241.g05762	S241:5306..6194
Botryllus leachi	Boleac.g00005474	Boleac.CG.SB_v3.S230.g05474	S230:164377..165976
Botryllus leachi	Boleac.g00005473	Boleac.CG.SB_v3.S230.g05473	S230:161165..161828
Botryllus leachi	Boleac.g00005467	Boleac.CG.SB_v3.S230.g05467	S230:68629..69421
Botryllus leachi	Boleac.g00005466	Boleac.CG.SB_v3.S230.g05466	S230:65562..66363
Botryllus leachi	Boleac.g00008864	Boleac.CG.SB_v3.S39.g08864	S39:212735..215345
Botryllus leachi	Boleac.g00007139	Boleac.CG.SB_v3.S3.g07139	S3:899066..899867
Botryllus schlosseri	Boschl.g00029432	Boschl.CG.Botznik2013.chrUn.g29432	chrUn:193160623..193166016
Botryllus schlosseri	Boschl.g00060878	Boschl.CG.Botznik2013.chrUn.g60878	chrUn:357237416..357239005
Botryllus schlosseri	Boschl.g00064236_1*	Boschl.CG.Botznik2013.chr1.g64236	chr1:2088629..2089670
Botryllus schlosseri	Boschl.g00064236_2*	Boschl.CG.Botznik2013.chr1.g64236	chr1:2088629..2089670
Botryllus schlosseri	Boschl.g00001820_1*	Boschl.CG.Botznik2013.chrUn.g01820	chrUn:11836552..11843494
Ciona robusta	Cirobu.g00000727	KH2012:KH.C1.484	KhC1:1605969..1609507
Ciona robusta	Cirobu.g00003356	KH2012:KH.C13.140	KhC13:1098737..1101581
Ciona robusta	Cirobu.g00010723	KH2012:KH.L114.12	KhL114:7379..13469
Ciona robusta	Cirobu.g00015934	NCBI:KH.108950643	KhS1551:737..5316
Ciona robusta	Cirobu.g00014126	KH2012:KH.S1907.1	KhS1907:2012..5334
Ciona robusta	Cirobu.g00014577	KH2012:KH.S489.7	KhS489:7848..12228
Ciona robusta	Cirobu.g00014575	KH2012:KH.S489.5	KhS489:3711..7783
Ciona robusta	Cirobu.g00010730	KH2012:KH.L114.8	KhL114:13968..19948
Ciona robusta	Cirobu.g00013381	KH2012:KH.L96.38	KhL96:527939..531762
Ciona robusta	Cirobu.g00013319	KH2012:KH.L95.13	KhL95:45622..51666
Ciona robusta	Cirobu.g00012729	KH2012:KH.L50.14	KhL50:46737..50041
Ciona robusta	Cirobu.g00012445	KH2012:KH.L37.79	KhL37:438453..442190
Ciona robusta	Cirobu.g00012427	KH2012:KH.L37.61	KhL37:443060..446305
Ciona robusta	Cirobu.g00012421	KH2012:KH.L37.56	KhL37:450645..458025
Ciona robusta	Cirobu.g00012378	KH2012:KH.L37.17	KhL37:446550..450645
Ciona robusta	Cirobu.g00012377	KH2012:KH.L37.16	KhL37:435518..438127
Ciona robusta	Cirobu.g00011972	KH2012:KH.L18.35	KhL18:904539..907458
Ciona robusta	Cirobu.g00011373	KH2012:KH.L15.4	KhL15:0..5422
Ciona robusta	Cirobu.g00008774	KH2012:KH.C8.165	KhC8:4381436..4384101
Ciona robusta	Cirobu.g00006128	KH2012:KH.C3.912	KhC3:5430031..5431722
Ciona robusta	Cirobu.g00006124	KH2012:KH.C3.909	KhC3:5426212..5429293
Ciona robusta	Cirobu.g00005790	KH2012:KH.C3.598	KhC3:5418744..5421149
Ciona robusta	Cirobu.g00005375	KH2012:KH.C3.212	KhC3:5423329..5425573
Ciona robusta	Cirobu.g00004442	KH2012:KH.C2.261	KhC2:6384592..6390017
Ciona robusta	Cirobu.g00004593	KH2012:KH.C2.4	KhC2:6389842..6395777
Ciona robusta	Cirobu.g00001167	KH2012:KH.C1.884	KhC1:1599237..1605242
Ciona robusta	Cirobu.g00000918	KH2012:KH.C1.659	KhC1:7381952..7385085
Ciona robusta	Cirobu.g00000884	KH2012:KH.C1.628	KhC1:1594584..1599237
Ciona robusta	Cirobu.g00000825	KH2012:KH.C1.574	KhC1:1610480..1616927
Ciona savignyi	Cisavi.g00002069	Cisavi.CG.ENS81.R1340.4963-14208	R1340:4962..14208
Ciona savignyi	Cisavi.g00003309	Cisavi.CG.ENS81.R17.1137329-1143119	R17:1137328..1143119
Ciona savignyi	Cisavi.g00001274	Cisavi.CG.ENS81.R1168.18860-25242	R1168:18859..25242
Ciona savignyi	Cisavi.g00004900	Cisavi.CG.ENS81.R238.864240-870586	R238:864239..870586
Ciona savignyi	Cisavi.g00004897	Cisavi.CG.ENS81.R238.793896-797113	R238:793895..797113
Ciona savignyi	Cisavi.g00004896	Cisavi.CG.ENS81.R238.779969-784983	R238:779968..784983
Ciona savignyi	Cisavi.g00011498	Cisavi.CG.ENS81.R96.66405-71182	R96:66404..71182
Ciona savignyi	Cisavi.g00010447	Cisavi.CG.ENS81.R72.297012-304829	R72:297011..304829
Ciona savignyi	Cisavi.g00010445	Cisavi.CG.ENS81.R72.274231-281481	R72:274230..281481
Ciona savignyi	Cisavi.g00010390	Cisavi.CG.ENS81.R72.110517-115529	R72:110516..115529
Ciona savignyi	Cisavi.g00009323	Cisavi.CG.ENS81.R56.567788-570673	R56:567787..570673
Ciona savignyi	Cisavi.g00000827	Cisavi.CG.ENS81.R1.815287-819959	R1:815286..819959
Ciona savignyi	Cisavi.g00000826	Cisavi.CG.ENS81.R1.811362-815258	R1:811361..815258
Ciona savignyi	Cisavi.g00000113	Cisavi.CG.ENS81.R0.2156633-2160313	R0:2156632..2160313
Halocynthia aurantium	Haaura.g00006236	Haaura.CG.MTP2014.S683.g06236	S683:836..1670
Halocynthia aurantium	Haaura.g00002003	Haaura.CG.MTP2014.S97.g02003	S97:98232..101063
Halocynthia aurantium	Haaura.g00001705	Haaura.CG.MTP2014.S78.g01705	S78:502..1336
Halocynthia aurantium	Haaura.g00001706	Haaura.CG.MTP2014.S78.g01706	S78:2508..3345
Halocynthia aurantium	Haaura.g00004400_1*	Haaura.CG.MTP2014.S336.g04400	S336:51330..52830
Halocynthia aurantium	Haaura.g00004400_2*	Haaura.CG.MTP2014.S336.g04400	S336:51330..52830
Halocynthia roretzi	Harore.g00005028_1*	Harore.CG.MTP2014.S231.g05028	S231:50601..52098
Halocynthia roretzi	Harore.g00005028_2*	Harore.CG.MTP2014.S231.g05028	S231:50601..52098
Halocynthia roretzi	Harore.g00012923	Harore.CG.MTP2014.S32.g12923	S32:79339..80170
Halocynthia roretzi	Harore.g00003599	Harore.CG.MTP2014.S32.g03599	S32:84219..85053
Halocynthia roretzi	Harore.g00005144	Harore.CG.MTP2014.S32.g05144	S32:86237..87075
Halocynthia roretzi	Harore.g00012525	Harore.CG.MTP2014.S31.g12525	S31:297992..299981
Halocynthia roretzi	Harore.g00015316	Harore.CG.MTP2014.S22.g15316	S22:315171..319113
Molgula occidentalis	Moocci.g00010729	Moocci.CG.ELv1_2.S375837.g10729	S375837:967..5638
Molgula occidentalis	Moocci.g00008258	Moocci.CG.ELv1_2.S319321.g08258	S319321:6183..12880
Molgula occidentalis	Moocci.g00017042	Moocci.CG.ELv1_2.S481489.g17042	S481489:277..2450
Molgula occidentalis	Moocci.g00017046	Moocci.CG.ELv1_2.S481525.g17046	S481525:5090..8015
Molgula occidentalis	Moocci.g00017878	Moocci.CG.ELv1_2.S489200.g17878	S489200:12779..26167
Molgula occidentalis	Moocci.g00024436	Moocci.CG.ELv1_2.S597736.g24436	S597736:2920..6567
Molgula occidentalis	Moocci.g00027903	Moocci.CG.ELv1_2.S634979.g27903	S634979:7802..10978
Molgula occulta	Moocul.g00012680	Moocul.CG.ELv1_2.S112991.g12680	S112991:37724..39692
Molgula occulta	Moocul.g00014061	Moocul.CG.ELv1_2.S118871.g14061	S118871:6223..7117
Molgula occulta	Moocul.g00012612	Moocul.CG.ELv1_2.S112849.g12612	S112849:2645..4054
Molgula occulta	Moocul.g00008648	Moocul.CG.ELv1_2.S99288.g08648	S99288:1990..5361
Molgula occulta	Moocul.g00008014	Moocul.CG.ELv1_2.S95750.g08014	S95750:32262..39631
Molgula occulta	Moocul.g00004119	Moocul.CG.ELv1_2.S64008.g04119	S64008:10286..14706
Molgula occulta	Moocul.g00003041	Moocul.CG.ELv1_2.S52149.g03041	S52149:15650..18180
Molgula occulta	Moocul.g00001460	Moocul.CG.ELv1_2.S30458.g01460	S30458:24550..27723
Phallusia fumigata	Phfumi.g00005311	Phfumi.CG.MTP2014.S2965.g05311	S2965:3293..5502
Phallusia fumigata	Phfumi.g00005291	Phfumi.CG.MTP2014.S2949.g05291	S2949:4426..14710
Phallusia fumigata	Phfumi.g00002448	Phfumi.CG.MTP2014.S630.g02448	S630:5137..7724

Phallusia mammalita	Phmamm.g00018199	Phmamm.CG.MTP2014.S4077.g18199	S4077:4749..8112
Phallusia mammalita	Phmamm.g00003413	Phmamm.CG.MTP2014.S110.g03413	S110:205293..208044
Phallusia mammalita	Phmamm.g00006033	Phmamm.CG.MTP2014.S244.g06033	S244:30112..33449
Phallusia mammalita	Phmamm.g00017658	Phmamm.CG.MTP2014.S3336.g17658	S3336:8644..12197
Phallusia mammalita	Phmamm.g00018375	Phmamm.CG.MTP2014.S4382.g18375	S4382:2684..6848
Phallusia mammalita	Phmamm.g00016143	Phmamm.CG.MTP2014.S2074.g16143	S2074:8029..14084
Phallusia mammalita	Phmamm.g00013947	Phmamm.CG.MTP2014.S1167.g13947	S1167:23785..26865
Phallusia mammalita	Phmamm.g00013192	Phmamm.CG.MTP2014.S1000.g13192	S1000:53668..57678
Phallusia mammalita	Phmamm.g00010988	Phmamm.CG.MTP2014.S658.g10988	S658:9338..11369
Phallusia mammalita	Phmamm.g00009351	Phmamm.CG.MTP2014.S489.g09351	S489:106339..112868
Phallusia mammalita	Phmamm.g00005433	Phmamm.CG.MTP2014.S210.g05433	S210:136665..144457
Phallusia mammalita	Phmamm.g00002791	Phmamm.CG.MTP2014.S84.g02791	S84:6810..7617

*Unique gene ID with an asterisk means that different MS4a proteins are encoded by a single long transcript, with the last number referring to their order of appearance from the 5'end.

Chapter 4 Supplementary materials

S4.1 Microfluidic chip design

The electronic file is attached in the Additional Material
File name:

- Microfluidic_Chip_Ciona.dwg

S4.2 Microfluidic chip holder

The electronic file is attached in the Additional Material
File name:

- Microfluidic_Chip_Holder.stl

S4.3 Plasmid maps for live calcium imaging

The electronic files are attached in the Additional Material and can be visualized with SnapGene viewer

File names:

- DMRT_GCAMP6m.dna
- B_Y_CRYS_GCAMP6m.dna
- PVGLUT_GCAMP6m.dna

S4.3 Videos

The electronic file is attached in the Additional Material
File name:

- Video Fig4.S1.MP4

S4.4 Supplementary tables

Table 4.S1. Measurements of 50 Ciona larvae's heads.

Number of embryos	Head width (microM)		
1	84.77	**MEAN**	88.9296327
2	94.6	**SD**	10.7148986
3	89.3	**ST ERROR**	3.38834846
4	109.3	**CI**	6.77669691
5	92.33		
6	77.64		
7	92.9		

8	84.18
9	86.62
10	80.22
11	83.88
12	71.51
13	86.03
14	79.26
15	81.68
16	104.47
17	95.086
18	85.83
19	64.68
20	78.22
21	101.86
22	91.03
23	89.83
24	94.83
25	88.47
26	88.46
27	102.6
28	96.94
29	103
30	107.91
31	83.76
32	79.84
33	88.4
34	83.06
35	71.94
36	75.22
37	77.6
38	98.59
39	102.45
40	89.68
41	115.39
42	79.266
43	87.18
44	92.55
45	85.13
46	81.49
47	105.04
48	94.56
49	78.97
50	98.23

Publications

- Poncelet, G. and S. M. Shimeld (2020). "The evolutionary origins of the vertebrate olfactory system." Open Biol **10**(12): 200330.

Comment: Publication as a first author in Open biology (Impact Factor 2020=6.41) and cited 3 times now.

- Lara-Ramirez, R., **G. Poncelet**, C. Patthey and S. M. Shimeld (2017). "The structure, splicing, synteny and expression of lamprey COE genes and the evolution of the COE gene family in chordates." Dev Genes Evol **227**(5): 319-338.

Comment: Publication as a contributing author in Development genes and evolution (Impact Factor 2017=2.125) and cited 10 times now.

OPEN BIOLOGY

royalsocietypublishing.org/journal/rsob

Review

Cite this article: Poncelet G, Shimeld SM. 2020 The evolutionary origins of the vertebrate olfactory system. *Open Biol.* **10**: 200330. https://doi.org/10.1098/rsob.200330

Received: 13 October 2020
Accepted: 26 November 2020

Subject Area:
developmental biology

Keywords:
evolutionary, vertebrates, olfactory, tunicate, lamprey

Author for correspondence:
Sebastian M. Shimeld
e-mail: sebastian.shimeld@zoo.ox.ac.uk

THE ROYAL SOCIETY PUBLISHING

The evolutionary origins of the vertebrate olfactory system

Guillaume Poncelet and Sebastian M. Shimeld

Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ, UK

SMS, 0000-0003-0195-7536

Vertebrates develop an olfactory system that detects odorants and pheromones through their interaction with specialized cell surface receptors on olfactory sensory neurons. During development, the olfactory system forms from the olfactory placodes, specialized areas of the anterior ectoderm that share cellular and molecular properties with placodes involved in the development of other cranial senses. The early-diverging chordate lineages amphioxus, tunicates, lampreys and hagfishes give insight into how this system evolved. Here, we review olfactory system development and cell types in these lineages alongside chemosensory receptor gene evolution, integrating these data into a description of how the vertebrate olfactory system evolved. Some olfactory system cell types predate the vertebrates, as do some of the mechanisms specifying placodes, and it is likely these two were already connected in the common ancestor of vertebrates and tunicates. In stem vertebrates, this evolved into an organ system integrating additional tissues and morphogenetic processes defining distinct olfactory and adenohypophyseal components, followed by splitting of the ancestral placode to produce the characteristic paired olfactory organs of most modern vertebrates.

1. Introduction: olfaction and chemosensation

Olfaction is a form of chemosensation. It is colloquially equated to the sense of smell, the specific sensing of chemicals in the air via the nose and the relaying of this information to the brain via olfactory nerves. However, the precise evolutionary and developmental delineation of the olfactory system becomes blurry when one considers the details. Many vertebrates have a related chemosensory system in the vomeronasal organ, which shares a developmental origin with the main olfactory system but has generally been thought to be devoted to chemical communication between conspecifics. In aquatic vertebrates, such as fish and amphibians, a homologous olfactory system to that of terrestrial vertebrates detects waterborne rather than airborne chemicals, while insects possess a well-described system in their antennae that senses airborne chemicals and is usually called an olfactory system, but is convergently evolved at the system level. Furthermore, the development of the vertebrate olfactory system includes the formation of cells associated with other functions, including that of the pituitary, and there are many vertebrate chemosensory cells that relay information to the brain but that are not part of olfactory systems in the conventional sense. Taste is an obvious example.

Sensing chemicals on the outer side of the cell membrane is a fundamental feature of all cells and sensing environmental chemicals has obvious adaptive advantages. It is therefore not surprising that a diversity of chemosensory mechanisms and systems have evolved in animals. We will not attempt to cover this diversity here, but will focus specifically on the evolution of the vertebrate olfactory system. We will combine two levels of comparison: first, the types of neural cells that develop in the olfactory system (both chemosensory and neurosecretory cell types). Second, the mechanisms that control

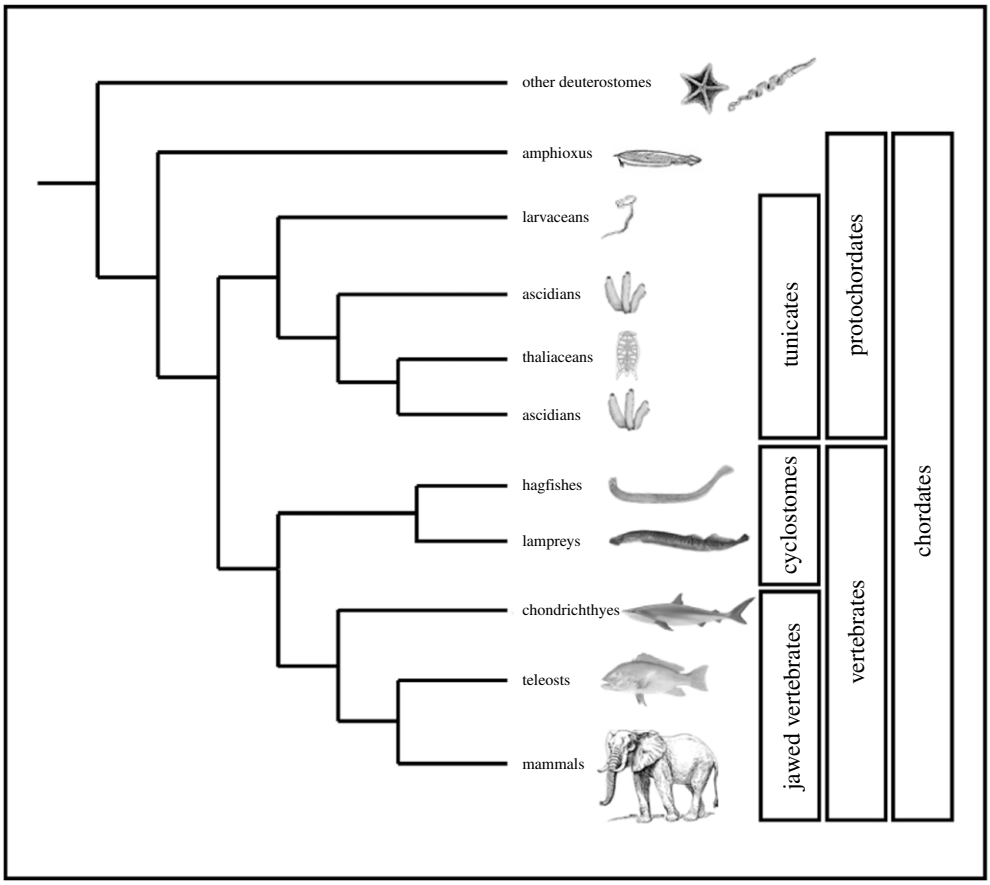


Figure 1. A phylogeny of the chordates showing the relationships of the major lineages discussed in this review. Note that the ascidians appear twice in the tree as they are paraphyletic.

the specification of olfactory cells and organs. Since vertebrate olfactory cells develop from an ectodermal placode that shares a developmental and evolutionary history with other such placodes, we will also consider placode development and evolution more broadly.

We first summarize what is known about this in jawed vertebrates (also known as gnathostomes). We then compare this to olfactory systems and related cells and structures in other chordates (figure 1): the jawless vertebrates (represented by lampreys and hagfishes, collectively the cyclostomes), the tunicates (including sea squirts and their allies) and the cephalochordates (represented by amphioxus). We will finish with a model for how the olfactory system in living vertebrates evolved.

2. The olfactory systems of jawed vertebrates and their neural cell-type derivatives

There are two major olfactory subsystems in jawed vertebrates: the main olfactory system (MOS) and the accessory olfactory system (AOS) (figure 2), and relevant neuronal cell types are summarized in table 1. The MOS is historically said to be important for the detection of odorants and the AOS to mainly sense pheromones. However, the systems may overlap functionally and act synergistically [1]. When a chemical stimulus flows into the MOS or AOS, it is detected by olfactory sensory neurons (OSNs) through specific membrane chemoreceptors. Nearly all these chemoreceptors are coupled to a specific G protein subunit α, encoded by genes of the GNAL and GNAS families [2]. The different Gα proteins mediate signal transduction pathways that open cyclic nucleotide-gated (CNG) ion channels in the MOS or transient receptor potential (TRP) channels in the AOS. These channels trigger a calcium influx in the olfactory neuron cytosol, promoting the opening of calcium-gated chloride channels. The combined effect of calcium and chloride efflux triggers OSN depolarization [3]. OSNs expressing the same chemoreceptor gene send their axon projections via the olfactory nerve to a specific glomerulus in the olfactory bulb. The OSNs of the MOS transmit the chemosensory signal through the main olfactory bulb, which then connects to higher brain centres for the processing of a behavioural response. The OSNs of the AOS target their axons to the accessory olfactory bulb in the rostral telencephalon, which then projects towards the amygdala and hypothalamus, which are involved in aggression and mating behaviours [4].

The MOS includes the main olfactory epithelium, which is composed of typical ciliated OSNs. The cilia of the OSNs harbour seven-transmembrane domain G-protein-coupled receptors from the olfactory receptor family (OR), or receptors of the trace amine-associated receptor family (TAAR), on their surface [5]. The ORs are the largest gene family in vertebrates and there may be more than a thousand different genes in the genomes of some species [6]. The TAARs and ORs are coupled to G protein subunit Gαolf. In the recesses of the mammalian main olfactory epithelium, there is an expression of chemoreceptors from the guanylate cyclase D receptor family, and the MS4A gene family, in the cilia of a specific group of OSNs known as the necklace OSNs. These latter receptor families are not coupled to specific G proteins [7].

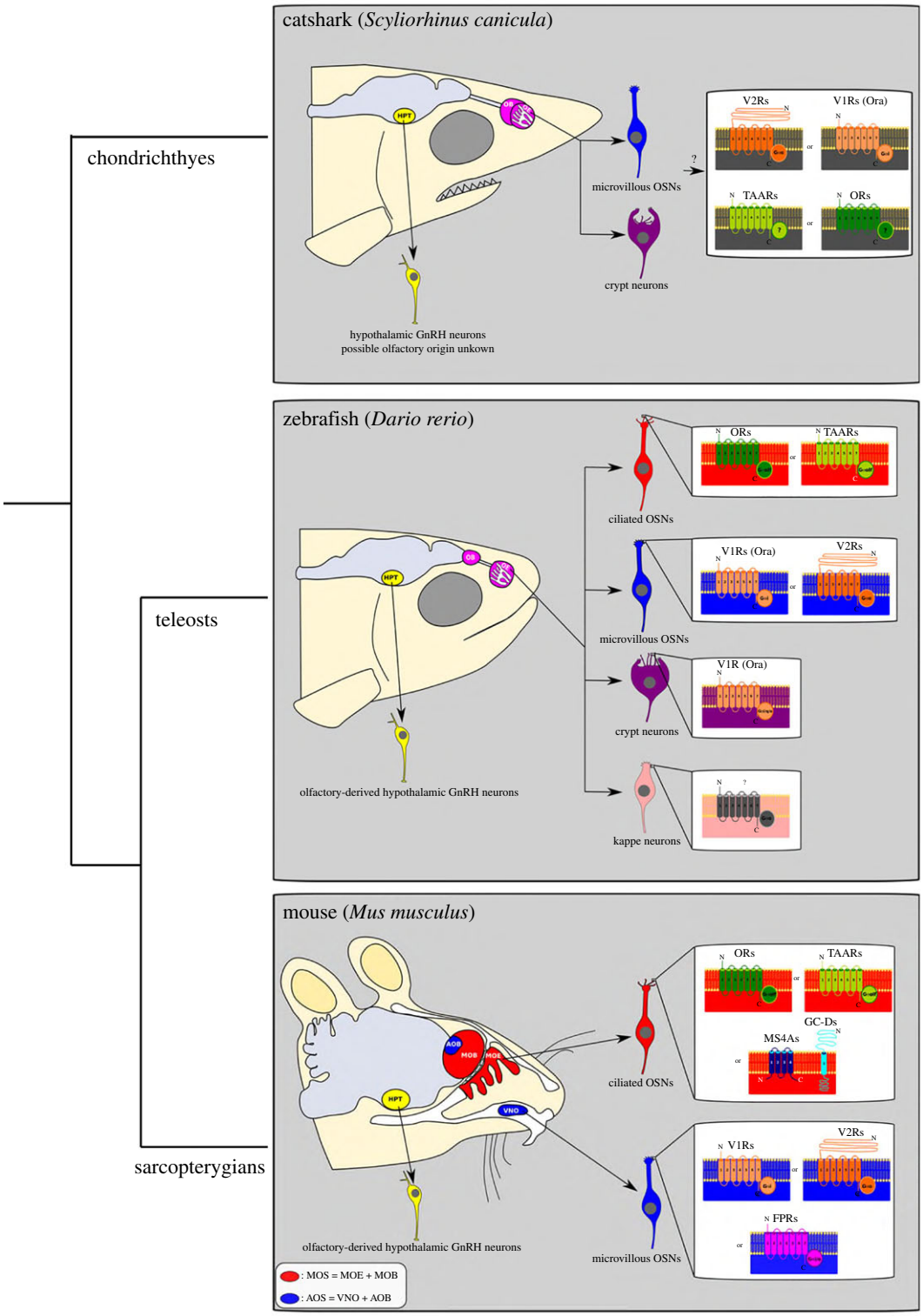


Figure 2. Schematic diagrams of adult organization of the main and accessory olfactory systems in shark, teleost and mammal lineages, including some of the neural cell types discussed in the text and shown in table 1. Note that olfactory-derived GnRH neurons of the terminal nerve and NPY neurons are not represented here. AOB, accessory olfactory bulb; GC-D, guanylate cyclase D; FPR, formyl peptide receptor; GnRH, gonadotropin-releasing hormone; HPT, hypothalamus; MOB, main olfactory bulb; MOE, main olfactory epithelium; OR, olfactory receptor; TAAR, trace amine-associated receptor family; V1R, vomeronasal type 1; V2R, vomeronasal type 2; VNO: vomeronasal organ.

The AOS is also known as the vomeronasal system in tetrapods and includes a distinct sensory epithelium containing OSNs bearing microvilli instead of cilia. Like ciliated MOS OSNs, these microvillous OSNs express seven-transmembrane domains G-protein-coupled receptors, but in the AOS, these are chemoreceptors of the vomeronasal type 1 (V1R) or type 2 (V2R) families. V1Rs and V2Rs are associated with the G protein subunits Gαi and Gαo, respectively. In rodents, some microvillous sensory neurons of the vomeronasal organ also express members of the formyl peptide receptor family associated with the identification of pathogens and infections [8]. The description of an accessory olfactory system in lungfish suggests that all sarcopterygians primitively had such a dual system [9]; however, the vomeronasal system is not preserved in all tetrapod groups and is absent or vestigial in some lineages like archosaurs (birds and crocodilians) and higher primates [1]. While molecular studies have yet to extend across the diversity of tetrapods, there is a long history of histological and ultrastructural studies of tetrapod olfactory and vomeronasal systems covering many different species. There

Table 1. Summary of neural cell types developing from the olfactory placode and its derivatives. Abbreviations: MOS, main olfactory system; AOS, accessory olfactory system; FPR, formyl peptide receptor; GC-D, guanylate cyclase D; GnRH, gonadotropin-releasing hormone; OSN, olfactory sensory neuron; OR, olfactory receptor; TAAR, trace amine-associated receptor; MS4A, membrane-spanning 4A receptor; V1R, vomeronasal type 1 receptor; V2R, vomeronasal type 2 receptor.

taxonomic group	location/ embryo origin	sensory cell types and receptors	G protein used	additional comments
sarcopterygians	MOS	ciliated OSNs expressing ORs	Gαolf	
		ciliated OSNs expressing TAARs	Gαolf	
		ciliated OSNs expressing GC-D family receptors and MS4As family receptors	not G protein coupled	demonstrated only in rodents
	AOS	microvillous OSNs expressing V1Rs	Gαi	AOS lost in some tetrapod lineages
		microvillous OSNs expressing V2Rs	Gαo	
		microvillous OSNs expressing FPRs family	Gαi/o	demonstrated only in rodents
teleosts	single epithelium	ciliated OSNs expressing ORs	Gαolf	
		ciliated OSNs expressing TAARs	Gαolf	
		microvillous OSNs expressing V1Rs	Gαi	
		microvillous OSNs expressing V2Rs	Gαq	
		crypt cells with cilia and microvilli expressing V1Rs	Gαi/o/q	
		cap cells, receptor unknown	Gαo	
chondrichthyes	AOS only	microvillous OSNs likely expressing V2Rs	Gαo	dominance of V2Rs, with few TAARs, V1Rs and ORs genes. It is not precisely known which receptor is expressed by each sensory cell type
		crypt cells with cilia and microvilli likely expressing V1Rs	unknown	
possibly all gnathostomes	olfactory placode	neuropeptide Y neurons	N/A	migrate from placode to hypothalamus, but this has only been demonstrated in chicken
		GnRH1 neurons	N/A	migrate from olfactory placode to hypothalamus and pre-optic area. It has been demonstrated in osteichthyes but not to date in chondrichthyes. In some species, they also migrate to form the terminal nerve if GnRH3 is absent (functional compensation of paralogue)
		GnRH3 neurons	N/A	migrate from olfactory placode to form the terminal nerve. In some species, they also migrate to the hypothalamus and pre-optic area if GnRH1 is absent (functional compensation of paralogue)

is insufficient space to detail these studies here and they have been well-reviewed by Eisthen [10]. It is important to note, however, that the tetrapod group may harbour system diversity beyond the simple division of ciliated OSNs in the MOS and microvillous OSNs in the AOS familiar from mammals. For example, both types of OSN are found in the MOS of some urodele amphibians and ciliated OSNs in some lizard and bird species may have their cilia surrounded by microvilli [10].

