UNIVERSITAT de BARCELONA

Deconstruction of the cardiopharyngeal gene regulatory network in appendicularians, a paradigmatic study of *Oikopleura dioica* as an evolutionary knockout model

Alfonso Ferrández Roldán

ALFONSO FERRÁNDEZ ROLDÁN

DECONSTRUCTION OF THE CARDIOPHARYNGEAL GENE REGULATORY NETWORK IN APPENDICULARIANS, A PARADIGMATIC STUDY OF *OIKOPLEURA DIOICA* AS AN EVOLUTIONARY KNOCKOUT MODEL

UNIVERSITAT DE
BARCELONA

Tesis Doctoral

Universidad de Barcelona

Deconstruction of the cardiopharyngeal gene regulatory network in appendicularians, a paradigmatic study of *Oikopleura dioica* as an evolutionary knockout model

Memoria presentada por

Alfonso Ferrández Roldán

Para optar al grado de

Doctor por la Universidad de Barcelona

Programa de Genética

Departamento de Genética, Microbiología y Estadística

Codirector:

Cristian Cañestro García

Tutor:

Jordi Garcia Fernàndez

Codirector:

Jordi Garcia Fernàndez

Doctorando:

Alfonso Ferrández Roldán

Barcelona, diciembre 2020

AGRADECIMIENTOS

Han pasado ya unos años desde que inicié esta aventura. Recuerdo que todo comenzó con una lista de doce genes; esa pequeña lista, como una semilla en un bosque, fue creciendo y evolucionando hasta convertirse en un sólido proyecto cargado tanto de momentos irritables como inolvidablemente buenos. Y son estos buenos momentos los que más ahondan en mi memoria junto con el recuerdo de las personas y las instituciones que los hicieron posibles y a los cuales estoy ampliamente agradecido.

En primer lugar la realización de este trabajo no hubiese sido posible sin la ayuda FPU recibida por el Ministerio de Educación, Cultura y Deporte (FPU014/02654) así como sin la subvención del proyecto por parte del Ministerio de Ciencia e innovación (BFU2016-80601). De la misma forma, quiero agradecer a la Universidad de Barcelona por acogerme en su programa de doctorado y en especial al departamento de Genética, Microbiología y Estadística por las reuniones, debates y congresos en los que he podido mostrar y discutir mis resultados.

El mayor de mis agradecimientos a mis directores Jordi García y en especial a Cristian Cañestro por todo lo que he aprendido a su lado. Agradecer también al doctor Ricard Albalat todas las discusiones y ayuda en este proyecto. No me puedo olvidar de todas las personas que durante estos años han pasado por el laboratorio y tantas cosas me han enseñado: Josep, Nuria, Alba, Marcos, Paula, Enya, Marc, Gaspar, Vittoria, Sebastián y Sara. Todos ellos de una manera más o menos directa son copartícipes de los resultados obtenidos.

Finalmente, no sería quien soy y no habría llegado hasta aquí sin el apoyo de mi familia. Muchas gracias a mis padres por quererme tal y como soy y por luchar por sacar lo mejor de mí, a mi hermana, por aguantarme tantos años y apoyarme en todo y a Samuel, por hacer tan especiales estos últimos tres años.

ABSTRACT

The bloom of genomic data has revealed a vast amount of gene losses across all life kingdoms. However, the impact of gene loss on the evolution of the mechanisms of embryo development remains an important challenge. In this work, we have used the successful gene loser *Oikopleura dioica*, to study the impact of gene loss on the evolution of the cardiopharyngeal gene regulatory network (GRN), and we have extrapolated our results to decipher the ancestral condition of tunicates as free-living or sessile, a hot topic of discussion.

To address this question, we have searched for gene losses by combining best reciprocal blast hit (BRBH) with exhaustive phylogenetic reconstructions of the gene family of interest. We have also performed expression analyses of the present orthologs to test for their cardiac function as well as, in the case of lost orthologs, with paralogs trying to detect potential events of function shuffling. Finally, we performed functional analyses by inhibiting the FGF and BMP signaling pathways and started the implementation of a microinjection facility for future functional analyses by gene targeting.

Our results show a clear deconstruction of the cardiopharyngeal GRN with the loss of many genes (*Mesp*, *Ets1/2a*, *Gata4/5/6*, *Mek1/2*, *Tbx1/10*, and RA- and FGF-signaling related genes) and cardiac subfunctions (*FoxF*, *Islet*, *Ebf*, *Mrf*, *Dach*, and Bmp signaling) crucial for cardiopharyngeal development in ascidians and vertebrates. All these losses have led to the dismantling of two genetic modules related to the maintenance of multipotency in the cardiopharyngeal precursors. This has been accompanied by the loss of the second heart field and pharyngeal muscles in appendicularians, which has been phenotypically translated into an open bilaminar heart with an accelerated development compared to the tubular heart present in the rest of chordates. The deconstruction of the cardiopharyngeal GRN in appendicularians can therefore be interpreted as an evolutionary adaptation to the transition from a sessile to a free-living lifestyle based on the innovation of the filter-feeding house.

Therefore, our results show *O. dioica* as a paradigmatic example of the advantages of using species that along their evolution has lost many genes (evolutionary knockout models, eKO) to better understand the evolution of GRNs, mechanisms of embryo development, or any physiological adaptation in the absence of any given gene of interest

INDEX

ABBREVIATIONS

ActnC	cytoplasmic actin
ActnM	muscular actin
AhR	aryl hydrocarbon receptor
AhRR	aryl hydrocarbon receptor repressor
aNHEJ	alternative non-homologous end joining
AP	anterior-posterior
ASMF	atrial siphon muscle field
ATM	anterior tail muscle cell
ATP	adenosine triphosphate
B-actin	body-wall actin
BaP	benzo[a]pyrene
BLAST	basic local alignment search tool
BMP	bone morphogenetic protein
BRBH	best reciprocal blast hit
cDNA	complementary DNA
cNHEJ	canonical non-homologous end joining
CNS	central nervous system
CTF	cardiac transcription factor
CYP	cytochrome P450
DAV	diagnostic-actin-value
DD	trans,trans-2,4-decadienal
DDC	duplication-degeneration-complementation
DIC	differential interference contrast
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNAi	interference DNA
DSB	double strand breaks
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
DV	dorso-ventral
eKO	evolutionary knockout
EM	early multipotent
EST	expressed sequence tag
FGF	fibroblast growth factor
FHF	first heart field

FHP	first heart precursors
GABA	gamma-aminobutyric acid
GRN	gene regulatory network
HIV	human immunodeficiency virus
HMG	high-mobility group
hpf	hours post-fertilization
LCA	last common ancestor
LM	late multipotent
LTR	long terminal repeats
miRNA	micro RNA
mpf	minutes post-fertilization
mRNA	messenger RNA
N-actin	notochord actin
P-actin	paraxial actin
PCR	polimerase chain reaction
pETS	phosphorylated ETS
PUA	polyunsaturated aldehyde
qRTPCR	quantitative RTPCR
RA	retinoic acid
RNA	ribonucleic acid
RNA-Seq	RNA sequencing
RNAi	RNA interference
rpm	revolutions per minute
RTPCR	reverse transcription PCR
SHF	second hear field
SHP	second heart precursors
SL RNA	spliced-leader RNA
snRNA	small nuclear RNA
Sox	Sry-type HMG box
sSW	sterile sea water
STVC	second trunk ventral cell
TE	transposable element
TVC	trunk ventral cell
WMISH	whole-mount in situ hybridization

INTRODUCTION

1 EVODEVO

One of the main challenges in biology is to decipher the evolutionary mechanisms that have favored the emergence of such vast range of living organisms. The answer to this matter brings together many different fields of biology as paleontology, comparative anatomy, comparative embryology, development and evolution. The last two areas, development and evolution, have gone together since 1866 when Ernst Haeckel put forth the theory of recapitulation inspired by the publication of The Origin of Species by Darwin (1859) and the idea, exposed by Karl Ernst Von Baer, that embryos look much more similar than adults. This theory, stating that ontogeny recapitulates phylogeny, has helped to phylogenetically reclassify many metazoan groups just by studying their embryo development. For example, tunicates were moved from the Phylum Mollusca to the Phylum Chordata after Kowalevsky discovered the presence of the notochord in the ascidian larvae (Kowalevsky A., 1866). During the 20th century, the connection between development and evolution became even stronger, and both fields were finally combined on a new area of biology known as EvoDevo. The

EvoDevo studies how evolution has affected the developmental mechanisms to produce the diversification of the species. Among those evolutionary changes, gene duplication has traditionally been considered one of the main mechanisms governing evolution (Force et al., 1999; Ohno, 1970). However, the recent discovery of a significant amount of gene losses among all life kingdoms is revealing gene loss as a major mechanism generating genetic diversity, and therefore, with a great potential to generate phenotypic diversity (Albalat and Cañestro, 2016; Fernández and Gabaldón, 2020; Guijarro-Clarke et al., 2020; Miller et al., 2005; Simakov et al., 2013).

1.1 THE PARADOX OF EVODEVO

During the last decades of the 20th century, the discovery that homologs of *Drosophila* developmental genes were present and had the same function in evolutionarily distant species revealed what can be considered one of the first paradox of EvoDevo (Jacob, 1977; Lewis, 1978; Nüsslein-volhard and Wieschaus, 1980). This paradox brought to attention that phylogenetically distant organisms, despite showing different characteristics (phenotypic diversity), shared similar

genetic developmental toolkits (genotypic unit) (Jacob, 1977; King and Wilson, 1975). For instance, the transcription factor eyeless in *Drosophila* and its *Pax-6* homolog in the mouse is essential in both organisms for eye development (Quiring et al., 1994), and their function is so conserved that expression of the mouse *Pax-6* in atypical organs of *Drosophila* causes the appearance of ectopic eyes (Halder et al., 1995). The conservation of the *Hox* cluster to organize the anterior-posterior body axis across bilaterians (Garcia-Fernàndez, 2005) and the presence of homologs of the *tinman/Nkx2.5* transcription factor in all developing hearts (Evans et al., 1995; Lints et al., 1993; Tonissen et al., 1994) are other examples corroborating that phenotypically different organisms share very similar "toolkits" of developmental genes.

If distant species use the same genetic toolkits, one of the main questions in EvoDevo has been to discover how the phenotypical variability between species has evolved. Numerous investigations have established that these differences between species result from minor changes in genetic toolkits that accumulate through evolution. Therefore, although, in general, genetic toolkits are conserved, they suffer mutations in coding and regulatory regions, gene duplications, and gene losses that, ultimately, cause phenotypic alterations. For example, mutations in the coding

sequence of developmental genes can change the function of the protein and transform an activator into a repressor. This happened with the *Ultrabithorax* gene during arthropod evolution. A mutation in the coding region of this abdominal gene in some crustacean/insect ancestor uncovered a limb-repression function that culminated with the origin of the hexapods (Galant and Carroll, 2002; Ronshaugen et al., 2002). However, alterations in regulatory regions of a gene, or different enhancers, usually have a more substantial effect on animal diversification due to their ability to evolve independently. Mutations in enhancers, moreover, can produce a wider variety of phenotypes, such as changes in transcription levels or temporal or spatial alterations (**Fig. I1 A-D**). Therefore, changes in gene regulatory elements are postulated to contribute more to morphological evolution than changes in coding regions (Carroll, 1995).

Gene duplication also plays a fundamental role in the generation of genetic diversity. Ever since in 1970 Susumo Ohno published "Evolution by gene duplication," numerous studies have supported gene duplication as one of the main driving forces of evolution (Lynch, 2002; Pastor-Satorras et al., 2003; Teichmann and Babu, 2004). The classical model for the evolution of duplicate genes predicts that sometimes rare beneficial mutations provide a novel function to one of the duplicates

(neofunctionalization) while the second copy maintains the original function (Ohno, 1970) (**Fig. I1 E**). However, as beneficial mutations are unusual, this model is unable to explain the abundant preservation of gene duplicates across evolution. To address this problem, Force et al. (1999) proposed the duplication-degeneration-complementation (DDC) model. This model states that degenerative mutations in regulatory regions, much more likely than beneficial mutations, can partition the ancestral gene function by forcing the maintenance of both duplicates (subfunctionalization) (**Fig. I1 F**). Although neo and subfunctionalization are possible, the most common fate for a duplicate copy is the fixation of a loss-of-function mutation and its subsequent loss (nonfunctionalization) (**Fig. I1 G**). However, loss-of-function mutations are not exclusive of duplicated genes, but also, they can affect any gene on the genome. In fact, during the last years, the increasing availability of genome data has elucidated that gene losses are pervasive in all life kingdoms and that, in many groups of organisms, their incidence is over the amount of novelty (Albalat and Cañestro, 2016; Guijarro-Clarke et al., 2020). Therefore, despite the vast amount of gene losses have rested hidden on unsequenced genomes, they have recently arisen as one of the main evolutionary forces that can generate genetic diversity.
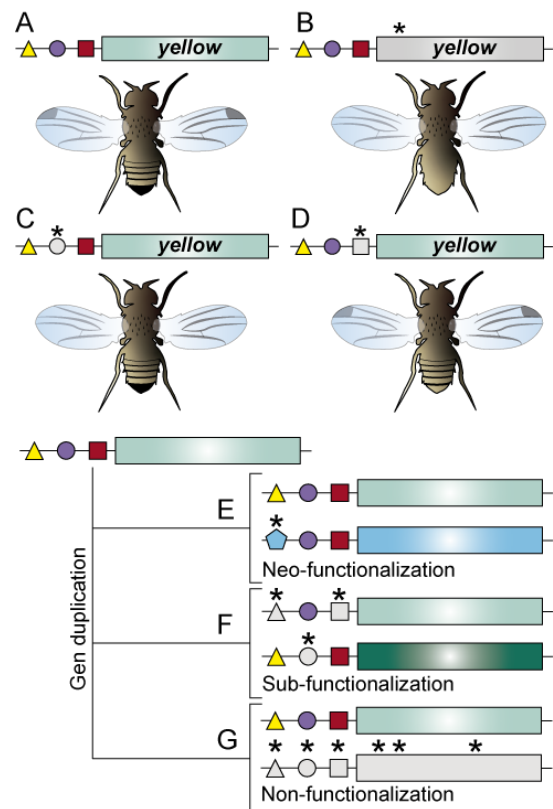
Fig. I1. Alterations in coding and regulatory sequences generate diversity. The *yellow* (*y*) gene in drosophila produces black pigment in different body parts. Various *yellow* enhancers control *y* expression in different body parts like the abdominal stripes, the posterior abdominal segments, or the wings (**A**). A mutation in the *y* coding region prevents *y* expression in all three body parts, wings, posterior abdomen, and abdomen strips (**B**). Mutations in different enhancers produce the absence of pigmentation in wings (**C**) or posterior abdomen (**D**). When a gene duplicates, one of the copies can change its fate. In non-functionalization, one of the copied genes accumulates mutations in coding and regulatory sequences, resulting in the loss of the gene (**E**). In sub-functionalization, mutations in the enhancers segregate the ancestral function between both genes (**F**). In neo-functionalization, one copy acquires new tasks by the gain of new regulatory elements (**G**). Asterisks represent mutations. Regulatory and coding regions in grey represent sequences that have lost their function.

1.2 Gene loss, breaking assumptions

During the 20th century, the pre-Darwinian concept of *scala naturae,* in which a series of organisms progresses from simple to complex, still had defenders. Although Darwin had already recognized that "becoming more complex did not necessarily mean progress and that simplification was a common state of affairs" (Darwin, 1876 p.176), the complexification idea used to hide in the background of many evolutionary hypotheses. This complexification idea was assumed to correlate with an increase in gene number, which presupposed that eukaryotes contained more genes than bacteria, animals more genes than plants, and vertebrates more genes than invertebrates (reviewed in Szathmáry et al., 2001). This perspective lost its credence after the sequencing of the genome of various cnidarian species. Cnidarian genomes uncovered that the ancestral eumetazoan genome was much more complex than initially expected and included many genes previously considered to be vertebrate innovations. The result was the revelation that gene losses do not seem to be occasional or insignificant, but a process that has widely occurred across all animal phyla (Kortschak et al., 2003; Putnam et al., 2007; Suga et al., 2013; Technau et al., n.d.).

Although mutations that cause gene losses are random events, the patterns of gene loss do not appear to be stochastic but show obvious biases depending on the genomic position, or the gene function (Albalat and Cañestro, 2016; Holland et al., 2017) Regarding genomic position, after whole-genome duplications, there is an asymmetric distribution of gene loss between ohnologons (e.g. chromosomic regions originated after genome duplication). However, the forces that influence these losses remain unknown (Albalat and Cañestro, 2016; Makino and McLysaght, 2012). Concerning gene function, there exist significant differences in gene retention between different groups of organisms according to Gene Ontology categories. For example, categories such as "protein modification," "protein metabolism," and "catabolism" are more prone to be lost in fish than in vertebrates, whereas genes involved in "catalytic activity" show the opposite trend (Blomme et al., 2006; Koonin et al., 2004).

A particular and exciting case of functional bias is the co-elimination of functionally connected genes that, progressively, dismantle a pathway after the relaxation of environmental constraints (**Fig. I2**). Numerous examples of gene losses due to co-elimination have been discovered in
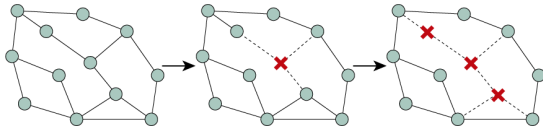
Fig. I2. Co-elimination of a gene pathway. The loss of the central gene of a pathway made the related genes to become useless and subsequently lost.

different organisms, such as spliceosome components in yeast (Aravind et al., 2000), the ubiquitin-proteasome-signalosome system in nematodes (Koonin et al., 2004), or proteins related to the metabolism and signaling of retinoic acid (RA) in *Oikopleura dioica* (Martí-Solans et al., 2016). In this last example, the co-elimination affected a broad array of enzymes related to the synthesis, degradation, and signaling of this metabolic pathway making the development of this chordate independent of this morphogen (Martí-Solans et al., 2016).

1.3 EVOLUTION BY GENE LOSS

One of the most surprising findings of the post-genomic era has been to what extent organisms can tolerate the inactivation of a significant number of genes (Papp et al., 2011). Large single-gene deletion screens on multiple organisms have demonstrated that from 65 to 90% of their protein-coding genes can be lost without altering the organism viability under laboratory conditions (Baba et al., 2006; Dietzl et al., 2007; Giaever et al., 2002; Kamath et al., 2003; Kim et al., 2010; Korona, 2011). This fact, referred to as the knockout paradox

(Papp et al., 2011), may indicate that the organisms have high mutational robustness with alternative genes supplying the lost ones, but also that under some environmental conditions, many genes become dispensable and prone to be lost (Albalat and Cañestro, 2016). When a gene becomes dispensable, it can be lost without affecting the fitness of the organism and favoring a scenario of regressive evolution (neutral loss), or it can be adaptatively lost as a frequent evolutionary response to environmental change (adaptive loss) (Fig. I3).

1.3.1 ADAPTATIVE GENE LOSSES: THE LESS-IS-MORE HYPOTHESIS

The less-is-more hypothesis proposes that gene loss can be a significant evolutionary adaptive strategy with great potential to rapidly occur after, for example, drastic shifts of environmental conditions (Olson and Varki, 2003; Olson, 1999). There exist reports of adaptive non-functional mutations in a vast range of organisms, from bacteria to vertebrates, including some paradigmatic cases in humans. For example, the loss-of-function mutation of *DUFFY* makes the carrier resistant to vivax malaria, which has become this allele variant predominant in sub-Saharan African populations as a result of natural selection (Howes et al., 2011). The loss-of-function mutation of the *CCR5* membrane receptor, which confers resistance to HIV, also
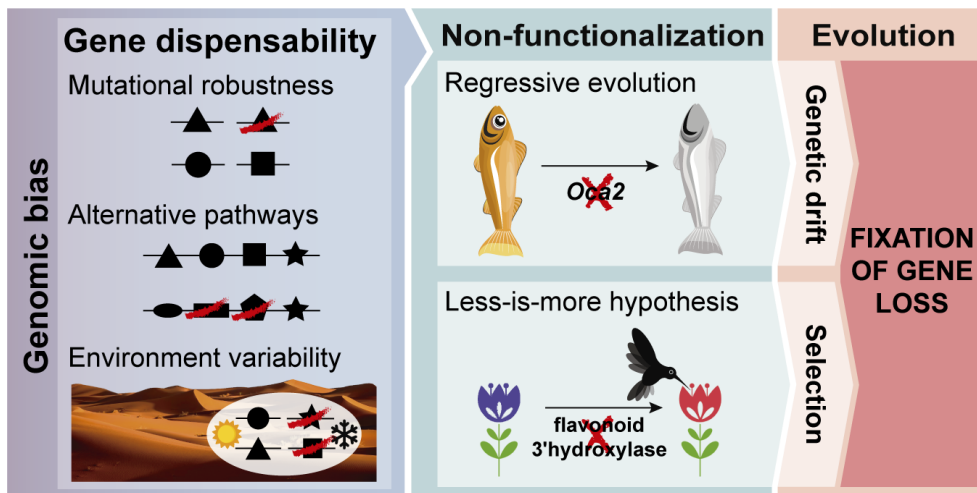
Fig. I3. Conceptual framework for gene loss. The probability of the fixation of a gene loss depends on how its non-functionalization affects the fitness of the organism. Mutational robustness, alternative pathways, or genes that become useless after an environmental change, are highly dispensable (cross out figures) and might be lost, without affecting the fitness of the organism as a case of regressive evolution (e. g. the loss of the eyes and pigmentation on the Mexican cavefish that lives in dark caves without light). These neutral non-functional mutations are fixed by genomic drift. The less-is-more hypothesis states that the loss of a gene after an environmental change can improve the fitness of the organism, making this loss adaptive and evolutionarily selected (e. g. the loss of the flavonoid 3'hydroxylase on *Ipomoea quamonlit* derived on a change of the color of the flower as an adaptation to bird pollination (Zufall and Rausher, 2004). The potential of a gene to be lost, apart from its dispensability, is also influenced by its position on the genome or its function (genomic bias).

shows a wide distribution, but in this case, across the European population. However, as HIV is a new disease, other viral infections could be the reason for its actual high frequency in Europe (Novembre et al., 2005).

1.3.2 NEUTRAL GENE LOSSES: REGRESSIVE EVOLUTION

Classically, gene losses have been studied in a scenario of regressive evolution in which losses of useless characteristics happen over time. There are several examples of regressive evolution such as the loss of genes related to vitamin C biosynthesis in different vertebrate lineages that have adopted vitamin C rich diet (Drouin et al., 2011), or the absence of

genes related to vision in the Mexican cavefish after colonization of dark environments (Jeffery, 2009). Those gene losses related to regressive evolution have been usually considered evolutionary neutral and driven by genetic drift since the loss of the dispensable characteristic does not offer a biological advantage. However, recent studies are questioning the neutrality of some regressive evolution examples. Back to the Mexican cavefish, for instance, recent analyses have suggested that the loss of vision could be adaptive because it reduces the amount of energy spent on the development and maintenance of expensive neural tissue. That decrease in energy waste could represent a selected advantage in the poor nutrient

environments where it lives (Moran et al., 2015).

1.4 GENE LOSSES ACROSS THE TREE OF LIFE

Gene losses affect all life kingdoms and might have had a great impact on most of them. For example, in some archaea and prokaryote species, gene loss has a three-times higher rate than gene gains (Koskiniemi et al., 2012; Nelson-Sathi et al., 2015; Puigbò et al., 2014), and in the case of plants and fungi, gene loss has been more frequent after whole-genome duplication events. The evolution of protostomes has experienced an extensive process of gene loss with the removal of 17% of the ancestral eumetazoan gene families (Albalat and Cañestro, 2016; Guijarro-Clarke et al., 2020). In deuterostomes, gene loss may have played a fundamental role in the evolution of the different body plans (Albalat and Cañestro, 2016; De Robertis, 2008), although it has had a lower general impact, except for tunicates (a.k.a. urochordates). Tunicates are a subphylum of the chordates, our phylum, whose genomes appear to have a "liberal" evolutionary pattern of gene loss (Dehal et al., 2002; Holland and Gibson-Brown, 2003; Hughes and Friedman, 2005; Somorjai et al., 2018), which contrasts with the "conservative" pattern of cephalochordates and vertebrates (Somorjai et al., 2018).

2 *OIKOPLEURA DIOICA*, AN ANIMAL MODEL TO STUDY THE IMPACT OF GENE LOSS ON THE EVOLUTION OF MECHANISMS OF DEVELOPMENT IN CHORDATES

Tunicates are cosmopolitan marine filter feeders distributed around many marine habitats that have undergone rapid evolution and speciation together with numerous gene losses (Holland, 2016). Tunicates have been classically divided into three classes (Fig. I4): Appendicularia (a.k.a. Larvacea), Ascidiacea, and Thaliacea. Appendicularians are free-living planktonic organisms that maintain the larval morphology and the chordate synapomorphies throughout the entire life cycle, and comprise 70 species distributed in three families (Oikopleuridae, Fritillariidae, and Kowalevskiidae) (Christen and Braconnot, 1998). The Ascidians (commonly known as sea squirts) have free-living larvae with a typical chordate body plan, but in contrast to appendicularians, at the end of the juvenile phase suffer a drastic metamorphosis that transforms them into sessile adults attached to a substrate and loose most of the chordate synapomorphies (i.e. postanal tail, notochord, dorsal nerve cord, endostyle, and gill slits). Ascidians have gone success radiation and they account for more than 3000 species. The Thaliacea class is composed of approximately 100 species with a free-living planktonic lifestyle. Recent phylogenomic analyses classify Thaliacea within ascidians as the sister group of Phlebobranchia, which transforms Ascidiacea into a paraphyletic group (Delsuc et al., 2018). The discovery that tunicates are the sister group of vertebrates, and therefore that the cephalochordate branching is basal within chordates (Bourlat et al., 2006; Delsuc et al., 2006), has provided a novel view of the last common ancestor of chordates as a free-living organism similar to cephalochordates or appendicularians (Fig. I4). This idea contrast with the traditional view proposed by Garstang (1928) in which the chordate ancestor had a sessile ascidian-like adult lifestyle (Garstang, 1928). According to this novel scenario, free-living appendicularians might represent the ancestral condition of tunicates (Berrill, 1950; Braun et al., 2020; Swalla et al., 2000). However, theories supporting an ascidian sessile tunicate ancestor have also been proposed (Stach, 2007; Stach et al., 2008; Stach and Turbeville, 2002). Therefore, whether the lifestyle of the last common ancestor of tunicates was sessile like in ascidians or free-living as in appendicularians remains unsolved (Fig. I4).

Despite tunicates, in general, are characterized by their "liberal" pattern of genome evolution, appendicularians, as

revealed by the sequencing of *O. dioica*, has likely undergone the most drastic alteration of genome-organization features accompanied by genome compaction and prevalent gene losses (Denoeud et al., 2010). These characteristics have prompted us to develop *O. dioica* as a model in the field of EvoDevo to decipher the mechanisms that lead to gene loss, as well as, the impact of gene losses on the evolution of mechanisms of embryo development in chordates and their impact on the divergence of phyla. Moreover, *O. dioica* has a simple and accessible morphology, a short generation time and life span, affordable culture in the laboratory, and amenable experimental manipulation (Albalat and Cañestro, 2016; Ferrández-Roldán et al., 2019).

