UNIVERSIDADE DE LISBOA
FACULDADE DE CIENCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

The Eph /ephrin gene family in the European Amphioxus – an Evo-Devo Approach

Stéphanie Correia de Matos David Bosne

Mestrado em Biologia Evolutiva e do Desenvolvimento

2009/2010

UNIVERSIDADE DE LISBOA
FACULDADE DE CIENCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

The Eph /ephrin gene family in the European Amphioxus – an Evo-Devo Approach

Stéphanie Correia de Matos David Bosne

Dissertation to obtain the Master's degree under the supervision of:
Dr. Jordi Garcia Fernàndez, Facultat de Biologia de la Universitat de Barcelona
Dr. Elio Sucena, Faculdade de Biologia da Universidade de Lisboa

Mestrado em Biologia Evolutiva e do Desenvolvimento
2009/2010

INDEX

Resumo 4
Abstract. 7
1. State of the Art 8
1.1.An Evo-Devo Approach 9
1.2.The Model Organism..10
1.3. Amphioxus – Biology and Ecology..12
1.4. Ephrin receptors and ligands - roles & structure 14
2. Objectives. 18
3. Materials and Methods. 19
3.1. Annotation of the Ephrin ligands and receptors..20
3.2. Amplification, Cloning and Sequencing of *Branchiostoma lanceolatum* Ephrin receptors and ligands..21
3.3. Ribosynthesis..21
3.4. Phylogenetic Analysis 22
3.5. Laboratory Maintenance of Amphioxus, Spawning and Embryos Collection 22
3.6. Whole-mount *In situ* Hybridization 23
4. Results. 19
4.1. phylogenetic analysis of the Ephrin gene family (Receptors and Ligands) 24
4.1.1. Nomenclature and structure of the protein of Ephrin Receptors and ligands 24
4.1.2. Phylogenetic analysis of the Eph Receptors 26
4.1.3. Phylogenetic analysis of the efn ligands..27
4.2. Expression of Ephrin receptors and ligands in the European Amphioxus, *Branchiostoma lanceolatum*..28
4.2.1. Expression pattern of the Ephrin receptors..28
4.2.2. Expression pattern of the ephrin ligands..29
5. Discussion. 31
5.2. phylogenetic analysis of the Ephrin gene family (Receptors and Ligands) 31
5.1.1. Nomenclature and structure of the protein of Ephrin Receptors and ligands 31
5.1.2. Phylogenetic analysis of the Eph Receptors 32
5.1.3. Phylogenetic analysis of the efn ligands..34
5.2. Expression of Ephrin receptors and ligands in the European species *Branchiostoma lanceolatum* 35

6. Conclusion. 37
7. Acknowledgments..38
Bibliography 39
Supplementary Data. 43

RESUMO

A Evolução e o Desenvolvimento (Evo-Devo) é uma area da biologia que tem por objectivo o estudo e a interpretação de conhecimentos de ambas as areas de evolução e da biologia do desenvolvimento. Tenta assim, explicar e testar as teorias evolutivas ao nivel morfologico. (David 2001)

A transição de invertebados para vertebrados é um marco importante na historia evolutiva das especies. A relação filogenética no Filo Chordata é importante para compreender esta transição. O Filo é constituido de tres subfilos Vertebrata, Cephalocordata e Urochordata. As caracteristicas que os une neste Filo é a presença de uma notocorda, um tubo nervoso dorsal, tubo nervoso dorsal, fendas branquiais, um endostélio e uma cauda pós-anal, em pelo menos uma fase de sua vida. Contudo, a relação filogenética dentro do Filo Chordata foi controversa. No século XIX os urocordados foram colocados na base da filogenia do Filo, estando deste modo, os cefalocordados mais proximos dos vertebrados (Fig.1). No entanto, dados moleculares recentes levaram a reversao destas posições estipulando que os cefalocordados se separam antes, localizando-se basalmente ao grupo Urochordata-Vertebrata (Fig.1) (Delsuc 2006).

O Filo Cephalochordata compreende dois generos (*Branchiostoma* and *Epigonichtys*). O anfioxo (*Branchiostoma floridae* - especie americana - *Branchiostoma lanceolatum* -especie europeia) possede uma faringe com fendas branquiais, uma cauda pos anal, um sistema circulatorio ausente de coração e um sistema excretor rudimentar (Fig.2). Contudo carecem de algumas caracteristicas de vertebrado como as células migratorias da crista neural, um endoesqueleto, um cérebro regionalizado e orgãos sensorias pares.

Por apresentar um genoma não duplicado, uma copia unica para a maioria das familias multigenicas dos vertebrados, pelas suas caracteristicas morfologicas e do desenvolvimento transitorias aos vertebrados, pela disponibilidade de ferramentas de manipulação génica e vantagens de manutenção em laboratorio e do genoma de *Branchiostoma floridae* estar completamente sequenciado, o anfioxo é considerado como sendo um bom organismo modelo para estudar a transição de invertebrado a vertebrado.

Em 1987, o primeiro receptor de efrina foi clonado. Os receptors (Eph) e ligandos (efn) de efrinas formam a maior das 14 subfamilias de receptores do tipo tirosina kinase. Estas possuem um dominio extracellular N-terminal que permite a interacção com o ligando; um dominio intracelular com funçao de kinase; um dominio SAM; e um dominio PDZ (Fig.4). Os ligandos estão subdivididos em dois grupos, as efnA que se encontram ancoradas à membrana via GPI (glycosylphosphatidylinositol) e as efn B que estão ancoradas por um dominio transmembranar (Fig.4). Os receptores EphA interagem preferencialemente com as efnA e as EphB com as efnB. Esta interacção ligando-receptor tem um papel crucial em processos celulares como a adesão, comunicação, rearranjos de citoesqueleto, divisão, migração, a activação de vias de sinalização citoplasmaticas que promovem a expressão génica. Deste modo, estão implicados em processos de desenvolvimento como a

migração das células da crista neural, segmentação e somitogénese, formação de sinapses e proriedades promotoras de tumores.

Os receptores e ligandos de efrinas estão presents em todos os Metazoa (de esponjas a vertebrados) em numero (tab. 1), padrões de expressão e funções distintas. Varias das funções desempenhadas pelas efrinas em vertebrados são também encontradas em invertebrados. Foi sugerido que as funções das efrinas teriam evoluido de desempenhar um papel mais simples em processos cellular, ainda mantido em grupos mais basais na filogenia, para papeis mais versateis e diversificados, como os observados em cordados.

Neste projecto propusemos um estudo de Evo-Devo da familia de genes das efrinas, pelo seu importante papel no desenvolvimento e pela sua presença em todas as espécies de Metazoa estudadas até a data, usando como modelo experimental o anfioxo europeu, devido à sua posiçao filogenética e caracteristicas. Com o objecctivo de melhor compreender a historia evolutiva e as funções durante o desenvolvimento deste genes os principais passos desenvolvidos foram: i) anotar e descrever os receptores e ligandos de efrinas em *Branchiostoma floridae;* ii) clonar e sequênciar os genes de efrina na especie europeia *Branchiostoma lanceolatum ;* iii) elucidar a relação filogenética da familia génica das efrinas em genomas de Metazoa disponiveis (Tab.1) ; iv) determinar os padroes de expressão, em diversos estadios de desenvolvimento, em ambos ligandos e receptores de efrina em *Branchiostoma lanceolatum* por hibridação *in situ*.

A relação filogenética dos receptores de efrinas esta descrita na Fig.9 desta dissertação e foi realizada com sequências de proteinas no programa Mega4. Nesta analise pode-se verificar que as Eph de vertebrados agrupam-se monofiléticamente juntas. Os genes de cefalocordados estão basais aos de vertebrados, agrupando-se com os dos não-cordados. As Eph de urocordados formam um grupo polifilético, sugerindo a ocorrência de duplicações independentes nesta linhagem, estando CiEphE, CiEphD e CiEphG mais proximos dos vertebrados. Os receptores de efrina dos Cefalocordados parecem ter divergido mais cedo do que os receptores de efrina de urocordados CiEphE, CiEphD and CiEphG, estando de acordo com a topologia proposta por Delsuc (2006) para o Filo Chordata. Os genes de *Nematostella vectensis* agrupam-se com o outgrupo *Ephydatia fluviatilis,* sugerindo que esta especie esteve sujeita a duplicaçoes independentes.

A topologia obtida para os ligandos de efrina esta representada na Fig.10. É de notar que os ligandos dos vertebrados formam dois grupos monofiléticos distintos, um que agrupa as efnA e outro as efnB. As efnA de vertebrados parecem estar filogenéticamente mais proximas das efnA de urocordados (CiefnAa, CiefnAb, CiefnAc and CiefnAd) e efnB mais proximas das efn de cefalocordados (Bfefn1, Bfefn2 e Bfefn3). Tal como observado com os receptors, os ligandos dos não-cordados formam um grupo monofilético.

O estudo da expressão dos receptores (BlEph1 e BlEph2) e ligandos (Blefn1 e Blefn2) de efrina foi feita pela técnica de hibridação *in situ* em embriões de *Branchiostoma lanceolatum* durante varios estadios de desenvolvimento (morulas de 32 celulas até larvas de 60horas). No estadio de morula todos os genes expressam-se no embrião inteiro.

Durante a neurulação BlEph1 expressava-se na mesoderme lateral, nos somitos em formação e possivelmente no precursor da vesicular cerebral. Na neurula tardia encontrava-se expresso na notocorda e no tubo neural e na future região oral. No estadio de *pré-boca* expressava-se na notocorda anterior e posterior, na boca e faringe em formação. No estadio de larva este gene expressava-se na notocorda anterior e posterior, na vesicular cerebral, na boca, no endostélio, na glandula *club-shaped*, no diverticulum. (Fig.11)

Durante a gastrulação, BlEph2 parecia expressar-se em todos os tecidos. Um padrão segmentado parecia aparecer durante a neurulação que parece seguir o padrão dos somitos em formação. No estadio de *pré-boca* BlEph2 expressava-se na boca em formação e na notocorda anterior. Nas larvas a expressão deste gene parecia localizar-se na boca e na região da faringe. (Fig.12)

O padrão de expressão dos ligandos de efrina (Blefn1 e Blefn2) parecia estar de acordo com o padrão observado nos receptores. Em gastrulas Blefn1 expressava-se maioritariamente na mesoderme justo ao blastoporo. Durante a neurulação a expressão era mesodermica, junto ao fecho do tubo neural e também parecia expressar-se em populacões de neuronios. No estadio *pré-boca* parecia expressar-se na futura região oral, na notocorda posterior e a endoderme posterior. No estado de larva a expressão parecia mais restringida a zona da boca e da faringe, mas ligeiramente expresso nos restantes tecidos, exceptuando a vesicula cerebral. (Fig.13)

Durante a gastrulação Blefn2 parecia expressar-se na gastrula inteira. Durante a neurulação a expressão parecia restringida a mesoendoderme, exceptuando o dominio mais posterior. Em embriões *pré-boca* parecia localizar-se na região da faringe em formação, na glandula de muco e na notocorda posterior. Em larvas a expressão parecia localizar-se na região da boca e da faringe e também na notocorda posterior (Fig.5F).

Na literatura esta descrito que as efrinas expressam-se em cordados nas zonas dos somitos em formação, na notocorda, no tubo neural, na mesoderme paraxial e nos nervos perifericos, como demonstrado pelas imagens obtidas por hibridação *in situ* (Fig.11 a 14). Em vertebrados estes genes também participam na formação de varias estruturas faciais incluido a região oral e os seus padrões de inervação, padrão também observado nos receptores e ligandos de efrinas de *Branchiostoma lanceolatum* sugerindo a co-opção destes genes para esta função durante o desenvolvimento em Chordados (Fig.11-14).

Com este trabalho foi possivel elucidar algumas das questoes sobre a expressão e a posição filogenética dos membros da familia genica das efrinas no anfioxo europeu (*Branchiostoma lanceolatum*). A estrutura destes genes parece ser conservada desde as esponjas aos vertebrados, mas também as funções desempenhadas durante no desenvolvimento. Entre estes incluem-seos processos a nivel cellular que, ao longo da evolução sofreram fenomenos de complexificação, na transição para cordados. Adicionado a estas observações, a topologia filogenética sugere fenomenos de co-opção para as novas funçoes de duplicações especificas de linhagem. Existem também evidências para convergência evolutiva entre espécies e de evolução paralela dentro da mesma espécie. Os cefalocordados parecem ocupar, relativamente a esta analise, uma posição entre os não-cordados e os vertebrados o que suporta a nova filogenia sugerida por Delsuc *et al* em 2006. No geral, pode-se sugerir que os genes da familia das efrinas diversificaram varias vezes ao longo da evolução dos

Metazoa e também que o ancestral dos urocordados e vertebrados apresentaria um unico receptor de efrina e dois ligandos. Contudo, dados adicionais sobre as funções, padrões de expressão génica e filogenia dos Metazoa seriam necessarios para melhor estabelecer uma relação evolutiva da familia génica das efrinas e o seu papel durante a evolução.

Palavras-chave: *Branchiostoma lanceolatum*, efrina, Eph, hibridacao *in situ*, filogenia

ABSTRACT

The Ephrin receptor and ligand gene families are implicated in several cellular processes such as cellular adhesion, communication, division, migration, and compartmentalization. These play an important role in development including, for instance, neural crest cell migration, somitogenesis, axon guidance, and have even been shown to have tumor promoting properties. They are described to be present from sponges to vertebrates in multiple copies, maintaining a conserved genomic structure. Cephalochordates, recently placed at the base of the Chordata, are considered to be the living animal that best approximates the ancestor at the transition from invertebrates to vertebrates. The phylogenetic analysis of this project shows that Eph and efn families have been independently expanded in amphioxus, vertebrate, and other Metazoan clades suggesting that each have convergently evolved complex Eph/efn complements. In the European amphioxus (Branchiostoma lanceolatum) the Eph/ephrin gene family seems to express mostly in the forming mouth apparatus, somites and notochord, as assessed by whole-mount in situ hybridization (ISH). The Eph/ephrin ISH expression pattern corresponds to some vertebrate characteristics for these genes, but they are also implied in several cellular processes similarly to invertebrates. In spite of the evidence, it is difficult to assess the exact evolutionary relationship and developmental role of these genes in amphioxus with these data, or to extrapolate to Metazoa.

Key-words: *Branchiostoma lanceolatum*, ephrin, Eph, whole-mount *in situ* hybridization, phylogeny

1. STATE OF THE ART

1.1. AN EVO-DEVO APPROACH

Evo-Devo is the field of biology that studies and assembles knowledge from both evolutionary and developmental biology, and tries to explain and test the evolutionary theories at the morphological level. Genotype is generally translated into the phenotype via the process of development. As such, within one generation, the genes are exposed to evolutionary forces and selection may sort the variability that arises. Like this, these changes are printed in the genome, are susceptible of being inherited, and the characteristics evolve through the generations. (David 2001)

To interpret evolutionary and developmental changes it is important to take into account the patterns and means by which these can occur, such as gene loss or novelty, convergence, and canalization. These may be originated by processes such as co-option, gene duplication, and heterochrony. These modifications are confined within biological limits, constraints in their several forms, as they are required to respect modularity at the different levels of interactions and pathways. However, there exist constraints underlying this process. These can be due to the historical background, compromises to normal development and/or lack of genetic variability. Then, a characteristic that is impracticable or that compromises the survival of an individual, being disadvantageous, cannot therefore be viably maintained. (von Dassow 1999, Hall 2003, Zhang 2003)

Several research lines point to the existence of conserved gene pathways, suggesting that evolution occurs through "tinkering" of molecules and pathways previously existent, rather than by creating absolute "novelties" (Jacob 1977). These conserved pathways, integrated into modules, are re-utilized due to *cis*-regulatory modification that control both spatial and temporal patterning, via processes like, for example, co-option or heterochrony. One of the best studied examples is the Hox gene cluster, which conserves its relative positions in body plan patterning and functional equivalence, the physical co-localization of the Hox genes in the genome (synteny), and protein structure across animal groups. In some organisms, such as vertebrates, the cluster has undergone gene duplication. (Garcia-Fernàndez 2009)

Gene duplication can be another process by which variability is created and upon which selection can act. By neofunctionalization of the duplicated gene, one copy is maintained allowing the new one to freely evolve a new function without constraints. In this way the ancient function is maintained, and the relaxation of the purifying selection can allow the occurrence of mutations. Subfunctionalization allows the sharing of the ancestral function by both duplicates, such that this expression can be regulated in space and time. However, these processes lead to problems of dosage decompensation until the duplicates diverge or the ancestral

function is partitioned. In any case the duplicated gene can undergo degeneration of one of the copies over time. (Force 1999, Lynch 2000, Kondrashov 2002, True 2002, Zhang 2003, Carroll 2005)

On a macroevolutionary scale, gene duplication may be an important process. As changing development output, it creates important modifications that lead to overcoming some of the steps achieved by small evolutionary changes, leading to speciation and evolutionary jumps. Microevolution generates population differences, so called polymorphisms, and their subsequent adaptation (Decaestecker 2007). These are characterized by changes that occur over relatively short periods of time, which includes changes in allele frequencies, genotypic composition or gene expression, and which can occur either within or between populations (Luikiart 2003).