In teleost fishes, there is no proper vomeronasal organ but a single olfactory epithelium showing morphological features of both the main and accessory systems and with intermingled ciliated and microvillous OSNs [11]. Teleost ciliated OSNs express the ORs and TAARs associated with Gαolf, similar to the main olfactory epithelium of tetrapods. Similarly, the teleost microvillous OSNs express V1Rs and V2Rs associated with Gαi and Gαo, as seen in the vomeronasal organ of mammals. It was also demonstrated in zebrafish that microvillous OSNs form a neural circuit via the dorsomedial olfactory bulb and intermediate ventral telencephalic nucleus (the putative teleost medial amygdala) to the tuberal hypothalamus, similar to the vomeronasal circuit of tetrapods [12]. Teleosts have a third class of OSNs, the crypt cells, which possess microvilli and cilia and express V1R-type (ORA) receptors [2,13]. In zebrafish, there is

also a fourth type of OSN, the cap (kappe) cell, whose receptor type is unidentified but known to associate with G*α*o [14]. In Chondrichthyes, it was observed that the sense of smell relies primarily on microvillous OSNs coupled to G*α*o [15]. The chemosensory receptor repertoire of cartilaginous fishes is dominated by the expanded V2R family, although there are also a few OR, TAAR and V1R genes [16]. As in teleosts, there is also the presence of crypt neuron-like cells but their exact receptor is unknown [17]. This peculiar feature and the absence of ciliated OSNs suggest that the cartilaginous fish olfactory system is just an accessory system [18].

Some neurosecretory cells also delaminate from the olfactory placode of gnathostomes. The most well-known are the gonadotropin-releasing hormone (GnRH) neurons involved in the reproductive axis [19,20]. There are three distinct populations of GnRH neurons in the brain of gnathostomes expressing one of the three existing *GnRH* genes: *GnRH1*, *GnRH2* and *GnRH3*. Some species have lost some of these paralogues, but most vertebrates express at least two. Neurons expressing either *GnRH1* or *GnRH2* have been identified in most gnathostomes, while *GnRH3*-expressing neurons are fish-specific [21,22]. The prevailing view is that the GnRH1 neurons and GnRH3 neurons originate from the olfactory placodes, while the GnRH2 neurons are of non-placodal origin and develop within the central nervous system, mainly the midbrain [23,24]. The GnRH1 neurons are key regulators of fertility as an essential part of the hypothalamic–pituitary–gonadal axis (HPG). These cells migrate along axons of the terminal nerve/olfactory pathways up to the forebrain where they settle inside the pre-optic and hypothalamic areas [25]. Once settled within the hypothalamus, GnRH1 neurons send their axons to the median eminence and secrete GnRH1 into the portal vessels, where it travels to the adenohypophysis [26]. Here, GnRH1 activates specific receptors of the gonadotrope cells, which then release two crucial hormones for sexual maturation and reproduction, luteinizing hormone and follicle-stimulating hormone [27]. In chick, it was shown that the neuropeptide Y (NPY) neurons also derive from the olfactory placodes and migrate to the hypothalamus along with GnRH1 neurons. NPY neurons control the secretion of GnRH1 by acting directly on GnRH1 neurons [28].

GnRH3 neurons migrate from the olfactory placodes and become components of the terminal nerve whose processes extend anteriorly to the nasal cavity and posteriorly to various brain regions, mediating chemosensory processing and reproduction [29]. The GnRH3 neurons of the terminal nerve (TN-GnRH3) have neuromodulatory effects on the OSNs [30,31]. In addition, the TN-GnRH3 neurons of zebrafish have been demonstrated to be chemosensory, detecting CO_2 in order to avoid incoming predators [32]. This latter finding suggests that olfactory-derived neurons with dual GnRH/chemosensory abilities could exist in vertebrates (as is advocated for the aATENs of ascidians, see §9 below for details) [33]. In species that lack either *GnRH1* or *GnRH3*, the remaining gene compensates functionally for the lost paralogue by being expressed by the relevant cells. For example in zebrafish, where *GnRH1* is lost, the *GnRH3* gene is expressed in the pre-optic area and hypothalamus (the *GnRH1* territory) in addition of the terminal nerve (the *GnRH3* territory).

It needs to be made clear that the olfactory epithelia of the MOS and AOS also contain non-neural cells that surround the OSNs, such as supporting cells and basal cells. The supporting cells provide physical and metabolic support to the olfactory epithelium. The basal cells are stem cells used to constantly replenish the olfactory epithelium as they can differentiate into either OSNs or supporting cells. The MOS also contains the mucus-producing olfactory (Bowman's) glands whose proteinaceous secretion allows solubilization of odorants in the nasal cavity [25,34]. These cells are important for vertebrate olfactory system function but will not be the focus of this review as they are not easily compared between vertebrates and other chordates.

3. Development of the jawed vertebrate olfactory system from the olfactory placodes

The vertebrate cranial placodes are transient ectodermal thickenings of embryonic head and contribute to the developing cranial sensory systems. In jawed vertebrates, the olfactory system develops from a pair of cranial placodes, the olfactory placodes. Other cranial placodes such as the lens, vestibulo-acoustic, trigeminal, epibranchial and lateral line placodes contribute to other cranial senses, including sight, hearing, balance, somatosensation, gustation and internal physiological monitoring: their respective cell types and functions have been extensively reviewed elsewhere and will not be further considered here [35–38]. However, one other placode develops in intimate association with the olfactory placodes and warrants further mention. The adenohypophyseal placode forms between the paired olfactory placodes, in front of the extreme anterior of the neural plate. During subsequent development, it becomes internalized through the mouth [39], eventually forming the adenohypophysis and thus having a direct functional connection with olfactory placode-derived GnRH neurons in the hypothalamus.

All the cranial placodes arise from the pre-placodal ectoderm (PPE), a U-shaped cell field around the edge of the anterior neural plate (figure 3). The PPE is specified by fibroblast growth factor (FGF) signalling and bone morphogenetic protein (BMP)- and Wingless-related integration site (Wnt)-antagonism [40]. It is characterized by the expression of pre-placodal competence factors such as the transcription factors (TFs) *Six1/2*, *Six4/5*, *Eya1-4*, *Dlx3/5/6*, *Gata3* and *Foxi1* [41]. During the development, the PPE subdivides in specific regions along the anteroposterior axis to give rise to individual cranial placodes. In particular, the lens, adenohypophysis and olfactory placodes are defined anteriorly through the combinatorial expression of TFs such as *Dmrt4*, *Otx2/5*, *Pax6*, *FoxE*, *Six3/6* and *Pitx1/2* [42]. BMP signalling promotes specification of an olfactory fate, while extended BMP exposure time promotes lens fate [43]. FGF signalling from the anterior neural plate is known to block expression of the lens marker gene *Pax6* and to promote the expression of *Dmrt4*, an olfactory placodal gene [44,45]. Olfactory placodes are further characterized by the expression of the transcription factor genes *Emx2* and *COE2* (*Ebf2*), among others [39].

In addition to the ectodermal cells from the PPE, the olfactory placodes become associated with migratory neural crest cells [25,35]. These give rise to the olfactory ensheathing cells, which are glial cells that envelop the bundles of olfactory axons [34,46,47]. The possible contribution of neural crest to other cell types in the olfactory placode such as the GnRH neurons is debated and controversial, and there is no coherent vision on the lineage origin of the major neural cell types

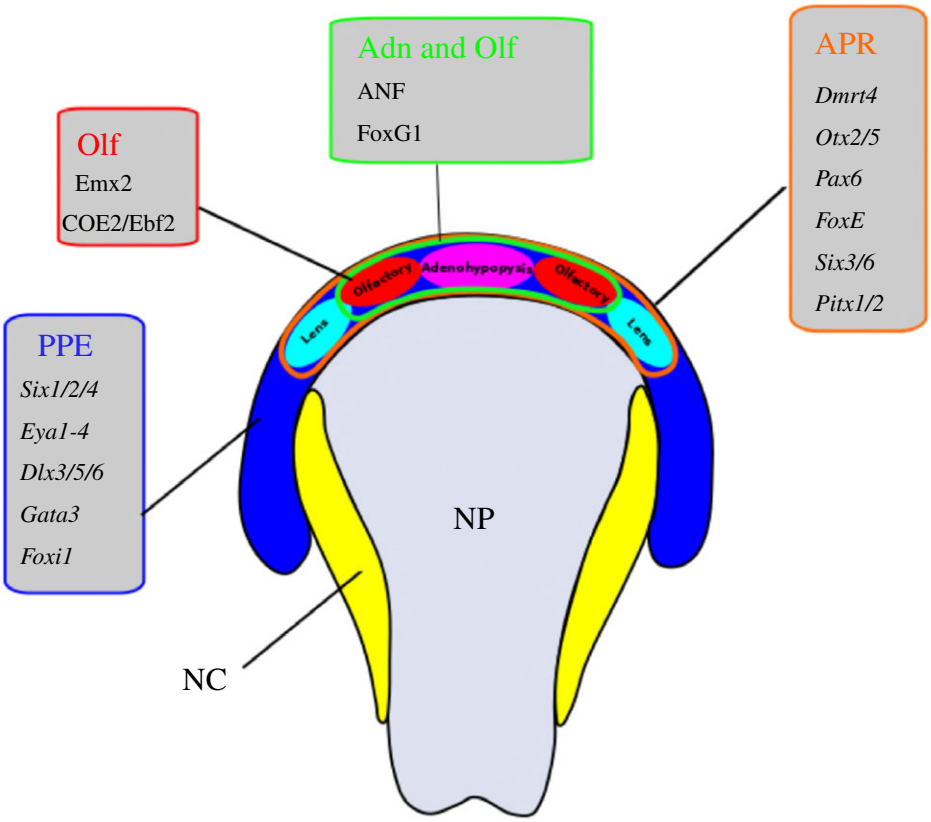


Figure 3. Anterior ectodermal patterning and origin of individual anterior placodes from the pre-placodal ectoderm (PPE) in jawed vertebrates. The PPE is specified by FGF signalling, BMP antagonism and Wnt-antagonism, all coming from the underlying mesoderm. These induce transcription factors within the PPE (blue), which specify precursor regions for multiplacodal areas (coloured outlines) and individual placodes (coloured ovals). Known transcription factors for each are shown (for a more detailed discussion of these genes and genes marking other placodes see [38]). Abbreviations: Adn, adenohypophyseal placode; APR, anterior placodal region; NC, neural crest; NP, neural plate; Olf, olfactory placode.

associated with the olfactory sensory epithelia of vertebrates [20,46,48]. Precise lineage cell-tracing data in zebrafish argue against a contribution from the neural crest and support the view that all the different neuronal populations within the olfactory epithelium originate from overlapping pools of progenitors in the PPE [19]. However, a recent analysis of GnRH1 neurons in the mouse olfactory placode argued for a heterogeneous origin, with neural crest and PPE-derived GnRH1 neurons [49]. It is possible these apparently conflicting reports reflect genuine differences between species.

4. The olfactory system in jawless vertebrates

There are only two surviving lineages of jawless vertebrates, the lampreys and the hagfishes (figure 1), though fossils show a much wider diversity of extinct lineages [50]. There are many similarities in head development between jawed and jawless vertebrates, including sensory systems and the central nervous system. For example, most jawed vertebrate placodes are clearly identifiable in lampreys [51,52]; cranial nerve organization is similar [51] and gross brain organization well-conserved [53,54]. There is, however, an important difference in the olfactory system. Jawed vertebrates have paired nostrils leading to paired olfactory sacs and derived from the paired olfactory placodes. Lampreys and hagfish have a single median nostril, a condition known as monorhiny. This develops to form a single anterior medial placode known as the nasohypophyseal placode, which combines characters of both olfactory and adenohypophyseal placodes and produces a single, medial nasal (olfactory) sac.

Despite this difference, both lampreys and hagfishes have well-developed olfaction. Adult sea lamprey uses odours from conspecific larvae (including dihydroxylated tetrahydrofuran fatty acids and some bile acids) to select the best streams for spawning based on their larval population [55–59]. Hagfishes are usually found in deep water and their olfactory organ seems particularly efficient as they have been shown to be among the first fish species to locate chemical signals of decaying prey [60,61]. Electrophysiological recordings indicate that their olfactory sensory neurons are particularly sensitive to amino acids [62]. In addition to the conventional olfactory system, hagfishes have specialized chemosensory structures named 'Schreiner organs' all over the body epidermis [63]. Their ecological role is not known, though we can speculate they may help with directional chemosensation in the absence of paired olfactory membranes.

5. The olfactory system of lamprey and its neural cell-type derivatives

In larval and adult lamprey, the single olfactory organ is composed of three elements: the nasal duct, the nasal sac and the nasopharyngeal pouch (figure 4) [67]. The nasal tube opens externally as a single nostril on the dorsal head surface. In adult lampreys, the nasal tube contains a valve that serves to introduce and expel water into the entrance of the nasal

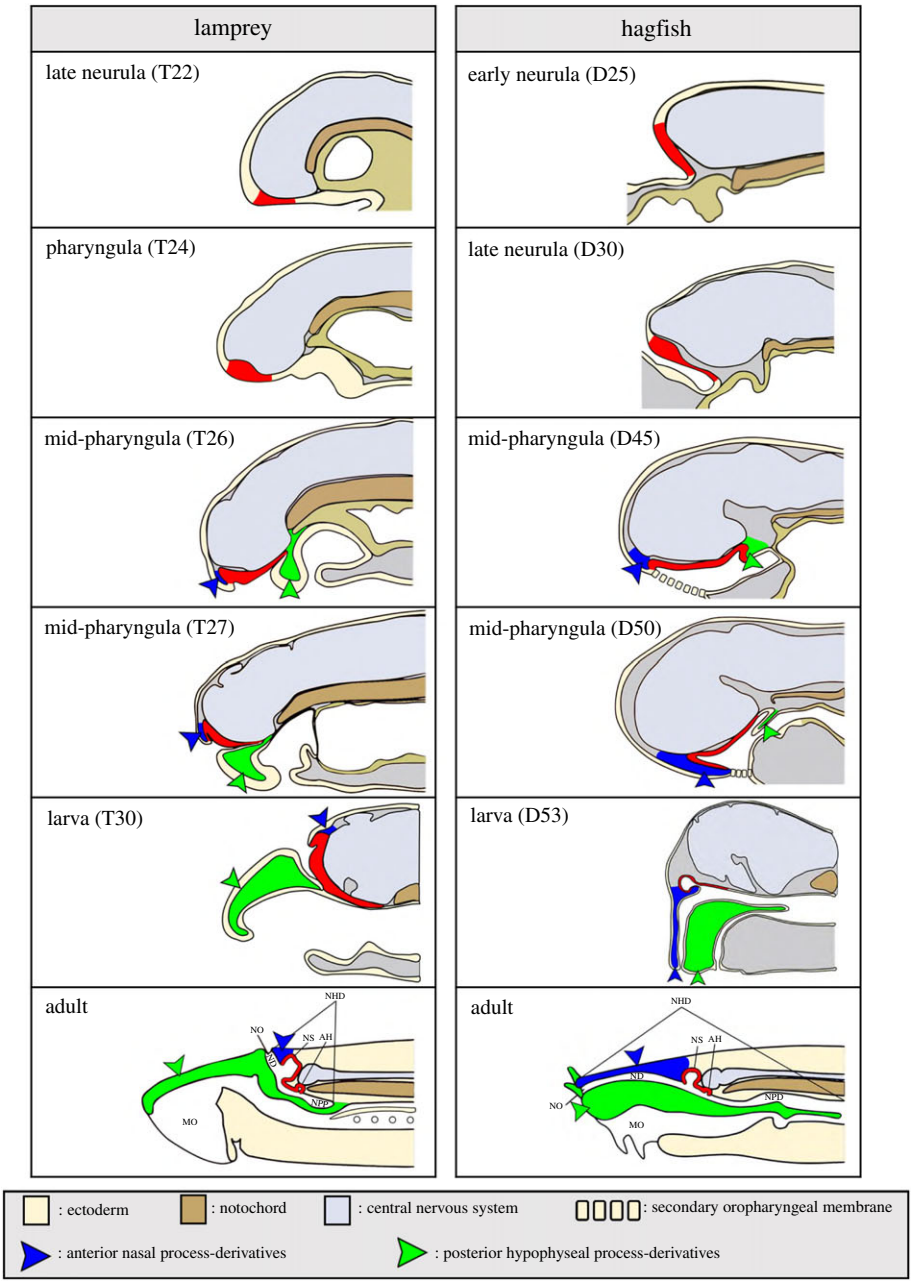


Figure 4. Comparative development of nasohypophyseal placode (NHP) in lampreys and hagfishes. The NHP is labelled in red. At the mid-pharyngula stage, when development is the most similar in cyclostomes, the NHP is rostrocaudally bordered by ventral growth of two ectomesenchymal processes, the anterior nasal process (blue) and the posterior hypophyseal process (green). Abbreviations: AH, adenohypophysis; MO, mouth; ND, nasal duct; NHD, nasohypophyseal duct; NO, nostril; NPD, nasopharyngeal duct; NPP, nasopharyngeal pouch; NS, nasal sac; Figure adapted from [64] with permission. T and D denote lamprey and hagfish embryo staging, respectively [65,66].

sac, the chemosensory part of the organ [68]. When larval lamprey metamorphos into adults, the olfactory organ extends and changes from an epithelial lined tube to a nasal sac with lamellar folds [69]. In the sea lamprey *Petromyzon marinus*, the nasal sac wall is divided into 25 folds [68]. Each fold is lined with the main olfactory epithelium, which is mainly composed of tall, narrow, ciliated OSNs [70,71]. However, OSNs in the main olfactory epithelium display at least three distinct morphotypes, with some not necessarily ciliated but with microvillar-like protrusions. These different OSNs have been proposed to be similar to the ciliated OSNs, microvillous OSNs and crypt cells found in teleost fishes [72]. The OSNs in the main olfactory epithelium express the three chemoreceptor gene families identified in the sea lamprey genome, the ORs, TAARs and V1Rs; the V2R gene family seems to be gnathostome-specific as it is apparently absent in lamprey [73–76]. The OSNs of the main olfactory epithelium send projections mainly to the non-medial region of the olfactory bulbs, but also send some projections to its medial region [77]. In the caudoventral portion of the peripheral olfactory organ, there is an accessory olfactory organ [78], which is covered with the accessory olfactory epithelium containing short, rounded, ciliated neurons [79,80]. The accessory olfactory epithelium sends projections exclusively to the medial olfactory bulb, which connects to the posterior tuberculum creating a motor response from olfactory inputs [80,81].

It has been hypothesized based on anatomical and molecular evidence that the lamprey accessory olfactory epithelium, coupled with the dorsomedial telencephalic neuropil, is the putative homologue of the tetrapod vomeronasal system [82,83]. However, it remains questionable

whether the so-called 'main' and 'accessory' olfactory epithelia of lamprey are indeed homologous to the main and accessory olfactory epithelia of sarcopterygians. Characters that point to homology are that the lamprey main and accessory olfactory epithelia have differences in their respective pathways and that distinct G protein subtypes are used for signal transduction, with Gαolf located only in the OSNs projecting to the non-medial olfactory bulb [84]. Hence, different types of G proteins are used from those of OSNs projecting to the medial glomeruli, a similarity shared with the vomeronasal organ. However, there is a notable difference in that ORs, TAARs and V1Rs are not differently expressed between the lamprey olfactory epithelia as opposed to the tetrapod vomeronasal organ and main olfactory epithelium [80,82]. This difference could represent an intermediate and ancestral condition before the exclusive shift to vomeronasal receptor genes, as seen in the AOS of gnathostomes.

In lampreys, current evidence suggest that GnRH neurons of the pre-optic area and hypothalamus are not derived from the nasohypophyseal (NHP) placode, contrary to what is observed in jawed vertebrates [85,86]. Immunohistochemical investigation concluded that lamprey GnRH neurons were never seen in association with the NHP placode during embryonic development, and it was hence hypothesized that GnRH neurons originate within proliferative zones of the diencephalon in developing lamprey, not in the olfactory system [86]. However, other data supporting this difference to jawed vertebrates are lacking.

6. The olfactory system of hagfish and its neural cell-type derivatives

As in lampreys, the olfactory system of adult hagfish is composed of three main parts: a nasal duct, a nasal sac and a nasopharyngeal duct (figure 4) [87,88]. Hagfishes also possess a single nostril just above the mouth, surrounded by two nasal tentacles on each side and by a dorso-median lip. The nasal duct leads to the nasal sac anterior to the brain [88,89]. A valve is present in an oblique position inside the nasal duct and serves to manage water flow in the duct towards the nasal sac [90]. The latter receives a continuous flow of water during respiration as the water flows from the nostril and is ejected through the gill openings as, unlike in lamprey, the nasopharyngeal duct does not end blindly but opens into the pharynx [88,89]. The nasal sac is divided into seven olfactory laminae and the olfactory epithelium is composed of two types of OSNs, ciliated or microvillous. In adult hagfishes, GnRH neurons have been identified in the diencephalon [91–93]. However, the embryonic development of these cells has not been investigated and nothing is known about the potential association or shared origin of hagfish GnRH neurons with the olfactory system.

7. Olfactory system development in jawless vertebrates

The developmental trajectories of lamprey and hagfish systems are shown in figure 4. Early in development the single median nasohypophyseal placode is characterized by orthologous molecular markers to those seen in the gnathostome olfactory placode. The entire nasohypophyseal placode territory expresses *Six3/6A* and *Soxb1* in hagfish [64] and *Pax6* in lamprey [94,95], consistent with the expression of the gnathostome orthologues *Pax6*, *Six3/6* and *Sox2/3* in both the olfactory and adenohypophyseal placodes [37]. At the late neurula stage of lamprey and hagfish embryos, the anterior part of the nasohypophyseal placode becomes the likely olfactory territory as it expresses *FGF8/17/18*, the orthologue of *FGF8* in gnathostomes. Reciprocally, the posterior part becomes the adenohypophyseal territory as it shows *PitxA* expression [64,96]. At the pharyngula stage in lamprey, the nasohypophyseal placode forms a thickened area of the ventral ectoderm, anterior to the mouth cavity. Morphogenesis of this region is coordinated with that of two ectomesenchymal processes (figure 4) and has been described elsewhere [64]. Important points to note are that the nasohypophyseal placode extends an epithelial cell process posteriorly to establish close contact with the definitive hypothalamic region, thus forming a pituitary similar to that of jawed vertebrates in combining central nervous system and placode-derived parts. The anterior part of the nasohypophyseal placode, which will form the future olfactory epithelium, is characterized by the expression of olfactory developmental gene markers such as *OtxA*, *CoeA*, *CoeB*, *EmxA* and *EmxB* [95–99].

As lamprey embryos approach the larval stage, the anterior part of the nasohypophyseal placode differentiates as the nasal sac, composed of a thick columnar epithelium and covering the rostral aspect of the telencephalon, whereas the posterior (as the future adenohypophysis and consisting of an epithelium of a few cell layers) extends caudally to the level of the optic chiasma [96]. In late hagfish embryos, the nasohypophyseal duct and oral cavity grow posteriorly relative to the position of the adenohypophysis. The presumptive nasohypophyseal duct is tilted inward in an oblique position unlike that of lamprey, which is situated more vertically. In addition, the posterior end of the nasohypophyseal duct in hagfishes ruptures into the pharynx (figure 4). Thus, the nasohypophyseal duct in hagfishes opens secondarily into the pharynx [64].

To summarize, there are many similarities in development, gene expression and cell-type derivatives between the olfactory systems of jawless and jawed vertebrates. There are also some important differences. Most notably, jawless vertebrate systems develop from a single medial placode combining olfactory and adenohypophyseal progenitors that separate during morphogenesis, and form a single medial olfactory system and not the paired systems of jawed vertebrates. Fossil data suggest monorhiny is the ancestral condition [100–102], so a single medial olfactory/adenohypophyseal placode as seen in living jawless fish is probably also ancestral (though note the description of paired nasal sacs in some vertebrate stem lineage fossils, and that despite monorhiny, lampreys and hagfishes have a pair of olfactory nerves like gnathostomes, meaning that some questions remain over this inference [103–105]). It is also unclear whether GnRH neurons are olfactory placode derivatives in jawless fish like GnRH1 and GnRH3 neurons in gnathostomes, or are born within the central nervous system like gnathostome GnRH2 neurons.

8. Potential olfactory system homologues in protochordates

The tunicates and cephalochordates are collectively known by the paraphyletic term 'protochordates' (figure 1). They

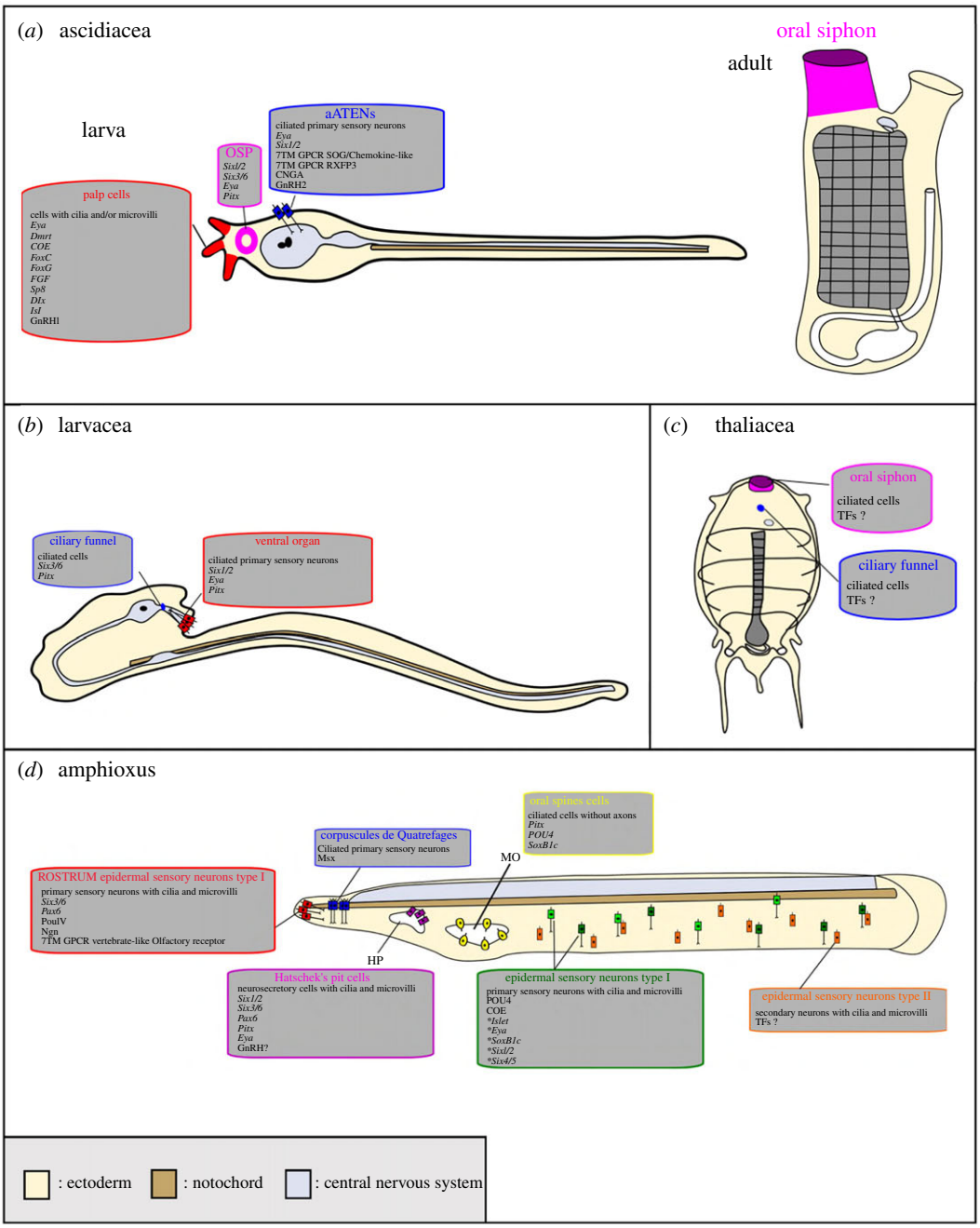


Figure 5. Possible olfactory sensory cells in protochordates and the genes they express. (*a*) Schematic drawing of an ascidian larva and adult of *Ciona intestinalis*. The ascidian swimming larva has three regions suspected to hold or give rise to putative chemosensory cells: the palp cells (red), the oral siphon primordium (OSP, pink) which becomes the oral siphon in adults, and the anterior trunk epidermal neurons (aATENs, blue). (*b*) Schematic drawing of the adult larvacean, *Oikopleura dioica*. Two main regions are believed to host chemosensory cells: the ciliary funnel (blue) and the ventral organ (red). (*c*) Schematic drawing of the adult thaliacean, *Thalia democratica*. Chemosensory cells are thought to be present in two places: the oral siphon (pink) and ciliary funnel (blue). (*d*) Schematic drawing of an amphioxus larva and the cells suspected to be chemosensory. The types of cells and relevant gene expression are indicated. Note that the Type 1 sensory cells are heterogeneous as indicated by different colours. Genes listed with an asterisk are only expressed in subsets of sensory cells. For more detail see [107,114]. Abbreviations: HP, Hatschek's pit. MO, mouth. Panel D adapted from [36] with permission.

are usually considered not to have cranial placodes in a strict sense [106]. However, they do have some ectodermal patterning mechanisms that appear conserved with those of vertebrate placodes. These have been best studied in the ascidians, which have two areas of ectoderm postulated to be placode homologues: one just anterior to the neural plate and a potential homologue of the olfactory and adenohypophyseal placodes (discussed more below) and the other paired and lateral to a more posterior neural plate. Data on ectodermal patterning have been recently reviewed in detail elsewhere [36,38,107] and the reader is referred here for more discussion on this aspect. Protochordates also have several types of morphologically distinct ectodermal sensory cells. Some have been proposed to be chemosensory, though this is mainly based on cytological features and the expression of marker genes, and no cell advocated as chemosensory has had this experimentally evaluated [107–112].

