Contribution of homeobox genes to the development of the oikoplastic epithelium, a major novelty of tunicate Larvaceans

Yana Mikhaleva

Dissertation for the degree of philosophiae doctor (PhD)

at the University of Bergen

2016

Dissertation date: 23.11.2016

Scientific environment.

The work presented in this Ph.D. thesis was carried out in the research group of Dr.Daniel Chourrout at the *Sars International Centre for Marine Molecular Biology* as part of the Ph.D. program at the *Department of Molecular Biology* at the *Faculty of Mathematics and Natural Science* of the *University of Bergen*.

This work is entirely reported in four research articles, including three already published and one submitted for publication. Other members of the group who contributed are listed in the authors of these reports. I am the first author and obtained most results in two of them, with external collaborations (Pr. Joel Glover and Dr. Orsolya Kreneisz, University of Oslo; Pr. Eric Thompson from Sars International Centre for Marine Molecular Biology, Bergen). The main collaborating institute in the other papers has been Genoscope, Genome Center in Evry (France) with essential contributions from Dr. Patrick Wincker and Dr. France Denoeud. The funding from Sars International Centre, the Research Council of Norway, Evo-Net grants and Genoscope supported this research project.

Acknowledgements

I am extremely grateful to everyone who supported and helped me during all the years of work on the thesis. It was a long way with inspiring successes and devastating failures, small wins and big challenges.

And first of all, I would like to thank my supervisor Daniel Chourrout for his enthusiasm and belief in this project. He has been the excellent supervisor, with never ending inspiration and encouragement, especially when the things have been difficult. Thank you for all these years I worked in your lab, for your guidance, pieces of advice, support in my scientific work and always the open door in your office.

I wish to thank my co-supervisor Hee-Chan Seo, who introduced me to molecular biology methods and supported me in the beginning of my lab-work.

I also appreciate the kind cooperation of Eric Thompson and Joel Glover for the effort to improve manuscripts, comments, and discussions during my work.

Big thanks also go to former and current members of the S1 lab for their excellent scientific and technical advice, friendly atmosphere and willingness to help whenever it needed. I am very grateful to Marit Flo Jensen for her help with the huge amount of *in situ* hybridizations we performed during this work and for all our "non-scientific" discussions and "do like locals" advice, that made my life in Norway even more comfortable.

I sincerely appreciate continuous support from all Appy-park people who maintain *Oikopleura* culture and provide animals for experiments. I also would like to thank all Sarsians for the good social environment and friendly support, and my special thanks to Kevin Pang for helping with thesis proofreading.

Finally, I would like to express tons of gratitude to my “big family” in Russia, and especially to my mom, who always supported me and took care of my health during this long journey. I'm very thankful to my best friend from the school age - Julia Haaland. I'm so lucky to have a friend like you, with your support I'll never need a help from a psychologist! Above all I grateful to my "small family" in Norway, who never give me forget that the real world exists outside the working place and most important people in my life with their definitive and unlimited love, are always on my side.

Contents

Summary.

The development of complex organisms is controlled by numerous genes, including developmentally regulated transcription factors (TF) encoded by multiple gene families. A well diversified group of TFs is the superclass of homeodomain proteins, among which some have a well studied and relatively well conserved function during the development of metazoans. The most popular homeodomain TFs are the *Hox* genes, core players in the AP axis patterning, with a peculiar genome organization in clusters which ensures the coordination of their spatio-temporal expression. However, the function of Hox and many other conserved homeobox genes has independently evolved in distinct metazoan lineages, contributing to changes of their body plan. This divergent evolution also led to their participation in lineage specific innovations. Such radical changes are governed by conserved genes whose expression patterns were importantly modified, as well as by new genes appearing in specific lineages. Gene duplication is one mechanism leading to genes with divergent functions.

When changes in coding sequences are scored among distinct genomes, it clearly appears that some metazoan lineages have evolved more rapidly than others. This is the case for tunicates, compared to other chordates. Within tunicates, larvaceans (=appendicularians) also evolved more rapidly than others classes, as judged from genome data of *Oikopleura dioica*, a cosmopolitan species that has become easily bred in the laboratory. The rapid genome evolution in this group is also visible through profound changes of the complement of genes (loss and duplication of genes), their intron-exon organization or their order on the chromosomes. A very speaking example is the Hox cluster disintegration in this lineage.

We contributed to show that in *O. dioica*, an important fraction of the homeobox genes has been lost, or at least fails detection by classical gene alignment tools. We also noticed that in the homeobox genes that remain present, a relatively large proportion have been duplicated once or several times. We then revealed, using *in situ* hybridization (ISH) in embryos and young larvae that the members of most amplified homeobox gene groups are expressed in the epithelium of the trunk. This happens very rarely for homeobox genes that have no duplicate in the genome. Since the expression in the trunk epithelium occurs when it is gradually transformed in the oikoplastic epithelium that later on generates the house (a large and complex filter-feeding apparatus typical

of larvaceans), we postulated that duplicated genes play a role in the formation of this novel structure. If this is true, the question of whether and how gene duplication events have been involved in the evolution of the oikoplastic epithelium and the house also needed to be addressed.

When we began our work, the function of *O. dioica* genes could be apprehended only with ISH and quantification of gene expression. We therefore devoted major efforts to experimentally modify their expression levels in adapting from other species a technique of RNA interference (RNAi) based on the introduction of double stranded RNA matching the gene sequence of interest. We contributed to validate this technique in knocking down the expression of three genes, chosen because of their easily predictable loss of function phenotype. These genes are *Brachyury*, which encodes a T-Box transcription factor and has a major role in the formation of the notochord, and two genes encoding essential enzymes of the cholinergic (ChAT=choline acetyltransferase) and GABA-ergic (GAD= glutamic acid decarboxylase) pathways, respectively. After injection of dsRNA for each of these three genes in defined conditions, we indeed observed an aborted tail development for *Brachyury*, a severe inhibition of tail locomotion for *ChAT*, and an uncoordinated swimming for *GAD*.

Before applying this technique to homeobox genes expressed in the trunk epithelium, we morphologically described the epithelium development during the hours after hatching. We could observe the formation from a rather uniform cell layer of individual cellular fields that characterize the oikoplastic epithelium and we monitored the concomitant expression initiation of several house protein genes (oikosins) during this process. A variable number of cell divisions occur in each forming field, and we observe no migration of cells between distinct fields. Homeobox genes revealed in the initial screen were intensely and transiently expressed during the formation of the fields and we tentatively mapped their expression domain at an early and a late stage of the process. As mentioned above, most of these genes are duplicated and it is noteworthy that duplicates of a given gene are expressed in similar regions, usually together or near each other. Consequently, the epithelial expression very likely predated the duplication events. If these homeobox genes are involved in the transition towards an oikoplastic epithelium, their duplication was probably not essential for its emergence during evolution. Duplications may have contributed to increase the complexity of this structure through neofunctionalization, and/or resulted in a division of the original gene function among duplicates.

To prove the involvement of homeobox genes in the formation of the oikoplastic epithelium we attempted RNA interference for genes that are expressed in epithelium at the early larval stage: three *otx* genes and two *prop* genes. The epithelial expression of *prop* genes was efficiently knocked down. We observed the malformation and disorganization of nuclei of epithelial cells around the dorsal midline, where both *prop* genes are normally expressed. The knockdown of *prop* genes also abolished the post-metamorphosis expression of one oikosin gene (*oik41a*) in the Anterior rosette field after metamorphosis, confirming their importance for cell differentiation in this region. In contrast, the attempts of RNAi for the three *otx* genes were not conclusive, partly because they provoked severe developmental abnormalities during embryogenesis.

Overall, our expression study supports that multiple homeodomain transcription factors have been recruited in genetic pathways governing the development of the oikoplastic epithelium, which synthesizes and secretes diverse components of the larvacean house. We bring evidence for the involvement of at least two homeobox genes (*prop* duplicates) in this process, upstream of the ultimate cell differentiation in a specific region. Other transcription factors including Fox genes are also expressed in the trunk epithelium during early larval stages. Although most homeobox genes transiently expressed in the epithelium of the larva trunk are duplicated, there is no indication that duplications proper were at the origin of this evolutionary innovation.

Introduction

The origin of living species diversity has been a major enigma for centuries, with an increased interest from every new generation of scientists. Biodiversity arises through the emergence of new organs and structures that favorize adaptation to specific environmental conditions. Numerous forms of adaptations exist. For example, the acquisition of a given color pattern can allow an organism to become less visible on its background. Harmless animals can get better protected from predators if they can imitate the color and shape of the body of their poisonous relatives. Other examples include biochemical adaptations (ability to produce poisons) and behavioral adaptations (optimization of offspring care). Adaptations to extreme environmental conditions are well-documented in desert and polar animals. A special place in this list of adaptations belongs to morphological adaptation. Examples are the webbing between the toes found in various aquatic animals (amphibians, birds, and others), the long legs and neck in marsh birds, and the flattened body of bottom-dwelling fish (rays, flatfishes, etc.). All these phenotypically visible adaptations are inscribed in the genome and faithfully transmitted to the next generation. Comparing the genomes of closely related animals is now possible at a relatively low cost, and partly aims at identifying genome changes that underly morphological diversity.

1.1. Mysterious animals

Tunicates comprise very diverse forms of marine animals, already identified by Aristotle (384-322 B.C.), who placed them together with sponges and holothurians in his classification, based on their degree of perfection and their ability to move. Tunicate adults looked so different from other animals that their exact position in the animal phylogeny remained unclear until the 19th century. The Russian embryologist Alexander Kovalevsky discovered that ascidian tadpole-like larvae (ascidians are sessile tunicates) display all features of chordates - a notochord, a hollow dorsal nerve cord, pharyngeal slits, an endostyle and a post-anal tail, - and he placed them into the Chordata phylum. Appendicularians (= larvaceans) form one class of pelagic tunicates. They were first mentioned in reports of the 1815-1818 Russian worldwide expedition (Fenaux, 1993). However, Huxley (1851) was the first scientist who recognized them as members of the

Tunicata subphylum, based on the "organization of the creature, its wide respiratory sac, its nervous system, its endostyle…". In contrast to ascidians, they remain free-swimming and keep a larval-like body plan until the adult stage. This is the reason why appendicularians are also named larvaceans.

The most striking peculiarity of larvaceans is a transparent house that the animal synthesizes and in which it is enclosed most of the time. Mertens (1831) noted that it is hard to keep captured appendicularians inside "the membranous shell, that can be reformed many times per day" and "the process takes approximately half an hour, and the animal can, whenever it wants, get rid of this strange production" (see Fenaux, 1993). He also suggested that the "house" had a respiratory function because it consisted of a network of filters. The function of the "house" was debated for many years. In 1859 Allman suggested that it served to protect the eggs during their development. In 1872, Fol produced the first classification of larvaceans. He described their feeding behavior by filtration through an intrapharyngeal filter, as well as the house of one species in each family of appendicularians: oikopleurids, fritillariids, and kowalewskiids. The house is specific of this class of tunicates. It is produced by the epithelium of the animal trunk, which has a unique morphology and function. The number and arrangement of distinct cell types in this epithelium is a criterion used for the classification of species in the *Oikopleura* genus. The transparent house, and the epithelium that produces it, can be legitimately considered as an evolutionary novelty born with the appendicularian class.

1.2. The oikoplastic epithelium

The oikoplastic epithelium is a monolayer of cells that covers most of the trunk. Its cells vary in size and shape, and they gather in distinct and easily recognizable groups. These groups of cells were named after the zoologists who had a pioneer contribution to the study of appendicularians. As already mentioned, the details of oikoplastic epithelium organization are species-specific, but oikopleurids all share the same global organization based on the same type of epithelial fields. Figure 1 is a schematic drawing of these epithelial fields (Flood, 2005).

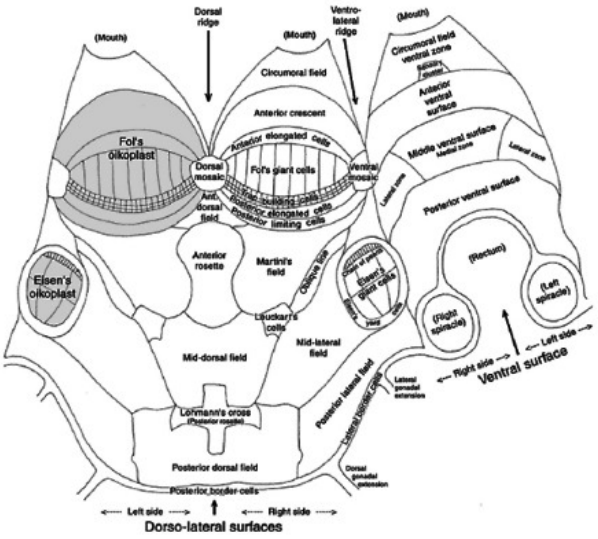

Figure 1. Line drawing of the "idealized" Oikopleuridae epithelium by Flood P.R. (2005)

The epithelium shows left-right symmetry with respect to the dorsal edge (or dorsal midline). As for many of the other organs in Larvaceans, the epithelium keeps a constant number of cells from the metamorphosis stage (Flood, 2005; Spada et al., 2001). As Lohmann and Buckmann (1926) noted, counting the largest and easily recognizable cells in the epithelium - Fol and Eisen giant cells - helps to identify species of the oikopleurid family. For example, *O. dioica* has seven giant Fol cells and seven giant Eisen cells, while Eisen cells are absent in *O. longicauda*. *O. gorskyi* has nine giant Fol cells and eight giant Eisen cells, and *O. labradoriensis* only has six giant Eisen cells (Aravena and Palma, 2002; Flood, 2000; Flood, 2005). Other species-specific epithelium fields were discovered using nuclear staining methods (Fig. 2 (Flood, 2005)). Both dorsal midline structures – the Anterior rosette and the Posterior rosette – are identifiable as highly organized cell groups in all species. In these fields, the number of cells, the shape of their nuclei, and their size and position vary among species (Flood, 2005). Finally, the number of cells, the shape of their nuclei, and their positions in the tissue are invariant among individuals of a given population.

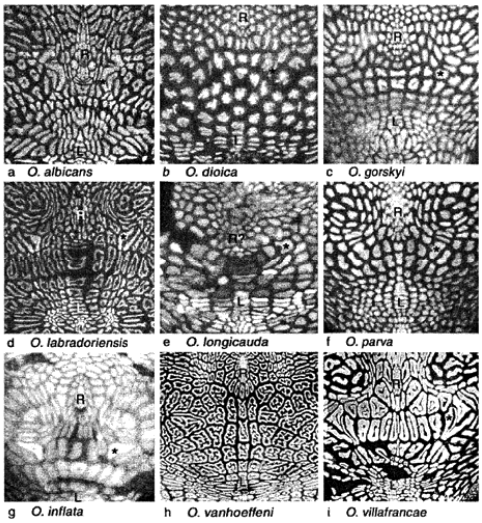

Figure 2. The nuclear pattern of the dorsal epithelium between the Anterior and Posterior rosettes; anterior - up, posterior – down. R – Anterior rosette, L – Posterior rosette. (Flood, 2005).