The comparative method is used to study evolutionary processes and, as the name suggests, it is preceded by the observation of the so said ancestral state and further comparison with the derived state of the features. Because evolution does not occur in a straight line, it is also important to understand the steps that were required to lead to the derived state of the feature, in order to determine either if it is a parallel or convergent evolution when compared with the same characteristic in other organisms, for example. The problem of this kind of approach is that one needs to screen and closely follow the evolutionary process of the group, which is often impossible when studying extant groups, which may lack a fossil record or have anciently diverged. (Collier 1993)

When facing the interpretation of evolutionary patterns one can be taken to determine the process by which certain traits appeared. Similar features that are present in different species can be have appeared by convergent evolution or parallel evolution. In the first case the same biological trait would have appeared independently in two different unrelated species; whilst in the second case, two related species, descending from a common ancestor, would have acquired the same trait. Another pattern can be defined as co-evolution and is defined as a change of a biological trait triggered by the change of a related trait, such as described in evolving host-parasite interaction (Mitchell 1991, Zang 1997, Yip 2008). It is also important to define if either the genes are analogous or homologous, and in this last category if they are orthologous or paralogous genes to be able to construct a robust evolutionary relationship and annotation. Homologous genes are sequences that share a common ancestor and they can be either orthologous (when originating from a single ancestral gene of the last common ancestor) or paralogous (when originated from duplication events) (Koonin 2005). To be able to answer a determined biological question, it is important to select the model organism that for one possesses traits suitable for the testing of the hypothesis, such as phylogenetic position, and that is also amenable to experimental manipulation.

1.2. THE MODEL ORGANISM

The transition from invertebrates to vertebrates is an important mark in the evolution of species. It is important to consider the evolutionary patterns and processes mentioned above to be able to understand how

it occurred, the steps required, to analyze the changes and novelties that emerged at the origin of this group. (Graham 2000, Garcia-Fernàndez 2008)

The phylogenetic relationship within the chordate phylum is important to understand this transition although it is controversial (Garcia-Fernàndez 2008). The phylum is comprised of three subphyla: Vertebrata, Cephalochordata and Urochordata. The common characteristics that include these groups in the Chordata phylum are the presence of a notochord, a hollow dorsal nerve cord, pharyngeal gill slits, endostyle, and post-anal tail. The adult of urochordates appears to be secondarily less complex than chordates, possessing a large notochord flanked by muscle blocks in the tail region, and a dorsal nervous system composed by relatively few neurons. Cephalochordates (amphioxus) on the other hand seem to be more complex organisms as they possess lateral muscles organized in somites, a nervous system and a pharyngeal region with a more complex structure. The discovery of the notochord in this organism united the ascidians, amphioxus and vertebrates in the same phylum Chordata (Sharman 1999, Graham 2000, Garcia-Fernàndez 2009). The vertebrates are the most intricate since they have a highly structured forebrain, external sensory organs, skeletal support, dorsal root ganglia, paired limbs, lateral muscle blocks, regionalized anterior brain, and neural crest cells, among other features. All of these are the fruit of the acquisition of novel modifications in the developmental program during the origin of the vertebrates (Graham 2000).

Near the late 19th century it was established that either cephalochordates or urochordates were located at the base of the phylogenetic tree of chordates (Kowalevsky 1867). Urochordates possess an unduplicated and smaller genome than amphioxus due to genetic loss (Seo 2004) and, along with its simplified morphology, made it difficult to establish homologies between its structures and vertebrates. From there, it was proposed that urochordates would be the most basal branch of chordate group and that the ancestor would have been a sessile animal similar to an ascidian adult. Amphioxus and vertebrates would therefore be neotonenic (Sly 2003). However, recent molecular data led to the reversing of the positions stipulating that cephalochordates branched earlier, localizing basally to Urochordata-Vertebrata clade (Delsuc 2006). Currently the ancestral chordate is thought to have been a free-living worm or fish-like shaped organism, with a series of pharyngeal gill slits. This implies that the origin of vertebrates may have not occurred by neotony of a urochordate-like sessile animal. These data also indicate that urochordates and echinoderms have a derived body plan, whereas amphioxus and vertebrates may have retained body plans more comparable to the ancestral deuterostome (Delsuc 2006, Garcia-Fernàndez 2008, Kaltenbach 2009). In this way, it was established that tunicates are the closest extant invertebrates to vertebrates. Illustrative examples are the presence of a vertebrate-like midbrain-hindbrain boundary, based upon the expression of Pax2/5/8, and the existence of migratory neural crest-like cells in tunicates (Jeffery

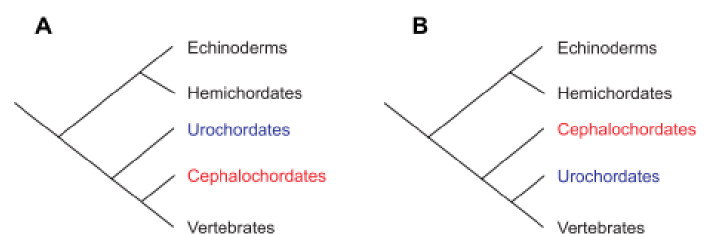

Fig. 1 - Deuterostome phylogeny, classical and modern topology. (A) Classical phylogeny before the proposition of Delsuc and colleagues 2006, Cephalochordates are a sister group of vertebrates, and Urochordtaes are basal to Chordates. (B) Present phylogeny in which Cephalochordates are basal to the Chordates and Urochordates closer to Vertebrates. Adapted from Garcia-Fernàndez et al, 2008.

2004, 2006). The absence of such features in amphioxus (Kozmik 1999, Garcia-Fernàndez 1994) could only be explained by secondary loss, based on the old phylogeny. The new phylogeny provides a more rational sequence of events in the generation of these and other key traits of vertebrates, by counting fewer secondary losses in amphioxus and considering the ascidian characteristics to be derived traits. As Grove and Newell (1961) said: "If amphioxus had not been discovered, it would have to have been invented". (Garcia-Fernàndez 2009)

All these suppositions are concordant with the ideas of Susumu Ohno (1970), who proposed that the vertebrate's genome is the result of two full genome duplications (polyploidization) – ideas that latter evolved into the 2R hypothesis. These are strongly supported by molecular data analysis such as the ones made in the Hox cluster and other conserved regulatory genes in amphioxus, (Garcia-Fernàndez 1994, Holland 1994) together with the sequencing of the Human genome (*Int. J. Human Genome Consortium 2001*) and its first synteny analyses (Lundin 1993). Amphioxus have many single members for most vertebrate multigenic families, as they diverged prior to the whole genome duplications characteristic of vertebrates. As genetic network conservation suggests, amphioxus possess many genes that are common to chordate biological processes, which suggests a co-option in vertebrates for new roles for these genes. (Holland 2008) Due to this and also to the exon-intron distribution and chromosomal organization, together with morphological body plan and embryonic development, amphioxus is a good candidate for the living ancestor of chordates. Amphioxus possess other additional vertebrate-like features (e.g. kidneys, notochord and segmental paraxial muscles) (Sharman 1999, Graham 2000, Garcia-Fernàndez 2008 and 2009).

After the two rounds of genome duplication, nearly 20–25% of the duplicates were kept in the vertebrate genome, and mostly biased towards the retention of genes involved in transcriptional regulation, signal transduction, development and neuronal processes. This was probably due to subfunctionalization (Force 1999) of genes with complex regulatory regions, or perhaps the doublings also produced genomic flexibility in duplicated coding and *cis*-regulatory sequences, facilitating the emergence of morphological and physiological novelties in vertebrates – neofunctionalization (Garcia-Fernàndez 2009).

However, the amphioxus may be a good candidate for studying the transition from invertebrates to vertebrates; its genome has not stopped in time. It continued to evolve, like all other organisms, until the present. It has suffered all kinds of alterations to its development including secondary gain and loss of genes and even independent duplications (examples in Minguillon 2003 and D'Aniello 2008).

1.3. Amphioxus – Biology and Ecology

The amphioxus was first described by Peter Simon Pallas in 1771 as a mollusk. Only in 1834, Gabriel Costa positioned the amphioxus phylogenetically near agnathan vertebrates. However, in his classification he confused the oral tentacles and described them as mouth gills, thus naming these creatures *Branchiostoma*

(branchio=gills, stoma=mouth), the genus name that persists until nowadays. Johannes Müller (1841) was the first to describe them as chordates, classification maintained since then. (Garcia-Fernàndez 2009)

The name "amphioxus" can find its origins in the Latin, meaning "sharped at both ends" (*amphis*=both; *Oxys*=sharp). (Garcia-Fernàndez 2009) Its particularly elongated shape is due to the presence of a dorsal notochord localized along the anterior-posterior axis, exceeding the length of the neural tube in the head. This unique characteristic gave the name to the subphylum in which it is grouped, the Cephalochordata (*kephale*=head, *khorde*=chord), which includes 29 species classified in two genera (*Branchiostoma* and *Epigonichtys*). It is also the only support structure that these animals possess and it is maintained through adulthood, unlike vertebrates, and is constituted of a particular type of muscle cells. They possess neurochordal synaptic contacts along the anterior-posterior axis, but not associated to ganglia, unlike vertebrates. Like other chordates they possess a perforated pharynx with pharyngeal slits (gill slits), a post-anal tail, a circulatory system deprived of heart, the functions of which are replaced by large contractile blood vessels, and a rudimentary excretory system (Fig.2) (Holland 2008). They lack some key characteristics of vertebrates such as migratory neural crest cells, an endoskeleton, a highly regionalized brain, and paired sense organs (Shimeld 2000, Garcia-Fernàndez 2009). Amphioxus have separated sexes that are distinguished by the color of their lateral-ventral gonads and the morphology of the gametes in the mature gonads. The embryonic development is very similar to the one of echinoderms and tunicates, since fecundation to blastulation occurs via repeated radial holoblastic cleavage. (Garcia-Fernàndez 2009, Holland 1996) However, the subsequent developmental stages are more like those of a vertebrate. They are characterized by the appearance of the dorsal nerve cord, the notochord and the somites during neurulation (Holland 1999). They undergo a metamorphosis when passing to the adult stage that causes redistribution of the musculature, nerves and the left-right asymmetry of the mouth structures. These migrate from the left position that they occupy during larva stage, to a median-ventral positioning in the adult. The developmental stages of the Amphioxus *Branchiostoma lanceolatum*

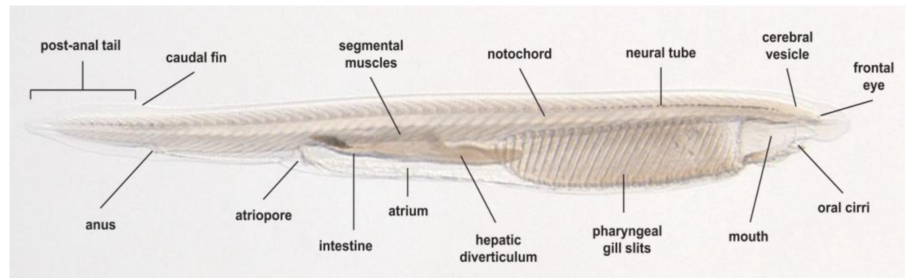

Fig.2 – Morphology of an amphioxus juvenile (*Branchiostoma lanceolatum*) highlighting the main key features. The anterior end is to the right of the picture. Adapted from Garcia-Fernandez et al, 2008.

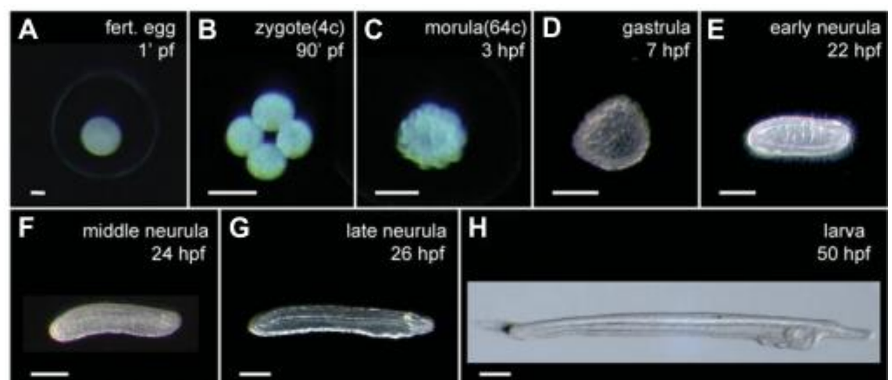

Fig.3 - Developmental stages of *Branchiostoma lanceolatum* grown at 19ºC at the facility of the Universitad the Barcelona. (A) fertilized egg; (B) zygote at four cell stage; (C) morula at 64 cells stage; (D) gastrula; (E) early neurula stage; (F) middle neurula stage; (G) late neurula stage; (H) larva stage with one gill slit. Adapted from Garcia-Fernàndez et al, 2008.

undergoes are shown in Fig.3. The adults live buried in the sand and their alimentation is mainly of unicellular algae acquired by filtration through their jawless mouth. (Gans 1996, Garcia-Fernàndez 2008)

Research in amphioxus has expanded because of the availability of genomic data and tools. The genome of *Branchiostoma floridae* is completely sequenced (Putnam 2008), and also the facility of obtaining embryos, maintenance of the animals in captivity for low cost and in limited space, constitutes several of the advantages for research in this organism. The species used in research are *Branchiostoma floridae* (American species), *Branchiostoma belcheri* (Asian species), and *Branchiostoma lanceolatum* (European species).

1.4. EPHRIN RECEPTORS AND LIGANDS - ROLES & STRUCTURE

In 1987, the first Eph receptor was cloned by Hirai and colleagues from an erythropoietin producing hepatocellular carcinoma cell line. At first they were called “blind receptors” as the nature of their ligands was unknown, as well as their specific functions (Flanagan 1998; Nakamoto 1999). Nowadays their function is mostly associated with axon guidance, as it was their first role described.

There exist about 14 subfamilies of receptor tyrosine kinases (RTKs), and the Ephrin receptors (Eph) and their ligands (ephrins) form the largest of these subfamilies. The RTKs are membrane spanning proteins with an intracellular kinase domain and an extracellular ligand binding domain, N-terminal that is highly conserved, and which is necessary and sufficient for specific ligand recognition and binding (Fig.4) (Arvanitis 2009; Holder 1998; Himanen 2003; Kalo 1999; Xu 2000). The SAM domain seems to modulate receptor dimerization, oligomerization, binding of adaptor proteins or even clustering at cell surface. Then Eph receptors and ephrins possess a binding site for proteins containing a PDZ domain (Fig.4), which are important for organization of membrane proteins and clustering; they also assemble complexes that allow interaction between components of signal transduction pathways, recruiting downstream cytoplasmic elements such as GTPases (Rac/Rho/Cdc42) and Rho kinases, which are important to control actin cytoskeleton polymerization. The activation of Eph receptors activates ten or more specific tyrosine residues within the cytoplasmic membrane.