Amphioxus species are all quite similar in gross morphology. Tunicates are more disparate. The majority of species fall into the 'Ascidiacea', a paraphyletic grouping

united by an ascidian type life cycle, with a motile tadpole larva and a sessile adult that may be solitary, or may form colonies by asexual reproduction [113]. Other tunicates are motile as adults. This includes the larvaceans, which maintain the tadpole body plan throughout life, and thaliaceans (figures 1, 5*a–c*).

9. Putative olfactory cell homologues in Ascidiacea

Several studies have reported a failure to identify orthologues of the vertebrate olfactory receptor genes in tunicate genomes, and orthologues of chemosensory receptor genes used by insects and nematodes could also not be found [115–117]. Tunicate orthologues of V1R and V2R genes are also missing [74,75,118]. Furthermore, although TAAR-like genes have been proposed to be present in tunicates and amphioxus [75], these authors did not give details and others have disagreed with this as they failed to identify TAAR orthologues [74,119]. It would be surprising if tunicates do not sense any chemical cues, especially with an active swimming larval stage that in most species needs to choose an appropriate settlement site for metamorphosis. Further, in adult tunicates, there is evidence that the oral siphon can sense chemicals as sensitivity to acids, bases, salts and anaesthetics has been shown [120]. It is therefore probable that alternatives to olfactory receptor genes are used in these organisms. Recently, some seven-transmembrane protein encoding genes have been identified and proposed to fulfil the role (discussed more below), though protein function is not established [33,121].

The model ascidian *Ciona intestinalis* is by far the best-studied species in this field. During neurulation, the border region to the neural plate gives rise to several types of sensory cell, two of which are relevant to the discussion of olfactory placode evolution: a subset of anterior trunk epidermal neurons (the aATENs) and the palp sensory cells (PSCs) (figure 5*a*). Both arise from the cells just anterior to the neural-plate, in the region mentioned above as a prospective olfactory/adenohypophyseal homologue. The developmental pathway leading to the specification of these sensory cells has been expertly reviewed elsewhere [122], and the reader is referred here for a detailed description of how they form and then acquire distinct identities.

The aATENs are four ciliated primary sensory neurons that in *Ciona* larvae come to be located in the epidermis dorsal to the cavity of the sensory vesicle, their simple equivalent of the brain [123,124]. These neurons express *GnRH*, a cyclic nucleotide-gated channel (*CNGA*) and, as mentioned above, two seven-transmembrane G-protein-coupled receptors: a relaxin-3 receptor (*RXFP3*) and a somatostatin/opioid/galanin/chemokine-like receptor (*SOG/Chemokine-like*) [33]. This led the authors to propose that the aATENs had dual chemosensory and neurosecretory activities, combining functions of vertebrate olfactory-derived OSNs and GnRH neurons in a single cell [33]. However, caveats to this are that functional chemosensory activity has not been experimentally shown, and that the two seven-transmembrane G-protein-coupled receptors identified in these cells are not orthologous to the vertebrate olfactory receptors. Testing sensory cells for chemical stimulation has been an experimental challenge in a developmental system like *Ciona* as the small cells make electrophysiology difficult. The recent development of fluorescent cell activity reporters will probably resolve this technical bottleneck [121,125].

Ascidian tadpoles also have ciliated primary sensory neurons in each of the palps, the anterior adhesive organs by which larvae appear to sense and bond to attachment sites [112,126]. Like aATEN cells, the palps develop from the area anterior to the neural plate postulated to be an olfactory/adenohypophyseal placode homologue and express many regulatory genes that are important for olfactory placode development in vertebrates, like *Eya*, *COE*, *Dmrt*, *FoxC*, *FoxG*, *FGF*, *Sp8*, *Dlx* and *Isl* [127–130]. It is possible that the palp sensory neurons are involved in tadpole settlement site selection via chemoreception, though again this is not experimentally validated and others have suggested that a different cell type in the palps may be chemosensory [121]. The palps are also known to produce GnRH in cells likely to be neuronal as they seem to possess long axons [131], suggesting an evolutionary link to the olfactory placode-derived GnRH neurons in gnathostomes

In addition to these larval cell types, the oral siphon primordium (OSP; figure 5*a*) develops in the same region, with its progenitors sandwiched between those that give rise to the aATENs and those that give rise to the palps. The primordium maintains the expression of placode marker genes including *Six1/2*, *Six3/6* and *Pitx* into the larval stage and at metamorphosis forms the oral siphon, which in adults includes sensory cells in many ascidian species. Some of these cells may be chemosensory [120] but their developmental origin has yet to be traced so it is still possible they do not derive from the primordium. A structure called the ciliated funnel opens into the oral siphon and connects to a gland associated with a ganglion, the combination of which is termed the neural complex. This dual structure is reminiscent of the pituitary and homology has been considered [132], with the ciliated funnel sensing water entering the oral siphon and perhaps chemosensory. As for other ascidian sensory cells, this has not been experimentally validated.

10. Putative olfactory cells in larvaceans

It has been suggested that the ventral organ (figure 5*b*) of the larvacean tunicate *Oikopleura dioica* is homologous to the vertebrate olfactory organ. The ventral organ possesses about 30 primary sensory cells with cilia that protrude externally into sea water. These neurons are located in an ectodermal slit-like pocket and send their axons to the rostral-most CNS [133]. Furthermore, developmental genes like *Eya*, *Pitx* and *Six1/2*, which are important in the development of vertebrate olfactory and adenohypophyseal placodes, are expressed in the primordia of the larvacean ventral organ. Therefore, the *Oikopleura* ventral organ placode was said to be homologous to the ectoderm of the ascidian palps based on gene expression and structure [134]. Again, the sensory function of the cells has not been tested.

11. Putative olfactory cells in thaliacea

Thaliaceans comprise the pelagic tunicates salps, doliolids and pyrosomes. There has been less research in thaliaceans in comparison to other tunicate classes. In salps and doliolids, the cells located around the oral lips (figure 5*c*) have been suggested to be chemoreceptors based on the

observation that salps respond to chemical stimuli positioned in proximity of the oral opening [135]. Another possible chemosensory structure is the ciliated funnel, as discussed above with respect to ascidians and which is present in thaliaceans. In the thaliacean *Thalia democratica*, it has been suggested that the ciliated funnel could possibly collect odorants from the environment. [136]. Developmental and genetic confirmation of this hypothesis is currently lacking, however.

12. A summary of olfactory system homology in tunicates

Experiments show tunicates respond to chemical stimuli, and position and developmental genetic data point to the ectoderm just anterior to the neural plate as the homologue of the vertebrate olfactory and adenohypophyseal placodes. This area also produces sensory neurons, at least some of which express GnRH. It remains to be experimentally shown whether cells are chemosensory, and if so what receptors they use considering homologues of vertebrate receptor families are lacking. However, assuming they are chemosensory, the data strongly support the contention that the ancestor of the tunicates and vertebrates had a chemosensory system developing from the ectoderm alongside the anterior neural plate and that has evolved into the systems we see today in living tunicates and vertebrates. It is less clear how complicated this ancestral system was. In tunicate larvae, it consists of scattered sensory cells that may combine multiple functions, rather than a larger organ system with the morphogenesis and specialized cell types of vertebrates. This points to a simple grade of organization in the common ancestor. Adult tunicates have morphologically more complex organs with more cells, which could point to a more complex ancestral state. However, cell functions are again unknown, developmental genetics poorly understood and cell lineages unclear. This makes discriminating between conservation and parallelism or convergence challenging. Until additional data are available, the analysis of outgroups like amphioxus provides the alternative route to inferring ancestral states.

13. Putative olfactory cells in amphioxus

Amphioxus are known to exhibit sensitivity to chemicals dissolved in sea water, with most triggering an escape response [137,138]. Ectodermal territories with the gene expression, morphogenetic processes or focused neurogenesis characteristic of vertebrate placodes have not been identified, with the possible exception of Hatschek's pit (see below). Ectodermal sensory cells are present but are broadly scattered throughout the general epidermis (figure 5*d*), and at least some of these cells develop in the ventral ectoderm of the early embryo under BMP signalling [139]. It is not known which sensory modalities are mediated by each cell but some have been suggested to be chemosensory based on cytology and molecular markers (reviewed in [36,107,140,141]). There are two major subtypes of these epidermal sensory cells: type I and type II. The type I sensory cells are primary sensory cells with a long cilium surrounded by microvilli. The type II sensory cells are secondary neurons and have a short cilium encircled by a collar of microvilli [142–144].

The type I cells have been proposed to be mechano- and/or chemoreceptors [145]. The population of type I sensory neurons is molecularly heterogeneous and subsets express orthologues of transcription factor genes seen during vertebrate olfactory placode development such as *COE*, *Islet*, *POU4*, *SoxB1c*, *Six1/2*, *Six4/5* and *Eya* [114,146–149]. Some may be chemosensory, for instance, the expression of *COE* by type I sensory cells located caudally along the flanks of the amphioxus body might suggest this as COE genes are expressed in the olfactory placodes of vertebrates and chemosensory neurons of the organism such as *Caenorhabditis elegans* [97,150]. In particular, the anterior of amphioxus is interesting as it includes type I ciliated primary sensory neurons coming from an ectodermal region expressing olfactory placodal markers like *Pax6*, *Six3/6*, *Ngn* and *POU4* [146,148,151]. Furthermore, in the amphioxus *Branchiostoma floridae* genome, there are more than 50 genes orthologous to the vertebrate ORs [115,117,152,153], and some of these anterior type I cells express at least one of these genes suggesting they may be cell-type homologues of OSNs [154].

Type II cell morphology, with a collar of external projecting microvilli supported by a modified cilium, also suggests chemosensory rather than mechanosensory function. The cells could be homologous to vertebrate primary olfactory sensory cells, which means they would have secondarily lost their axons. They could also be homologous to secondary vertebrate chemosensory cells like the taste buds or solitary chemoreceptor cells. However, the molecular identity of the cells is unknown as their late appearance in larval development [143] has so far hindered molecular studies. All these hypotheses also need to be confirmed by physiological studies as none of these cells has had their sensory modality validated.

In addition of type I and type II epidermal sensory cells, there are other specialized sensory cell types in amphioxus that have been inferred to be chemosensory; the cells from the corpuscles de Quatrefages, the oral spine cells and Hatschek's pit cells (figure 5*d*). The corpuscles de Quatrefages are located in the rostrum and form two clusters of specialized ciliated primary sensory neurons that have one to four sensory cells with two cilia each, surrounded by up to seven sheath cells [155,156]. They have been speculated to form a mechanosensory organ but could well be chemosensory [157]. The ciliated oral spine cells around the mouth opening do not possess axons or microvilli, though they do express *Pitx*, *POU4* and *SoxB1c*, which are vertebrate anterior placodal markers [114,158]. These specialized cells have been suggested to be mechanoreceptors but also proposed to be homologous to vertebrate taste cells [159].

In adult amphioxus, Hatschek's pit is a structure located in the roof of the oral cavity and sends a projection dorsally around the notochord to contact the ventral brain, an organization similar to the relationship between adenohypophysis and hypothalamus of the vertebrate pituitary system [36,160]. The idea that Hatschek's pit and pituitary may be homologous organs is over 100 years old, and molecular analysis has added some support to this since the developmental precursor to Hatschek's pit (known as the pre-oral pit) expresses the vertebrate anterior placode markers *Six1/2*, *Eya*, *Pitx*, *Pax6* and *Six3/6*. There has also been a suggestion that Hatschek's pit includes chemosensory cells because they carry microvilli and cilia, and are exposed to water flowing into the mouth and so are in contact with potential odorants [161]. This hypothesis requires experimental corroboration as

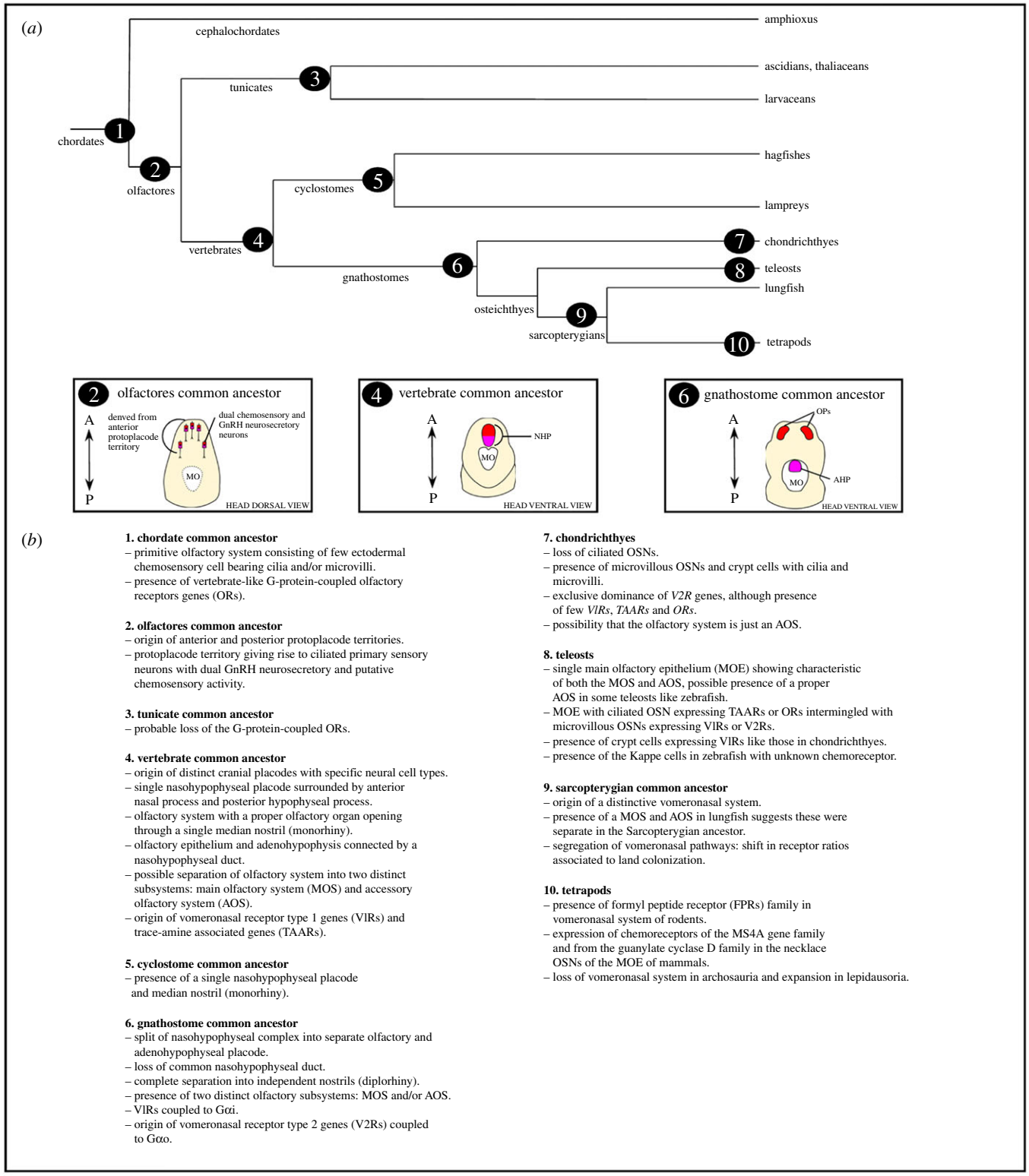


Figure 6. Schematic phylogeny of chordate lineages. Numbers in (*a*) relate to the key in (*b*) and indicate where key events in the evolution of the olfactory system are likely to have occurred. Abbreviations: AHP, adenohypophyseal placode. MO, mouth. NHP, nasohypophyseal placode. OP, olfactory placode.

the cytoarchitecture of the Hatschek's pit cells suggests that they are neurosecretory, and they do not have axons.

Immunochemistry with antibodies raised to vertebrate GnRH proteins has been used to suggest that GnRH neurosecretory cells are present in the amphioxus neural tube, and possibly also Hatschek's pit [162,163]. Neural tube GnRH cells are candidates for homologues of the olfactory-derived GnRH neurons of vertebrates, but nothing is currently known about their development: they might migrate in from an ectodermal territory like vertebrate GnRH1/3 cells, but also could be born within the central nervous system like vertebrate GnRH2 cells. The expression of GnRH in Hatschek's pit has also been questioned as it could not be detected in a later study, which only identified GnRH immunoreactivity in the central canal of the anterior nerve cord [164]. This difference in the detection of GnRH could be owing to genuine biological causes (for example, a difference in the reproductive state of the animals examined might affect GnRH expression), or might reflect experimental artefacts from fixation or antibody cross reactivity. Additional work is needed to clarify this.

Overall, Hatschek's pit appears similar to the vertebrate adenohypophysis but not the olfactory system. However, even here there is an important difference in that Hatschek's pit derives from the pre-oral pit of the larva, which itself develops from an anterior head cavity, an endodermal derivative. The adenohypophysis is an ectodermal placode derivative. The lineages of the relevant Hatschek's pit cell types through this developmental process need to be confirmed for certainty, but if they too derive from endoderm then homology would imply that the capacity to build this organ and its associated cell types has transferred from one germ layer to another in chordate evolution. Such a shift in cell fate could have occurred through the shifting in the expression domains of the relevant transcription factors, as proposed by Schlosser [37]. This possibility has had recent support from cell lineage studies in zebrafish, which have demonstrated that in this species some adenohypophysis cells may naturally derive from the endoderm during normal development [165]. While this could be a derived character of zebrafish and not reflect the ancestral condition, it irrespectively shows confinement of pituitary cell fates by germ layer is not as strict as historically imagined.

14. A summary of olfactory system homology in amphioxus

In general, the scattered ectodermal sensory neurons of amphioxus are poor candidates for olfactory cell homologues. Most are not anteriorly located and while some express genes that mark vertebrate placode cells, there are also substantial differences in their developmental programmes. Most importantly their early induction and regulation are different as they form far from the neural plate in ventral ectoderm under high Bmp signalling [139]. As such they more resemble another type of sensory neuron in tunicates, those in the ventral tail fin [122], and not aATENs, palp cells or vertebrate placode cells, which all originate in the neural plate boundary. Sensory cells in the anterior of amphioxus larvae may be the exception to this owing to their possible expression of amphioxus OR genes and anterior location. However, as yet their developmental genetics, cell lineage and sensory functions have not been determined and these data would be needed to convincingly test a hypothesis of homology. Hatschek's pit remains the best candidate for an adenohypophysis homologue, though this too has counterarguments as discussed above, and the expression of GnRH here remains to be convincingly established.

With respect to inferring the organization of olfactory sensation in the common ancestor of amphioxus and other chordates, the data do not support the presence of morphogenetic processes building an olfactory organ at this stage in evolution. Rather, chemosensation may have been mediated by scattered sensory cells. Amphioxus OR gene expression data suggest assigning some of these as cell-type homologues of vertebrate OSNs but more work on their development is needed to support this. It would be especially important to know what specifies OR-expressing cells and if their lineage traces back to the anterior neural plate border.

15. A model for vertebrate olfactory system evolution

By plotting genes, developmental processes, cell types and tissues onto a phylogeny of the chordates we arrive at a model for how the olfactory systems of living vertebrates evolved (figure 6). Chemosensation is an ancient sense and was likely mediated by scattered sensory cells in the epidermis of the chordate common ancestor (figure 6–1). Some of these cells probably expressed orthologues of the vertebrate OR genes, as still seen in living amphioxus. By the common ancestor of the tunicates and vertebrates (collectively the Olfactores), specialized ectodermal territories homologous to vertebrate placodes had evolved. We have not discussed the detailed evidence for this here as it has been recently and extensively reviewed [36,38,107]; however, the data suggest the Olfactores ancestor had two placode-like territories and that the anterior of these is the source of the olfactory and adenohypophyseal placodes of vertebrates (figure 6–2). In living tunicates, the anterior placode territory persists and is the source of GnRH and probably chemosensory cells, though the tunicate lineage also lost conventional OR genes.

By the common ancestor of the vertebrates (figure 6–4) a well-defined anterior placode combining olfactory and adenohypophyseal progenitors had evolved, as well as the morphogenetic processes by which it built distinct olfactory and adenohypophyseal systems. It is possible this included distinct main and accessory olfactory system components, including the evolution and deployment of *V1R* receptor genes. We do not yet know whether it also included the progenitors of GnRH neurons that then migrated to the pre-optic area and hypothalamus as this has not been determined in cyclostomes. Key innovations that map to the origin of vertebrates therefore include: (i) the evolution of new types of receptor gene; (ii) as proposed by Schlosser [107] for placode evolution more generally, changes in the control of progenitor proliferation turning single neurons into a neurogenic organ; (iii) mechanisms for specifying subpopulations of OSNs expressing different types of receptor; (iv) changes in morphogenesis including the incorporation of neural crest cells and interactions between placode ectoderm and cranial ectomesenchyme.

In the common ancestor of jawed vertebrates (figure 6–6), this combined placode separated into a medial adenohypophyseal placode and paired olfactory placodes, possibly facilitating the evolution of paired nostrils as seen in all living jawed vertebrates. Subsequently the loss of main or accessory olfactory systems has occurred in some vertebrate lineages, with concomitant shifts in dependence on ORs, V1Rs and V2Rs (figure 6–7, 6–10). It is also interesting that studies of model systems have now identified additional olfactory chemosensory mechanisms beyond these well-known receptors, such as the MS4As, TAARs and FPRs. The evolutionary ancestries of these are less well known and worthy of more study. These findings come from just a handful of species, raising the possibility that the many thousands of less well-studied vertebrate species may harbour additional surprises on this front and that the diversity of olfactory mechanisms deployed by vertebrates may be far higher than currently understood.

Data accessibility. This article has no additional data.

Authors' contributions. Both authors drafted the manuscript and collaborated in developing the figures and table.

Competing interests. We declare we have no competing interests.

Funding. We received no funding for this study.

References

1. Suarez R, Garcia-Gonzalez D, de Castro F. 2012 Mutual influences between the main olfactory and vomeronasal systems in development and evolution. *Front. Neuroanat.* **6**, 50. (doi:10.3389/fnana.2012.00050)
2. Oka Y, Korsching SI. 2011 Shared and unique G alpha proteins in the zebrafish versus mammalian senses of taste and smell. *Chem. Senses* **36**, 357–365. (doi:10.1093/chemse/bjq138)
3. Buck L, Axel R. 1991 A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187. (doi:10.1016/0092-8674(91)90418-X)
4. Dulac C, Torello AT. 2003 Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nat. Rev. Neurosci.* **4**, 551–562. (doi:10.1038/nrn1140)
5. Bear DM, Lassance JM, Hoekstra HE, Datta SR. 2016 The evolving neural and genetic architecture of vertebrate olfaction. *Curr. Biol.* **26**, R1039–R1R49. (doi:10.1016/j.cub.2016.09.011)
6. Niimura Y, Matsui A, Touhara K. 2014 Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res.* **24**, 1485–1496. (doi:10.1101/gr.169532.113)
7. Greer PL *et al.* 2016 A family of non-GPCR chemosensors defines an alternative logic for mammalian olfaction. *Cell* **165**, 1734–1748. (doi:10.1016/j.cell.2016.05.001)
8. Riviere S, Challet L, Fluegge D, Spehr M, Rodriguez I. 2009 Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. *Nature* **459**, 574–577. (doi:10.1038/nature08029)
9. Gonzalez A, Morona R, Lopez JM, Moreno N, Northcutt RG. 2010 Lungfishes, like tetrapods, possess a vomeronasal system. *Front. Neuroanat.* **4**, 1–11. (doi:10.3389/fnana.2010.00130)
10. Eisthen HL. 1992 Phylogeny of the vomeronasal system and of receptor cell types in the olfactory and vomeronasal epithelia of vertebrates. *Microsc. Res. Tech.* **23**, 1–21. (doi:10.1002/jemt.1070230102)
11. Hussain A. 2011 The olfactory nervous system of terrestrial and aquatic vertebrates. *Nat. Precedings.* (doi:10.1038/npre.2011.6642.1)
12. Biechl D, Tietje K, Ryu S, Grothe B, Gerlach G, Wullimann MF. 2017 Identification of accessory olfactory system and medial amygdala in the zebrafish. *Sci. Rep.* **7**, 44295. (doi:10.1038/srep44295)
13. Saraiva LR, Korsching SI. 2007 A novel olfactory receptor gene family in teleost fish. *Genome Res.* **17**, 1448–1457. (doi:10.1101/gr.6553207)
14. Ahuja G, Bozorg Nia S, Zapilko V, Shiriagin V, Kowatschew D, Oka Y, Korsching SI. 2014 Kappe neurons, a novel population of olfactory sensory neurons. *Sci. Rep.* **4**, 4037. (doi:10.1038/srep04037)
15. Ferrando S, Gallus L, Gambardella C, Vacchi M, Tagliafierro G. 2010 G protein alpha subunits in the olfactory epithelium of the holocephalan fish *Chimaera monstrosa*. *Neurosci. Lett.* **472**, 65–67. (doi:10.1016/j.neulet.2010.01.059)
16. Sharma K, Syed AS, Ferrando S, Mazan S, Korsching SI. 2019 The chemosensory receptor repertoire of a true shark is dominated by a single olfactory receptor family. *Genome Biol. Evol.* **11**, 398–405. (doi:10.1093/gbe/evz002)
17. Ferrando S, Bottaro M, Gallus L, Girosi L, Vacchi M, Tagliafierro G. 2006 Observations of crypt neuron-like cells in the olfactory epithelium of a cartilaginous fish. *Neurosci. Lett.* **403**, 280–282. (doi:10.1016/j.neulet.2006.04.056)
18. Ferrando S, Gallus L. 2013 Is the olfactory system of cartilaginous fishes a vomeronasal system? *Front. Neuroanat.* **7**, 37. (doi:10.3389/fnana.2013.00037)
19. Aguillon R, Batut J, Subramanian A, Madelaine R, Dufourcq P, Schilling TF, Blader P. 2018 Cell-type heterogeneity in the early zebrafish olfactory epithelium is generated from progenitors within preplacodal ectoderm. *Elife* **7**, e32401. (doi:10.7554/eLife.32041)
20. Sabado V, Barraud P, Baker CV, Streit A. 2012 Specification of GnRH-1 neurons by antagonistic FGF and retinoic acid signaling. *Dev. Biol.* **362**, 254–262. (doi:10.1016/j.ydbio.2011.12.016)
21. Gaillard AL *et al.* 2018 Characterization of gonadotropin-releasing hormone (GnRH) genes from cartilaginous fish: evolutionary perspectives. *Front. Neurosci.* **12**, 607. (doi:10.3389/fnins.2018.00607)
22. Duan C, Allard J. 2020 Gonadotropin-releasing hormone neuron development in vertebrates. *Gen. Comp. Endocrinol.* **292**, 113465. (doi:10.1016/j.ygcen.2020.113465)
23. Umatani C, Oka Y. 2019 Multiple functions of non-hypophysiotropic gonadotropin releasing hormone neurons in vertebrates. *Zool. Lett.* **5**, 23. (doi:10.1186/s40851-019-0138-y)
24. Northcutt RG, Muske LE. 1994 Multiple embryonic origins of gonadotropin-releasing hormone (GnRH) immunoreactive neurons. *Dev. Brain Res.* **78**, 279–290. (doi:10.1016/0165-3806(94)90037-X)
25. Cho HJ, Shan Y, Whittington NC, Wray S. 2019 Nasal placode development, GnRH neuronal migration and Kallmann syndrome. *Front. Cell Dev. Biol.* **7**, 121. (doi:10.3389/fcell.2019.00121)
26. Wierman ME, Kiseljak-Vassiliades K, Tobet S. 2011 Gonadotropin-releasing hormone (GnRH) neuron migration: initiation, maintenance and cessation as critical steps to ensure normal reproductive function. *Front. Neuroendocrinol.* **32**, 43–52. (doi:10.1016/j.yfrne.2010.07.005)
27. Wray S. 2002 Molecular mechanisms for migration of placodally derived GnRH neurons. *Chem. Senses* **27**, 569–572. (doi:10.1093/chemse/27.6.569)
28. Hilal EM, Chen JH, Silverman A-J. 1996 Joint migration of gonadotropin-releasing hormone (GnRH) and neuropeptide Y (NPY) neurons from olfactory placode to central nervous system. *J. Neurobiol.* **31**, 487–502. (doi:10.1002/(SICI)1097-4695(199612)31:4<487::AID-NEU8>3.0.CO;2-5)
29. Wirsig-Wiechmann CR. 2001 Function of gonadotropin-releasing hormone in olfaction. *Keio J. Med.* **50**, 81–85. (doi:10.2302/kjm.50.81)
30. Eisthen HL, Delay RJ, Wirsig-Wiechmann CR, Dionne VE. 2000 Neuromodulatory effects of gonadotropin releasing hormone on olfactory receptor neurons. *J. Neurosci.* **20**, 3947–3955. (doi:10.1523/JNEUROSCI.20-11-03947.2000)
31. Kawai T, Oka Y, Eisthen H. 2009 The role of the terminal nerve and GnRH in olfactory system neuromodulation. *Zool. Sci.* **26**, 669–680. (doi:10.2108/zsj.26.669)
32. Koide T, Yabuki Y, Yoshihara Y. 2018 Terminal nerve GnRH3 neurons mediate slow avoidance of carbon dioxide in larval zebrafish. *Cell Rep.* **22**, 1115–1123. (doi:10.1016/j.celrep.2018.01.019)
33. Abitua PB *et al.* 2015 The pre-vertebrate origins of neurogenic placodes. *Nature* **524**, 462–465. (doi:10.1038/nature14657)
34. Katoh H *et al.* 2011 The dual origin of the peripheral olfactory system: placode and neural crest. *Mol. Brain.* **4**, 34. (doi:10.1186/1756-6606-4-34)
35. Baker CV, Bronner-Fraser M. 2001 Vertebrate cranial placodes I. Embryonic induction. *Dev. Biol.* **232**, 1–61. (doi:10.1006/dbio.2001.0156)
36. Patthey C, Schlosser G, Shimeld SM. 2014 The evolutionary history of vertebrate cranial placodes–I: cell type evolution. *Dev. Biol.* **389**, 82–97. (doi:10.1016/j.ydbio.2014.01.017)
37. Schlosser G. 2005 Evolutionary origins of vertebrate placodes: insights from developmental studies and from comparisons with other deuterostomes. *J. Exp. Zool. B (Mol. Dev. Evol.)* **304**, 347–399. (doi:10.1002/jez.b.21055)
38. Schlosser G, Patthey C, Shimeld SM. 2014 The evolutionary history of vertebrate cranial placodes II. Evolution of ectodermal patterning. *Dev. Biol.* **389**, 98–119. (doi:10.1016/j.ydbio.2014.01.019)
39. Saint-Jeannet JP, Moody SA. 2014 Establishing the pre-placodal region and breaking it into placodes with distinct identities. *Dev. Biol.* **389**, 13–27. (doi:10.1016/j.ydbio.2014.02.011)
40. Schlosser G. 2006 Induction and specification of cranial placodes. *Dev. Biol.* **294**, 303–351. (doi:10.1016/j.ydbio.2006.03.009)
41. Aguillon R, Blader P, Batut J. 2016 Patterning, morphogenesis, and neurogenesis of zebrafish cranial sensory placodes. *Methods Cell Biol.* **134**, 33–67. (doi:10.1016/bs.mcb.2016.01.002)
42. Toro S, Varga ZM. 2007 Equivalent progenitor cells in the zebrafish anterior preplacodal field give rise to adenohypophysis, lens, and olfactory placodes. *Semin. Cell Dev. Biol.* **18**, 534–542. (doi:10.1016/j.semcdb.2007.04.003)
43. Sjodal M, Edlund T, Gunhaga L. 2007 Time of exposure to BMP signals plays a key role in the specification of the olfactory and lens placodes ex

vivo. *Dev. Cell.* **13**, 141–149. (doi:10.1016/j.devcel.2007.04.020)