2.1 *O. DIOICA* HAS A COMPACT GENOME ACCOMPANIED BY MASSIVE GENE LOSSES

2.1.1 GENOME COMPACTION AND MALLEABILITY IN *O. DIOICA*

With only 70 Mb, the genome of *O. dioica* is the smallest chordate genome and one of the smallest genomes in chordates (Danks et al., 2013; Denoeud et al., 2010; Seo et al., 2001). The small size of the genome of *O. dioica* is the result of a vast compaction process (Berna and Alvarez-Valin, 2014; Chavali et al., 2011; Denoeud
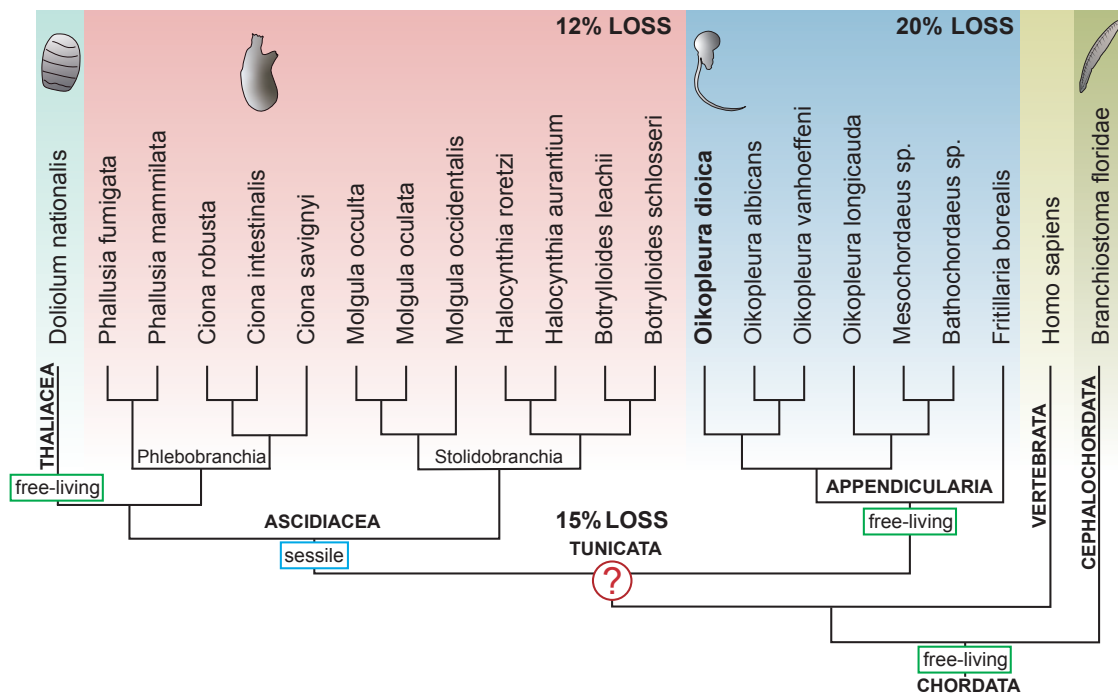
Fig. I4. Phylogenetic relationship between different members of the chordate phylum. Appendicularians, ascidians, and Thaliacea group inside the Tunicata subphylum, which is the sister group of vertebrates. Therefore, cefalochordates branching is basal inside the chordate phylum. The fact that ascidians are sessile and appendicularians are free-living cast doubt on how was the ancestor of tunicates. Tunicates have suffered massive gene losses with the loss of 15% of ancestral gene families. Ascidians have lost an additional 12%, while appendicularians have lost an additional 20% becoming the chordate group in which gene loss has had the most significant impact.

et al., 2010). One of the reasons for this compaction is the reduction of the intergenic and intron regions. Concerning the intergenic regions, 53% of these sequences are smaller than 1 kb, and 27% of the genes lack intergenic regions as they enclose in 1800 polycistronic transcription units or operons (Denoeud et al., 2010). In the case of the introns, 62% of the intronic regions are smaller than 50 nucleotides, and only 2.4% are larger than 1 kb (Denoeud et al., 2010; Seo et al., 2001). Transposable elements or transposons have also decreased their quantity and diversity (Cañestro and Albalat, 2012; Chalopin et al., 2015; Denoeud et al., 2010; Volff et al., 2004) contributing, together with the intragenic and intron reduction, to the compaction of *O. dioica* genome.

Besides being highly compacted, the genome of *O. dioica* is an extreme case of animal genome malleability, with negligible synteny to other animal genomes and the disintegration of the *Hox* cluster, generally conserved in all bilaterians (**Fig. I5**) (Cañestro et al., 2007; Denoeud et al., 2010; Seo et al., 2004). All these chromosomal rearrangements and genomic compaction may be the consequence of the high evolutionary rate of this species, which is supported by a high population mutation rate and a reduced purifying selection (Denoeud et al., 2010; Edvardsen et al., 2005).
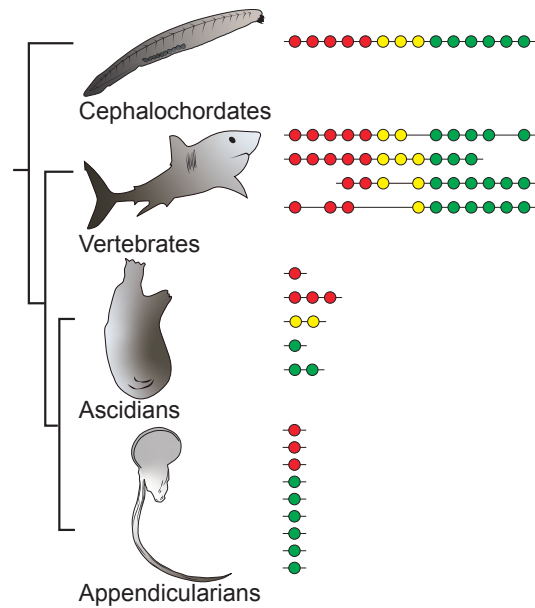
Fig. I5. **Evolution of the *Hox* cluster in chordates**. Cephalochordates contain a copy of 14 *Hox* genes grouped in a unique cluster. In vertebrates, the *Hox* cluster was multiplied and some paralogs were lost in some clusters. The *Hox* cluster in tunicates has been dramatically degenerated and ultimately disappeared in appendicularians, which besides, have lost all the central *Hox* genes.

2.1.2 GENE LOSSES IN O. DIOICA

Among tunicates, *O. dioica* appears to have pushed gene losses to its limits. This species has lost many genes related to genetic structure and expression, cellular and physiological functions, and embryonic development (Ferrández-Roldán et al., 2019; Ferrier, 2011).

Regarding the losses associated with cellular and physiological functions, Denoeud et al., 2010 identified the lack of the non-homologous end joining repair system that is in charge of fixing double-strand breaks of the DNA in all eukaryotes. The epigenetic machinery has also

encountered the effects of gene loss and lacks histone acetyltransferases, many core components of the canonical Polycomb complexes, genes for several proteins of the trithorax group, and some methyltransferases (Albalat et al., 2012; Cañestro et al., 2007; Navratilova et al., 2017). Finally, the minor spliceosome, in charge of removing introns with noncanonical AT-AC boundaries across most eukaryotic lineages, is missing in *O. dioica* (Denoeud et al., 2010). Overall, these losses might have contributed to increasing the malleability of the *O. dioica* genome by accumulating rearrangements and changing the genome architecture.

Many losses of genes related to cellular and physiological functions have also been reported in *O. dioica*. For example, its immune system has suffered and extreme simplification. It has lost almost all genes with domains corresponding to typical immune receptors or effectors and lacks interleukins and cytokines involved in the immunity response in other chordates (Denoeud et al., 2010). Caspases have a crucial role in the maturation of immune system proteins as well as in apoptosis. The caspase family has also experienced a significant reduction with the presence of only three members compared to more than ten present in other chordates (Weill et al., 2005). Not only the immune machinery for the elimination of invaders has suffered from gene losses, but also the machinery

for removing toxic compounds. Cytochrome P450 genes, which participate in the elimination of xenobiotics (Goldstone et al., 2006), show the smallest repertoire among sequenced metazoan genomes in *O. dioica* (Yadetie et al., 2012). Moreover, *O. dioica* is so far the only aerobic non-parasitic organism without peroxisomes, organelles in charge of the elimination of toxic compounds produced in the cells (Žárský and Tachezy, 2015). The loss of the machinery to remove invaders or toxic substances has accompanied the loss of proteins in charge of intracellular traffic amongst organelles. Thereby, *O. dioica* has lost 50% of the Rab toolkit, being the metazoan species with the smallest number of Rab subfamilies described to date (Coppola et al., 2019).

Gene losses have also affected essential pathways for chordate embryonic development in *O. dioica*. Homeobox and Sox families, encoding transcription factors related to development, have lost many of their families (Edvardsen et al., 2005; Heenan et al., 2016; Torres-Águila et al., 2018). Noteworthy, *O. dioica* has lost *Hox3* and all the central *Hox* genes (Seo et al., 2004), and the remaining *Hox* genes are dispersed on the genome and have lost their cluster organization. Despite the crucial role of retinoic acid in chordate axial patterning, *O. dioica* has lost most of the genes for retinoic acid production, degradation, and signaling becoming the

first and only chordate able to develop without this morphogen (Cañestro et al., 2007; Cañestro and Postlethwait, 2007; Martí-Solans et al., 2016). miRNAs have also suffered from gene loss in *O. dioica* with the absence of 18 highly conserved families (Fu et al., 2008; Wang et al., 2017).

Despite the vast amount of gene losses found in *O. dioica*, the genome of this species contains 18,020 predicted genes a similar number to other tunicates (e.g., *C. robusta* ≈15,300 genes) and only slightly below other chordates (e.g., the cephalochordate *B. floridae* ≈22,000 genes; the vertebrate *Fugu rubripes* ≈18,300 genes). This astonishing number of genes is the result of a balance between gene losses and gene gains because the genome of *O. dioica* is prone to lose genes, but also to gain them.

2.2 *O. DIOICA*, FROM A DETERMINATIVE DEVELOPMENT TO A FREE-LIVING LIFESTYLE

O. dioica is a semi-cosmopolite species that can be found in almost all marine coasts. *O. dioica* and other appendicularian species are the second most abundant species, after copepods, in marine mesozooplankton (Capitanio et al., 2008; Gorsky and Fenaux, 1998). Appendicularians have a crucial ecological relevance because they graze about ten percent of the ocean's primary production (Acuna et al., 2002), they serve as sustenance for fish larvae, and they significantly contribute to vertical carbon transport (Davoll and Youngbluth, 1990; Robison et al., 2005; Troedsson et al., 2013).

Although being so small, *O. dioica* has a simple but typical chordate body plan, which includes: a notochord anchoring muscle cells along the post-anal tail, a hollow neural tube which becomes in the central nervous system, a pair of gill slits, and an endostyle. Those chordate synapomorphies remain throughout the whole life cycle because *O. dioica* suffers only a mild metamorphosis in contrast to the drastic metamorphosis that undergoes in ascidian tunicates.

2.2.1 EMBRYO DETERMINATIVE DEVELOPMENT

The determinative development of *O. dioica* is speedy, and at 19ºC only takes 10 hours from fertilization to a juvenile. As *O. dioica* is the only dioecious species so far reported in the tunicate subphylum (Fol, 1872; Nishida, 2008), the development begins after the external fertilization between the sperm released by males and the eggs spawned by females. Those gametes are produced in the gonads during the last hours of *O. dioica*'s life (Ganot et al., 2007; Nishino and Morisawa, 1998). Both spermatogenesis and oogenesis, have a syncytial nuclear proliferation phase. During
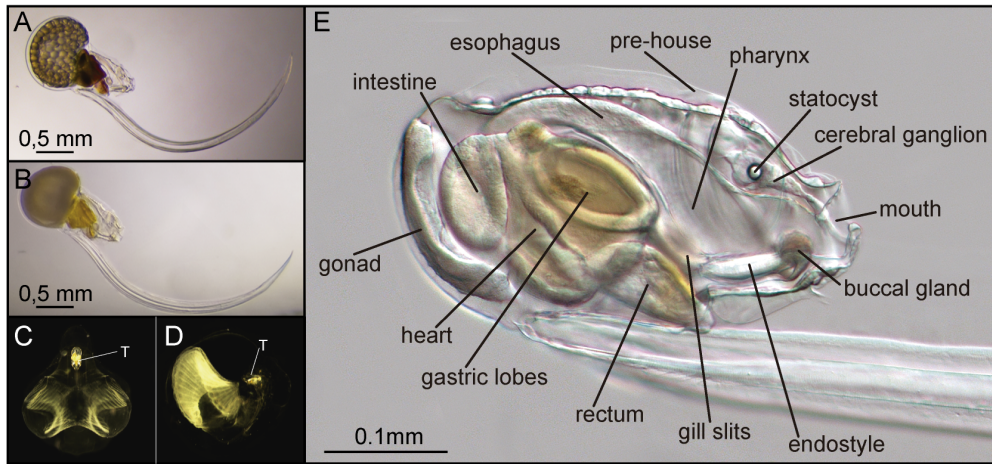
Fig. I6. Mature female of *O. dioica*, the gonad is full of eggs (**A**). Mature male of *O. dioica* the gonad is full of sperm (**B**). Dorsal (**C**) and lateral (**D**) views of a house of *O. dioica*. T, trunk. Immature adult specimen of *O. dioica* showing the main organs (**E**). Differential interference contrast (DIC) micrographs show the right lateral view of the animal with anterior to the left and ventral down.

male gonad development, all the nuclei are identical and become individual spermatozoa, whereas in the ovary, half of the syncytial nuclei become polyploid nurse nuclei, and the other half enter meiosis, surround by actin filaments, and begin to grow until they differentiate into mature eggs (Ganot et al., 2007; Martinucci et al., 2005; Onuma et al., 2017) (Fig. I6 A, B).

Following egg fertilization, the determinative development of *O. dioica* has been well described (Fujii et al., 2008; Stach et al., 2008). The first division takes place after 20 minutes. At 60 minutes post-fertilization, the 32-cell stage embryo gastrulates, which consists of the ingression of the vegetal blastomeres (Fujii et al., 2008; Stach et al., 2008). Neurulation starts when the embryo arrives at the 64-cell stage 80 minutes post-fertilization (mpf). Throughout neurulation, eight cells from the anterior

region align in two rows of four cells, forming a matrix that internalizes. Tailbud starts 135 mpf when the tail and the trunk start becoming distinguishable. During the tailbud stage until the hatchling of the larva, at 3.6 hours post-fertilization, the tail elongates, and the embryo bents ventrally (Ferrández-Roldán et al., 2019; Fujii et al., 2008; Stach et al., 2008) (Fig. I7).

Larval development lasts 6 hours at 19°C. During the first hours, the tail starts moving, and the organ primordia begin to develop. Towards the end of development, the water starts flowing inside the digestive tract of the larva, and eventually, the larva suffers a mild metamorphosis called tailshift. This process only consists of a change of orientation of the tail 120°, which completes the development of the embryo and begins the juvenile phase (Delsman, 1910; R

13

Fenaux, 1998; Galt and Fenaux, 1990; Nishida, 2008) (**Fig. I7**)

2.2.2 FROM JUVENILE TO ADULT FREE-LIVING PHASE

After tailshift, juveniles start soon to inflate the first house and feeding. The house is a filter-feeding device made of polysaccharides, proteins, and cellulose that traps unicellular algae and guides them into the mouth (**Fig. I6 C, D**) (Hosp et al., 2012; Kimura et al., 2001; Spada et al., 2001; Thompson et al., 2001). The house is synthesized by the epidermis of the trunk, known as oikoblast or oikoplastic epithelium, which contains a fixed number of cells organized in different domains defined by the shape of the cells and the

morphology of the nuclei. Each of these domains shows a bilateral symmetric pattern and correlate with the different structures of the house (Kishi et al., 2017; Mikhaleva et al., 2018; Spada et al., 2001; Thompson et al., 2001). After 4-5 hours of filtering material, the house gets obstructed, *O. dioica* abandons it and, immediately after, inflates a new one.

While *O. dioica* filtrates water, the food particles captured by the house arrive at the mouth, they enter the pharynx thanks to the water flux generated by the spiracles, a pair of gill slits with beating cilia (R. Fenaux, 1998). In the pharynx, the mucous secreted by the endostyle, homologous to the thyroid of vertebrates, traps the food particles and conducts them to the
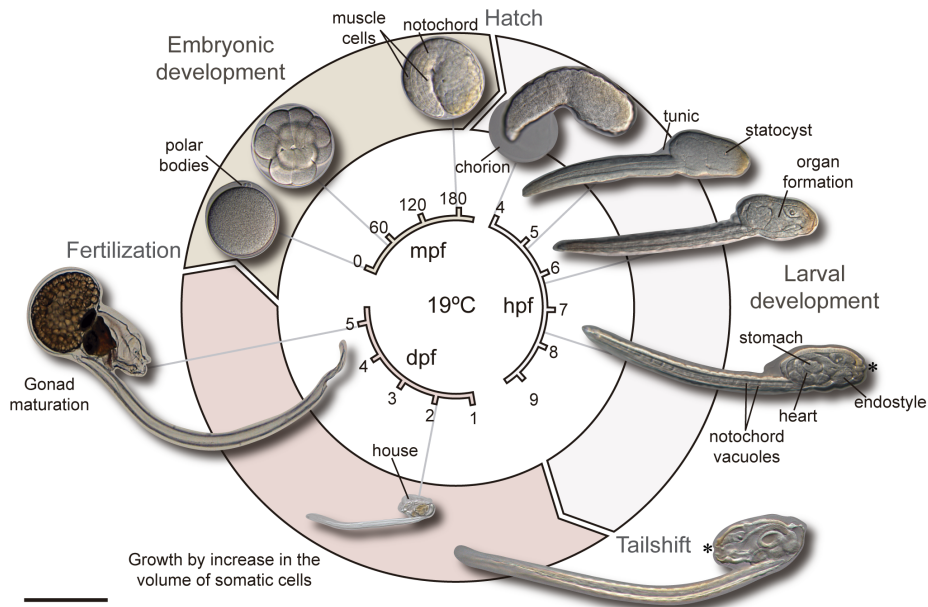
Fig. I7. Schematic representation of *O. dioica* life cycle. Embryonic development starts with the fertilization of the oocyte and ends when a larva hatches breaking the chorion (3.6 hpf at 19°C). Larval development lasts 6 hours (at 19°C) and ends when the tail of the larva changes its orientation 120˚ in a process called tailshift, comparable to metamorphosis in other tunicates. Notice that in the juvenile animal, the position of the mouth (asterisk) relative to the tail has changed. During the next 4.5 days (at 19°C), juvenile animals grow and become mature males or females to close the life cycle. Scale bar represents 100 μm for all stages except for day 2 and mature adults (1 mm).

digestive tract (Cañestro et al., 2008; Olsson, 1963). This tract comprises the esophagus, two stomach lobes, and the intestine which digest the food and absorb the nutrients (Burighel et al., 2001; Cima et al., 2002) (Fig. I6 E).

The movement of the food particles through the house to the mouth, as well as the process of inflating a new house, require a continuous water flow inside the house. This water current is created by tail beating, which are also used for locomotion, and demands a huge amount of energy. The tail is stiff and flexible thanks to the notochord, and its movements are driven by two strips of ten striated muscle cells on dorsal and ventral sides of the notochord (Nishino et al., 2000; Nishino and Satoh, 2001; Olsson, 1965; Soviknes et al., 2007). Muscle cells are innervated by small ganglia distributed along the nerve cord that extends through the left side of the tail. This nerve cord connects to the caudal ganglion at the anterior region of the tail, and continues into the trunk until the cerebral ganglion, that with only 70 neurons, is the most complex part of the central nervous system, and processes sensory stimuli provided by the statocyst, the mouth, the pharynx, and the ventral organ (Bollner et al., 1986; Cañestro et al., 2005; Holmberg, 1984). The movement of the tail has also been described to participate in hemolymph circulation helping the only muscular structure present

in the trunk of *O. dioica*, the heart (Fig. I8 A).

2.3 THE HEART OF *O. DIOICA*

At the beginning of this thesis, the knowledge about the heart of appendicularians was scarce and restricted to some descriptions published in the early 20th century. In these descriptions, Salensky represented the heart as an organ constituting of two closed pouches lying behind the posterior wall of the stomach (Salensky. W, 1903; Salensky, 1904). During my thesis, a modern scanning
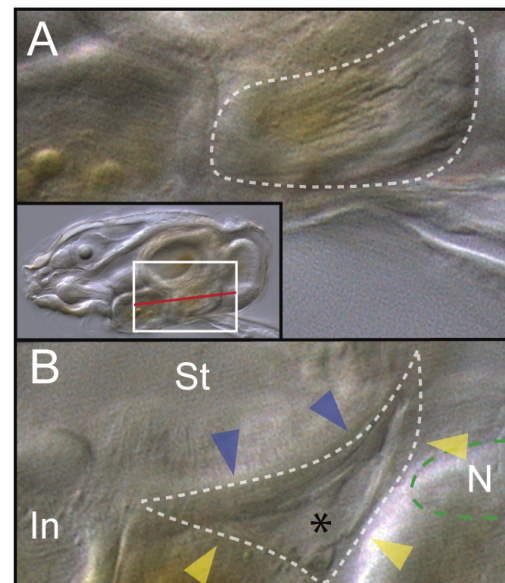
Fig. 8. **Anatomy of the adult heart of *O. dioica*.** Left DIC micrography of the trunk of *O. dioica*. The white square indicates the amplificated region of the trunk in which the striated muscle of the heart is showed (**A**). Ventral image of the trunk showing the bilaminar structure of the heart (**B**). The pericardium (blue arrows) beats against the stomach wall. The myocardium (yellow arrows) protects the myocardium. Discontinuous white lines delimit the shape of the heart during organogenesis. Red line indicate the section amplified in B. Discontinuous green lines delimit the notochord. Asterisk indicates the lumen of the heart. Intestine (In), Notochord (N), Stomach (St).

microscopy study has been published describing the morphology of the body of adult specimens of *O. dioica*, including the heart (Onuma et al., 2017) and corroborating the original bilaminar description made by Salensky.

The heart of *O. dioica* consists of a posterior ventral bag wedged between the left stomach wall and the intestine. This bag is lined by two types of flat mesodermal tissues, the myocardium and the pericardium (**Fig. I8 B**). The myocardium includes striated muscle cells connected by cell junctions with the cytoplasm full of actin filaments (Onuma et al., 2017). The pericardium is not muscular and protects the myocardium. The heart of *O. dioica* produces peristaltic contractions that periodically reverse, which together with tail movements, contribute to the circulation of the hemolymph between the ectoderm and the internal organs in an open circulatory system (R. Fenaux, 1998). The heart of *O. dioica* adults*,* therefore*,* can be considered the simplest heart described on a chordate (Ferrández-Roldán et al.*,* 2019). The development of the heart in *O. dioica* or any other appendicularian species remains totally unknown, and as a case study, it will become the main subject of this PhD project.

3 EVOLUTIONARY CONSERVATION OF THE CARDIAC GENETIC NETWORK

The presence of hearts is widespread throughout the animal kingdom. Hearts are often characterized by being a muscle-powered pumping organ consisting of various cell types that work together to propel fluid inside the animal facilitating the uniform access to oxygen and nutrients for all the cells in the organism, as well as the elimination of carbon dioxide and excreta. Throughout the evolution, hearts have developed an impressive array of morphologies, from simple vessels that propel fluid by peristaltic contractions to multi-chambered architectures that coordinately beat to efficiently propel the blood circulation throughout the body (Xavier-Neto et al., 2007).

A central question in the study of the evolution of hearts is if all these cardiac structures are homologous evolving from a common organ present in the last common ancestor of metazoans, or on the contrary if hearts are analogous and they have independently evolved in different lineages although converging in similar structures. The fact that hearts appear in members of all the major bilaterian groups (Deuterostomes, Ecdysozoans and Lophotrochozoans) and all of them share a surprising number of genetic circuits, supports a common origin of the circulatory

pumps. However, the similarities in morphology and functionality in the hearts of very distantly related animals such as the chambered hearts of vertebrates and mollusks, or the pulsating peristaltic vessels of annelids and cephalochordates suggest that, although having a similar design, hearts were independently created to perform analogous essential functions (Xavier-Neto et al., 2007). These disparate views were integrated on a new hypothesis that proposes that all hearts derive from parallel improvements of an ancestral, peristaltic design represented by a layer of myocytes at the external walls of primitive vessels (Bishopric, 2005; Xavier-Neto et al., 2007) (Fig. I9).

The assumption that all the hearts derive from an ancestral layer of myocytes present in the common ancestor of protostomes and deuterostomes agrees with the existence of a shared cardiac regulatory genetic kernel. This handful of orthologs – *Gata4/5/6*, *Nk4*, *Hand1/2*, *Tbx*– control cardiac cell fates, the expression of genes encoding contractile proteins, and the morphogenesis of cardiac structures in all studied metazoans (Davidson and Erwin, 2006; Olson, 2006; Poelmann and Gittenberger-de Groot, 2019; Wijesena et al., 2017) (Fig. I9). Cardiogenesis starts with the transcription of *Gata4/5/6* and *Nkx4* orthologs in the mesodermal precardiac tissue. This tissue is simultaneously prompted by different inductive signals as

	NK4	GATA	Tbx	Hand		
Vertebrates	≥2	3	≥7	2		
Urochordates	1	1	≥2	1		
Cephalochordates	1	2	?	1		
Echinoderms						
Hemichordates						
Tardigrades						
Onychophorans						
Arthropods	1	2	3	1		
Nematodes						
Sipunculans						
Molluscs						
Anelids						
Platyhelminthes						
Cnidarians						
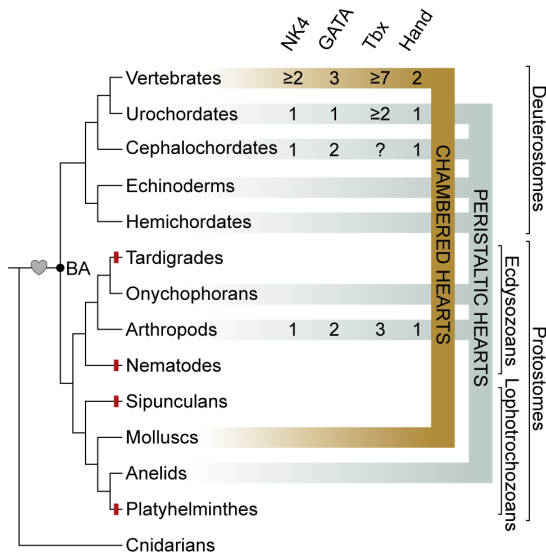
Fig. I9. **Origin and evolution of the hearts**. The new hypothesis of the origin of the heart suggests that the bilaterian ancestor already had a simple heart that evolved parallelly in the different bilaterian lineages. The numbers of cardiac kernel genes known to be expressed in the cardiac structures of each organism are shown. Red lines indicate the loss of the heart in that lineage. BA; bilaterian ancestor. Modified from Xavier-Neto 2007.

retinoic acid (RA), fibroblast growth factor (FGF), Wnt, Hedgehog, bone morphogenetic proteins (BMP), or Notch, with different roles in different species. Then, this genetic environment regulates the activation of *Tbx* and *Hand* orthologs, followed by various sets of patterning and muscle genes. (Olson, 2006) (Fig. I10). Although all metazoans share the same genetic kernel during the first steps of cardiac development, differences in pattern and timing of gene expression and co-option of upstream and downstream regulators have been the major forces responsible for the many different heart designs (Diogo et al., 2015).

3.1 CARDIOPHARYNGEAL DEVELOPMENT IN CHORDATES

During chordate evolution, the duplication of many cardiac kernel genes and the co-option of new transcription factors have favored cardiac genetic complexity. This genetic complexity has been translated into the emergence of the branchiomeric or pharyngeal muscles from the same mesodermal progenitors that give rise to the heart. Therefore, in chordates, cardiac progenitors are included in a larger mesodermal population called cardiopharyngeal progenitors, which contributes to both, the heart and the skeletal muscles of the head and neck (Diogo et al., 2015).

3.1.1 VERTEBRATES

Vertebrates are the only animal group, together with mollusks, that develop a heart with different chambers. This distinctive characteristic was used by Haeckel in 1866 to designate vertebrates as the
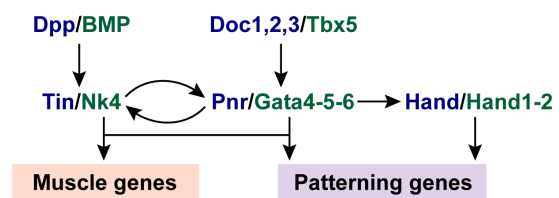
Fig. I10. Schematic representation of the conserved heart regulatory kernel. A core set of regulatory orthologs participate in the induction of the cardiac tissue in evolutionarily distant organisms. The interactions between them are also highly conserved. In blue orthologs of *D. melanogaster*, in green orthologs of chordates.

pachycardia, referring to them as the group of animals with a skull and a thick heart. The cardiopharyngeal field in vertebrates includes two adjacent mesodermal populations the first heart field (FHF) and the second heart field (SHF). The lateral anterior splanchnic mesoderm generates the FHF that produces the linear heart tube in the embryo and the ventricle and parts of the atria in the adult. After the formation of the linear heart tube, the pharyngeal mesoderm adds SHF precursors that generate the myocardium from the outflow tract, the right ventricle, and parts of the atria (Kelly and Evans, 2010). SHF precursors share a clonal relationship with branchiomeric progenitors that give rise to the craniofacial skeletal muscles, a group of muscles that were crucial for evolutionary innovations in vertebrates as the emergence of the tetrapod neck (Diogo et al., 2015; Lescroart et al., 2010).

3.1.2 CEPHALOCHORDATES

Cephalochordates have a close circulatory system that proper the hemolymph with the help of peristaltic pulsating vessels surrounded by smooth muscle (Hirakow and Kajita, 1994; Xavier-Neto et al., 2010). In contrast to vertebrates, amphioxus presents a decentralized cardiac domain likely correlated to the absence of a central heart (Holland et al., 2003; Panopoulou et al., 1998; Pascual-Anaya et al., 2013). However, although lacking a central heart,

amphioxus pterygial muscles seem to be homologs to the vertebrate branchiomeric muscles, suggesting that an ancestral configuration of these muscles were already present in the common ancestor of the extant chordates (Sambasivan et al., 2011).

3.2 ASCIDIANS, A MODEL TO STUDY THE BASES OF CHORDATE CARDIAC DEVELOPMENT

For more than 20 years, the tunicate *Ciona robusta* has emerged as a simple model to study the basis of cardiopharyngeal development in chordates. Its low number of cells, its remarkably reduce gene redundancy, and the ability to accurately manipulate its genetic components provide high spatiotemporal resolution throughout development (Beh et al., 2007; Christiaen et al., 2010; Davidson, 2007; Wang et al., 2019). Studies using *C. robusta* have profoundly increased our understanding of the origin and evolution of the heart and the branchiomeric muscles by revealing interactions between genes, and the activity of regulatory elements.

In *C. robusta*, the heart starts beating just after metamorphosis, 2 or 3 days after fertilization. However, the heart requires two more weeks to acquire its fully functional adult morphology (Davidson, 2007). The *Ciona* adult heart is located just posterior to the pharynx and anterior to the stomach and consists of a valve-less, U-

shaped myocardial tube encased in a pericardial sac (**Fig. I11A, B**) (Davidson, 2007). The raphe, a dense extracellular matrix, connects the one cell layer thick myocardium and pericardium. The heart of *Ciona* alternates pump direction, driving the fluid through two large vessels that are attached to both its ends and branch out throughout the body of the adult animal (Davidson, 2007).