Ephs are subdivided in two groups, the A and the B classes. The Eph A class of receptors interact preferentially with GPI (glycosylphosphatidylinositol) anchored ephrin-A ligands, whilst the Eph B class of receptors usually binds to transmembrane anchored ephrin-B ligands (Fig.4). There is some promiscuity in ligand-receptor interaction between A type ligands and B type receptors or *vice versa,* although only occur in specific contexts (Surawska 2004). They also present so-called bi-directional signaling because ephrin ligands are anchored to the cell membrane, enabling a signalization performed by both receptors and ligands in the cell to which they are anchored. Conventionally, Eph receptors act in forward whilst ephrin ligands activate a reverse signal (Gale 1997, Brückner 1998, Kalo 1999, Coulthard 2002, Drescher 2002, Kullander 2002, Himanen 2003, Blits-Huizinga 2004, Davy 2004, Poliakov 2004). Contrarily to other RTK subfamilies, Eph receptors bind to soluble monomeric ligands but they do not induce receptors phosphorylation (Kalo 1999). When the Eph receptors are activated

they induce a forward signaling event. Reverse signaling is established when the tyrosines of the cytoplasmic portion of the ephrin ligands are phospohorylated (Nakamoto 2000, Klein 2002, Kullander 2002, Himanen 2003, Blits-Huizinga 2004). Studies have revealed that both types of signaling can act independently and, therefore, be responsible for different developmental functions (Davy 2004; Poliakov 2004). Both Eph A and B perform the same general functions, so it is complicated to assess a range of functions for each receptor class. However, there exists specificity for each class of receptor, even if only at the level of spatio-temporal patterning. (Mellot 2008)

Depending on Eph activation, the extent and regulation of Eph receptor and ephrins clustering in the plasma membrane, the kind of downstream pathway they induce, the cytoskeleton modifications, or the nature of external cell-type specific factors, these molecules can trigger different developmental mechanisms, although they undergo the same type of ligand interaction and phosphorylation as other members of the RTK family. (O'Leary 1999, Schmucker 2001, Wilkinson 2001, Drescher 2002, Kullander 2002, Santiago 2002; Himanen 2003, Poliakov 2004)

The interaction between Eph receptors and ephrins only occurs at cell-cell contact regions. Eph receptors can dimerize or tetramerize in *cis* (within the same cell membrane) and mediate interactions with other dimers/tetramers of ephrin ligands, between interacting cells (Sela-Donenfeld 2005; Arvanitis 2009).

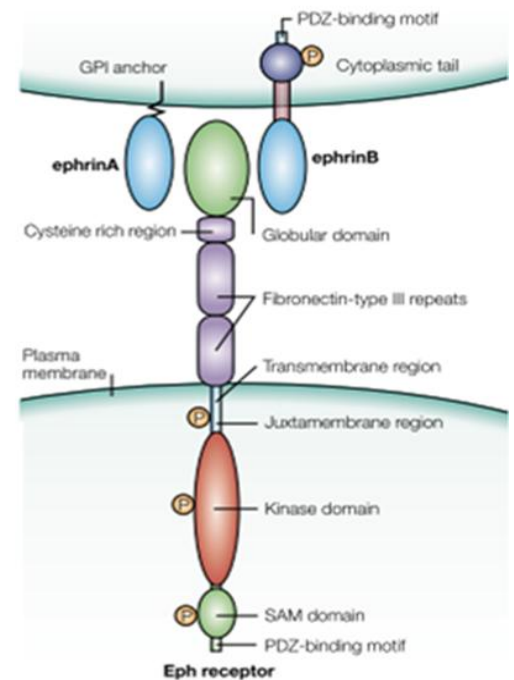

Fig. 4 – Schematic representation of the domain structure of an Eph receptor and the anchorage to the membrane of the two types of ephrins ligands. SAM, sterile alpha motif; PDZ, PDZ-binding motif; GPI, glycosylphosphatidylinositol linkage. Adapted from Kullander and Klein, 2002.

It is now known that Eph receptors and ephrins play an important role in several cellular processes. The role played by various ephrins is different and tissue specific. They are capable of interfering with processes such as cellular behavior, adhesion, communication, cytoskeleton rearrangements, divisions, migration, gap junction communication, boundary formation (by repulsive or adhesive forces by rearrangements of the actin cytoskeleton as in somitogenesis) (Coulthard 2002)., compartmentalization, and activation of downstream cytoplasmic pathways that promote the regulation of gene expression. As these are important to undertake major developmental processes, Eph receptors and ephrins are implicated in neural crest cell migration, bone formation, regulation of insulin secretion, vascular morphogenesis and angiogenesis, segmentation and somitogenesis, axon guidance, synapse formation, gradient establishment in several topographic mappings, and even tumor promoting properties. The correct execution of these processes is extremely important to obtain a correct morphogenesis of the organism. (Drescher 1995, Wang 1997, Holder 1999, Kalo 1999, Krull 1999, Drescher 2002, Davy 2005, Himanen 2007)

One of the explanations for ephrins being involved in cell adhesion, guidance and migration is that they are tightly interacting with integrins, influencing cell adhesions, and possibly having the same substrate or

functions as these in focal adhesions. However, it is still unclear what type of relation exists between these signals and the adhesions. (Drescher 2002; Arvanitis 2009) Different mediators of signal transduction in different cell types or alternatively, quantitative and qualitative differences between ephrins and Eph receptors could be at the origin of these distinct signals for repulsion or adhesion. (Wilkinson 2001)

Ephrin receptors and ligands are present in a wide variety of animals. Their number and expression patterns, as well as their function, vary from species to species. An Ephrin receptor was recently described in sponges, exhibiting the same structural domains of Eph receptors of vertebrates. This suggests that these genes must have appeared in an early branching at the origin of the Animal kingdom. However their primordial role was unlikely to have been axon guidance as they possess neither neurons nor axons, but perhaps instead might have had a function in cell movements' regulation (Suga 1999; Coulthard 2002; Drescher 2002; Mellot 2008). In invertebrate bilaterians, like *Drosophila sp*. and *C. elegans*, only a single gene is present, probably representing the ancestral form. This suggests that the large number of genes present in vertebrates is probably an innovation. In *C. elegans*, it is expressed at neural boundaries playing a role in axon guidance and in tissue segregation. In *D. melanogaster* this gene is implicated in guiding cortical axons to the optic ganglion and restricting interneuronal axons from crosssing borders. Several functions performed by vertebrate Ephs are found in invertebrates as well (Mellot 2008). The more diverse group of Eph is found in teleosts because they underwent additional whole genome duplication beyond the two duplications rounds that occurred at the base of vertebrates. In vertebrates there are a larger number of receptors and ligands and these seem to have become also expressed in all tissues during development, unlike the tissue-specifity encountered in invertebrates. For example, their expression can be found in the retina, the central nervous system, the posterior notochord, the neural tube, the neural crest cells, the forming somites, the mesoderm, and the forming vasculature (Helbling 2000, Mellitzer 2000, Naruse-Nakajima 2001, Santiago 2002, Coulthard 2002, Poliakov 2004, Davy 2005). It has been suggested that Eph receptors and ligands would have evolved from possessing a single a role in cell movements processes, that is still maintained in clades positioned lower in the phylogeny, to a more enriched, versatile and diversified roles. (Wilkinson 2001)

Phylogenetic relationships for Ephrin receptors and ligands have been established, although these are hard to ascertain between vertebrates and invertebrates, due to large evolutionary gaps. Nevertheless, Mellott and Burke (2008) published a phylogeny in which vertebrate receptors appear as a lineage-specific expansion and therefore divergence of the two families A and B would be a vertebrate innovation (Fig.5). Ephrin ligands (efn), by contrast, would have already diversified into the two groups A and B in the common ancestor of urochordates and vertebrates. Furthermore, the efn B group clustered with the ligands of other metazoans, separated from the A ephrins, suggesting that they are probably less derived and that the ancestral chordate efn, (prior to the divergence of A and B) was probably more efn-B-like in terms of its primary sequence (Fig.6). Then both the whole genome duplications and additional gene duplication events would have led to the current complexity of the Eph/efn complement. This also suggests that the ancestral Eph receptor, before diverging into A and B type, was able to interact with both ephrin ligands (already existent); with the subsequent duplication and evolutionary divergence, it would more likely interact with one of the two ephrin

types. Following this idea, it was proposed that perhaps *H. sapiens* Eph A4 and B2 are related to the ancestral Eph receptor, due to their ability to interact with both ephrin A and B ligand types. (Mellot 2008)

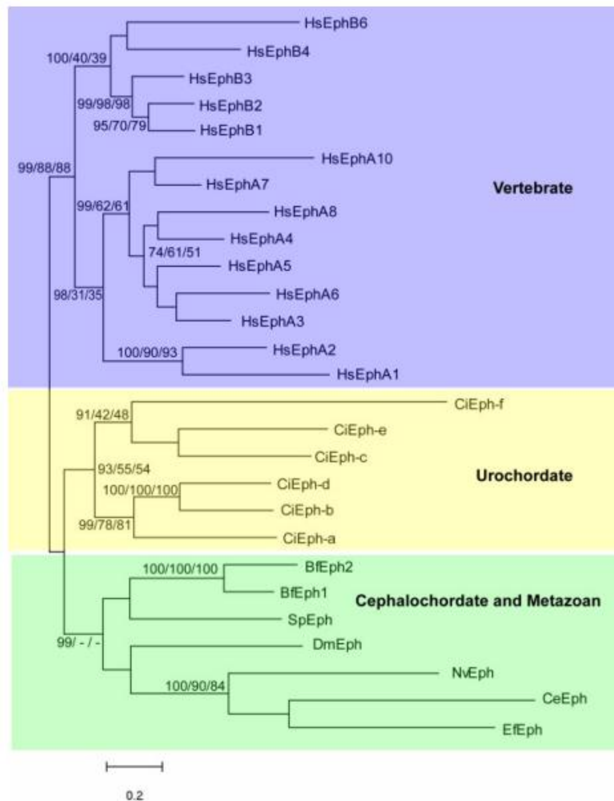

Fig.5-Unrooted phylogenetic tree for Eph receptors constructed with the Maximum Likelihood method. Numbers represent percentage bootstrap values for 1000 replicates (Maximum likelihood/Neighbour Joining/Minimum Evolution); unlabelled branches or/- indicates a value less that 50%. Three major clades are present: Vertebrate, Urochordate, and a group containing Cephalochordates, Echinoderms, Ecdysozoans, Coelenterate and sponges. Using the ecdysozoan branch to root the tree does not change the composition of the major groups. *Bf, Brachiostoma floridae; Ce, Caenorhabditis elegans; Ci, Ciona intestinalis; Dm, Drosophila melanogaster; Ef, Ephydatia fluviatilis; Hs, Homo sapiens; Nv, Nematostela vectensis; Sp, Strongylocentrotus purpuratus.* Adapted from Mellott and Burke 2008

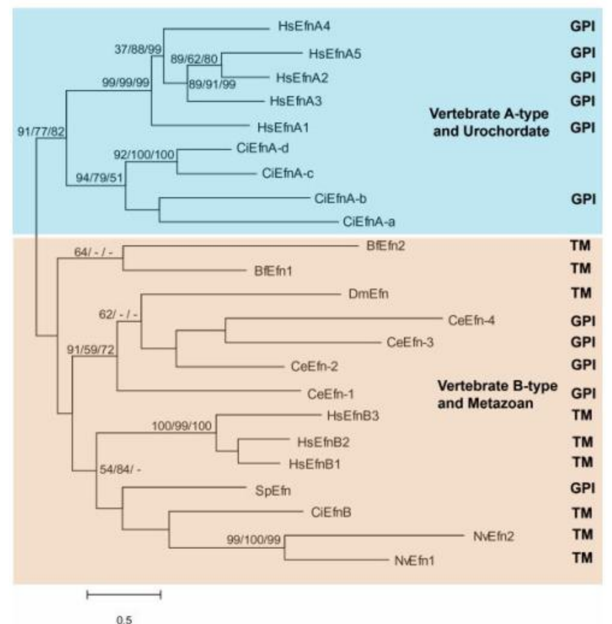

Fig.6-Phylogenetic tree for ephrin ligands constructed with Maximum Likelihood method. Numbers represent percentage bootstrap values for 1000 replicates (Maximum likelihood/Neighbour Joining/Minimum Evolution); unlabelled branches or/- indicates a value less that 50%. Two major clades are present: Vertebrate A-type and Urochordate ephrins and a group containing Vertebrate B-type, Cephalochordate, Ecdysozoan, Echinoderm, Coelenterate, Sponge and one Urochordate ephrin. Using the ecdysozoan branch to root the tree does not change the composition of the major groups. *Efn, ephrin; Bf, Brachiostoma floridae; Ce, Caenorhabditis. elegans; Ci, Ciona intestinalis; Dm, Drosophila melanogaster; Hs, Homo sapiens; Nv, Nematostella vectensis; Sp, Strongylocentrotus purpuratus. TM, transmembrane domains; GPI, glycosylphosphatidylinositol modification*. Adapted from Mellott and Burke 2008

During the evolutionary process it seems that Eph receptors and ligands have not conserved all of their roles. Perhaps due to the promiscuous binding between receptors and ligands characteristic of the group, or even because of the common downstream pathways, these roles have been less constrained. (Mellot 2008) In some species, a given Eph receptor can have the same or a distinct expression site which suggests an overlapping or synergisticity of roles in some tissues. Depending on the species, orthologous Eph receptors and ephrins can be

expressed in different tissue or at distinct developmental stages, suggesting that some members within the same class may be functionally interchangeable but biochemically similar. (Xu 2000)

The appearance of tyrosine kinase (TK) superfamily might have appeared by modular recombination of preexisting domains and regulatory pathways in the ancestors of Metazoa. The TKs would have only expanded and diversified in metazoan and choanoflagellates, their closest unicellular relatives, and present a high similarity in the kinase domain through evolution. The amphioxus seems to have kept a representing copy of each TK family present on the common ancestor of protostomes and deuterostomes, unlike vertebrates. But its genome has also undergone several large expansions such as described in the EXTK family and the RET receptor. Because of its genomic plasticity to evolve and having retained most of the gene match of its ancestor, amphioxus is considered to possess the richest TK collection amongst all metazoan studied and a good model to study evolutionary processes, at least in this gene family (d'Aniello 2008).

In this project, we propose an Evo-Devo study of a gene family that plays major roles in developmental processes – the Eph/ephrins – using amphioxus as a model organism. The members of the *Branchiostoma* genus are phylogenetically positioned at the base of the chordate phylum. Since amphioxus possesses a genome and a body plan that appear to retain the characteristics of a prototypical chordate, it represents a good system in which to study the evolutionary origin of chordate, and by extension vertebrate, novelties (Holland 2008). To better understand the evolutionary history and the developmental functions of the Eph/ephrin gene family we proposed to: construct a phylogenetic relationship of amphioxus Eph/ephrin genes with those of other Metazoa, as well as the expression pattern of the gene family by whole-mount *in situ* hybridization at several developmental stages, we should be able to gain insight into the evolutionary history of these genes in chordates, at the invertebrate-vertebrate transition.

2. OBJECTIVES

The purpose of this project was to investigate whether amphioxus could provide insight into the evolution of the transition from invertebrate to vertebrates through the analysis of the Eph/ephrin gene family. These were selected both for their implications in important developmental processes and for their presence in all Metazoa. In this context, the specific aims for this thesis were as follows:

1. Annotation and description of the Ephrin receptors (Eph) and ephrin ligands in the American species *Branchiostoma floridae*, by analysis of the genome browser http://www.jgi.doe.gov/
2. Cloning and sequencing of the Eph/efn genes in the European species *Branchiostoma lanceolatum*.
3. Elucidation of phylogenetic relationship of the Eph/ephrin gene family in available metazoan genomes.
4. Determination of the expression patterns, at several developmental stages, of both ligands and receptors of the Ephrin gene family in *Branchiostoma lanceolatum* by whole-mount in situ hybridization.

3. MATERIALS AND METHODS

3.1. Annotation of the Ephrin ligands and receptors

To find and annotate the Ephrin ligands (efn) and receptors (Eph), tBlastn and Blastn against the *Branchiostoma floridae* genome (JGI, V1.0: http://genome.jgi-psf.org/Brafl1/Brafl1.home.html) were performed, using efn and Eph sequences of *Mus muscullus*. The hits were analyzed, (confirming with ESTs (expressed sequence tags) and predicted gene models) to obtain the best gene annotation. Auxiliary programs were used to predict missing regions of the genes like GenomeScan (Yeh 2001) or GeneWise2 (Birney 2004). The final models were verified by Blast and compared with the ESTs and the genome, and also by determination of the structures of the protein domains (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Finally to discard the existence of any other gene, the models were blasted against the genome by lowering the e-value to diminish the restrictions of the analysis.