44. Bailey AP, Bhattacharyya S, Bronner-Fraser M, Streit A. 2006 Lens specification is the ground state of all sensory placodes, from which FGF promotes olfactory identity. *Dev. Cell.* **11**, 505–517. (doi:10.1016/j.devcel.2006.08.009)
45. Huang X, Hong CS, O'Donnell M, Saint-Jeannet JP. 2005 The doublesex-related gene, *XDmrt4*, is required for neurogenesis in the olfactory system. *Proc. Natl Acad. Sci. USA* **102**, 11 349–11 354. (doi:10.1073/pnas.0505106102)
46. Forni PE, Taylor-Burds C, Melvin VS, Williams T, Wray S. 2011 Neural crest and ectodermal cells intermix in the nasal placode to give rise to GnRH-1 neurons, sensory neurons, and olfactory ensheathing cells. *J. Neurosci.* **31**, 6915–6927. (doi:10.1523/JNEUROSCI.6087-10.2011)
47. Barraud P, Seferiadis AA, Tyson LD, Zwart MF, Szabo-Rogers HL, Ruhrberg C, Liu KJ, Baker CVH. 2010 Neural crest origin of olfactory ensheathing glia. *Proc. Natl Acad. Sci. USA* **107**, 21 040–21 045. (doi:10.1073/pnas.1012248107)
48. Whitlock KE, Wolf CD, Boyce ML. 2003 Gonadotropin-releasing hormone (GnRH) cells arise from cranial neural crest and adenohypophy seal regions of the neural plate in the zebrafish, *Danio rerio*. *Dev. Biol.* **257**, 140–152. (doi:10.1016/S0012-1606(03)00039-3)
49. Shan Y, Saadi H, Wray S. 2020 Heterogeneous origin of gonadotropin releasing hormone-1 neurons in mouse embryos detected by islet-1/2 expression. *Front. Cell Dev. Biol.* **8**, 35. (doi:10.3389/fcell.2020.00035)
50. Johanson Z, Bosivert CA, Trinajstic K. 2019 Early vertebrates and the emergence of jaws. In *Heads, jaws, and muscles: anatomical, functional, and developmental diversity in chordate evolution* (eds JM Ziermann, RE Diaz Jr, R Diogo), pp. 23–44. Berlin, Germany: Springer.
51. Kuratani S, Ueki T, Aizawa S, Hirano S. 1997 Peripheral development of cranial nerves in a cyclostome, *Lampetra japonica*: morphological distribution of nerve branches and the vertebrate body plan. *J. Comp. Neurol.* **384**, 483–500. (doi:10.1002/(SICI)1096-9861(19970811)384:4<483::AID-CNE1>3.0.CO;2-Z)
52. Modrell MS, Hockman D, Uy B, Buckley D, Sauka-Spengler T, Bronner ME, Baker CVH. 2014 A fate-map for cranial sensory ganglia in the sea lamprey. *Dev. Biol.* **385**, 405–416. (doi:10.1016/j.ydbio.2013.10.021)
53. Parker HJ, Bronner ME, Krumlauf R. 2014 A *Hox* regulatory network of hindbrain segmentation is conserved to the base of vertebrates. *Nature* **514**, 490–493. (doi:10.1038/nature13723)
54. Sugahara F *et al.* 2016 Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain. *Nature* **531**, 97–100. (doi:10.1038/nature16518)
55. Bjerselius R, Li W, Teeter JH, Seelye JG, Johnsen PB, Maniak PJ, Grant GC, Polkinghorne CN, Sorensen PW. 2000 Direct behavioral evidence that unique bile acids released by larval sea lamprey (*Petromyzon marinus*) function as a migratory pheromone. *Can. J. Fish. Aquatic Sci.* **57**, 557–569. (doi:10.1139/f99-290)
56. Li K, Brant CO, Huertas M, Hessler EJ, Mezei G, Scott AM, Hoye TR, Li W. 2018 Fatty-acid derivative acts as a sea lamprey migratory pheromone. *Proc. Natl Acad. Sci. USA* **115**, 8603–8608. (doi:10.1073/pnas.1803169115)
57. Li W, Sorensen PW, Gallaher DD. 1995 The olfactory system of migratory adult sea lamprey (*Petromyzon marinus*) is specifically and acutely sensitive to unique bile acids released by conspecific larvae. *J. Gen. Physiol.* **105**, 569–587. (doi:10.1085/jgp.105.5.569)
58. Siefkes MJ, Li W. 2004 Electrophysiological evidence for detection and discrimination of pheromonal bile acids by the olfactory epithelium of female sea lampreys (*Petromyzon marinus*). *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **190**, 193–199. (doi:10.1007/s00359-003-0484-1)
59. Zielinski BS, Osahan JK, Hara TJ, Hosseini M, Wong E. 1996 Nitric oxide synthase in the olfactory mucosa of the larval sea lamprey (*Petromyzon marinus*). *J. Comp. Neurol.* **365**, 18–26. (doi:10.1002/(SICI)1096-9861(19960129)365:1<18::AID-CNE2>3.0.CO;2-M)
60. Glover CN, Newton D, Bajwa J, Goss GG, Hamilton TJ. 2019 Behavioural responses of the hagfish *Eptatretus stoutii* to nutrient and noxious stimuli. *Sci. Rep.* **9**, 13369. (doi:10.1038/s41598-019-49863-x)
61. Martinez I, Jones EG, Davie SL, Neat FC, Wigham BD, Priede IG. 2011 Variability in behaviour of four fish species attracted to baited underwater cameras in the North Sea. *Hydrobiologia* **670**, 23–34. (doi:10.1007/s10750-011-0672-x)
62. Døving KB, Holmberg K. 1974 A note on the function of the olfactory organ of the hagfish *Myxine glutinosa*. *Acta Physiol. Scand.* **91**, 430–432. (doi:10.1111/j.1748-1716.1974.tb05698.x)
63. Braun CB. 1998 Schreiner organs: a new craniate chemosensory modality in hagfishes. *J. Comp. Neurol.* **392**, 135–163. (doi:10.1002/(SICI)1096-9861(19980309)392:2<135::AID-CNE1>3.0.CO;2-3)
64. Oisi Y, Ota KG, Kuraku S, Fujimoto S, Kuratani S. 2013 Craniofacial development of hagfishes and the evolution of vertebrates. *Nature* **493**, 175–180. (doi:10.1038/nature11794)
65. Dean B. 1899 On the embryology of *Bdellostoma stouti*. A genera account of myxinoid development from the egg and segmentation to hatching. In *Festschrift zum 70ten Geburstag Carl von Kupffer*, pp. 220–276. Jena, Germany: Gustav Fischer Verlag.
66. Tahara Y. 1988 Normal stages of development in the lamprey, *Lampetra reissneri* (Dybowski). *Zool. Sci.* **5**, 109–118.
67. Leach JW. 1951 The hypohysis of lampreys in relation to the nasal apparatus. *J. Morph.* **89**, 217–255. (doi:10.1002/jmor.1050890203)
68. Kleerekoper H, van Erkel GA. 1960 The olfactory apparatus of *Petromyzon marinus* L. *Can. J. Zool.* **38**, 209–223. (doi:10.1139/z60-025)
69. VanDenbosshe J, Youson JH, Pohlman D, Wong E, Zielinski BS. 1997 Metamorphosis of the olfactory organ of the sea lamprey (*Petromyzon marinus* L.): morphological changes and morphometric analysis. *J. Morphol.* **231**, 41–52. (doi:10.1002/(SICI)1097-4687(199701)231:1<41::AID-JMOR4>3.0.CO;2-R)
70. Thornhill RA. 1967 The ultrastructure of the olfactory epithelium of the lamprey *Lampetra fluviatilis*. *J. Cell Sci.* **2**, 591–602.
71. VanDenbosshe J, Seelye JG, Zielinski BS. 1995 The morphology of the olfactory epithelium in larval, juvenile and upstream migrant stages of the sea lamprey, *Petromyzon marinus*. *Brain Behav. Evol.* **45**, 19–24. (doi:10.1159/000113382)
72. Laframboise AJ, Ren X, Chang S, Dubuc R, Zielinski BS. 2007 Olfactory sensory neurons in the sea lamprey display polymorphisms. *Neurosci. Lett.* **414**, 277–281. (doi:10.1016/j.neulet.2006.12.037)
73. Freitag J, Beck A, Gunther L, von Buchholtz L, Breer H. 1999 On the origin of the olfactory receptor family: receptor genes of the jawless fish (*Lampetra fluviatilis*). *Gene* **226**, 165–174. (doi:10.1016/S0378-1119(98)00575-7)
74. Grus WE, Zhang J. 2006 Origin and evolution of the vertebrate vomeronasal system viewed through system-specific genes. *Bioessays* **28**, 709–718. (doi:10.1002/bies.20432)
75. Libants S, Carr K, Wu H, Teeter JH, Chung-Davidson YW, Zhang Z, Wilkerson C, Li W. 2009 The sea lamprey *Petromyzon marinus* genome reveals the early origin of several chemosensory receptor families in the vertebrate lineage. *BMC Evol. Biol.* **9**, 180. (doi:10.1186/1471-2148-9-180)
76. Ubeda-Banon I *et al.* 2011 Cladistic analysis of olfactory and vomeronasal systems. *Front. Neuroanat.* **5**, 3. (doi:10.3389/fnana.2011.00003)
77. Green WW, Basilious A, Dubuc R, Zielinski BS. 2013 The neuroanatomical organization of projection neurons associated with different olfactory bulb pathways in the sea lamprey, *Petromyzon marinus*. *PLoS ONE* **8**, e69525. (doi:10.1371/journal.pone.0069525)
78. Scott WB. 1887 Notes on the development of Petromyzon. *J. Morphol.* **1**, 253–310. (doi:10.1002/jmor.1050010203)
79. de Beer GR. 1924 On a problematical organ of the lamprey. *J. Anat.* **59**, 97–107.
80. Ren X, Chang S, Laframboise A, Green W, Dubuc R, Zielinski B. 2009 Projections from the accessory olfactory organ into the medial region of the olfactory bulb in the sea lamprey (*Petromyzon marinus*): a novel vertebrate sensory structure? *J. Comp. Neurol.* **516**, 105–116. (doi:10.1002/cne.22100)
81. Derjean D, Moussaddy A, Atallah E, St-Pierre M, Auclair F, Chang S, Ren X, Zielinski B, Dubuc R. 2010 A novel neural substrate for the transformation of olfactory inputs into motor output. *PLoS Biol.* **8**, e1000567. (doi:10.1371/journal.pbio.1000567)
82. Chang S, Chung-Davidson Y-W, Libants SV, Nanlohy KG, Kiupel M, Brown C, Li W. 2013 The sea lamprey has a primordial accessory olfactory system.

BMC Evol. Biol. **13**, 1–11. (doi:10.1186/1471-2148-13-172)

83. Hagelin L, Johnels AG. 1955 On the structure and function of the accessory olfactory organ in lampreys. *Acta Zool.* **36**, 113–125. (doi:10.1111/j.1463-6395.1955.tb00376.x)
84. Frontini A, Zaidi AU, Hua H, Wolak TP, Greer CA, Kafitz KW, Li W, Zielinski BS. 2003 Glomerular territories in the olfactory bulb from the larval stage of the sea lamprey *Petromyzon marinus*. *J. Comp. Neurol.* **465**, 27–37. (doi:10.1002/cne.10811)
85. Muske LE. 1993 Evolution of gonadotropin-releasing hormone (GnRH) neuronal systems. *Brain Behav. Evol.* **42**, 215–230. (doi:10.1159/000114156)
86. Tobet SA, Chickering TW, Sower SA. 1996 Relationship of gonadotropin-releasing hormone (GnRH) neurons to the olfactory system in developing lamprey (*Petromyzon marinus*). *J. Comp. Neurol.* **376**, 97–111. (doi:10.1002/(SICI)1096-9861(19961202)376:1<97::AID-CNE6>3.0.CO;2-J)
87. Døving KB. 1998 The olfactory system of hagfishes. In *The biology of hagfishes* (eds JM Jørgensen, JP Lomholt, RE Weber, H Malte), pp. 533–540. Dordrecht, The Netherlands: Springer. (doi:10.1007/978-94-011-5834-3_33)
88. Theisen B. 1973 The olfactory system in the hagfish *Myxine glutinosa*. *Acta Zool.* **54**, 271–284. (doi:10.1111/j.1463-6395.1973.tb00462.x)
89. Theisen B. 1976 The olfactory system in the pacific hagfishes *Eptatretus stoutii*, *Eptatretus deani*, and *Myxine circifrons*. *Acta Zool.* **57**, 167–173. (doi:10.1111/j.1463-6395.1976.tb00224.x)
90. Holmes WM, Cotton R, Xuan VB, Rygg AD, Craven BA, Abel RL, Slack R, Cox JPL. 2011 Three-dimensional structure of the nasal passageway of a hagfish and its implications for olfaction. *Anat. Rec. (Hoboken)* **294**, 1045–1056. (doi:10.1002/ar.21382)
91. Braun CB, Wicht H, Northcutt GR. 1995 Distribution of gonadotropin-releasing hormone immunoreactivity in the brain of the pacific hagfish, *Eptatretus stouti* (Craniata: Myxinoidea). *J. Comp. Neurol.* **353**, 464–476. (doi:10.1002/cne.903530313)
92. Sower SA, Nozaki M, Knox CJ, Gorbman A. 1995 The occurence and distribution of GnRH in the brain of Atlantic hagfish, an agnatha, determined by chromatography and immunocytichemistry. *Gen. Comp. Endocrinol.* **97**, 300–307. (doi:10.1006/gcen.1995.1030)
93. Blähser S, King JA, Kuenzel WJ. 1989 Testing of Arg-8-gonadotropin-releasing hormone-directed antisera by immunological and immunocytochemical methods for use in comparative studies. *Histochemistry* **93**, 39–48. (doi:10.1007/BF00266845)
94. Derobert Y, Baratte B, Lepage M, Mazan S. 2002 *Pax6* expression patterns in *Lampetra fluviatilis* and *Scyliorhinus canicula* embryos suggest highly conserved roles in the early regionalization of the vertebrate brain. *Brain Res. Bullet.* **57**, 277–280. (doi:10.1016/S0361-9230(01)00695-5)
95. Murakami Y, Ogasawara M, Sugahara F, Hirano S, Satoh N, Kuratani S. 2001 Identification and expression of the lamprey *Pax6* gene: evolutionary origin of the segmented brain of vertebrates. *Development* **128**, 3521–3531.
96. Uchida K, Murakami Y, Kuraku S, Hirano S, Kuratani S. 2003 Development of the adenohypophysis in the lamprey: evolution of epigenetic patterning programs in organogenesis. *J. Exp. Zool.* **300**, 32–47. (doi:10.1002/jez.b.44)
97. Lara-Ramirez R, Poncelet G, Patthey C, Shimeld SM. 2017 The structure, splicing, synteny and expression of lamprey *COE* genes and the evolution of the *COE* gene family in chordates. *Dev. Genes Evol.* **227**, 319–338. (doi:10.1007/s00427-017-0591-6)
98. Myojin M, Ueki T, Sugahara F, Murakami Y, Shigetani Y, Aizawa S, Hirano S, Kuratani S. 2001 Isolation of *Dlx* and *Emx* gene cognates in an Agnathan species, *Lampetra japonica*, and their expression patterns during embryonic and larval development: conserved and diversified regulatory patterns of homeobox genes in vertebrate head evolution. *J. Exp. Zool.* **291**, 68–84. (doi:10.1002/jez.6)
99. Tank EM, Dekker RG, Beauchamp K, Wilson KA, Boehmke AE, Langeland JA. 2009 Patterns and consequences of vertebrate *Emx* gene duplications. *Evol. Dev.* **11**, 343–353. (doi:10.1111/j.1525-142X.2009.00341.x)
100. Gai Z, Donoghue PC, Zhu M, Janvier P, Stampanoni M. 2011 Fossil jawless fish from China foreshadows early jawed vertebrate anatomy. *Nature* **476**, 324–327. (doi:10.1038/nature10276)
101. Janvier P. 1996 *Early vertebrates*. New York, NY: Oxford Scientific Publications.
102. Kuratani S, Nobusada Y, Horigome N, Shigetani Y. 2001 Embryology of the lamprey and evolution of the vertebrate jaw: insights from molecular and developmental perspectives. *Phil. Trans. R. Soc. B.* **356**, 1615–1632. (doi:10.1098/rstb.2001.0976)
103. Shu D-G *et al.* 2003 Head and backbone of the Early Cambrian vertebrate *Haikouichthys*. *Nature* **421**, 526–529. (doi:10.1038/nature01264)
104. Morris SC, Caron JB. 2014 A primitive fish from the Cambrian of North America. *Nature* **512**, 419–422. (doi:10.1038/nature13414)
105. Pombal MA, Megías M. 2019 Development and functional organization of the cranial nerves in lampreys. *The Anatomical Rec.* **302**, 512–539. (doi:10.1002/ar.23821)
106. Gans C, Northcutt RG. 1983 Neural crest and the origin of vertebrates: a new head. *Science* **220**, 268–274. (doi:10.1126/science.220.4594.268)
107. Schlosser G. 2017 From so simple a beginning–what amphioxus can teach us about placode evolution. *Int. J. Dev. Biol.* **61**, 633–648. (doi:10.1387/ijdb.170127gs)
108. Caicci F, Zaniolo G, Burighel P, Degasperi V, Gasparini F, Manni L. 2010 Differentiation of papillae and rostral sensory neurons in the larva of the ascidian *Botryllus schlosseri* (Tunicata). *J. Comp. Neurol.* **518**, 547–566. (doi:10.1002/cne.22222)
109. Groppelli S, Pennati R, Scarì G, Sotgia C, De Bernardi F. 2003 Observations on the settlement of *Phallusia mammillata* larvae: effects of different lithological substrata. *Italian J. Zool.* **70**, 321–326. (doi:10.1080/11250000309356537)
110. Imai JH, Meinertzhagen IA. 2007 Neurons of the ascidian larval nervous system in *Ciona intestinalis*: II. Peripheral nervous system. *J. Comp. Neurol.* **501**, 335–352. (doi:10.1002/cne.21247)
111. Torrence SA, Cloney RA. 1982 Nervous system of ascidian larvae: caudally primary sensory neurons. *Zoomorphology* **99**, 103–115. (doi:10.1007/BF00310303)
112. Torrence SA, Cloney RA. 1983 Ascidian larval nervous system: primary sensory neurons in adhesive papillae. *Zoomorphology* **102**, 111–123. (doi:10.1007/BF00363804)
113. Delsuc F *et al.* 2018 A phylogenomic framework and timescale for comparative studies of tunicates. *BMC Biol.* **16**, 39. (doi:10.1186/s12915-018-0499-2)
114. Meulemans D, Bronner-Fraser M. 2007 The Amphioxus *SoxB* Family: implications for the evolution of vertebrate placodes. *Int. J. Biol. Sci.* **3**, 356–364. (doi:10.7150/ijbs.3.356)
115. Churcher AM, Taylor JS. 2009 Amphioxus (*Branchiostoma floridae*) has orthologs of vertebrate odorant receptors. *BMC Evol. Biol.* **9**, 242. (doi:10.1186/1471-2148-9-242)
116. Kamesh N, Aradhyam GK, Manoj N. 2008 The repertoire of G protein-coupled receptors in the sea squirt *Ciona intestinalis*. *BMC Evol. Biol.* **8**, 129. (doi:10.1186/1471-2148-8-129)
117. Niimura Y. 2009 Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. *Hum. Genomics* **4**, 107–118. (doi:10.1186/1479-7364-4-2-107)
118. Nordstrom KJ, Fredriksson R, Schioth HB. 2008 The amphioxus (*Branchiostoma floridae*) genome contains a highly diversified set of G protein-coupled receptors. *BMC Evol. Biol.* **8**, 9. (doi:10.1186/1471-2148-8-9)
119. Eyun SI *et al.* 2017 Evolutionary history of chemosensory-related gene families across the arthropoda. *Mol. Biol. Evol.* **34**, 1838–1862. (doi:10.1093/molbev/msx147)
120. Hecht S. 1918 The physiology of Ascidia atra lesueur. *J. Exp. Zool.* **25**, 229–259. (doi:10.1002/jez.1400250108)
121. Johnson CJ, Razy-Krajka F, Stolfi A. 2020 Expression of smooth muscle-like effectors and core cardiomyocyte regulators in the contractile papillae of *Ciona*. *EvoDevo* **11**, 15. (doi:10.1186/s13227-020-00162-x)
122. Liu B, Satou Y. 2020 The genetic program to specify ectodermal cells in ascidian embryos. *Dev. Growth Differ.* **62**, 301–310. (doi:10.1111/dgd.12660)
123. Imai JH, Meinertzhagen IA. 2007 Neurons of the ascidian larval nervous system in *Ciona intestinalis*: I. Central nervous system. *J. Comp. Neurol.* **501**, 316–334. (doi:10.1002/cne.21246)
124. Yokoyama TD, Hotta K, Oka K. 2014 Comprehensive morphological analysis of individual peripheral neuron dendritic arbors in ascidian larvae using the photoconvertible protein Kaede. *Dev. Dyn.* **243**, 1362–1373. (doi:10.1002/dvdy.24169)

125. Okawa N, Shimai K, Ohnishi K, Ohkura M, Nakai J, Horie T, Kuhara A, Kusakabe TG. 2020 Cellular identity and Ca^{2+} signaling activity of the non-reproductive GnRH system in the *Ciona intestinalis* type A (*Ciona robusta*) larva. *Sci. Rep.* **10**, 18590. (doi:10.1038/s41598-020-75344-7)
126. Zeng F, Wunderer J, Salvenmoser W, Hess MW, Ladurner P, Rothbacher U. 2019 Papillae revisited and the nature of the adhesive secreting collocytes. *Dev. Biol.* **448**, 183–198. (doi:10.1016/j.ydbio.2018.11.012)
127. Cao C *et al.* 2019 Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349–354. (doi:10.1038/s41586-019-1385-y)
128. Liu B, Satou Y. 2019 *Foxg* specifies sensory neurons in the anterior neural plate border of the ascidian embryo. *Nat. Commun.* **10**, 4911. (doi:10.1038/s41467-019-12839-6)
129. Mazet F, Hutt JA, Milloz J, Millard J, Graham A, Shimeld SM. 2005 Molecular evidence from *Ciona intestinalis* for the evolutionary origin of vertebrate sensory placodes. *Dev. Biol.* **282**, 494–508. (doi:10.1016/j.ydbio.2005.02.021)
130. Wagner E, Stolfi A, Gi Choi Y, Levine M. 2014 Islet is a key determinant of ascidian palp morphogenesis. *Development* **141**, 3084–3092. (doi:10.1242/dev.110684)
131. Kusakabe TG *et al.* 2012 A conserved non-reproductive GnRH system in chordates. *PLoS ONE.* **7**, e41955. (doi:10.1371/journal.pone.0041955)
132. Manni L, Agnoletto A, Zaniolo G, Burighel P. 2005 Stomodeal and neurohypophysial placodes in *Ciona intestinalis*: insights into the origin of the pituitary gland. *J. Exp. Zool. Part B* **304b**, 324–339. (doi:10.1002/jez.b.21039)
133. Bollner T, Holmberg K, Olsson R. 1986 A rostral sensory mechanism in *Oikopleura dioica* (Appendicularia). *Acta Zool.* **67**, 235–241. (doi:10.1111/j.1463-6395.1986.tb00868.x)
134. Bassham S, Postlethwait JH. 2005 The evolutionary history of placodes: a molecular genetic investigation of the larvacean urochordate *Oikopleura dioica*. *Development* **132**, 4259–4272. (doi:10.1242/dev.01973)
135. Madin LP. 1995 Sensory ecology of salps (Tunicata, thaliacea): more questions than answers. *Mar. Fresh Behav. Physiol.* **26**, 175–195. (doi:10.1080/10236249509378938)
136. Pennati R, Dell'Anna A, Zega G, De Bernardi F. 2012 Immunohistochemical study of the nervous system of the tunicate *Thalia democratica* (Forsskal, 1775). *Euro. J. Histochem.* **56**, 96–101. (doi:10.4081/ejh.2012.16)
137. Parker GH. 1908 The sensory reactions of amphioxus. *Proc. Am. Acad. Arts Sci.* **43**, 414–423. (doi:10.2307/20022358)
138. Zieger E, Garbarino G, Robert NSM, Yu JK, Croce JC, Candiani S, Schubert M. 2018 Retinoic acid signaling and neurogenic niche regulation in the developing peripheral nervous system of the cephalochordate amphioxus. *Cell Mol. Life Sci.* **75**, 2407–2429. (doi:10.1007/s00018-017-2734-3)
139. Lu TM, Luo YJ, Yu JK. 2012 BMP and delta/Notch signaling control the development of amphioxus epidermal sensory neurons: insights into the evolution of the peripheral sensory system. *Development* **139**, 2020–2030. (doi:10.1242/dev.073833)
140. Holland LZ. 2005 Non-neural ectoderm is really neural: evolution of developmental patterning mechanisms in the non-neural ectoderm of chordates and the problem of sensory cell homologies. *J. Exp. Zool.* **304B**, 304–323. (doi:10.1002/jez.b.21038)
141. Lacalli TC. 2004 Sensory systems in amphioxus: a window on the ancestral chordate condition. *Brain Behav. Evol.* **64**, 148–162. (doi:10.1159/000079744)
142. Baatrup E. 1981 Primary sensory cells in the skin of amphioxus (*Branchiostorna lanceolatum* (P)). *Acta Zool.* **62**, 147–157. (doi:10.1111/j.1463-6395.1981.tb00624.x)
143. Lacalli TC, Hou S. 1999 A reexamination of the epithelial sensory cells of amphioxus (*Branchiostoma*). *Acta Zool.* **80**, 125–134. (doi:10.1046/j.1463-6395.1999.80220005.x)
144. Stokes MD, Holland ND. 1995 Embryos and larvae of a lancelet, *Branchiostoma floridae*, from hatching through metamorphosis: growth in the laboratory and external morphology. *Acta Zool.* **76**, 105–120. (doi:10.1111/j.1463-6395.1995.tb00986.x)
145. Bone Q, Best ACG. 1978 Ciliated sensory cells in amphioxus (*Branchiostoma*). *J. Mar. Biol. Assoc. United Kingdom* **58**, 479–486. (doi:10.1017/S0025315400028137)
146. Candiani S, Oliveri D, Parodi M, Bertini E, Pestarino M. 2006 Expression of *AmphiPOU-IV* in the developing neural tube and epidermal sensory neural precursors in amphioxus supports a conserved role of class IV POU genes in the sensory cells development. *Dev. Genes Evol.* **216**, 623–633. (doi:10.1007/s00427-006-0083-6)
147. Jackman WR, Langeland JA. 2000 Kimmel CB. *islet* reveals segmentation in the Amphioxus hindbrain homolog. *Dev. Biol.* **220**, 16–26. (doi:10.1006/dbio.2000.9630)
148. Kozmik Z *et al.* 2007 *Pax-Six-Eya-Dach* network during amphioxus development: conservation *in vitro* but context specificity *in vivo*. *Dev. Biol.* **306**, 143–159. (doi:10.1016/j.ydbio.2007.03.009)
149. Mazet F, Masood S, Luke GN, Holland ND, Shimeld SM. 2004 Expression of *AmphiCoe*, an amphioxus COE/EBF gene, in the developing central nervous system and epidermal sensory neurons. *Genesis* **38**, 58–65. (doi:10.1002/gene.20006)
150. Kim K, Colosimo ME, Yeung H, Sengupta P. 2005 The UNC-3 Olf/EBF protein represses alternate neuronal programs to specify chemosensory neuron identity. *Dev. Biol.* **286**, 136–148. (doi:10.1016/j.ydbio.2005.07.024)
151. Holland LZ, Schubert M, Holland ND, Neuman T. 2000 Evolutionary conservation of the presumptive neural plate markers *AmphiSox1/2/3* and *AmphiNeurogenin* in the invertebrate chordate amphioxus. *Dev. Biol.* **226**, 18–33. (doi:10.1006/dbio.2000.9810)
152. Churcher AM, Taylor JS. 2011 The antiquity of chordate odorant receptors is revealed by the discovery of orthologs in the cnidarian *Nematostella vectensis*. *Genome Bio.l Evol.* **3**, 36–43. (doi:10.1093/gbe/evq079)
153. Niimura Y. 2009 On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biol. Evol.* **1**, 34–44. (doi:10.1093/gbe/evp003)
154. Satoh G. 2005 Characterization of novel GPCR gene coding locus in amphioxus genome: gene structure, expression, and phylogenetic analysis with implications for its involvement in chemoreception. *Genesis* **41**, 47–57. (doi:10.1002/gene.20082)
155. Baatrup E. 1982 On the structure of the corpuscles of de Quatrefages (*Branchiostorna lanceolatum* (P)). *Acta Zool.* **63**, 39–44. (doi:10.1111/j.1463-6395.1982.tb00757.x)
156. de Quatrefages MA. 1845 Sur le systeme nerveux et sur l'histologie du Branchiostome ou amphioxus. *Anal. Sci. Nat. Zool.* **4**, 197–248.
157. Baker CV, Bronner-Fraser M. 1997 The origins of the neural crest. Part II: an evolutionary perspective. *Mech. Dev.* **69**, 13–29. (doi:10.1016/S0925-4773(97)00129-9)
158. Boorman C, Shimeld SM. 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. (doi:10.1046/j.1525-142X.2002.02021.x)
159. Lacalli TC, Gilmour THJ, Kelly SJ. 1999 The oral nerve plexus in amphioxus larvae: function, cell types and phylogenetic significance. *Proc. R. Soc. Lond. B* **266**, 1461–1470. (doi:10.1098/rspb.1999.0801)
160. Hatschek B. 1881 Studien über die Entwickelung des Amphioxus. *Arb. Zool. InstWien.* **4**, 1–88.
161. Nozaki M, Gorbman A. 1992 The question of functional homology of Hatschek's pit of amphioxus (*Branchiostoma helcheri*) and the vertebrate Adenohypophysis. *Zool. Sci.* **9**, 387–395.
162. Yongqiang F, Weiquan H, Lei C, Jinshan Z. 1999 Distribution of gonadotropin-releasing hormone in the brain and Hatschek's pit of amphioxus (*Branchiostoma belcheri*). *Acta Zool. Sinica.* **45**, 106–111.
163. Chang CY, Liu YX, Zhu YT, Zhu HH. 1985 The reproductive endocrinology of Amphioxus. In *Frontiers in physiological research* (eds DG Carlick, PI Korner), pp. 70–86. Canberra, Australia: Autstralian Academy of Science.
164. Roch GJ, Tello JA, Sherwood NM. 2014 At the transition from invertebrates to vertebrates, a novel GnRH-like peptide emerges in amphioxus. *Mol. Biol. Evol.* **31**, 765–778. (doi:10.1093/molbev/mst269)
165. Fabian P, Tseng KC, Smeeton J, Lancman JJ, Dong PDS, Cerny R, Crump JG. 2020 Lineage analysis reveals an endodermal contribution to the vertebrate pituitary. *Science* **370**, 463–467. (doi:10.1126/science.aba4767)

Dev Genes Evol (2017) 227:319–338
DOI 10.1007/s00427-017-0591-6

ORIGINAL ARTICLE

The structure, splicing, synteny and expression of lamprey *COE* genes and the evolution of the *COE* gene family in chordates

Ricardo Lara-Ramírez[1,2] · **Guillaume Poncelet**[1] · **Cédric Patthey**[1,3] · **Sebastian M. Shimeld**[1]

Received: 13 September 2016 / Accepted: 20 August 2017 / Published online: 5 September 2017

Abstract *COE* genes encode transcription factors that have been found in all metazoans examined to date. They possess a distinctive domain structure that includes a DNA-binding domain (DBD), an IPT/TIG domain and a helix-loop-helix (HLH) domain. An intriguing feature of the *COE* HLH domain is that in jawed vertebrates it is composed of three helices, compared to two in invertebrates. We report the isolation and expression of two *COE* genes from the brook lamprey *Lampetra planeri* and compare these to *COE* genes from the lampreys *Lethenteron japonicum* and *Petromyzon marinus*. Molecular phylogenetic analyses do not resolve the relationship of lamprey *COE* genes to jawed vertebrate paralogues, though synteny mapping shows that they all derive from duplication of a common ancestral genomic region. All lamprey genes encode conserved DBD, IPT/TIG and HLH domains; however, the HLH domain of lamprey *COE-A* genes encodes only two helices while *COE-B* encodes three helices. We also identified *COE-B* splice variants encoding either two or three helices in the HLH domain, along with other *COE-A* and *COE-B* splice variants affecting the DBD and C-terminal transactivation regions. In situ hybridisation revealed expression in the lamprey nervous system including the brain, spinal cord and cranial sensory ganglia. We also detected expression of both genes in mesenchyme in the pharyngeal arches and underlying the notochord. This allows us to establish the primitive vertebrate expression pattern for *COE* genes and compare this to that of invertebrate chordates and other animals to develop a model for *COE* gene evolution in chordates.