3.2.1 MORPHOGENESIS OF THE CARDIOPHARYNGEAL LINEAGE IN CIONA ROBUSTA

Cardiopharyngeal progenitors develop during embryonic and larval stages, before the profound metamorphosis that makes *C. robusta* to lose the chordate synapomorphies and give rise to the adult. The B7.5 pair of blastomeres constitute the cardiopharyngeal precursors in the 64-cell stage embryo (Davidson, 2007; Razy-Krajka and Stolfi, 2019). Those blastomeres undergo two rounds of mitosis, resulting in two clusters of 4 cells at the border between the future trunk and tail regions. As the tail extends, the four posterior daughters separate, becoming into the anterior tail muscle cells (ATMs), and the anterior ones give rise to the trunk ventral cells (TVCs) (Davidson, 2007). Throughout the tailbud stage, the right and left TVCs migrate as pairs until they meet in the ventral midline of the trunk. Then, cardiopharyngeal progenitors divide and

generate the first heart precursors (FHP) and the second heart precursors (SHP) that contribute to the heart, and the atrial siphon muscle field (ASMF) or pharyngeal precursors, which are homologous to the branchiomeric muscles of vertebrates and gives rise to the atrial siphon muscles and longitudinal muscles of the body wall in the adult (Hirano and Nishida, 1997; Stolfi et al., 2010) (**Fig. I11 C**). After dividing, cardiopharyngeal precursors arrest their development until metamorphosis, when they complete both cardiac and pharyngeal muscle development.

3.2.2 THE CARDIOPHARYNGEAL GENETIC PATHWAY IN CIONA

The cardiopharyngeal gene regulatory network (GRN) in *Ciona* can be divided into three parts: (1) induction of the cardiopharyngeal lineage, (2) migration of cardiopharyngeal precursors, and (3) specification of cardiac and pharyngeal precursors.

3.2.2.1 INDUCTION OF THE CARDIOPHARYNGEAL LINEAGE

The induction of the cardiopharyngeal lineage in ascidians occurs early in development when B7.5 blastomeres, in response to the maternal effectors macho-1 and ß-catenin, activate *Tbx6* and *Lhx3*, respectively (Christiaen et al., 2010). The simultaneous expressions of these transcription factors upregulate *Mesp,* a bHLH transcription factor that acts as a

master regulator for heart development in olfactores (Davidson et al., 2005; Saga et al., 2000; Satou et al., 2004). Although *Mesp* is expressed in the cardiac progenitors of vertebrates, its specific role is unclear (Devine et al., 2014; Kitajima et al., 2000; Lescroart et al., 2014; Saga et al., 2000). In *C. robusta,* however, *Mesp* is known to upregulate *Ets1/2*, a transcriptional effector of the FGF-MAPK pathway (Christiaen et al., 2010; Davidson et al., 2006).

Fibroblast growth factor (FGF) and retinoic acid (RA) signaling pathways have a crucial role in the induction of the chordate cardiopharyngeal development (Krieg and Warkman, 2015; Perl and Waxman, 2019; Reifers et al., 2000). In vertebrates, RA plays an essential role in delineating the posterior border of the emerging heart field (Simões-Costa et al., 2005). This function is conserved in ascidians where RALDH2 – enzyme that catalyzes the synthesis of RA– is expressed in the progeny of B7.5. However, the RALDH2 expression is only maintained in the prospective tail muscle cells (Nagatomo and Fujiwara, 2003). The precursors destined to be TVCs downregulate RALDH2 expression (Bernard et al., 1998; Simões-Costa et al., 2005), while activating the FGF-MAPK signaling pathway. FGF-MAPK signaling pathway phosphorylates Ets, previously transcribed as a result of *Mesp* activation. Finally, phosphorylated Ets (pEts) triggers
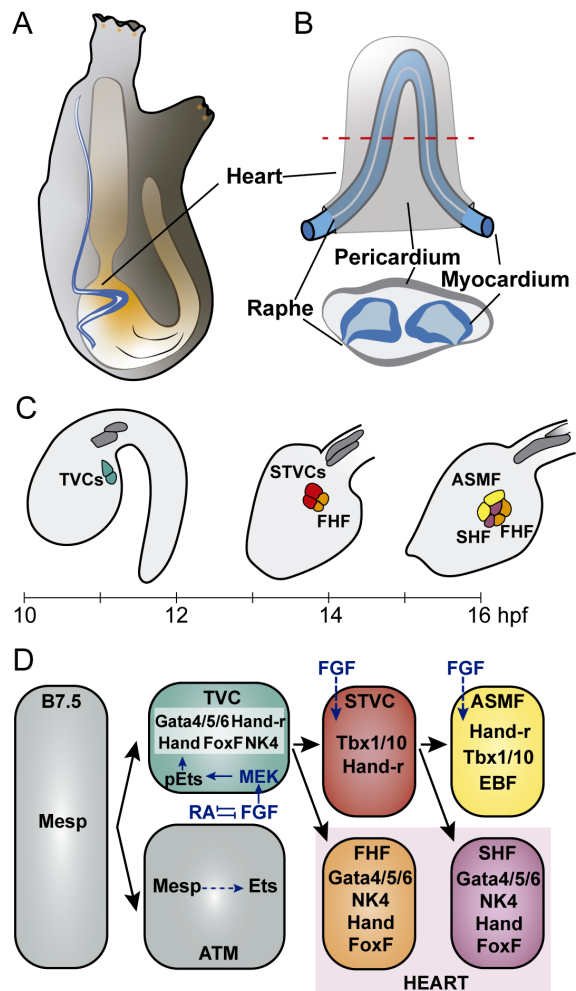
Fig. I11. **Heart development in *C. robusta*.** Scheme of a *C. robusta* adult, the heart is represented in blue on the left side of the organism (**A**). Cartoon and cross-sectional diagram of the U-shaped *Ciona* heart. The myocardium, in blue, is protected by the pericardium, in grey. The raphe connects myocardium and pericardium (**B**). Early cardiopharyngeal development in *Ciona* from 10 to 16 hpf. Cardiopharyngeal-lineage cells are shown for only one side and different precursors are differentially colored. Green, Trunk ventral cells (TVCs); Red, Second trunk ventral cells (STVCs); Orange, First heart field (FHF); Yellow, Atrial siphon muscle field (ASMF); Purple, Second heart field (SHF) (**C**). Cardiopharyngeal-linage cells are shown for only one side. Cell-type-specific marker genes are indicated inside each cell, as well as the inductive signals (**D**).

the expression of the heart regulatory kernel and migratory factors (**Fig. I11D**).

3.2.2.2 MIGRATION OF THE CARDIOPHARYNGEAL LINEAGE

Downstream of the FGF/MAPK signaling, pEts upregulates *FoxF* and *Ddr*, which, simultaneously, activates other cardiac transcription factors (Beh et al., 2007; Christiaen et al., 2008). *FoxF* is required for FGF-induced cardiac cell migration, but not for heart muscle specification (Beh et al., 2007). TVCs migrate as a polarized pair of cells with one anterior or leader and one posterior or trailer. This distinct polarized state of the TVCs is established by *Ddr*, which also maintains their adhesion to the cell-matrix (Bernadskaya et al., 2019).

During migration, TVCs are thought to follow a localized cue of BMP2/4 ligand originated in the trunk ventral epidermis (Christiaen et al., 2010) that also participates in the induction of *Gata4/5/6* and *Nk4*, genes from the heart regulatory kernel. *Gata4/5/6* is upregulated at the beginning of the migration when TVCs are far from the ventral epidermis, and the levels of BMP2/4 ligand are low. Gata4/5/6 is supposed to maintain *FoxF* expression during migration, and together with high levels of BMP2/4 ligand, Gata4/5/6 upregulates *Nk4* at the end of the migration (Christiaen et al., 2010; Tolkin and Christiaen, 2012; Wang et al., 2013) (**Fig. I11 D**).

3.2.2.3 SPECIFICATION OF CARDIOPHARYNGEAL LINEAGE

Following the migration, TVCs divide asymmetrically to create secondary trunk ventral cells (STVCs) and first heart precursors (FHPs). STVCs maintain *Nk4* expression and the FGF-MAPK signaling pathway activated, promoting the expression of *Hand-r*, *Tbx1/10*, and *Ebf* (Wang et al., 2019). By contrast, FGF-MAPK signaling stops in FHPs, where *Nk4* reactivates *Gata4/5/6* expression and upregulates *Hand* and other cardiac-specific genes (Wang et al., 2019). Subsequent division of the STVCs generates the second heart precursors (SHPs) and the atrial siphon muscle founder cells (ASMFs). The antagonism *Nk4* versus *Tbx1/10* and the activation of the FGF pathway govern the distinction between cardiac and pharyngeal muscle (Wang et al., 2019, 2013). In SHPs both, *Tbx1/10* inactivation by *Nk4*, the onset of *Dach* and the cease of the FGF-MAPK signaling allow the expression of cardiac-specific genes. In contrast, in ASMFs, the FGF-MAPK signaling and Tbx1/10 keep activated, allowing the expression of muscle-specific genes as *MyoD*, *Ebf* or *Islet1* (Wang et al., 2019) (**Fig. I11 D**).

In this thesis, we will analyze the impact of gene loss in the cardiopharyngeal GRN in the tunicate *O. dioica*, and the results will be discussed on the context of the

evolution of free-living lifestyle of appendicularians.

O BJECTIVES

The overall aim of this PhD project was to better understand the impact of gene loss in the evolution of the mechanisms of development, particularly in the context of chordate evolution, which is the main area of interest of the lab where this project has taken place. In particular, this project aimed to study the impact of the loss of developmental genes on the evolution of the free-living appendicularians in contrast to other sessile tunicates such as ascidians. Among appendicularians, we aim to develop the study of *Oikopleura dioica* as an eKO model among chordates, in particular considering previous results from our laboratory showing that *O. dioica* appears as the only chordate so far known to have lost the RA signaling pathway. Thus, considering the fundamental role of RA signaling in cardiac development in ascidians and vertebrates, and the relevance of the cardiopharyngeal gene regulatory network in the evolution of free-living styles, the primary focused of this PhD project was to study heart development in appendicularians using *O. dioica* as a RA eKO model system. Second, in line with our interest to develop the study of *O. dioica* as an eKO model, and taking advantage of the uniqueness of *O. dioica* as a RA eKO model, we became interested in a better understanding of the developmental genetic response of chordate embryos to environmental threats, such for instance

diatom bloom-derived biotoxins such as PUAs, which in other organisms have been suggested to affect the development of structures regulated by RA during development, including mesodermal derivatives such muscle and heart. Thus, the specific objectives of the PhD are:

- **Objective 1**: To generate an anatomical atlas of the development of the heart of *O. dioica* and to describe its morphology.

- **Objective 2**: To characterize the origin and map the lineage of the cardiac cells.

- **Objective 3**: To perform a comprehensive genome survey in *O. dioica* and a phylogenetic analysis to characterize its homologs to the components of the cardiopharyngeal GRN in ascidians and vertebrates.

- **Objective 4**: To perform an expression analysis of cardiopharyngeal genes in *O. dioica* to characterize the cardiac expression dynamics during development.

- **Objective 5**: To perform functional analyses by inhibitory treatments or gene knockdown approaches to understand cardiogenic fate cell differentiation.

- **Objective 6**: To further develop the study of *O. dioica* as an eKO model, we investigated genetic responses to environmental threats such as diatom bloom-derived biotoxins in the absence of RA signaling.

SUPERVISORS' REPORT

Thesis supervisors' report about authorship and impact factor of the publications of this doctoral thesis presented by Alfonso Ferrández Roldán

Dr. Cristian Cañestro and Dr. Jordi García-Fernàndez, supervisors of the PhD thesis entitled "**Deconstruction of the cardiopharyngeal gene regulatory network in appendicularians, a paradigmatic study of *Oikopleura dioica* as an evolutionary knockout model**" by Alfonso Ferrández Roldán, certify that the results obtained have been or will be submitted to peer-reviewed international journals. The entire thesis comprises four articles: three of them already published, and one finished and ready to be submitted. None of the articles have been used for the elaboration of other PhD thesis. The details of the papers, journal and their impact factor (Journal Citation Reports) are detailed below:

Publication 1: Developmental atlas of appendicularian *Oikopleura dioica* actins provides new insights into the evolution of the notochord and the cardio-paraxial muscle in chordates. Almazán* A, Ferrández-Roldán* A, Albalat R, Cañestro C. (2019) *Dev Biol.* 448(2):260-270. doi: 10.1016/j.ydbio.2018.09.003
Impact factor (JCR2018): 2.936
Category JCR: Molecular Biology and Genetics
Rank: 31/309; Quartile: Q1

*The first two authors equally contributed. The PhD candidate did the genome survey, identified all actin genes, performed phylogenetic analysis, developed the DAV methodology, participated and co-supervised the first author TFG project on the Actin expression analyses, and contributed to the writing and making of the figures of the manuscript.

Publication 2: Adaptive evolution by modular deconstruction of the cardiopharyngeal gene regulatory network in appendicularians reveals that ancestral tunicates were sessile Ferrández-Roldán A., Fabregà-Torrus M., Sánchez-Serna G., Durán-Bello E., Joaquín-Lluís M., Bujosa P., Plana-Carmona M., Garcia-Fernàndez J., Albalat R., Cañestro C. (Finished and ready for submission)

This manuscript collects the main body of the results of the project of the PhD candidate, in which he has done the genome survey for all the members of the cardiopharyngeal gene regulatory network, performed phylogenetic analysis, inferred gene losses and duplications, performed most of the colorimetric whole mount in situ hybridization, performed the 4D nuclei reconstruction and the cardiopharyngeal cell lineage fate map, performed the BMP inhibitory treatments, co-supervised the contribution of six other co-authors during their TFG or Master projects, and contributed to the writing and making of the figures of the manuscript.

Publication 3: *Oikopleura dioica*: **An Emergent Chordate Model to Study the Impact of Gene Loss on the Evolution of the Mechanisms of Development**. Ferrández-Roldán* A, Martí-Solans* J, Cañestro C, Albalat R. *Results Probl Cell Differ.* (2019) 68:63-105. doi: 10.1007/978-3-030-23459-1_4.
Impact factor (JCR2018): 0.650
Category Scopus: Biochemistry, Genetics and Molecular Biology; Cell Biology
Rank: 204/274; Quartile: Q3

*The first two authors equally contributed. The PhD candidate contributed to design of the review, and contributed to the writing and making of the figures of the manuscript.

Publication 4: Diatom bloom-derived biotoxins cause aberrant development and gene expression in the appendicularian chordate *Oikopleura dioica.* Torres-Águila NP, Martí-Solans J, Ferrández-Roldán A, Almazán A, Roncalli V, D'Aniello S, Romano G, Palumbo A, Albalat R, Cañestro *C. Commun Biol.* (2018) 1:121. doi: 10.1038/s42003-018-0127-2.
Impact factor (JCR2019): 4.165
Category JCR: Biology-Sciences
Rank: 18/93; Quartile: Q1
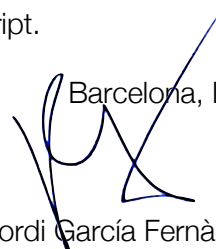
The PhD candidate contributed to identification, cloning and expression analyses of cardiac markers, and contributed to the conceptualization of the use of evolutionary knockout models, and participated in the writing of the manuscript.

Barcelona, December 9th 2020

Cristian Cañestro García   Jordi García Fernàndez

CHAPTER I

DEVELOPMENTAL ATLAS OF APPENDICULARIAN *OIKOPLEURA DIOICA* ACTINS PROVIDES NEW INSIGHTS INTO THE EVOLUTION OF THE NOTOCHORD AND THE CARDIO-PARAXIAL MUSCLE IN CHORDATES

Contents lists available at ScienceDirect

Developmental Biology

CrossMark

Developmental atlas of appendicularian *Oikopleura dioica* actins provides new insights into the evolution of the notochord and the cardio-paraxial muscle in chordates

Alba Almazán[1], Alfonso Ferrández-Roldán[1], Ricard Albalat*, Cristian Cañestro*

Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Catalonia, Spain

A R T I C L E   I N F O

A B S T R A C T

Locomotion by tail beating powered by a system of bilateral paraxial muscle and notochord is likely one of the key evolutionary innovations that facilitated the origin and radiation of chordates. The innovation of paraxial muscle was accompanied by gene duplications in stem chordates that gave rise to muscular actins from cytoplasmic ancestral forms, which acquired contractile capability thanks to the recruitment of the myosin motor-machinery. To better understand the role of actin diversification during the evolution of chordates, in this work we have characterized the complete actin catalogue of the appendicularian *Oikopleura dioica*, an urochordate that maintains a chordate body plan throughout its life, including the notochord in a muscled tail that confers an active free-living pelagic style. Our genomic survey, phylogenetic analyses and Diagnostic-Actin-Values (DAVs) reveal that *O. dioica* has four muscular actins (*ActnM1−4*) and three cytoplasmic actins (*ActnC1−3*), most of which originated by independent gene duplications during the evolution of the appendicularian lineage. Detailed developmental expression atlas of the complete actin catalogue of *O. dioica* reveals differences in the temporal-regulation and tissue-specificity of different actin paralogs, suggesting complex processes of subfunctionalization during the evolution of urochordates. Our results suggest the presence of a "cardio-paraxial" muscular actin at least in the last common ancestor of Olfactores (i.e. vertebrates +urochordates). Our results reveal highly dynamic tissue-specific expression patterns for some cytoplasmic actins, including the notochord, ciliated cells and neurons with axonal projections, which challenge the classic housekeeping notion ascribed to these genes. Considering that previous work had demonstrated the existence of notochord-specific actins in cephalochordates, the tissue-specific expression of two cytoplasmic actins in the notochord of *O. dioica* suggests that this pattern plausibly reflects the ancestral condition of chordates, and provides new insights to better understand the evolutionary origin of the notochord.

1. Introduction

Actins are highly conserved globular proteins, and many of their biological functions derive from their ability to form filaments by linear polymerization (reviewed in Dominguez and Holmes, 2011). Actin microfilaments are one of the major components of the cytoskeleton, being involved in providing structural strength to cellular architecture, ordering and shaping intracellular compartments and organelles, as well as contributing to physically distribute molecules within the cell (Blanchoin et al., 2014; Gunning et al., 2015). Thus, actins have been described to play important roles in cell-cycle control (Heng and Koh, 2010; Tripathi, 1989), chromosome segregation (Ramkumar and Baum, 2016), cell membrane dynamic modulation (Bezanilla et al., 2015; Chang et al., 2017; Fritzsche et al., 2016), regulation of mRNA transport and translation (Besse and Ephrussi, 2008; Vidaki et al., 2017; Zarnack and Feldbrugge, 2010), protein transport and activity modulation (Farwell et al., 1990; Kondrikov et al., 2010; Silva et al., 2016) and regulation of steady-state length of cilia (Avasthi et al., 2014). In addition, the nuclear fraction of actins has been involved in the maintenance of the architecture of the nucleus (Hofmann, 2009), as well as in the regulation of gene transcription either by direct interaction with RNA-polymerases (Philimonenko et al., 2004), by chromatin remodeling (Kapoor et al., 2013), or by direct regulation of the activity of transcription factors (Vartiainen et al., 2007).

One of the most dramatic specializations that cytoskeleton actin filaments have acquired is contractility thanks to the recruitment of myosin II motor proteins (Lodish et al., 2000). Vertebrate striated muscle cells represent the evolution of one of the most sophisticated actomyosin contractile complexes, in which bundles of filaments organized in sarcomeres can contract quickly, repetitively through long distance and with enough force to drive movement (Lodish et al., 2000). The acquisition of contractility by the recruitment of myosin II motors into actin filaments has not only occurred in muscle cells, but other non-muscle cells have also evolved analogous systems (Conti and Adelstein, 2008; Vicente-Manzanares et al., 2009; Zaidel-Bar et al., 2015). For instance, actin filaments and myosin-II assemble into a contractile ring during cytokinesis that provides mechanical forces for cell cleavage at the end of mitosis (Pollard, 2010). Similarly, other ring-like or circumferential belt contractile structures of actomyosin in the inner surface of cell membranes are fundamental for cell-cell adhesion, wound healing and epithelial extrusions (Nakajima and Tanoue, 2012; Schwayer et al., 2016).

Actins are one of the most conserved proteins in eukaryotes, with a 97% conserved sequence identity across distant species such as sea anemone and human. Classically, actins have been classified into muscle actins and cytoplasmic (or non-muscle) actins, depending on whether they were specialized in muscle contraction or, on the contrary, they form part of the cytoskeleton and were involved in many other cellular functions (Hightower and Meagher, 1986; Vandekerckhove and Weber, 1978). In mammals, for instance, from the six major actin genes that have been identified, four are muscle-specific (alpha-skeletal *ACTA1*, alpha-cardiac *ACTC1*, alpha-smooth muscle *ACTA2*, gamma-smooth muscle *ACTG2*), and two are cytoplasmic (beta-cytoplasmic *ACTB* and gamma-cytoplasmic isoactin *ACTG1*) (Vandekerckhove and Weber, 1978). Muscle-specific actin genes have also been described in non-chordate deuterostomes and protostomes (see Hooper and Thuma, 2005 for an extensive review). In *Drosophila melanogaster*, for instance, four out of its six actin genes (*act57A*, *act87E*, *act88F* and *act79B*) are expressed in muscle tissues of different structures during different times of the life cycle (Fyrberg et al., 1983). Recent phylogenetic analysis based on accurate tree-based orthology estimation has shown that despite their conserved muscular expression, chordate and non-chordate muscle actin genes are not orthologous, but they have a polyphyletic origin due to independent duplications from cytoplasmic genes in different lineages (Inoue and Satoh, 2018). This finding is consistent with previous analyses that based on comparisons of diagnostic positions and exon-intron structures had classified muscle-specific genes of non-chordate animals together with the cytoplasmic actin genes (Chiba et al., 2003; Fyrberg et al., 1981; T. Kusakabe et al., 1997; Vandekerckhove and Weber, 1984).

In cephalochordates, in addition to muscle and cytoplasmic actins, a third group has been described as notochord-type actin genes (N-actin) because of their specific expression in the notochord (Suzuki and Satoh, 2000). The notochord of amphioxus has the peculiarity to be formed by muscle fibers, which in 1870, Muller already described as arranged birefringent myofilaments capable of contracting and altering its mechanical properties upon nervous stimulation (Flood et al., 1969; Müller, 1870). Large scale analysis of notochord ESTs revealed indeed that 11% of its cDNAs was related to muscle genes –including N-actin and many other muscle genes such as tropomyosin, troponin I, myosin regulatory light and heavy chains (Suzuki and Satoh, 2000). The study of the actin gene family has become useful to investigate the origin and evolution of the notochord and the paraxial muscle, and thus, to better understand the origin of the fish-like ancestral chordate body plan that probably was characterized by its swimming capability powered by tail beating (Satoh, 2016). In this work, we focus on the characterization of actin gene family in the appendicularian *Oikopleura dioica*. Appendicularians (a.k.a. larvaceans), ascidians and thaliaceans belong to the urochordate phylum (Satoh et al., 2014), which is the sister group of vertebrates, and together with cephalochordates constitute the

chordate superphylum. Appendicularians, in contrast to other urochordates, do not suffer a drastic metamorphosis and maintain their body plan, including an actively motile tail that confers them a free-living pelagic style during their entire life cycle (Mikhaleva et al., 2015; Nishida, 2008; Soviknes et al., 2007; Soviknes and Glover, 2008). Our work reveals how muscle and cytoplasmic actin genes have undergone extensive gene duplications during the evolution of the appendicularian lineage, and reveals cardio-paraxial and notochord-specific expression domains of some muscular and cytoplasmic actin genes that help to better understand the role of actins during the origin and evolution of chordates.

2. Material and methods

2.1. Biological material

O. dioica specimens were obtained from the Mediterranean coast of Barcelona (Catalonia, Spain), and cultured in our animal facility at the University of Barcelona, in which embryos have been collected as described (Marti-Solans et al., 2015).

2.2. Genome survey, phylogenetic analysis, Diagnostic-Actin-Value (DAV) and exon-intron organization

To identify *O. dioica* actin homologs, we made BLAST searches (Altschul et al., 1997) in the genome database of *O. dioica* (Oikobase, http://oikoarrays.biology.uiowa.edu/Oiko) using as queries known actin sequences from ascidians, amphioxus and two cDNAs described in *Oikopleura longicauda* (Nishino et al., 2000). Gene annotations were manually corrected using ESTs and sequences from PCR fragments we have cloned during our analyses.

Phylogenetic trees were made by maximum-likelihood inferences calculated with PhyML v3.0 using automatic Akaike Information Criterion for the substitution model and aLRt SH-like for branch support (Lefort et al., 2017), and by Bayesian inferences calculated by MrBayes 3.2.6 using invgamma model for among-site rate variation, WAG amino acid rate matrix, 2 million generations, in two parallel runs sampling every 1000 generations, and discarding 0.25 of sampled values as burnin to calculate branch support for posterior probabilities (Ronquist et al., 2012).

To analyze the muscular or cytoplasmic nature of actin genes, we have developed an algorithm that calculates the "Diagnostic-Actin-Value" (DAV) according to the conservation of previously described diagnostic positions (Nishino et al., 2000; Vandekerckhove and Weber, 1984) (for details see Sup. File 1). Briefly, according to the classification as muscular or cytoplasmic actins in the phylogenetic tree (branching support = 0.98), first we calculated the frequency of residue conservation of each diagnostic position among all 95 analyzed actins of vertebrates and ascidians. Then, to calculate the DAV, each position contributed with +1 or -1 if the residue was conserved in at least 20% of the muscular or cytoplasmic sequences, respectively. In case that a conserved residue for a group of actins was also found in sequences of the other group, the absolute value of the contribution for each position (i.e. +1 or -1) was corrected by subtracting the frequency of that residue in the other group. According to this algorithm, DAV can range from +16 to -16 depending on the total conservation of muscular or cytoplasmic diagnostic positions, respectively.

To analyze the exon-intron structure of actin genes we used the software GECA (Fawal et al., 2012), defining at least 70% of similarity in the 20 amino acids surrounding the intron position.

2.3. Cloning and expression analysis

Probes for whole-mount in situ hybridization (WMISH) of *O. dioica* actins were PCR amplified and cloned with the Topo TA Cloning® Kit of Invitrogen (primers are listed in Sup. File 2 Table S1). Probes were

designed on the 3′ untranslated regions in order to avoid cross-hybridization among paralogs. Cloning, sequencing, and WMISH experiments on fixed embryos at selected developmental stages were performed as previously described (Marti-Solans et al., 2015).

2.4. Nuclear and phalloidin staining

For nuclear and phalloidin staining, embryos were fixed in 4% paraformaldehyde in fixation buffer (0.1 M MOPS, 0.5 M NaCl, 2 mM MgSO4, 1 mM EGTA). After 1 h at room temperature, fixation buffer was replaced by washing twice in PBST (PBS Tween-20, 0,2%). Phalloidin and nuclei stainings were performed incubating the embryos in staining solution (1 μM Hoeschst-33342 Invitrogen-62249, 0.1 μM TRITC-Phalloidin Sigma-P-1951 in PBST) for 1 h at 37 °C in the dark. Stained embryos were washed in PBST and mounted in glycerol 80% in PBS, and photographed in a Fluorescent microscope (Zeiss Aixophot).

3. Results

3.1. Characterization and evolutionary history of Oikopleura dioica actin genes

Our survey of genomic and transcriptomic databases of *O. dioica* identified seven actin genes (labeled in red in Fig. 1**A and** Sup. Fig. S1) distributed in six scaffolds, whose annotations were manually curated and fixed according to available ESTs and our PCR clones (Sup. File 2 Table S2). Phylogenetic analysis suggested that four *O. dioica* actin genes were homologs to chordate muscle actin genes (namely *ActnM1*, *ActnM2*, *ActnM3* and *ActnM4*), whereas the other three grouped with the cytoplasmic ones (*ActnC1*, *ActnC2* and *ActnC3*). The fact that all *O. dioica* ActnM grouped in a single cluster together with a cDNA actin previously described in *Oikopleura longicauda* (Nishino et al., 2000), but without including any muscular actin from other urochordate species, suggested that *ActnMs* of *O. dioica* were paralogs that have been originated by independent gene duplications during the evolution of appendicularians, after their split from ascidians. Cytoplasmic *ActnC* genes of *O. dioca*, on the other side, grouped in two different clusters with moderate support (i.e. 0.91 and 0.88), one including *ActnC1* and *ActnC2*, and the other *ActnC3* (red circles in Fig. 1A). The fact that these two clusters also included actins from phlebobranchia and stolidobranchia ascidian species (dark and light blue circles, respectively, in Fig. 1**A**) suggested that their origin might come from ancient duplications that occurred at least before the split of the appendicularian and the ascidian lineages, followed by further duplications in the appendicularian lineage that gave rise to *ActnC1* and *ActnC2*.