To construct a phylogeny, the orthologous ephrin ligands and receptors were annotated in other organisms by Blast against the published genomes, using genes of phylogenetically close organisms. The names of the genes were maintained according to the bibliography or by numbers if they were not annotated before (in thes speciesor in a sister species). The genes were all verified and corrected against *Branchiostoma floridae* genes. The organisms and genome browser used were:

Species	Genome browser
Homo sapiens	*http://www.ncbi.nlm.nih.gov/*
Xenopus tropicalis	*http://genome.jgi-psf.org/Xentr4/Xentr4.home.html*
Ciona intestinalis	*http://genome.jgi-psf.org/Cioin2/Cioin2.home.html*
Branchiosotma floridae	*http://genome.jgi-psf.org/Brafl1/Brafl1.home.html*
Strongylocentrotus purpuratus	*http://www.hgsc.bcm.tmc.edu/project-species-o-Strongylocentrotus%20purpuratus.hgsc?pageLocation=Strongylocentrotus%20purpuratus*
Saccoglossus kowalevskii	*http://www.hgsc.bcm.tmc.edu/project-species-o-Acorn%20worm.hgsc?pageLocation=Acorn%20worm*
Drosophila melanogaster	*http://flybase.org/*
Apis mellifera	*http://www.ncbi.nlm.nih.gov/*
Nasonia vitripennis	*http://www.ncbi.nlm.nih.gov/*
Tribolium castaneum	*http://www.ncbi.nlm.nih.gov/*
Daphnia pulex	*http://genome.jgi-psf.org/Dappu1/Dappu1.home.html*
Ixodes scapularis	*http://www.ncbi.nlm.nih.gov/*
Lottia gigantea	*http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html*
Capitella teleta	*http://genome.jgi-psf.org/Capca1/Capca1.home.html*
Nematostella vectensis	*http://genome.jgi-psf.org/Nemve1/Nemve1.home.html*
Trichoplax adhaerens	*http://genome.jgi-psf.org/Triad1/Triad1.home.html*
Ephydatia fluviatilis	*http://www.ncbi.nlm.nih.gov/*
Monosiga brevicollis	*http://genome.jgi-psf.org/Mobr1/Mobr1.home.html*

3.2. AMPLIFICATION, CLONING AND SEQUENCING OF *BRANCHIOSTOMA LANCEOLATUM* EPHRIN RECEPTORS AND LIGANDS

The PCR amplification of the Eph/efn genes was done from a cDNA library of *Branchiostoma lanceolatum* constructed by the laboratory and using Invitrogen primers (Forward and Reverse) and Promega PCR reagents, Illustra GFX PCR DNA and Gel Band Purification Kit (Amersham) was used to purify the PCR products according to the manufacturer's instructions (sample capture and binding, washes and dry and a final elution).

The fragments were transformed into vectors pGEM T-easy (Promega) and cloned into a vector using the TOPO TA cloning Kit DH5α cells (Invitrogen), in a Petri dish with LB medium with ampicillin (50μg/ml). X-Gal at 20mg/mL is added to detect the clones that positively incorporated the plasmid with the insert.

The PureYield Plasmid Miniprep System kit (Promega) was used according to the manufacturer's instructions (lyses of the bacterial culture, washes and a final elution) to extract the plasmids from the positive clones.

The sequencing of each clone is done using the sequencing kit BigDye 3.1 (Applied Biosystems) and the M13 F (5' GTAAAACGACGGCCAGT 3') and M13 R (5' AACAGCTATGACCATG 3') universal primers and the sequencing services of the Universitat de Barcelona. To determine the sequences at 5' and 3' of the gene the specific Forward and Reverse Primers combined with the M13F and R (present in the vector in which the fragment is cloned) are used. The sequences are then visualized, assembled, compared and characterized with the BioEdit v7.0.1 program (Hall 1999).

The primers used to amplify the Ephrin/efn of the European species *Branchiostoma lanceolatum* were designed using the gene predictions of the American species *B. floridae* in highly conserved regions, determined by the protein domain structure and Blasts against other genomes. (Mellot 2008)

Primers from 5' to 3':

	BfEphA	BfEphB	BfefnA	BfefnB
Forward	CGTCAGAACGTACCAAGTGTG	TGTCAGAACGTACCAAGTCTGC	TTTATACGGGACACTGTTGCT	GACAAGCTCGACATCATCTG
Reverse	TGGGCACTGTTCATGATCTTC	ATGGAGTTCAGGATCTTCTTCTG	ACCAATCAGCATCTGGGAT	GAACACTGTGGCCAGCAAG

3.3. RIBOSYNTHESIS

The mold to ribosynthesize *in vitro* a digoxigenine marked probe is obtained by amplification of the purified clone (see 3.2) using the universal M13F or M13R primers, depending on the orientation of the insert in the vector, and a specific primer localized in the inserted fragment. The *in vitro* transcription is undertaken by adding 100ng/mL of template, DIGrNTPmix (Roche), TSC buffer (Roche), RNAse inhibitor (Promega), and T3/T7/SP6 polymerase (Roche), depending on the orientation of the fragment in the template. Then the mold

is degraded by adding DNAse I (Roche) and finally the probe is eluted and precipitated with mini Quick Spin Columns (Roche). The probe is stored in 50% formamide at -20ºC.

3.4. Phylogenetic Analysis

The sequences of the genes of interest were annotated as described in point 3.1. For several genes splicing isoforms were found but only products with orthologous regions in other species were used for the analysis. This is similar to the case of *Drosophila melanogaster* which possesses four isoforms for the Ephrin receptor; therefore only the isoform A was analyzed. The protein sequences obtained were aligned with ClustalX 2.0 (Larkin 2007) and manually curated. Since some genes, mainly in Nematostella, only contained the ligand binding domain (lbd), two approaches were undertaken for the Ephrin receptor, one using only the ligand binding domain (lbd); and another tree was made using the whole protein, excluding the ones that only presented the lbd (NevEph10, NevEph2, TaEph, XtEphb5, XtEpha5) (see Fig. 6). In the case of the ephrin ligands two trees were also obtained, one using all the annotated ephrin genes, and another one in which the *Nematostella vectensis* genes were excluded to obtain a less altered topology, because they are very divergent.

The phylogenetic trees were constructed using Mega4 (Tamura 2007) by analysis of the several topologies given by different methods (Maximum parsimony, Minimum Evolution, and Neighbor Joining). The number of bootstrap replicates and additional parameters used are indicated in the figure legend of each of the resulting trees.

3.5. Laboratory Maintenance of Amphioxus, Spawning and Embryos Collection

Adults of *Branchiostoma lanceolatum* were collected at Argelès-sur-mer (France), thanks to the tools offered by the oceanologic observatory of the University of Pierre et Marie Curie of Paris, during May and June. They were counted, classified and separated according to their gonad status (animal with visible gonads and in post-spawning stages). Individuals from both sexes were kept in common aquariums without sand and filtered sea water at 17-19ºC, with controlled day-night cycle and are fed algae. The water was regularly changed. The spawning season is in June-July, when the gonads fill with mature gametes. To induce the release of the gametes, the temperature was raised to 23ºC for 36h. Both sexes are maintained together during 24h and then each individual is separated for the last 12h. When the night period starts, the gametes are released. As males and females are separated, the gametes can be collected and a controlled *in vitro* fertilization can be performed and, by monitoring the time post-fertilization of the eggs, embryos of each developmental stage can be fixed with 4% paraformaldehyde (PFA) and stored in 70% ethanol at 4ºC. (Fuentes 2007)

3.6. WHOLE-MOUNT *IN SITU* HYBRIDIZATION

The embryos were rehydrated in 70% ethanol then washed several times in PBT (PBS 1x, 0.1% Tween 20). After they were treated with proteinase K (7.5ug/mL in PBT) – 3minutes for gastrulae, 5 minutes for early and mid neurula, 8minutes for late neurula, 10 minutes for larvae younger than 36h, and 15minutes for larvae older than 36h – the digestion was stopped by washing with 2mg/mL of Glycine in PBT. Then they are post fixed with 4% paraformaldehyde (PFA) in PBT for 30 minutes at room temperature (RT). Then they were washed with triethanolamine 0.1M (pH8), followed by a treatment with 2.5uL/mL and then 5uL/mL of acetic anhydride in 0.1M triethanolamine pH8 (without shaking), washed in PBT and pre-incubated in hybridization buffer (50% formamide, 25% 20x SSC, 2% yeast RNA 50mg/mL, 1% heparin, 2% 50x Denhart, 1% EDTA 0.5M, 0.1% Tween) for 1 hour at 65ºC. At last they suffer hybridization with the pre-denaturalized probe (100-200ng/mL) at 70ºC and kept in the hybridization solution over night at 65ºC. After they were washed first with preheated hybridization buffer and then with wash solution (50% formamide, 25% 20xSSC, 1%EDTA 0.5M, 0.1% Tween) at 65ºC. Then washes with MABT (50mg/mL maleic acid, 40mg/mL NaCl, 0.1% Tween) are performed at RT before incubation with 2% blocking reagent (Boehringer-Mannheim) in MABT for 1hour. The incubation with the antibody is done at a concentration of 1:3000 of Fab-antiDIG (Roche) in the previous blocking solution for 3hours at RT or over night at 4ºC. Then several washes with MABT are performed before treating the embryos with Alkaline Phosphatase (AP buffer) (10% Tris 1M pH9.5, 5% MgCl2 1M, 2% NaCl 5M) first at 50% in MABT, then without MgCl2 and at last with MgCl2. Finally the embryos are placed in the revealer BMP purple (AP substrate precipitating, NBT/BCIP ready-to-use solution, Roche) and kept in total absence of light, at RT, with no shaking until the signal appears. Once the signal is satisfyingly visible the embryos are washed in MABT and post-fixed in 4% PFA, washed in PBT and kept at 4ºC in 80%Glycerol in PBT.

4. RESULTS

4.1. PHYLOGENETIC ANALYSIS OF THE EPHRIN GENE FAMILY (RECEPTORS AND LIGANDS)

4.1.1. NOMENCLATURE AND STRUCTURE OF THE PROTEIN OF EPHRIN RECEPTORS AND LIGANDS

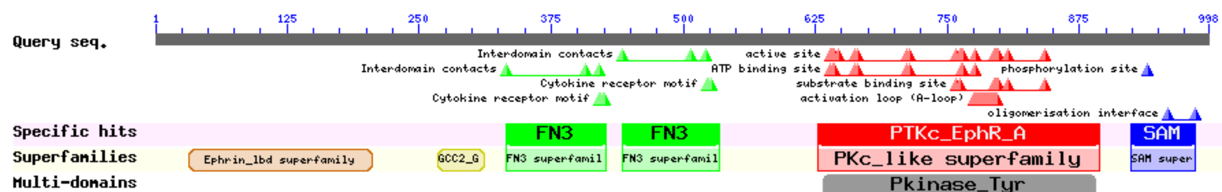

Fig. 7– Structure of the conserved domains of an Ephrin receptor. These topologies can be obtained by analyzing a protein sequence with the *Conserved Domain Search* tool of the NCBI website (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). These genes typically present an ephrin ligand binding domain (lbd) at 5', a GCC domain, two fibronectin type III domains (FN3), protein tyrosine kinase domain (PTK), and a SAM domain at 3'.

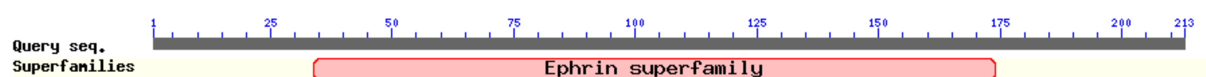

Fig. 8 – Structure of the conserved domains of an ephrin ligand. Topologies can be obtained by analyzing a protein sequence with the *Conserved Domain Search* tool of the NCBI website (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). These genes present a predominant conserved domain that is included in the ephrin superfamily.

Deuterostome species	Eph Receptors	Efn ligands
Homo sapiens	HsEphA1 HsEphB1 HsEphA2 HsEphB2 HsEphA3 HsEphB3 HsEphA4 HsEphB4 HsEphA5 HsEphB6 HsEphA6 HsEphA7 HsEphA8 HsEphA10	HsefnA1 HsefnB1 HsefnA2 HsefnB2 HsefnA3 HsefnB3 HsefnA4 HsefnA5
Xenopus tropicalis	XtEphA2 XtEphB1 XtEphA3 XtEphB3 XtEphA4a XtEphB4 XtEphA4b XtEphB5 XtEphA5 XtEphA6 XtEphA7a XtEphA7b	XtefnA1 XtefnB1 XtefnA3 XtefnB3 XtefnA5 XtefnA6
Ciona intestinalis	Ci-Eph epsilon (CiEphE) Ci-Eph gamma (CiEphG) Ci-Eph delta (CiEphD) Ci-Eph like (CiEphL) Ci-EphB3	Ci-ephrinA a Ci-ephrinA b Ci-ephrinA c Ci-ephrinA d Ci-ephrinB
Branchiostoma floridae	BfEph1 BfEph2	Bfefn1 Bfefn2 Bfefn3
Branchiosotma lanceolatum	BlEph1 BlEph2	Blefn1 Blefn2
Strongylocentrotus purpuratus	Sp-Eph	Sp-efn
Saccoglossus kowalevskii	SkEph	Skefn1 Skefn2

Protostome species	Eph Receptors	Efn ligands
Drosophila melanogaster	DmEph	Dmefn
Apis mellifera	AmEph	Amefn1 Amefn2
Nasonia vitripennis	NavEph	Navefn
Tribolium castaneum	TcEph	Tcefn1 Tcefn2
Daphnia pulex	DpEph	Dpefn1 Dpefn2
Ixodes scapularis	IsEph	Isefn
Lottia gigantea	LgEph	Lgefn
Capitella teleta	CtEph	Ctefn
Nematostella vectensis	NevEph1 NevEph6 NevEph2 NevEph7 NevEph3 NevEph8 NevEph4 NevEph9 NevEph5 NevEph10	Nevefn1 Nevefn2 Nevefn3 Nevefn4
Trichoplax adhaerens	TaEph	0
Ephydatia fluviatilis	EfEph	0
Monosiga brevicollis	0	0

Tab.1 – Nomenclature of the receptors and ligands of the gene family of the Ephrins of the organisms used in the phylogenetic analysis. The abbreviations of the genes are the one used in the phylogenetic tree. The genes were annotated as described in the 3.1 of the *Methodology* section.

In *Branchiostoma floridae* we found three ephrin ligands *in silico* (Bfefn1, Bfefn2 and Bfefn3), although it was only possible to amplify and clone two ephrins in *Branchiostoma lanceolatum* (Blefn1 and Blefn2).

The number of Eph receptors seems to increase in *Branchiostoma sp.* (2 Eph)*, Ciona intestinalis* (6 Eph) and vertebrates (*Xenopus tropicalis* with 12 and *Homo sapiens* with 14). The non-chordates analyzed possess a single copy of the Ephrin receptor except for *Nematostella vectensis*, which has 10 Eph receptors (Nev1-10) and *Monosiga brevicollis,* which doesn't appear to possess either Ephrin receptors or ligands (Tab. 1). By analysis of the genomes by Blast, for *Trichoplax adhaerens,* and by search in the NCBI database, for *Ephydatia fluviatilis*, it seemed that these species do not possess ephrin ligands. It was also possible to observe that the number of ephrin ligands varied independently in each species from *Monosiga brevicollis* to *Strongylocentrotus purpuratus* without a clear pattern for the increase in number of ligands present in each genome, with *Nematostella vectensis* presenting four ephrin ligands. In the Chordata phylum the number of ephrin genes seemed to have a tendency to increase (*Branchiostoma floridae* 3efn, *Ciona intestinalis* 5efn, *Xenopus tropicalis* 8efn, *Homo sapiens* 8efn) (Tab. 1).

After annotation of all the genes encountered in the analyzed genomes, curiously not all the Ephrin receptors presented the typical structure observed in Fig.7. In some species these proteins lacked the parte corresponding to the tyrosine kinase domain, FN3 and SAM. This was the case of the genes of *Nematostella vectensis* NevEph2 and NevEph10, the genes of *Xenopus tropicalis* XtEphA5 and XtEphB5, and the gene of *Trichoplax adherens* TaEph.