1.3. The house of larvaceans

The oikoplastic epithelium produces and secretes the components of the larvacean "house" (Spada et al., 2001). The gelatinous house is a complex spheroid system, which contains multiple filters and chambers. The house is mainly an instrument for filter feeding, which permits to select, capture and concentrate food particles. With a sinusoidal movement of the tail, the animal generates a water flow into the house via the inlet filters that exclude the too large particles. Inside the house, fine meshed filters collect small bacteria and planktonic eukaryotes that are then concentrated and directed into the digestive tract.

From metamorphosis onwards, the house is periodically replaced by inflating a new house rudiment named pre-house. The first pre-house is produced by the oikoplastic epithelium around the time of metamorphosis. Up to three pre-houses can be stacked on the trunk before the most outer one is expanded into a functional house. The rates of house production are species-dependent (Sato et al., 2001; Sato et al., 2003). Thus, at room temperature, *O. longicauda* and *O. fusiformis* produce a new house every 1.5 hours, *O. rufensis* generates a new house once every six hours, and *O. cophocerca* only renews its house twice during a 24-hour period (Sato et al.,

2003). The walls of discarded houses aggregate multiple particles coming from the seawater. The clogged filters accumulate phytoplankton cells, bacteria, flagellates, mineral grains, and other particles that were blocked by the filter mesh. Due to the accumulation of organic matter, the discarded houses of larvaceans serve as a suitable food for zooplankton (Hansen et al., 1996) as well as benthic organisms (Robison et al., 2005).

The houses of larvaceans markedly differ in their global shape (spherical, elongated, egg-shaped, or rhomboid in some projections) and size (from 5 mm in *O. dioica* and *O. longicauda* to the size of a chicken egg in *O. vanhoeffni*), as well as in the shape of their filters (Fig.3) (Flood, 2005). The variable number and shape of the Eisen field cells correlate with architectural variations in the inlet filters. In *O. dioica,* the bilateral inlet filters are situated in grooves that are 1 mm deep. In *O. longicauda*, the inlet filter cannot be identified, while the inlet funnels are present. In *O. fusiformis,* two lateral "wings" make the house fairly unique, and these function as a very large, elongated inlet filter. In *O. labradoriensis*, the inlet filters are relatively small and situated in shallow grooves. In *O. vanhoeffeni,* the inlet filters are situated deep within the inlet funnels. Another correlation observed is between the number of giant cells in the Fol region corresponds and the number of filter ridges in the food concentrating filter (Flood, 2005)

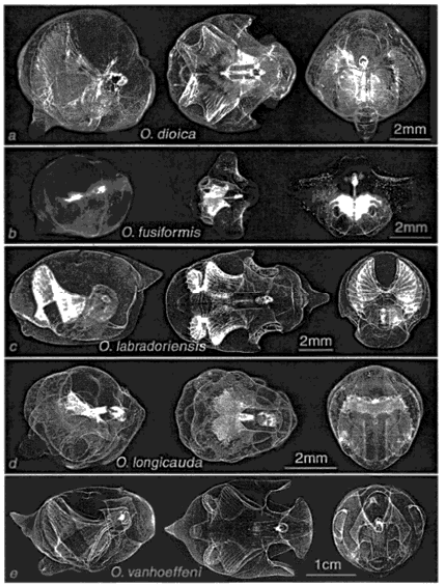

Figure 3. Houses of different *Oikopleura* species in three different projections (Flood, 2005)

The house of *O. dioica* is almost spherical in lateral and dorsal projection and more rhomboid in the frontal projection. Its diameter is 6-8 mm. Walls of the house are transparent and contains five lines of bioluminescent inclusion bodies. These bodies have a pillow shape and are all oriented in A-P direction (Flood, 2005). The animal is situated inside the central chamber, and sinusoidal tail movements create a water flow through the system of filters (Fig.4). First, water enters the house through two inlet filters that are situated in 1 mm deep groves on each side of the house. Then water passes through the inlet funnel down to the tail chamber and is further directed through the food-concentrating filter via two bilateral supply passages. Food particles are trapped by the filter net and directed into the mouth, while the water flow goes into the posterior chamber and out through the exit spout of the house, which is 2-3 mm long and conical in shape. (Bone, 1998; Bouquet et al., 2009; Sato et al., 2001; Troedsson et al., 2009)

The house is replaced on average every 4 hours throughout the life cycle of *O. dioica*. House renewal appears to depend on the house rudiment secretion rate and not on the clogging of the filter, as it was believed. Instead, an increase in temperature and decrease of water salinity will accelerate house renewal (Sato et al., 2001).

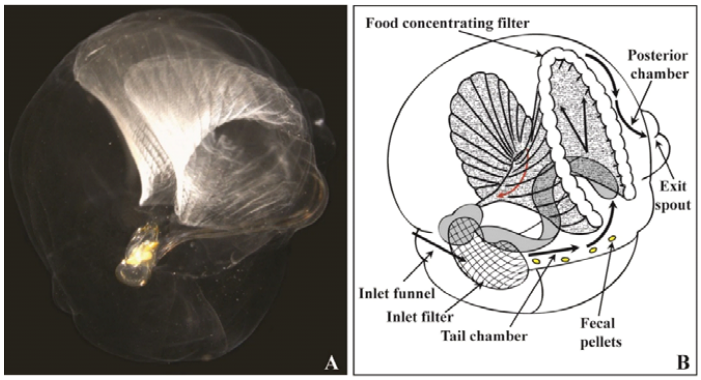

Figure 4. from (Bouquet et al., 2009) (A) O. dioica inside a fully inflated gelatinous filter-feeding house. The ribbed food-concentrating filter is visible at the top. The animal at the center of the panel is oriented with mouth to the top, gonad to the bottom and tail projecting to the right. (B) Schematic representation of the oikopleurid house modified after Flood and Deibel (1998) and Spada et al. (2001). Water flow through the house is indicated by black arrows and food particle concentration toward the mouth by a red arrow.

This type of large complex structure may have evolved primarily to optimize the filter feeding performance in water containing low concentrations of food. Authors who have admired its complexity agree to consider it as a crucial adaptation of appendicularians that has allowed their success (Acuña, 2001; Troedsson et al., 2009). From an evolutionary viewpoint, how the house has appeared and then needed to be refined not only to optimize its function but perhaps also to exert it in the first place is typically a challenging enigma for evolutionary biologists. When more knowledge will be available on how the house structures are genetically engineered and which ones are necessary for the filter feeding to occur, the way the genome has evolved to produce complex houses should stimulate interesting debates.

1.4. Evolutionary innovations.

An important source of morphological diversity in living organisms is the emergence of new structures. We and others use the term "evolutionary innovations" or "novelties" for those structures that display no clear homology with more ancient tissues or organs. Novelties can appear through various mechanisms, including changes of gene content (from duplicates of extant genes to the *de novo* creation of lineage specific genes), changes in coding sequences or regulatory elements of extant genes.

1.4.1. Gene duplication.

Gene duplication can ultimately lead to a novel function. Duplication creates a genetic redundancy for some time. It sometimes allows one of the duplicates to evolve a new function. The duplication occurs through various mechanisms (e.g. unequal crossing over, retroposition, and large-scale duplication). Several fates are possible for the duplicated genes.

Pseudogenes resulting from duplication.

Most genes resulting from a duplication accumulate degenerative mutations and deletions. As a consequence, new non-functional pseudogenes appear (Vanin 1984). The pseudogenes carry

multiple genetic lesions, significant alterations of their 3' UTR, or complete lack of introns (if they arose through retroposition). The biological roles of pseudogenes remain largely unknown. Many efforts were devoted to its understanding until recent times (Mighell et al., 2000), and their biological importance may be real. In chicken, immunoglobulin diversity is generated by gene conversion of the *vh1* gene. A single functional gene encodes the heavy chain variable region of immunoglobulins, whereas pseudogenes on the 5' side of the *vh 1* gene are a resource for enhanced diversity (Ota and Nei, 1995). Also, some pseudogenes can be transcribed, like human 5-HT7 pseudogene (Olsen and Schechter, 1999) and even alternatively spliced, like two human transcripts, yHLAO1, and yHLAO2, of a copper-containing monoamine oxidase pseudogene (Cronin et al., 1998). Although pseudogene transcripts may yield translation products, they are not expected to result in a functional protein. However, the early assumption that transcribed pseudogenes could have a specific function has been confirmed in the work of Hirotsune (Hirotsune et al., 2003): a reduced transcription of a Makorin1-p1 pseudogene, resulting in destabilization of Makorin1 mRNA led to specific pleiotropic phenotypes.

Conservation of gene function.

Unless the duplication has been partial, we can expect that the two daughter genes are functionally redundant immediately after the duplication event. S. Ohno (1970) put forward the theory that one reason for which a duplicated copy with unchanged function could be retained is an increase of the amount of functional protein, for example ribosomal RNA and histone genes (Braastad et al., 2004). Observations from Qian et al. (2010) support that a reduction of gene expression after duplication may facilitate the retention of duplicates. The maintenance of function among duplicated paralogous genes may occur through two mechanisms that limit their sequence divergence – concerted evolution or strong purifying selection (Chen et al., 2010a; Liao, 1999; Zhang, 2003).

Subfunctionalization.

One general process leading to keep both daughter genes in the genome, called subfunctionalization, is the division of ancestral gene function among them. Subfunctionalization as a mechanism for the preservation of both duplicated copies was described by Hughes (Hughes,

1994; Hughes, 2005). In accumulating point mutations each duplicate may keep only part of the function of the ancestral gene, while the other keeps the other part. Subfunctionalization may take different forms. One is illustrated by is the evolution of *crystallin αB* protein in the vertebrate lineage. Human *crystallin* αB combines optical and chaperone functions, whereas in the zebrafish the gene is duplicated and the duplicates took one or the other function (Smith et al., 2006). It was proposed that subfunctionalization could serve as an interim step to achieving the emergence of a new gene function (He and Zhang, 2005; Rastogi and Liberles, 2005). This phenomenon termed "subneofunctionalization" was supported by experimental work from Deng et al. (2010) who showed that the *antifreeze protein gene* of Antarctic zoarcid fish evolved from an old *sialic acid synthase (SAS)* gene by duplication. In one duplicate, the N-terminal SAS domain was deleted and replaced with a nascent signal peptide and the rudimentary ice-binding function was optimized on the C-terminal domain, leading to a secreted protein with non-colligative freezing-point depression. Changes in the coding sequence had taken place in both examples cited above. Another form of subfunctionalization is a modification of CRE (cis - regulatory elements) after gene duplication. It leads to dividing the expression domain among duplicates. One of the examples is the *pax6* gene. The expression pattern for the single copy of *pax6* in birds and mammals represent the sum of expression domains of *pax6.1* and *pax6.2* genes in zebrafish (Nornes et al., 1998). Studies of *pax6* CRE regions in mouse and zebrafish confirmed loss and retention of specific cis-regulatory elements, that strongly correlates with the diverged expression of co-orthologs (Kleinjan et al., 2008). A confirmation of DDC (duplication-degeneration-complementation) theory (Force et al., 1999; Lynch and Force, 2000a), was also obtained by comparison of conserved functional regions in noncoding DNA of the HoxA clusters among several species of fish and mammals. The analysis has shown that distinct putative regulatory sequences are retained in duplicated Hox clusters in the fish lineages (Santini et al., 2003).

Neofunctionalization.

The idea that gene duplication is a potent source of evolutionary innovations was expressed as early as the 1930s (Haldane, 1933). However, it was not until 1970 that Ohno provided evidence for a link between gene duplication and the acquisition of a new function. In Ohno's theory, the emergence of "extra" gene copies opens up freedom for evolutionary

experiments. Mutations that occur in one of the two duplicates, will not be screened against by selection because the second gene copy will ensure the ancestral function. Therefore, one of the copies is likely to remain more or less unchanged and the other will accumulate random mutations. High chances exist that accumulation of random mutations will lead to non-functionality of this gene copy. However, the possibility to evolve new properties will increase in case of advantageous mutations. In the classical version of neofunctionalization, one of the duplicates loses its original function. Since this event is considered to be extremely rare, an expanded hypothesis of neofunctionalization (NF) was proposed: 1) after duplication one gene copy keeps all the old functions and acquires new features (NF-I); 2) a new gene loses all the old functions (NF-II); 3) a new gene retains some of the old features (NF-III) (He, Zhang, 2005). In recent years, many examples of gene duplication and new function acquisition were described in the literature, for example in plants (Flagel and Wendel, 2009), in yeast (van Hoof, 2005), in *Drosophila* (Assis and Bachtrog, 2013; O'Neill and Clark, 2013; Wang et al., 2004; Zhou et al., 2008), in worms (Lercher et al., 2003; Rubin et al., 2000), in fish (Deng et al., 2010; Robinson-Rechavi and Laudet, 2001) and in apes (Rosso et al., 2008).

1.4.2. Changes in the coding sequence.

Making a duplicate is however not required for giving a gene a new function. A powerful way is to change/mutate the coding sequence, as observed for some Hox genes. Such changes can make the protein function diverge, creating new interactions, ultimately causing significant morphological variation. A classical example is that of the Ultrabithorax (Ubx) protein of *Drosophila*. The transcriptional repression domain found in the carboxy-terminal region of the Ubx protein is highly conserved in insects but is absent in other arthropods and onychophorans. The evolution of this domain can explain the morphology of posterior thoracic and anterior abdominal segments of modern insects (Galant and Carroll, 2002). Another example is for the Drosophila *ftz* gene. The conserved role of *ftz* in CNS development (*eve* activation) is mediated by the homeodomain or other conserved sequences, while the segmentation function of *ftz* in *Drosophila* is mainly mediated by the divergent regions of the protein outside of the homeodomain, probably through protein-protein interactions (Alonso et al., 2001).

1.4.3. Evolution of CRE and morphological novelties.

Another way to generate novelties without duplicating genes is to modify cis-regulatory elements (CRE) of the genes. Experimental work on fly wing pigmentation has shown that homeobox genes with conserved functions acquired a new role in the *Drosophila biarmipes* wing development via changes in the regulatory regions (Gompel et al., 2005). The homeobox transcription factor *Engrailed*, which plays a well conserved role in arthropod segmentation, represses the expression of the gene *yellow* in the posterior wing compartment, thereby preventing the formation of the pigmentation spot. Thus, *Engrailed* was co-opted in the establishment of a new regulatory circuit and acquired a new function in pigmentation regulation through specific binding sites in the regulatory region of the *yellow* gene in *D. biarmipes*. Similarly, the Hox gene *Abd-B* directly controls the formation of abdominal pigmentation in *D.melanogaster* males, by directly binding CRE of *yellow* and activating its expression (Prud'homme et al., 2007). Also, the recruitment of the *hedgehog* regulatory circuit in butterfly eyespot evolution is a good example of how a whole preexisting circuit could govern evolutionary novelties (Keys et al., 1999).