Amino-acid sequence identity within the muscular and cytoplasmic actins of *O. dioica* was high, being > 99.5% among ActnMs and > 95% among ActnCs (Sup. File 2 Table S2). The fact that ActnM2 and ActnM3 were identical, and that the other *O. dioica* ActnMs did not differ in more than two amino acids was compatible with either the possibility that duplications that originated these paralogs were recent, or that all ActnM sequences were so similar because they have shared high functional redundancy and have been evolving under the same selective restrictions. Among cytoplasmic actins, ActnC1 and ActnC2 differed in 4 amino acids, while ActnC3 was the most divergent, showing up to 18 amino acid differences. Comparison between *O. dioica* muscle and cytoplasmic actins revealed that their lengths differed in 6 amino acids, due to a 18 nucleotides expansion of the first exon in the muscle actins during the evolution of the *O. dioica* lineage, absent in *O. longicauda*. *O. dioica* muscle and cytoplasmic actins showed 28–33 amino acid differences, suggesting differential selective restrictions for sequence variability between the two groups.

The high sequence similarity among *O. dioica* actins prompted us to analyze their exon-intron structures to look for further support that they were different genes rather than allelic variants. Differences in the exon-intron structures (Sup. File 2 Fig. S2), together with the lack of

sequence similarity between introns or untranslated regions (Sup. File 2 Table S3), discarded the possibility that the small number of amino acid differences among genes could be polymorphic variants, and corroborated therefore that the actin catalogue in *O. dioica* is made of seven genes. The exon-intron organization of most *O. dioica* actin genes was, indeed, considerably divergent from the structure that is broadly conserved in actin genes of other species (Sup. File 2 Fig. S2), which was consistent with the high intron turnover described for most genes in *O. dioica* (Edvardsen et al., 2004). Interestingly, the four muscle actin genes (*ActnM1−4*) shared a synapomorphic intron 1, absent in cytoplasmic actins, which corroborated their paralogy and that all *ActnM* genes originated by gene duplication during the appendicularian evolution (intron 1 in red Sup. File 2 Fig. S2). Likewise, the presence of at least two introns in *O. dioica ActnC1* and *ActnC2* that were synapomorphically conserved in other chordate cytoplasmic actin genes supported their position in the phylogenetic tree. The absence of introns in *ActnC3* could be explained either by events of intron loss, a common phenomena observed in many *O. dioica* genes (Denoeud et al., 2010), or by an evolutionary origin related to retrotranscription and integration of a transcript from an ancestral *ActnC* gene. The presence of an intron in *ActnM1* in the same position that in the *ActnC* genes prompted us to analyze the possibility of gene conversion by unequal crossing-over between the two groups of actins. Comparisons of the sequences of the intron and surrounding exons, however, revealed no evidence of gene conversion, concluding that the origin of that intron in *ActnM1* might be due to a convergent intron gain, a phenomenon also observed in several *O. dioica* genes (Denoeud et al., 2010).

To further corroborate the muscular and cytoplasmic nature of *O. dioica ActnM* and *ActnC* genes, we also analyzed the diagnostic positions that had been previously described to differentiate muscle and cytoplasmic actins (Vandekerckhove and Weber, 1984). To perform this analysis, we developed a novel algorithm that calculates the "Diagnostic-Actin-Value" (DAV) according to the conservation of diagnostic positions across an extended list of vertebrates and urochordate actins, adopting more positive or negative values, between + 16 and -16, according to the more muscle or more cytoplasmic characteristics of their sequences, respectively (Sup. File 1). In *O. dioica*, all muscular ActnM shared a DAV of +15.38, while the cytoplasmic actins ActnC1 and ActnC2 had a DAV of -15.42, and ActnC of -9.1 (Fig. 1B). These values strongly supported the assignment of their muscle or cytoplasmic nature reflected in the phylogenetic tree (Fig. 1**A**).

Results from our phylogenetic analysis, in which we included new actins from a broad panel of ascidian species, corroborated previous work showing how muscle actins experienced a gene duplication that originated two paralogous groups during the evolution of ascidians, classically referred as "larval" and "adult body-wall" ascidian actins (reviewed in Kusakabe, 1997), and recently renamed as ascidian-P (for Paraxial) or ascidian-B (for Body-wall) actins, respectively (Inoue and Satoh, 2018). However, none of our phylogenetic analyses, neither using maximum likelihood nor Bayesian approaches, unfortunately, provided robust support to clarify the relationship between *O. dioica* and ascidian-P and -B groups, in many cases collapsing in a trichotomy (Fig. 1**A** and Sup. File 2 Fig. S3). Interestingly, the DAV of all *O. dioica ActnM* paralogs (average 15.38) was closer to those of ascidian-P actins (average 14.69 ± 0.78) than those of ascidian-B actins (average of 12.41 ± 0.56), suggesting therefore that *O. dioica* ActnMs had a closer structural and likely functional relationship to ascidian-P actins than to ascidian-B actins (Fig. 1**B**). Further sequencing of actin genes in additional urochordate groups, especially in doliolids and salps as well as in other appendicularian species, will be necessary to solve the evolution of muscle actin genes in urochordates, and to clarify whether the duplication that gave rise to the P and B paralogs occurred before or after the appendicularian-ascidian split.
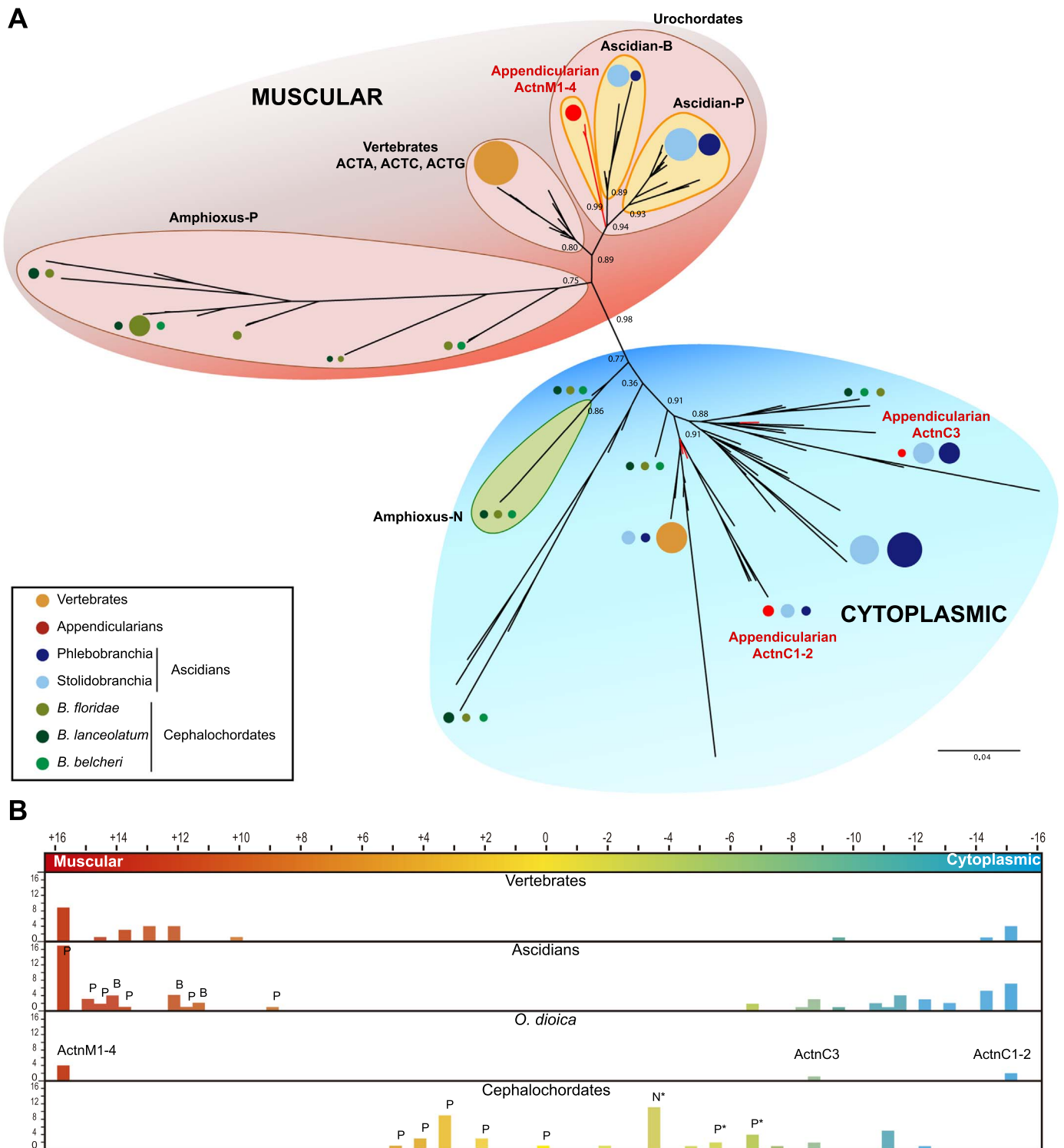
Fig. 1. Evolutionary history and analysis of diagnostic positions of muscular and cytoplasmic actins in chordates. (A) Phylogenetic tree inferred by maximum-likelihood of muscular (in red background) and cytoplasmic (in blue background) actins in chordates. Appendicularian ActnM1−4 and ActnC1−3 are labeled in red. For the sake of clarity, sequence species are depicted as circles with sizes proportional to the number of represented sequences and color coded according to species taxonomy as indicated in the legend (vertebrates in ochre, cephalochordate species in three different green tones, Phlebobranchia and Stolidobranchia ascidians in dark and light blue tones, respectively). Major actin groups are indicated, and values for the approximate likelihood-ratio test (aLRT) are shown in main nodes. Scale bar indicates amino acid substitutions. Fully detailed tree with all sequence names and branch supports are provided in Sup. File 2 Fig. S1. **(B)** Gradient colored representation from totally muscular (+16 in red) to totally cytoplasmic (-16 in blue) DAVs (Diagnostic-Actin-Values) of vertebrate, ascidian, *O. dioica*, and cephalochordate actins. Bars represent the number of sequences (y-axis) included in DAV ranges (x-axis on top). *O. dioica* ActnMs and ActnCs, ascidian-P and -B actins, and amphioxus-P and -N actins are labeled on top of the bars. Asterisks indicate that in addition to -P or -N actins, other cytoplasmic actins having the same DAV are included in the bar.
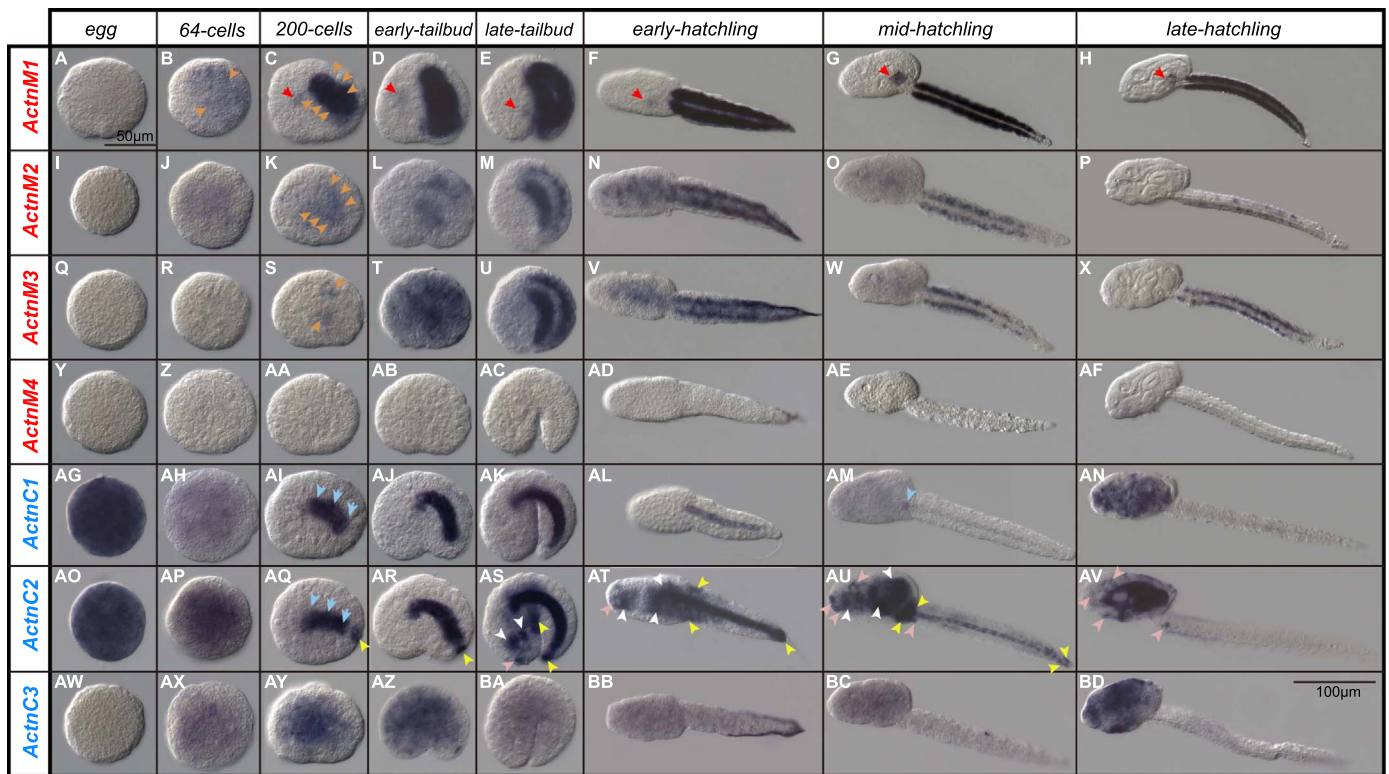
Fig. 2. Expression atlas of muscular and cytoplasmic actins during *Oikopleura dioica* development. Developmental (columns from egg to late-hatchling stage,) expression analysis by whole-mount in situ hybridization of *O. dioica* muscular (**A-H**, *ActnM1*; **I-P**, *ActnM2*; **Q-X**, *ActnM3*; **Y-AF**, *ActnM4*) and cytoplasmic (**AG-AN**, *ActnC1*; **AO-AV**, *ActnC2*; **AW-BD**, *ActnC3*) actin paralogs. Arrowheads: heart (red), paraxial muscle (orange), notochord (blue), epidermis (yellow), nervous system (pink), endoderm (white). Scale bars for eggs to late-tailbud (**A**, 50 μm), from early- to late-hatchling (**BD**, 100 μm).

3.2. Developmental expression atlas of Oikopleura dioica actin genes

In order to create a detailed developmental atlas of actin expression in *O. dioica*, we investigated the expression patterns of the seven actin genes by whole mount in situ hybridization (WMISH) throughout eight developmental stages from egg to late hatchling (Fig. 2**A-BD**). The high sequence similarity among the seven actin genes, however, hampered these analyses since we noticed that probes designed in coding regions could cross-hybridize among different actins (Sup. File 2 Table S2). To avoid cross-hybridization, therefore, we were forced to design all actin probes in regions of the 3′ untranslated regions that showed lower similarity to other actin genes (Sup. File 2 Table S3), which in some cases rendered small probes (< 100 bp) that required long staining periods up to three weeks in order to detect the signal.

WMISH results revealed that *ActnM* genes had tissue specific expression, which consistent with their 'phylogenetic' muscular nature appeared to be restricted to the muscle cell lineage (Fig. 2**A-AF**). Despite *ActnM* paralogs showed important overlap in their expression domains, they also displayed noticeable dissimilarities that suggested differences in their spatio-temporal regulation, confirming the specificity of the probes used in our WMISH experiments. Temporal differences were shown, for instance, by the time of the expression onset, in which *ActnM1* signal was the earliest to be observed by the 64-cell stage (Fig. 2**B**), followed by *ActnM2* and *ActnM3*, which did not show obvious signal up until the 200-cell stage and early-tailbud stage, respectively, in all cases labeling paraxial muscle cells (Fig. 2**K, S, T**, orange arrowheads). From the late tail-bud up to mid-hatchling stage, the signal of *ActnM1*, *ActnM2*, and *ActnM3* remained constant. By late hatchling stage, however, *ActnM2* seemed to downregulate as its expression signal appeared to be fainter (Fig. 2**P**). Spatial differences in the expression of *ActnM* genes were shown by *ActnM1*, which besides paraxial muscle, it was also expressed in the cardiac cell lineage (Fig. 2**C-H**). We did not observe any signal of *ActnM4* in any of the

many WMISH experiments we performed, not even decreasing the stringency of the conditions of the hybridization nor after long staining periods. It is possible that *ActnM4* might be expressed during adult stages, although we cannot discard that the limited small size of the specific probe of this paralog (i.e. 96 nucleotides) might be hampering the detection of signal, especially if this gene was lowly expressed during development in comparison with the other *ActnM* paralogs.

In the case of cytoplasmic genes (Fig. 2**AG-BD**), only one of the three paralogs, namely *ActnC3*, showed a generalized broad expression pattern after long staining periods. The expression of this gene therefore, could be consistent with a basic housekeeping function, which is expected from ubiquitously expressed cytoplasmic actins related to components of the cytoskeleton (Fig. 2**AX-BD**). However, the other two *ActnC* paralogs, namely *ActnC1* and *ActnC2*, showed highly dynamic tissue-specific expression patterns changing over time in different organs (Fig. 2**AG-AV**). Thus, unfertilized eggs displayed strong signal for *ActnC1* and *ActnC2*, suggesting that these two paralogs were maternally transcribed (Fig. 2**AG, AO**). By the 64-cell stage, most cells still showed *ActnC1* and *ActnC2* signal, but weaker than in eggs, suggesting that it likely was part of the still remaining maternal component (Fig. 2**AH, AP**). By the 200-cell stage, while the maternal signal of *ActnC1* and *ActnC2* was not detected anymore, the entire notochord appeared obviously stained by both paralogs, suggesting that the first zygotic expression of these two paralogs was notochord specific (Fig. 2**AI, AQ** blue arrowheads). This strong signal remained in the notochord throughout tailbud and early-hatchling stages, but in mid-hatchlings, while *ActnC2* signal was still obvious in the entire notochord, *ActnC1* signal appeared to have fainted in most of the notochord except in the first most rostral cell (Fig. 2**AM** blue arrowhead). By the 200-cell stage, in addition to the notochord, the *ActnC2* expression signal was also detected in an ectodermal domain restricted to the most distal cells of the tail, corresponding to the developing tailbud region (Fig. 2**AI, AQ** yellow arrowhead). This *ActnC2* expression domain in the tailbud remained strongly labeled

during tailbud and hatchling stages, but disappeared by late-hatchling stage coinciding with the time at which the tail stops elongating (Fig. 2**AV**). By the late-tailbud stage, *ActnC2* signal showed a new expression domain broadly extended in the trunk, including endodermal-, neural- and epidermal-derived tissues, which remained throughout all analyzed hatchling stages (Fig. 2**AS-AV** white, pink and yellow arrowheads for endodermal, neural and epidermal domains, respectively; for details see below next result section). By the late-hatchling stage, in addition to *ActnC2*, the expression signal of *ActnC1* and *ActnC3* also appeared for the first time throughout the trunk (Fig. 2**AN, BD**).

Overall, the developmental atlas of actin expression revealed that both muscle and cytoplasmic actin genes have gone through a complex process of sub-functionalization and/or neofunctionalization after extensive gene duplications during the evolution of appendicularians, in which different paralogs would have acquired notorious differences in their spatio-temporal regulation.

3.3. Detection of actin filaments during Oikopleura dioica development

To visualize actin filaments we used phalloidin staining. While in stages prior to tailbud phalloidin stained most cells (data not shown), by tailbud stage, phalloidin signal was specially obvious in the membranes of notochordal cells, as well as in the membranes of internal cells of the trunk, the internal part of the trunk-tail junction, and in a very restricted area in the most distal part of the tail corresponding to the tailbud (Fig. 3**A**). Most regions that were stained for phalloidin recapitulated the observed expression domains of *ActnC1* and *ActnC2* (Fig. 3**B and** Fig. 2**AG-AV**). Muscle cells at tailbud stage, however, despite the strong expression of three *ActnM* paralogs (namely *ActnM1*, *M2* and *M3*), did not show any phalloidin staining, suggesting therefore that actin had not polymerized yet, and myofilaments had not been formed at this developmental stage (Fig. 3**C,D**).

After hatching, phalloidin staining revealed the typical "striated" organization of actin myofilaments of the skeletal muscle throughout the tail (Fig. 3**E-G**). The fact that the tail switches 90° respect to the trunk during *O. dioica* development (Cañestro et al., 2005) makes the muscle cells lie in the dorsal and ventral sides of the tail. Thus, from a ventral view, phalloidin-labeled myofilaments displayed a striated pattern spanning the width of the notochord (Fig. 3**E**), whereas from a lateral view, they appeared as two dense lines close to the membrane near the notochord (Fig. 3**G**). In late hatchling stages, phalloidin staining revealed dark and bright bands of the striated pattern corresponding to the presence of sarcomeres in the paraxial muscle in the tail, as well as a clear dark demarcation corresponding to the border of muscle cells (Fig. 3**K**). The actin of myofilaments likely corresponded to the products of *ActnM* paralogs (Fig. 2).

In late hatchlings, in addition to the broad phalloidin signal labeling most of the cytoskeleton attached to cell membranes, we observed strong signal in different parts of the digestive system (i.e. the mid-dorsal part of the endostyle, the floor and roof of the esophagus, the internal membranes of stomach lobules, the rectum and the ciliary rings within the gills), the nervous system (i.e. rostral paired nerves n1, the sensory vesicle, the ciliary funnel, the ventral organ and some cells in the anterior third of the caudal ganglion), as well as the base of the placode derived structures of the Langerhans receptors (Fig. 3**G-K**). All these rich actin domains expressed *ActnC2*, and all of them have in common either the presence of cilia (Fig. 3**K**) or axonal projections.

4. Discussion

The present study identifies the complete catalogue of actin genes in *O. dioica*, which based on phylogenetic analyses, intron positions, DAVs and expression patterns is made of four muscular *ActnM1−4* and three cytoplasmic *ActnC1−3* genes. Our work also suggests that the multiple paralogs that belong to each group of actins have been mostly

originated by lineage-specific duplications during the evolution of the appendicularians. Our results corroborate the muscular and cytoplasmic homologies assigned to two actin cDNAs that had previously been isolated in *O. longicauda* (Nishino et al., 2000), as well as to six *O. dioica* actins predicted by an automatic orthology pipeline designed to identify muscle genes in deuterostome genomes (Inoue and Satoh, 2018). Previous analyses of ascidian actins (Beach and Jeffery, 1990; Chiba et al., 2003; Kovilur et al., 1993; Kusakabe et al., 1995, 1992, 1996; Swalla et al., 1994; Tomlinson et al., 1987) together with our genomic survey and phylogenetic analyses suggest that different ascidian lineages seem to have also expanded their actin catalogues independently (i.e. at least 14 in *Ciona robusta* −formerly *C. intestinalis*−, 9 in *Ciona savignyi*, 16 in *Phallusia mammillata*, 8 in *Molgula oculata*, 9 in *Molgula occidentalis*, 16 in *Halocynthia roretzi*, and 11 in *Botryllus schlosseri*) (Sup. File 2 Fig. S2 and S3).

4.1. Ancestral "cardio-paraxial" muscular actin in stem olfactores

In the case of ascidian muscle actins, our phylogenetic study shows that ascidian-P and -B paralogous groups are present in all ascidian species analyzed in both Phlebobranchia and Stolidobranchia orders. This fact reinforces results from previous work that had suggested that the origin of these paralogous groups was due to an ancient gene duplication that occurred during early ascidian evolution (Inoue and Satoh, 2018). Unfortunately, current available sequences do not contain enough phylogenetic signal to provide robust enough support to determine if that duplication occurred before or after the appendicular-ascidian split. Our findings, however, favor the scenario that the duplication occurred after the appendicularian-ascidian split because *O. dioica* ActnMs seem to share a mix of characteristics with both ascidian-P and -B groups, which may represent the ancestral urochordate condition. On one side, the DAVs of *O. dioica* ActnMs are closer to those of ascidian-P than to those of ascidian-B actins. The sequence resemblance with ascidian-P may be due to similar molecular interactions with specific actin-binding proteins of the myosin motor machinery (Perrin and Ervasti, 2010), which are necessary to drive tail movement in both *O. dioica* and ascidian larvae. On the other side, the expression pattern of *O. dioica* ActnM1 shares expression domains with ascidian-P and -B in paraxial muscle and heart (see a summary of *C. robusta* actin expression data Sup. File 2 Fig. S4 from ANISEED (www.aniseed.cnrs.fr); although cross-hybridization between probes of ascidian-P and -B actin genes, which share an overall 87% of identity, cannot be discarded (Brozovic et al., 2018)). Our results, therefore, are consistent with an evolutionary scenario in which stem urochordates had an ancestral "cardio-paraxial" actin gene that was involved both in heart and paraxial muscle development (Fig. 4). Then, in agreement with recent work by Inoue and Satoh (Inoue and Satoh, 2018), after the split between the appendicularian and ascidian lineages, the ancestral "cardio-paraxial" actin was duplicated in ascidians, experiencing a process of subfunctionalization in which ascidian-P actins mostly evolved under selective restrictions related to motor-tail movement, and ascidian-B actins mostly evolved under the selection of post-metamorphic innovations, such as adult body-wall muscle and adult heart slow paced contractility (Fig. 1**B**). Studies of further actins in additional urochordate groups, especially in doliolids and salps and other appendicularians species, will be necessary to corroborate or disprove this hypothesis. To our knowledge, appendicularians do not have any tissue homologous to the body-wall muscle present in ascidians or some thaliaceans, so further studies of the phylogeny and expression of actins from these different groups will possibly help to better understand the origin and evolution of body-wall muscle in urochordates (Degasperi et al., 2009).

Analyses in vertebrates had hypothesized that the last common ancestor of vertebrates also had an actin that was expressed both in paraxial muscle and heart (Vandekerckhove and Weber, 1984). Our work comparing urochordates and vertebrates points to an older origin
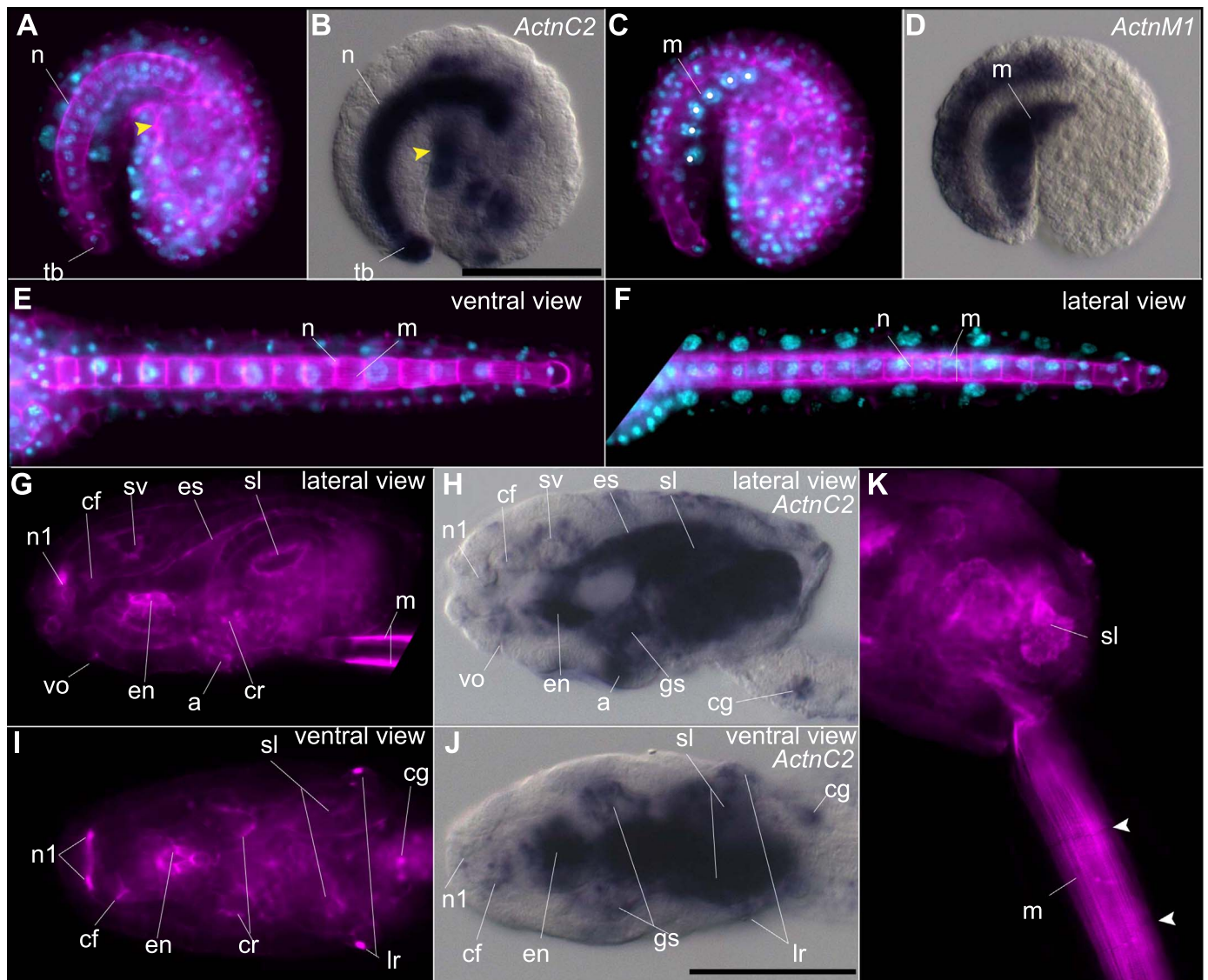
Fig. 3. Correlation between actin-filament detection by phalloidin-staining and actin paralog gene expression. (A-D) At tailbud stage (lateral views), phalloidin staining (**A, C**, magenta) labeled cells in the trunk, specially of the internal membrane of the epidermal cells and the cells in the internal area of the trunk-tail junction (yellow arrowhead), notochord (n) and the tailbud (tb), consistently with *ActnC2* expression (**B**), but it did not label muscle (m) cells (white dots on Hoeschst-stained nuclei in blue) despite the strong expression of *ActnM* genes **(D). (E-F)** After hatching, however, phalloidin-stained myofilaments in muscle cells could be observed from a ventral view displaying a striated pattern **(E)**, and forming two lines near the notochord from a lateral view **(F)**. (**G-J**) In late hatching stages, phalloidin labeled endodermal tissues such the endostyle (en), esophagus (es), stomach lobes (sl), the ciliary rings in the gills (gs) and anus (a), neural derivatives such as the paired nerves 1 (n1), ciliary funnel (cf), sensory vesicle (sv), ventral organ (vo), caudal ganglion (cg), and the placode derived Langerhans receptors (lr). The presence of actin in these organs was consistent with the expression of *ActnC2*. All these structures had in common either the presence of cilia or axonal projections. In tail shift stage **(K)**, phalloidin-stained cilia were easily observed within the stomach lobe. At this stage, phalloidin staining also revealed the presence of dark and light bands typical of sarcomeres of the striated muscle, and it also labeled the limits of each muscle cell (white arrowheads).

of the "cardio-paraxial" actin present at least in the last common ancestor of olfactores (urochordates + vertebrates), since *ActnM1* in *O. dioica,* the ascidian-B *Ci-Ma2* in *C. robusta,* and the alpha-cardiac *ACTC1* and the alpha-skeletal *ACTA1* in vertebrates share expression in both cardiac and paraxial muscle cell precursors during embryo development (Fig. 4) (Mayer et al., 1984; Ordahl, 1986; Schwartz et al., 1986). While *ActnM1* seems to be the only *ActnM* paralog expressed during heart development in *O. dioica,* in vertebrates, three alpha muscle actins are sequentially expressed during cardiac development. *ACTA2* expression starts marking the onset of cardiomyocyte differ- entiation, which is sequentially replaced first by *ACTA1* and later by *ACTC1.* This process has been related to the regulation of the specific functional activities of each actin isoforms in different developmental stages (Bertola et al., 2008; Ruzicka and Schwartz, 1988; Woodcock- Mitchell et al., 1988). During skeletal muscle development, sequential expression of different actins is also observed in vertebrates and *O.*

dioica. First *ACTC1* and next *ACTA1* are expressed in vertebrates (Mayer et al., 1984; Schwartz et al., 1986), and *ActnM1* expression is followed by *ActnM2* and *ActnM3* in *O. dioica.* Altogether, our work highlights similarities as well as differences in the parallel evolution that actin paralogs have experienced after being independently dupli- cated in both urochordates and vertebrates, including subfunctionali- zation leading to temporal regulation (i.e. differential expression onsets) and tissue specificity (i.e. paraxial vs cardiac) specializations.