Keywords COE · Ebf · Neurogenesis · Lamprey · Cranial ganglia · Pharyngeal arch · Brain

Communicated by Karen E. Sears

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00427-017-0591-6) contains supplementary material, which is available to authorized users.

✉ Sebastian M. Shimeld
sebastian.shimeld@zoo.ox.ac.uk

[1] Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

[2] Present address: Centro de Investigación en Ciencias Biológicas Aplicadas, Instituto Literario No. 100, Colonia Centro, CP 50000 Toluca, México

[3] Present address: Umeå Center for Molecular Medicine, Umeå University, Umeå, Sweden

Introduction

Lampreys are jawless vertebrates. Together with hagfishes they form the cyclostomes, a lineage that separated early in vertebrate evolution from the lineage that gave rise to the jawed vertebrates (gnathostomes). While lampreys have core vertebrate features such as a dorsal tubular nervous system, neural crest cells, placode-derived sensory ganglia and a cranial and axial skeleton, they lack gnathostome characters such as hinged jaws and paired appendages. As such, they have been an important model system for understanding the early morphological evolution of vertebrates (Shimeld and Donoghue 2012). The expression of genes involved in developmental processes has been a critical piece of evidence in such studies, allowing insight into the evolution of new characters.

The placement of lampreys relative to the two rounds of genome duplication (2R) thought to have occurred early in vertebrate evolution (Putnam et al. 2008) is also important for understanding vertebrate character evolution. Jawed vertebrate genomes are characterised by large paralogous blocks

of genes deriving from duplications, each traceable to a small number of ancestral linkage groups in a reconstructed pre-duplication ancestor (Nakatani et al. 2007). Consequently, many gene families comprise multiple paralogous genes in jawed vertebrates but a single copy gene in the vertebrates' nearest relatives, amphioxus and urochordates (Putnam et al. 2008). However, it is currently unclear whether lampreys share the 2R duplications. While lamprey genomes have clearly undergone genome duplication, such that multiple paralogues of many gene families are found and leading some authors to suggest both genome duplications are shared (Smith et al. 2013), in molecular phylogenetic analyses, lamprey genes rarely group robustly with jawed vertebrate paralogue groups (Kuraku et al. 2009). This raises the possibility that lampreys and jawed vertebrates might have undergone parallel genome duplication, or, as a recent study based on an improved genome map suggests, there may have been only one genome duplication coupled with a number of segmental duplications (Smith and Keinath 2015). Furthermore, some or all of these duplications may have occurred at a similar time to lineage separation, such that gene relationships are obscured and/or the return from tetraploidy to diploidy happened independently in the two lineages (Furlong and Holland 2002). These uncertainties mean evolutionary comparisons involving lamprey genes need to be made at the level of gene families rather than individual orthologues.

The *COE* genes (also known as *Ebf* genes) are a family of HLH transcription factors that are involved in many developmental aspects of the vertebrate nervous system. Their restricted expression patterns reflect important aspects of the structure of the nervous system in different vertebrate species. For example, in mice, *COE* genes highlight the regionalisation of the brain and spinal cord, particularly marking post-mitotic neurons. In the spinal cord, they are expressed ventrolaterally in a region corresponding to motor neurons (Garel et al. 1997). In the PNS, mouse *COE* genes are expressed in the olfactory epithelium, vomeronasal organ, trigeminal and glossopharyngeal cranial ganglia, inner ear, dorsal root ganglia (DRG) and peripheral glia (Corradi et al. 2003; Malgaretti et al. 1997; Wang et al. 1997). Also, in keeping with their expression sites in the nervous system, COE proteins have been shown to induce neuronal differentiation and associated cell cycle exit (Garcia-Dominguez et al. 2003). Altogether, the expression of *COE* genes aids both the identification of areas of neuronal differentiation and the assessment of neuronal cell types in the nervous system of species with less well-described anatomy and developmental processes.

Apart from the nervous system, vertebrate *COE* genes are also expressed in mesodermal structures during development and in mesodermal derivatives at late developmental stages and in adulthood (Hesslein et al. 2009; Jimenez et al. 2007; Kieslinger et al. 2005). *COE* genes are expressed in somites (El-Magd et al. 2014a; El-Magd et al. 2014b; Green and Vetter 2011) and have an important role in the commitment and differentiation of muscle cells (Green and Vetter 2011; Jin et al. 2014). *COE* genes are also expressed in the mesodermal and neural crest components of pharyngeal arches at embryonic stages (Dubois et al. 1998; El-Magd et al. 2014b; Jin et al. 2014; Kieslinger et al. 2005; Pozzoli et al. 2001). In the lymphocyte lineage, *Ebf1*/*COE1* participates in the specification of B-cell lymphocytes (Hagman et al. 1995; Lin and Grosschedl 1995; Treiber et al. 2010). Also, *COE* genes seem to be involved in the regulation of adipocyte and osteoblast commitment and differentiation (Akerblad et al. 2002; Akerblad et al. 2005; Hesslein et al. 2009; Kieslinger et al. 2005) and in the specification of brown versus white adipocyte identity (Rajakumari et al. 2013).

While *COE* genes encode an HLH domain with a typical helix1 (H1)-linker (L)-helix2 (H2) structure, they lack the DNA-binding basic region typical of most HLH proteins. Instead, they possess a large DNA-binding domain (DBD) which includes an atypical Zn-finger (Fields et al. 2008; Hagman et al. 1995). Between the DBD and HLH domains, an IPT/TIG (immunoglobulin-like, plexins, transcription factors/transcription factor immunoglobulin) domain is also present and has been suggested to be involved in protein-protein interactions, dimerization and even DNA binding (Siponen et al. 2010; Treiber et al. 2010). In addition, COE proteins have a transactivation (TA) domain at the carboxy terminus (Hagman et al. 1995). *COE* genes have been found in a wide variety of metazoans (Crozatier and Vincent 1999; Demilly et al. 2011; Jackson et al. 2010; Mazet et al. 2004; Pang et al. 2004), with presence in ctenophores and sponges showing they date from at least the last common ancestor of animals (Daburon et al. 2008; Jackson et al. 2010; Pang et al. 2004). Invertebrates, including cephalochordates and tunicates, generally have a single *COE* gene (Dubois and Vincent 2001; Jackson et al. 2010; Mazet et al. 2004; Pang et al. 2004).

Invertebrate and jawed vertebrate *COE* genes have another key difference: all invertebrate *COE* genes analysed to date have a typical H1-L-H2 organisation in the HLH domain, while all jawed vertebrate *COE* genes encode a duplicated H2 (H2d) such that the organisation of their HLH domain is H1-L-H2d-H2a (Crozatier et al. 1996; Dubois and Vincent 2001; Mazet et al. 2004; Pang et al. 2004). Daburon and colleagues characterised two *COE* genes, which they named *COE-A* and *COE-B*, from the lamprey *Petromyzon marinus* (Daburon et al. 2008). *COE-A* appeared to lack the duplicated H2 (H2d), possessing only one H2 (H2a), though this was not definitive due to poor quality genome data in this region. *COE-B*, however, had the duplicated H2. Furthermore, expressed sequence tag (EST) data from a second lamprey species, *Lampetra fluviatilis*, showed *COE-B* transcripts are alternately spliced, such that transcripts could have the structure H1-L-H2d-H2a (as in jawed vertebrates) or H1-L-H2a (as

in invertebrates). These studies present an intriguing picture of *COE* family evolution in chordates but leave unanswered questions; for example, it is unclear when the H2 duplication occurred, relative to the timing of genome duplications, and how this relates to gene orthology.

Even though significant progress has been made in the understanding of *COE* gene expression and function in jawed vertebrates, particularly in mice, and new information of *COE* gene expression and function in several metazoans is emerging (Jackson et al. 2010; Pang et al. 2004), there is still a gap in understanding of *COE* gene expression and function at the invertebrate-vertebrate transition. *COE* gene expression in the cephalochordate *Branchiostoma floridae* has been described in detail and shown to be expressed in scattered cells throughout the brain and nerve cord, as well as in the ventral part of the anterior somites and in scattered ectodermal cells presumed by these authors and others to be peripheral epidermal sensory neurons (Mazet et al. 2004; Schubert et al. 2004). In urochordates, *COE* expression in the ascidian *Ciona intestinalis* is first detected at the gastrula stage in the A9.32 cell pair (Imai et al. 2004), which gives rise to central nervous system cells. By the neurula and tailbud stages, *COE* expression is more widespread in the central nervous system, as well as in palp cells that may be sensory neurons (Mazet et al. 2005). *C. intestinalis COE* is also expressed in mesodermal cells, specifically in the pharyngeal muscle lineage (Razy-Krajka et al. 2014). When compared to studies of other invertebrate taxa, these data support a primitive role for *COE* genes in neural differentiation and a possible ancient role for *COE* genes in mesodermal cells (Jackson et al. 2010; Pang et al. 2004). However, little is known about *COE* gene expression and function in cyclostomes, which can bridge the gap between what is known in invertebrates and jawed vertebrates and clarify important aspects of expression patterns and gene evolution within the *COE* family.

To gain more insight into *COE* gene evolution and expression at the invertebrate-vertebrate transition, we identified and studied two lamprey *COE* genes in the brook lamprey *Lampetra planeri*. We address the presence/absence of the duplicated H2 in lamprey *COE* genes (Daburon et al. 2008) and their relation to vertebrate paralogy groups. To accomplish this, we used genome data from *P. marinus* and *Lethenteron japonicum* (also known as *Lethenteron camtschaticum*) (Mehta et al. 2013; Smith et al. 2013), as well as transcript data from *L. planeri*, and show that lamprey *COE-B* has the duplicated H2 in all these species, while *COE-A* lacks a duplicated H2. We further explored alternative splicing of both *COE-A* and *COE-B* in *L. planeri* by RT-PCR and transcriptomics, showing alternate splicing of both genes. We also expand previous molecular phylogenetic analyses and couple this with synteny comparisons to develop evolutionary models of *COE* loci in chordates. These analyses show all four gnathostome *COE* loci, and both lamprey *COE* loci share syntenic characters with each other and with the amphioxus *COE* locus. In addition, we describe the expression patterns of both *L. planeri COE* genes during embryogenesis by in situ hybridisation and show widespread expression in central and peripheral nervous systems, as well as in pharyngeal mesenchyme and other mesodermal populations. Our data suggest conservation of *COE* gene function in neuronal differentiation and pharyngeal development in all vertebrate lineages, but highlight differences within the vertebrate lineage such as the absence of expression in dorsal root ganglia and muscle derivatives during lamprey embryonic development with respect to jawed vertebrates.

Materials and methods

Embryo collection, fixation, in situ hybridisation and gene cloning

We extracted total RNA from stage 24–26 *L. planeri* embryos. For PCR, we based our primer design on *P. marinus* genomic information from the Ensembl website. For *LpCOE-A*, we used sense primer 5′-CTAGCGCGGGCCCACTTCGAGAA-3′ and antisense primer 5′-TGGGAGGCCTCGGACACGCTGAT-3′. For *LpCOE-B*, we used sense primer 5′-GAGG GCACACTTTGAGAAGCAGCCA-3′ and antisense primer 5′-GCCGGGCTCCGAAACGCTCAC-3′. Cloned sequences have been deposited in GenBank, with accession numbers MF539934 (COE-A) and MF539935 (COE-B). *L. planeri* embryos were collected from a shallow river in the New Forest National Park, United Kingdom, with permission from the Forestry Commission. Embryos were brought to the laboratory and placed in Petri dishes with filtered river water from the same river where they were caught. They were kept at 13–15 °C and fixed at different stages of development following the staging system of Tahara (1988). All experiments were performed under local ethical approval. When necessary, embryo chorions were removed with fine forceps before fixation. Embryos were fixed in 4% paraformaldehyde (PFA) pH 7.5 in phosphate-buffered saline (PBS). PFA was cooled on ice before use. Embryos were fixed in an approximately 10× excess volume of 4% PFA-PBS with respect to river water at 4 °C overnight or longer. After fixation, embryos were washed twice in diethyl pyrocarbonate (DEPC)-treated 1× PBS for 10 min each and then dehydrated through a graded series of PBS:methanol (25, 50 and 75% of methanol in 1× PBS) once for 10 min each. Finally, they were washed twice in 100% methanol for 10 min each and stored in fresh methanol at −20 °C. In situ hybridisation experiments and histology were carried out as previously described (Lara-Ramirez et al. 2015).

Molecular phylogenetic analysis

Sequence analysis and manipulation were performed using BioEdit (Hall 1999). Accession numbers for sequences used

for phylogenetic analysis are shown in Fig. 1. We selected sequences from representatives of major vertebrate groups (mammal *Homo sapiens*; lepidosaur *Anolis carolinensis*; archosaur *Gallus gallus*; amphibian *Xenopus tropicalis*; sarcopterygian fish *Latimeria chalumnae*; actinopterygian *Lepisosteus oculatus*; teleost *Danio rerio*; elasmobranch: *Callorhincus milii*, *Raja eglantaria*) plus selected invertebrate outgroups. Multiple sequence alignments were carried out using MAFFT software version 6.864b (Katoh et al. 2002; Katoh and Toh 2008). The parameter for strategy for MAFFT alignments was set as "auto". All other parameters were as the defaults. This alignment was trimmed using GBLOCKS to regions included in all sequences, and a third alignment which was further adjusted by eye, removing both helix 2s from all sequences because of uncertainty over alignment in this region. For phylogenetic tree construction, the maximum likelihood (ML) and Bayesian algorithms were used, and tree construction was conducted using MEGA version 5.2 (Tamura et al. 2011) for ML, and MrBayes version 3.2 (Ronquist and Huelsenbeck 2003) for Bayesian analyses. For ML, we used the Whelan and Goldman (WAG) amino acid substitution matrix (Whelan and Goldman 2001) and 100 bootstrap replicates to obtain support values at each node. Bayesian inference was performed using the Markov chain Monte Carlo method. Two independent Markov chains were run, each with 1 million iterations with default heating parameters. The first 25% of the trees were discarded as burn-in before compiling consensus trees and summary statistics. Posterior probabilities at each internal branch were taken as a measure of statistical support. Both methods produced the same tree topology at key nodes, though with differing support values. Initial analyses included the COE sequence from *C. intestinalis*, and this was consistently placed outside the vertebrate genes as previously reported (Daburon et al. 2008). However, in our analyses, inclusion of this sequence reduced resolution within the vertebrate genes, so we removed it for subsequent analyses.

Fig. 1 (**a**) Molecular phylogenetic analysis of deuterostome *COE* sequences. The tree shown was constructed by Bayesian inference. Shown above key nodes are posterior probabilities from this analysis, while numbers below these nodes are percentage bootstrap support for the same node derived from Maximum Likelihood analysis. Gnathostome and cyclostome *COE* orthology groups are boxed. Support for orthology of *HsCOE4* with *CmCOE4* and *LoCOE4* is weak; however, *CmCOE4* and *LoCOE4* are well-supported as orthologues, and synteny supports orthology of *HsCOE4* and *LoCOE4* (Fig. S3). Species abbreviations: Ac, *Anolis carolinensis*; Bf, *Branchiostoma floridae*; Cm, *Callorhincus milii*; Dr, *Danio rerio*; Eb, *Eptatretus burgeri*; Gg, *Gallus gallus*; Hs, *Homo sapiens*; Lc, *Latimeria chalumnae*; Lj, *Lethenteron japonicum*; Lo, *Lepisosteus oculatus.* Lp, *Lampetra planeri*; Pm, *Petromyzon marinus*; Re, *Raja eglanteria*; Sk, *Saccoglossus kowalevski*; Sp, *Strongylocentrotus purpuratus*; Xt, *Xenopus tropicalis*. (**b**) Schematic maps of *COE* locus paralogy and synteny in jawed vertebrate, lamprey and amphioxus genomes. *COE* genes are in red, colour coding of other genes is as follows: Grey, genes with no orthologues or paralogues identified in the analysed regions. White, genes with syntenic orthologues and/or paralogues in human and shark. Black, genes linked to *COE* in amphioxus and their orthologue positions in other species. Other colours, genes linked to lamprey *COE* genes and their orthologues and/or paralogues in other species. Discontinuities shown as angled bars indicate where genes map to the same chromosome arm or scaffold but are separated by multiple intervening genes which are not shown. While the current amphioxus (*B. floridae*) genome assembly has tandem *COE* gene models, sequence comparisons (not shown) indicate these are derived from a single gene as depicted. Additional synteny data for *L. oculatus* are in Fig. S3

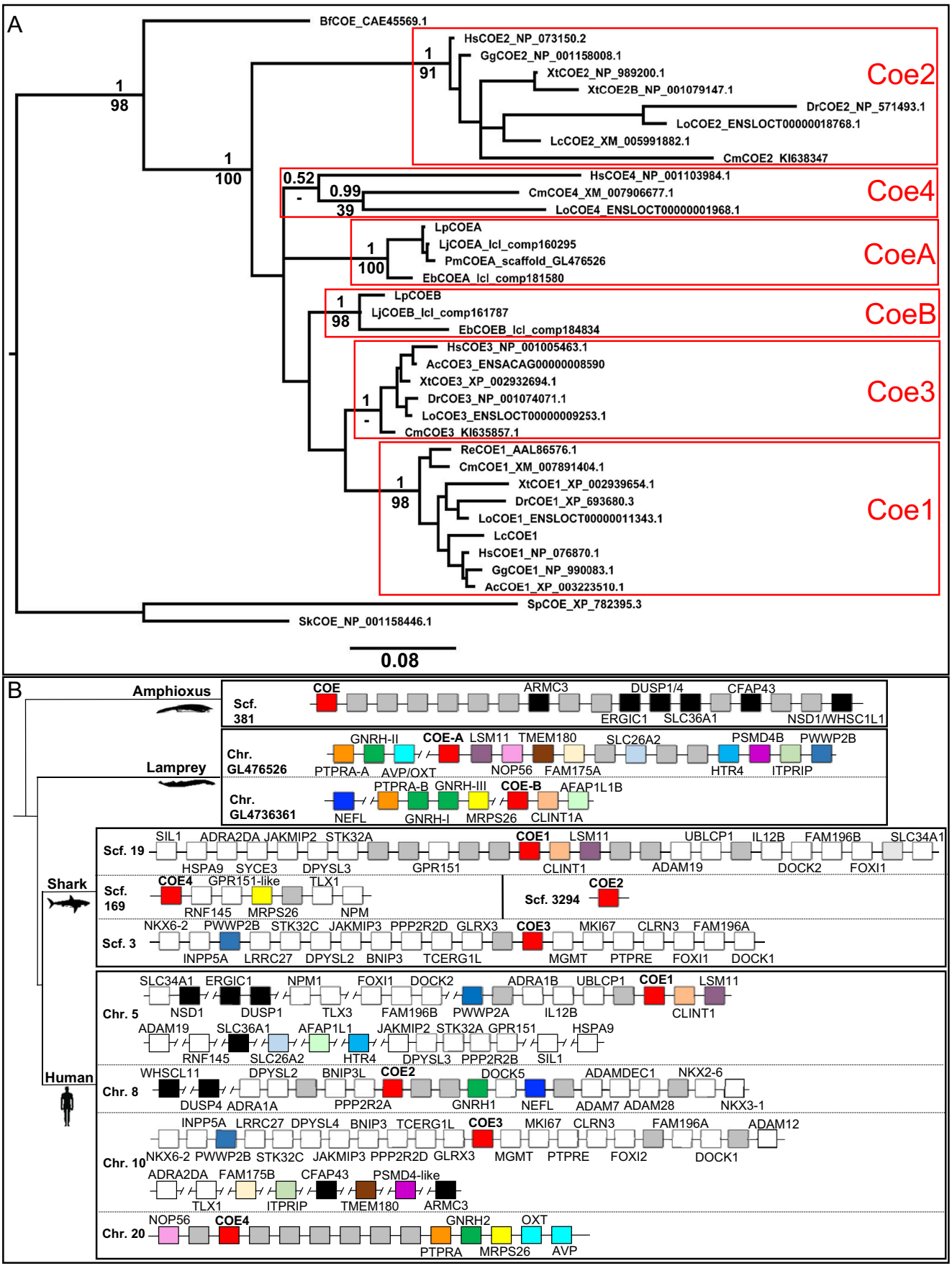
Intron-exon organisation and synteny analysis

COE genomic loci were identified in the genomes of *B. floridae* (amphioxus), *L. japonicum*, *P. marinus*, *C. milii* (elephant shark) and *H. sapiens* by BLAST. In addition, we compared *COE* loci in *H. sapiens* and *L. oculatus* using Genomicus v89.01 (Louis et al. 2015). We chose *H. sapiens* as an extensively annotated vertebrate genome and *C. milii* as a member of the earliest diverging jawed vertebrate lineage; together, they encompass extant jawed vertebrate diversity. Intron-exon structures were extracted from gene models or determined by comparing transcripts to the genomic sequence. To map synteny and paralogy relationships, genes adjacent to the *COE* loci in each species were searched by TBLASTN of their predicted proteins across the other genomes. The top chromosomal or scaffold hits were recorded to predict orthology. Human paralogues were extracted from Ensembl predictions of paralogy.

Transcriptome analysis and RT-PCR

Total embryo RNA from *L. planeri* embryo stages 25, 26 and 27 (Tahara 1988) was pooled and sequenced by Illumina HiSeq. Raw data have been deposited in the SRA (Bioproject PRJNA371458). Reads were assembled using Trinity (Grabherr et al. 2011) to persevere candidate splice variants and putative *COE* transcripts extracted from the assembly. We also remapped the raw transcriptome data back onto the *COE* gene models to examine read distribution across the splice variants; briefly, all reads mapping to each gene were extracted using BLASTN with default parameters. A kmer search strategy was used to identify individual fragments (derived from the paired-end reads) that crossed exon boundaries, taking into account single nucleotide variations. All fragments matching to the 12 last or 12 first nucleotides of each exon were examined to see whether they corresponded to the canonical splice form or to potential alternate splice forms. For each putative exon-exon junction, fragments containing a predicted sequence of 7 nucleotides of the upstream exon followed by 20 nucleotides of the downstream exon, or the converse, were counted as occurrences of the splice junction.

Splice variants predicted from the transcriptome data and represented by >1 fragment were verified by RT-PCR on *L. planeri* RNA deriving from staged lamprey embryos. Primers were


A
BfCOE_CAE45569.1
HsCOE2_NP_073150.2
GgCOE2_NP_001158008.1
XtCOE2_NP_989200.1
XtCOE2B_NP_001079147.1
DrCOE2_NP_571493.1
LoCOE2_ENSLOCT00000018768.1
LcCOE2_XM_005991882.1
CmCOE2_KI638347
Coe2
HsCOE4_NP_001103984.1
CmCOE4_XM_007906677.1
LoCOE4_ENSLOCT00000001968.1
Coe4
LpCOEA
LjCOEA_lcl_comp160295
PmCOEA_scaffold_GL476526
EbCOEA_lcl_comp181580
CoeA
LpCOEB
LjCOEB_lcl_comp161787
EbCOEB_lcl_comp184834
CoeB
HsCOE3_NP_001005463.1
AcCOE3_ENSACAG00000008590
XtCOE3_XP_002932694.1
DrCOE3_NP_001074071.1
LoCOE3_ENSLOCT00000009253.1
CmCOE3_KI635857.1
Coe3
ReCOE1_AAL86576.1
CmCOE1_XM_007891404.1
XtCOE1_XP_002939654.1
DrCOE1_XP_693680.3
LoCOE1_ENSLOCT00000011343.1
LcCOE1
HsCOE1_NP_076870.1
GgCOE1_NP_990083.1
AcCOE1_XP_003223510.1
Coe1
SpCOE_XP_782395.3
SkCOE_NP_001158446.1
0.08
B
Amphioxus
Lamprey
Shark
Human
Scf. 381
COE
ARMC3
DUSP1/4
CFAP43
ERGIC1
SLC36A1
NSD1/WHSC1L1
Chr. GL476526
GNRH-II
COE-A
LSM11
TMEM180
SLC26A2
PSMD4B
PWWP2B
PTPRA-A
AVP/OXT
NOP56
FAM175A
HTR4
ITPRIP
Chr. GL4736361
PTPRA-B
GNRH-III
COE-B
AFAP1L1B
NEFL
GNRH-I
MRPS26
CLINT1A
Scf. 19
SIL1
ADRA2DA
JAKMIP2
STK32A
COE1
LSM11
UBLCP1
IL12B
FAM196B
SLC34A1
HSPA9
SYCE3
DPYSL3
GPR151
CLINT1
ADAM19
DOCK2
FOXI1
Scf. 169
COE4
GPR151-like
TLX1
RNF145
MRPS26
NPM
Scf. 3294
COE2
Scf. 3
NKX6-2
PWWP2B
STK32C
JAKMIP3
PPP2R2D
GLRX3
COE3
MKI67
CLRN3
FAM196A
INPP5A
LRRC27
DPYSL2
BNIP3
TCERG1L
MGMT
PTPRE
FOXI1
DOCK1
Chr. 5
SLC34A1
ERGIC1
NPM1
FOXI1
DOCK2
ADRA1B
UBLCP1
COE1
LSM11
NSD1
DUSP1
TLX3
FAM196B
PWWP2A
IL12B
CLINT1
ADAM19
SLC36A1
AFAP1L1
JAKMIP2
STK32A
GPR151
HSPA9
RNF145
SLC26A2
HTR4
DPYSL3
PPP2R2B
SIL1
Chr. 8
WHSCL11
DPYSL2
BNIP3L
COE2
DOCK5
ADAMDEC1
NKX2-6
DUSP4
ADRA1A
PPP2R2A
GNRH1
NEFL
ADAM7
ADAM28
NKX3-1
Chr. 10
INPP5A
LRRC27
DPYSL4
BNIP3
TCERG1L
COE3
MKI67
CLRN3
FAM196A
ADAM12
NKX6-2
PWWP2B
STK32C
JAKMIP3
PPP2R2D
GLRX3
MGMT
PTPRE
FOXI2
DOCK1
ADRA2DA
FAM175B
CFAP43
PSMD4-like
TLX1
ITPRIP
TMEM180
ARMC3
Chr. 20
NOP56
COE4
GNRH2
OXT
PTPRA
MRPS26
AVP

designed to span introns to exclude the possibility of genomic DNA contamination confounding the results. Bands were cloned and sequenced, confirming all the splice variants. Primer sequences used in the RT-PCR are as follows: COEA TIG 5′; AATAACTCCAAGCACGGGCG. COEA TIG 3′; CTGATGGCTTTGATGCACGG. COEB ZNF 5′; AGAATCCGGAGATGTGTCGG. COEB ZNF 3′; CGATGGGGTCTCGTTTCTGT. COEB TIG 5′; GACAACTTCTTCGACGGGCT. COEB TIG 3′; AGGGTGACCTCCACCACTC. COEB TA1 5′; GTGAGCGTTTCGGAGCCTG. COEB TA1 3′; GGGACACGCTGCTCGTATT. COEB Helix1 5′; TCTGAACGAGCCCACCATTGACTAC. COEB Helix2a 3′; TGGTGGCGGGGCATGCTGTACAGAGCT.