4.1.2. Phylogenetic analysis of the Eph Receptors

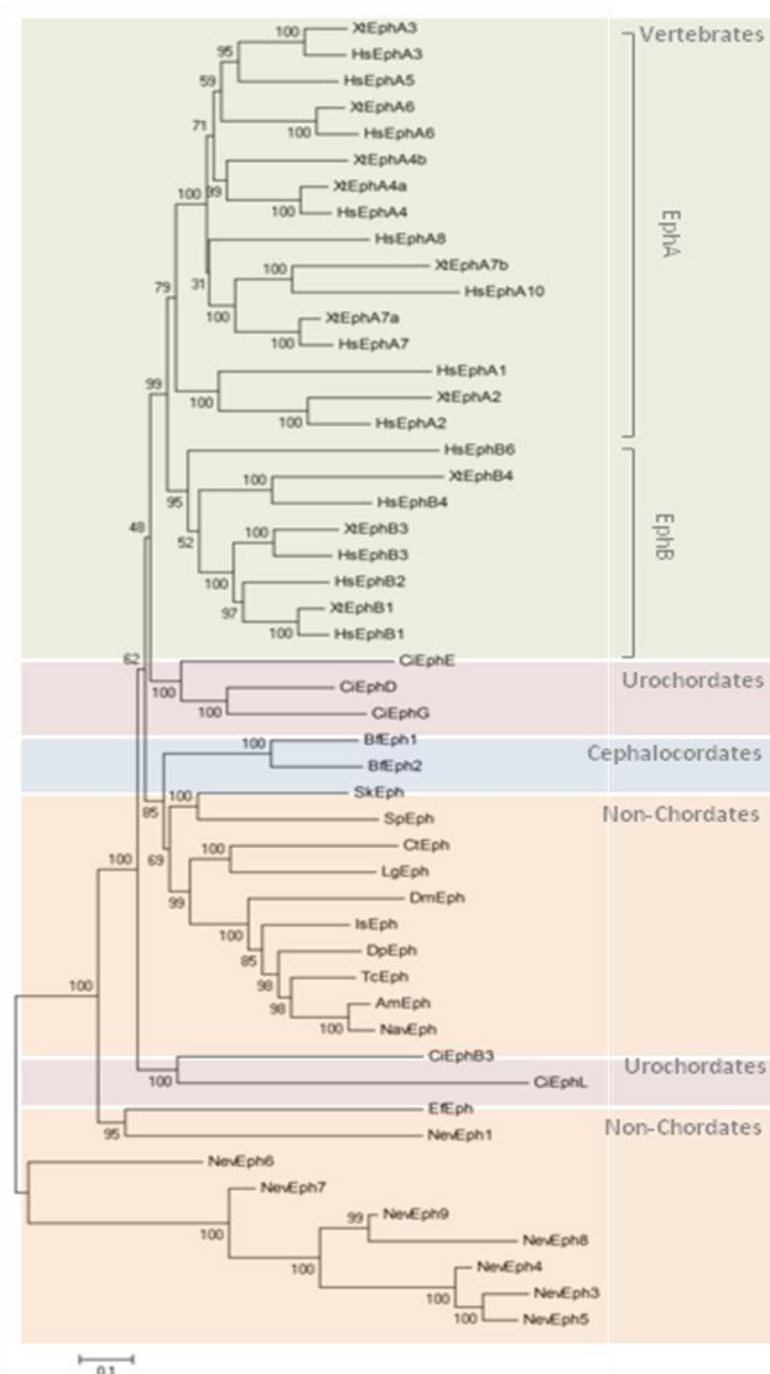

Fig.9- Unrooted phylogenetic tree for the Eph receptors (whole protein) constructed by Neighbor-Joining method with Gamma distributed rates among sites (1.0), pairwise deletion of the gaps/missing data and Poisson correction. Numbers represent bootstrap values for 1000 replicates. In this analysis only the genes that included all the Eph domains (see Fig. 7) were included, the other genes were excluded from the analysis. Four major groups were created to facilitate the description (vertebrates, Urochordates, Cephalocordates, and Non-Chordates). Am, *Apis mellifera;* Bf, *Branchiostoma floridae;* Ci, *Ciona intestinalis;* Ct, *Capitella teleta;* Dm, *Drosophila melanogaster ;* Dp, *Daphnia pulex ;* Ef, *Ephydatia fluviatilis ;* Hs, *Homo sapiens;* Is, *Ixodes scapularis;* Lg, *Lottia gigantea;* Nav, *Nasonia vitripennis;* Nev, *Nematostella vectensis;* Sk, *Saccoglossus kowalevskii;* Sp, *Strongylocentrotus purpuratus;* Tc, *Tribolium castaneum;* Xt, *Xenopus tropicalis.* The scale bar corresponds to the number of substitutions per site. Branch length indicates the amount of divergence inferred.

The phylogenetic relationship of the Eph receptors between clades is shown in Fig. 9 and has been separated in several groups for clarity.

Upon analysis we see that Ephrin receptors A and B of *Xenopus tropicalis* and *Homo sapiens* group together in the same clade (vertebrates) but also that Ephrin receptors A (EphA) of *Xenopus tropicalis* and *Homo sapiens* form a monophyletic group separated from the EphB, which form another distinct monophyletic group.

Urochordate Ephs seem to form two a polyphyletic group with two distinct clades, one closely related to vertebrates and which includes CiEphE, CiEphD and CiEphG. The other group, which includes the CiEphB3 and CiEphL, is basal to the split that includes vertebrates, cephalochordates, the remaining urochordates and non-chordates, probably due to Long-branch-attraction. The fact that the cephalochordate genes group with those of other metazoans and not with chordates probably indicates that the urochordates and vertebrate Eph are more divergent, whereas the amphioxus genes resemble the ancestral situation.

Apart from the problems with the paraphyletic grouping of chordates and deuterostomes, the rest of the topology follows the phylogenetic relationships known for these clades. Echinoderms and hemichordates group together as ambulacrarians and protostomes appear monophyletic, with well supported lophotrochozoans (mollusks and annelids) and arthropod (crustacean, chelicerate and insect) clades. Non bilateral animals clearly separate from the rest (bootstrap value of 100), with most of the *Nematostella* genes clustering in a monophyletic group.

4.1.3. Phylogenetic analysis of the efn ligands

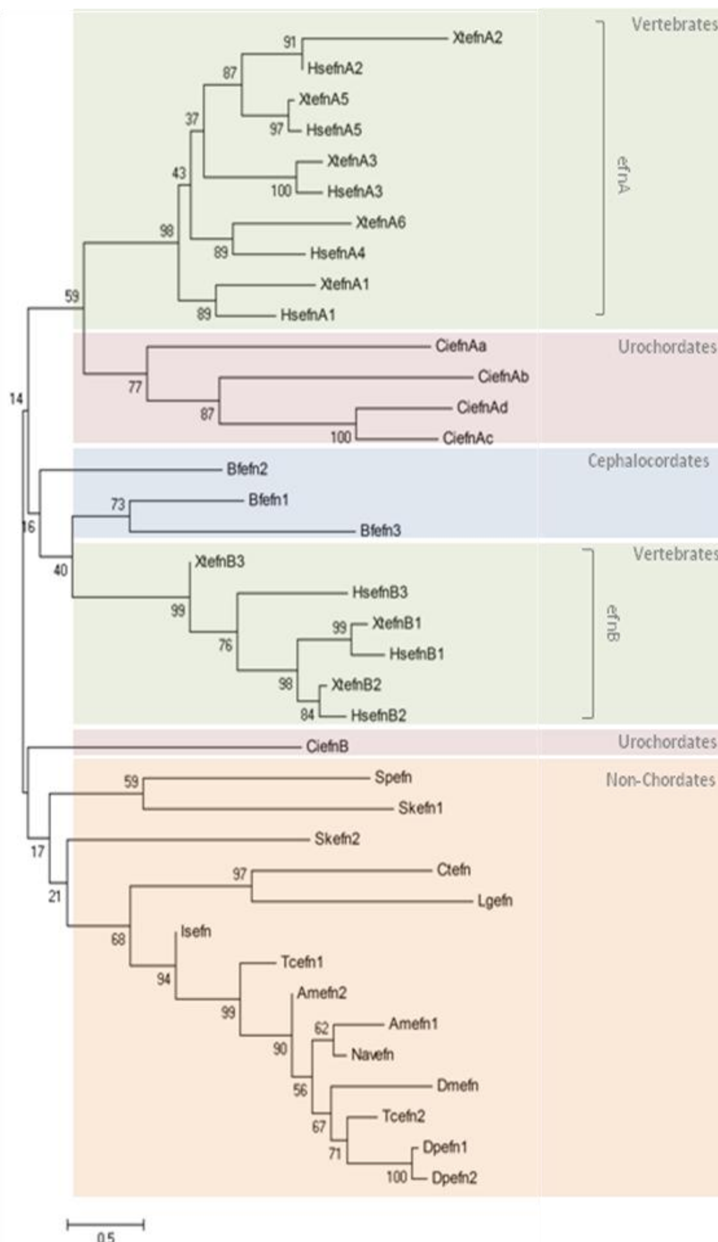

Fig. 10- Unrooted phylogenetic tree for the ephrin ligands (efn) constructed by the Neighbor-Joining method with Gamma distributed rates among sites (1.0), pairwise deletion of the gaps/missing data and Poisson correction. Numbers represent bootstrap values for 1000 replicates. Four major groups were created to facilitate the description (vertebrates, Urochordates, Cephalocordates, and Non-Chordates). Am, *Apis mellifera;* Bf, *Branchiostoma floridae;* Ci, *Ciona intestinalis;* Ct, *Capitella teleta;* Dm, *Drosophila melanogaster ;* Dp, *Daphnia pulex ;* Hs, *Homo sapiens;* Is, *Ixodes scapularis;* Lg, *Lottia gigantea;* Nav, *Nasonia vitripennis;* Sk, *Saccoglossus kowalevskii;* Sp, *Strongylocentrotus purpuratus;* Tc, *Tribolium castaneum;* Xt, *Xenopus tropicalis.* The scale bar corresponds to the number of substitutions per site. Branch length indicates the amount of divergence inferred.

The topology of the phylogenetic tree represented in Fig.10 shows the relationship between the ephrin ligands (efn) of several species. We can observe that the vertebrate ephrins form two monophyletic groups, the efnA and the efnB group. The vertebrate efnAs seem to be closely related to the efnA genes of urochordates (CiefnAa, CiefnAb, CiefnAc and

CiefnAd) and the efnBs appear to be closer to the cephalochordate ephrins (Bfefn1, Bfefn2 and Bfefn3). Curiously a urochordate ephrin (CiefnB) seems to group with the Non-chordates ephrins and not with the other ephrin ligands of the species. As observed with the receptors, the ephrin ligands of the non-chordates form a monophyletic group. Importantly, the bootstrap values are generally low and only those in protostome branches are more robust. The clustering of some lineage-specific duplications (like the expansion of vertebrate EphA and EphB, *Ciona* EphA and *Daphnia* EphA and B) are also well supported.

4.2. Expression of Ephrin receptors and ligands in the European Amphioxus, *Branchiostoma lanceolatum*

After annotating and designing the primers for the ephrin receptors and ligands (BlEph1, BlEph2, Blefn1, Blefn2 and Blefn3) in the American species *Branchiostoma floridae,* the genes were amplified and cloned in the European species *B. lanceolatum.* Once this was achieved, the expression of the genes was detected by the technique of whole-mount *in situ* hybridization in embryos of the *B. lanceolatum*, in several early developmental stages (32 cells morula to 60hrs larvae), as described in section 3.6 of the *Materials and Methods*.

4.2.1. Expression pattern of the Ephrin receptors

At the morula stage, all the Ephrin receptors and ligands expressed in the whole embryo (data not shown). BlEph1was expressed at the blastopore region (Fig.11A) at gastrula stage. During neurulation BlEph1 was expressed in the lateral mesoderm, surrounding the closing neural tube, in the forming somites, and possibly in the precursor of the cerebral vesicle (Fig.11B-E). In late neurula BlEph1 was expressed also in the notochord and neural tube and starts expressing in the future oral region (Fig.11 F and G, arrowhead). In pre-mouth stage BlEph1 was expressed in the anterior notochord, the forming mouth apparatus and pharynx (arrowheads in Fig.11 H and I) and the posterior notochord (arrowheads in I) (Fig.11H and I). In larvae this gene was expressed in the anterior notochord, in the cerebral vesicle, the mouth , the endostyle, the *club-shaped* gland, the preoralpit, the diverticulum, and the posterior notochord (arrowheads) (Fig.11 J) The sense probe didn’t present any specific pattern (Data not shown)

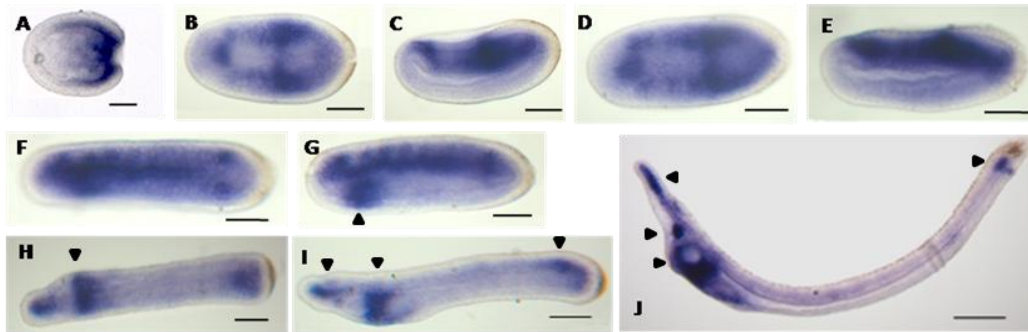

Fig.11 – Expression pattern of BlEph1 in embryos of *Branchiostoma lanceolatum* by whole-mount *in situ* hybridization (antisense probe BlEph1). (A) 12hrs gatrula; (B, C) 18hrs mid neurula, dorsal and lateral view, respectively; (D, E) 21hrs mid neurula, dorsal and lateral view, respectively; (F, G) 24hrs late neurula, dorsal and lateral view, respectively; (H, I) 36hrs pre-mouth embryo, dorsal and lateral view, respectively; and (J) 60hrs larva, lateral view. The arrowheads indicate the future mouth apparatus region (G-I), the anterior and posterior notochord (I, J) and the diverticulum and mouth (J). Anterior is to the right, posterior to the left. Scale bars: 50µm

The BlEph2 receptor seemed to express in all tissues through gastrulation to the stage of mid neurula (Fig.12 A-C). A segmented pattern seemed to appear during neurulation that follows the pattern established by the forming somites (Fig.12 B, C). In pre-mouth embryos the signal was localized in the forming mouth apparatus and the anterior notochord (Fig.12 D, E). In larvae the expression suggested an expression in the mouth apparatus and pharynx region (signalized by an arrowheads in Fig.12 F). The sense probe didn't present any specific pattern (Data not shown).