Our phylogenetic analysis corroborates that muscular actins were originated by a gene duplication from an cytoplasmic actin at the base of chordates, before the split between cephalochordates and olfactores, which was followed by the recruitment of the myosin motor-machinery that conferred contractile capability to muscle cells (Inoue and Satoh, 2018). During the evolution of cephalochordates, muscular actins followed complex and divergent evolutionary histories as proposed by previous works describing lineage-specific duplications and events of
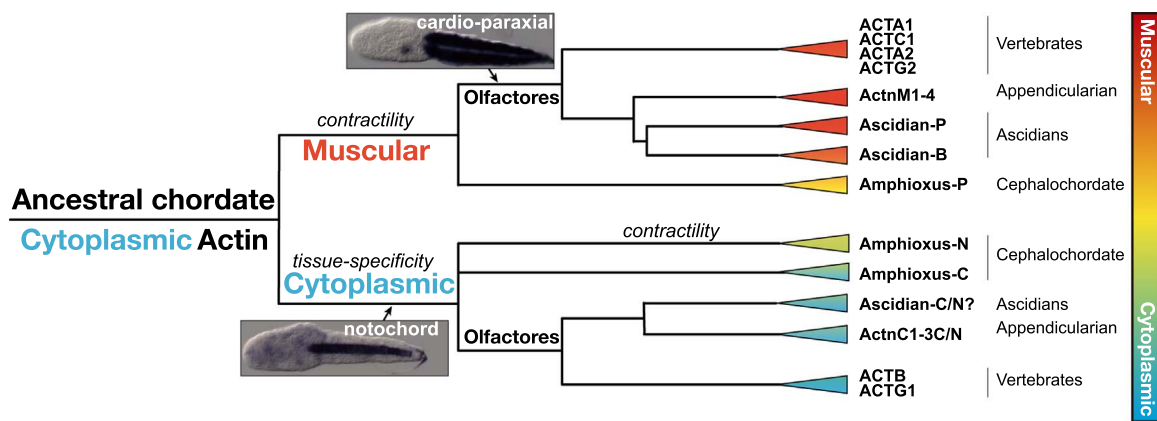
Fig. 4. Schematic representation of the actin evolution in chordates. Muscular actins were originated by an ancestral duplication from a cytoplasmic gene at the base of the chordate superphylum, followed by the acquisition of contractility by the recruitment of the myosin motor-machinery in muscle cells. The ancestral muscle actin of olfactores already played "cardio-paraxial" roles, as currently reflected by *ActnM1* in *O. dioica*. Extensive gene duplications independently expanded muscular actins in vertebrates, appendicularians and ascidians, in the latter originating two paralogous groups, ascidian-P and -B actins. Cytoplasmic actins in chordates show tissue-specific expression patterns, challenging the housekeeping notion classically ascribed to these cytoplasmic (C) actins. Interestingly, our results showing notochord-specific cytoplasmic actins in *O. dioica* likely represents the ancestral condition of chordates. During the evolution of cephalochordates, the recruitment of the myosin motor-machinery to the ancestral cytoplasmic-N actins led also to the acquisition of contractility of the notochord in this lineage. DAV analysis likely reflects the different selective restrictions that apply to the evolution of diagnostic positions according to protein-protein interactions of each group of actins (color code is labeled as in Fig. 1**B).**

gene conversion in cephalochordates (Fig. 1**A**) (Inoue and Satoh, 2018; R. Kusakabe et al., 1997a, 1999; Suzuki and Satoh, 2000). Our analysis of DAVs supports a distinctive evolutionary pattern of cephalochordate actins in which none of the muscular actins of amphioxus shows the characteristic high positive values of muscle actins in olfactores (Fig. 1**B and** Sup. File 1**)**, but they have lower positive DAVs close to a "neutral" region, in which actins likely share characteristics of both muscular and cytoplasmic actins (Fig. 1**B**). Overall, our results suggest that the evolutionary restrictions of muscle actins related to interactions with motor actin-binding proteins differ between cephalochordates and olfactores.

4.2. Ancestral notochord-specific cytoplasmic actin in stem chordates

Cephalochordates have a third type of actins, named amphioxus-N (notochord) actins that are expressed in the notochord together with many other motor-proteins that confer contractile capability to the notochord (Fig. 4) (Suzuki and Satoh, 2000; Urano et al., 2003). Our phylogenetic analyses and relatively not too negative DAV results of amphioxus-N actins are compatible with the work by Inoue and Satoh that suggested that these actins originated from a duplication of a cytoplasmic actin gene (Inoue and Satoh, 2018), and likely suffered a convergent selective process towards contractile muscular actins. Considering the notochord-specific expression of amphioxus-N actins, our finding that *O. dioica ActnC1* and *ActnC2* are tissue-specific with a prominent expression in the notochord, allow us to infer that the expression of cytoplasmic actins in the notochord likely represents the ancestral condition already present at least in stem chordates **(**Fig. 4**)**. In ascidians, several studies have shown a fundamental role of actin filaments in notochord convergent extension processes during morphogenesis as well as providing stiffness to notochord structure (Jiang and Smith, 2007; Munro and Odell, 2002a, 2002b). Our results detecting actin filaments in the notochord (Fig. 3) are compatible with a potential conservation of actin functions between *O. dioica* and ascidians. To our knowledge, however, no evidence has been found supporting the expression of muscle genes in the notochord of ascidians or appendicularians (Hotta et al., 2000; Jiang and Smith, 2007; Kugler et al., 2011; Takahashi et al., 1999 an our unpublished data). Our findings, therefore, are consistent with the recently postulated hypothesis by Inoue and Satoh, in which the evolution of the amphioxus-N actin and recruitment of muscle genetic machinery in the notochord of cephalochordates would be an innovation of this lineage

(Inoue and Satoh, 2018), whereas the urochordate non-contractile actin-expressing notochord, as it is currently found in *O. dioica*, would reflect the ancestral condition of stem chordates (Fig. 4).

In the last years, the evolutionary origin of the notochord has been a hot topic of discussion (Annona et al., 2015; Brunet et al., 2015; Hejnol and Lowe, 2014), especially after the axochord hypothesis suggesting that the notochord could have evolved by modification of a ventromedian muscle axochord present in stem bilaterians (Lauri et al., 2014). During the characterization of the axochord in different bilaterians, phalloidin staining has shown the presence of actin expression in this structure (Lauri et al., 2014). In the light of our new data showing that the notochord-specific expression of cytoplasmic actins is an ancestral chordate condition, future comparative studies of the regulatory mechanisms of actin expression in the axochord and in the notochord, as well as investigations on whether homologous actins are responsible for the process of midline convergence of both axochord (Lauri et al., 2014) and notochord (Jiang and Smith, 2007) could shed some light on the evolution of the notochord, and therefore in the origin of chordates.

4.3. Tissue-specificity and temporal regulation of cytoplasmic actins related to cilia and axonal projections

The notion that cytoplasmic actins are housekeeping genes with constitutive and ubiquitous expression patterns is an extended idea that lately has been challenged, for instance, by cases in which actin expression has been shown not to be appropriate to normalize RNA or protein quantification (Lin and Redies, 2012; Ruan and Lai, 2007). Our results show that from the three ActnC paralogs of *O. dioica*, only *ActnC3* displays a broad non-tissue specific expression pattern consistent with a housekeeping role. Its ubiquitous expression could be related to its possible retrotranscriptional origin and subsequent insertion in a genomic location driving its transcription in a non-tissue-specific manner. In contrast to *ActnC3*, the other two paralogs *ActnC1* and *ActnC2* show remarkable tissue-specific and temporally-regulated expression patterns, not compatible with housekeeping roles (Figs. 2 and 3). In *O. dioica*, the expression domains of these two *ActnC* paralogs –besides the notochord– correlates with regions characterized by either the presence of cilia or axonal projections, as well as regions with presumptive high cell proliferation. In the endostyle, for instance, the expression domain of *ActnC2* is restricted to the central part of the endostyle (mostly dorsal, but also ventral; Fig. 3**H, J**). This region is characterized by the presence of ciliated cells that project giant cilia

that can be easily observed with phalloidin staining towards the corridor of this organ (Fig. 3**G,I**), in which they wind the food trapping secretions towards the pharynx lumen. In the digestive system, *ActnC2* is highly expressed in the esophagus, left and right stomach lobes, vertical and mid intestine, and rectum, which all of them are characterized by the presence of ciliated cells (Burighel and Brena, 2001)(Fig. 3**G-J**). In the pharyngo-branchial region, the expression domains of *ActnC1* and *ActnC2* also correlates with the presence of cilia in the pharynx and the ciliary rings in the gills, the beating of which induce water circulation from the mouth to the gill slits, or in opposite direction depending on the orientation of the beating (Fig. 3**G-J**). In the epidermis, the expression domains of *ActnC2* correlates with the presence of ciliated sensory cells, including the mechanosensory cells in the circumoral organ, the ventral organ and the mechanoreceptors of the Langerhans cells (Fig. 3**G-J**). These sensory structures have been described to be evolutionary related to placodes, which have been hypothesized to be key innovations for the origin and evolution of sensory organs in olfactores (Bassham and Postlethwait, 2005; Manni et al., 2004). In the nervous system, the expression domains of *ActnC2* correlates with either the presence of cilia (i.e. coronet cells in the sensory vesicle whose cilia touch the statocyte likely related to responses to positional cues such as vibration, acceleration or gravity (Bone, 1998; Olsson, 1975)) or axonal projections, such as those forming the rostral paired nerves n1, nerves innervating the ventral organ or neurons in the caudal ganglia projecting towards the motor system (Bollner et al., 1986; Burighel et al., 2011; Cañestro et al., 2005; Soviknes et al., 2007; Soviknes and Glover, 2007). Interestingly, most of these *ActnC2* expression domains with cilia and axonal projections also express one of the two *Pax2/5/8* paralogs of *O. dioica* (Bassham et al., 2008; Cañestro et al., 2008; Cañestro and Postlethwait, 2007), suggesting a potential role of this transcription factor upstream of *ActnC2* during the development of cilia and axonal projections. Finally, we observe *ActnC2* expression in the most distal part of the tail (Fig. 2**AQ-AU**), a region in which presumably high cell proliferation and cell differentiation may be occurring related to the process of tail elongation. Our finding of non-ubiquitous expression of *ActnC* genes in *O. dioica* is consistent with the tissue-specificity already described for a cytoplasmic actin gene in the ascidian *Halocynthia roretzi*, which showed prominent expression in mesenchymal cells of the trunk, in epidermal mechanoreceptor cells, neurons, cells with cytoplasmic extensions, as well as faint expression in the notochord (Araki et al., 1996). We can conclude, therefore, that the tissue-specificity of cytoplasmic actin genes is likely an ancestral condition of urochordates.

Overall, our results highlight that during the evolution of chordates, cytoplasmic actins did not simply play housekeeping roles, but they also displayed tissue-specific and temporally-regulated expression patterns during a complex evolutionary history with extensive gene duplications. The role of actins in axon projections and the formation of cilia and flagella seems to be of ancient origin (Quarmby, 2014; Roblodowski and He, 2017). However, the recruitment of cytoplasmic actin genes for the development of ciliated cells present in placode-like sensory organs and pharyngo-branchial structures, the recruitment of cytoplasmic actin genes for the development of the notochord, and the duplication of cytoplasmic genes that gave rise to muscular actins involved in cardio-paraxial development appear as crucial events associated to key evolutionary innovations of the chordate body plan.

Authors' contributions

A.A. carried gene cloning and WMISH experiments. Nuclear and phalloidin stainings were carried out by A.A. and A.F-R. A.F-R. performed gene identifications, phylogenetic analyses, exon-intron analyses, and development of DAV algorithm. A.A. and A.F-R interpreted the data and made the figures. R.A. and C.C. supervised the experiments. C.C. conceptualized the project and wrote the MS with

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ydbio.2018.09.003.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Annona, G., Holland, N.D., D'Aniello, S., 2015. Evolution of the notochord. Evodevo 6, 30.

Araki, I., Tagawa, K., Kusakabe, T., Satoh, N., 1996. Predominant expression of a cytoskeletal actin gene in mesenchyme cells during embryogenesis of the ascidian Halocynthia roretzi. Dev. Growth Differ. 38, 401–411.

Avasthi, P., Onishi, M., Karpiak, J., Yamamoto, R., Mackinder, L., Jonikas, M.C., Sale, W.S., Shoichet, B., Pringle, J.R., Marshall, W.F., 2014. Actin is required for IFT regulation in Chlamydomonas reinhardtii. Curr. Biol. 24, 2025–2032.

Bassham, S., Cañestro, C., Postlethwait, J.H., 2008. Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. BMC Biol. 6, 35.

Bassham, S., Postlethwait, J.H., 2005. The evolutionary history of placodes: a molecular genetic investigation of the larvacean urochordate Oikopleura dioica. Development 132, 4259–4272.

Beach, R.L., Jeffery, W.R., 1990. Temporal and spatial expression of a cytoskeletal actin gene in the ascidian Styela clava. Dev. Genet. 11, 2–14.

Bertola, L.D., Ott, E.B., Griepsma, S., Vonk, F.J., Bagowski, C.P., 2008. Developmental expression of the alpha-skeletal actin gene. BMC Evol. Biol. 8, 166.

Besse, F., Ephrussi, A., 2008. Translational control of localized mRNAs: restricting protein synthesis in space and time. Nat. Rev. Mol. Cell Biol. 9, 971–980.

Bezanilla, M., Gladfelter, A.S., Kovar, D.R., Lee, W.L., 2015. Cytoskeletal dynamics: a view from the membrane. J. Cell Biol. 209, 329–337.

Blanchoin, L., Boujemaa-Paterski, R., Sykes, C., Plastino, J., 2014. Actin dynamics, architecture, and mechanics in cell motility. Physiol. Rev. 94, 235–263.

Bollner, T., Holmberg, K., Olsson, R., 1986. A rostral sensory mechanism in Oikopleura dioica (Appendicularia). Acta Zool. 67, 235–241.

Bone, Q., 1998. Nervous system, sense organs, and excitable epithelia. In: Bone, Q. (Ed.), The Biology of Pelagic Tunicates. Oxford University Press, New York, 55–80.

Brozovic, M., Dantec, C., Dardaillon, J., Dauga, D., Faure, E., Gineste, M., Louis, A., Naville, M., Nitta, K.R., Piette, J., Reeves, W., Scornavacca, C., Simion, P., Vincentelli, R., Bellec, M., Aicha, S.B., Fagotto, M., Gueroult-Bellone, M., Haeussler, M., Jacox, E., Lowe, E.K., Mendez, M., Roberge, A., Stolfi, A., Yokomori, R., Brown, C.T., Cambillau, C., Christiaen, L., Delsuc, F., Douzery, E., Dumollard, R., Kusakabe, T., Nakai, K., Nishida, H., Satou, Y., Swalla, B., Veeman, M., Volff, J.N., Lemaire, P., 2018. ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. Nucleic Acids Res 46, D718–D725.

Brunet, T., Lauri, A., Arendt, D., 2015. Did the notochord evolve from an ancient axial muscle? The axochord hypothesis. Bioessays 37, 836–850.

Burighel, P., Brena, C., 2001. Gut ultrastructure of the appendicularian Oikopleura dioica (Tunicata). Invertebr. Biol. 120, 278–293.

Burighel, P., Caicci, F., Manni, L., 2011. Hair cells in non-vertebrate models: lower chordates and molluscs. Hear. Res. 273, 14–24.

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

Cañestro, C., Bassham, S., Postlethwait, J.H., 2008. Evolution of the thyroid: anterior-posterior regionalization of the Oikopleura endostyle revealed by Otx, Pax2/5/8, and Hox1 expression. Dev. Dyn. 237, 1490–1499.

Cañestro, C., Postlethwait, J.H., 2007. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. Dev. Biol. 305, 522–538.

Chang, C.H., Lee, H.H., Lee, C.H., 2017. Substrate properties modulate cell membrane roughness by way of actin filaments. Sci. Rep. 7, 9068.

Chiba, S., Awazu, S., Itoh, M., Chin-Bow, S.T., Satoh, N., Satou, Y., Hastings, K.E., 2003. A genomewide survey of developmentally relevant genes in Ciona intestinalis. IX. Genes Muscle Struct. Proteins Dev. Genes Evol. 213, 291–302.

Conti, M.A., Adelstein, R.S., 2008. Nonmuscle myosin II moves in new directions. J. Cell Sci. 121, 11–18.

Degasperi, V., Gasparini, F., Shimeld, S.M., Sinigaglia, C., Burighel, P., Manni, L., 2009. Muscle differentiation in a colonial ascidian: organisation, gene expression and evolutionary considerations. BMC Dev. Biol. 9, 48.

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J.M., Da Silva, C., Brinkmann, H., Mikhaleva, J., Olsen, L.C., Jubin, C., Cañestro, C., Bouquet, J.M., Danks, G., Poulain, J., Campsteijn, C., Adamski, M., Cross, I., Yadetie, F., Muffato, M., Louis, A., Butcher, S., Tsagkogeorga, G., Konrad, A., Singh, S., Jensen, M.F., Cong, E.H., Eikeseth-Otteraa, H., Noel, B., Anthouard, V., Porcel, B.M., Kachouri-Lafond, R., Nishino, A., Ugolini, M., Chourrout, P., Nishida, H., Aasland, R., Huzurbazar, S., Westhof, E., Delsuc, F., Lehrach, H., Reinhardt, R., Weissenbach, J., Roy, S.W., Artiguenave, F., Postlethwait, J.H., Manak, J.R., Thompson, E.M., Jaillon, O., Du Pasquier, L., Boudinot, P., Liberles, D.A., Volff, J.N., Philippe, H., Lenhard, B., Roest Crollius, H., Wincker, P., Chourrout, D., 2010. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science 330, 1381–1385.

Dominguez, R., Holmes, K.C., 2011. Actin structure and function. Annu. Rev. Biophys. 40, 169–186.

Edvardsen, R.B., Lerat, E., Maeland, A.D., Flat, M., Tewari, R., Jensen, M.F., Lehrach, H., Reinhardt, R., Seo, H.C., Chourrout, D., 2004. Hypervariable and highly divergent intron-exon organizations in the chordate Oikopleura dioica. J. Mol. Evol. 59, 448–457.

Farwell, A.P., Lynch, R.M., Okulicz, W.C., Comi, A.M., Leonard, J.L., 1990. The actin cytoskeleton mediates the hormonally regulated translocation of type II iodothyronine 5'-deiodinase in astrocytes. J. Biol. Chem. 265, 18546–18553.

Fawal, N., Savelli, B., Dunand, C., Mathe, C., 2012. GECA: a fast tool for gene evolution and conservation analysis in eukaryotic protein families. Bioinformatics 28, 1398–1399.

Flood, P.R., Guthrie, D.M., Banks, J.R., 1969. Paramyosin muscle in the notochord of Amphioxus. Nature 221, 87–88.

Fritzsche, M., Erlenkamper, C., Moeendarbary, E., Charras, G., Kruse, K., 2016. Actin kinetics shapes cortical network structure and mechanics. Sci. Adv. 2, e1501337.

Fyrberg, E.A., Bond, B.J., Hershey, N.D., Mixter, K.S., Davidson, N., 1981. The actin genes of Drosophila: protein coding regions are highly conserved but intron positions are not. Cell 24, 107–116.

Fyrberg, E.A., Mahaffey, J.W., Bond, B.J., Davidson, N., 1983. Transcripts of the six Drosophila actin genes accumulate in a stage- and tissue-specific manner. Cell 33, 115–123.

Gunning, P.W., Ghoshdastider, U., Whitaker, S., Popp, D., Robinson, R.C., 2015. The evolution of compositionally and functionally distinct actin filaments. J. Cell Sci. 128, 2009–2019.

Hejnol, A., Lowe, C.J., 2014. Animal evolution: stiff or squishy notochord origins? Curr. Biol. 24, R1131–R1133.

Heng, Y.W., Koh, C.G., 2010. Actin cytoskeleton dynamics and the cell division cycle. Int. J. Biochem. Cell Biol. 42, 1622–1633.

Hightower, R.C., Meagher, R.B., 1986. The molecular evolution of actin. Genetics 114, 315–332.

Hofmann, W.A., 2009. Cell and molecular biology of nuclear actin. Int. Rev. Cell Mol. Biol. 273, 219–263.

Hooper, S.L., Thuma, J.B., 2005. Invertebrate muscles: muscle specific genes and proteins. Physiol. Rev. 85, 1001–1060.

Hotta, K., Takahashi, H., Asakura, T., Saitoh, B., Takatori, N., Satou, Y., Satoh, N., 2000. Characterization of Brachyury-downstream notochord genes in the Ciona intestinalis embryo. Dev. Biol. 224, 69–80.

Inoue, J., Satoh, N., 2018. Deuterostome genomics: lineage-specific protein expansions that enabled chordate muscle evolution. Mol. Biol. Evol..

Jiang, D., Smith, W.C., 2007. Ascidian notochord morphogenesis. Dev. Dyn. 236, 1748–1757.

Kapoor, P., Chen, M., Winkler, D.D., Luger, K., Shen, X., 2013. Evidence for monomeric actin function in INO80 chromatin remodeling. Nat. Struct. Mol. Biol. 20, 426–432.

Kondrikov, D., Fonseca, F.V., Elms, S., Fulton, D., Black, S.M., Block, E.R., Su, Y., 2010. Beta-actin association with endothelial nitric-oxide synthase modulates nitric oxide and superoxide generation from the enzyme. J. Biol. Chem. 285, 4319–4327.

Kovilur, S., Jacobson, J.W., Beach, R.L., Jeffery, W.R., Tomlinson, C.R., 1993. Evolution of the chordate muscle actin gene. J. Mol. Evol. 36, 361–368.

Kugler, J.E., Kerner, P., Bouquet, J.M., Jiang, D., Di Gregorio, A., 2011. Evolutionary changes in the notochord genetic toolkit: a comparative analysis of notochord genes in the ascidian Ciona and the larvacean Oikopleura. BMC Dev. Biol. 11, 21.

Kusakabe, R., Kusakabe, T., Satoh, N., Holland, N.D., Holland, L.Z., 1997a. Differential gene expression and intracellular mRNA localization of amphioxus actin isoforms throughout development: implications for conserved mechanisms of chordate development. Dev. Genes Evol. 207, 203–215.

Kusakabe, R., Satoh, N., Holland, L.Z., Kasakabe, T., 1999. Genomic organization and evolution of actin genes in the amphioxus Branchiostoma belcheri and Branchiostoma floridae. Gene 227, 1–10.

Kusakabe, T., 1997. Ascidian actin genes: developmental regulation of gene expression and molecular evolution. Zool. Sci. 14, 707–718.

Kusakabe, T., Araki, I., Satoh, N., Jeffery, W.R., 1997b. Evolution of chordate actin genes: evidence from genomic organization and amino acid sequences. J. Mol. Evol. 44, 289–298.

Kusakabe, T., Hikosaka, A., Satoh, N., 1995. Coexpression and promoter function in two muscle actin gene complexes of different structural organization in the ascidian Halocynthia roretzi. Dev. Biol. 169, 461–472.

Kusakabe, T., Makabe, K.W., Satoh, N., 1992. Tunicate muscle actin genes. Structure and organization as a gene cluster. J. Mol. Biol. 227, 955–960.

Kusakabe, T., Swalla, B.J., Satoh, N., Jeffery, W.R., 1996. Mechanism of an evolutionary change in muscle cell differentiation in ascidians with different modes of development. Dev. Biol. 174, 379–392.

Lauri, A., Brunet, T., Handberg-Thorsager, M., Fischer, A.H., Simakov, O., Steinmetz, P.R., Tomer, R., Keller, P.J., Arendt, D., 2014. Development of the annelid axochord: insights into notochord evolution. Science 345, 1365–1368.

Lefort, V., Longueville, J.E., Gascuel, O., 2017. SMS: smart model selection in PhyML. Mol. Biol. Evol. 34, 2422–2424.

Lin, J., Redies, C., 2012. Histological evidence: housekeeping genes beta-actin and GAPDH are of limited value for normalization of gene expression. Dev. Genes Evol. 222, 369–376.

Lodish, H., Berk, A., Zipursky, S.L., 2000. al., e. Cell motility and shape I: microfilaments, In: Freeman, W.H. (Ed.), Molecular Cell Biology, 4th ed., New York.

Manni, L., Lane, N.J., Joly, J.S., Gasparini, F., Tiozzo, S., Caicci, F., Zaniolo, G., Burighel, P., 2004. Neurogenic and non-neurogenic placodes in ascidians. J. Exp. Zool. B Mol. Dev. Evol. 302, 483–504.

Marti-Solans, J., Ferrandez-Roldan, A., Godoy-Marin, H., Badia-Ramentol, J., Torres-Aguila, N.P., Rodriguez-Mari, A., Bouquet, J.M., Chourrout, D., Thompson, E.M., Albalat, R., Cañestro, C., 2015. Oikopleura dioica culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. Genesis 53, 183–193.

Mayer, Y., Czosnek, H., Zeelon, P.E., Yaffe, D., Nudel, U., 1984. Expression of the genes coding for the skeletal muscle and cardiac actions in the heart. Nucleic Acids Res. 12, 1087–1100.

Mikhaleva, Y., Kreneisz, O., Olsen, L.C., Glover, J.C., 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..

Müller, W., 1870. Beobachtungen des pathologischen Instituts zu Jena.

Munro, E.M., Odell, G., 2002a. Morphogenetic pattern formation during ascidian notochord formation is regulative and highly robust. Development 129, 1–12.

Munro, E.M., Odell, G.M., 2002b. Polarized basolateral cell motility underlies invagination and convergent extension of the ascidian notochord. Development 129, 13–24.

Nakajima, H., Tanoue, T., 2012. The circumferential actomyosin belt in epithelial cells is regulated by the Lulu2-p114RhoGEF system. Small GTPases 3, 91–96.

Nishida, H., 2008. Development of the appendicularian Oikopleura dioica: culture, genome, and cell lineages. Dev. Growth Differ. 50 (Suppl 1), S239–S256.

Nishino, A., Satou, Y., Morisawa, M., Satoh, N., 2000. Muscle actin genes and muscle cells in the appendicularian, Oikopleura longicauda: phylogenetic relationships among muscle tissues in the urochordates. J. Exp. Zool. 288, 135–150.

Olsson, R., 1975. Primitive coronet cells in the brain of Oikopleura (Appendicularia, Tunicata). Acta Zool. 56, 155–161.

Ordahl, C.P., 1986. The skeletal and cardiac alpha-actin genes are coexpressed in early embryonic striated muscle. Dev. Biol. 117, 488–492.