Results

Molecular phylogenetic and synteny analysis of lamprey *COE-A* and *COE-B*

cDNA fragments encoding *LpCOE-A* and *LpCOE-B* were initially isolated from stages 24–26 *L. planeri* embryo RNA using primers based on *COE* sequences identified in the *P. marinus* genome. To extend these sequences, we generated and searched an *L. planeri* transcriptome assembly, allowing us to deduce the whole open reading frame for each gene. Alignment of their predicted amino acid sequences with mouse COE proteins COE1–4 revealed high sequence similarity in the DBD, IPT/TIG and HLH domains (Supplementary file 1). This process also identified splice variants, which are described further in the following texts.

We also analysed *COE* sequences from other cyclostomes (Supplementary file 1): specifically *P. marinus COE-A* (Daburon et al. 2008), sequences from another lamprey species (*L. japonicum* (Mehta et al. 2013)) and sequences from a transcriptome from the hagfish *Eptatretus burgheri* (kindly provided by Juan Pascal Anaya). Molecular phylogenetic analysis of these cyclostome sequences, including jawed vertebrate COE sequences plus invertebrate deuterostome COE sequences as outgroups, placed the cyclostome genes with *COE* genes from jawed vertebrates, though failed to clarify relationships to jawed vertebrate *COE1–4* (Fig. 1a). The cyclostome sequences fell into two clades with reasonable support (Fig. 1a), which we call *COE-A* and *COE-B* following previous nomenclature (Daburon et al. 2008).

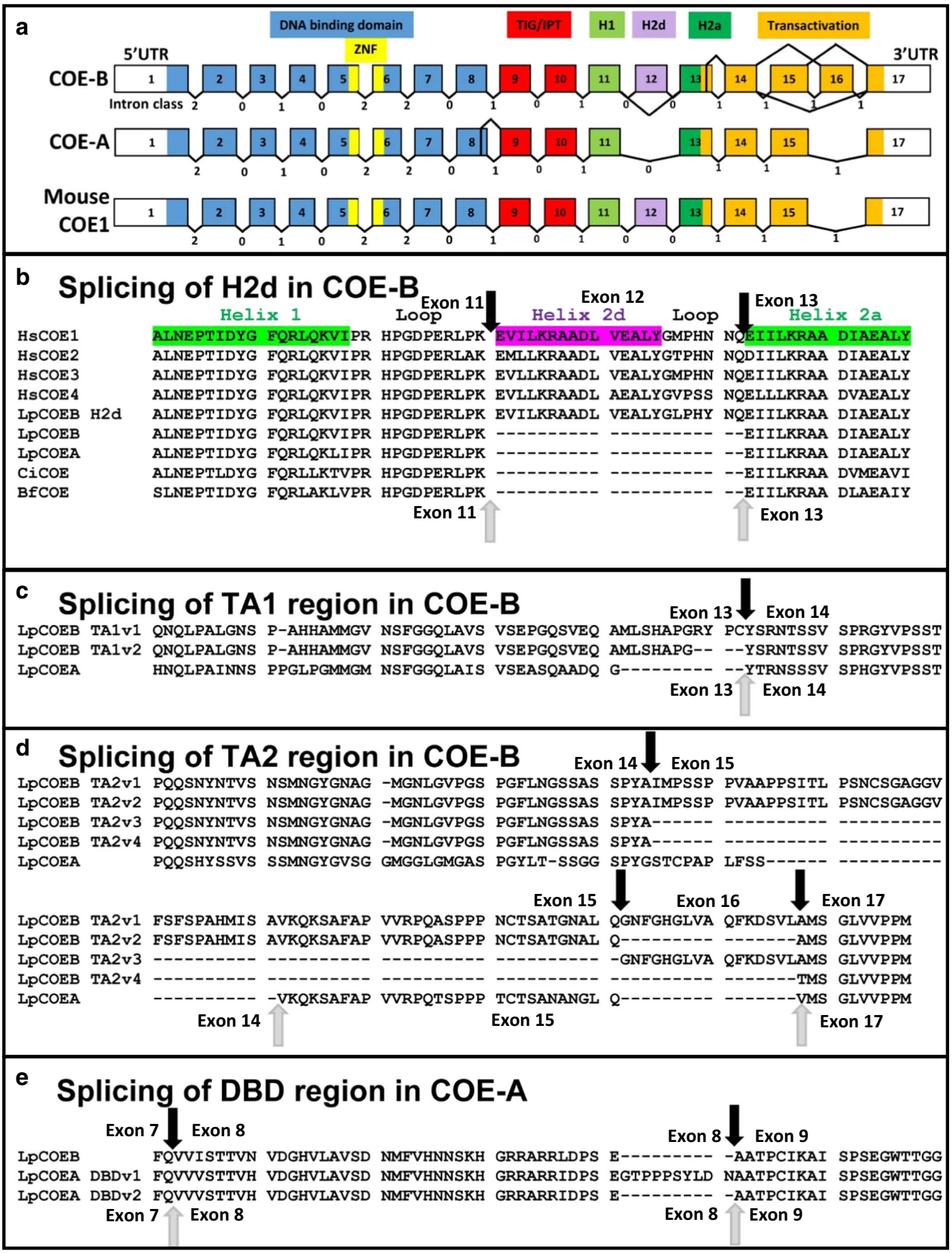
To further examine the relationships between lamprey and other chordate *COE* loci, we examined the synteny surrounding *COE* loci of human, spotted gar, elephant shark, lamprey and amphioxus. Jawed vertebrate *COE1–4* lie in paralogous regions of the genome (Fig. 1b, S3), with sufficient similarity in neighbouring genes to conclude the loci evolved by block duplication. For example, human *COE2* and *COE4* are linked to *GNRH* paralogues, while *COE1* and *COE3* are linked to *FoxI* paralogues. Moreover, *COE1–4* were predicted to form a family of whole genome duplication paralogues in a pre-computed whole genome assessment based on synteny (Singh et al. 2015). We found weak evidence for syntenic organisation of these regions with the amphioxus *COE* locus. Genes linked to amphioxus *COE* on scaffold 381 had orthologues on the same chromosome as human *COE* loci, though the genomic distance was relatively large (Fig. 1b). Both lamprey *COE* loci also showed evidence of similarity in organisation to jawed vertebrate *COE* loci. For example, both are linked to *GNRH* paralogues, lamprey *COE-B* and elephant shark *COE1* are linked to *CLNT1A*, and lamprey *COE-A* and human *COE4* are linked to *NOP56*. However, there is no clear one-to-one relationship between the lamprey loci and the four jawed vertebrate loci such that orthology can be deduced (Fig. 1b). We hence conclude that both jawed vertebrate and lamprey *COE* regions evolved by block duplications from a single ancestral locus as seen in amphioxus, but we cannot determine whether these are shared duplications or occurred in parallel.

Fig. 2 Alternate splicing of lamprey *COE* genes. (**a**) Schematic intron-exon maps of *COE-A* and *COE-B*, generated by mapping *L. planeri* sequence to the *L. japonicum* genome. The various protein domains are colour coded, and alternate splicing verified by transcriptome and RT-PCR is indicated. These are shown in further detail in panels d–e below. Intron phase class is also shown, and the structure of Mouse *COE1*, as previously described (Daburon et al. 2008), is shown for comparison. Exons are numbered sequentially in *COE-B*, with numbering preserved in *COE-A* and mouse *COE1* to indicate exon homology. An expansion of this with accompanying RT-PCR data is in Fig. S2. (**b**) Splicing site in the HLH domain. *COE-B* shows alternate splicing of the H2d exon. Black arrows at the top indicate intron sites in *COE-B*, and grey arrows at the bottom indicate intron sites in COE-A. (**c**) Splicing within the TA1 region of *LpCOE-B*. Variant 2 removes 4 amino acids from the predicted ORF adjacent to the exon 13-exon 14 boundary. (**d**) Splicing within the TA2 region of *LpCOE-B*. Four splice variants were identified. Comparison to intron positions illustrates this occurs by skipping of exon 15, exon 16, or both. *LpCOE-A* is shown for comparison and has a relatively different sequence and a different exon structure in this region. (**e**) Splicing at the junction between exons 8 and 9 of *LpCOE-A*. This results in the insertion of 10 amino acids, as compared to *LpCOE-B*

Alternate splicing of lamprey *COE* genes

Daburon et al. (2008) previously provided evidence that the *P. marinus COE-B* locus included a duplicated helix 2 as found in all jawed vertebrate *COE* genes. They also showed *COE-B* to be alternately spliced in *L. fluviatilis*, yielding transcripts encoding one or both H2 copies, but failed to identify similar splicing for *COE-A*. Since the presence or absence of the duplicated H2 appears to be a major structural difference between jawed vertebrate and invertebrate *COE* genes, we sought to clarify the structure and splicing of both lamprey *COE* genes.

Schematic intron-exon maps of the two lamprey *COE* loci, inferred by mapping *L. planeri* transcripts to the *L. japonicum* genome, are shown in Fig. 2a. Exons are numbered


a
DNA binding domain
ZNF
TIG/IPT
H1
H2d
H2a
Transactivation
5'UTR
3'UTR
COE-B
COE-A
Mouse COE1
Intron class
b
Splicing of H2d in COE-B
Helix 1
Loop
Helix 2d
Loop
Helix 2a
Exon 11
Exon 12
Exon 13
HsCOE1 ALNEPTIDYG FQRLQKVIPR HPGDPERLPK EVILKRAADL VEALYGMPHN NQEIILKRAA DIAEALY
HsCOE2 ALNEPTIDYG FQRLQKVIPR HPGDPERLAK EMLLKRAADL VEALYGTPHN NQDIILKRAA DIAEALY
HsCOE3 ALNEPTIDYG FQRLQKVIPR HPGDPERLPK EVLLKRAADL VEALYGMPHN NQEIILKRAA DIAEALY
HsCOE4 ALNEPTIDYG FQRLQKVIPR HPGDPERLPK EVLLKRAADL AEALYGVPSS NQELLLKRAA DVAEALY
LpCOEB H2d ALNEPTIDYG FQRLQKVIPR HPGDPERLPK EVILKRAADL VEALYGLPHY NQEIILKRAA DIAEALY
LpCOEB ALNEPTIDYG FQRLQKVIPR HPGDPERLPK ---------- ---------- --EIILKRAA DIAEALY
LpCOEA ALNEPTIDYG FQRLQKLIPR HPGDPERLPK ---------- ---------- --EIILKRAA DIAEALY
CiCOE ALNEPTLDYG FQRLLKTVPR HPGDPERLPK ---------- ---------- --EIILKRAA DVMEAVI
BfCOE SLNEPTIDYG FQRLAKLVPR HPGDPERLPK ---------- ---------- --EIILKRAA DLAEAIY
Exon 11
Exon 13
c
Splicing of TA1 region in COE-B
Exon 13
Exon 14
LpCOEB TA1v1 QNQLPALGNS P-AHHAMMGV NSFGGQLAVS VSEPGQSVEQ AMLSHAPGRY PCYSRNTSSV SPRGYVPSST
LpCOEB TA1v2 QNQLPALGNS P-AHHAMMGV NSFGGQLAVS VSEPGQSVEQ AMLSHAPG-- --YSRNTSSV SPRGYVPSST
LpCOEA HNQLPAINNS PPGLPGMMGM NSFGGQLAIS VSEASQAADQ G--------- --YTRNSSSV SPHGYVPSST
Exon 13
Exon 14
d
Splicing of TA2 region in COE-B
Exon 14
Exon 15
LpCOEB TA2v1 PQQSNYNTVS NSMNGYGNAG -MGNLGVPGS PGFLNGSSAS SPYAIMPSSP PVAAPPSITL PSNCSGAGGV
LpCOEB TA2v2 PQQSNYNTVS NSMNGYGNAG -MGNLGVPGS PGFLNGSSAS SPYAIMPSSP PVAAPPSITL PSNCSGAGGV
LpCOEB TA2v3 PQQSNYNTVS NSMNGYGNAG -MGNLGVPGS PGFLNGSSAS SPYA------ ---------- ----------
LpCOEB TA2v4 PQQSNYNTVS NSMNGYGNAG -MGNLGVPGS PGFLNGSSAS SPYA------ ---------- ----------
LpCOEA PQQSHYSSVS SSMNGYGVSG GMGGLGMGAS PGYLT-SSGG SPYGSTCPAP LFSS------ ----------
Exon 15
Exon 16
Exon 17
LpCOEB TA2v1 FSFSPAHMIS AVKQKSAFAP VVRPQASPPP NCTSATGNAL QGNFGHGLVA QFKDSVLAMS GLVVPPM
LpCOEB TA2v2 FSFSPAHMIS AVKQKSAFAP VVRPQASPPP NCTSATGNAL Q--------- -------AMS GLVVPPM
LpCOEB TA2v3 ---------- ---------- ---------- ---------- -GNFGHGLVA QFKDSVLAMS GLVVPPM
LpCOEB TA2v4 ---------- ---------- ---------- ---------- ---------- -------TMS GLVVPPM
LpCOEA ---------- -VKQKSAFAP VVRPQTSPPP TCTSANANGL Q--------- -------VMS GLVVPPM
Exon 14
Exon 15
Exon 17
e
Splicing of DBD region in COE-A
Exon 7
Exon 8
Exon 8
Exon 9
LpCOEB FQVVISTTVN VDGHVLAVSD NMFVHNNSKH GRRARRLDPS E--------- -AATPCIKAI SPSEGWTTGG
LpCOEA DBDv1 FQVVVSTTVH VDGHVLAVSD NMFVHNNSKH GRRARRIDPS EGTPPPSYLD NAATPCIKAI SPSEGWTTGG
LpCOEA DBDv2 FQVVVSTTVH VDGHVLAVSD NMFVHNNSKH GRRARRIDPS E--------- -AATPCIKAI SPSEGWTTGG
Exon 7
Exon 8
Exon 8
Exon 9

sequentially based on *COE-B*, with numbering preserved in the other genes to illustrate exon homology. The intron-exon structure and intron phase class of chordate *COE* genes have been previously shown to be relatively conserved (Daburon et al. 2008), and our analysis confirmed this, with *COE-A* and *COE-B* very similar to each other and to jawed vertebrate *COE* genes, with the exception of the duplicated H2, and some variation in C terminal exon structure (Fig. 2a). This confirms the previous report (Daburon et al. 2008) of a duplicated H2 in *COE-B*, and RT-PCR of *L. planeri* demonstrated alternate splicing here in this species (Fig. 2, S2). We found no evidence for a similar exon duplication in *COE-A*; a second H2 was not present in either the *L. japonicum* or *P. marinus* genome assemblies, and neither RT-PCR across this region (not shown) or transcriptome analysis (see following texts) identified alternate splicing in this region for *LpCOE-A*. Neither *COE-A* nor *COE-B* hagfish transcripts contained a duplicated H2, though in the absence of a hagfish genome, we cannot determine if this is due to splicing or if hagfish lack the duplicated exon.

Among the RT-PCR clones generated for *LpCOE-B* were two that encoded truncated ORFs. One contained a small insertion of 14 bp in the ZNF of the DBD. This altered the reading frame of the ORF, leading to premature truncation of the predicted protein within the DBD. Comparison of this sequence to the *L. japonicum* genome indicates that it is derived from splicing at a site located 14 bp upstream of the canonical exon 6 splice acceptor site, within the intron between exons 5 and 6 (Fig. 2a, S1). The other contained a deletion in the IPT/TIG domain, also altering the ORF and truncating the predicted protein within this domain. This derived from alternate splicing between exons 9 and 10, removing 23 bp from exon 10 (Fig. S1). To examine whether these were biologically meaningful, and to test for other splice variants, we extracted predicted *COE* transcripts from an *L. planeri* embryo transcriptome assembly (Supplementary file 2). In addition, we mined individual fragments (derived from the paired-end reads) from the sequence data used to construct this assembly, identifying and counting fragments that bridged between exons (Tables 1 and 2). Both these analyses confirmed the H2 splice variants of *LpCOE-B*. Neither showed evidence of *LpCOE-B* splicing in either the ZNF or IPT/TIG domains, with 18 and 24 fragments, respectively, reflecting the canonical splice, and none reflecting the putative alternate splices discussed previously (Tables 1 and 2). However, these analyses did identify other splice products for both *LpCOE-A* and *LpCOE-B* (Tables 1 and 2, Fig. 2a–e). *LpCOE-A* showed a single alternate product, at the boundary between the DBD and IPT/TIG domains. This resulted from use of an alternate splice site in exon 8 which maintained the reading frame, with the two forms represented by 19 and 10 fragments, respectively (Tables 1 and 2; Fig. 2a, e). *LpCOE-B* showed multiple alternate products in the C terminal TA domain, all supported by multiple fragments (Tables 1 and 2; Fig. 2a, c, d). One, which we named TA1, results from loss of 12 bp at the junction between exons 13 and 14. In addition to this, four alternate variants (jointly named TA2) were detected from splicing of exons 14, 15, 16 and 17, resulting in the loss of varying amounts of sequence from the C-terminal region of the protein, with the 5′ end of exon 15 also represented by fragments mapping to two closely spaced acceptor sites (Fig. 2d; Tables 1 and 2). Two additional variants were reflected by single fragments in the raw transcriptome data (Tables 1 and 2) but not found in the assembly; these were not considered further.

Finally, we sought to test alternate splicing experimentally (with the exception of the variants represented only by single transcriptome fragments, which we did not address further). We designed primers for each and amplified them from staged *L. planeri* embryo RNA, confirming band identity by sequencing (Fig. S2). This confirmed all the variants predicted by the transcriptome analysis were genuine, with all amplified from embryo and larval RNA. However, as with the transcriptome, it failed to confirm the variants detected in the ZNF and IPT/TIG domains *of LpCOE-B*, instead only amplifying a single band encoding the canonical ORF from each region (Fig. S2). Since these variants were originally identified as clones from an amplification-cloning experiment, they may reflect rare events difficult to detect by RT-PCR or transcriptomics, or PCR amplification artefacts. As their biological relevance is questionable, we have not sought to investigate this further.

In summary, analysis of splice variants and intron-exon structures in multiple lamprey species indicates that the two lamprey *COE* loci are structurally distinct. *COE-B* has a duplicated H2 and alternately splices this to produce jawed vertebrate-like and invertebrate-like *COE* transcripts. *COE-A* encodes only one H2 and has no alternate splicing in this region. In *L. planeri*, both genes are spliced at other points, either by skipping exons, or by use of closely spaced alternate splice sites near intron-exon boundaries resulting in the loss/gain of small numbers of amino acids.

Expression of *LpCOE-A* in *L. planeri* embryos

Expression of *LpCOE-A* was analysed from stages 21 to 28. At stage 21, expression is seen in the developing mandibular arch as well as in a domain under the notochord (Fig. 3a). At stage 22, expression appears in the nascent second and third pharyngeal arches (Fig. 3b). At stage 23, transcripts are seen in the spinal cord and ventral regions of the diencephalon, midbrain and hindbrain, in the trigeminal, geniculate and posterior lateral line placode/ganglia, as well as in the developing fourth pharyngeal arch and in an extending domain below the notochord (Fig. 3c, d). At this stage, faint expression starts to be seen in the nasohypophyseal plate (NHP) (Fig. 3d, arrowhead). At stage 24, *LpCOE-A* expression is seen as

Table 1 Sequences marking the 3′ and 5′ ends of *COE-A* exons and the number of transcriptome fragments confirming each

Exon-exon junction	3′ of upstream exon	5′ of downstream exon	Count
Exon 1–exon 2	GCCAGCACCGCGGCACAGAG	TGGCATTGCGCTAGCGCGGG	29
Exon 2–exon 3	ACTTCGTGGAGAAGGATCGG	GAACCCAACAATGAAAAGAC	31
Exon 3–exon 4	ACAGCTGTTGTACAGTAATG	GTGTGCGGACGGAGCAAGAT	29
Exon 4–exon 5	TCGACTCCATGAACAAACAG	GCCATTATCTATGAAGGGCA	21
Exon 5–exon 6	ACGCACGAGATCATGTGCAG	CCGCTGCTGCGACAAGAAGA	28
Exon 6–exon 7	GATCCGGTGATAATAGACAG	ATTCTTTCTGAAGTTCTTTC	29
Exon 7–exon 8	GGGACATGCGCCGATTTCAG	GTGGTTGTCTCGACGACAGT	38
Exon 8 short–exon 9 canonical	CAGAATCGACCCCTCCGAAG	CAGCCACACCGTGCATCAAA	8
Exon 8 short–exon 9 short	CAGAATCGACCCCTCCGAAG	CCACACCGTGCATCAAAGCC	2
Exon 8 canonical–exon 9 canonical	ACCGTCTTATCTGGACAATG	CAGCCACACCGTGCATCAAA	19
Exon 8 canonical–exon 9 short	ACCGTCTTATCTGGACAATG	CCACACCGTGCATCAAAGCC	0
Exon 9–exon 10	CCATGCTGGTGTGGAGCGAG	CTCATCACGCCCCATGCCAT	33
Exon 10–exon 11	AGGGCGATTCGTGTACACTG	CTCTGAATGAGCCAACAATT	23
Exon 11–exon 13	ACCCAGAGAGGCTTCCCAAG	GAAATTATCCTGAAGCGAGC	31
Exon 13–exon 14	CTCCCAGGCTGCTGACCAGG	GGTACACGCGCAACAGCAGC	27
Exon 14 short–exon 15	CGGCGGCTCTCCCTACGGCA	TGAAACAGAAGAGTGCATTC	1
Exon 14 canonical–exon 15	TGCGCCCCTCTTTTCATCGG	TGAAACAGAAGAGTGCATTC	26
Exon 15–exon 17	CAACGCCAACGGTCTGCAAG	TCATGTCTGGACTGGTGGTC	17

two parallel lateral stripes from a dorsal view. This expression is found in the mantle layer of the neural tube, and it extends from the diencephalon posteriorly all along the growing spinal cord (Fig. 3e). Expression progresses dorsally in the diencephalon, midbrain and hindbrain with respect to stage 23, leaving a gap of expression at the level of the midbrain-hindbrain boundary (MHB) (Fig. 3f, g). Expression also increases in the NHP and ophthalmic, maxillomandibular and posterior lateral line placodes/ganglia, and also appears in the developing fifth pharyngeal arch (Fig. 3f, g).

At stage 25, expression expands in most of the CNS, running the entire length of the spinal cord (Fig. 4a). In the brain, expression covers the hindbrain, midbrain and most of the forebrain, except its ventral-most region which possibly corresponds to the hypothalamus (Fig. 4d). At this stage, transcripts are now present in the telencephalon (Fig. 4d). Faint expression is also seen in the epiphysis. In the hindbrain, a big patch of expression is seen at the level of r3–r5 dorsally (Fig. 4d, asterisk). Notably, the MHB remains unstained. Outside the brain, expression increases in the forming olfactory epithelium and trigeminal, geniculate and petrosal ganglia (Fig. 4d). At this stage, transcripts are also seen in the eight pharyngeal arches and in the upper lip (Fig. 4d). From a ventral view, expression is distinguished in the mesoderm—and possibly neural crest—of each pharyngeal arch but is clearly excluded from the endoderm and epidermis (Fig. 4f). From a dorsal view, staining is seen as two lateral stripes along the neural tube as in stage 24 embryos (Fig. 4e). Expression under the notochord also progresses posteriorly as the embryo elongates (Fig. 4a, i, arrows). Cross-sections of stage 25 embryos reveal strong staining on the lateral sides of the neural tube and a thin, unstained domain in the middle demarcating the mantle layer and the ventricular zone, respectively (Fig. 4b, c, g, h). At the level of the otic vesicle, there is a gap of expression in the middle of the mantle layer with respect to the dorso-ventral axis (Fig. 4c, white asterisks). Interestingly, there is also a relatively wide unstained area in the dorsal-most part of the brain (Fig. 4b, c, black asterisks), whereas in the forming spinal cord, it is mostly seen in the ventral-most part (Fig. 4h, black asterisk). Outside the nervous system, expression is seen in mesenchyme of the pharyngeal arches (Fig. 4b, f). Expression is also seen surrounding roughly the ventral half of otic vesicles (Fig. 4c, arrows). Medial expression in this domain possibly corresponds to the acoustic ganglion. In the body, transcripts are found in mesenchyme ventro-lateral to the notochord and in a more lateral domain just above the vitellum (Fig. 4g, black and white arrows, respectively). This expression is not observed more posteriorly (Fig. 4h).

At stage 26, *LpCOE-A* expression diminishes both in the head and spinal cord (Fig. 4j). In the head, reduction of expression is seen in the ventral midbrain, dorsal diencephalon and olfactory epithelium (Fig. 4k). *LpCOE-A* is no longer observed in the epiphysis. A marked patch of expression is maintained in the dorsal hindbrain at the level of the otic vesicle (Fig. 4k, white asterisk). Expression in pharyngeal arches remains relatively strong, and staining is now seen in a forming ninth pharyngeal arch depicting eight pharyngeal

Table 2 Sequences marking the 3′ and 5′ ends of *COE-B* exons and the number of transcriptome fragments confirming each

Exon-exon junction	3′ of upstream exon	5′ of downstream exon	Count
Exon 1–exon 2	GCCAACACGGCCGCCCAGAG	CGGAGTCGCTCTGGCGAGGG	12
Exon 2–exon 3	ACTTTGTCGAGAAGGACAGA	GAACCAAACAGTGAAAAAAC	8
Exon 3–exon 4	ACAGTTACTCTACAGCAATG	GCATCCGCACGGAGCAAGAC	15
Exon 4–exon 5	TCGACTCCATGACTAAGCAG	GCGATCATTTACGATGGGCA	14
Exon 5–exon 6 long	ACGCACGAGATCATGTGCAG	TGTGCAACTCACAGTCGCTG	0
Exon 5–exon 6 canonical	ACGCACGAGATCATGTGCAG	TCGCTGCTGCGACAAGAAGA	18
Exon 6–exon 7	GACCCGGTCATCATCGACAG	GTTTTTTCTCAAGTTCTTCC	8
Exon 6–exon 9 short	GACCCGGTCATCATCGACAG	CTACGCCGTGCATCAAAGCA	1
Exon 7–exon 8	GAGACATGAGGCGGTTCCAG	GTCGTCATCTCCACGACCGT	14
Exon 8–exon 9 canonical	GAGACTCGACCCTTCGGAAG	CAGCTACGCCGTGCATCAAA	6
Exon 8–exon 9 short	GAGACTCGACCCTTCGGAAG	CTACGCCGTGCATCAAAGCA	19
Exon 9–exon 10 canonical	CCATGCTCGTGTGGAGTGAG	CTGATCACACCCCATGCCAT	24
Exon 9–exon 10 short	GGGACGCTTCGTCTACACCG	GGTGCAGACACCTCCCCGCC	0
Exon 10–exon 11	GGGACGCTTCGTCTACACCG	CTCTGAACGAGCCCACCATT	18
Exon 11–exon 12	ACCCCGAGAGGCTACCCAAG	GAGGTGATCTTGAAGCGCGC	5
Exon 11–exon 13	ACCCCGAGAGGCTACCCAAG	GAAATCATTCTGAAGAGAGC	8
Exon 12–exon 13	GCCTGCCTCATTACAACCAG	GAAATCATTCTGAAGAGAGC	4
Exon 13 short–exon 14	CATGCTGAGCCACGCCCCCG	GTTACAGCCGCAATACGAGC	12
Exon 13 canonical–exon 14	CGCCCCCGGTCGGTACCCCT	GTTACAGCCGCAATACGAGC	8
Exon 14–exon 15 canonical	CGCCAGCTCGCCCTATGCCA	TCATGCCGTCAAGCCCCCCC	6
Exon 14–exon 15 short	CGCCAGCTCGCCCTATGCCA	AAACAGAAGAGCGCCTTCGC	3
Exon 14–exon 16	CGCCAGCTCGCCCTATGCCA	GGAATTTTGGACATGGTCTT	2
Exon 14–exon 17	CGCCAGCTCGCCCTATGCCA	CCATGTCCGGCTTGGTCGTC	6
Exon 15–exon 16	TACCGGCAATGCCCTCCAAG	GGAATTTTGGACATGGTCTT	8
Exon 15–exon 17	TACCGGCAATGCCCTCCAAG	CCATGTCCGGCTTGGTCGTC	4
Exon 16–exon 17	GTTTAAAGATTCGGTTTTAG	CCATGTCCGGCTTGGTCGTC	5

pouches (Fig. 4k). From a dorsal view, expression is restricted to the lateral side of the brain and spinal cord delimiting the extent of the ventricular zone in the middle (Fig. 4l). In the brain, the ventricular zone expands at the level of the epiphysis, the MHB and posterior hindbrain (Fig. 4l, white asterisks). Expression is still seen in trigeminal and geniculate ganglia. In the spinal cord, expression is considerably downregulated with respect to previous stages, and expression under the notochord disappears (Fig. 4m, arrows). At stage 27, expression diminishes even more with respect to stage 26 becoming more restricted, although it maintains the same general expression pattern (data not shown). At stage 28, expression is more reduced and confined to the head (Fig. 4n). In the brain, the same pattern is maintained overall, with transcripts still observed in the telencephalon, very faintly in the dorsal diencephalon, dorsal midbrain and restricted regions of the hindbrain (Fig. 4o). The trigeminal, geniculate and petrosal ganglia are still stained, and in the olfactory epithelium transcripts are observed in the ventral half of the epithelium (Fig. 4o). Expression in all pharyngeal arches is maintained relatively strong along their dorso-ventral axes (Fig. 4o).