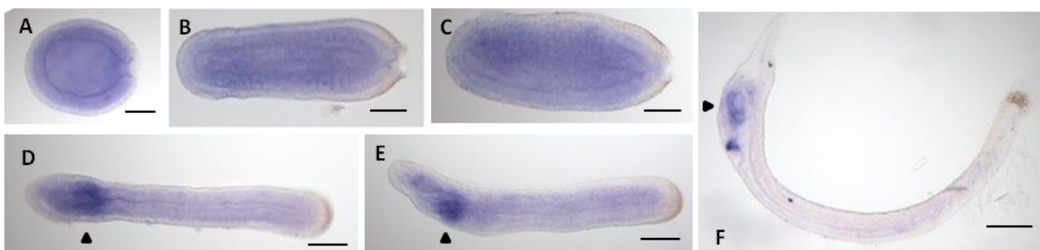

Fig.12 – Expression pattern of BlEph2 in embryos of *Branchiostoma lanceolatum* by whole-mount *in situ* hybridization (antisense probe BlEph2). (A) 12hrs gatrula; (B, C) 24hrs mid neurula, dorsal and lateral view, respectively; (D, E) 36hrs pre-mouth embryo, dorsal and lateral view, respectively; and (F) 60hrs larva, lateral view. The arrowheads indicate the region of the future mouth (D, E) and the formed mouth (F). Anterior is to the right, posterior to the left. Scale bars: 50µm

4.2.2. Expression pattern of the ephrin ligands

The expression pattern of the ephrin ligands (efn 1 and 2) seemed to be in agreement with the receptors expression. Blefn1 was expressed mostly in the mesendoderm around the blastopore region in gastrula embryos (Fig.13A). During neurulation the expression was mesodermic, surrounding the closing neural tube (Fig.13B-E) and it also seemed to express in populations of neurons (arrowheads) (Fig.13D, E). In pre-mouth

stage it was expressed in the region of the future mouth apparatus and the posterior notochord, as well as posterior endoderm (signalized with arrowheads) (Fig.13F, G). Already in the larva stage the expression was mostly confined to the region of the mouth apparatus and pharynx (arrowheads), and weakly elsewhere, with the exception of the cerebral vesicle (Fig.13H). The sense probe didn't present any specific pattern (Data not shown)

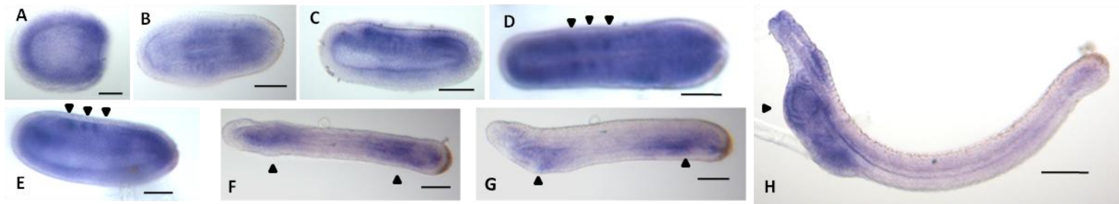

Fig.13 – Expression pattern of Blefn1 in embryos of *Branchiostoma lanceolatum* by whole-mount *in situ* hybridization (antisense probe Blefn1). (A) 12hrs gatrula; (B, C) 21hrs mid neurula, dorsal and lateral view, respectively; (D, E) 24hrs late neurula, dorsal and lateral view, respectively; (F, G) 36hrs pre-mouth embryo, dorsal and lateral view, respectively; and (H) 48hrs larva, lateral view. The arrowheads indicate the future mouth apparatus region and the posterior notochord (F, G), and the mouth apparatus (H). Anterior is to the right, posterior to the left. Scale bars: 50µm

During gastrulation, Blefn2 seemed to express in the entire gastrula (Fig.14A). During neurulation it was confined to the mesoendoderm, with the exception of the most posterior domain (Fig.14 B and C). In pre-mouth embryos a clear pattern was localized in the region of the forming pharynx, and the mucus gland and posterior notochord (arrowheads) (Fig.14 D and E). In larvae the expression seemed to be localized in the region of the mouth apparatus and pharynx (signalized by arrowheads) and also the posterior notochord (Fig.14F). The sense probe didn't present any specific pattern (Data not shown)

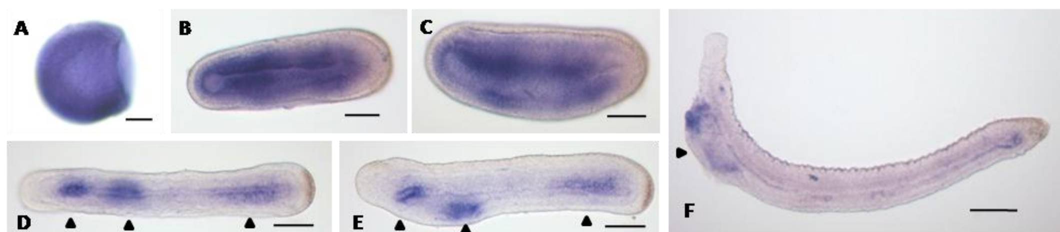

Fig.14 – Expression pattern of Blefn2 in embryos of *Branchiostoma lanceolatum* by whole-mount *in situ* hybridization (antisense probe Blefn2). (A) 12hrs gatrula; (B, C) 24hrs late neurula, dorsal and lateral view, respectively; (D, E) 36hrs pre-mouth embryo, dorsal and lateral view, respectively; and (F) 48hrs larva, lateral view. The arrowheads indicate the anterior notochord, the future mouth apparatus region and the posterior notochord (D, E), and the mouth apparatus (F). Anterior is to the right, posterior to the left. Scale bars: 50µm

5. DISCUSSION

5.1. PHYLOGENETIC ANALYSIS OF THE EPHRIN GENE FAMILY (RECEPTORS AND LIGANDS)

5.1.1. NOMENCLATURE AND STRUCTURE OF THE PROTEIN OF EPHRIN RECEPTORS AND LIGANDS

By annotation of the Eph genes we were able to find two *Branchiostoma sp*. Eph receptor genes and that this number increases in urochordates and vertebrates, whilst in the non-chordates only one Eph receptor gene is present, suggesting intensive duplication events at the origin of Cephalochordata, Urochordata and also at the Vertebrata lineages. If we analyze closely Table 1, we see that *Xenopus tropicalis* possesses 12 Eph whilst *Homo sapiens* 14Eph. This could be due to an independent duplication event specifically in the human lineage or at the base of this lineage, but after the split that lead to *Xenopus tropicalis*. Alternatively, this may result from an independent loss in *Xenopus tropicalis* or to poor annotation of the genes in this species. In this case I have chosen to keep the nomenclature proposed by the data found, but it is known that XtEphA4b does not exist in Humans which probably is due to a secondary loss in these, and that there are genes that were badly annotated in Xenopus such as XtEphA7b, that should be called XtEphA10 and XtEphB5, that should be XtEphB6. However, to be sure of this it would be important to realize more analysis of the genes and compare with other vertebrates. Teleosts were not included due to the difficulties in annotating the genes because of the specific whole genome duplications that occurred in this lineage, possessing the more diverse group of Eph found (Coulthard 2002).

Interestingly in *Trichoplax adhaerens* and *Ephydatia fluviatilis* ephrin ligands were not found. There is no genome available for *Ephydatia* and therefore it is not possible to know if this absence is real. In the case of *Trichoplax*, this could be because the receptor would have appeared before the ligand during the evolution or perhaps these species would have suffered a secondary loss of the ligand or it was simply bad annotated in both species. In any case more studies would have to be done to show if this receptor has a function in these species and try to explain the cause of the absence of the ephrin gene. If we observe the other species of non-chordates it is possible to see that specific expansions of the ligands occurred in some species that present two genes (*Apis mellifera, Tribolium castaneum*, *Daphnia pulex, Saccoglossus kowalevskii* and *Nematostella vectensis*) in comparison with the others that only possess one. The curious fact is that this expansion event seems to have occurred several times and in a convergent fashion and in the cases of the sea anemone, amphioxus, sea urchin and vertebrates it would have happened several times and in parallel with the

receptors. This suggests that on one hand there could have been a co-evolution of the ligands and the receptors; and on the other hand that possessing several Eph and efn seems to be an evolutionary advantageous characteristic, perhaps by conferring a more performing developmental process in space and time, resulting in more complex and plastic structures by enhancing the ability for cell migration, compartmentalization, formation of a central nervous system, muscular organization, and other features.

5.1.2. Phylogenetic analysis of the Eph Receptors

The phylogenetic analysis was performed with protein sequences due to the large phylogenetic distance between the organisms used and to avoid the phenomenon produced by saturation of the substitution at the DNA level. The phylogenetic analysis for the Eph receptors was performed using two different approaches. At first only the Eph sequences corresponding to the ligand-binding-domain (lbd) were considered to be able to include and align all the proteins available for analysis (Supplementary data Fig. 1). The second approach was to use the whole protein and thus exclude the genes (NevEph10, NevEph2, TaEph, XtEphb5, XtEpha5)that lack the other domains (tyrosine kinase, FN3 and SAM – Fig. 9) in order to be able to increase the phylogenetic signal (see point 3.5 of the *Materials and Methods* section). This approach brought more information because included more sites that were also susceptible to evolve, besides the conserved domains such as the lbd, and provided what was considered to be a better topology of the tree, as well as stronger bootstrap values (Fig. 9 and Fig.1 of the Supplementary data). The truncation of some of the Eph receptor genes observed may be due to problems with the genome assembly or, in the case of duplicates, subfunctionalization of the duplicated copy by loss of the Tyrosine kinase property or simply bad annotation of the genes. To be able to assess which would be the right option, more analysis of the genomes would have to be performed. *Ephydatia fluviatilis* was considered as the outgroup for the Eph analysis because it seems to be the living animal that possesses the ancestral form of a Ephrin receptor (Suga 1999) and because no Ephrin receptor was found in the metazoan *Monosiga brevicollis* (Suga 2008), a species that has comparatively a more basal position in the tree of life.

Urochordate Eph receptors CiEphE, CiEphD and CiEphG seem to have diverged earlier than the vertebrates Eph lineage, and cephalochordates (BfEph1 and BfEph2) diverged even earlier than the urochordates. These results are in accordance with the current view on the phylogenetic relationship of chordates (Delsuc et al 2006) in which urochordates are closer to vertebrates than cephalochordates, suggesting that these would have been a transitional form at the invertebrate-vertebrate split. (Fig. 9)

Urochordates Eph do not group in a monophyletic group suggesting that independent duplications would have occurred in this species, such as also suggested by Satou and collaborators in 2003. It is observable in the phylogenetic analysis that the CiEphE, CiEphD and CiEphG are closer to each other and to the Vertebrate Eph receptors then to the other members of Ephrin receptors of *Ciona intestinalis* (CiEphB3 and CiEphL) that group closer to the Non-Chordates and basal to the lineage of vertebrates, cephalochordates, the other urochordates genes and even some Non-Chordates genes. When looking at the alignment the CiEphB3 and CiEphL have a

relatively divergent sequence in comparison to the CiEphE, CiEphD and CiEphG but they align with higher homology with *Nematostella vectensis* (data not shown).This is perhaps due to the divergence between the copies generated by specific duplications in the species lineage that evolved independently after, such as what also seems to have occurred in *Nematostella vectensis* (Fig. 9). When looking at the Fig.1 of the Supplementary data, the CiEphL and CiEphB3 group with CiEphE, CiEphD and CiEphG, cephalochordates and vertebrates. Therefore these strange groupings are due to Long-branch-attraction artifacts.

It is also worth noting that the arthropods group monophyletically together and with lophotrochozoan protostomes, as suggested by the topology found in the tree of life. However, *Nematostella vectensis* is a particular case because it doesn't seem to group with the other Metazoa and is located basally to the phylogenetic tree. This suggests that this species have been exposed to independent duplications specific to the lineage which is also concordant on one hand with the possession of many more Eph receptors genes than the other non-chordate metazoans (Tab.1), with most of them truncated and possessing only the ligand-binding domain (data not shown). Moreover, the sequences are highly divergent and are more similar to each other than to those of other metazoans (data not shown). Finally, *Nematostella vectensis* (sea anemone, Cnidaria) genes group with *Ephydatia fluviatilis* genes (sponge, Porifera), which are reported to possess the ancestral form of the Ephrin receptor and which are more basal in the tree of life than Cnidarians (http://tolweb.org/tree/). This may be due to the highly divergent sequence observed in sponges and in the duplicates of *Nematostella vectensis.* But as *Nematostella* genes are so divergent they can be influencing the topology of the tree by a phenomenon called Long-branch-attraction (Felsenstein 2004). If one excludes the *Nematostella vectensis* Eph genes, the topology of the tree remains except for the *Ciona intestinalis* genes that group all together, basal to the tree and closer to the *Ephydatia fluviatilis* Eph, suggesting once more that independent duplications have occurred in the urochordates lineage. (Supplementary data Fig. 2) When constructing a phylogenetic tree by only excluding the *Ciona intestinalis* Eph genes, because these have suffered specific independent duplication that could influence the topology by Long-branch-attraction: the vertebrates Eph group together (as observed in Fig.9); cephalochordates genes are basal to this branching, suggesting to have diverged earlier than vertebrates as it is known by the literature (Delsuc 2006); the non-chordates group monophyletically together (as observed in Fig.9); and the *Nematostella vectensis* genes together with the *Ephydatia fluviatilis* gene group at the base of the tree(as observed in Fig.9)(Supplementary data, Fig.3).

By observing the topology of the tree it seems that cephalochordates are more closely related to non-chordates than they are to urochordates or vertebrates. (Fig.8 and Fig.2 of Supplementary data) But as they are included in the Chordata phylum, this may suggest that this group possesses a more ancestral slowly evolving form, more similar to the non-chordate genes than to the more divergent vertebrate orthologs (Delsuc 2006).

The bootstrap values are generally low, what may compromise the reliability of the phylogenetic results obtained. However by analysis with other methods, such as Minimum-evolution and maximum parsimony, the topologies were not more satisfying nor were the bootstrap value more significant (data not shown). To be

able to assess a better topology other methods (Bayesian approach) should be used and perhaps more information should be included (such as genes from other species). Another interesting perspective would be to determine the functional nature of the receptors in *Branchiostoma sp*, to see if they behave as an A or B type receptors.

5.1.3. Phylogenetic analysis of the EFN ligands

It is important to refer that in this analysis the ephrins of *Nematostella vectensis* (Nevefn1-4) were excluded due to important independent duplications occurring in this species that was causing a phenomenon of Long-branch-attraction (Felsenstein 2004) and affecting the topology of the tree that was expected by the analysis of the bibliography. However, the topology of the tree with these ephrin genes included can be found in the Supplementary data Fig. 4, but it is apparently badly supported by low bootstrap values.

As expected, the chordates (cephalochordates, urochordates and vertebrates) are phylogenetically closer between them than they are to the non-chordates, except for the gene CiefnB. Urochordates CiefnAs group with vertebrates efnAs suggesting a higher similitude between these, such as Mellot and Burke have suggested in 2008. It is also in concordance with the structure of the protein because the A type (GPI anchored) of both urochordates and vertebrates group together. However, in this analysis the transmembranar B type of urochordates (CiefnB) is not grouping with the vertebrates ephrin B but is closer to non-chordates ephrins. This is perhaps due to the high divergence encountered in this group, known to be fast-evolving (Fig.10) (Satou 2003).

Mellot et al suggested that the EfnB would be closer to the ancestral form. In this study vertebrate efnB do not group with non-chordate efnB, but they do group with amphioxus sequences. Therefore it could be argued that they in fact represent more ancestral forms or that they are less divergent than efnA.

The limits of this study are the bootstrap values of the internal branches of the trees, which are very low, thereby diminishing the reliability of our interpretations, and making it more difficult to infer strongly supported conclusions. To be able to have a better idea of the phylogenetic relationship of this gene and the amphioxus' Eph/ephrin phylogenetic position, more methods would have to be compared (such as a Bayesian approach), calculation of a model for the evolution of these sequences would have to be undertaken, or perhaps the analysis would benefit from including other species in the analysis.

5.2. Expression of Ephrin receptors and ligands in the European species *Branchiostoma lanceolatum*

The development of the amphioxus mouth is particular. It undergoes an asymmetrical opening of the mouth on the left side, starting at the first mesodermal block (or somite) of the developing embryo. An extension and a posterioriarization occur and, after metamorphosis, the mouth adopts its final anterio-ventral position. Amphioxus share, however, a common developmental pattern in this process with vertebrates. It consists in the innervations of the mouth by axons that come from several regions of the organism, including the CNS (central nervous system). In amphioxus this only occurs later on during the development due to the later development of the nerve networks. (Yasui 2008)

The formation of teeth in mammals has its origins in the first branchial arch and the neural crest-derived ectomesenchymal cells (Corbourne 2005). Also derived from migrating neural crest cells (NCC) are the upper lip, maxillary, medial nasal, and lateral nasal prominences and the primitive oral cavity (Jiang 2005). In the development of the pharynx and branchial arches also participate the NCC (Graham 2001) and also in the neural tube formation. The Eph/ephrin signaling is important in the formation of several mouth structures in vertebrates such as teeth and palate (Luukko 2005, Arthur 2009, Risley 2009).

As it is known from the literature, the Eph/ephrin interaction is implicated in the patterning of the nervous system and in the process of innervations, in the formation of the notochord, the tail notochord, the neural tube, the somites and paraxial mesoderm in vertebrates (Chan 2001; Nakamoto 1999; Naruse-Nakajima 2001).

As it was observable by whole-mount *in situ* hybridization, the expression pattern of neurulas from Ephrin receptor 1 (BlEph1 Fig.11 B-G) and ephrin ligand 1 and 2 of *B. lanceolatum* (Blefn1 Fig.13 B-E, Fig.14 B, C) was localized in the forming somites and paraxial mesoderm, which was concordant with what is known from the literature (Chan 2001; Nakamoto 1999; Naruse-Nakajima 2001). The expression pattern of the same developmental stage found in BlEph2 seemed to follow the segmentation of the somites and would perhaps be the peripheral nerves that would innervate the future musculature (Fig.12 B, C). To be able to determine if this pattern was real and eliminate the background it should be performed other tests such as designing better probes, perhaps in the region of the 3’ UTR that is more specific and so, more selective of each gene. Perhaps the hybridization temperature could also be raised from the 70ºC used to increase the specificity of hybridization. Or perhaps this gene is expressed ubiquitously in the entire organism through development having an important role in tissue formation.