Perrin, B.J., Ervasti, J.M., 2010. The actin gene family: function follows isoform. Cytoskeleton 67, 630–634.

Philimonenko, V.V., Zhao, J., Iben, S., Dingova, H., Kysela, K., Kahle, M., Zentgraf, H., Hofmann, W.A., de Lanerolle, P., Hozak, P., Grummt, I., 2004. Nuclear actin and myosin I are required for RNA polymerase I transcription. Nat. Cell Biol. 6, 1165–1172.

Pollard, T.D., 2010. Mechanics of cytokinesis in eukaryotes. Curr. Opin. Cell Biol. 22, 50–56.

Quarmby, L., 2014. Cilia assembly: a role for F-actin in IFT recruitment. Curr. Biol. 24, R796–R798.

Ramkumar, N., Baum, B., 2016. Coupling changes in cell shape to chromosome segregation. Nat. Rev. Mol. Cell Biol. 17, 511–521.

Roblodowski, C., He, Q., 2017. Drosophila Dunc-115 mediates axon projection through actin binding. Invertebr. Neurosci. 17, 2.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Ruan, W., Lai, M., 2007. Actin, a reliable marker of internal control? Clin. Chim. Acta 385, 1–5.

Ruzicka, D.L., Schwartz, R.J., 1988. Sequential activation of alpha-actin genes during avian cardiogenesis: vascular smooth muscle alpha-actin gene transcripts mark the onset of cardiomyocyte differentiation. J. Cell Biol. 107, 2575–2586.

Satoh, N., 2016. Chapter 7 - The New Organizers Hypothesis for Chordate Origins, Chordate Origins and Evolution. Academic Press, San Diego, 97–120.

Satoh, N., Rokhsar, D., Nishikawa, T., 2014. Chordate evolution and the three-phylum system. Proc. Biol. Sci. 281, 20141729.

Schwartz, K., de la Bastie, D., Bouveret, P., Oliviero, P., Alonso, S., Buckingham, M., 1986. Alpha-skeletal muscle actin mRNA's accumulate in hypertrophied adult rat hearts. Circ. Res. 59, 551–555.

Schwayer, C., Sikora, M., Slovakova, J., Kardos, R., Heisenberg, C.P., 2016. Actin rings of power. Dev. Cell 37, 493–506.

Silva, R.C., Sattlegger, E., Castilho, B.A., 2016. Perturbations in actin dynamics reconfigure protein complexes that modulate GCN2 activity and promote an eIF2 response. J. Cell Sci. 129, 4521–4533.

Soviknes, A.M., Chourrout, D., 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.

Soviknes, A.M., Glover, J.C., 2007. Spatiotemporal patterns of neurogenesis in the appendicularian Oikopleura dioica. Dev. Biol. 311, 264–275.

Soviknes, A.M., Glover, J.C., 2008. Continued growth and cell proliferation into adulthood in the notochord of the appendicularian Oikopleura dioica. Biol. Bull. 214, 17–28.

Suzuki, M.M., Satoh, N., 2000. Genes expressed in the amphioxus notochord revealed by EST analysis. Dev. Biol. 224, 168–177.

Swalla, B.J., White, M.E., Zhou, J., Jeffery, W.R., 1994. Heterochronic expression of an adult muscle actin gene during ascidian larval development. Dev. Genet. 15, 51–63.

Takahashi, H., Hotta, K., Erives, A., Di Gregorio, A., Zeller, R.W., Levine, M., Satoh, N., 1999. Brachyury downstream notochord differentiation in the ascidian embryo. Genes Dev. 13, 1519–1523.

Tomlinson, C.R., Beach, R.L., Jeffery, W.R., 1987. Differential expression of a muscle actin gene in muscle cell lineages of ascidian embryos. Development 101, 751–765.

Tripathi, S.C., 1989. A possible role of actin in the mechanical control of the cell cycle. Biol. Cell 67, 351–353.

Urano, A., Suzuki, M.M., Zhang, P., Satoh, N., Satoh, G., 2003. Expression of muscle-related genes and two MyoD genes during amphioxus notochord development. Evol. Dev. 5, 447–458.

Vandekerckhove, J., Weber, K., 1978. At least six different actins are expressed in a higher mammal: an analysis based on the amino acid sequence of the amino-

terminal tryptic peptide. J. Mol. Biol. 126, 783–802.

Vandekerckhove, J., Weber, K., 1984. Chordate muscle actins differ distinctly from invertebrate muscle actins. The evolution of the different vertebrate muscle actins. J. Mol. Biol. 179, 391–413.

Vartiainen, M.K., Guettler, S., Larijani, B., Treisman, R., 2007. Nuclear actin regulates dynamic subcellular localization and activity of the SRF cofactor MAL. Science 316, 1749–1752.

Vicente-Manzanares, M., Ma, X., Adelstein, R.S., Horwitz, A.R., 2009. Non-muscle myosin II takes centre stage in cell adhesion and migration. Nat. Rev. Mol. Cell Biol. 10, 778–790.

Vidaki, M., Drees, F., Saxena, T., Lanslots, E., Taliaferro, M.J., Tatarakis, A., Burge, C.B., Wang, E.T., Gertler, F.B., 2017. A requirement for mena, an actin regulator, in local mRNA translation in developing neurons. Neuron 95, 608–622, (e605).

Woodcock-Mitchell, J., Mitchell, J.J., Low, R.B., Kieny, M., Sengel, P., Rubbia, L., Skalli, O., Jackson, B., Gabbiani, G., 1988. Alpha-smooth muscle actin is transiently expressed in embryonic rat cardiac and skeletal muscles. Differentiation 39, 161–166.

Zaidel-Bar, R., Zhenhuan, G., Luxenburg, C., 2015. The contractome–a systems view of actomyosin contractility in non-muscle cells. J. Cell Sci. 128, 2209–2217.

Zarnack, K., Feldbrugge, M., 2010. Microtubule-dependent mRNA transport in fungi. Eukaryot. Cell 9, 982–990.

41

CHAPTER I: SUPPLEMENTAY INFORMATION

Supplemetary File 1. Algorithm to calculate the Diagnostic aminoacid values (DAV) according to the conservation of diagnostic positions.

Summary of the excel file:

The first tab calculates the Matrix of DAVs per poition according to ascidian and vertebrate actins

The second tab calculates the DAV values for all analyzed sequences

The third tab shows a list of all chordate actins according to their DAVs from muscular to cytoplasmic

Correspondence between the 16 positions and those used to differentiate between muscular and citoplasmatic actins in chordates (in parenthesis) after Kusakabe et al., 1999

| 1 (5) | 2(6) | 3(10) | 4(16) | 5(17) | 6(76) | 7(103) | 8(129) | 9(162) | 10(176) | 11(201) | 12(260) | 13(267) | 14(287) | 15(297) | 16(365) |

Reference sequence: Bbe_BAA13444-BbCA1-Kusakabe99-CYTO

Kusakabe, R., Satoh, N., Holland, L.Z., Kasakabe, T., 1999. Genomic organization and evolution of actin genes in the amphioxus Branchiostoma belcheri and Branchiostoma floridae. Gene 227, 1-10

Matrix A: amino acids in the 16 positions of muscular (red) and cytoplasmic (blue) actins in olfactores
according to phylogenetic analysis used to calculate the frequency of each amino acid in each position

Name\positions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Hsa_ACTC1_ENSP0	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Hsa_ACTA1_ENSP0	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Olo_BAA86215_Oil	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Odi_Sc1.2_GSOIDT0	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Odi_Sc39_GSOIDT0	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Odi_Sc41_GSOIDT0	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Odi_Sc75_GSOIDT0	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Gga_ENSGALP0000	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Lch_acta1a_ENSLA	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Gga_ENSGALP0000	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Lch_ENSLACT0000	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Loc_zgc_86709_EN	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Loc_actc1a_ENSLO	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Lch_ENSLACT0000	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Hsa_ACTG2_ENSP0	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A
Hsa_ACTA2_ENSP0	S	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A
Cro_ENSCINP00000	Q	T	C	L	V	V	V	N	M	V	T	I	I	N	A	
Csa_ENSCSAVP000	Q	T	C	L	V	V	V	N	M	V	T	I	I	N	A	
Hro_CG.MTP2014.S	Q	T	C	L	V	V	V	N	M	V	T	I	I	N	A	
Pmamm.CG.MTP20	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Cro_ENSCINP00000	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Csa_ENSCSAVP000	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Pmamm.CG.MTP20	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Cro_ENSCINP00000	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Cro_GRAIL227-9-1_	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Pmamm.CG.MTP20	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Hro_CG.MTP2014.S	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Hro_D29014_HrMA	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Hro_D10887_MTP2	T	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Hro_CG.MTP2014.S	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Mocci.CG.ELv1_2.S	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Mocci.CG.ELv1_2.S	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Mocci.CG.ELv1_2.S	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Scl_X61040_ScTb1-	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Csa_ENSCSAVP000	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Mocul.CG.ELv1_2.S	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Mocul_D78190_Mc	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A
Gga_ACTG2_ENSGA	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A
Gga_ENSGALP0000	S	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A
Loc_ENSLOCT0000	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A
Xtr_Act3_NP_9890	T	T	C	L	V	I	T	V	N	Q	V	T	I	I	N	A
Hro_CG.MTP2014.S	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X
Hro_CG.MTP2014.S	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X
Hro_CG.MTP2014.S	T	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X
Hro_CG.MTP2014.S	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X
Hro_CG.MTP2014.S	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X
Lch_ENSLACT0000	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	X
Lch_ENSLACT0000	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	X
Mocul.CG.ELv1_2.S	Q	T	C	L	V	I	V	V	N	M	V	T	I	I	N	S
Mocul_L21915_Mc	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	S
Mocci.CG.ELv1_2.S	T	T	C	L	V	I	V	V	N	M	V	T	I	I	N	S
Mocci.CG.ELv1_2.S	Q	T	C	L	V	X	T	V	N	S	V	T	I	I	N	A
Psi_ENSPSIT00000	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A
Oan_ENSOANT000	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A
Loc_ENSLOCT0000	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A
Lch_ENSLACT0000	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A
Scl_X61042_SpMA1	Q	T	C	L	V	I	V	V	N	M	V	T	I	I	N	S
Bsc_CAX48983_BsN	Q	T	C	L	V	I	V	V	N	M	V	T	V	I	N	S
Bsc_botctg010875g	Q	T	C	L	V	I	V	V	N	M	V	T	V	I	N	S
Mocul.CG.ELv1_2.S	Q	T	C	L	V	I	T	T	N	A	V	T	I	I	N	S
Gga_ENSGALT0000	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	S
Mocci.CG.ELv1_2.S	X	X	X	X	X	T	V	N	S	V	T	I	I	N	A	
Pmamm.CG.MTP20	V	P	V	M	C	V	T	T	X	I	T	T	I	V	X	S
Odi_Sc50_GSOIDT0	V	A	I	M	C	V	T	T	T	I	T	T	I	V	T	S
Pmamm.CG.MTP20	I	T	V	M	C	V	I	T	T	I	T	T	I	V	T	S
Csa_ENSCSAVP000	V	A	V	M	C	V	V	T	L	T	T	I	V	T	S	
Cro_ENSCINP00000	V	A	V	M	C	V	V	T	I	T	T	I	V	T	S	
Csa_ENSCSAVP000	T	V	V	M	C	V	V	A	T	I	T	V	I	V	T	S
Mocci.CG.ELv1_2.S	I	T	I	M	C	V	V	T	T	L	T	A	L	V	A	A
Pmamm.CG.MTP20	T	T	V	M	C	V	I	T	T	L	T	A	L	V	T	S
Pmamm.CG.MTP20	S	P	V	M	C	V	V	T	T	M	T	A	L	V	T	S
Bsc_botctg025474g	E	A	V	M	C	V	C	S	T	I	T	T	I	V	T	S
Bsc_botctg025474g	V	A	V	M	C	V	C	T	T	I	T	T	I	V	T	S
Hsa_ACTBL2_ENSP0	L	S	V	M	C	V	I	T	I	L	T	A	L	V	T	A
Bsc_CAX48982_Deg	S	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Bsc_botctg033661g	S	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Bsc_botctg073109g	S	S	V	M	C	V	C	S	T	L	T	A	L	V	T	S
Bsc_botctg087748g	V	A	V	M	C	V	C	A	T	I	T	A	I	V	T	S
Mocul.CG.ELv1_2.S	V	T	I	M	C	V	V	T	T	L	T	T	I	V	T	S
Pmamm.CG.MTP20	T	A	V	M	C	V	V	C	T	L	T	A	L	V	T	S
Bsc_botctg031514g	X	X	X	X	X	V	C	A	T	I	T	A	I	V	T	S
Hsa_ACTB_ENSP00	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Olo_BAA86216_Oil	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Odi_Sc1.1_GSOIDT0	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Odi_Sc49_GSOIDT0	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Hsa_ACTG1_ENSP0	I	A	I	M	C	V	V	T	T	L	T	A	L	V	T	S
Gga_ENSGALP0000	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Gga_ACTB_ENSGAL	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Gga_ENSGALP0000	I	A	I	M	C	V	V	T	T	L	T	A	L	V	T	S
Lch_ACTB_ENSLACI	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Pma_ENSPMAT000	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Pma_ENSPMAT000	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Mocul.CG.ELv1_2.S	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Bsc_botctg021406_	X	X	X	X	X	V	C	S	T	L	T	A	L	V	T	S
Cro_ENSCINP00000	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Csa_ENSCSAVP000	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Cro_GRAIL176-22-1	V	A	V	M	C	V	I	T	T	L	T	A	L	V	T	S
Hro_CG.MTP2014.S	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Mocci.CG.ELv1_2.S	V	A	V	M	C	V	V	A	T	S	T	A	L	V	T	S
Pmamm.CG.MTP20	A	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Hro_CG.MTP2014.S	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S
Cro_GRAIL2-31-1_N	I	A	I	M	C	V	V	A	T	L	T	A	L	V	T	S
Pmamm.CG.MTP20	V	A	I	M	C	V	V	T	T	L	T	A	L	V	T	S
Mocci.CG.ELv1_2.S	I	S	V	M	C	V	V	X	T	L	T	S	L	V	T	S

Count of amino acid for each position in Matrix A for muscular actins

Muscular sequences (62 sequences)   *threshold value (to be present in at least 20% of the sequences)*   12,4

24T	61T	61C	61L	49V	60I	47T	61V	62N	36M	62V	62T	60I	62I	62N	47A
35Q	X	X	X	12C	X	15V	1T		23A			2V			X
2S				X					1Q						8S
X									2S						

Select amino acid N: 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 62 | 0 | 0 | 0 | 0 | 0 | 62 | 0 | Counter of aminoacid selected in S11

Count of amino acid for each position in Matrix A for cytoplasmic actins

Cytoplasmic (42 sequences)   *threshold value (to be present in at least 20% of the sequences)*   8,4

18V	2P	33V	40M	40C	42V	2T	32T	X	9I	42T	7T	10I	42V	X	40S
12I	30A	7I	X	X		4I	5A	40T	31L		1V	32L		40T	2A
3T	4T	X				30V	3S	1I	1M		33A			1A	
4S	1V					6C	1C		1S		1S				
1E	3S						X								
1L	X														
X															
1A															

Select amino acid N: 0 | ... | 0 | Counter of aminoacid selected in S23

Matrix of DAVs per position:
If the frequency of an aa in a given position is greater than 20% (12.4 for muscular actin and 8.4 for cytoplasmic), the DAV per position is +1 for muscular and -1 for cytoplasmic
If the aa is also present in the other group, the DAV is corrected by subtracting the frequency of the aa in the other group
   aa muscular: 1-(count of the aa in cytoplasmic/42); aa cytoplasmic: -1+(count of the aa in muscular/62)
If the aa frequency in a given position is lower than 20%, the DAV per position is zero

	Pos1	Pos2	Pos3	Pos4	Pos5	Pos6	Pos7	Pos8	Pos9	Pos10	Pos11	Pos12	Pos13	Pos14	Pos15	Pos16
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,87
T	0,929	0,905	0	0	0	0	0,95	-0,98	-1	0	-1	0,833	0	0	-1	0
N	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Q	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	1	0	-0,806	0	0	0	0	0	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	-1	0	0	0	0	0	0	0	1	0	-1	0	0	0	0,952
I	-1	0	0	0	0	1	0	0	0	-1	0	0	0,762	1	0	0
L	0	0	0	1	0	0	0	0	0	-1	0	0	-1	0	0	0
M	0	0	0	-1	0	0	0	0	0	0,976	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V	-1	0	-1	0	1	-1	-0,758	1	0	0	0	1	0	0	-1	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

43

Summation of total DAVs for all chordate analyzed sequences

Name\positions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	al DAVs
Hsa_ACTC1_ENSP	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,31
Hsa_ACTA1_ENSP	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,31
Olo_BAA86215_C	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,38
Odi_Sc1.2_GSOID	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,38
Odi_Sc39_GSOID	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,38
Odi_Sc41_GSOID	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,38
Odi_Sc75_GSOID	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,38
Hsa_ACTG2_ENSP	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	13,50
Hsa_ACTA2_ENSP	S	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A	12,57
Hsa_ACTB_ENSP0	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Olo_BAA86216_C	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Odi_Sc1.1_GSOID	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Odi_Sc49_GSOID	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Hsa_ACTG1_ENSP	I	A	I	M	C	V	V	T	T	L	T	A	L	V	T	S	-14,42
Hsa_ACTBL2_ENS	L	S	V	M	C	V	I	T	I	L	T	A	L	V	T	A	-9,84
Gga_ENSGALP00	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Gga_ACTB_ENSG	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Gga_ENSGALP00	I	A	I	M	C	V	V	T	T	L	T	A	L	V	T	S	-14,42
Lch_ACTB_ENSLA	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Pma_ENSPMAT00	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Pma_ENSPMAT00	I	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Bbe_BAA13445_E	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,86
Bbe_BAA13351_K	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,86
Bfl_XP_00261079	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,86
Bfl_XP_00261079	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,86
Bfl_XP_00261080	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,86
Bla_BL18017_even	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,86
Bla_BL24512_even	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,86
Bfl_XP_00261080	A	T	C	L	V	V	T	S	T	L	V	A	L	I	N	S	1,99
Bfl_XP_00261080	A	T	V	L	V	V	S	A	V	L	V	A	L	I	N	A	1,86
Bfl_XP_00261080	A	T	V	L	V	V	S	A	V	L	V	A	L	I	N	A	1,86
Bfl_XP_00259649	A	T	V	M	C	V	T	T	V	L	I	A	L	I	T	C	-5,93
Bla_BL08899_even	A	T	V	M	C	V	T	T	V	L	I	A	L	I	T	S	-6,80
Bfl_XP_00260193	A	T	C	M	C	V	T	S	T	L	V	A	L	I	T	S	-3,82
Bla_BL03268_even	A	T	C	M	C	V	T	S	T	L	V	A	L	I	T	S	-3,82
Bla_BL19911_even	A	T	C	L	V	V	T	S	V	L	T	A	L	I	I	S	-0,01
Bbe_BRBE02586	A	V	M	C	V	V	C	T	L	V	T	I	I	I	T	S	-6,84
Bbe_BRBE02587	V	A	I	M	C	V	V	T	T	L	V	T	I	I	I	S	-6,82
Bla_BL19005_even	V	A	I	M	C	V	V	T	T	L	V	T	I	I	I	S	-6,82
Bbe_BRBE24941	V	A	V	M	C	V	V	S	T	L	T	S	L	I	T	S	-11,44
Bfl_XP_00260607	V	A	V	M	C	V	V	S	T	L	T	S	L	I	T	S	-11,44
Bla_BL10072_even	V	A	V	M	C	V	V	S	T	L	T	S	L	I	T	S	-11,44
Bbe_BAA13444_B	V	A	V	M	C	V	V	S	T	L	T	S	L	I	T	S	-11,44
Mocul.CG.ELv1_2	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Bsc_CAX48982_D	S	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-14,42
Bsc_botctg03366	S	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-14,42
Bsc_botctg02140	X	X	X	X	V	C	S	T	L	V	A	L	V	T	S		-8,87
Bsc_botctg07310	S	S	V	M	C	V	C	S	T	L	T	A	L	V	T	S	-11,68
Cro_ENSCINP000	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Csa_ENSCSAVP00	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Cro_GRAIL176-22	V	A	V	M	C	V	I	T	T	L	T	A	L	V	T	S	-14,66
Csa_ENSCSAVP00	V	A	V	M	C	V	V	T	T	L	T	T	I	V	T	S	-11,82
Hro_CG.MTP2014	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Mocci.CG.ELv1_2	V	A	V	M	C	V	V	A	T	S	T	A	L	V	T	S	-13,44
Cro_ENSCINP000	V	A	V	M	C	V	V	T	I	T	T	I	V	T	S		-11,82
Csa_ENSCSAVP00	T	V	V	M	C	V	V	A	T	V	T	V	I	V	T	S	-8,75
Odi_Sc50_GSOID	V	A	I	M	C	V	T	T	T	I	T	T	I	V	T	S	-9,11
Pmamm.CG.MTP2	A	V	A	M	C	V	V	T	T	L	T	A	L	V	T	S	-14,42
Bsc_botctg03151	X	X	X	X	V	C	A	T	I	T	A	I	V	T	S		-7,11
Bsc_botctg08774	V	A	V	M	C	V	C	A	T	I	T	A	I	V	T	S	-11,92
Hro_CG.MTP2014	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-15,42
Bbe_BRBE20270	V	A	V	M	C	V	V	S	T	L	T	A	L	I	A	S	-11,44
Bfl_XP_00259201	V	A	V	M	C	V	V	S	T	I	T	A	L	I	I	S	-12,44
Bfl_XP_00259201	V	A	V	M	C	V	V	S	T	L	T	T	I	I	S	S	-7,84
Bla_BL95821_even	V	A	V	M	C	V	V	S	T	L	T	T	I	I	T	S	-8,84
Bla_BL95820_even	V	A	V	M	C	V	V	S	T	L	T	T	I	I	T	S	-8,84
Bbe_BRBE02588	V	A	V	M	C	I	V	T	T	M	V	T	I	I	T	S	-3,85
Bfl_XP_00259015	V	A	V	M	C	I	V	T	T	M	V	T	I	I	T	S	-3,85
Bla_BL19846_even	V	A	V	M	C	I	V	T	T	M	V	T	I	I	T	S	-3,85
Bbe_BRBE21587	V	A	V	M	C	V	T	N	L	V	T	I	I	I	N	S	-3,82
Bfl_XP_00259052	V	A	V	M	C	V	T	N	L	V	T	I	I	I	N	S	-3,82
Bla_BL08383_even	V	A	V	M	C	V	T	N	L	V	T	I	I	I	N	S	-3,82
Bla_BL25564_even	V	A	V	M	C	V	T	N	L	V	T	I	I	I	N	S	-3,82
Bfl_XP_00260313	V	A	V	M	C	V	T	N	L	V	T	I	I	I	N	S	-3,82
Cro_GRAIL2-31-1	I	A	I	M	C	V	V	A	T	L	T	A	L	V	T	S	-13,44
Pmamm.CG.MTP2	V	A	V	M	C	V	V	T	T	L	T	A	L	V	T	S	-14,42
Mocci.CG.ELv1_2	I	T	I	M	C	V	V	T	T	L	T	A	L	V	A	A	-9,69
Mocul.CG.ELv1_2	V	T	I	M	C	V	V	T	T	L	T	A	L	V	T	S	-12,51
Pmamm.CG.MTP2	T	T	V	M	C	V	I	T	T	I	T	T	I	V	T	S	-10,83
Pmamm.CG.MTP2	I	T	V	M	C	V	I	T	T	I	T	T	I	V	T	S	-9,16
Bsc_CAX48983_B	Q	T	C	L	V	I	V	V	N	M	V	T	V	I	N	S	11,09
Bsc_botctg01087	Q	T	C	L	V	I	V	V	N	M	V	T	V	I	N	S	11,09
Cro_ENSCINP000	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	13,67
Csa_ENSCSAVP00	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	13,67
Hro_CG.MTP2014	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	13,67
Mocul.CG.ELv1_2	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	S	11,85
Pmamm.CG.MTP2	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	13,67
Mocul_L21915_N	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	S	11,85
Mocci.CG.ELv1_2	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	S	11,78
Cro_ENSCINP000	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Csa_ENSCSAVP00	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Pmamm.CG.MTP2	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40

DAVs per position for all chordate analyzed sequences

Name\positions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Hsa_ACTC1_EN	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Hsa_ACTA1_EN	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Olo_BAA86215_	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Odi_Sc1.2_GSO	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Odi_Sc39_GSOI	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Odi_Sc41_GSOI	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Odi_Sc75_GSOI	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Hsa_ACTG2_EN	0,93	0,90	1,00	1,00	-0,81	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Hsa_ACTA2_EN	0,00	0,90	1,00	1,00	-0,81	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Hsa_ACTB_ENS	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Olo_BAA86216_	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Odi_Sc1.1_GSO	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Odi_Sc49_GSOI	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Hsa_ACTG1_EN	-1,00	-1,00	0,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Hsa_ACTBL2_EN	0,00	0,00	0,00	-1,00	-0,81	-1,00	0,00	-0,98	0,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	0,95
Gga_ENSGALP0	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Gga_ACTB_ENS	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Gga_ENSGALP0	-1,00	-1,00	0,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Lch_ACTB_ENSL	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Pma_ENSPMAT	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Pma_ENSPMAT	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Bbe_BAA13445_	0,00	0,90	1,00	1,00	1,00	1,00	0,00	0,00	-1,00	-1,00	0,00	-1,00	-1,00	1,00	1,00	0,95
Bbe_BAA13351_	0,00	0,90	1,00	1,00	1,00	1,00	0,00	0,00	-1,00	-1,00	0,00	-1,00	-1,00	1,00	1,00	0,95
Bfl_XP_0026107	0,00	0,90	1,00	1,00	1,00	1,00	0,00	0,00	-1,00	-1,00	0,00	-1,00	-1,00	1,00	1,00	0,95
Bfl_XP_0026107	0,00	0,90	1,00	1,00	1,00	1,00	0,00	0,00	-1,00	-1,00	0,00	-1,00	-1,00	1,00	1,00	0,95
Bfl_XP_0026108	0,00	0,90	1,00	1,00	1,00	1,00	0,00	0,00	-1,00	-1,00	0,00	-1,00	-1,00	1,00	1,00	0,95
Bla_BL18017_ev	0,00	0,90	1,00	1,00	1,00	1,00	0,00	0,00	-1,00	-1,00	0,00	-1,00	-1,00	1,00	1,00	0,95
Bla_BL24512_ev	0,00	0,90	1,00	1,00	1,00	1,00	0,00	0,00	-1,00	-1,00	0,00	-1,00	-1,00	1,00	1,00	0,95
Bfl_XP_0026108	0,00	0,90	1,00	1,00	1,00	1,00	0,95	0,00	-1,00	-1,00	0,00	-1,00	-1,00	1,00	1,00	-0,87
Bfl_XP_0026108	0,00	0,90	-1,00	1,00	1,00	1,00	0,00	0,00	0,00	-1,00	0,00	-1,00	-1,00	1,00	1,00	0,95
Bfl_XP_0026108	0,00	0,90	-1,00	1,00	1,00	1,00	0,00	0,00	0,00	-1,00	0,00	-1,00	-1,00	1,00	1,00	0,95
Bfl_XP_0025964	0,00	0,90	-1,00	-1,00	-0,81	0,95	-0,98	-1,00	0,00	0,00	-1,00	-1,00	-1,00	1,00	-1,00	0,00
Bla_BL08899_e	0,00	0,90	-1,00	-1,00	-0,81	1,00	0,95	-0,98	0,00	-1,00	-1,00	-1,00	-1,00	1,00	-1,00	-0,87
Bfl_XP_0026019	0,00	0,90	1,00	-1,00	-0,81	1,00	0,95	0,00	-1,00	-1,00	0,00	-1,00	-1,00	1,00	-1,00	-0,87
Bla_BL03268_e	0,00	0,90	1,00	-1,00	-0,81	1,00	0,95	0,00	-1,00	-1,00	0,00	-1,00	-1,00	1,00	-1,00	-0,87
Bla_BL19911_e	0,00	0,90	1,00	1,00	1,00	1,00	0,95	0,00	0,00	-1,00	-1,00	-1,00	-1,00	1,00	0,00	-0,87
Bbe_BRBE0258	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	1,00	0,83	0,76	1,00	-1,00	-0,87
Bbe_BRBE0258	-1,00	-1,00	0,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	1,00	0,83	0,76	1,00	-1,00	-0,87
Bla_BL19005_e	-1,00	-1,00	0,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	1,00	0,83	0,76	1,00	-1,00	-0,87
Bbe_BRBE2494	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	-1,00	-1,00	1,00	-1,00	-0,87
Bfl_XP_0026060	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	-1,00	-1,00	1,00	-1,00	-0,87
Bla_BL10072_e	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	-1,00	-1,00	1,00	-1,00	-0,87
Bbe_BAA13444	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	-1,00	-1,00	1,00	-1,00	-0,87
Mocul.CG.ELv1_	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Bsc_CAX48982_	0,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Bsc_botctg0336	0,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Bsc_botctg0214	0,00	0,00	0,00	0,00	-0,81	-1,00	0,00	0,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Bsc_botctg0731	0,00	0,00	-1,00	-1,00	-0,81	-1,00	0,00	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Cro_ENSCINP00	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Csa_ENSCSAVP	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Cro_GRAIL176-	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	0,00	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Csa_ENSCSAVP	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	0,83	0,76	-1,00	-1,00	-0,87
Hro_CG.MTP20	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Mocci.CG.ELv1_	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	0,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Cro_ENSCINP00	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	0,00	-1,00	0,83	0,76	-1,00	-1,00	-0,87
Csa_ENSCSAVP	0,93	0,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	0,00	-1,00	0,76	-1,00	-1,00	-1,00	-0,87
Odi_Sc50_GSOI	-1,00	-1,00	0,00	-1,00	-1,00	-0,81	0,95	-0,98	0,00	-1,00	-1,00	0,83	0,76	-1,00	-1,00	-0,87
Pmamm.CG.MT	-1,00	-1,00	0,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Bsc_botctg0315	0,00	0,00	0,00	0,00	-0,81	-1,00	0,00	0,00	0,00	0,00	0,76	-1,00	-1,00	-1,00		-0,87
Bsc_botctg0877	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	0,00	0,00	0,00	0,00	0,76	-1,00	-1,00	-1,00	-1,00	-0,87
Hro_CG.MTP20	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Bbe_BRBE2027	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	-1,00	-1,00	0,00	0,00	-0,87
Bfl_XP_002592	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	0,00	-1,00	-1,00	-1,00	0,00	0,00	-0,87
Bfl_XP_002592	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	0,83	0,76	0,00	0,00	-0,87
Bla_BL95821_e	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	0,83	0,76	1,00	-1,00	-0,87
Bla_BL95820_e	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	0,83	0,76	1,00	-1,00	-0,87
Bbe_BRBE0258	-1,00	-1,00	-1,00	-1,00	-0,81	1,00	-0,76	-0,98	-1,00	0,98	0,00	0,83	0,76	1,00	-1,00	-0,87
Bfl_XP_002590	-1,00	-1,00	-1,00	-1,00	-0,81	1,00	-0,76	-0,98	-1,00	0,98	0,00	0,83	0,76	1,00	-1,00	-0,87
Bla_BL19846_e	-1,00	-1,00	-1,00	-1,00	-0,81	1,00	-0,76	-0,98	-1,00	0,98	0,00	0,83	0,76	1,00	-1,00	-0,87
Bbe_BRBE2158	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	0,00	-1,00	0,83	0,76	1,00	1,00	-0,87
Bfl_XP_002590	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	0,00	-1,00	0,83	0,76	1,00	1,00	-0,87
Bla_BL08383_e	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	0,00	-1,00	0,83	0,76	1,00	1,00	-0,87
Bla_BL25564_e	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	0,00	-1,00	0,83	0,76	1,00	1,00	-0,87
Bfl_XP_002603	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	0,00	-1,00	0,83	0,76	1,00	1,00	-0,87
Cro_GRAIL2-31-	-1,00	-1,00	0,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Pmamm.CG.MT	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Mocci.CG.ELv1_	-1,00	0,90	0,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	0,00	0,95
Mocul.CG.ELv1_	-1,00	0,90	0,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Pmamm.CG.MT	0,93	0,90	-1,00	-1,00	-0,81	-1,00	0,00	-0,98	-1,00	0,00	-1,00	0,83	0,76	-1,00	-1,00	-0,87
Pmamm.CG.MT	-1,00	0,90	-1,00	-1,00	-0,81	-1,00	0,00	-0,98	-1,00	0,00	-1,00	0,83	0,76	-1,00	-1,00	-0,87
Bsc_CAX48983_	1,00	0,90	1,00	1,00	1,00	1,00	-0,76	1,00	1,00	0,98	1,00	0,83	0,00	1,00	1,00	-0,87
Bsc_botctg0108	1,00	0,90	1,00	1,00	1,00	1,00	-0,76	1,00	1,00	0,98	1,00	0,83	0,00	1,00	1,00	-0,87
Cro_ENSCINP00	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Csa_ENSCSAVP	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Hro_CG.MTP20	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Mocul.CG.ELv1	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	-0,87
Pmamm.CG.MT	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Mocul_L21915_	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	-0,87
Mocci.CG.ELv1_	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	-0,87
Cro_ENSCINP00	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Csa_ENSCSAVP	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Pmamm.CG.MT	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95