A theory of morphological changes that do not involve gene duplication, but are based on cis-regulatory element evolution has been proposed by Caroll S.B. and co-authors (Carroll, 2000; Carroll, 2005). This theory conquered a certain authority during the last decade. However, it is not clear yet whether new protein functions acquired in such a way can yield major morphological changes. The more significant alteration of body plan explained by genetic changes seems due to modifications of coding sequences (Galant and Carroll, 2002; Lynch and Wagner, 2008). Gene duplication creates a potent source for changes in both the CRE and/or coding sequence of the genes, but its level of impact on morphological diversity is not clear. It may have played a role when the duplications were numerous, for example in the vertebrate lineage after the whole genome duplication events (Panopoulou and Poustka, 2005; Putnam et al., 2008; Peer et al., 2009, 2010). Lineage-specific small-scale duplications (SSD) (e.g. tandem duplications, retroposition, and slippage of DNA polymerase during replication) give a scope of available genetic material for genotypic and morphological evolution (Andersson et al., 2015; Dittmar and Liberles, 2010; Ohno S., 1970).

1.5. Candidate gene approach and RNA interference

Forward genetics is the most elegant approach for determining the genetic basis of phenotypic differences, because it is not biased for any candidate. It is however almost reserved to advanced model systems (for animals, *Drosophila melanogaster*, *Caenorhabditis elegans, Mus musculus,* etc.) and their closely related species. Although *Oikopleura dioica* is a species with significant potentialities for genetic analysis (short life cycle, high fecundity), no mutant screen has been implemented in this species yet. We therefore operated a candidate gene approach and attempted to elucidate their function through reverse genetics, targeting gene knockdowns by RNA interference (RNAi).

RNA interference is a method for gene knockdown that allows to reveal the biological function of a targeted gene and, in favorable cases, to study its influence on phenotype. First experiments with antisense RNA injections into cells were performed as early as the 1980s (Izant and Weintraub, 1984; Rebagliati and Melton, 1987; Weintraub et al., 1985). However, it is only in 1998 that Andrew Fire and Craig Mello (Fire et al., 1998) found that double-stranded RNA is more effective at producing interference than are single-stranded antisense RNAs. The RNAi mechanism allows specific binding of double-stranded RNA (dsRNA) to the mRNA of targeted gene. First, long dsRNA are cut into short fragments of 19-21 bp by the enzyme Dicer (Bernstein et al., 2001; Elbashir et al., 2001; Park et al., 2011). These short interfering RNAs (siRNA) form a complex with certain cellular proteins. dsRNA gets unwinded, the sense strand is degraded, and the antisense strand stays bound to the RNA-induced silencing complex (RISC). Then, short single-stranded RNA of RISC binds to their complementary sites on the mRNA of the targeted gene. This is the signal for cutting the mRNA by the enzyme complex. Destroyed mRNA is no longer able to provide valuable protein synthesis. Thus, the synthesis of protein is disturbed, so that phenotypic changes can be induced (Hutvagner and Zamore, 2002; Novina and Sharp, 2004). In plants and nematodes, dsRNA have the ability of self-amplification and only a few molecules of initial dsRNA are sufficient to give an effect (Baulcombe, 2004; Fire et al., 1998). In mammals, a dosage-dependent effect of RNAi was shown (Novina and Sharp, 2004; Yang et al., 2001; Zimmermann et al., 2006).

It is important to emphasize that most studies of non-model organisms aiming to find out the determinants of macroevolutionary changes also use the candidate gene approach, with the candidate genes chosen based on studies in model organisms. It is a *priori* difficult to guess how many genetic changes can have occurred before a new and complex structure such as the appendicularian house appeared. We postulated that the molecular mechanisms that primarily shape the oikoplastic epithelium and are in ultimately charge of the house production, must include a number of conserved transcription factors that were secondarily recruited for this new programme. We more or less arbitrarily focused our candidate transcription factor approach on the homeobox genes, which play multiple essential roles during animal development.

1.6. Homeobox genes

The most popular homeobox genes, because having a similar and major function in distantly related taxa, are the *Hox* genes. In contrast to most homeobox gene families, the *Hox* genes are usually clustered in the genome and they play concerted role in patterning the body along the anterior-posterior (A-P) axis (Duboule, 1994; Duboule and Dolle, 1989; Graham et al., 1989; Lewis, 1978). The comparison of Hox cluster organization among various phyla suggests that the common ancestor of bilaterians had an intact Hox cluster, either actively maintained for functional reasons or passively maintained due to low rates of genome rearrangements. However, a number of bilaterian species, beginning with the fruitfly, have a split Hox cluster or sometimes an even more degenerated one, for example the nematode *C. elegans* (Fig. 5) (Aboobaker and Blaxter, 2003; Berná and Alvarez-Valin, 2014; Hejnol and Martindale, 2009; Lemons and McGinnis, 2006; McGinnis et al., 1984). The Hox clusters have been described in relatively good details in the ascidian *Ciona intestinalis* and in *O. dioica*. In *C. intestinalis*, the cluster has been strongly rearranged though some linkages between Hox genes remain (Ikuta et al., 2004), while the Hox cluster in *O. dioica* has been totally disintegrated (Seo et al., 2004). Both species have nine Hox genes, but their Hox gene complements are different. In the absence of experimental studies of Hox gene function in tunicates, interpreting such alterations of their cluster was not easy. On one hand, the rapid genome evolution in this subphylum may have been a cause if the cluster organization was no longer required. On the other hand, the main reason for cluster

integrity in the mouse seems to reside in the coordination of Hox gene expression timing during development (temporal collinearity). This temporal coordination may not be needed in rapidly developing embryos such as those of tunicates and of other species including the fruitfly. It is also important to mention that several if not all Hox genes of tunicates may have evolved to novel functions, based on the observation of their expression patterns or of the discrete phenotypes observed after gene knockout (Seo et al., 2004; Ikuta et al., 2004; Sasakura et al., 2012). This example indicates that even the most conserved developmental genes may have evolved to a rather derived fate in tunicates.

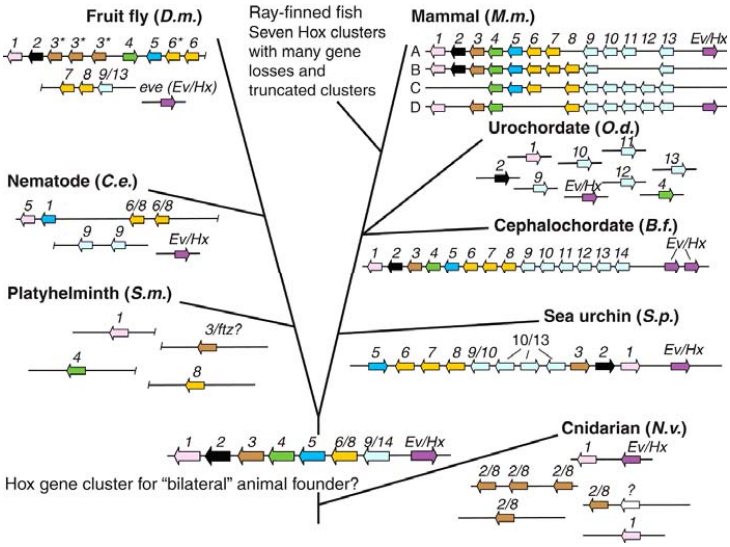

Figure 5. (from Lemons & McGinnis, 2006) Schematic representation of Hox gene clustering in metazoan evolution.

In ancestral chordates and probably before them, the function of Hox genes in AP axis patterning does not affect the most anterior structures. In the central nervous system, the most anteriorly expressed Hox genes (*Hox1* or *Hox2*) are not expressed ahead of the hindbrain. Other homeobox genes play essential roles more anteriorly, and these include the *Otx* genes. *Otx* genes have been discovered in cnidarians (Mazza et al., 2007; Smith et al., 1999) and in all bilaterian animals studied thus far (e.g. Boncinelli and Morgan 2001; Lanjuin et al. 2003; Martínez-Morales

et al. 2003; Saló et al. 2002; Urbach 2007; Wada, Sudou, and Saiga 2004; Yamamoto et al. 2007). The widely conserved function of the *otx* genes in bilaterians (both protostomes and deuterostomes) is in patterning the anterior CNS. At the molecular level, the borders between brain compartments are defined by TF expression: *otx1/otx2* are expressed in the forebrain; *pax258, en, fgf8* are expressed in the midbrain/MHB; *hox1* is expressed in the hindbrain, and other *hox* genes expressed more posteriorly. Comparative study of *otx* gene expression patterns together with other anterior CNS markers as *gbx, pax6, pax258, en,* reveals similarities in their expression patterns and functions in different phyla (Fig.6) (Cañestro et al., 2005; Castro et al., 2006; Harada et al., 2000; Satoh, 2003).

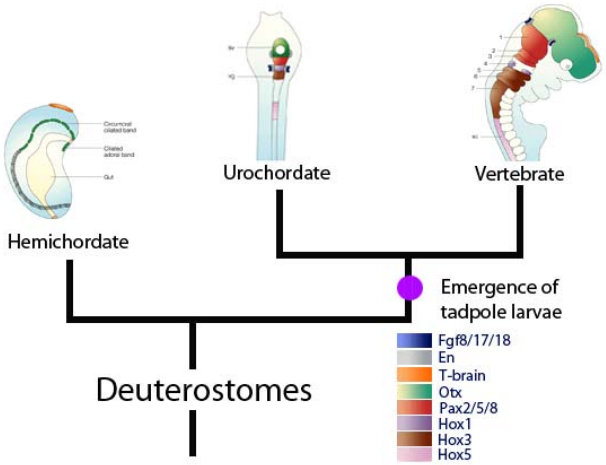

Figure 6. (modified from (Satoh, 2003)) Similarities of homeobox gene expression patterns and functions in different phyla.

Examination of homeobox gene expression patterns in the *O. dioica* CNS has shown its regionalization along the A-P axis in a manner comparable to that of other animals (Cañestro et al., 2005). The early expression of *otxB* and *hox1* was detected during cleavages of the *O.dioica* embryo. It is likely that the A-P axis appears during early gastrulation. Other CNS markers (including *otx* genes) are expressed sequentially, and based on their temporal and spatial expression, it was suggested that the *O.dioica* cerebral ganglion is homologous to the vertebrate forebrain. The posterior overlap of *otx* and *pax6* expression, the loss of *gbx*, and the variability or

absence of the expression of vertebrate midbrain markers (such as *engrailed* and *pax2/5/8*), suggest that the midbrain was modified or lost independently in the urochordate lineage. The absence of *engrailed* and *pax2/5/8* expression immediately posterior to the *otx* expression domain argues for the absence of an MHB (mid-hindbrain boundary) organizer homolog in the larvacean CNS. The caudal ganglion and the posterior nerve cord in *Oikopleura* derive from a posterior CNS region expressing *hox1*. These structures were considered to be homologous to at least part of the vertebrate hindbrain (fig.7 from (Cañestro et al., 2005)).

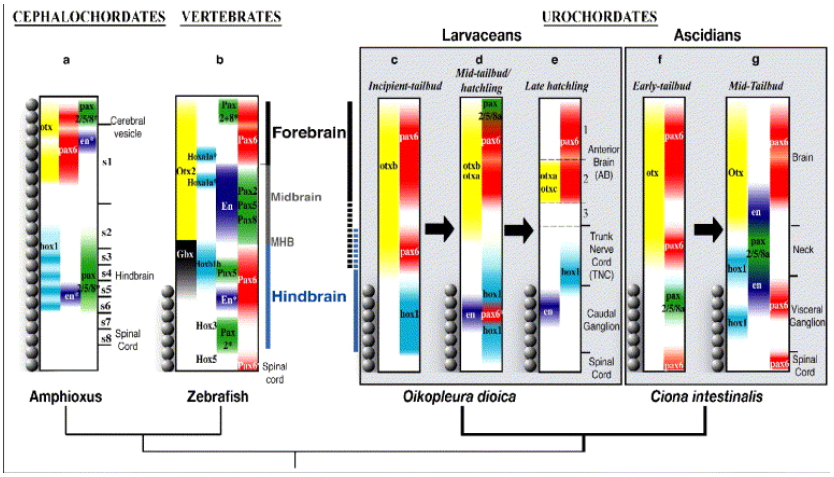

Figure 7. Comparative expression of homeobox genes in the CNS of different chordates.

Thus, it was shown that some of the *Oikopleura* homeobox genes play a conserved role in CNS development. However, some of these genes are also expressed outside the CNS, such as *otx* duplicates, and their epithelial expression was mentioned previously but not described in details. In our work, we concentrated our attention on the epithelial expression of *otx* genes and tried to study their function with RNAi. Other candidate genes for our functional studies were the *prop* genes. The role of *prop* genes was studied mainly in mammals, even though *prop* orthologs were also found in the genomes of several invertebrates (*Drosophila melanogaster*, *Tribolium castaneum*, *Saccoglossus kowalevskii* and *Caenorhabditis elegans*). There is really a lack of information about *prop* evolution and function in invertebrates. In vertebrates, the single *prop* gene plays a major role in pituitary gland development. *Prop1* is important for proliferation,

specialization, and differentiation of highly specialized hormone-producing cells (Davis et al., 2011; Scully and Rosenfeld, 2002). Mutations in *Prop1* are a common genetic cause of combined pituitary hormone deficiency in human, mice and fish (Kelberman et al., 2009; Angotzi et al., 2011).

Homeobox genes are present in all known metazoan genomes and often have a conserved function during evolution. However, apparently minor modifications of such powerful developmental genes, in their coding sequence or regulatory elements, can result in significant evolutionary changes.

1.7. Oikopleura dioica as a model system

1.7.1. Phylogenetic positions of Appendicularians.

Traditionally, Tunicates are divided into three classes based on morphology, life history and mode of reproduction: **1) Ascidians** are by far the most diversified group, with more than 2500 species. Ascidians have tadpole-like larvae with a typical chordate body plan, and after a particularly traumatic metamorphosis become sessile adults, that can be either solitary or colonial; **2) Thaliaceans** comprise pyrosomes (free-floating conical shaped colonial tunicates), salps (solitary generation reproducing asexually and colonial generation reproducing sexually) and doliolids (solitary and colonial). Thaliacean have pelagic adults and are particularly diverse in their anatomy, locomotion, buoyancy, sensory system and embryonic development. However they share a metagenic life cycle involving alteration of generation (Bone, 1998) and **3) Larvaceans** (appendicularians) have pelagic adults grossly displaying the larval body plan, resulting from a discrete metamorphosis that includes a shift of tail orientation with respect to the trunk (Bone, 1998). During the last two decades, genome sequencing and genome annotation had a strong impact on our knowledge of the phylogeny in various taxons. In tunicates, phylogenetic analysis supported evolutionary relationships between ascidians, thaliaceans, and larvaceans that are more complex than traditionally proposed, although the results are debated (Satoh, 2009; Stach and Turbeville, 2002; Tsagkogeorga et al., 2009). A predominant view is that the common ancestor of chordates was a free-living animal with a notochord, dorsal neural tube, myotomes, and a post-anal tail (Brown et al., 2008; Cameron et al., 2000; Satoh, 2008). Tunicates are

confirmed by phylogenetic analysis to form a monophyletic subphylum, in which Appendicularians diverged first, before other tunicates finally split into two classes of Stolidobranch Ascidians and Phlebobranch Ascidians/Thaliaceans, respectively. (Fig.8) (Satoh, 2009; Swalla et al., 2000).