The expression observed in the anterior and posterior notochord and in the neural tube in all four genes (Fig.11-14) was concordant with the expression pattern of these genes in the other Chordates (Picco 2007). This could be an evidence for the phylogenetic position of the cephalochordate basally in the chordate group. Even though amphioxus do not express bona fide migrating neural crest cells they present a structure similar to

this and that expresses several genes common to the vertebrates NCC. For this reason and having in account that NCC are also involved in the development of several facial structure including elements of the oral apparatus and its innervations, it justifies the strong expression observed in both Ephrin receptors (BlEph1 and BlEph2) and ligands (Blefn1 and Blefn2) in pre-mouth embryos and larvae (Fig.11-14). But it was also observable that these genes were expressed in later developmental stages (from 36hrs post-fertilization), perhaps because this structure only develops much later or due to the establishment of the nerve networks later during development (Yasui 2008). This may suggest that at least at the split that established the lineage leading to Chordates, Eph/ephrin genes were co-opted to play a role in migrating NCC and in the formation of the mouth. To be able to investigate these hypotheses other tests should be performed, such as an immunohystochemistry or histological studies and also expression experiments of Eph/ephrin genes in other organisms such as urochordates and other non-chordates.

To be able to have a better interpretation of these results it would also be an interesting approach to perform treatments with ephrin inhibitors. Some of the possible drugs would be Nilotinib (AMN107 from Novartis Pharma, Basel, Switzerland) (Piccaluga 2007); PD173955 (Pyrido[2,3-d]pyrimidine) (Caligiuri 2006); or even peptides constructed to target specifically the ligand-binding-cleft of the Eph receptors such as the previously reported SNEW and TNYL for EphB receptors (Chrencik 2007), and KYL for EphA receptors (Noberini 2008) tested in humans cells. Also other synthesized compounds may be used such as anilinopyrimidine or xanthine derivates (Pasquale 2010) or other general kinase inhibitors that target ephrins as well. As many exist it is important to select the ones that do not target too many other kinases to ensure specificity of the results. Some examples of these compounds are ABT-869, MLN-8054, ZD-6474, VX-745, CI-1033 (canertinib), BIRB-796 or BAY-43-9006 (sorafinib). (Fabian 2005, Karaman 2008, Manning 2002) Once these essays are done it is also important to detect the expression of the Ephrin receptors and ligands in the treated animals.

6. CONCLUSION

With this work we were able to elucidate some questions about the expression and phylogenetic position of the Eph/ephrin gene family in the European amphioxus (*Branchiostoma lanceolatum*). The Ephrin receptors and ligands were identified and annotated in *Branchiostoma floridae* (BfEph1, BfEph2 for the receptors and Bfefn1, Bfefn2 and Bfefn3 for the ligands), cloned in *Branchiostoma lanceolatum* (BlEph1, BlEph2 for the receptors and Blefn1 and Blefn2 for the ligands) and their expression by whole-mount *in situ* hybridization determined in embryos of several stages (from 32 cells morula to 60hrs larvae). The Ephrin receptors and ligands seemed to express in amphioxus in patterns mostly localized in the forming mouth apparatus and somites but also in the notochord, such as observed in vertebrates, which suggests a developmental role in these regions that might have been at least co-opted in the transition to chordates. But also, by consequence, they would have to play a role that is observed from sponges to vertebrates such as cell migration, positioning, and tissue arrangement. Even more, the Eph and ephrin protein structure seemed to be maintained across species. All of this, added to the phylogenetic topology, suggests a conservation of structure and some developmental roles through evolution, co-option by duplication for the new developmental roles. Moreover there is some evidence of convergent evolution across species and parallel evolution within species at the level of the ephrin receptors and ligands. Cephalochordates seem to occupy, relative to the analysis of these genes, a position in between non-chordates and vertebrates, as suggested by their expression sites and phylogenetic position, supporting the new phylogeny proposed by Delsuc *et al* in 2006. It is also apparent that several lineage specific duplications have occurred across species for both receptors and ligands. Generally, it can be suggested that Eph/ephrin genes diversified several times during metazoan evolutionary, and that the ancestor of urochordates and vertebrates likely possessed a single Eph receptor and two ligands. However, additional data on the function, gene expression patterns and phylogeny of the Metazoa would be necessary to better address the evolutionary relationships of the Ephrin receptors and ligand family, and their role during evolution.

7. ACKNOWLEDGEMENTS

I would like to thank all the team of Dr. Jordi Garcia-Fernàndez for receiving and supporting me during the elaboration and dissertation of this Masters' thesis project.

To Nacho Maeso for all the help with the phylogenetic analysis.

To Dr. Ildiko Somorjai for support with the *in situs* and the research for the ephrin inhibitor drug.

To Dr. Salvatore d'Aniello, Manuel Irimia, Demian Burguera, Dr. Juan Pascual, Hannah Katz, and Willian Silva for all the support during the time I passed in the group.

To Dr. Vitor Almada, Dr. André Levy, Dr. Pedro David, Dr. Pierre Brezellec for the help and support with the phylogenetic analysis.

To the Gabinete de relaçoes internacionais da FCUL and the Gabinete Erasmus of the Universitat de Barcelona that made this experience possible.

To all the people that were part of this international and scientific experience, family and friends, for their paciente, comprehension, presence and emotional support.

BIBLIOGRAPHY

Arthur, et al Eph/ephrinB Mediate Dental Pulp Stem Cell Mobilization and Function. Journal of Dental Research (2009).

Arvanitis, et al. Eph/ephrin signaling: networks. Genes & Development (2008).

Birney, et al. GeniWise and Genome wise. Genome Res. May. (2004)

Blits-Huizinga., et al. Ephrins and their receptors: Binding versus biology. Iubmb Life (2004).

Bruckner, et al. Signaling by Eph receptors and their ephrin ligands. Current Opinion in Neurobiology (1998).

Carroll, S. Evolution at Two Levels: On Genes and Form. PLoS Biol (2005).

Chan, t al. Morphogenesis of prechordal plate and notochord requires intact Eph/ephrin B signaling. Developmental Biology (2001).

Chrencik,. et al. Three-dimensional structure of the EphB2 receptor in complex with an antagonistic peptide reveals a novel mode of inhibition. Journal of Biological Chemistry (2007).

Coulthard, et al. The role of the Eph-ephrin signalling system in the regulation of developmental patterning. International Journal of Developmental Biology (2002).

D'Aniello. et al. Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. Molecular Biology and Evolution (2008).

David, Evolution and development: some insights from evolutionary theory. Anais Da Academia Brasileira De Ciencias (2001).

Davy, . et al. Ephrin signaling in vivo: Look both ways. Developmental Dynamics (2005).

Decaestecker, et al. Host-parasite 'Red Queen' dynamics archived in pond sediment. Nature (2007).

Delsuc, et al. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, (2006).

Drescher, Eph family functions from an evolutionary perspective. Current Opinion in Genetics & Development (2002).

Drescher, et al. IN-VITRO GUIDANCE OF RETINAL GANGLION-CELL AXONS BY RAGS, A 25 KDA TECTAL PROTEIN RELATED TO LIGANDS FOR EPH RECEPTOR TYROSINE KINASES. Cell (1995).

Fabian, et al. A small molecule-kinase interaction map for clinical kinase inhibitors. Nature Biotechnology (2005).

Felsenstein,. Inferring Phylogenies. Sinauer Associates, Sunderland, MA. (2004)

Flanagan. et al. The ephrins and Eph receptors in neural development. Annual Review of Neuroscience (1998).

Force, et al. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. (1999)

Fuentes, et al. Insights into spawning behavior and development of the European amphioxus (Branchiostoma lanceolatum). Journal of Experimental Zoology Part B-Molecular and Developmental Evolution (2007).

Gale,et al. Ephrins and their receptors: a repulsive topic? Cell and Tissue Research (1997).

Garcia-Fernandez, et al., Archetypal organization of the amphioxus Hox gene cluster. Nature. (1994)

Garcia-Fernàndez, et al. From the American to the European amphioxus: towards experimental Evo-Devo at the origin of chordates. The international journal of Developmental Biology. (2008).

Garcia-Fernandez, et al. It's a long way from amphioxus: descendants of the earliest chordate. Bioessays (2009).

Garcia-Fernandez,et al. From the American to the European amphioxus: towards experimental Evo-Devo at the origin of chordates. International Journal of Developmental Biology (2009).

Graham, The evolution of the vertebrates - genes and development. Current Opinion in Genetics & Development (2000).

Graham, The development and evolution of the pharyngeal arches. Journal of Anatomy (2001).

Hall.. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser (1999)

Hall. Evo-Devo: evolutionary developmental mechanisms. International Journal of Developmental Biology (2003).

Helbling, et al. The receptor tyrosine kinase EphB4 and ephrin-B ligands restrict angiogenic growth of embryonic veins in Xenopus laevis. Development (2000).

Himanen, et al. Eph receptors and ephrins. International Journal of Biochemistry & Cell Biology (2003).

Himanen. et al. Eph signaling: a structural view. Trends in Neurosciences (2003).

Himanen, et al. Cell-cell signaling via Eph receptors and ephrins. Current Opinion in Cell Biology (2007).

Holder, et al. The Eph receptor tyrosine kinases and the ephrins - roles in nervous system development. European Journal of Neuroscience (1998).

Holder, et al. Eph receptors and ephrins: effectors of morphogenesis. Development (1999).

Holland, et al. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Research (2008).

Hollandet al. Evolution of neural crest and placodes: amphioxus as a model for the ancestral vertebrate? Journal of Anatomy (2001).

Holland, et al. The chordate amphioxus: an emerging model organism for developmental biology. Cellular and Molecular Life Sciences (2004).

Holland, et al. AmphiPax3/7, an amphioxus paired box gene: insights into chordate myogenesis, neurogenesis, and the possible evolutionary precursor of definitive vertebrate neural crest. Evolution & Development (1999).

Holland, Gene duplication: Past, present and future. Seminars in Cell & Developmental Biology (1999).

Holland, et al. Sequence and developmental expression of AmphiDII, an amphioxus Distal-less gene transcribed in the ectoderm, epidermis and nervous system: Insights into evolution of craniate forebrain and neural crest. Development (1996).

Holland, et al., Gene duplications and the origins of vertebrate development. Development. Supp (1994)

Jacob. EVOLUTION AND TINKERING. Science (1977).

Jeffery, et al.Migratory neural crest-like cells form body pigmentation in a urochordate embryo. Nature (2004)

Jeffery,, Ascidian neural crest-like cells: phylogenetic distribution, relationship to larval complexity, and pigment cell fate. J Exp Zool B Mol Dev Evol. (2006)

Kalo, et al. Signal transfer by Eph receptors. Cell and Tissue Research (1999).

Kaltenbach, et al. The origin and migration of the earliest-developing sensory neurons in the peripheral nervous system of amphioxus. Evolution & Development (2009).

Karaman, et al. A quantitative analysis of kinase inhibitor selectivity. Nature Biotechnology (2008).

Klein. Excitatory Eph receptors and adhesive ephrin ligands. Current Opinion in Cell Biology (2001).

Kondrashov, et al. Selection in the evolution of gene duplications. Genome Biology (2002).

Koonin, Orthologs, paralogs, and evolutionary genomics. Annual Review of Genetics (2005).

Koonin, et al. Is evolution Darwinian or/and Lamarckian? Biology Direct (2009)

Kowalevsky. Entwicklungsgeschichte des Amphioxus lanceolatus. Mém. Acad. Sci. St. Petersb. (1867)

Krull, et al. Interactions of Eph-related receptors and ligands confer rostrocaudal pattern to trunk neural crest migration. Current Biology (1997).

Kullander, & Klein. Mechanisms and functions of EPH and ephrin signalling. Nature Reviews Molecular Cell Biology (2002).

Larkin, et al. Clustal W and Clustal X version 2.0. Bioinformatics. (2007)

Lundin. Evolution of the vertebrate genome as reflected in para- logous chromosomal regions in man and the house mouse. Genomics. (1993)

Luukko, et al. Expression of ephrin-A ligands and EphA receptors in the developing mouse tooth and its supporting tissues. Cell and Tissue Research (2005).

Lynch, et al. The evolutionary fate and consequences of duplicate genes. Science (2000).

Lynch, et al. The probability of duplicate gene preservation by subfunctionalization. Genetics (2000).

Manning, et al. The protist, Monosiga brevicollis, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. Proceedings of the National Academy of Sciences of the United States of America (2008).

Mellitzer, et al. Control of cell behaviour by signalling through Eph receptors and ephrins. Current Opinion in Neurobiology (2000).

Mellott, & Burke. Divergent roles for Eph and Ephrin in Avian Cranial Neural Crest. Bmc Developmental Biology (2008).

Mellott, Burke,. The molecular phylogeny of eph receptors and ephrin ligands. Bmc Cell Biology (2008).

Minguillon, et al. The amphioxus Hairy family: differential fate after duplication. Development (2003).

Nakamoto, M. Eph receptors and ephrins. International Journal of Biochemistry & Cell Biology 32, 7-12 (2000).

Naruse-Nakajima, et al. Involvement of EphA2 in the formation of the tail notochord via interaction with ephrinA1. Mechanisms of Development (2001).

Noberini, et al. Small Molecules Can Selectively Inhibit Ephrin Binding to the EphA4 and EphA2 Receptors. Journal of Biological Chemistry (2008).

Ohno, S., Evolution by gene duplication. London, New York, Allen & Unwin; Springer-Verlag, (1970)

O'Leary, et al. Eph receptors and ephrins in neural development. Current Opinion in Neurobiology (1999).

Pasquale. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. Nature Reviews Cancer (2010).

Piccaluga, et al. Tyrosine kinase inhibitors for the treatment of Philadelphia chromosome-positive adult acute lymphoblastic leukemia. Cancer (2007).

Picco, et al. Ephrin-Eph signalling drives the asymmetric division of notochord/neural precursors in Ciona embryos. Development (2007).

Poliakov, et al. Diverse roles of Eph receptors and ephrins in the regulation of cell migration and tissue assembly. Developmental Cell (2004).

Putnam, et al. The amphioxus genome and the evolution of the chordate karyotype. Nature (2008).

Risley, et al. EphB2 and EphB3 forward signalling are required for palate development. Mechanisms of Development (2009).

Santiago, et al. Ephrin-B ligands play a dual role in the control of neural crest cell migration. Development (2002).

Satou, et al. A genomewide survey of developmentally relevant genes in Ciona intestinalis - V. Genes for receptor tyrosine kinase pathway and Notch signaling pathway. Development Genes and Evolution (2003).

Schmucker, et al. Signaling downstream of Eph receptors and ephrin ligands. Cell (2001).

Sela-Donenfeld, et al. Eph receptors: Two ways to sharpen boundaries. Current Biology (2005).

Seo, et al. Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. Nature. (2004)

Sharman, et al. An amphioxus Msx gene expressed predominantly in the dorsal neural tube. Dev Genes Evol, (1999)

Shimeld, The evolution of dorsoventral pattern formation in the chordate neural tube. American Zoologist (1999).

Sly, et al. Who came first—larvae or adults? origins of bilaterian metazoan larvae. Int. J. Dev. Biol. (2003)

Suga,et al. Protein tyrosine kinase cDNAs from amphioxus, hagfish, and lamprey: Isoform duplications around the divergence of cyclostomes and gnathostomes. Journal of Molecular Evolution (1999).

Suga, et al. Extensive gene duplication in the early evolution of animals before the parazoan-eumetazoan split demonstrated by G proteins and protein tyrosine kinases from sponge and hydra. Journal of Molecular Evolution (1999).

Suga, et al. Ancient divergence of animal protein tyrosine kinase genes demonstrated by a gene family tree including choanoflagellate genes. Febs Letters (2008).

Suga, et al. Ancient divergence of animal protein tyrosine kinase genes demonstrated by a gene family tree including choanoflagellate genes. Febs Letters (2008).

Surawska, et al. The role of ephrins and Eph receptors in cancer. Cytokine & Growth Factor Reviews (2004).

Tamura, et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution (2007).

True, et al, Gene co-option in physiological and morphological evolution. Annual Review of Cell and Developmental Biology (2002).

von Dassow, et al. Modularity in animal development and evolution: Elements of a conceptual framework for EvoDevo. Journal of Experimental Zoology (1999).

Wang. et al. Eph family transmembrane ligands can mediate repulsive guidance of trunk neural crest migration and motor axon outgrowth. Neuron (1997).

Wilkinson. Multiple roles of Eph receptors and ephrins in neural development. Nature Reviews Neuroscience (2001).