Left alignment table

Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Score
Cro_ENSCINP000	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Cro_GRAIL227-9-	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Pmamm.CG.MTP.	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Hro_CG.MTP2014	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X	14,45
Hro_CG.MTP2014	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Hro_D29014_HrN	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Hro_CG.MTP2014	T	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X	14,38
Hro_CG.MTP2014	T	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X	14,38
Hro_D10887_MTI	T	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,33
Hro_CG.MTP2014	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X	14,45
Hro_CG.MTP2014	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Mocci.CG.ELv1_2	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Mocci.CG.ELv1_2	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Mocci.CG.ELv1_2	X	X	X	X	X	X	T	V	N	S	V	T	I	I	N	A	8,50
Mocci.CG.ELv1_2	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Scl_X61040_ScTb	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Csa_ENSCSAVP00	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Mocul.CG.ELv1_2	Q	T	C	L	V	I	T	T	N	A	V	T	I	I	N	S	11,60
Mocul.CG.ELv1_2	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Mocul_D78190_N	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,40
Mocci.CG.ELv1_2	Q	T	C	L	V	X	T	V	N	S	V	T	I	I	N	A	13,40
Gga_ACTG2_ENS	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A	13,50
Gga_ENSGALP00	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,31
Lch_acta1a_ENSL	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,31
Gga_ENSGALP00	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,31
Lch_ACTC1_ENSL	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,31
Loc_zgc_86709_E	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,31
Loc_actc1a_ENSL	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,31
Gga_ENSGALP00	S	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A	12,57
Lch_ENSLACT000	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	X	13,50
Loc_ENSLOCT000	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	X	13,50
Lch_ENSLACT000	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	X	12,55
Lch_ENSLACT000	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,31
Xtr_Act3_NP_989	T	T	C	L	V	I	T	V	N	Q	V	T	I	I	N	A	14,33
Gga_ENSGALT00	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	S	9,97
Psi_ENSPSIT0000	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A	11,79
Oan_ENSOANT00	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A	11,79
Loc_ENSLOCT000	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A	11,79
Lch_ENSLACT000	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A	11,79
Scl_X61042_SpM	Q	T	C	L	V	I	V	V	N	M	V	T	I	I	N	S	11,85
Bbe_BRBE213730	T	T	C	L	C	V	C	V	T	M	V	A	L	I	N	S	3,13
Bfl_XP_00261080	T	T	C	L	C	V	C	V	T	M	V	A	L	I	N	S	3,13
Bbe_BRBE213760	T	T	V	L	V	V	T	V	T	M	V	A	L	I	N	S	3,89
Bfl_XP_00261080	T	T	V	L	V	V	T	V	T	M	V	A	L	I	N	S	3,89
Pmamm.CG.MTP.	S	V	M	C	V	V	T	T	M	T	A	L	V	T	S		-11,44
Mocci.CG.ELv1_2	I	S	V	M	C	V	V	X	T	L	T	S	L	V	T	S	-12,44
Pmamm.CG.MTP.	T	A	V	M	C	V	V	C	T	L	T	A	L	V	T	S	-12,51
Pmamm.CG.MTP.	P	V	M	C	V	T	T	X	I	T	I	V	X	S			-7,11
Bfl_XP_00261080	A	T	I	L	C	V	I	A	A	M	X	T	I	N	S		3,80
Bla_BL10644_evn	A	T	I	L	C	V	I	A	A	M	V	T	I	I	N	S	4,80
Bsc_botctg02547	E	A	V	M	C	V	C	S	T	I	T	I	V	V	T	S	-9,08
Bsc_botctg02547	V	A	V	M	C	V	C	T	T	I	T	I	V	T	S		-11,07
Bfl_XP_00258800	D	I	V	M	C	V	V	S	T	E	V	T	I	I	T	S	-3,84
Bla_BL11551_evn	D	I	V	M	C	V	V	S	T	E	S	T	I	M	T	S	-5,84
Bfl_XP_00259016	X	X	X	X	V	I	S	T	E	A	A	I	V	S			-2,11
Bla_BL06827_evn	V	S	V	M	C	V	I	S	T	Q	V	A	I	I	I	S	-4,92

Right score table

Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Cro_ENSCINP00	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Cro_GRAIL227-9	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Pmamm.CG.MT	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Hro_CG.MTP20	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,00
Hro_CG.MTP20	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Hro_D29014_H	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Hro_CG.MTP20	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,00
Hro_CG.MTP20	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,00
Hro_D10887_M	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Hro_CG.MTP20	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,00
Hro_CG.MTP20	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Mocci.CG.ELv1_	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Mocci.CG.ELv1_	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Mocci.CG.ELv1_	0,00	0,00	0,00	0,00	0,00	0,00	0,95	1,00	1,00	0,00	1,00	0,83	0,76	1,00	1,00	0,95
Mocci.CG.ELv1_	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Scl_X61040_ScT	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Csa_ENSCSAVP	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Mocul.CG.ELv1_	1,00	0,90	1,00	1,00	1,00	1,00	0,95	-0,98	1,00	1,00	1,00	0,83	0,76	1,00	1,00	-0,87
Mocul.CG.ELv1_	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Mocul_D78190_	1,00	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	0,83	0,76	1,00	1,00	0,95
Mocci.CG.ELv1_	1,00	0,90	1,00	0,00	1,00	0,00	0,95	1,00	1,00	0,00	1,00	0,83	0,76	1,00	1,00	0,95
Gga_ACTG2_EN	0,93	0,90	1,00	1,00	-0,81	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Gga_ENSGALP0	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Lch_acta1a_ENS	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Gga_ENSGALP0	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Lch_ACTC1_ENS	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Loc_zgc_86709	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Loc_actc1a_ENS	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Gga_ENSGALP0	0,00	0,90	1,00	1,00	-0,81	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Lch_ENSLACT00	0,93	0,90	1,00	1,00	-0,81	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,00
Loc_ENSLOCT00	0,93	0,90	1,00	1,00	-0,81	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Lch_ENSLACT00	0,93	0,90	1,00	1,00	-0,81	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,00
Lch_ENSLACT00	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Xtr_Act3_NP_9	0,93	0,90	1,00	1,00	1,00	1,00	0,95	1,00	0,00	1,00	0,83	0,76	1,00	1,00	0,95	
Gga_ENSGALT0	0,93	0,90	1,00	1,00	-0,81	1,00	-0,76	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	-0,87
Psi_ENSPSIT000	0,93	0,90	1,00	1,00	-0,81	1,00	-0,76	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Oan_ENSOANT0	0,93	0,90	1,00	1,00	-0,81	1,00	-0,76	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Loc_ENSLOCT00	0,93	0,90	1,00	1,00	-0,81	1,00	-0,76	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Lch_ENSLACT00	0,93	0,90	1,00	1,00	-0,81	1,00	-0,76	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	0,95
Scl_X61042_Sp	1,00	0,90	1,00	1,00	1,00	1,00	-0,76	1,00	1,00	0,98	1,00	0,83	0,76	1,00	1,00	-0,87
Bbe_BRBE2137	0,93	0,90	1,00	1,00	-0,81	-1,00	0,00	1,00	-1,00	0,98	1,00	-1,00	-1,00	1,00	1,00	-0,87
Bfl_XP_0026108	0,93	0,90	1,00	1,00	-0,81	-1,00	0,00	1,00	-1,00	0,98	1,00	-1,00	-1,00	1,00	1,00	-0,87
Bbe_BRBE2137	0,93	0,90	-1,00	1,00	1,00	-1,00	0,95	1,00	-1,00	0,98	1,00	-1,00	-1,00	1,00	1,00	-0,87
Bfl_XP_0026108	0,93	0,90	-1,00	1,00	1,00	-1,00	0,95	1,00	-1,00	0,98	1,00	-1,00	-1,00	1,00	1,00	-0,87
Pmamm.CG.MT	0,93	0,00	-1,00	-1,00	-0,81	-1,00	-0,76	-0,98	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Mocci.CG.ELv1_	-1,00	0,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	0,00	-1,00	-1,00	-1,00	-0,87
Pmamm.CG.MT	0,93	-1,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-0,87
Pmamm.CG.MT	-1,00	0,00	-1,00	-1,00	-0,81	1,00	0,95	-0,98	0,00	-1,00	0,83	0,76	0,00	1,00	0,00	-0,87
Bfl_XP_0026108	0,00	0,90	1,00	1,00	-0,81	-1,00	0,00	0,00	-1,00	0,98	0,00	0,83	0,76	1,00	-1,00	-0,87
Bla_BL10644_e	0,00	0,90	0,00	1,00	-0,81	-1,00	0,00	0,00	-1,00	0,98	1,00	0,83	0,76	1,00	1,00	-0,87
Bsc_botctg0254	-1,00	0,00	-1,00	-1,00	-0,81	-1,00	0,00	0,00	-1,00	-1,00	-1,00	0,83	0,76	-1,00	-1,00	-0,87
Bsc_botctg0254	-1,00	-1,00	-1,00	-1,00	-0,81	-1,00	0,00	-0,98	-1,00	-1,00	-1,00	0,83	0,76	-1,00	-1,00	-0,87
Bfl_XP_0025880	0,00	0,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	0,00	1,00	0,83	0,76	1,00	-1,00	-0,87
Bla_BL11551_e	0,00	0,00	-1,00	-1,00	-0,81	-1,00	-0,76	0,00	-1,00	0,00	0,00	0,83	0,76	0,00	-1,00	-0,87
Bfl_XP_0025901	0,00	0,00	0,00	0,00	-1,00	-1,00	0,76	0,00	-1,00	0,00	0,00	-1,00	0,76	1,00	0,00	-0,87
Bla_BL06827_e	-1,00	0,00	-1,00	-1,00	-0,81	-1,00	0,00	0,00	-1,00	0,00	1,00	-1,00	0,76	1,00	0,00	-0,87

Matrix of DAVs per position: (mirror)

	Pos1	Pos2	Pos3	Pos4	Pos5	Pos6	Pos7	Pos8	Pos9	Pos10	Pos11	Pos12	Pos13	Pos14	Pos15	Pos16
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,87
T	0,929	0,905	0	0	0	0	0,95	-0,98	-1	0	-1	0,833	0	0	-1	0
N	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
Q	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	1	0	-0,81	0	0	0	0	0	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	-1	0	0	0	0	0	0	0	1	0	-1	0,70	0	0	0,952
I	-1	0	0	0	0	1	0	0	0	-1	0	0	0,762	1	0	0
L	0	0	0	1	0	0	0	0	0	-1	0	0	-1	0	0	0
M	0	0	0	-1	0	0	0	0	0,976	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V	-1	0	-1	0	1	-1	-0,76	1	0	0	1	0	0	-1	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

List of chordate actins sorted from greater to smaller DAVs

name\positions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Color coded DAVs
Cro_ENSCINP00000002686-ENSCINP00000012660-ENSCINP00000014519-muscle	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Csa_ENSCSAVP00000006787-ENSCSAVP00000006981-ENSCSAVP00000007718ENSCSAVP00000013947_muscle	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Pmamm.CG.MTP2014.S1453.g14830/g14833/g14834-S439.g08747-S1024.g13313.01.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Cro_ENSCINP00000014691_muscle	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Cro_GRAIL227-9-1_Ci-MA3-embryonicLarval	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Pmamm.CG.MTP2014.S1024.g13315/g13314.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Hro_CG.MTP2014.S1.g06718/g06881.01.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Hro_D29014_HrMA1a-Kusakabe95-larva	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Hro_CG.MTP2014.S348.g09995.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Mocci.CG.ELv1_2.S115016.g01801.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Mocci.CG.ELv1_2.S466348.g15701.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Mocci.CG.ELv1_2.S289045.g07141.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Scl_X61040_ScTb1-Kovilur93-larva	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Csa_ENSCSAVP00000014706_muscle-ACT	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Mocul.CG.ELv1_2.S65459.g04258-S74264.g05155.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Mocul_D78190_MocuMA1-Kusakabe96-larva	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,4047619
Olo_BAA86215_OilMA1-BAA86217_OilMA2-Nishino00_muscle	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3809524
Odi_Sc1.2_GSOIDT00000756001_muscle	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3809524
Odi_Sc39_GSOIDT00011141001_muscle	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3809524
Odi_Sc41_GSOIDT00012400001*_muscle	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3809524
Odi_Sc75_GSOIDT00016817001_muscle	Q	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3809524
Hro_D10887_MTP2014.S575.g07280.01.p_MTP2014.S575.g13659.01.p_HrMA4a-Kuwakabe92-larva	T	T	C	L	V	I	T	V	N	A	V	T	I	I	N	A	15,3333333
Hsa_ACTC1_ENSP00000290378_actin-alpha-cardiac-muscle-1	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3095238
Hsa_ACTA1_ENSP00000355645_actin-alpha-1-skeletal-muscle	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3095238
Gga_ENSGALP00000018042_Actin-alpha-skeletal-muscle	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3095238
Lch_acta1a_ENSLACP00000013687	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3095238
Gga_ENSGALP00000015988_Actin-alpha-cardiac-muscle-1	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3095238
Lch_ACTC1_ENSLACP00000013402	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3095238
Loc_zgc_86709_ENSLOCP00000014975	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3095238
Loc_actc1a_ENSLOCP00000015557-ENSLOCT00000019072_ENSLOCP00000019040	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3095238
Lch_ENSLACT00000004449_ENSLACP00000004411	T	T	C	L	V	I	T	V	N	M	V	T	I	I	N	A	15,3095238
Hro_CG.MTP2014.S1.g06477.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X	14,4523810
Hro_CG.MTP2014.S22.g04056.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X	14,4523810
Hro_CG.MTP2014.S14.g02421.01.p	Q	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X	14,4523810
Hro_CG.MTP2014.S40.g11771.01.p	T	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X	14,3809524
Hro_CG.MTP2014.S45.g15766.01.p	T	T	C	L	V	I	T	V	N	A	V	T	I	I	N	X	14,3809524
Xtr_Act3_NP_989076.1_actin__alpha_sarcomeric/skeletal	T	T	C	L	V	I	T	V	N	Q	V	T	I	I	N	A	14,3333333
Cro_ENSCINP00000015480_muscle-GRAIL32-59-1_Ci-MA2-adultHeart-bodyWall	Q	T	C	L	V	I	V	V	N	M	V	T	I	I	N	A	13,6705069
Csa_ENSCSAVP00000014853_muscle-ACT	Q	T	C	L	V	I	V	V	N	M	V	T	I	I	N	A	13,6705069
Hro_CG.MTP2014.S4.g08850.01.p_BAC53766_HrMAa-Kawahara02-adult	Q	T	C	L	V	I	V	V	N	M	V	T	I	I	N	A	13,6705069
Pmamm.CG.MTP2014.S102.g03221.01.p	Q	T	C	L	V	I	V	V	N	M	V	T	I	I	N	A	13,6705069
Hsa_ACTG2_ENSP00000295137_actin-gamma-2-smooth-muscle-enteric	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A	13,5030722
Gga_ACTG2_ENSGALP00000048103_pep_protein_coding	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A	13,5030722
Loc_ENSLOCT00000007323_ENSLOCP00000007315	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A	13,5030722
Mocci.CG.ELv1_2.S260282.g06079.01.p	Q	T	C	L	V	X	T	V	N	S	V	T	I	I	N	A	13,4047619
Hsa_ACTA2_ENSP00000224784_actin-alpha-2-smooth-muscle-aorta	S	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A	12,5745008
Gga_ENSGALP00000010239_ENSGALT00000010253_actin-aortic-smooth-muscle	S	T	C	L	C	I	T	V	N	M	V	T	I	I	N	A	12,5745008
Lch_ENSLACT00000004521_ENSLACP00000004482	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	X	12,5506912
Lch_ENSLACT00000005620_ENSLACP00000005571	T	T	C	L	C	I	T	V	N	M	V	T	I	I	N	X	12,5506912
Mocul.CG.ELv1_2.S44049.g02319.01.p-BAA12860_MocuMA2-Kusakabe97-bodyWall	Q	T	C	L	V	I	V	V	N	M	V	T	I	I	N	S	11,8471582
Mocul_L21915_McMA1-Swalla94-adult	Q	T	C	L	V	I	V	V	N	M	V	T	I	I	N	S	11,8471582
Scl_X61042_SpMA1-Kovilur93-adult	Q	T	C	L	V	I	V	V	N	M	V	T	I	I	N	S	11,8471582
Psi_ENSPSIT00000006049_ENSPSIP00000006014	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A	11,7926267
Oan_ENSOANT00000011791_ENSOANP00000011789	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A	11,7926267
Loc_ENSLOCT00000006366_ENSLOCP00000006358	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A	11,7926267
Lch_ENSLACT00000005143_ENSLACP00000005097	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	A	11,7926267
Mocci.CG.ELv1_2.S495159.g18580.01.p	T	T	C	L	V	I	V	V	N	M	V	T	I	I	N	S	11,7757296
Mocul.CG.ELv1_2.S110839.g11786.01.p	Q	T	C	L	V	I	T	T	N	A	V	T	I	I	N	S	11,5975422
Bsc_CAX48983_BsMA2a-Degasperi00-bodyWallCardiac	Q	T	C	L	V	I	V	V	N	M	V	T	V	I	N	S	11,0852535
Bsc_botctg010875g10356_CC_botctg069403g27308_CC_muscle-ACT	Q	T	C	L	V	I	V	V	N	M	V	T	V	I	N	S	11,0852535
Gga_ENSGALT00000081189__ENSGALP00000044190-ENSGALP00000023523	T	T	C	L	C	I	V	V	N	M	V	T	I	I	N	S	9,9692780
Mocci.CG.ELv1_2.S637331.g28276.01.p	X	X	X	X	X	X	T	V	N	S	V	T	I	I	N	A	8,5000000
Bla_BL10644_evm0_gene=ACTA2	A	T	I	L	C	V	I	A	A	M	V	T	I	I	N	S	4,7987711
Bbe_BRBE213760F-t1-BRBE213740F-t1_muscle	T	T	V	L	V	V	T	V	T	M	V	A	L	I	N	S	3,8909370
Bfl_XP_002610808.1__BRAFL115345_muscle-ACT	T	T	V	L	V	V	T	V	T	M	V	A	L	I	N	S	3,8909370
Bfl_XP_002610807.1_	A	T	I	L	C	V	I	A	A	M	X	T	I	I	N	S	3,7987711
Bbe_BRBE213730F-t1_muscle	T	T	C	L	C	V	C	V	T	M	V	A	L	I	N	S	3,1321045
Bfl_XP_002610809.1__BRAFL126328_muscle-ACT	T	T	C	L	C	V	C	V	T	M	V	A	L	I	N	S	3,1321045
Bbe_BAA13445_BbMA1-Kusakabe99-muscle	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,8571429
Bbe_BAA13351-KusakabeR-muscle	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,8571429
Bfl_XP_002610797.1_XP_002610798.1	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,8571429
Bfl_XP_002610799.1_	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,8571429
Bfl_XP_002610801.1_	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,8571429
Bla_BL18017_evm0_gene=ACTC1	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,8571429
Bla_BL24512_evm0_gene=ACTC1	A	T	C	L	V	V	C	S	T	L	V	A	L	I	N	A	2,8571429
Bfl_XP_002610800.1_	A	T	C	L	V	V	T	S	T	L	V	A	L	I	N	S	1,9861751
Bfl_XP_002610802.1_	A	T	V	L	V	V	S	A	V	L	V	A	L	I	N	A	1,8571429
Bfl_XP_002610803.1_	A	T	V	L	V	V	S	A	V	L	V	A	L	I	N	A	1,8571429
Bla_BL19911_evm0_gene=ACTC1	A	T	C	L	V	V	T	A	L	I	T	A	L	I	I	S	-0,0138249
Bfl_XP_002590160.1*_NW_003101374.1_332249..373196_	X	X	X	X	X	V	I	S	T	E	A	A	I	I	V	S	-2,1090630
Bfl_XP_002601937.1_	A	T	C	M	C	V	T	S	T	L	V	A	L	I	T	S	-3,8202765
Bla_BL03268_evm0_gene=ENSG00000267807	A	T	C	M	C	V	T	S	T	L	V	A	L	I	T	S	-3,8202765
Bbe_BRBE215870R-BRBE294130F-BRBE215890R-BAA96545/6/7_BbNA3BbNA1/2/3-Suzuki00-notochord	V	A	V	M	C	V	V	T	N	L	V	T	I	I	N	S	-3,8241167
Bfl_XP_002590525.1__BRAFL115237-BRAFL114966-BRAFL114967_notochord-ACT	V	A	V	M	C	V	V	T	N	L	V	T	I	I	N	S	-3,8241167

Name	Sequence	Value
Bla_BL08383_evm0_gene=ACTG2_Sc0000322=268752-270053-BL05535-BL12029-BL12109-Notochord	V A V M C V V T N L V T I I N S	-3,8241167
Bla_BL25564_evm0_gene=ACTG2	V A V M C V V T N L V T I I N S	-3,8241167
Bfl_XP_002603133.1__BRAFL116948_notochord-ACT	V A V M C V V T N L V T I I N S	-3,8241167
Bfl_XP_002588007.1_	D I V M C V V S T E V T I I T S	-3,8402458
Bbe_BRBE025880R-t1_cyto	V A V M C I V T T M V T I I T S	-3,8479263
Bfl_XP_002590158.1_	V A V M C I V T T M V T I I T S	-3,8479263
Bla_BL19846_evm0_gene=ENSG00000267807	V A V M C I V T T M V T I I T S	-3,8479263
Bla_BL06827_evm0_gene=ACTB	V S V M C V I S T Q V A I I I S	-4,9155146
Bla_BL11551_evm0_gene=ACTB	D I V M C V V S T E S T I M T S	-5,8402458
Bfl_XP_002596497.1_NW_003101441.1	A T V M C V T T V L I A L I T C	-5,9331797
Bla_BL08899_evm0_gene=ACTB	A T V M C V T T V L I A L I T S	-6,8041475
Bbe_BRBE025870R-t1_cyto	V A I M C V V T T L V T I I T S	-6,8241167
Bla_BL19005_evm0_gene=ACTB	V A I M C V V T T L V T I I T S	-6,8241167
Bbe_BRBE025860R-t1_cyto	I A V M C V V C T L V T I I T S	-6,8402458
Bsc_botctg031514g7387_CC_botctg064563g24641_CC	X X X X X V C A T I T A I V T S	-7,1090630
Pmamm.CG.MTP2014.S269.g06421.01.p	V P V M C V T T X I T T I V X S	-7,1136713
Bfl_XP_002592018.1__BRAFL79607_cytoplasmic-ACT	V A V M C V V S T L T T I I S S	-7,8402458
Csa_ENSCSAVP00000012175_cytoplasmic-ACT	T V V M C V V A T I T V I V T S	-8,7450077
Bla_BL95821_evm1_gene=ACTB	V A V M C V V S T L T T I I T S	-8,8402458
Bla_BL95820_evm0_gene=ACTB	V A V M C V V S T L T T I I T S	-8,8402458
Bsc_botctg021406_g2746_CC_	X X X X X V C S T L T A L V T S	-8,8709677
Bsc_botctg025474g4510_CC_	E A V M C V C S T I T T I V T S	-9,0821813
Odi_Sc50_GSOIDT00013080001*cyto	V A I M C V T T T I T T I V T S	-9,1136713
Pmamm.CG.MTP2014.S269.g06422.01.p	I T V M C V I T T I T T I V T S	-9,1612903
Mocci.CG.ELv1_2.S652301.g30535.01.p	I T I M C V V T T L T A L V A A	-9,6912442
Hsa_ACTBL2_ENSP00000416706	L S V M C V I T I L T A L V T A	-9,8379416
Pmamm.CG.MTP2014.S269.g06424.01.p	T T V M C V I T T L T A L V T S	-10,8279570
Bsc_botctg025474g4512_CC_	V A V M C V C T T I T T I V T S	-11,0660522
Bbe_BRBE249410R-t1_cytoplasmic	V A V M C V V S T L T S L I T S	-11,4354839
Bfl_XP_002606072.1_BAA13350_BRAFL56687-Kusabake97_cytoplasmic	V A V M C V V S T L T S L I T S	-11,4354839
Bla_BL10072_evm0_gene=ACTB	V A V M C V V S T L T S L I T S	-11,4354839
Bbe_BAA13444-BbCA1-Kusabake99-CYTO	V A V M C V V S T L T S L I T S	-11,4354839
Bbe_BRBE202700F-t1_NONE	V A V M C V V S T L T A L I A S	-11,4354839
Pmamm.CG.MTP2014.S269.g06426.02.p	S P V M C V V T T M T A L V T S	-11,4431644
Bsc_botctg073109g29085_CC_cytoplasmic-ACT	S S V M C V C S T L T A L V T S	-11,6774194
Csa_ENSCSAVP00000014972_cytoplasmic-ACT	V A V M C V V T T L T T I V T S	-11,8241167
Cro_ENSCINP00000017903-GRAIL176-23-1-CYTO	V A V M C V V T T I T T I V T S	-11,8241167
Bsc_botctg087748g38113_CC-contig_9376_g24641CC_cytoplasmic-ACT	V A V M C V C A T I T A I V T S	-11,9155146
Bfl_XP_002592016.1__BRAFL114535_cytoplasmic-ACT	V A V M C V V S T I T A L I T S	-12,4354839
Mocci.CG.ELv1_2.S82356.g01111.01.p	I S V M C V V X T L T S L V T S	-12,4354839
Pmamm.CG.MTP2014.S327.g07316.01.p	T A V M C V V C T L T A L V T S	-12,5069124
Mocul.CG.ELv1_2.S10781.g00320.01.p	V T I M C V V T T L T A L V T S	-12,5145929
Mocci.CG.ELv1_2.S138720.g02381.01.p	V A V M C V V A T S T A L V T S	-13,4354839
Cro_GRAIL2-31-1_NONE	I A I M C V V A T L T A L V T S	-13,4354839
Hsa_ACTG1_ENSP00000331514_actin-gamma-1	I A I M C V V T T L T A L V T S	-14,4193548
Gga_ENSGALP00000042680_actin-gamma-1	I A I M C V V T T L T A L V T S	-14,4193548
Bsc_CAX48982_Degasperi09-CYTO	S A V M C V V T T L T A L V T S	-14,4193548
Bsc_botctg033661g8425_CC_cytoplasmic-ACT	S A V M C V V T T L T A L V T S	-14,4193548
Pmamm.CG.MTP2014.S269.g06423.02.p	A A V M C V V T T L T A L V T S	-14,4193548
Pmamm.CG.MTP2014.S269.g06427.01.p	V A I M C V V T T L T A L V T S	-14,4193548
Cro_GRAIL176-22-1-CYTO	V A V M C V I T T L T A L V T S	-14,6612903
Hsa_ACTB_ENSP00000349960_actin-beta	I A V M C V V T T L T A L V T S	-15,4193548
Olo_BAA86216_OilCA1-Nishino00_cyto	V A V M C V V T T L T A L V T S	-15,4193548
Odi_Sc1.1_GSOIDT00000372001_cyto	V A V M C V V T T L T A L V T S	-15,4193548
Odi_Sc49_GSOIDT00013012001_cyto	V A V M C V V T T L T A L V T S	-15,4193548
Gga_ENSGALP00000039176_Actin-cytoplasmic-type-5	I A V M C V V T T L T A L V T S	-15,4193548
Gga_ACTB_ENSGALP00000015657_Actin-cytoplasmic-1-Actin-cytoplasmic-1-	I A V M C V V T T L T A L V T S	-15,4193548
Lch_ACTB_ENSLACP00000019270	I A V M C V V T T L T A L V T S	-15,4193548
Pma_ENSPMAT00000011113_ENSPMAP00000011067	I A V M C V V T T L T A L V T S	-15,4193548
Pma_ENSPMAT00000004586_ENSPMAP00000004567	I A V M C V V T T L T A L V T S	-15,4193548
Mocul.CG.ELv1_2.S97945.g08376.01.p	V A V M C V V T T L T A L V T S	-15,4193548
Cro_ENSCINP00000001538-GRAIL176-17-1-CYTO	V A V M C V V T T L T A L V T S	-15,4193548
Csa_ENSCSAVP00000014485_cytoplasmic-ACT	V A V M C V V T T L T A L V T S	-15,4193548
Hro_CG.MTP2014.S250.g09381.01.p	V A V M C V V T T L T A L V T S	-15,4193548
Hro_CG.MTP2014.S422.g05364.01.p	V A V M C V V T T L T A L V T S	-15,4193548