At a higher taxonomic level, it was found and confirmed that cephalochordates diverged first from other chordates, that later on split into tunicates and vertebrates, respectively (Fig.8) (Bertrand and Escriva, 2011; Bourlat et al., 2006; Delsuc et al., 2006). Such a result initially came as a surprise, especially for the community of morphologists, since cephalochordates are for multiple criteria more similar to vertebrates than are tunicates. One way to interpret the new phylogeny of chordates is that their common ancestor may have been more complex that earlier thought, and that tunicates underwent a drastic simplification after they split from the vertebrate lineage.

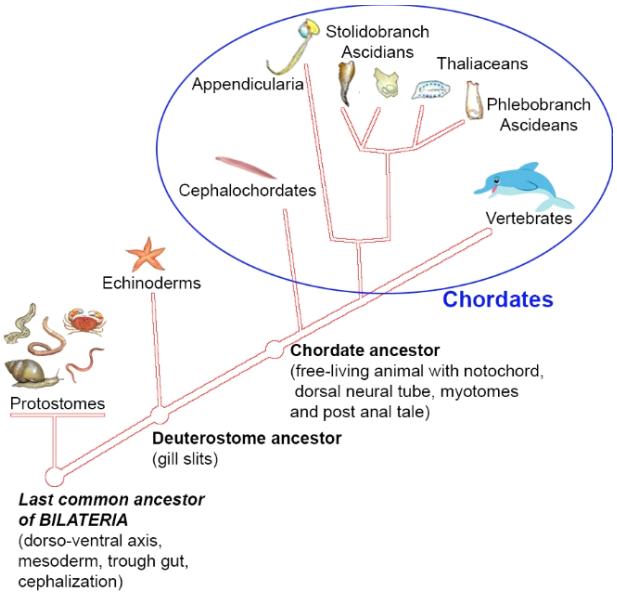

Figure 8. Phylogenetic position of appendicularians. Note that the branches leading to tunicates are longer than others, especially for appendicularians, indicating a more rapid evolution of coding sequences in this group.

1.7.2. Life cycle

O. dioica is currently the only larvacean species that can be routinely cultured in the laboratory. After continuous improvements of the culture procedures, it can indeed be bred in captivity for hundreds of consecutive generations (Bouquet et al., 2009; Spada et al., 2001). The life cycle of *O. dioica* lasts for only 6 days at 13-14^{0}C. *O. dioica* is the only appendicularian species that has separate sexes. Males and females become mature at Day 6. Individuals that are almost ready to spawn are manually transferred into 6-liter beakers of filtered sea water where food (algae) is added. After a few hours they spawn in group and provide the next generation. Alternatively, the gametes (sperm and oocytes) can be collected and mixed *in vitro*, allowing precise timing and observation of the development. The embryogenesis is rapid. The first cleavage occurs after 30 minutes post-fertilization, with following cleavages coming every 5 minutes at room temperature. Gastrulation begins with the 5th cleavage (32 cell stage), with internalization of vegetal cells by ingression and/or epiboly. Neurulation overlaps with gastrulation and starts at the 64 cell stage. Thus, gastrulation and neurulation take place early when the embryo has only few cells. At 3 hours post-fertilization (hpf), the tailbud begins to form. Hatching occurs at 4 hpf, and larvae stretch their tail. Organogenesis proceeds during the first six hours of the larval life. After 10 hpf, while the first pre-house is secreted, the metamorphosis takes place. The tail changes orientation from a straight to an almost perpendicular position with respect to the trunk (tailshift). The first house is produced by the oikoplastic epithelium soon after the tailshift metamorphosis is completed. At this stage, the number of somatic cells is already fixed for most organs and juvenile larvae essentially grow through an increment of cell size. The first signs of sexual maturation (growing gonads) are visible at Day 5 at 14^{0}C. When the males are mature at Day 6, spermatozoa exit the gonads through a orifice located on the dorsal side of the testis. Females release the eggs into the water after rupture of the body wall. Animals die soon after spawning (after (Bone, 1998; Bouquet et al., 2009; Nishida, 2008; Stach et al., 2008)).

1.7.3. House secretion and the oikoplastic epithelium.

The house is produced by the secretory activity of a monolayer oikoplastic epithelium that covers the trunk. The epithelial cells of the adult trunk are variable in size, shape and ploidy (Ganot and Thompson, 2002). At metamorphosis, when the first house is secreted, the epithelium consists of several characteristic cellular fields (Fig. 9). The Field of Fol is situated in the first half of the epithelium, anteriorly and consists of five different cells type: Anterior cells, Giant Fol cells, Nasse cells, posterior cells, and small cells of the Fiber rosette that are situated on the trunk midline between rows of Nasse cells. The Field of Eisen is represented by two lateral fields situated symmetrically around the midline. Each of them consists of three types of oikoplasts: seven Giant Eisen cells, a chain of pearl cells situated at the anterior border of the Eisen region, and border cells that surround the Eisen field. We can also identify the Martini fields, situated posterior to Posterior Fol cells, the Anterior rosette situated between both Martini fields, and the Posterior rosette situated on the midline at the posterior end of the trunk. Certain house structures appear to be spatially related to defined groups of cells in the epithelium. Thus, it is well known that the Fol and Eisen regions are related to the production of food concentrating filter and inlet filter, respectively (Bone, 1998; Sagane et al., 2011). The Posterior rosette is said to be related to the escape slot, through which the animal abandons the house. The Anterior rosette is in charge of "keel" or "sail" production. The circle of epithelial cells surrounding the mouth is related to the exit spout of the house. Finally, the crescent-shaped region of the posterior ventral oikoplastic epithelium is related to the tail chamber of the inflated house (Flood, 2005; Flood et al., 1998).

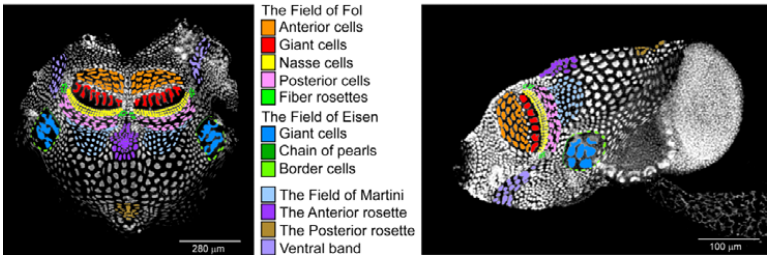

Figure 9. from (Hosp et al., 2012). The epithelial fields of O.dioica adults. Left - dorsal view: Anterior - up, posterior – down. Right - lateral view: anterior to left, posterior to the right.

1.7.4. The house components.

The house comprises cellulose (Sagane et al., 2010), mucopolysaccharides, and proteins called oikosins which are secreted from specific cellular territories in the oikoplastic epithelium (Ganot and Thompson, 2002; Hosp et al., 2012; Spada et al., 2001). The ability to synthesize cellulose is a very special feature of tunicates, setting them apart from other metazoans. Ascidian and thaliacean tunics consists of tunicin - cellulose-like substances (Hirose et al. 1999; Hirose 2009). Larvaceans produce a house instead of a tunic, that also contains cellulose (Ganot and Thompson, 2002; Kimura et al., 2001). It is unlikely that cellulose biosynthesis was lost in all animals except tunicates. Therefore either tunicates independently evolved it, or their common ancestor acquired it by horizontal transfer from another organism. A putative cellulose synthase-like gene was described for *Ciona intestinalis* and *Ciona savignyi* (Matthysse et al., 2004), and *cellulose I-beta* was found in the house of *Oikopleura rufescens,* supporting some degree of homology between the house and the tunic (Kimura et al., 2001). The presence of two *cellulose synthase A (CesA)* genes in *O. dioica* that have distinct roles, one pre-metamorphic and the other post-metamorphic is an illustration of a lineage-specific duplication of the cellulose synthase gene with division of functions over time (Sagane et al., 2010). All the data support a single lateral cellulose synthase gene transfer, from a prokaryote to tunicates.

Mucopolysaccharides (or glycosaminoglycans) are long unbranched polymeric carbohydrate-protein complexes with a primary content of carbohydrates (70-80%). The repeated unit consists of an amino sugar (N-acetylglucosamine or N-acetylgalactosamine) along with a uronic sugar (glucuronic acid or iduronic acid) or galactose. With the ability to bind and retain water, acidic mucopolysaccharides serve as a natural lubricant. They determine the elasticity of structures and may serve as a shock absorber. Mucopolysaccharides also possess antibacterial properties. Oikosins are novel proteins produced by the trunk epithelium that make a complex with glycosaminoglycans in the house wall (Spada et al., 2001). These genes and their encoded proteins have no identifiable orthologs in metazoans other than larvaceans. The oikosin genes are specifically expressed in the epithelial territories in a patchwork manner. Oikosins are the proteins that are secreted from the surface of the trunk in a non-classical manner by a combination of four hypothetical pathways: secretion by endolysosomes, exosomes, membrane blebbing or transporters (Hosp et al., 2012). They are then incorporated into the house wall in a way that correlates with their expression patterns, as revealed by immunolocalization studies

(Sagane et al., 2011). Thus, for example, Oikosin 1 (with expression in the seven giant Fol cells), Oikosin 2 (with expression in the Nasse Cells and Posterior Fol cells) and Oikosin 3 (with expression in the Anterior Fol) are incorporated into the food concentrated filter. Around 80 oikosins, highly glycosylated proteins (when counting all known isoforms) were recently described. Around one third of them appeared through serial gene duplication, with up to eight paralogs in one family (Hosp et al., 2012).

1.7.5. Early development of epithelium.

The development of tunicates is often said to be mosaic, although the development of several tissues depends on major cell-to-cell interactions (Conklin 1905; Berrill 1932; Crowther and Whittaker 1983; Gilbert 2000; Satoh 2013). Embryos with mosaic development cannot substitute lost blastomeres by genetic reprogramming of neighbors cells. Already in the fertilized egg, different parts of egg cytoplasm contain maternal mRNAs that specify the future determination of the blastomeres, depending on the specific determinants inherited (Abercrombie and Brachet, 2013; Gilbert 2000). The development of *O. dioica* embryos is extremely rapid in cleavage and morphogenetic events, and fate restriction occurs very early even compared to ascidians. Already at the 8 cell stage, we can recognize the four animal blastomeres (a4.2, *a4.2*, b4.2, *b4.2*) that will give rise to all epithelial cells. After the fifth round of cell divisions (32 cells, blastula stage), gastrulation begins and most adult tissue types can be assigned to specific fate-restricted cells (Stach et al., 2008). Neurulation is overlapping in time with gastrulation, and prospective nerve cells are aligning along the dorsal midline. At the 128 cell stage, after the seventh division, all nerve cells are covered with epidermal cells, and the prospective ectodermal cells close over the blastopore. During tailbud formation, the epidermal cells migrate from the dorsal part of embryo towards the prospective anterior part of the hatched animal. It is the main migration of epidermal cells during development. At the moment of hatching, all epidermal cells have taken their places in the larval epithelium. The molecular mechanisms that underly the early epidermal development are still unknown.

Aims of study

The "house" of Larvaceans is seen as a spectacular innovation, because neither its components nor the tissue in charge of its production seem to have equivalents (or homologs) in other phylogenetic groups. We would like to know which genetic programmes, built during evolution, govern the formation of the oikoplastic epithelium during embryo and larval development. Of particular interest are the level of novelty of the molecular players and their interactions within these programmes. Various methods are in principle available for such a purpose, but only part of them could be adapted to *Oikopleura*. In the absence of unbiased genetic screen, we had to begin with a candidate gene approach. A substantial part of our work consisted in the validation of a method leading to gene expression knockdown. Our main aims were as follows:

- **To elect candidate genes encoding transcription factors** that may play upstream of gene regulatory networks. We collected them mainly in the superclass of homeobox genes. Our contribution to the inventory of these genes was the production and analysis of their normal expression patterns.
- **To describe the formation of the oikoplastic epithelium.** During a few hours following hatching, well defined territories of cells having specific shape and size appear and begin the synthesis of oikosins, the proteins of the house. The criterion for selecting candidate genes is an expression in this epithelium, possibly transient, during the first four hours of the larval life.
- **To validate a method of RNA interference.** Work from Nishida's and our lab showed that the expression level of specific genes, including the T-box transcription factor Brachyury can be markedly reduced by injecting dsRNA into either the ovary or the fertilized egg. We wished to examine whether this type of RNAi technique could work with other genes. We focused on two genes that play major role in the synthesis of neurotransmitters.
- **To assess the role of homeobox genes in the oikoplastic epithelium formation.** We finally used RNAi to examine whether some of the candidate homeobox genes selected based on their expression were involved in the oikoplastic epithelium development. The knockdown of two duplicates of the *prop* gene indeed led to very local morphological alterations in this tissue, as well as to an inhibition of expression for one oikosin gene.

Articles relevant to the thesis work

Article I.

Edvardsen, R. B., Seo, H.-C., Jensen, M. F., Mialon, A., Mikhaleva, J., Bjordal, M., Cartry, J., Reinhardt, R., Weissenbach, J., Wincker, P., et al. (2005). **Remodelling of the homeobox gene complement in the tunicate Oikopleura dioica**. Curr. Biol. 15, R12–13.

Article II.

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J.-M., Da Silva, C., Brinkmann, H., Mikhaleva, Y., Olsen, L. C., Jubin, C., Cañestro, C., et al. (2010). **Plasticity of Animal Genome Architecture Unmasked by Rapid Evolution of a Pelagic Tunicate.** Science. 330, 1381–1385.

Article III.

Mikhaleva, Y., Kreneisz, O., Olsen, L. C., Glover, J. C. and Chourrout, D. (2015). **Modification of the larval swimming behavior in Oikopleura dioica, a chordate with a miniaturized central nervous system by dsRNA injection into fertilized eggs.** J. Exp. Zool. B. Mol. Dev. Evol. 324, 114–127.

Article IV.

Mikhaleva Y., Skinnes R., Sumic S., Thompson E. M. and Chourrout D.
Development of the house secreting epithelium, a major innovation of tunicate larvaceans, involves multiple homeodomain transcription factors. Submitted.

Summary of results

4.1. Homeobox gene loss gain in Oikopleura dioica (article I)

The classification of the superclass of homeobox genes is mostly based on early surveys of vertebrate and fruitfly genomes. Based on phylogenetic analyses of the homeodomain and other domains, most homeobox genes are allocated to 10 distinct classes (ANTP, PRD, POU, LIM, TALE, SIX, CUT, ZFH, HNF1, PROX) that contain approximately 89 gene groups. Most of these gene groups are present in invertebrates and vertebrates and therefore appeared before the earliest radiations of metazoans. Examining two tunicate genomes, from *Ciona intestinalis* and *Oikopleura dioica* failed to detect a substantial number of homeobox gene groups, a rather important deviation from observations of other invertebrate genomes. Approximately half of these gene groups are absent in both species. After the split of both tunicate lineages, other homeobox genes were lost, mainly in the *Oikopleura* lineage.