Xu, et al. Roles of Eph receptors and ephrins in segmental patterning. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences (2000).

Yasui, et al. The Lancelet and Ammocoete Mouths. Zoological Science (2008).

Yip,. et al. An integrated system for studying residue coevolution in proteins. Bioinformatics (2008).

Yeh, et al. Computational Inference of Homologous Gene Structures in the Human Genome. Genome Research (in press). (2001)

Zhang,. Evolution by gene duplication: an update. Trends in Ecology & Evolution (2003).

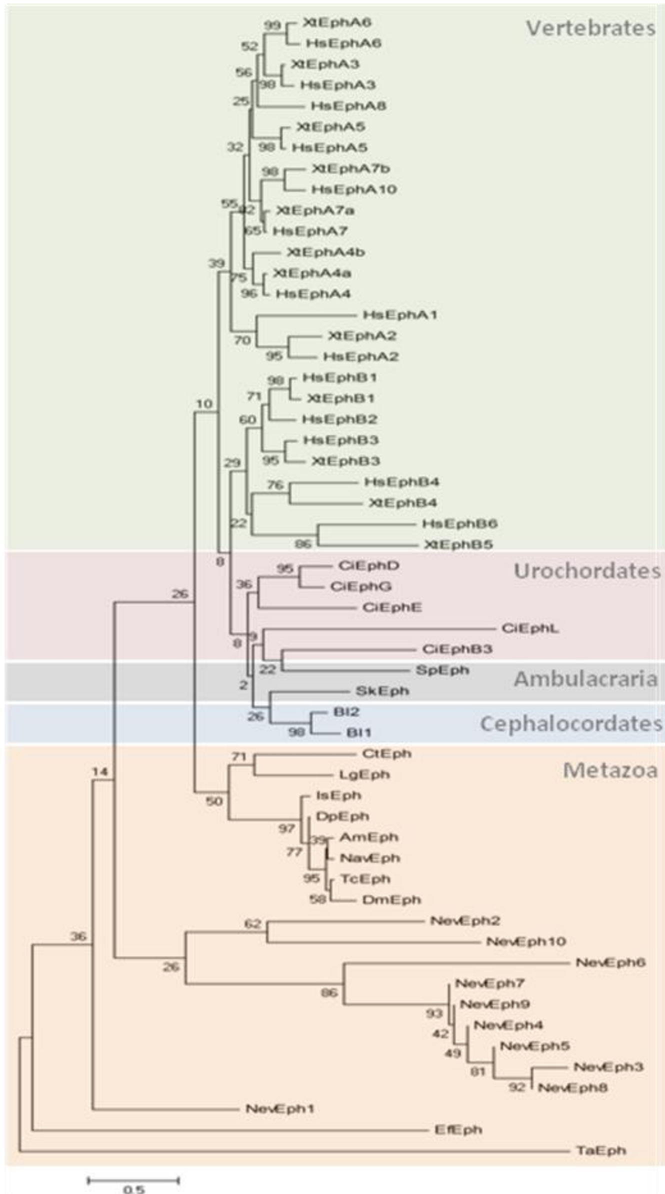

Fig. 1- Unrooted phylogenetic tree for the Eph receptors (ligand-binding domain) constructed by the Minimum-Evolution method, Pairwise deletion, Poisson correction, Gamma distributed (1.0). Numbers represent bootstrap values for 1000 replicates. Four major groups are represented (Vertebrates, Urochordates, Ambulacraria, Cephalocordates, Metazoa). Am, *Apis mellifera;* Bl, *Branchiostoma lanceolatum;* Ci, *Ciona intestinalis;* Ct, *Capitella teleta;* Dm, *Drosophila melanogaster ;* Dp, *Daphnia pulex ;* Ef, *Ephydatia fluviatilis ;* Hs, *Homo sapiens;* Is, *Ixodes scapularis;* Lg, *Lottia gigantea;* Nav, *Nasonia vitripennis;* Nev, *Nematostella victensis;* Sk, *Saccoglossus kowalevskii;* Sp, *Strongylocentrotus purpuratus;* Ta, *Trichoplax adhaerens;* Tc, *Tribolium castaneum;* Xt, *Xenopus tropicalis.* The scale bar corresponds to the number of substitutions per site. Branch length indicates the amount of divergence inferred.

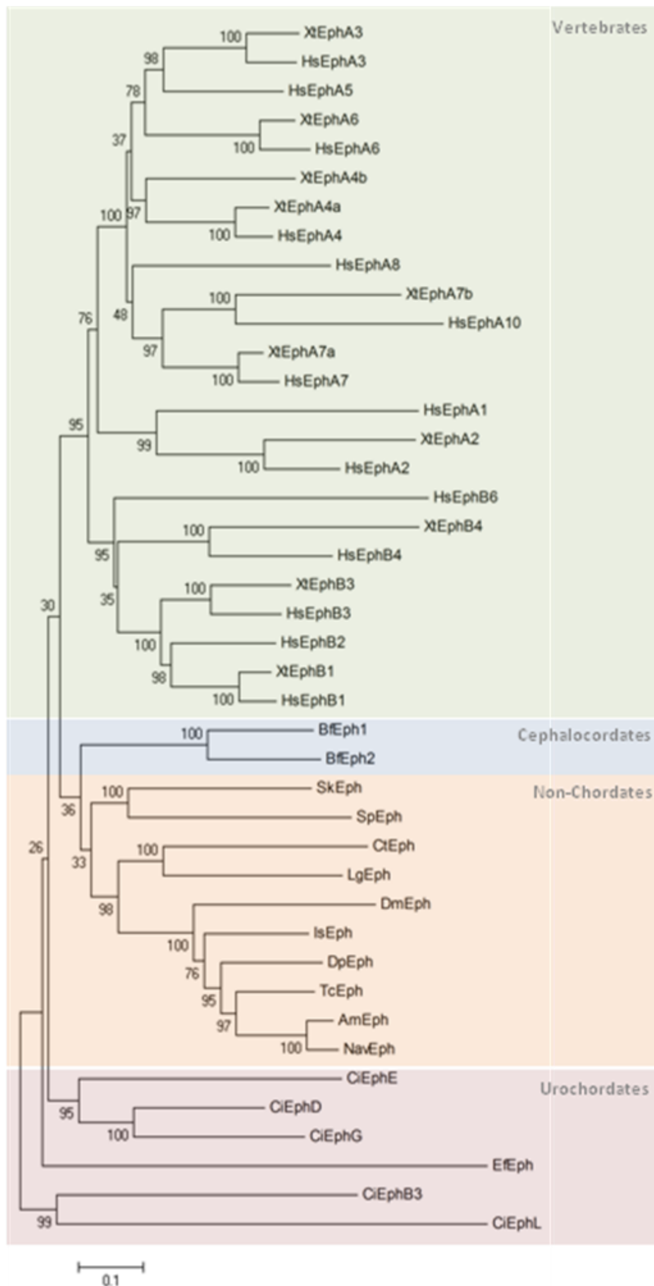

Fig. 2- Unrooted phylogenetic tree for the Eph receptors (whole protein, excluding *Nematostella vectensis* genes) constructed by Neighbor-Joining method with Gamma distributed rates among sites (1.0), pairwise deletion of the gaps/missing data and Poisson correction. Numbers represent bootstrap values for 1000 replicates. In this analysis only the genes that included all the Eph domains (see Fig. 7) were included, the other genes were excluded from the analysis such as *Nematostella vectensis,* to reduce the phenomenon of long-branch attraction. Four major groups were created to facilitate the description (Vertebrates, Urochordates, Cephalocordates, and Non-Chordates). Am, *Apis mellifera;* Bf, *Branchiostoma floridae;* Ci, *Ciona intestinalis;* Ct, *Capitella teleta;* Dm, *Drosophila melanogaster ;* Dp, *Daphnia pulex ;* Ef, *Ephydatia fluviatilis ;* Hs, *Homo sapiens;* Is, *Ixodes scapularis;* Lg, *Lottia gigantea;* Nav, *Nasonia vitripennis;* Sk, *Saccoglossus kowalevskii;* Sp, *Strongylocentrotus purpuratus;* Tc, *Tribolium castaneum;* Xt, *Xenopus tropicalis.* The scale bar corresponds to the number of substitutions per site. Branch length indicates the amount of divergence inferred.

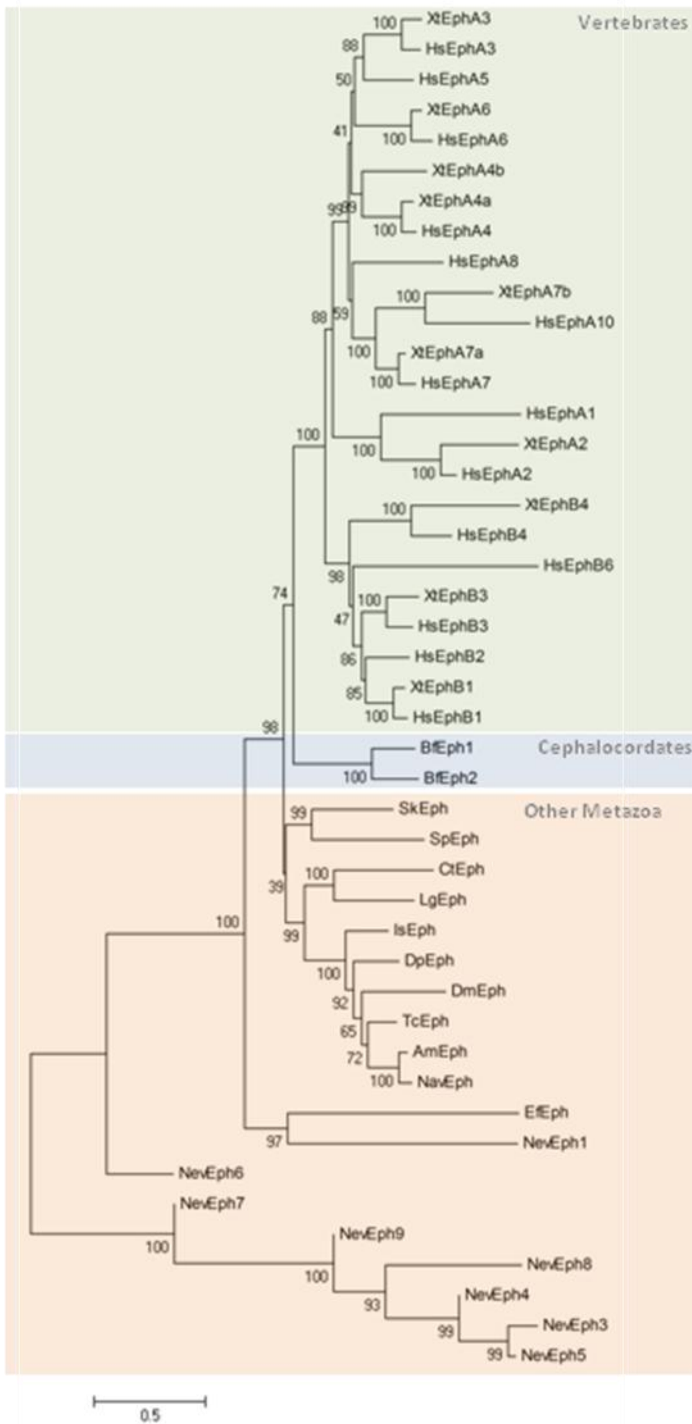

Fig.3 - Unrooted phylogenetic tree for the Eph receptors (whole protein, excluding *Ciona intestinalis* genes) constructed by Neighbor-Joining method with Gamma distributed rates among sites (1.0), pairwise deletion of the gaps/missing data and Poisson correction. Numbers represent bootstrap values for 1000 replicates. In this analysis only the genes that included all the Eph domains (see Fig. 7) were included, the other genes were excluded from the analysis as well as *Ciona intestinalis* genes, due to their specific independent duplication that could be in the origin of a phenomenon of Long Branch Attraction. Three major groups were created to facilitate the description (Vertebrates, Cephalocordates, and Non-Chordates). Am, *Apis mellifera;* Bf, *Branchiostoma floridae;* Ct, *Capitella teleta;* Dm, *Drosophila melanogaster ;* Dp, *Daphnia pulex ;* Ef, *Ephydatia fluviatilis ;* Hs, *Homo sapiens;* Is, *Ixodes scapularis;* Lg, *Lottia gigantea;* Nav, *Nasonia vitripennis;* Nev, *Nematostella vectensis;* Sk, *Saccoglossus kowalevskii;* Sp, *Strongylocentrotus purpuratus;* Tc, *Tribolium castaneum;* Xt, *Xenopus tropicalis.* The scale bar corresponds to the number of substitutions per site. Branch length indicates the amount of divergence inferred.

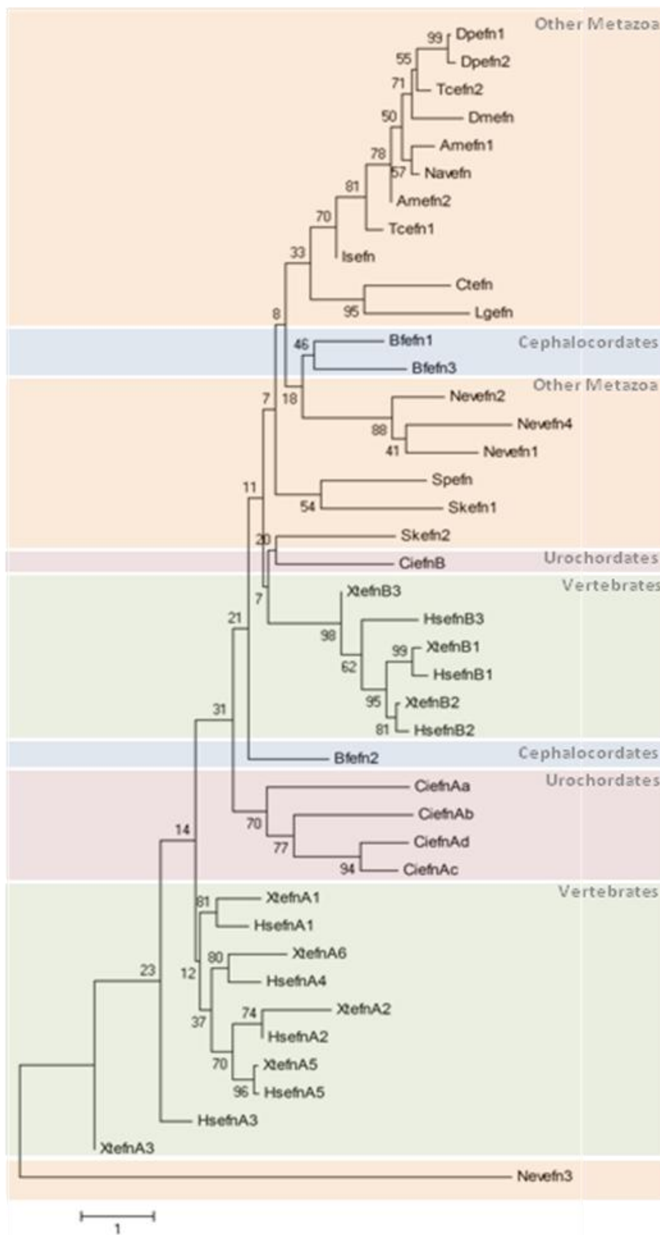

Fig. 4- Unrooted phylogenetic tree for the ephrin ligands (efn) constructed by the Neighbor-Joining method with Gamma distributed rates among sites (1.0), pairwise deletion of the gaps/missing data and Poisson correction. Numbers represent bootstrap values for 1000 replicates. Four major groups were created to facilitate the description (Vertebrates, Urochordates, Cephalocordates, and Other Metazoa). Am, *Apis mellifera;* Bf, *Branchiostoma floridae;* Ci, *Ciona intestinalis;* Ct, *Capitella teleta;* Dm, *Drosophila melanogaster ;* Dp, *Daphnia pulex ;* Ef, *Ephydatia fluviatilis ;* Hs, *Homo sapiens;* Is, *Ixodes scapularis;* Lg, *Lottia gigantea;* Nav, *Nasonia vitripennis;* Nev, *Nematostella victensis;* Sk, *Saccoglossus kowalevskii;* Sp, *Strongylocentrotus purpuratus;* Tc, *Tribolium castaneum;* Xt, *Xenopus tropicalis.* The scale bar corresponds to the number of substitutions per site. Branch length indicates the amount of divergence inferred.