Expression of *LpCOE-B* in *L. planeri* embryos

Expression of *LpCOE-B* was analysed from stages 21 to 28. At stage 21, weak expression is observed in the anterior spinal cord as distinct, widely spaced dots forming two lateral stripes as seen from a dorsal view (Fig. 5a, b, arrows). Later at stage 22, this expression pattern increases in intensity and extends posteriorly along the growing spinal cord (Fig. 5c). Expression is still seen as widely spaced dots from a dorsal view forming two parallel stripes of staining (Fig. 5d). At this stage, expression is also observed in the trigeminal (maxillomandibular) placodes and presumptive NHP (Fig. 5c, black and white arrowheads, respectively). Expression in the presumptive NHP is observed as two relatively large lateral dots in its posterior facet, which weakly extend towards the anterior until they progressively meet in the middle (Fig. 5c, inset). At stage 23, expression is maintained in the trigeminal placode, and it now appears in the developing geniculate and petrosal placodes (Fig. 5e). At this stage, expression persists in the spinal cord. At stage 24, *LpCOE-B* appears in the hindbrain, midbrain and diencephalon (Fig. 5f). Expression

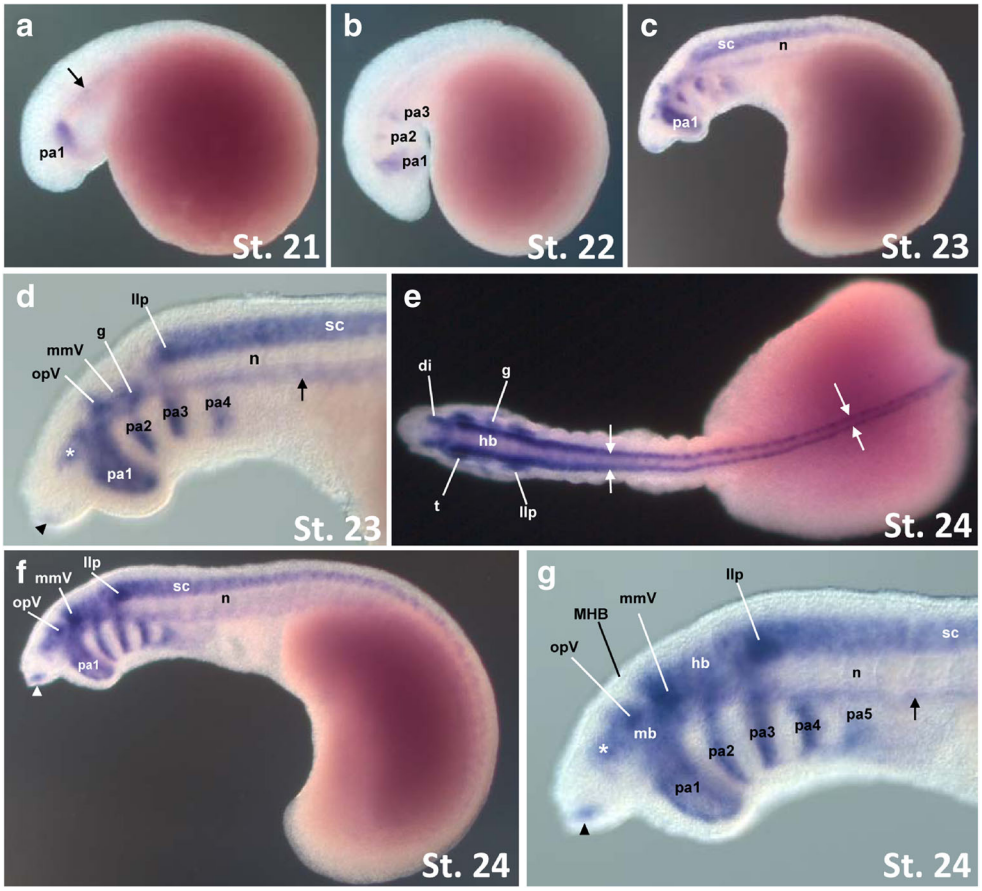
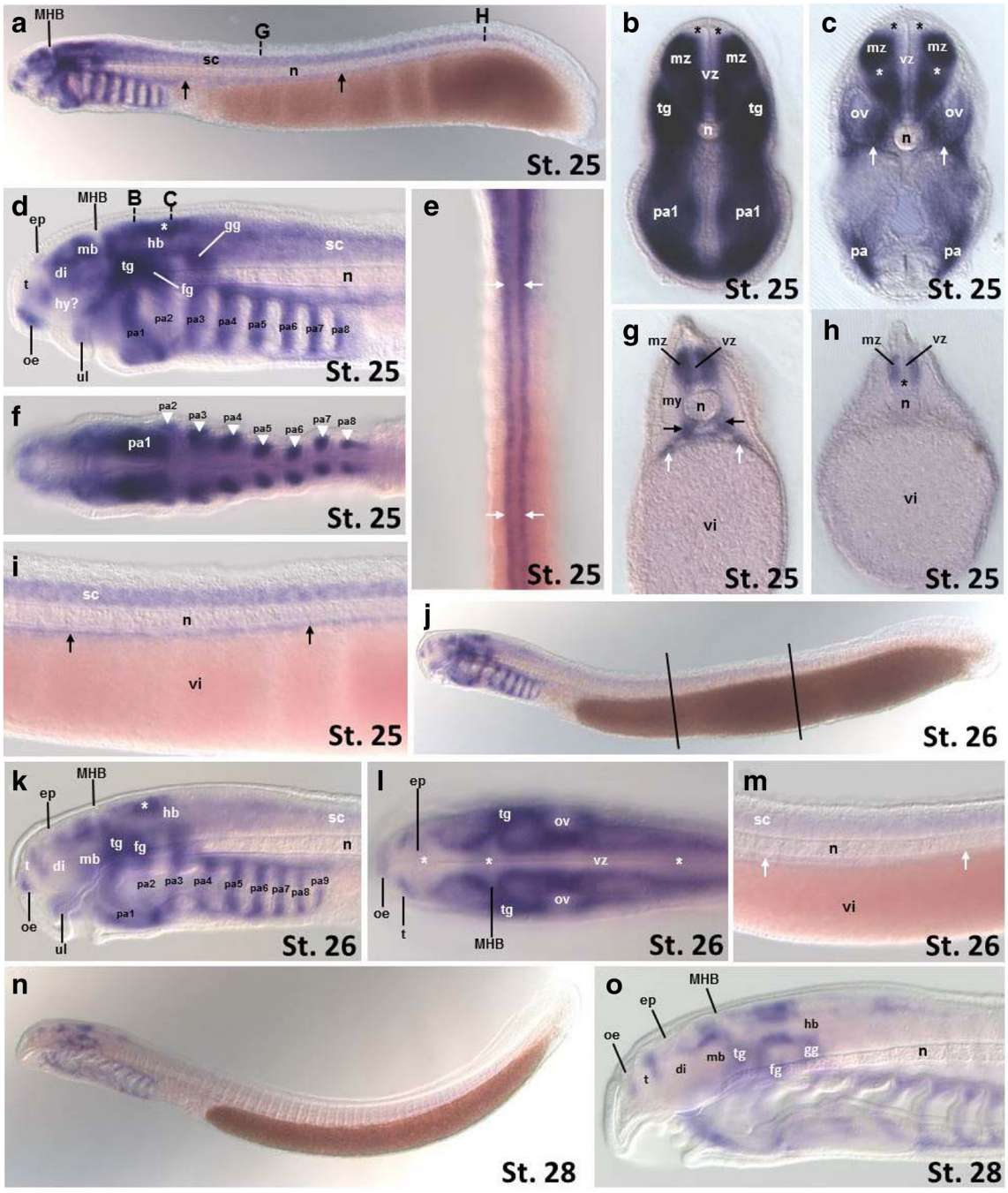


Fig. 3 *LpCOE-A* expression during *L. planeri* development at stages 21–24. (**a**–**d**, **f** and **g**) are lateral views, and (**e**) dorsal view. In all images, anterior is to the left. (**a**) At stage 21, expression is seen in the mandibular arch (pa1) and below the notochord (n: arrow). (**b**) At stage 22, expression appears in the second and third pharyngeal arches (pa2, pa3). (**c**, **d**) At stage 23, expression appears in the spinal cord (sc), ventral diencephalon (di: white asterisk), midbrain (mb: behind the ophthalmic placode (opV)) and hindbrain (hb), in the maxillomandibular (mmV), geniculate (g) and posterior lateral line (llp) placodes, the olfactory/neurohypophyseal plate (o/NHP: black arrowhead), as well as in the fourth pharyngeal arch (pa4). Expression below the notochord extends posteriorly (arrow). (**e**) At stage 24, from a dorsal view, expression is seen as two lateral stripes (arrows) all along the neural tube except in the telencephalon. (**f**, **g**) In the brain, expression progresses dorsally in the diencephalon (white asterisk), midbrain and hindbrain. Expression increases in the o/NHP (arrowheads) and in the opV, mmV and llp placodes/ganglia. At this stage, transcripts are also observed in the nascent fifth pharyngeal arch (pa5) and below the notochord (**g**, arrow). Additional abbreviation: MHB, midbrain-hindbrain boundary

intensifies in the forming maxillomandibular, geniculate and petrosal ganglia, as well as in the NHP (Fig. 5f, h). Expression in the spinal cord is considerably upregulated with respect to earlier stages and is observed roughly in the ventral half (Fig. 5g), preserving its position on the lateral sides of the neural tube as seen from a dorsal view (Fig. 5h). Outside the nervous system, strong expression is seen in mesenchyme of the first pharyngeal arch and in an extending domain above the pharynx (Fig. 5f, g).

At stage 25, *LpCOE-B* is strongly expressed in restricted regions of the diencephalon, midbrain and hindbrain, as well as in trigeminal, geniculate and petrosal ganglia and olfactory epithelium (Fig. 6a). In the hindbrain, transcripts are mostly observed dorsally in an anterior and a strong posterior patch (Fig. 6a, asterisks), with diffuse staining in between. Faint expression is observed in the epiphysis (Fig. 6a, black arrowhead). Notably, *LpCOE-B* is not expressed in the telencephalon at this stage. Expression in the spinal cord becomes very restricted to the dorsal side and is continuous with the posterior patch of expression in the hindbrain (Fig. 6b). *LpCOE-B* is not observed under the notochord in the trunk region as with *LpCOE-A* (Fig. 6b, arrows; compare with Fig. 4i, arrows). At this stage, transcripts are seen in the five anterior-most pharyngeal arches with strongest staining in the first arch (Fig. 6a). Like *LpCOE-A*, an expression domain between the notochord and pharyngeal arches is also present, although it does not extend as far to the posterior (Fig. 6a, arrow). From a dorsal view, staining is clearly restricted to the lateral sides of the neural tube similar to *LpCOE-A*, delineating the ventricular zone in the middle (Fig. 6c).

At stage 26, expression dramatically increases in the head and expression in the spinal cord is maintained dorsally (Fig. 6d). In the brain, expression is more refined, and transcripts are now detected in the telencephalon (Fig. 6f, white arrow).



In the diencephalon, mild staining is observed in the epiphysis (Fig. 6f, black arrowhead) and in a small region next to the telencephalon and two more dorsal domains next to the midbrain (Fig. 6f, black asterisks). In the midbrain, two large expression domains, one dorsal and one ventral, are separated by an unstained region (Fig. 6f, white open arrowheads). In the hindbrain, two big patches of staining are seen on the dorsal side, one at the level of rhombomeres 2–4 and another one at the transition with the spinal cord (Fig. 6f, white asterisks). Expression in the olfactory epithelium remains strong, and signals are now detected in the upper and lower lips (Fig. 6f). Eight pharyngeal arches are now stained all along their dorsoventral axes (Fig. 6f). The spinal cord maintains its dorsal expression domain although it is observed slightly weaker, and no expression is observed below the notochord in the spinal cord region (Fig. 6g). From a dorsal view, staining is observed at the lateral sides of the brain and spinal cord demarcating the ventricular zone in the middle (Fig. 6e, h). Similar to *LpCOE-A*, in the brain, the unstained medial region follows expansions of the ventricular zone at the level of the epiphysis, MHB and posterior hindbrain (Fig. 6e, white asterisks). Note, however, that at stage 25, these expansions

Fig. 4 *LpCOE-A* expression during *L. planeri* development at stages 25–28. (**a**, **d**, **i**–**k**, **m**–**o**) Lateral views. (**f**) Ventral view of the head of the embryo shown in (**a**). (**e**, **l**) Dorsal views of the anterior trunk and head, respectively. Anterior is to the left in all images except in (**e**) where anterior is to the top, and (**b**, **c**, **g**, **h**) are cross-sections of the embryo shown in (**a**) and (**d**). Levels of cross-sections are indicated in (**a**) and (**d**), and dorsal side is to the top. (**i**) and (**m**) are lateral views of the trunk at the levels marked by lines in (**a**) and (**j**), respectively. (**a**) At stage 25, expression expands in most of the CNS. (**d**) In the head, expression is seen in the telencephalon (tc), diencephalon, midbrain and hindbrain as well as in the epiphysis (ep). Strong expression in the dorsal hindbrain is seen at the level of rhombomeres 3–5 (asterisk). Transcripts are also observed in the forming olfactory epithelium and trigeminal, geniculate and petrosal ganglia. At this stage, transcripts are seen in the eight pharyngeal arches and in the upper lip (ul). (**f**) From a ventral view of the head, expression is distinguished in mesenchyme of each pharyngeal arch that possibly corresponds to the mesoderm and neural crest components of each arch. (**e**) From a dorsal view, staining is seen as two lateral stripes along the neural tube. Expression under the notochord progresses posteriorly as the embryo elongates (**a**, **i**, arrows). (**b**, **c**, **g**, **h**) Cross-sections reveal strong staining on the sides of the neural tube corresponding to the mantle layer (ml). (**b**) Expression is seen in trigeminal ganglia and mesenchyme of pharyngeal arches. (**c**) Transcripts are also seen surrounding the otic vesicle (ov) and possibly in the acoustic ganglia (white arrows). (**g**) In the body, transcripts are found in mesenchyme ventro-lateral to the notochord (black arrows) and in an extended domain (white arrows) just above the vitellum (vi). (**h**) This expression is not observed more posteriorly. (**j**, **k**, **m**) At stage 26, *LpCOE-A* expression is downregulated both in the head and spinal cord maintaining the same general pattern, but expression under the notochord disappears (**m**, arrows). All nine pharyngeal arches are stained at this stage (**k**). (**l**) From a dorsal view, expression is restricted to the lateral sides of the neural tube delimiting the extent of the ventricular zone (vz) in the middle, with expansions of the ventricular zone at the level of the epiphysis, the MHB, and posterior hindbrain (white asterisks). (**n**, **o**) At stage 28, expression is even more reduced and confined to the head. Additional abbreviations as previous figures plus: hy, hypothalamus; tc, telencephalon

are not so evident (Fig. 6c). Cross-sections at stage 26 confirm the lateral expression in the neural tube, generally stronger dorsally (Fig. 6i–k, white arrows). Cross-sections also reveal expression in maxillomandibular ganglia (Fig. 6i) as well as in mesenchyme of the pharyngeal arches and developing vellum (Fig. 6i, arrowheads and asterisks, respectively). A wide expression domain is observed at the dorsal edge of each pharyngeal arch (Fig. 6j, asterisks), which is contiguous with a line of stained cells located between the neural tube and myotomes, passing lateral to the notochord (Fig. 6j, black arrows). More posteriorly, expression is only observed in the spinal cord (Fig. 6k).

At stage 27, expression clearly diminishes in the head and in the spinal cord it is virtually absent (Fig. 6l, m). Expression in the diencephalon disappears almost completely (Fig. 6l). The two big patches of staining in the dorsal hindbrain persist although at a much lower level, and a third small patch is distinguished between them (Fig. 6l, asterisks). Expression remains strong in the olfactory epithelium, upper lip and mandibular arch, in dorsal domains of each pharyngeal arch and in the trigeminal, geniculate and petrosal ganglia. Also, expression in pharyngeal arches starts to fade, mostly ventrally, and expression disappears in the eighth pharyngeal arch (Fig. 6l). At stage 28, expression diminishes even more in the head, but the same expression sites are maintained, with strongest expression in the olfactory epithelium, upper lip, cranial ganglia and dorsal and ventral domains of each pharyngeal arch (Fig. 6n).

Discussion

The evolution of *COE* gene structure, splicing and duplication in vertebrates

Gene relationships between lampreys and jawed vertebrates have often proven difficult to decipher. While lamprey genome sequences provide evidence for at least one and possibly two genome duplications (Mehta et al. 2013; Smith and Keinath 2015; Smith et al. 2013), analysis of individual gene families often fails to clarify whether these are shared with the genome duplications of jawed vertebrates (Kuraku et al. 2009). Daburon and colleagues' original analysis of *COE* H2d evolution showed at least one lamprey *COE* gene had this duplicated exon, but poor genome quality prevented them from determining the status of the second *COE* gene (Daburon et al. 2008).

Combining molecular phylogenetics, synteny, intron-exon maps and splice form analysis helped us further explore this. First, molecular phylogenetics suggest cyclostomes have only two COE genes: our data includes three lamprey species, plus one hagfish, and all these sequences clearly fall into two orthologue groups. These genes do not appear as orthologues to specific jawed vertebrate *COE* genes in this analysis, and while synteny shows all these copies have derived by duplications of large gene blocks consistent with genome duplication, they too do not determine whether these are shared or evolved in parallel. However, the duplicated H2d is shared between lamprey *COE-B* and jawed vertebrate *COE* genes. Unless we consider this change has evolved in parallel, the exon duplication that formed H2d must have occurred prior to the separation of the two lineages, and prior to the duplications that formed *COE1–4* and *COE-A* and *COE-B*. This model, detailed in Fig. 7, implies that lamprey *COE-A* has lost the duplicated H2 exon and reverted to an invertebrate-like state. Intron phasing across these exons (Daburon et al. 2008), plus the presence in *COE-B* genes of an alternate splice product lacking H2d, indicates this loss could occur relatively easily while still preserving a functional ORF. The functional implications of having two versus three helices are not understood, though since the HLH region is involved in dimerization, COE protein heterodimers have been reported (Wang et al. 2002), and dimerization between two and three helix versions of COE proteins appears feasible (Daburon et al. 2008), we

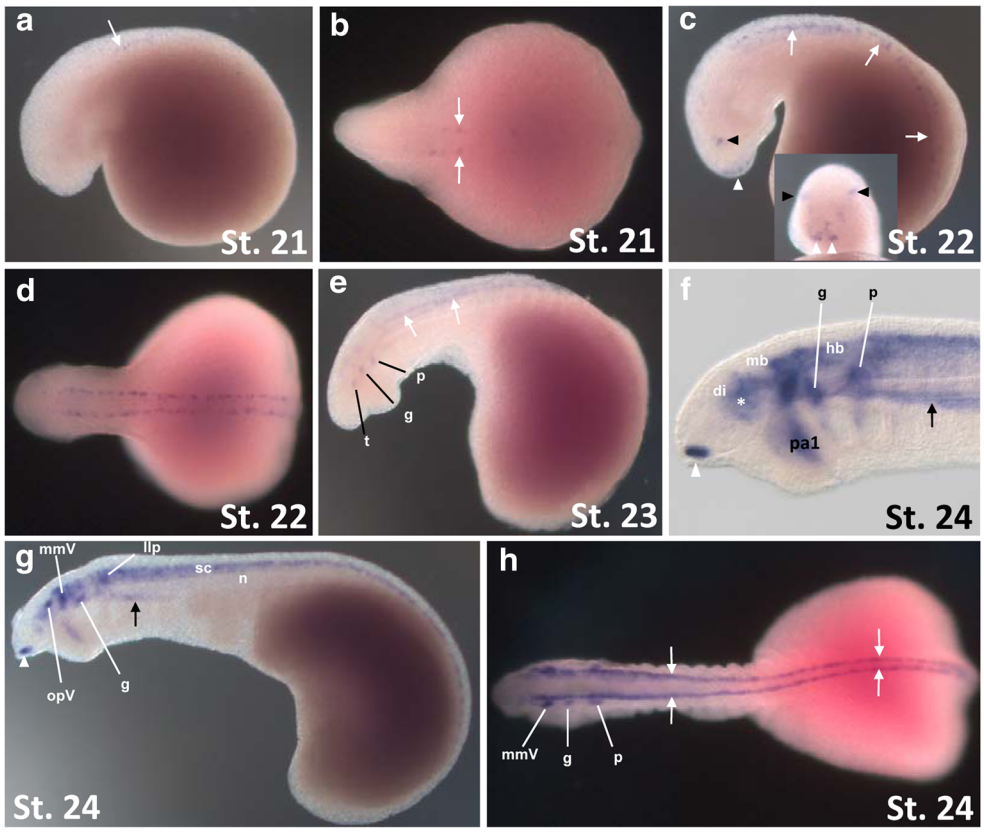


Fig. 5 *LpCOE-B* expression during *L. planeri* development at stages 21–24. (**a**, **c**, **e**, **f** and **g**) are lateral views. (**b**, **d** and **h**) are dorsal views. In all images, anterior is to the left. (**a**, **b**) At stage 21, *LpCOE-B* is observed in the anterior spinal cord as two lateral stripes (arrows). (**c**, **d**) At stage 22, expression extends posteriorly along the growing spinal cord (**c**, arrows), which is seen as two parallel stripes from a dorsal view (**d**). Transcripts are also observed in the trigeminal placodes (black arrowheads) and the o/NHP anlage (white arrowheads). Inset, frontal view of the embryo shown in (**c**). (**e**) At stage 23, expression is maintained in the spinal cord (arrows) and trigeminal placodes, and it now appears in the developing geniculate and petrosal placodes. (**f**–**h**) At stage 24, *LpCOE-B* appears in the hindbrain, midbrain and in a small region of the diencephalon (white asterisk) (**f**). Expression considerably intensifies in the trigeminal, geniculate and petrosal ganglia, as well as in the NHP (white arrowheads) and spinal cord (**g**). Outside the nervous system, strong expression is seen in mesenchyme of the first pharyngeal arch (pa1) and in an extending domain between the pharynx and notochord (black arrows). Expression in the neural tube is localised to the lateral sides as in previous stages (**h**, white arrows). Abbreviations as in previous figures

can speculate that *COE-B* H2d splicing in lampreys allows for a wider array of dimers.

We also observed other splice variants for both *LpCOE-A* and *LpCOE-B*. Splicing in the C-terminal TA domain of *LpCOE-B* resulted primarily from use of different exons, including what appears to be a new exon (exon 16 in Fig. 2a) which lacks an equivalent in either lamprey *COE-A* genes or jawed vertebrate *COE* genes. The sequence encoded by this exon was also absent from the hagfish *COE-B* sequence, though without a genome sequence, this is not conclusive. C-terminal splicing of jawed vertebrate *COE* genes has been little-studied, having only been experimentally verified for mouse *COE-4* (Wang et al. 2002). Other experimentally validated lamprey splice products resulted from use of alternative splice donor or acceptor sites, resulting in the addition or loss of small blocks of sequence. These splice products have no described counterparts in jawed vertebrates, so they likely represent lamprey innovations, as depicted in Fig. 7.

COE expression in the lamprey CNS and lineage-specific expression domains

The expression of *LpCOE-A* and *LpCOE-B* in the lamprey nervous system is in agreement with a conserved role of *COE* genes in neuronal differentiation. From the early stages of development, when expression of *LpCOE-A* and *LpCOE-B* appears in the rhombospinal region, both transcripts are observed at the lateral margins of the neural tube. This is similar to the expression of *COE* genes in post-mitotic neurons in the spinal cord of mouse, chicken, *Xenopus* and zebrafish (Bally-Cuif et al. 1998; Dubois et al. 1998; Garel et al. 1997). Expression of both *LpCOE-A* and *LpCOE-B* in the rhombospinal region appears complementary to that of *LpNgnA* (Lara-Ramirez et al. 2015), which is restricted to the ventricular zone of the neural tube.

In mouse and chicken, expression of *COE* genes is observed in the motor column at the ventrolateral margin of the developing spinal cord, and in a thinner domain that extends

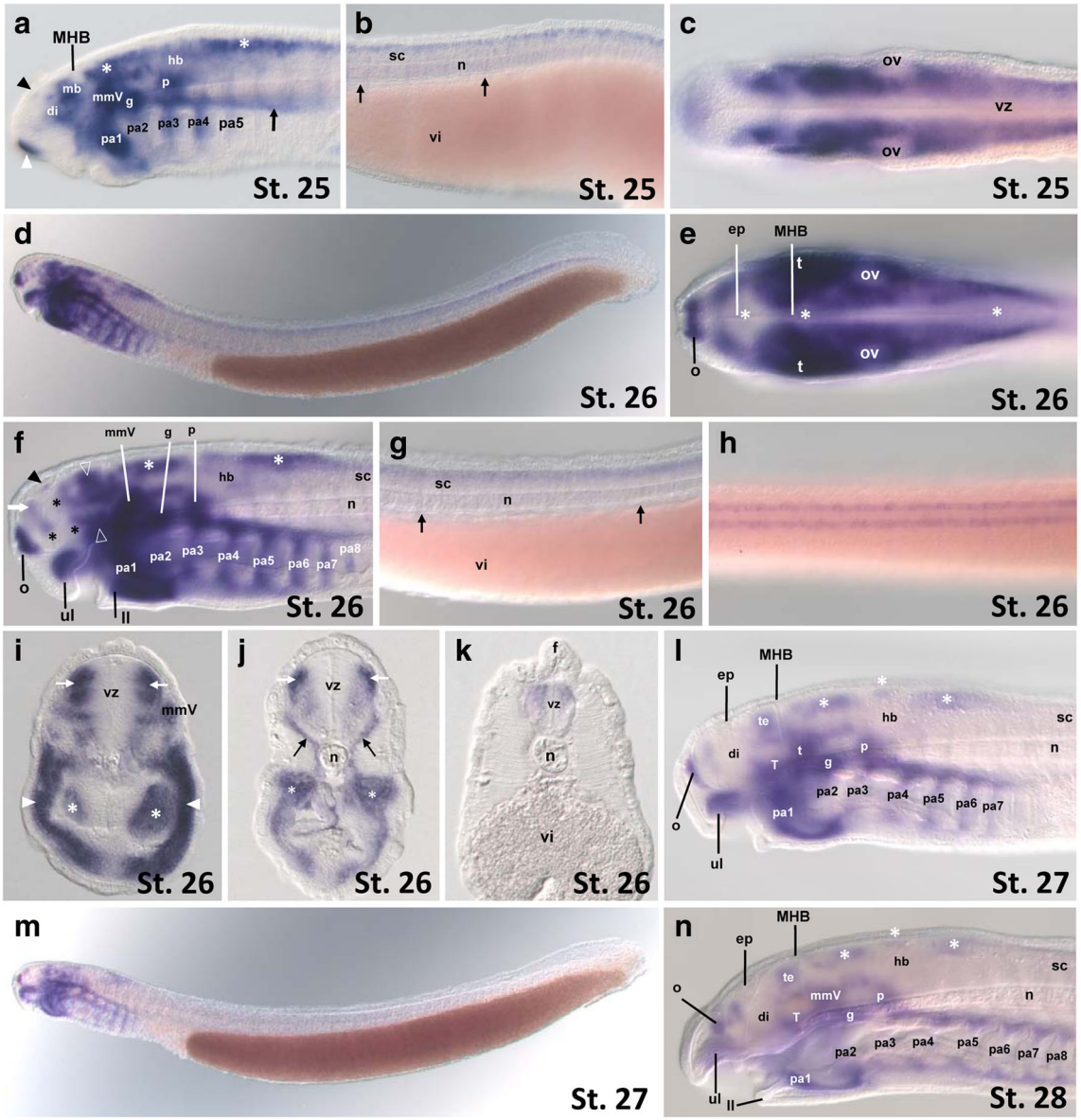


Fig. 6 *LpCOE-B* expression during *L. planeri* development at stages 25–28. (**a**, **b**, **d**, **f**, **g**, **l**–**n**) are lateral views; (**b**) and (**g**) are lateral views of the trunk region at the stages indicated. (**c**, **e**, **h**) are dorsal views of the head and trunk region. (**i**–**k**) are cross-sections of a stage 26 embryo, dorsal to the top. In all lateral and dorsal views, anterior is to the left. (**a**–**c**) At stage 25, *LpCOE-B* is strongly expressed in specific regions of the diencephalon, midbrain and hindbrain, as well as in the maxillomandibular, geniculate and petrosal ganglia and olfactory epithelium (**a**, white arrowhead). In the hindbrain, two patches of staining are observed dorsally (**a**, asterisks), with diffuse staining in between. Faint expression is also observed in the epiphysis (**a**, black arrowhead). *LpCOE-B* is not expressed in the telencephalon at this stage. Expression in the spinal cord becomes very restricted to the dorsal side, and transcripts are not observed under the notochord as with *LpCOE-A* (**b**, arrows). At this stage, transcripts are seen in the five anterior-most pharyngeal arches and in an extending domain just above the pharynx (**a**, black arrow). From a dorsal view, staining is still restricted to the lateral sides of the neural tube as with *LpCOE-A*, delineating the ventricular zone in the middle (**c**). (**d**–**k**) At stage 26, expression increases even more in the head. Staining is highly increased in the trigeminal (t), geniculate and petrosal ganglia as well as in the first pharyngeal arch (**d**, **f**). In the spinal cord, transcripts are restricted dorsally and are not observed under the notochord (**g**, arrows). Expression is localised to the lateral sides of the brain and spinal cord as seen in dorsal views (**e**, **h**). In the brain, expansions of the ventricular zone are observed at the level of the epiphysis, MHB and posterior hindbrain (**e**, asterisks). (**i**–**k**) Cross-sections reveal staining on the lateral sides of the neural tube corresponding to the mantle layer, with stronger expression dorsally (white arrows). Expression is also observed in pharyngeal arch mesenchyme possibly corresponding to both mesoderm and neural crest (arrowheads). Transcripts are also observed in the velum (asterisks) and in a stream of cells ventrolateral to the neural tube that reach pharyngeal expression (black arrows). Transcripts are also observed in maxillomandibular ganglia (**i**). (**l**, **m**) At stage 27, expression is clearly downregulated in the head, although maintaining the same pattern as stage 26 embryos, except in the diencephalon (**l**) and in the spinal cord (**m**) where transcripts are virtually absent. (**n**) At stage 28, expression is reduced even more but the same pattern as in previous stages is still observed in the head, with strongest expression seen in olfactory epithelium, upper lip, cranial ganglia and dorsal and ventral domains of each pharyngeal arch. Additional abbreviations as previous figures plus: ll, lower lip; T, tegmentum; te, optic tectum

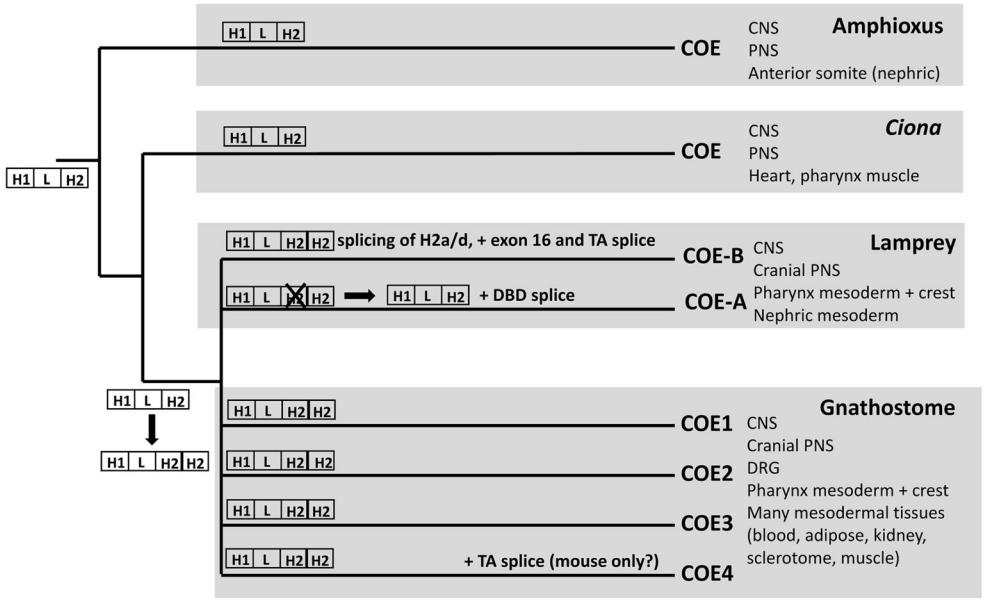


Fig. 7 A model for the evolution of the *COE* genes in chordates. Duplication of H2 is shown before the duplication and divergence of the vertebrate *COE* paralogues; we have shown vertebrate *COE* paralogue relationships as unresolved, as while synteny data shows it happened in both lineages by block duplication probably tied to genome duplication, lamprey-gnathostome orthology is not resolvable by either molecular phylogenetics or synteny. Lamprey *COE-A* then reverts to a 2 helix state. A summary of expression is shown next to each lineage; for lampreys and gnathostomes, this summarises the combined expression of all paralogues. Expression of *COE* genes in CNS and PNS is ancestral. Some mesodermal expression is also likely ancestral. We found no evidence for endodermal expression. Neural and mesodermal expression may both be older, as both are found in many other animals, though the type of mesoderm expressing *COE* genes is quite variable (Jackson et al. 2010). For sources of expression data, see respective sections of the "Discussion" section

dorsally at the level of the subventricular zone (El-Magd et al. 2014b; Garel et al. 1997). In the lamprey, however, expression of *LpCOE-A* is proportionately more extended, covering nearly all the mantle layer, and no subventricular zone-like domain of expression is observed with either lamprey *COE* gene. This expression could reflect an anatomical difference between lamprey and gnathostome spinal cords, or a gene regulation difference between lamprey and gnathostome *COE* genes. For example, given that *COE* genes are strongly expressed in the ventrolateral motor column of gnathostome spinal cords, it is possible that most of the lamprey mantle layer marked by *COE* genes corresponds to populations of motor neurons, which are proportionately more extended in the lamprey spinal cord than in jawed vertebrates. Alternatively, it is possible that the lamprey spinal cord also possesses a subventricular zone as in gnathostomes, but the expression of *COE* genes is activated in a bigger cell repertoire that also includes different interneuron subpopulations. In either case, a more detailed characterisation of neuronal populations of the developing lamprey spinal cord is needed to clarify the particular cell types where lamprey *COE* genes are expressed.