The total number of homeobox genes counted in tunicate genomes is however approaching that found in most other invertebrate taxons. This is because an unusually high proportion of the remaining homeobox genes have been duplicated with their duplicates retained. These amplifications of gene groups happened more frequently in the *Oikopleura* than in the *Ciona* lineages. Some genes were duplicated multiple times, leading to three genes (*otx*) or even four genes (*pax37*). Later genome surveys revealed that such a turnover of gene complement, with high rates of losses and duplications, also occurred for other families of transcription factors and other developmental genes (Denoeud et al., 2010). Recent unpublished work (Sumic and Chourrout) also showed that the homeobox gene complement is very similar among distinct *Oikopleura* species, suggesting that the losses and gain are relatively ancient.

4.2. Expression patterns of developmental genes, with emphasis on epithelium development (articles II and IV).

The study of expression patterns permits to formulate hypotheses on current gene functions. Their comparisons with expression patterns of their orthologs in other species gives insight on how this function has evolved. We studied the expression of multiple homeobox genes and Fox genes in *O. dioica*, with whole mount *in situ* hybridizations exclusively. We have no information on the protein expression.

Expression of duplicated homeobox genes in Oikopleura dioica.

The invertebrate and vertebrate orthologs of most genes that were duplicated in the *O.dioica* lineage are expressed in the CNS and its derivatives, probably indicating the ancestral function. In contrast, most duplicated homeobox genes of *Oikopleura* are expressed in the epithelial cells of the trunk, during the early larval life. This epithelial expression is for some genes added to an expression in the CNS (e.g. *otx* genes, *onecut* genes). For other genes, only one copy is expressed in the CNS (e.g. *meis* genes, *prop* genes). Finally, there are cases where both (or all) copies are expressed in the trunk epithelium, but not in the CNS (*pax37* genes). An epithelial expression of some of these genes (e.g. *otx*, *meis*, *pax37*) is also observed in ascidian embryos (Imai et al., 2004; Mazet et al., 2003; Wada et al., 2003; Wada et al., 2004). It may or may not be homologous to the expression in *Oikopleura*. However, their expression in *Oikopleura* larvae is highly regionalized and seems restricted to territories of the oikoplastic epithelium. The precise coincidence between the expression domains and these cellular territories is not clarified yet. In total, most copies of duplicated genes are expressed in the epithelium of the larval trunk. The central part of the larval trunk, where the Fol region is formed, concentrates expression signals from 19 of 20 duplicated gene groups. In most cases, both duplicates are expressed there, and in the same region. This suggests that the epithelial expression preceded the duplication event. The expression domains of copies of the same gene are in most cases partially different, suggesting further evolution through subfunctionalization, although neofunctionalization cannot be excluded.

Expression of non-duplicated homeobox genes in Oikopleura dioica.

While studying the expression patterns of non-duplicated homeobox genes, we found other interesting cases. These include a few gene groups that also have several members in *Ciona intestinalis* and therefore may result from duplications anterior to the split between both tunicate lineages. Examples are *nk2A/nk2B, pax258A/pax258B,* and *dlx1/dlx2/dlx3*. Both *nk2* genes show conserved expression patterns in the endoderm derivatives, but no expression in the CNS. A similar situation is observed for *pax258* genes, which play an important role in the regionalization of the vertebrate brain. In ascidians, *pax258* gene is only expressed in the CNS, whereas in *Oikopleura,* it is only expressed in the endodermal derivatives. This suggests that some ancestral gene functions have been partitioned differently among the duplicates in the two tunicate lineages. Both tunicates have three *dlx* genes, that may result from two duplications preceding the ascidian/larvacean split. Dlx genes in *Oikopleura* are expressed in anterior (mouth) and posterior epithelial cells, but not in the CNS. An expression is also found in the anterior epithelium of the amphioxus and of *Ciona* (Caracciolo et al., 2000; Holland et al., 1996), in which it probably contributes to the formation of the adhesive organ. It is difficult to say whether or not the anterior ectoderm expression in *Oikopleura* is homologous to those of amphioxus and *Ciona*, since *Oikopleura* does not have an adhesive organ. Some gene groups from non-amplified groups with a single member (such as *emx, lhx36, mbx*) are also expressed in the trunk epithelium. However, this expression takes place in the most anterior or posterior regions, and not in the central epithelial zone, that ultimately builds the food concentration filter. Finally, the gene *hox2* is exclusively expressed in epithelial cells of the ventral trunk. All other Hox genes are expressed in the tail, with *hox1* having there the most anterior expression. This suggests that the function of *hox2* has been changed, probably towards patterning the oikoplastic epithelium.

Expression of Fox genes.

Another large family of transcription factor genes – the Fox genes - was also studied with whole mount *in situ* hybridization. In invertebrates and vertebrates, they play important roles in a variety of developmental processes, including germ layer specification, gastrulation, cell fate determination, and morphogenesis. They have a conserved DNA-binding domain with winged helix structure. The Fox gene family has been divided into 19 subfamilies (from A to S). *Oikopleura* seems to possess not more than 11 subfamilies of Fox genes, from which we were

able to clone 20 out of 22 annotated Fox genes. *Ciona* would have 17 subclasses. These counts largely depend upon the robustness of phylogenetic tree topologies. Only 5 classes (A, C, H, I, M) of Oikopleura Fox genes were duplicated and three of those (A, H, I) are also duplicated in *Ciona*. We observed a zygotic expression for 11 genes only, and seven show *in situ* hybridization signals in the larval epithelium. Epithelial expression in the duplicated Fox gene groups was observed in the epithelium as often as in other tissues. Two genes of non-duplicated gene groups (*foxB*, *foxF*) also showed an epithelium-specific expression (article IV). Overall, many of the transcription factor genes that we studied, either duplicated or not duplicated, are expressed in the trunk epithelium during the early larval stage, and they consequently qualified for our candidate gene approach.

Oikoplastic epithelium development.

In a 4-hour old tadpole, that just hatched, a layer of epithelial cells covers the whole trunk. At this stage, all nuclei of epithelial cells have similar size and shape, and we assume that their cells are not differentiated. They actively divide, gradually giving shape to distinct cellular fields during the following 4-5 hours. Some fields become recognizable before the others. At 5 hpf, all Eisen cells are clearly present, though not at their final location yet. At 6 hpf the anterior Fol cells have become organized within an oval shape field. Giant Fol cells are easily recognizable in two symmetrical rows on each side of the dorsal midline. Nasse cells nuclei have become smaller, and gathered in three distinct rows on each side. The Eisen region is still not forming a circle, but it can be recognized due to specific nuclei shapes. At 7 hpf the epithelium has acquired well defined territories, even though the final number of cells is not reached yet. Eisen cells have now reached their final position. The numbers of cells continue to rapidly increase until 8 hpf, when the dorsal epithelium is resembling well that of the adult. The nuclei are at this stage densely packed. During the following three hours, the inter-nuclear spaces increase, and the cells slowly reach their final number and positions (article IV, Fig.1). No cell migration is observed during the oikoplastic epithelium development, at least between distinct territories. The only migration observed is a collective movement of Eisen cells from posterior to centro-lateral epithelium areas.

After we obtained this global description of the epithelium development, we tried to map the expression domains of homeobox genes on the prospective epithelial cell fields (article IV). The genes are expressed in three main domains: anterior - in epithelial cells around the mouth,

central – in the regions of FoI, Anterior rosette and Martini region (food concentration filter), and posterior – in the Posterior rosette and the posterior dorsal cells. In most if not all cases, the expression signals appear as either spots or stripes on the trunk surface, most often bilaterally symmetrical around the dorsal midline.

4.3. Establishment of gene expression knockdown (article III).

To reveal and further investigate the role of candidate genes in the oikoplastic epithelium development, we needed functional tools, that were until recently completely lacking in *Oikopleura*. We ideally needed a method repressing the expression of specific genes. Our first attempts to knock down gene expression made use of injected morpholino oligonucleotides. This method seems to be often efficient in *Ciona intestinalis* and at least one study also showed their effect in *Oikopleura* (Sagane et al., 2010). We tried it with variable concentrations of either translation-blocking and/or splice-blocking morpholinos on several homeobox genes (*propA, propB, pax37A, pax37B, hox*-genes) as well as on the T-box transcription factor *Brachyury*. Unfortunately, we did not observe convincing phenotypic changes and we therefore moved on to other methods of RNA interference. We began injections of dsRNA into fertilized eggs. We targeted the whole T-box region of Brachyury with 120 ng/µl dsRNA. Two abnormal phenotypes were observed: shortened or totally absent tail, with only a few remaining notochord cells (Fig. 3B,E) or no notochord cell at all (Fig. 3A,D,G), respectively. Whole mount *in situ* RNA–RNA hybridizations demonstrated strong reduction of Brachyury mRNA expression (Fig. 3A,B), which was confirmed with RT-PCR. These experiments were performed independently of those published earlier by Nishida's group in Japan (Omotezako et al., 2013), who injected the dsRNA solution in the ovary.

We then decided to test whether the method can be also efficient for other types of genes. In the absence of mutants, we chose genes whose expression knockdown were likely to result in specific and well predictable phenotypes. We already knew the expression sites of various neurotransmitters (Soviknes et al., 2005; Soviknes et al., 2007) and tried to interfere with their synthesis, in targeting the ChAT (choline acetyltransferase) and the GAD (glutamic acid decarboxylase) genes. Their enzymatic products are critical for the biosynthesis of acetylcholine

(ACh) and γ-Aminobutyric acid (GABA), respectively. ACh is known as a major excitatory neuromuscular transmitter in vertebrates. GABA is a major inhibitory neurotransmitter in the interneurons of vertebrate CNS and has inhibitory activities in motor neurons of invertebrates. After initial attempts, dsRNA for these genes were injected at relatively high concentrations (900ng/µl). The interference with ChAT expression abolished the normal bidirectional, propagating tail movement, permitting only repeated unilateral tail bends. This phenotype is most likely caused by the observed reductions of transcripts for both the targeted ChAT gene and the VAChT gene (Vesicular Acetylcholine Transporter) which are part of the same operon (Fig. 5 and Fig 6, Movies 3 and 4, Article III). Interference with GAD expression resulted in an uncoordinated movement leading to a spiral swimming trajectory (Movie 6, Article III), although made of episodes similar in duration and cycle frequency to those of normal swimming. These findings indicate an abnormal left - right segmental phase shift in muscle contraction within a normal Central Pattern Generator (CPG). Our results demonstrate that the locomotor functions in larvaceans is dependent on at least cholinergic excitatory peripheral neurons and central inhibitory GABAergic neurons. This approach could be extended further for a future genetic dissection of *Oikopleura* neuronal circuits, which are among the simplest in the chordate phylum.

4.4. Functional study of duplicated homeobox genes (article IV).

After the RNAi method was validated for a few and rather diverse genes, we could test it on candidate transcription factor genes expressed in the trunk epithelium of the larva. We chose two gene groups that had been duplicated in the *Oikopleura* lineage - Prop and Otx. The duplication of these groups is relatively old, as shown by the presence of the same duplicates in multiple larvacean genomes (Sumic and Chourrout, unpublished). In the genome, *propA* and *propB* are located next to each other in opposite orientation, with no intervening genes, unlike many gene duplicates of *Oikopleura.* Their separation via chromosome rearrangement may have been prevented by functional constraints, for example if they shared common regulatory elements (see Fig.S3,II). Their tandem arrangement in opposite orientation is also observed in *O.albicans* and in the more distantly-related *O.longicauda*. The genomic organization of otx genes is similar to that of prop genes. There are three paralogs *otxA*, *otxB* and *otxC* in *O. dioica* and other

oikopleurid species, with at least two of them linked in opposite orientation (the three genes are linked in *O. longicauda*) (Fig.S3, I). Using the RNAi knockdown method we were able to reduce the expression of both *prop* genes and to thus alter the oikoplastic epithelium development. The alterations include the morphology of specific areas where the *prop* genes are expressed, as well as a spatially restricted down-regulation of one oikosin gene (*oik41a*).

Normal expression of Prop genes

The *propA* gene becomes expressed at the tailbud stage in only two posterior cells of cerebral ganglion, where this expression continues after hatching. The expression in the epithelium begins at 6 hpf between the left and right groups of Nasse cells. The *propB* gene is expressed soon after hatching in several cells situated around the dorsal midline of the epithelium. Later on, its expressions expands to all cells of the middle part of dorsal epithelium, from the anterior Fol field to the posterior part of the Anterior rosette. The expression domains of *propA* and *propB* partly overlap at 6-8 hpf in dorsal epithelial cells at the level of Nasse cells (Fig.S2, I). They suggest that both duplicates are involved in the development of the Anterior rosette, a particularly well organized epithelial region in oikopleurids. A house protein gene that is well expressed in the Anterior rosette of adult animals is *Oikosin41a* (Hosp et al., 2012). We monitored its co-expression with *propB* during larval stages, with double ISH (Fig.3, E-L). The two genes are expressed in the area where the Anterior rosette will appear betweeen 6 and 10 hpf. Later on, the expression of *propB* vanishes and only *oik41A* is expressed in the Anterior rosette.

Effect of propA knockdown

Most animals injected with dsRNA-*propA* did not show any signal of *propA* expression (article IV, Fig.4, I, A), while *propB* expression seemed unchanged (article IV, Fig. 4, I, D). RT-qPCR showed a two-fold reduction of *propA* transcripts, and only a minor decrease of *propB* transcripts, that is also observed in negative controls with mRNA-EGFP injections (article IV, Fig.S3, IV). After dsRNA-*propA* injections, Day 1 animals display a malformed dorsal epithelium. The number of cells in the middle rows of Anterior Fol and Anterior rosette is significantly decreased. The Anterior rosette is not formed properly, although the size and shape of nuclei look similar to those of control animals (article IV, Fig.4, II).

Effect of propB knockdown

ISH with *propB*-probe after dsRNA-*propB* injections also showed a dramatic reduction of expression in most injected animals (article IV, Fig.4, I, E). RT-qPCR results indicated an almost three-fold reduction of mRNA-*propB* (article IV, Fig. S3, IV). dsRNA-*propB* injected animals displayed malformations of the dorsal midline (between anterior Fol cells), which became narrower than in controls animals, as well as alterations of the Anterior rosette field (article IV, Fig.4, II, DH). After injections of dsRNA-*propB,* the nuclei of the middle rows of Anterior Fol and Anterior rosette had a round shape and were smaller than in control animals. Such changes in nuclear shape and size, as well as an increased number of nuclei in the Anterior rosette field, suggested that cells of the Anterior rosette did not differentiate normally (article IV, Fig.4, II, H).

Effect of double propA and propB knockdown

The simultaneous injection of dsRNA for both *prop* genes induced an almost complete expression knockdown, observed with double fluorescent ISH. RT-qPCR showed a marked reduction of the amount of transcripts for each gene, similar to what was observed after individual gene knockdowns (article IV, Fig.4, II, J). In dsRNA-p*ropA+propB* injected animals (article IV, Fig.4, II, EI), the midline was also narrowed, with a single row of nuclei separating the left and right Anterior Fol fields. Strong malformations and disorganization of nuclei were observed in the region of Anterior rosette formation (article IV, Fig.4, II,I). The midline in this zone was almost absent. Nuclei had fuzzy contours and they did not evolve into a population of nuclei with variable shapes (elongated, triangle and square), as they do in controls. Since *propB* is co-expressed with *oik41a* in the Anterior rosette region, we compared the *oik41a* expression between Day 1 control animals and Day 1 animals resulting from dsRNA-p*ropA*+p*ropB* injections (article IV, Fig.4, II, KLMN). The expression of *oik41a* was undetectable in the region of Anterior rosette in injected animals, whereas the expressions in anterior mouth epithelium and in the posterior epithelium were maintained (article IV, Fig.4, II, LN). Strong *oik41a* expression was observed in the Anterior rosette region of control Day 1 animals (article IV, Fig.4, II, KM).

Otx genes: normal expression and attempts of knock down

The detailed expression of the three *O.dioica* otx genes was studied from before hatching till after metamorphosis (article IV, Fig.S2, III). All three paralogs are expressed in the cerebral ganglion, possibly reflecting the ancestral *otx* gene function. They are also expressed in the developing epithelium of the trunk. The expression of *otxA* does not overlap with those of *otxB* and *otxC*, which are co-expressed in a few lateral epithelial cells at 4-5 hpf. The expression territory of *otxC* is however much broader. The expression level of *otxB* is very high at the earliest stages and is not detectable in the epithelium from 6hpf. RNA interference was attempted for all three *otx* genes. After injections of dsRNA, the embryos were generally abnormal and died long before metamorphosis. The amounts of *otxA/B/C* transcripts measured by RT-qPCR were severely reduced, but the ISH expression signals remained surprisingly intense (article IV, Fig.S4). We are therefore unsure whether RNAi induced the knockdown of *otx* genes, and we could not progress further in their functional study.

Discussion

The beautiful and complex structure of the *Oikopleura* house provides us with an attrative object for studying the emergence of evolutionary novelties. The evolution of the house required functional and structural changes of the trunk epithelium. How did this tissue evolve and which molecular mechanisms were set up in this purpose? Our work brings for the first time partial answers to these questions.

5.1. Technical aspects of our candidate gene approach.

The candidate gene approach consists of proposing and then testing the involvement of given genes in a biological process, that may or may not be conserved among lineages. For an elaborate process like the formation of the oikoplastic epithelium, we assumed that transcription factors must have a contribution. Among transcription factors known to play a role in multiple developmental processes were the homeobox genes. These were inventoried in the *Oikopleura*

genome, through gene annotation and for some of them a characterization of transcripts. We contributed to study their expression patterns at several developmental stages, using the simple but potent technique of whole mount *in situ* hybridization. Our focus was rather arbitrarily targeted on homeobox genes, but we could easily recognize that several of them were good candidates for a role in the oikoplastic epithelium formation. They were indeed intensely expressed in various regions of the trunk epithelium in the hours following hatching. Interestingly, most of these genes had been duplicated in the lineage leading to *Oikopleura*, opening the possibility that gene duplications had either allowed or facilitated their recruitment for this lineage-specific developmental process. We also extended the research of candidates to Fox genes, other important transcription factors, and several of them were also transiently expressed in the trunk epithelium. An important part of this work has consisted in testing and validating a method of RNA interference in *Oikopleura*. The lack of knockdown methods for our model organism has for many years been an obstacle for functional studies. Even though morpholino oligonucleotides successfully knocked down gene expression in ascidians *Ciona sp*. and *Halocynthia roretzi* (Imai et al., 2004; Wada et al., 2004), our own attempts with this technique on *O. dioica Brachyury* and a couple of other transcription factors were fruitless. The efficiency of dsRNA injection for knocking down the *Brachyury* gene was shown by Nishida's group in Japan using injections in the ovary (Omotezako et al., 2013), and independently by us using injections into fertilized eggs. Before using such approach for epithelial expressed transcription factors, we wanted to evaluate its effect on other genes. *Brachyury* and genes involved in the synthesis of neurotransmitters were chosen because we could predict with a high level of confidence the phenotypes that would result from an efficient knockdown. Because the information collected during the functional analysis of these genes is somehow unrelated to our main topic, it is discussed elsewhere (see article III).

5.2. Frequent duplication of transcription factors recruited for the oikoplastic epithelium development.

An intriguing question appearing beyond this potential recruitment of old genes for the oikoplastic epithelium development is why many of them were duplicated with their duplicates

retained. Compared to other animals and chordates (including ascidians), *Oikopleura* lacks many homeobox genes and the remaining gene groups have been frequently duplicated. An inventory of various families of developmental genes has shown that such loss/duplication are by no means restricted to homeobox genes (Denoeud et al., 2010). A recent study of five other oikopleurid genomes also revealed homeobox gene complements almost identical to that of *O. dioica* (Sumic and Chourrout, personal communication). A rather high level of divergence is observed between the same coding sequences among those species, suggesting that they have evolved separately for a significant amount of time, though with little change of the homeobox gene complement. Consequently, the turnover of homeobox genes was not necessarily continuous and could instead have occurred in one or more episodes. A detailed study of the *O. dioica* genome organization had revealed a history of abundant chromosome rearrangements, leading to the almost complete loss of conserved macrosyntenic and microsyntenic associations (Denoeud et al., 2010). These rearrangements may have been more frequent than in other lineages due to the short life cycle and correlative large number of meiotic events per time unit. Interestingly, a survey of DNA repair genes in the *O. dioica* genome has supported the absence of all key players of the non-homologous end-joining pathway (NHEJ) of double-strand break (DSB) repair (Denoeud et al., 2010). Ongoing experiments in our lab indeed show that *O. dioica* unfertilized eggs and early embryos very frequently repair induced DSBs in exploiting sequence microhomologies around them (Deng and Chourrout, personal communication). The mechanism of repair may be the alternative NHEJ named MMEJ, which is more prone to chromosomal translocations than is NHEJ itself.

Highly regulated developmental genes are often larger that housekeeping genes, due to the abundance of regulatory elements in intergenic sequences and introns. Duplications are caused by rearrangements of genome segments that are in average relatively short (Lynch and Conery, 2003; Lynch and Force, 2000b; Lynch and Wagner, 2009), and consequently, large developmental genes have less chance to remain intact and functional after local rearrangements. This is also true for *O. dioica*, but its developmental genes tend to be smaller than their orthologs in other animals (Denoeud et al., 2010). The frequency of their duplications with retention of duplicates could be for this reason higher than in taxons with larger genomes. It was also proposed that the early fate predetermination and invariant development in the tunicate lineage gives additional freedom for genome reorganization including gene duplication, the disintegration

of the clusters and the rapid emergence of new genes (Holland, 2015). Specific gene families are also expanded in *Drosophila* genome (Rubin et al., 2000) and it was demonstrated that approximately 30% of young genes (most of them appeared through duplication) are essential for viability and contribute to the lineage-specific developmental program (Chen et al., 2010b). Another example of organism with high gene duplication rate is *C.elegans* (Lipinski et al., 2011), in which about one-third of the total gene complement is made of duplicates (Cutter et al., 2009).

5.3. Frequent recruitment of old transcription factors for a novel and lineage specific function.

The expression of multiple homeobox genes in the early larval epithelium suggested that these genes play a role in the oikoplastic epithelium patterning and the establishment of identities in its cellular territories. It also supported that such new functions had replaced or has been added to more ancestral functions in CNS development. To test the involvement of these genes in the oikoplastic epithelium formation, we devoted important efforts to validate a method of RNA interference, using other genes whose role in other developmental and physiological functions could be predicted with a high level of confidence. In the following lines, we focus the discussion on homeobox genes expressed in the developing oikoplastic epithelium, and for which the expression knockdown was attempted.

Orthologs of *prop* were found in the genomes of several invertebrates (*Drosophila melanogaster, Tribolium castateum, Saccoglossus kowalevskii,* and *Caenorhabditis elegans*), but we are not aware of detailed studies of their expression and function. As already mentioned, the single *prop* gene of vertebrates plays a central important role in the pituitary gland development. In mouse, it is expressed in the anterior part of the gland (adenohypophysis) where it contributes to the specialization and differentiation of lactotrophs, somatotropes, thyrotropes, and gonadotropes (Davis et al., 2011; Scully and Rosenfeld, 2002). Mutations of the gene are a common genetic cause of combined pituitary hormone deficiency in human and mouse (Kelberman et al., 2009). Prop1$^{-/-}$ mice fail to initiate proliferation of these cell types. Morpholino knock-down of the *prop1* gene in zebrafish, performed in our laboratory, reduced the expression of *pit1*, *prl* (prolactin) and *gh* (growth hormone) and induced abnormal shape, growth and

cellular organization in the developed adenohypophysis (Angotzi et al., 2011). In the genome of the ascidian *Ciona intestinalis*, the best studied tunicate species, a single *prop* gene is found and it is expressed only in the larval CNS (Imai et al., 2004, http://ghost.zool.kyoto-u.ac.jp/cgi-bin/txtgetkh.cgi?inkey=citb047e24&source=kh2013). In species other than *O. dioica,* we are not aware of *prop* genes being expressed in epithelia. *O. dioica* has two *prop* genes resulting from a lineage-specific duplication. This duplication occurred relatively early in the oikopleurid lineage since all closely related species have orthologs of these two *prop* genes. The duplicate *propA* is expressed in the posterior part of cerebral ganglion, possibly reflecting an ancient function. It is later on expressed in a defined territory of the trunk epithelium. The expression of *propB* is in contrast restricted to the trunk epithelium, in a region close to but broader than that of *propA*. Although there have been numerous genome rearrangements in *O. dioica* ancestors (Denoeud et al., 2010), both *prop* genes have remained placed next to each other. The maintenance of this physical association may have functional reasons, such as the utilization of common regulatory elements. Our RNA interference experiments supported our assumption that *prop* genes are involved in morphological and cell differentiation processes occurring during the development of the dorsal oikoplastic epithelium. This role seems confined to the areas where the *prop* genes are expressed, mainly the dorsal midline and the Anterior rosette. Until now, RNA interference was attempted for only five homeobox genes (two *prop* and three *otx* – see below), with a variable level of success. However, the types of spatio-temporal expression patterns observed for numerous homeobox (and Fox) genes authorize us to speculate that the oikoplastic epithelium formation involves a large and possibly much larger number of transcription factors.

Our functional studies of *otx* genes in *O. dioica* were unfortunately not as conclusive as for the *prop* genes. However, their lineage-specific duplications and expression patterns in *Oikopleura* also deserve some comments, since otx genes and their functions are much better known than prop genes in vertebrates and other invertebrates. Attempts of *otxA/B/C* knockdowns in *O. dioica* dramatically affected the morphology of embryos, possibly due to effects on functions other than the oikoplastic epithelium development. However, and in contrast with our RNAi experiments with *prop* genes, the ISH supported the persistence of important amounts of *otx* transcripts at the larval stage. Either very early developmental arrests followed efficient knockdowns when they occurred, or our RNAi technique did not knock down the expression of these genes. The two rounds of duplication for *otx* genes are also relatively ancient since all

larvacean species whose genome were sequenced possess three *otx* paralogs. Like the *prop* paralogs, those are also physically linked in the genome in several species. The three *otx* genes of *O. dioica* are expressed in the CNS where they may exert an ancestral function, and they are also highly expressed in specific regions of the trunk epithelium in young larvae. We believe that this epithelial expression is lineage-specific and novel, although the single *otx* gene of ascidians is also expressed in the trunk epithelium of larvae (Wada et al., 2004). The functional analysis of this ascidian gene revealed its involvement in the differentiation of both the sensory vesicle and the anterior trunk epidermis, including the region where the adhesive organ is formed. The epithelial expression of *otx* indeed occurs in the precursor region of the adhesive gland, which does not exist in larvaceans. Homology between the most anterior expression domain of *O. dioica otxA* and *otxC* genes (epithelial cells surrounding the mouth) and the sites of epithelial expression in ascidians cannot be ruled out. The expression of otx genes in the oral ectoderm was also observed in some invertebrates, including hemichordates (Harada et al., 2000), echinoderms (Hinman et al., 2003; Morris and Byrne, 2005) and annelids (Arendt et al., 2001), and it could therefore be very ancient (Mazza et al., 2010). The otx genes of vertebrate gnathostomes (at least two) are expressed in the CNS, in the developing neuroectoderm, and in the prospective sensory organs placodes. In two species of lampreys (agnathan), *otx* gene(s) are expressed in the upper and lower lips in addition to CNS and placodes (Tomsa and Langeland, 1999; Ueki et al., 1998). In contrast, *otx* expression in amphioxus seems restricted to the anterior CNS (Castro et al., 2006; Holland and Holland, 1998; Zhang and Mao, 2009). Overall, the most anterior expression of *O. dioica* otx genes in ectodermal cells surrounding the mouth may be ancient and conserved beyond larvaceans, whereas their more posterior and well-regionalized expression in the central part of the oikoplastic epithelium is probably novel and lineage-specific.

5.4. How did the new function of old genes evolve?

Genes can acquire a new function through changes in either their regulatory elements or their coding sequence (Lynch and Wagner, 2008; Prud'homme et al., 2007). The mechanisms underlying these changes are diverse. A new binding site can appear for an upstream transcription factor, or a new mode of regulation can result from novel interactions between regulatory

proteins. Changes in coding sequence include gain and loss of splice sites or short linear motifs, point mutations, or domain shuffling (Lynch and Wagner, 2008). The number of genes in a genome seems to have limits, and gene gains are expected to be compensated by gene losses (Tautz and Domazet-Lošo, 2011). A shift of gene expression into a new tissue, that we postulate for homeobox genes in the oikoplastic epithelium, could in principle be caused by an addition of regulatory elements. Gene duplications which are often observed for these genes should not be required for such changes to appear, and would be even less probable in species with a highly compact genome. The duplication of such transcription factors could have other consequences, such as a reduction of pleiotropy of the ancestral genes (Lynch and Wagner, 2008) with duplicates evolving new fine-tuned tissue-specific regulation, and/or functionally divergent proteins. To assess the extent of functional divergence between duplicates, we would need other tools such as transgenic technologies, allowing the insertion of promoter constructs or the knock-in of coding sequences. Unfortunately, transgenics were never produced in *Oikopleura*, despite important efforts. New hopes exist after our lab showed the efficiency of CRISPR-cas9 to induce targeted DSBs in the genome, which in principle offer a broad spectrum of opportunities for genetic manipulations (Deng and Chourrout, unpublished). At this point, there are no reasons to believe that the duplication of homeodomain transcription factors allowed their recruitment for the oikoplastic epithelium formation. The epithelial expression is most often found in both duplicates of the same gene, suggesting that it preceded the duplication event.