In the hindbrain, both lamprey *COE* genes may mark branchiomotor nuclei, which are morphologically discerned from stage 26 (Murakami et al. 2004). In the mouse, *COE1–3* are strongly expressed in facial branchiomotor neurons of r4, although at different time points (Garel et al. 2000). We also observed a clear distinction in hindbrain expression between lampreys and jawed vertebrates. In zebrafish and mouse, *COE* genes are first expressed in r2 and r4, leaving gaps of expression in r1, r3 and r5–7, although later in development they are activated in these remaining rhombomeres. In the lamprey, however, no such r2/r4 initial expression was observed. This is in keeping with the spatial distribution of branchiomotor nuclei with respect to rhombomeric boundaries, since, in lampreys, branchiomotor neurons are not in register with rhombomeres as in gnathostomes (Murakami et al. 2004). We also did not observe expression of either lamprey *COE* gene in rhombomeric boundaries as with mouse *COE* genes (Garel et al. 1997). These comparisons support the view that lampreys, while having a hindbrain that is fundamentally similar to gnathostomes in terms of broad rhombomeric organisation, differ with respect to the precise organisation of the cells that differentiate in each (Parker et al. 2016).

According to previous interpretations of the regionalisation of the lamprey diencephalon at stage 26, a time at which the embryonic lamprey brain acquires a well-defined compartmentalisation (Lara-Ramirez et al. 2015; Murakami et al. 2002; Murakami et al. 2001), both lamprey genes seem to be expressed in the alar (pretectum) and basal plates of prosomere (P) 1, but only *LpCOE-A* is expressed in P2 in the dorsal thalamus, except for a very small patch of *LpCOE-B* expression in the epiphysis. The presence or absence of a P3 territory in the embryonic lamprey brain is not resolved by either lamprey *COE* gene. We also note other differences in

expression: (i) The region of the embryonic lamprey hypothalamus has been clearly defined previously by expression of *TTF-1/Nkx2.1* (Osorio et al. 2005), and both lamprey *COE* genes seem to be absent from the hypothalamus, which is in sharp contrast with mouse *COE1–3* which are strongly expressed in this region (Garel et al. 1997). (ii) We did not detect specific expression of either lamprey *COE* gene in the midbrain-hindbrain boundary (MHB) as with zebrafish *COE2* and *COE3* (Li et al. 2010). (iii) We did not observe lamprey *COE* expression in the ventricular zone as occurs with mouse *COE* genes in the anterior hindbrain (Garel et al. 1997). (iv) We did not observe expression of either *LpCOE-A* or *LpCOE-B* in the lamprey retina, while retinal expression is seen with *Xenopus COE3* and mouse *COE1–4* (Garel et al. 1997; Pozzoli et al. 2001; Wang et al. 1997).

In the telencephalon, tetrapod *COE1* genes specifically mark the striatum of the lateral ganglionic eminence (LGE) and a cell corridor passing through the medial ganglionic eminence (MGE) to the ventral thalamus (Bielle et al. 2011; Lopez-Bendito et al. 2006). In the lamprey, we observe a thin domain of expression running continuously from the telencephalon to the ventral diencephalon close to the hypothalamus. The lamprey telencephalon has been shown to be divided into pallium and subpallium, and the lamprey subpallium has been proposed to be equivalent to the LGE only, with the MGE being a gnathostome innovation (Sugahara et al. 2011). However, a recent reanalysis of this question including study of the hagfish embryonic brain came to the alternative view, that the MGE is primitive and present in hagfishes and lampreys (Sugahara et al. 2016). Our data are consistent with this later interpretation of the lamprey telencephalon and suggest a conserved thalamo-striatal connection marked by *COE* expression in the vertebrate MGE.

COE expression in the vertebrate PNS

In the peripheral nervous system, both lamprey *COE* genes were observed in the olfactory placode/epithelium as well as in the placode-derived cranial ganglia. Jawed vertebrate olfactory epithelia and cranial ganglia also express *COE* genes, though there is variation in paralogue group(s) depending on the species examined (Bally-Cuif et al. 1998; Dubois et al. 1998; El-Magd et al. 2014b; Pozzoli et al. 2001; Wang and Reed 1993; Wang et al. 2002; Wang et al. 1997). The only exception is in lateral line placodes/ganglia where *LpCOE-A* was expressed, while to our knowledge, no other vertebrate *COE* gene has been reported to have lateral line expression. However, lateral line ganglia have not been well-studied as they are absent from some model species, so this probably just reflects missing data. Overall this likely reflects subfunctionalisation of *COE* expression and suggests cranial placode/ganglia expression preceded gene duplication. *COE* genes also show PNS expression in amphioxus and *Ciona* in cells postulated to be placode homologues (Mazet et al. 2005; Mazet et al. 2004), showing this is a chordate-wide character.

No expression of either gene was observed in lamprey DRG, a prominent expression site for mouse (Davis and Reed 1996; Wang et al. 1997) and chicken *COE* genes (El-Magd et al. 2014b). Lamprey embryonic DRG are visible at the stages examined and have been shown to express another bHLH gene, *LpNgnA*, at these stages (Lara-Ramirez et al. 2015). Since *COE* genes elsewhere mark differentiating neurons, this may reflect a delay in the differentiation of DRG neurons in lampreys compared to other vertebrates.

COE genes in mesodermal tissues

We observed lamprey *COE* gene expression in the mesenchyme of the pharyngeal arches and the mesenchyme dorsal to the pharynx that runs posteriorly, ventral to the notochord. Expression in pharyngeal arches has been observed with *Xenopus COE2* and *COE3* (Dubois et al. 1998; Pozzoli et al. 2001), chicken *COE1–3* (El-Magd et al. 2014b) and mouse *COE2* and *COE3* (Kieslinger et al. 2005) during embryonic development. Chicken *COE1* and *COE3* are mainly expressed in the neural crest component of pharyngeal arches, though weak expression in the mesodermal component was also observed (El-Magd et al. 2014b). In contrast, chicken *COE2* is mainly expressed in the mesodermal component of pharyngeal arches. We did not observe such a distinction with either lamprey *COE* gene. In addition, there is a time difference in pharyngeal expression of *COE* genes among tetrapods, as mouse *COE2* and *COE3* appear in pharyngeal arches before neural expression, whereas *Xenopus COE2* and *COE3* and chicken *COE1–3* are expressed after the onset of neural expression. In this sense, lamprey *COE* activation is more similar to *Xenopus* and chicken than to mouse. Zebrafish *COE2* expression has been reported in migrating cranial neural crest cells (Bally-Cuif et al. 1998), although its expression has not been followed at further developmental stages to determine if zebrafish *COE2* is activated in pharyngeal arches. Interestingly, in *Ciona*, *COE* is also expressed in pharyngeal mesoderm and has an important role in its specification (Razy-Krajka et al. 2014), but pharyngeal expression has not been reported from amphioxus. This suggests stepwise acquisition of *COE* pharyngeal mesenchyme expression in chordates: from no expression primitively to pharyngeal mesodermal expression in the vertebrate-tunicate ancestor, and acquisition of neural crest expression in vertebrates.

We also observed expression of both lamprey *COE* genes above the pharynx and running posteriorly ventral to the notochord, although their expression domains are not identical. Comparison with a recent anatomical description of lamprey mesoderm (Tulenko et al. 2013) reveals that *LpCOE-A* is

expressed in nephric tissues. Tunicates lack a homologue of the nephric system, but in amphioxus *COE* expression in the mesoderm is restricted to the ventral part of the left anterior somite (Mazet et al. 2004), where a structure known as Hatschek's nephridium will form. Hatschek's nephridium is proposed to be homologous to the vertebrate nephric system and expresses nephric marker genes (Kozmik et al. 1999). Expression of *COE* genes in the lamprey intermediate mesoderm suggests nephric expression may be ancestral for chordates. In agreement with this, mouse *COE4* is expressed in kidney (Wang et al. 2002).

No expression of either lamprey *COE* gene was observed in pre-somitic mesoderm, somites or their derivatives. Somitic expression of *COE* genes has been observed with mouse *COE2* (Kieslinger et al. 2005) and chicken *COE1–3* in the forming sclerotome, excluded from the dermomyotome (El-Magd et al. 2014b). Chicken *COE1* and *COE3* expression in these tissues is related to cartilage blastemas of the dorsal neural arches and proximal ribs, and to mesenchyme lateral to hypaxial and epaxial muscle precursors (El-Magd et al. 2014a). In mice, *COE* genes are expressed in osteoblasts and have been shown to have an important role in the differentiation of osteoclasts (Hesslein et al. 2009; Kieslinger et al. 2005). Thus, at least in mice and chicken, *COE* genes seem to have a role in the development of bone tissues. Absence of lamprey *COE* expression in somites might thus be explained by the lack of a mineralised skeleton.

In gnathostomes, *COE* genes are also expressed in the muscle lineage. In *Xenopus*, *COE2* and *COE3* are expressed in somites and in ventrally migrating hypaxial muscle, and direct expression of a number of genes that are involved in the commitment, migration and differentiation of muscle cells (Green and Vetter 2011). Also, mouse *COE1* is expressed in skeletal muscle and *COE3* is expressed in the diaphragm, where they have an important role in regulating muscle cell relaxation (Jin et al. 2014). However, we did not observe expression of either lamprey *COE* gene in any muscle derivatives. We also did not observe expression of lamprey *COE* genes associated to the heart as with chicken *COE3* (El-Magd et al. 2014b). Thus, our results suggest that all muscle-related expression of *COE* genes as observed in mouse and *Xenopus* is completely absent in lampreys and, to a broader extent, all somitic *COE* expression is absent in lampreys. Amphioxus *COE* is only expressed in the anterior-most somite, where the nephric system develops (Mazet et al. 2004) and so is also not muscle-related. However, in the tunicate *Ciona*, *COE* has a role in specifying the heart and pharyngeal muscle lineages (Razy-Krajka et al. 2014). It is therefore unclear whether muscle expression is an ancient character either lost or undetected in lampreys and amphioxus or has evolved independently in *Ciona* and gnathostomes. The role of the *Drosophila COE* orthologue, *Collier*, in muscle development (Crozatier and Vincent 1999) would suggest the former.

Summary and conclusions

Our analysis of lamprey *COE* gene identifies both similarities and differences in gene structure and expression between *COE* genes in vertebrates, other chordates and other animals. *COE* gene structure is generally conserved in vertebrates, though lamprey *COE-B* has a new exon towards its 3′ end. Intron phasing through this region is the same in both lamprey *COE* genes and in vertebrate *COE* genes (Fig. 2a; see Daburon et al. 2008 for other vertebrate *COE* genes). This means alternate splicing through this region is potentially viable for all these genes, though it has only been identified in lamprey *COE-B* and mouse *COE4*. Unless it has remained undetected for other *COE* genes, this would hence appear to have evolved in parallel. Splicing of the duplicated H2 also appears to be unique to lamprey *COE-B* genes. The model for the evolutionary origin of the duplicated H2 proposed by Daburon et al. (2008), in which duplication of the H2-encoding exon is followed by loss of the 3′ end of the most 5′ duplicate, is consistent with our data. However, we find no evidence for this in *COE-A* genes. This suggests an evolutionary scenario in which this exon has been lost by *COE-A* following genome duplication (Fig. 7).

We are also now able to map *COE* gene expression onto this model for the major chordate lineages: amphioxus, tunicates, cyclostomes and gnathostomes (further detailed in Fig. 7). Both CNS and PNS expression are ancestral for chordates, and in vertebrates, PNS expression has become incorporated into the peripheral ganglia (although on current data, DRG expression is in gnathostomes only). Some aspects of mesodermal expression also appear ancestral: nephric expression for all chordates and pharyngeal expression for the tunicate-vertebrate ancestor. General expression in the nervous system and some mesodermal derivatives seems to be ancestral for all metazoans (Jackson et al. 2010), though descriptions are insufficient to determine if the latter corresponds to nephric and/or pharyngeal-type cells. In gnathostomes, COE expression appears in other mesodermal cell types, and gnathostome COE expression in the brain differs to that of lampreys (discussed previously). While the latter could map onto the evolution of patterns of neurogenesis in the vertebrate brain, mesodermal expression may reflect diversification of cell types early in vertebrate evolution.

Acknowledgements We thank the Forestry Commission of England for permission to collect lamprey embryos and Dr. Jo Begbie for access to histology facilities. RL-R was supported by the Mexican National Council for Science and Technology (CONACYT). CP was supported by a Royal Society Newton International Fellowship and by an EMBO Long-Term Fellowship.

References

Akerblad P, Lind U, Liberg D, Bamberg K, Sigvardsson M (2002) Early B-cell factor (O/E-1) is a promoter of adipogenesis and involved in control of genes important for terminal adipocyte differentiation. Mol Cell Biol 22:8015–8025

Akerblad P et al (2005) Gene expression analysis suggests that EBF-1 and PPARgamma2 induce adipogenesis of NIH-3T3 cells with similar efficiency and kinetics. Physiol Genomics 23:206–216. https://doi.org/10.1152/physiolgenomics.00015.2005

Bally-Cuif L, Dubois L, Vincent A (1998) Molecular cloning of Zcoe2, the zebrafish homolog of Xenopus Xcoe2 and mouse EBF-2, and its expression during primary neurogenesis. Mech Dev 77:85–90

Bielle F et al (2011) Slit2 activity in the migration of guidepost neurons shapes thalamic projections during development and evolution. Neuron 69:1085–1098. https://doi.org/10.1016/j.neuron.2011.02.026

Corradi A et al (2003) Hypogonadotropic hypogonadism and peripheral neuropathy in Ebf2-null mice. Development 130:401–410

Crozatier M, Vincent A (1999) Requirement for the Drosophila COE transcription factor Collier in formation of an embryonic muscle: transcriptional response to notch signalling. Development 126: 1495–1504

Crozatier M, Valle D, Dubois L, Ibnsouda S, Vincent A (1996) Collier, a novel regulator of Drosophila head development, is expressed in a single mitotic domain. Curr Biol 6:707–718

Daburon V, Mella S, Plouhinec JL, Mazan S, Crozatier M, Vincent A (2008) The metazoan history of the COE transcription factors. Selection of a variant HLH motif by mandatory inclusion of a duplicated exon in vertebrates. BMC Evol Biol 8:131. https://doi.org/10.1186/1471-2148-8-131

Davis JA, Reed RR (1996) Role of Olf-1 and Pax-6 transcription factors in neurodevelopment. J Neurosci 16:5082–5094

Demilly A, Simionato E, Ohayon D, Kerner P, Garces A, Vervoort M (2011) Coe genes are expressed in differentiating neurons in the central nervous system of protostomes. PLoS One 6:e21213. https://doi.org/10.1371/journal.pone.0021213

Dubois L, Vincent A (2001) The COE—Collier/Olf1/EBF—transcription factors: structural conservation and diversity of developmental functions. Mech Dev 108:3–12

Dubois L, Bally-Cuif L, Crozatier M, Moreau J, Paquereau L, Vincent A (1998) XCoe2, a transcription factor of the Col/Olf-1/EBF family involved in the specification of primary neurons in Xenopus. Curr Biol 8:199–209

El-Magd MA, Saleh AA, El-Aziz RM, Salama MF (2014a) The effect of RA on the chick Ebf1-3 genes expression in somites and pharyngeal arches. Dev Genes Evol 224:245–253. https://doi.org/10.1007/s00427-014-0483-y

El-Magd MA, Saleh AA, Farrag F, Abd El-Aziz RM, Ali HA, Salama MF (2014b) Regulation of chick Ebf1–3 gene expression in the pharyngeal arches, cranial sensory ganglia and placodes. Cells Tissues Organs 199:278–293. https://doi.org/10.1159/000369880

Fields S, Ternyak K, Gao H, Ostraat R, Akerlund J, Hagman J (2008) The 'zinc knuckle' motif of Early B cell Factor is required for transcriptional activation of B cell-specific genes. Mol Immunol 45:3786–3796. https://doi.org/10.1016/j.molimm.2008.05.018

Furlong RF, Holland PW (2002) Were vertebrates octoploid? Philos Trans R Soc Lond Ser B Biol Sci 357:531–544. https://doi.org/10.1098/rstb.2001.1035

Garcia-Dominguez M, Poquet C, Garel S, Charnay P (2003) Ebf gene function is required for coupling neuronal differentiation and cell cycle exit. Development 130:6013–6025. https://doi.org/10.1242/dev.00840

Garel S, Marin F, Mattei MG, Vesque C, Vincent A, Charnay P (1997) Family of Ebf/Olf-1-related genes potentially involved in neuronal differentiation and regional specification in the central nervous system. Dev Dyn 210:191–205. https://doi.org/10.1002/(SICI)1097-0177(199711)210:3<191::AID-AJA1>3.0.CO;2-B

Garel S, Garcia-Dominguez M, Charnay P (2000) Control of the migratory pathway of facial branchiomotor neurones. Development 127: 5297–5307

Grabherr MG et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–U130. https://doi.org/10.1038/nbt.1883

Green YS, Vetter ML (2011) EBF proteins participate in transcriptional regulation of Xenopus muscle development. Dev Biol 358:240–250. https://doi.org/10.1016/j.ydbio.2011.07.034

Hagman J, Gutch MJ, Lin H, Grosschedl R (1995) EBF contains a novel zinc coordination motif and multiple dimerization and transcriptional activation domains. EMBO J 14:2907–2916

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hesslein DG et al (2009) Ebf1-dependent control of the osteoblast and adipocyte lineages. Bone 44:537–546. https://doi.org/10.1016/j.bone.2008.11.021

Imai KS, 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. https://doi.org/10.1242/dev.01270

Jackson DJ et al (2010) Developmental expression of COE across the Metazoa supports a conserved role in neuronal cell-type specification and mesodermal development. Dev Genes Evol 220:221–234. https://doi.org/10.1007/s00427-010-0343-3

Jimenez MA, Akerblad P, Sigvardsson M, Rosen ED (2007) Critical role for Ebf1 and Ebf2 in the adipogenic transcriptional cascade. Mol Cell Biol 27:743–757. https://doi.org/10.1128/MCB.01557-06

Jin S et al (2014) Ebf factors and MyoD cooperate to regulate muscle relaxation via Atp2a1. Nat Commun 5:3793. https://doi.org/10.1038/ncomms4793

Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9:286–298. https://doi.org/10.1093/bib/bbn013

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066

Kieslinger M et al (2005) EBF2 regulates osteoblast-dependent differentiation of osteoclasts. Dev Cell 9:757–767. https://doi.org/10.1016/j.devcel.2005.10.009

Kozmik Z, Holland ND, Kalousova A, Paces J, Schubert M, Holland LZ (1999) Characterization of an amphioxus paired box gene, AmphiPax2/5/8: developmental expression patterns in optic support cells, nephridium, thyroid-like structures and pharyngeal gill slits, but not in the midbrain-hindbrain boundary region. Development 126:1295–1304

Kuraku S, Meyer A, Kuratani S (2009) Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? Mol Biol Evol 26:47–59. https://doi.org/10.1093/molbev/msn222

Lara-Ramirez R, Patthey C, Shimeld SM (2015) Characterization of two neurogenin genes from the brook lamprey lampetra planeri and their expression in the lamprey nervous system. Dev Dyn. https://doi.org/10.1002/dvdy.24273

Li S, Yin M, Liu S, Chen Y, Yin Y, Liu T, Zhou J (2010) Expression of ventral diencephalon-enriched genes in zebrafish. Dev Dyn 239: 3368–3379. https://doi.org/10.1002/dvdy.22467

Lin H, Grosschedl R (1995) Failure of B-cell differentiation in mice lacking the transcription factor EBF. Nature 376:263–267. https://doi.org/10.1038/376263a0

Lopez-Bendito G et al (2006) Tangential neuronal migration controls axon guidance: a role for neuregulin-1 in thalamocortical axon navigation. Cell 125:127–142. https://doi.org/10.1016/j.cell.2006.01.042

Louis A, Nguyen NT, Muffato M, Roest Crollius H (2015) Genomicus update 2015: KaryoView and MatrixView provide a genome-wide perspective to multispecies comparative genomics. Nucleic Acids Res 43:D682–D689. https://doi.org/10.1093/nar/gku1112

Malgaretti N et al (1997) Mmot1, a new helix-loop-helix transcription factor gene displaying a sharp expression boundary in the embryonic mouse brain. J Biol Chem 272:17632–17639

Mazet F, Masood S, Luke GN, Holland ND, Shimeld SM (2004) Expression of AmphiCoe, an amphioxus COE/EBF gene, in the developing central nervous system and epidermal sensory neurons. Genesis 38:58–65. https://doi.org/10.1002/gene.20006

Mazet F, Hutt JA, Milloz J, Millard J, Graham A, Shimeld SM (2005) Molecular evidence from *Ciona intestinalis* for the evolutionary origin of vertebrate sensory placodes. Dev Biol 282:494–508. https://doi.org/10.1016/j.ydbio.2005.02.021

Mehta TK et al (2013) Evidence for at least six Hox clusters in the Japanese lamprey (Lethenteron japonicum). Proc Natl Acad Sci U S A 110:16044–16049. https://doi.org/10.1073/pnas.1315760110

Murakami Y, Ogasawara M, Sugahara F, Hirano S, Satoh N, Kuratani S (2001) Identification and expression of the lamprey Pax6 gene: evolutionary origin of the segmented brain of vertebrates. Development 128:3521–3531

Murakami Y, Ogasawara M, Satoh N, Sugahara F, Myojin M, Hirano S, Kuratani S (2002) Compartments in the lamprey embryonic brain as revealed by regulatory gene expression and the distribution of reticulospinal neurons. Brain Res Bull 57:271–275

Murakami Y, Pasqualetti M, Takio Y, Hirano S, Rijli FM, Kuratani S (2004) Segmental development of reticulospinal and branchiomotor neurons in lamprey: insights into the evolution of the vertebrate hindbrain. Development 131:983–995. https://doi.org/10.1242/dev.00986

Nakatani Y, Takeda H, Kohara Y, Morishita S (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res 17:1254–1265. https://doi.org/10.1101/gr.6316407

Osorio J, Mazan S, Retaux S (2005) Organisation of the lamprey (Lampetra fluviatilis) embryonic brain: insights from LIM-homeodomain, Pax and hedgehog genes. Dev Biol 288:100–112. https://doi.org/10.1016/j.ydbio.2005.08.042

Pang K, Matus DQ, Martindale MQ (2004) The ancestral role of COE genes may have been in chemoreception: evidence from the development of the sea anemone, *Nematostella vectensis* (Phylum Cnidaria; Class Anthozoa). Dev Genes Evol 214:134–138. https://doi.org/10.1007/s00427-004-0383-7

Parker HJ, Bronner ME, Krumlauf R (2016) The vertebrate Hox gene regulatory network for hindbrain segmentation: evolution and diversification: coupling of a Hox gene regulatory network to hindbrain segmentation is an ancient trait originating at the base of vertebrates. BioEssays 38:526–538. https://doi.org/10.1002/bies.201600010

Pozzoli O, Bosetti A, Croci L, Consalez GG, Vetter ML (2001) Xebf3 is a regulator of neuronal differentiation during primary neurogenesis in Xenopus. Dev Biol 233:495–512. https://doi.org/10.1006/dbio.2001.0230

Putnam NH et al (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071. https://doi.org/10.1038/nature06967

Rajakumari S et al (2013) EBF2 determines and maintains brown adipocyte identity. Cell Metab 17:562–574. https://doi.org/10.1016/j.cmet.2013.01.015

Razy-Krajka F, Lam K, Wang W, Stolfi A, Joly M, Bonneau R, Christiaen L (2014) Collier/OLF/EBF-dependent transcriptional dynamics control pharyngeal muscle specification from primed cardiopharyngeal progenitors. Dev Cell 29:263–276. https://doi.org/10.1016/j.devcel.2014.04.001

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Schubert M, Holland ND, Escriva H, Holland LZ, Laudet V (2004) Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). Proc Natl Acad Sci U S A 101:10320–10325. https://doi.org/10.1073/pnas.0403216101

Shimeld SM, Donoghue PC (2012) Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). Development 139:2091–2099. https://doi.org/10.1242/dev.074716

Singh PP, Arora J, Isambert H (2015) Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes. PLoS Comput Biol 11:e1004394. https://doi.org/10.1371/journal.pcbi.1004394

Siponen MI et al (2010) Structural determination of functional domains in early B-cell factor (EBF) family of transcription factors reveals similarities to Rel DNA-binding proteins and a novel dimerization motif. J Biol Chem 285:25875–25879. https://doi.org/10.1074/jbc.C110.150482

Smith JJ, Keinath MC (2015) The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome Res 25:1081–1090. https://doi.org/10.1101/gr.184135.114

Smith JJ et al (2013) Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat Genet 45:415–421, 421e411–412. https://doi.org/10.1038/ng.2568

Sugahara F, Aota S, Kuraku S, Murakami Y, Takio-Ogawa Y, Hirano S, Kuratani S (2011) Involvement of Hedgehog and FGF signalling in the lamprey telencephalon: evolution of regionalization and dorsoventral patterning of the vertebrate forebrain. Development 138:1217–1226. https://doi.org/10.1242/dev.059360

Sugahara F et al (2016) Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain. Nature 531:97–100. https://doi.org/10.1038/nature16518

Tahara Y (1988) Normal stages of development in the lamprey, Lampetra-reissneri (Dybowski). Zoological Science 5:109–118

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

Treiber N, Treiber T, Zocher G, Grosschedl R (2010) Structure of an Ebf1:DNA complex reveals unusual DNA recognition and structural homology with Rel proteins. Genes Dev 24:2270–2275. https://doi.org/10.1101/gad.1976610

Tulenko FJ et al (2013) Body wall development in lamprey and a new perspective on the origin of vertebrate paired fins. Proc Natl Acad Sci U S A 110:11899–11904. https://doi.org/10.1073/pnas.1304210110

Wang MM, Reed RR (1993) Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. Nature 364:121–126. https://doi.org/10.1038/364121a0

Wang SS, Tsai RY, Reed RR (1997) The characterization of the Olf-1/EBF-like HLH transcription factor family: implications in olfactory gene regulation and neuronal development. J Neurosci 17:4149–4158

Wang SS, Betz AG, Reed RR (2002) Cloning of a novel Olf-1/EBF-like gene, O/E-4, by degenerate oligo-based direct selection. Mol Cell Neurosci 20:404–414

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699