Changes of transcription factor expression patterns and/or their biochemical properties may create new gene regulatory networks in which new downstream effectors, such as the *oikosins,* can be incorporated. The house proteins – oikosins – have no clear orthologs in other species and are either new genes or genes that have rapidly diverged from their ancestors. Out of 80 identified *oikosins*, half possess known domain modules or show sequence similarity to known proteins. The other half show no such conservation and may have appeared *de novo* in larvaceans. One-third of the current *oikosin* complement probably arose through oikosin gene duplications (Hosp et al., 2012). A duplication of transcription factors may have happened in response to the growing diversity of oikosins. This hypothesis may be tested in the future, when we will know more about the set of transcription factors involved, allowing to establish fine correlations between their expression and those of most oikosins. RNA interference has shown a repression of *oik41a* consecutive to the knockdown of *prop* genes. There is at the moment no

reason to invoke a direct regulation of *oik41a* by *prop* genes, and the lack of *oik41a* expression may be due primarily to an inappropriate differentiation of *oik41a* producing cells. The oikoplastic epithelium is a tissue with highly specific programs of cell proliferation, growth, and endocycling (Ganot and Thompson, 2002). The change in the nucleus shape and size in the dorsal epithelium midline after RNAi-*prop* injections could indicate a violation of these programs. To assess the mechanistic relationships between candidate upstream transcription factors and the oikosins, we need to implement a molecular search of transcription factor direct targets.

As a conclusion ...

This work supports that quite a few homeobox genes are involved in the development of the oikoplastic epithelium. We have provided a brief and global description of this complex lineage-specific phenomenon that leads to the formation of well defined morphological and functional cellular territories. As these genes were arbitrarily considered in our initial candidate gene approach, we hypothesize that many other transcription factors and other developmentally regulated genes participate in the process. Most homeobox genes involved in the oikoplastic epithelium development have been duplicated, and both duplicates are expressed in a similar region of the epithelium. This suggests that the epithelial expression preceded the gene duplication events, which were not crucial for the genesis of the house via the oikoplastic epithelium emergence. Instead a high frequency of gene duplication may be a general feature of the *Oikopleura* genome, with no clear relationship to specific developmental features. Other homeobox genes expressed in the developing oikoplastic epithelium are not duplicated but were also recruited in this purpose. If the functional divergence of duplicated homeobox genes represents more than a simple division of work (subfunctionalization), they may have played a role in the complexification of the oikoplastic epithelium and of the house architecture.

References

Abercrombie, M. and Brachet, J. (Eds.) (2013). *Advances in morphogenesis (Vol. 1). The embryology of Ascidians.* Academic Press.

Aboobaker, A. A. and Blaxter, M. L. (2003). Hox Gene Loss during Dynamic Evolution of the Nematode Cluster. *Curr. Biol.* **13**, 37–40.

Acuña, J. L. (2001). Pelagic tunicates: why gelatinous? *Am. Nat.* **158**, 100–107.

Alonso, C. R., Maxton-Kuechenmeister, J. and Akam, M. (2001). Evolution of Ftz protein function in insects. *Curr. Biol*. **11**, 1473–1478.

Andersson, D. I., Jerlström-Hultqvist, J. and Näsvall, J. (2015). Evolution of New Functions De Novo and from Preexisting Genes. *Cold Spring Harb. Perspect. Biol.* **7**, a017996–.

Angotzi, A. R., Mungpakdee, S., Stefansson, S., Male, R., & Chourrout, D. (2011). Involvement of Prop1 homeobox gene in the early development of fish pituitary gland. *General and comparative endocrinology,* **171(3)**, 332-340.

Aravena, G. and Palma, S. (2002). Taxonomic identification of appendicularians collected in the epipelagic waters off northern Chile (Tunicata, Appendicularia). *Rev. Chil. Hist. Nat.* **75**, 307–325.

Arendt, D., Technau, U., & Wittbrodt, J. (2001). Evolution of the bilaterian larval foregut. *Nature,* **409(6816)**, 81-85.

Assis, R. and Bachtrog, D. (2013). Neofunctionalization of young duplicate genes in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17409–14.

Baulcombe, D. (2004). RNA silencing in plants. *Nature* **431**, 356–363.

Berná, L. and Alvarez-Valin, F. (2014). Evolutionary genomics of fast evolving tunicates. *Genome Biol. Evol.* **6**, 1724–38.

Bernstein, E., Caudy, A. A., Hammond, S. M. and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366.

Berrill, N. J. (1932). The mosaic development of ascidian egg. *Biol. Bull.* **63**, 381–386.

Bertrand, S. and Escriva, H. (2011). Evolutionary crossroads in developmental biology: amphioxus. *Development* **138**, 4819–30.

Boncinelli, E. and Morgan, R. (2001). Downstream of Otx2, or how to get a head. *Trends Genet.* **17**, 633–636.

Bone, Q. (1998). *The Biology of Pelagic Tunicates*. Oxford press.

Bouquet, J. M., Spriet, E., Troedsson, C., Ottera, H., Chourrout, D. and Thompson, E. M. (2009). Culture optimization for the emergent zooplanktonic model organism Oikopleura dioica. *J Plankt. Res* **31**, 359–370.

Bourlat, S. J., Juliusdottir, T., Lowe, C. J., Freeman, R., Aronowicz, J., Kirschner, M., Lander, E. S., Thorndyke, M., Nakano, H., Kohn, A. B., et al. (2006). Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88.

Braastad, C. D., Hovhannisyan, H., van Wijnen, A. J., Stein, J. L. and Stein, G. S. (2004). Functional characterization of a human histone gene cluster duplication. *Gene* **342**, 35–40.

Brown, F. D., Prendergast, A. and Swalla, B. J. (2008). Man is but a worm: chordate origins. *Genesis* **46**, 605–13.

Cameron, C. B., Garey, J. R. and Swalla, B. J. (2000). Evolution of the chordate body plan : New insights from phylogenetic analyses of deuterostome phyla. *Proceedings of the National Academy of Sciences,* **97(9),** 4469-4474.

Cañestro, C., Bassham, S. and Postlethwait, J. (2005). Development of the central nervous system in the larvacean Oikopleura dioica and the evolution of the chordate brain. *Dev. Biol.* **285**, 298–315.

Caracciolo, A., Di Gregorio, A., Aniello, F., Di Lauro, R. and Branno, M. (2000). Identification and developmental expression of three Distal-less homeobox containing genes in the ascidian Ciona intestinalis. *Mech. Dev.* **99**, 173–176.

Carroll, S. B. (2000). Endless forms: the evolution of gene regulation and morphological diversity. *Cell* **101**, 577–580.

Carroll, S. B. (2005). Evolution at Two Levels: On Genes and Form. *PLoS Biol.* **3**, e245.

Castro, L. F. C., Rasmussen, S. L. K., Holland, P. W. H., Holland, N. D. and Holland, L. Z. (2006). A Gbx homeobox gene in amphioxus: insights into ancestry of the ANTP class and evolution of the midbrain/hindbrain boundary. *Developmental biology* **295**, 40–51.

Chen, S., Armistead, J. S., Provost-Javier, K. N., Sakamoto, J. M., & Rasgon, J. L. (2010). Duplication, concerted evolution and purifying selection drive the evolution of mosquito vitellogenin genes. *BMC evolutionary biology*, **10(1)**, 1.

Chen, S., Zhang, Y. E. and Long, M. (2010b). New genes in Drosophila quickly become essential. *Science* **330**, 1682–1685.

Conklin, E. G. (1905). Mosaic development in ascidian eggs. *J. Exp. Zool.* **2**, 145–223.

Cronin, C. N., Zhang, X., Thompson, D. A. and McIntire, W. S. (1998). cDNA cloning of two splice variants of a human copper-containing monoamine oxidase pseudogene containing a dimeric Alu repeat sequence 1. *Gene* **220**, 71–76.

Crowther, R. J. and Whittaker, J. R. (1983). Developmental Autonomy of Muscle Fine Structure in Muscle Lineage Cells of Ascidian Embryos. *Dev. Biol.* **96**, 1–10.

Cutter, A. D., Dey, A. and Murray, R. L. (2009). Evolution of the Caenorhabditis elegans genome. *Mol. Biol. Evol.* **26**, 1199–1234.

Davis, S. W., Mortensen, A. H., & Camper, S. A. (2011). Birthdating studies reshape models for pituitary gland cell specification. *Developmental biology* ***352*****(2),** 215-227.

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

Deng, C., Cheng, C. C., Ye, H., He, X., Chen, L., Sean, B., Dengab, C., Yeab, H. and Heb, X. (2010). Evolution of an antifreeze protein by under escape neofunctionalization from adaptive conflict. *PNAS* **107**, 21593–21598.

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J. M., Da Silva, C., Brinkmann, H., et al. (2010). Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science*, **330(6009),** 1381-1385.

Dittmar, K. and Liberles, D. A. (2010). *Evolution after Gene Duplication.* Hoboken, NJ, USA: John Wiley & Sons, 2011.

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development,* **1994(Supplement)**, 135–142.

Duboule, D. and Dolle, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. *The EMBO journal*, **8(5)**, 1497–1505.

Elbashir, S. M., Lendeckel, W. and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200.

Fenaux, R. (1993). *The classification of appendicularia (Tunicata): history and current state*. Mémoirs de l'Institut Océanographique/Fondation Albert Ier, Prince de Monaco (ISSN 0304-5714. – 1993. – №. 17).

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998). Potent and specific genetic interference by double- stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–11.

Flagel, L. E. and Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* **183**, 557–64.

Flood, P. R. (2000). A new Appendicularian, Oikopleura gorskiu n.sp. (Tunicata) from norwegian fjord. *Belletin Zool. museum Univ. Amsterdam* **50**, 69–77.

Flood, P. R. (2005). *Towards a photographic atlas on a special taxonomic characters of oikopleurid Appendicularia (Tunicata). In book: Responce to Marine Ecosystems to Global Change. Ecologycal impact of Appendicularians.* (ed. Gorsky, G., Youngbluth, M., and Deibel, D.) Contemporary Publishing International.

Flood, P. R., Deibel, D. and Morris, C. (1998). The appendicularian house. In *The biology of pelagic tunicates.* (ed. Bone, Q.), pp. 105–124. Oxford University Press, Oxford.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. and Postlethwait, J. (1999). Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **151**, 1531–1545.

Galant, R. and Carroll, S. B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* **415**, 910–913.

Ganot, P. and Thompson, E. M. (2002). Patterning through Differential Endoreduplication in Epithelial Organogenesis of the Chordate, Oikopleura dioica. *Dev. Biol.* **252**, 59–71.

Gilbert, S. (2000). *Developmental biology, 6th edition*. 6th ed. Sunderland (MA): Sinauer Associates.

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. a and Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature* **433**, 481–487.

Graham, A., Papalopulu, N., Krumlauf, R., Ridgeway, T. and Hill, M. (1989). The Murine and Drosophila Homeobox Complexes Have Common Features of O rganization and Expression. *Cell* **57**, 367–378.

Haldane, J. B. S. (1933). The part played by recurrent mutation in evolution. *Am. natiralist* **67**, 5–19.

Hansen, J. L. S., Kiorboe, T. and Alldredge, A. L. (1996). Marine snow derived from abandoned larvacean houses: Sinking rates, particle content and mechanisms of aggregate formation. *Mar. Ecol. Prog. Ser.* **141**, 205–215.

Harada, Y., Okai, N., Taguchi, S., Tagawa, K., Humphreys, T. and Satoh, N. (2000). Developmental expression of the hemichordate otx ortholog. *Mech. Dev.* **91**, 337–339.

He, X. and Zhang, J. (2005). Rapid Subfunctionalization Accompanied by Prolonged and Substantial Neofunctionalization in Duplicate Gene Evolution. *Genetics,* **169(2),** 1157-1164

Hejnol, A. and Martindale, M. Q. (2009). Coordinated spatial and temporal expression of Hox genes during embryogenesis in the acoel Convolutriloba longifissura. *BMC Biol.* **7**, 65.

Hinman, V. F., Nguyen, A. T., & Davidson, E. H. (2003). Expression and function of a starfish Otx ortholog, AmOtx: a conserved role for Otx proteins in endoderm development that predates divergence of the eleutherozoa. *Mechanisms of development,* **120(10)**, 1165-1176.

Hirose, E., Kimura, S., Itoh, T. and Nishikawa, J. (1999). Tunic Morphology and Cellulosic Components Pyrosomas, Doliolids, and Salps (Thaliacea, Urochordata). *Biol. Bull.* **196**, 113–120.

Hirose, E. (2009). Ascidian tunic cells : morphology and functional diversity of free cells outside the epidermis. *Invertabrate Biol.* **128**, 83–96.

Hirotsune, S., Yoshida, N., Chen, A., Garrett, L., Sugiyama, F., Takahashi, S. and Yagami, K. (2003). An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* **423**, 91–100.

Holland, L. Z. and Holland, N. D. (1998). Developmental gene expression in amphioxus: new insights into the evolutionary origin of vertebrate brain region, neural crest, and rostrocaudal segmentation. *Am. Zool.* **38**, 647–658.

Holland, L. Z. (2015)**.** Genomics, evolution and development of amphioxus and tunicates: the goldilocks principle. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **324(4)**, 342-352.

Holland, N. D., Panganiban, G., Henyey, E. L. and Holland, L. Z. (1996). 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* **122**, 2911–2920.

Hosp, J., Sagane, Y., Danks, G. and Thompson, E. M. (2012). The evolving proteome of a complex extracellular matrix, the Oikopleura house. *PLoS One* **7**, e40172.

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* **256**, 119–124.

Hughes, A. L. (2005). Gene duplication and the origin of novel proteins. *Proceedings of the National Academy of Sciences of the United States of America,* **102**, 8791–8792.

Hutvagner, G. and Zamore, P. D. (2002). RNAi : nature abhors a double-strand. *Curr. Opin. Genet. Dev.* **12**, 225–232.

Ikuta, T., Yoshida, N., Satoh, N., & Saiga, H. (2004). Ciona intestinalis Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proceedings of the National Academy of Sciences of the United States of America,* **101(42),** 15118-15123.

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

Izant, J. G. and Weintraub, H. (1984). Inhibition of Thymidine Kinase Gene Expression by Anti-Sense RNA : A Molecular Approach to Genetic Analysis. *Cell* **36**, 1007–1015.

Keys, D. N., Lewis, D. L., Selegue, J. E., Pearson, B. J., Goodrich, L. V, Johnson, R. L., Gates, J., Scott, M. P. and Carroll, S. B. (1999). Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science (80).* **283**, 532–534.

Kelberman, D., Turton, J. P. G., Woods, K. S., Mehta, A., Al-Khawari, M., Greening, J., ... & Dattani, M. T. (2009). Molecular analysis of novel PROP1 mutations associated with combined pituitary hormone deficiency (CPHD). *Clinical endocrinology,* **70(1),** 96-103

Kimura, S., Ohshima, C., Hirose, E. and Nishikawa, J. (2001). Cellulose in the house of appendicularian Oikopleura rufescens. *Protoplasma* **216**, 71–74.

Kleinjan, D. A., Bancewicz, R. M., Gautier, P., Dahm, R. and Schonthaler, H. B. (2008). Subfunctionalization of Duplicated Zebrafish pax6 Genes by cis -Regulatory Divergence. *Plos Genet.* **4**, e29.

Lanjuin, A., VanHoven, M. K., Bargmann, C. I., Thompson, J. K. and Sengupta, P. (2003). Otx/otd Homeobox Genes Specify Distinct Sensory Neuron Identities in C. elegans. *Dev. Cell* **5**, 621–633.

Lemons, D. and McGinnis, W. (2006). Genomic Evolution of Hox Gene Clusters. *Science,* **313(5795)**, 1918-1922.

Lercher, M. J., Blumenthal, T., Hurst, L. D. (2003). Coexpression of neighboring genes in Caenorhabditis elegans is mostly due to operons and duplicate genes. *Genome research*, **13(2),** 238-243.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570.

Liao, D. (1999). Concerted Evolution : Molecular Mechanism and Biological Implications. *The American Journal of Human Genetics,* **64**, 24–30.

Lipinski, K. J., Farslow, J. C., Fitzpatrick, K. A., Lynch, M., Katju, V. and Bergthorsson, U. (2011). High spontaneous rate of gene duplication in Caenorhabditis elegans. *Curr. Biol.* **21**, 2621–2629.

Lynch, M. and Conery, J. S. (2003). The evolutionary demography of duplicated genes. *J. Struct. Funct. Genomics* **3**, 35–44.

Lynch, M. and Force, A. (2000a). The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics* **154**, 459–473.

Lynch, M. and Force, A. G. (2000b). The Origin of Interspecific Genomic Incompatibility via Gene Duplication. *Am. Nat.* **156**, 590–605.

Lynch, V. J. and Wagner, G. P. (2008). Resurrecting the role of transcription factor change in developmental evolution. *Evolution (N. Y).* **62**, 2131–2154.

Lynch, V. J. and Wagner, G. P. (2009). Multiple chromosomal rearrangements structured the ancestral vertebrate Hox-bearing protochromosomes. *PLoS Genet,* **5(1),** e1000349

Martínez-Morales, J. R., Dolez, V., Rodrigo, I., Zaccarini, R., Leconte, L., Bovolenta, P. and Saule, S. (2003). OTX2 activates the molecular network underlying retina pigment epithelium differentiation. *J. Biol. Chem.* **278**, 21721–31.

Matthysse, A. G., Deschet, K., Williams, M., Marry, M., White, A. R. and Smith, W. C. (2004). A functional cellulose synthase from ascidian epidermis. *Proceedings of the National Academy of Sciences of the United States of America*, **101(4),** 986-991.

Mazet, F., Hutt, J. A., Millard, J. and Shimeld, S. M. (2003). Pax gene expression in the developing central nervous system of Ciona intestinalis. *Gene Expr. Patterns* **3**, 743–745.

Mazza, M. E., Pang, K., Martindale, M. Q., & Finnerty, J. R. (2007). Genomic organization, gene structure, and developmental expression of three clustered otx genes in the sea anemone Nematostella vectensis. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution,* **308(4),** 494-506.

Mazza, M. E., Pang, K., Reitzel, A. M., Martindale, M. Q., & Finnerty, J. R. (2010). A conserved cluster of three PRD-class homeobox genes (homeobrain, rx and orthopedia) in the Cnidaria and Protostomia. *EvoDevo,* **1(1),** 1.

McGinnis, W., Garber, R. L., Wit-z, J., Kuroiwa, A. and Gehring, W. J. (1984). A Homologous Protein-Coding Sequence in Drosophila Homeotic Genes and Its Conservation in Other Metazoans. *Cell,* **37(2)**, 403–408.

Mighell, A. J., Smith, N. R., Robinson, P. A., Markham, A. F. (2000). Vertebrate pseudogenes. *FEBS letters*, **468(2-3)**, 109-114.

Morris, V. B., Byrne, M. (2005). Involvement of two Hox genes and Otx in echinoderm body-plan morphogenesis in the sea urchin Holopneustes purpurescens. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution,* **304(5)**, 456-467.

Nishida, H. (2008). Development of the appendicularian Oikopleura dioica: Culture, genome, and cell lineages. *Dev. Growth Differ.* **50**, S239–S256.

Nornes, S., Clarkson, M., Mikkola, I., Pedersen, M. and Bardsley, A. (1998). Zebrafish contains two Pax6 genes involved in eye development 1. *Mech. Dev.* **77**, 185–196.

Novina, C. D. and Sharp, P. A. (2004). The RNAi revolution. *Nature* **430**, 161–164.

O'Neill, R. S. and Clark, D. V (2013). Evolution of three parent genes and their retrogene copies in Drosophila species. *Int. J. Evol. Biol.* **2013**, 693085.

Ohno S. (1970). *Evolution by gene duplication*.

Olsen, M. a. and Schechter, L. E. (1999). Cloning, mRNA localization and evolutionary conservation of a human 5-HT7 receptor pseudogene. *Gene* **227**, 63–69.

Omotezako, T., Nishino, A., Onuma, T. a and Nishida, H. (2013). RNA interference in the appendicularian Oikopleura dioica reveals the function of the Brachyury gene. *Dev. Genes Evol.* **223**, 261–7.

Ota, T., & Nei, M. (1995). Evolution of immunoglobulin VH pseudogenes in chickens. *Molecular biology and evolution*, **12(1)**, 94-102.

Panopoulou, G. and Poustka, A. J. (2005). Timing and mechanism of ancient vertebrate genome duplications - the adventure of a hypothesis. *Trends Genet.* **21**, 559–67.

Park, J.-E., Heo, I., Tian, Y., Simanshu, D. K., Chang, H., Jee, D., Patel, D. J. and Kim, V. N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* **475**, 201–205.

Peer, Y. Van De, Maere, S. and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732.

Peer, Y. Van De, Maere, S. and Meyer, A. (2010). 2R or not 2R is not the question anymore. *Nat. Rev. Genet.* **11**, 166.

Prud'homme, B., Gompel, N. and Carroll, S. B. (2007). Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* **104 (Suppl 1)**, 8605–12.

Putnam, N. H., Butts, T., Ferrier, D. E. K., Furlong, R. F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071.

Qian, W., Liao, B. Y., Chang, A. Y. F., Zhang, J. (2010). Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends in Genetics*, **26(10)**, 425-430.

Rastogi, S., Liberles, D. A. (2005). Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC evolutionary biology,* **5(1)**, 1.

Rebagliati, M. R. and Melton, D. A. (1987). Antisense RNA Injections in Fertilized Frog Eggs Reveal an RNA Duplex Unwinding Activity. *Cell* **48**, 599–605.

Robinson-Rechavi, M., Laudet, V. (2001). Evolutionary rates of duplicate genes in fish and mammals. *Molecular Biology and Evolution*, **18(4)**, 681-683.

Robison, B. H., Reisenbichler, K. R. and Sherlock, R. E. (2005). Giant Larvacean Houses : Rapid Carbon Transport to the Deep Sea Floor. *Science* **308**, 1609–1611.

Rosso, L., Marques, A. C., Weier, M., Lambert, N., Lambot, M.-A., Vanderhaeghen, P. and Kaessmann, H. (2008). Birth and rapid subcellular adaptation of a hominoid-specific CDC14 protein. *PLoS Biol.* **6**, e140.

Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., et al. (2000). Comparative genomics of the eukaryotes. *Science* **287**, 2204–15.

Sagane, Y., Zech, K., Bouquet, J. M., Schmid, M., Bal, U. and Thompson, E. M. (2010). Functional specialization of cellulose synthase genes of prokaryotic origin in chordate larvaceans. *Development* **137**, 1483–1492.

Sagane, Y., Hosp, J., Zech, K. and Thompson, E. M. (2011). Cytoskeleton-mediated templating of complex cellulose-scaffolded extracellular structure and its association with oikosins in the urochordate Oikopleura. *Cell.Mol. Life Sci.* **68**, 1611–1622.

Saló, E., Pineda, D., Marsal, M., Gonzalez, J., Gremigni, V. and Batistoni, R. (2002). Genetic network of the eye in Platyhelminthes: expression and functional analysis of some players during planarian regeneration. *Gene* **287**, 67–74.

Santini, S., Boore, J. L. and Meyer, A. (2003). Evolutionary Conservation of Regulatory Elements in Vertebrate Hox Gene Clusters. *Genome Res.* **13**, 1111–1122.

Sasakura, Y., Kanda, M., Ikeda, T., Horie, T., Kawai, N., Ogura, Y., ... & Fujiwara, S. (2012). Retinoic acid-driven Hox1 is required in the epidermis for forming the otic/atrial placodes during ascidian metamorphosis. *Development,* **139(12),** 2156-2160.

Sato, R., Tanaka, Y. and Ishimaru, T. (2001). House production by Oikopleura dioica (Tunicata , Appendicularia) under laboratory conditions. *J. Plankt. reseach* **23**, 415–423.

Sato, R., Tanaka, Y. and Ishimaru, T. (2003). Species-specific house productivity of appendicularians. *Mar. Ecol. Prog. Ser.* **259**, 163–172.

Satoh, N. (2003). The ascidian tadpole larva: comparative molecular development and genomics. *Nat. Rev. Genet.* **4**, 285–95.

Satoh, N. (2008). An aboral-dorsalization hypothesis for chordate origin. *Genesis* **46**, 614–22.

Satoh, N. (2009). An Advanced Filter-Feeder Hypothesis for Urochordate Evolution An Advanced Filter-Feeder Hypothesis for Urochordate Evolution. *Zoological science,* **26**, 97–111.

Satoh, N. (2013). *Developmental genomics of ascidians.* John Wiley & Sons.

Scully, K. M., Rosenfeld, M. G. (2002). Pituitary development: regulatory codes in mammalian organogenesis. *Science*, **295(5563)**, 2231-2235

Smith, K. M., Gee, L., Blitz, I. L. and Bode, H. R. (1999). CnOtx, a member of the Otx gene family, has a role in cell movement in hydra. *Dev. Biol.* **212**, 392–404.

Seo, H. C., Edvardsen, R. B., Maeland, A. D., Bjordal, M., Jensen, M. F., Hansen, A., ... & Reinhardt, R. (2004). Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. *Nature,* **431(7004)**, 67-71.

Smith, A. A., Wyatt, K., Vacha, J., Vihtelic, T. S., Samuel Zigler, J., Wistow, G. J., & Posner, M. (2006). Gene duplication and separation of functions in αB-crystallin from zebrafish (Danio rerio). *FEBS* **273**, 481–490.

Soviknes, A. M., Chourrout, D. and Glover, J. C. (2005). Development of putative GABAergic neurons in the appendicularian urochordate Oikopleura dioica. *J Comp Neurol* **490**, 12–28.

Soviknes, A. M., Chourrout, D. and Glover, J. C. (2007). Development of the caudal nerve cord, motoneurons, and muscle innervation in the appendicularian urochordate Oikopleura dioica. *J Comp Neurol* **503**, 224–243.

Spada, F., Steen, H., Troedsson, C., Kallesoe, T., Spriet, E., Mann, M. and Thompson, E. M. (2001). Molecular patterning of the oikoplastic epithelium of the larvacean tunicate Oikopleura dioica. *J. Biol. Chem.* **276**, 20624–32.

Stach, T. and Turbeville, J. M. (2002). Phylogeny of Tunicata inferred from molecular and morphological characters. *Molecular phylogenetics and evolution,* **25(3),** 408–428.

Stach, T., Winter, J., Bouquet, J. M., Chourrout, D. and Schnabel, R. (2008). Embryology of a planktonic tunicate reveals traces of sessility. *Proc. Natl. Acad. Sci. USA* **105**, 7229–7234.

Swalla, B. J. S., Cameron, C. B. C., Corley, L. S. C. and Garey, J. R. G. (2000). Urochordates Are Monophyletic Within the Deuterostomes. *Syst. Biol.* **49**, 52–64.

Tautz, D., Domazet-Lošo, T. (2011). The evolutionary origin of orphan genes. *Nature Reviews Genetics,* 12(10), 692-702.

Tomsa, J. M., & Langeland, J. A. (1999). Otx Expression during Lamprey Embryogenesis Provides Insights into the Evolution of the Vertebrate Head and Jaw. *Developmental biology*, **207(1)**, 26-37.

Troedsson, C., Bouquet, J.-M., Skinnes, R., Acuna, J.-L., Zech, K., Frischer, M. E. and Thompson, E. M. (2009). Regulation of filter-feeding house components in response to varying food regimes in the appendicularian, Oikopleura dioica. *J. Plankt. reseach* **31**, 1453–1463.

Tsagkogeorga, G., Turon, X., Hopcroft, R. R., Tilak, K., Feldstein, T., Shenkar, N., Loya, Y., Huchon, D., Douzery, E. J. P. and Delsuc, F. (2009). An apdated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. *BMC Evol. Biol.* **16**, 1–16.

Ueki, T., Kuratani, S., Hirano, S., & Aizawa, S. (1998). Otx cognates in a lamprey, Lampetra japonica. Development genes and evolution, **208(4)**, 223-228.

Urbach, R. (2007). A procephalic territory in Drosophila exhibiting similarities and dissimilarities compared to the vertebrate midbrain/hindbrain boundary region. *Neural Dev.* **2**, 23.

van Hoof, A. (2005). Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. *Genetics* **171**, 1455–61.

Vanin, Elio F. (1985). "Processed pseudogenes: characteristics and evolution." *Annual review of genetics* **19.1**, 253-272.

Wada, S., Tokuoka, M., Shoguchi, E., Spagnuolo, A., Branno, M., Kohara, Y., Rokhsar, D., Levine, M., Saiga, H., Satoh, N., et al. (2003). A genomewide survey of developmentally relevant genes in Ciona intestinalis. II. Genes for homeobox transcription factors. *Dev. Genes Evol.* **213**, 222–234.

Wada, S., Sudou, N. and Saiga, H. (2004). Roles of Hroth, the ascidian otx gene, in the differentiation of the brain (sensory vesicle) and anterior trunk epidermis in the larval development of Halocynthia roretzi. *Mechanisms of development,* **121(5)**, 463-474.

Wang, W., Yu, H. and Long, M. (2004). Duplication-degeneration as a mechanism of gene fission and the origin of new genes in Drosophila species. *Nat. Genet.* **36**, 523–7.

Weintraub, H., Izant, J. G. and Harland, R. M. (1985). Anti-sense RNA as a molecular tool for genetic analysis. *Trends Genet.* 22–25.

Yamamoto, T., Kawamoto, R., Fujii, T., Sakamoto, N. and Shibata, T. (2007). DNA variations within the sea urchin Otx gene enhancer. *FEBS Lett.* **581**, 5234–40.

Yang, S., Tutton, S., Pierce, E. and Yoon, K. (2001). Specific Double-Stranded RNA Interference in Undifferentiated Mouse Embryonic Stem Cells. *Mol. Cell. Biol.* **21**, 7807–7816.

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**, 292–298.

Zhang, Y. and Mao, B. (2009). Developmental Expression of an Amphioxus *Branchiostoma belcheri* Gene Encoding a GATA Transcription Factor. *Zool. Res.* **30**, 137–143.

Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., ... & Wang, W. (2008). On the origin of new genes in Drosophila. *Genome research*, **18(9)**, 1446-1455.

Zimmermann, T. S., Lee, A. C. H., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M. N., Harborth, J., Heyes, J. A., Jeffs, L. B., John, M., et al. (2006). RNAi-mediated gene silencing in non-human primates. *Nature,* **441**, 111–114.