Supplementary Figure S1. Full detailed maximum-likelihood tree corresponding to Figure 1 in Nexus Format, including species names and sequence references identifications and databses as indicated in Inoue and Satoh (2018). Full resolution image https://www.sciencedirect.com/science/article/pii/S0012160617308886

Bbe Branchiostoma belcheri
Bfl Branchiostoma floridae
Bla Branchiostoma lanceolatum
Bsc Botryllus schlosseri
Gga Gallus gallus
Cro Ciona robusta (former Ciona intestinalis)
Csa Ciona savignyi
Hro Halocynthia roretzi
Hsa Homo sapiens
Lch Latimeria chalumnae (coelacanth)

Pma Petromyzon marinus (lamprey)
Loc Lepisosteus oculatus (spotted gar)
Mocul Molgula oculata
Mocci Molgula occidentalis
Odi Oikopleura dioica
Olo Oikopleura longicauda
Pmamm Phallusia mammillata
Psi Pelodiscus sinensis (Chinese softshell turtle)
Oan Ornithorhynchus anatinus (Platypus)
Scl Styela clava
Xtr Xenopus tropicalis
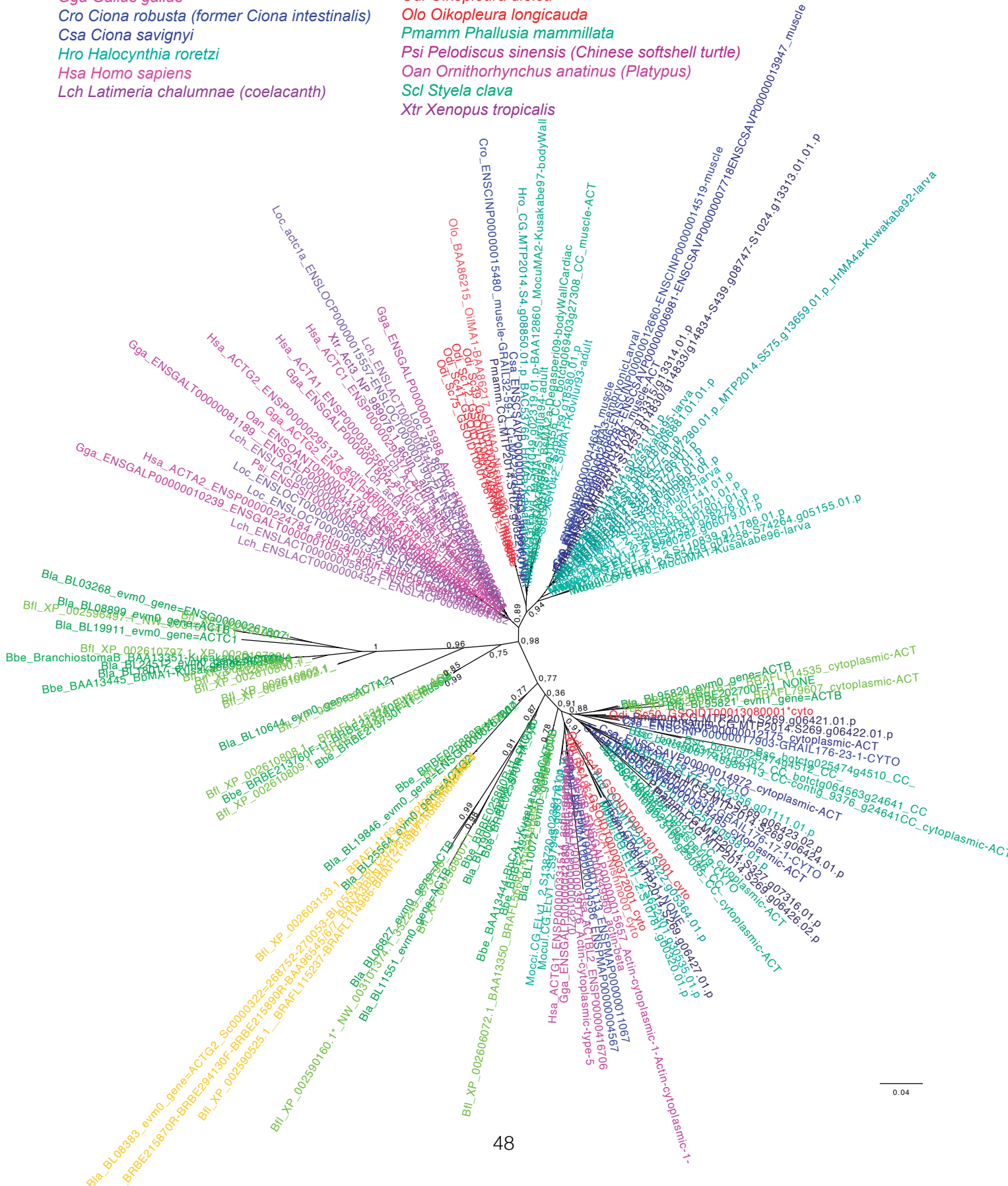
Supplementary Figure S2. Comparative analyses of exon-intron structures of cytoplasmic and muscle actins in olfactores. Representation generated with GECA software indicating with colored lines the positions of all 13 introns found in actins (numbers on top), according to their relative position in a sequence alignment of cytoplasmic and muscle actins (black lines). Homo sapiens (Hsa), Ciona robusta (Cro), Oikopleura longicauda (Olo) and Oikopleura
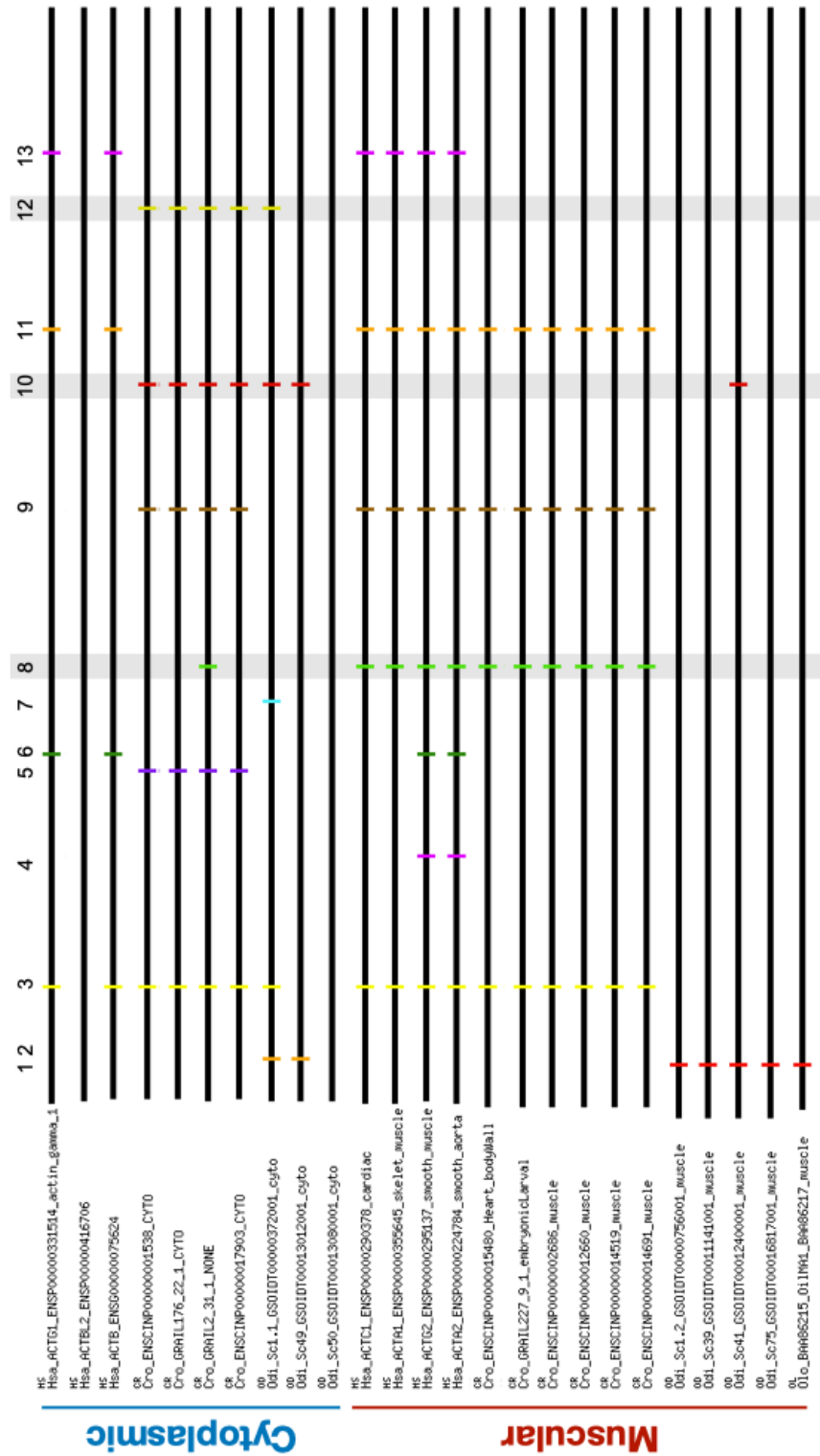
Supplementary Figure S3. Bayesian inferences of muscular actins from olfactores does not resolve the relationship between appendicularian ActnM actins and Ascidian-B and -P groups. (A) When fast evolving actins from non-mammalian actins are included among vertebrate sequences, ActnM from a cluster with Ascidian-B actins with a support of 0.9. (B) When fast evolving actins from non-mammalian actins are not included among vertebrates actins, Appendicularian ActnM, and Ascidian-B and -P groups collapse into a trichotomy. The fact that the tree topologies can change dramatically by the presence or absence of certain sequences suggest that there is not enough evolutionary signal (as indicated by the many trichotomies shown throughout the tree) to compensate internal tensions of long branch repellence, and therefore a confident topology cannot be determined. Larger samples with wider taxonomic representation, specially from thaliaceans, salps, further appendicularians species, as well as basally divergent gnathostomes would be required to better understand the evolutionary history of actins in urochordates. Species abbreviations are as indicated in Fig. Sup. 1. Nexus format tree files, and sequence alignments are provided.

50
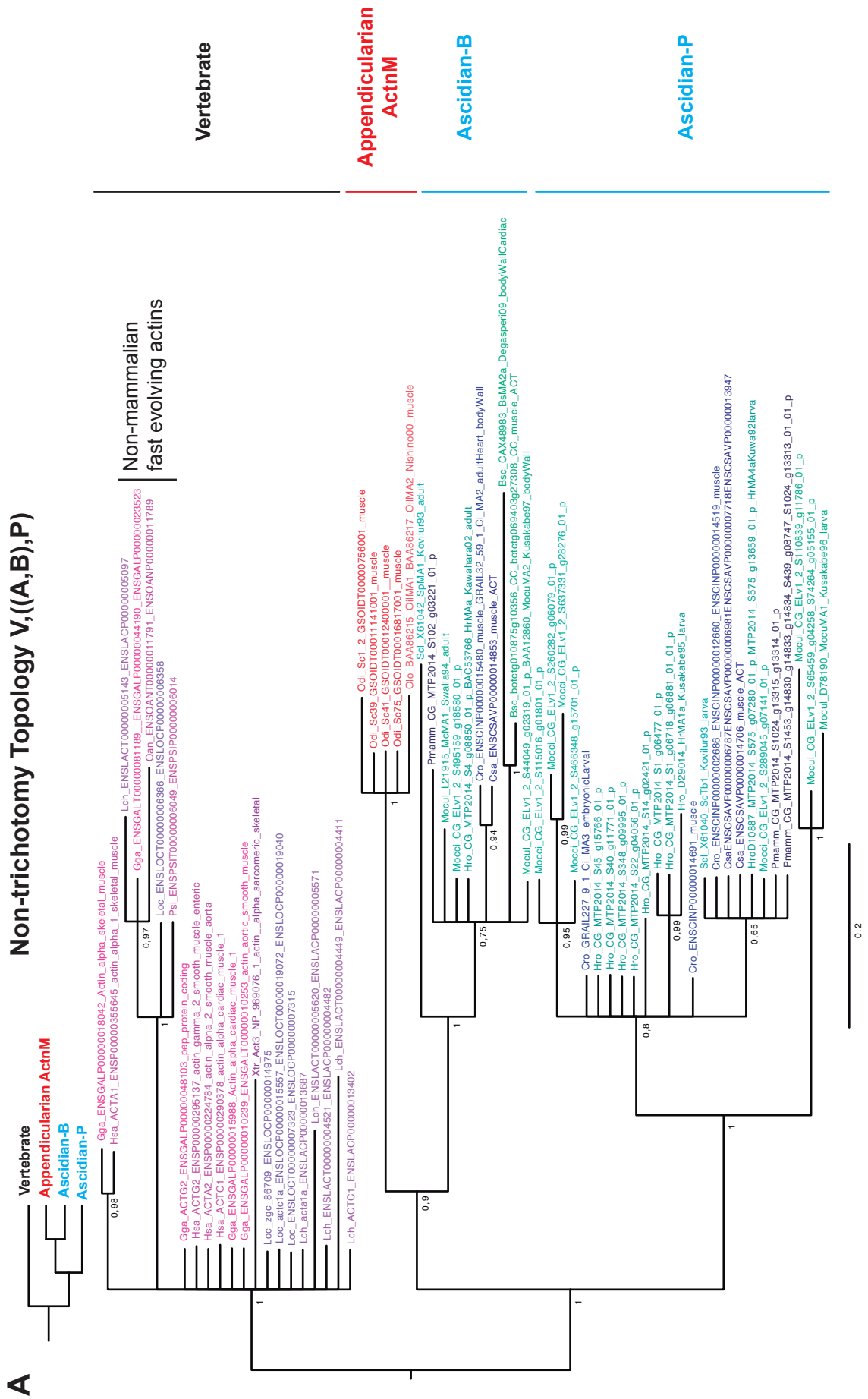
B

Tricotomy topology V,(A,B,P)

Vertebrate
Appendicularian ActnM
Ascidian-B
Ascidian-P

Supplementary Figure S4. Summary of WMISH expression data from cytoplasmic actins, and muscular actins from the -P and -B groups in the ascidian *C. robusta* curated in the ANISEED database (www.aniseed.cnrs.fr). Expression domains in paraxial muscle and heart precursors cells are indicated in the in situ hybridization experiments with ascidian-P and -B muscle actins. While heart precursor cells appear to be labeled by ascidian-B actin (e.g. *Ci-MA2*) – but not ascidian-P actins, e.g. KH.C1.242 Ci-Ma1, kh.c1.570 Ci-Ma3, KH.S1440.1–, paraxial muscle cells appear to be labeled by both ascidian-P and -B muscle actins. In the latter case, however, we cannot discard the possibility of cross-hybridization between probes of ascidian-P and -B genes, since their sequences share an overall 87-88% of similarity. Specific probes from the UTR regions of ascidian actin-P and -B genes should be used to test this possibility, and to reconcile the apparent conflictive results between in situ hybridization experiments and expression data from EST counts and RNA-seq transcriptomic profiles, which points for instance that ascidian-B genes are mostly expressed in heart and adults, but mostly absent in larval stages (e.g. https://www.aniseed.cnrs.fr/aniseed/gene/show_expression?unique_id=Cirobu.g00009302).

In situ expression data

https://www.aniseed.cnrs.fr/aniseed/

Cytoplasmic Actins

Cro_**KH2012:KH.L154.4**-ENSCINP00000001538-GRAIL176-17-1-CYTO
https://www.aniseed.cnrs.fr/aniseed/gene/show_expression/unique_id=Cirobu.g00011580

Cro_**KH2012:KH.L154.43**_ENSCINP00000017903-GRAIL176-23-1-CYTO   NO DATA
https://www.aniseed.cnrs.fr/aniseed/gene/show_expression/unique_id=Cirobu.g00011584

Cro_**KH2012:KH.L154.34**_GRAIL176-22-1-CYTO   NO DATA
https://www.aniseed.cnrs.fr/aniseed/gene/show_expression/unique_id=Cirobu.g00011574

Cro_**KH2012:KH.C3.563**_GRAIL2-31-1_NONE   NO DATA
https://www.aniseed.cnrs.fr/aniseed/gene/show_expression/unique_id=Cirobu.g00005752

Muscle Actins

Ascidian-P

Cro_KH2012:KH.C1.242-KH2012:KH.C1.570-KH2012:KH.S1440.1-ENSCINP00000002686-ENSCINP00000012660-ENSCINP00000014519-muscle (identical protein sequence)
ENSCINP00000014691 and GRAIL227-9-1 likely polymorphic variants

KH2012:KH.C1.242 (Ci-Ma1)
https://www.aniseed.cnrs.fr/aniseed/gene/show_expression/unique_id=Cirobu.g00014691

paraxial muscle

KH2012:KH.C1.570 (CiMA3)
https://www.aniseed.cnrs.fr/aniseed/gene/show_expression/unique_id=Cirobu.g00000821

paraxial muscle

KH2012:KH.S1440.1
https://www.aniseed.cnrs.fr/aniseed/gene/show_expression/unique_id=Cirobu.g00013872

paraxial muscle

paraxial muscle

Ascidian-B

Cro_**KH2012:KH.C8.649**_ENSCINP00000015480_muscle-GRAIL32-59-1_Ci-MA2-adultHeart-bodyWall
https://www.aniseed.cnrs.fr/aniseed/gene/show_expression/unique_id=Cirobu.g00009302

heart

paraxial muscle

cross-hybridization ?

Table S1. Primers, cloning and riboprobe labeling information.

	Gene	Forward	Reverse	Source; length; vector; RNA-pol/digestion enzyme
Specific probes	ActnM1	5'- GATCGTCCACCGAAAGTGC-3'	5'- GTCAGCAACTGTTTGAATATATTG-3'	genomic from exon 2 to 3'UTR; 325 bp ;pCR4-TOPO; T7/PstI
	ActnM2	5'- GGTCCCTCCATCGTCCA-3'	5'- GTTTTATTATTTTGTTTGTTTTGAAGGTG-3'	genomic from exon 2 to 3'UTR; 60 bp ; pCR4-TOPO; T7/PstI
	ActnM3	5'- GTCCACCGAAAGTGCTTCTAA-3'	5'- GAGGTAGTGTATTCGAAAAGTAC-3'	genomic from exon 3 to 3'UTR; 98 bp ; pCR4-TOPO; T3/NotI
	ActnM4	5'- GTCCACCGAAAATGCTTCTAA-3'	5'- GGACCCCTCATGATATTTTTTA-3'	genomic from exon 2 to 3'UTR; 96 bp ; pCR4-TOPO; T7/PstI
	ActnC1	5'- GCGCCGTTGTTATTTCCACG-3'	5'- GAATTAACATGAAGCGGGGTAAT-3'	genomic from exon 6 to 3'UTR; 486 bp ; pCR4-TOPO; T3/NotI
	ActnC2	5'- GTCCACCGAAAATGCTTCTAAAC-3'	5'- GCGCCGTTGGGCAAGCC-3'	genomic from exon 3 to 3'UTR; 229 bp ; pCR4-TOPO; T3/NotI
	ActnC3	5'-GACCTTCCATCGTCCACAG-3'	5'-GCTTTAAATCTCTCTCGTGTG-3'	genomic from exon 1 to 3'UTR; 117 bp ; pCR4-TOPO; T7/PstI
Cross-hybridizing probe	ActnM1	5'-GTCCCCGCCATGTACGTCTG-3'	5'-GCATCGGAATCGCTCGTTACCA-3'	genomic from the middle region of exon 2; 389 bp; pCR4-TOPO;T3/NotI

Table S2. Protein sequence identity among *O. dioica* Actins in percentage (above) and number of differences (below).
Muscular and Cytoplamatic are in red and blue, respectively, and comparisons within groups are labeled in bold. * indicate manual curation of annotations

Name	Annotation ID (oikobase)	Scaffold (oikobase)	ID Genbank	ActnM1	ActnM2	ActnM3	ActnM4	ActnC1	ActnC2	ActnC3
ActnM1	GSOIDT00000756001	1 (position 2,362 Kb)	gi|313191932|CABV01000065.1		99,7%	99,7%	99,5%	92,3%	92,0%	91,2%
ActnM2	GSOIDT00011141001	39	gi|313190120|CABV01001709.1	1		100,0%	99,7%	92,6%	92,3%	91,5%
ActnM3	GSOIDT00012400001*	41	gi|313190075|CABV01001754.1	1	0		99,7%	92,6%	92,3%	91,5%
ActnM4	GSOIDT00016817001	75	gi|313189377|CABV01002452.1	2	1	1		92,6%	92,3%	91,5%
ActnC1	GSOIDT00000372001	1 (position 1,275 Kb)	gi|313191976|CABV01000043.1	29	28	28	28		98,9%	95,2%
ActnC2	GSOIDT00013012001	49	gi|313189901|CABV01001928.1	30	29	29	29	4		95,5%
ActnC3	GSOIDT00013080001*	50	gi|313189893|CABV01001936.1	33	32	32	32	18	17	

Table S3. Probe length and sequence percentage identity among *O. dioica* ActnM and ActnC.
The values in brackets indicate the identities when the positions with gaps of reference sequence were excluded

Genbank ID	Position	Probe length (nt)	Probe	*ActnM1*	*ActnM2*	*ActnM3*	*ActnM4*	*ActnC1*	*ActnC2*	*ActnC3*
gi|313191932|CABV01000065.1	25912 - 26236	325	*ActnM1*	**100**	42(12)	46(19)	47(20)	32(49)	59(38)	47(20)
gi|313190120|CABV01001709.1	5331 - 5389	59	*ActnM2*	44(71)	**100**	87(80)	84(64)	9(76)	58(85)	86(75)
gi|313190075|CABV01001754.1	11104 - 11201	98	*ActnM3*	47(60)	87(40)	**100**	73(30)	12(57)	54(50)	80(44)
gi|313189377|CABV01002452.1	8809 - 8714	96	*ActnM4*	48(64)	83(28)	74(29)	**100**	13(67)	57(50)	78(40)
gi|313191976|CABV01000043.1	31892 - 31407	486	*ActnC1*	26(26)	3(3)	5(5)	9(9)	**100**	23(23)	9(9)
gi|313189901|CABV01001928.1	6064 - 5836	229	*ActnC2*	62(45)	54(14)	53(19)	57(19)	28(54)	**100**	58(21)
gi|313189893|CABV01001936.1	19931 - 19814	118	*ActnC3*	48(59)	85(41)	80(47)	77(42)	15(63)	59(64)	**100**
gi|313191932|CABV01000065.1	25193 - 25579	387	**Cross-hybridizing probe**	**100**	97(97)	96(96)	99(99)	86(86)	89(89)	79(79)

To calculate the identity for each Actn specific probe, only the range of the aligment corresponding to the probe was considered. Values in brackets correspond to the identities when the positions of the reference sequenced that had gaps were excluded. The fact that the specific probes mostly encompassed the 3'UTR, and their sequences and lengths were vey variable made that some of the identity values were very low, avoiding therefore crosshybridization. For instance, comparison between the sequence of the ActnC1probe with the corresponding region of the ActnM2 in the alignment revealed only 3% similarity, which corresponded to the 17 identical positions out of 488, while the other positions consisted in 11 mismatches and 460 gaps.

Whole mount in istu hybridization with the cross-hybridizin probe that recognizes different actins, displaying for instance the ActnM1 expression domains in paraxial muscle and heart primordium, as well ActnC1 expression domains in the notochord, trunk endoderm and mouth epidermis as shown in Fig. 2E and Fig. 2AS, respectively.

"cross-hybridizing probe"

notochord · heart · mouth · endoderm · epidermis · paraxial muscle
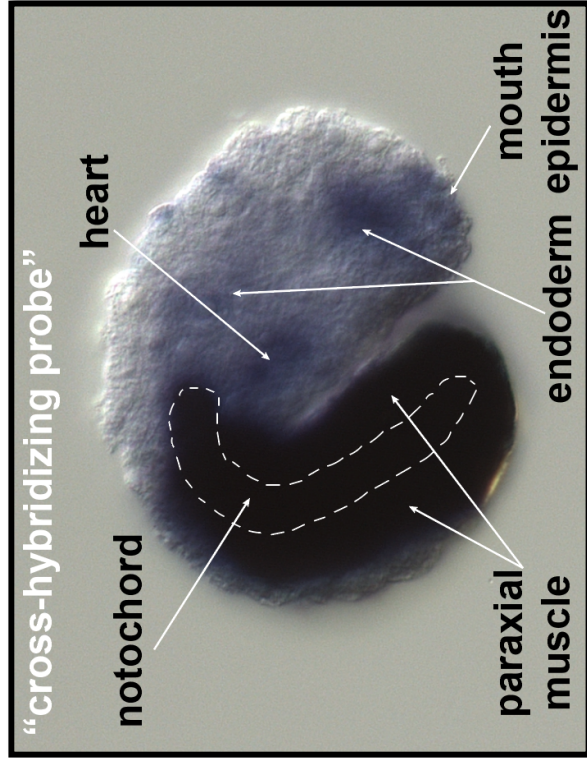
CHAPTER II

ADAPTIVE EVOLUTION BY MODULAR DECONSTRUCTION OF THE CARDIOPHARYNGEAL GENE REGULATORY NETWORK IN APPENDICULARIANS REVEALS THAT ANCESTRAL TUNICATES WERE SESSILE

Adaptive evolution by modular deconstruction of the cardiopharyngeal gene regulatory network in appendicularians reveals that ancestral tunicates were sessile

Ferrández-Roldán A., Fabregà-Torrus M.[1], Sánchez-Serna G.[1], Durán-Bello E., Joaquín-Lluís M., Bujosa P., Plana-Carmona M., Garcia-Fernàndez J., Albalat R., Cañestro C.*

Departament de Genètica, Microbiologia i Estadística, Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona (Spain).

[1]*Author equal contribution.*

Corresponding author: Email: canestro@ub.edu

ARTICLE INFO

ABSTRACT

The discovery that cephalochordates, rather than tunicates, diverged at the base of the chordates challenged the traditional view in which the ancestral chordate had a biphasic ascidian-like lifestyle with a pelagic larva and a sessile adult, and suggested that the ancestral chordate was entirely free-living as modern amphioxus or appendicularians. A key problem in understanding tunicate evolution has been the origin of sessility of ascidians, and whether the appendicularian free-living style represents a primitive or derived condition of tunicates. To address this problem, we have performed comprehensive developmental and genomic comparative analyses of the cardiopharyngeal gene regulatory network (GRN) between appendicularians and ascidians. Our results reveal that the cardiopharyngeal GRN has suffered a process of evolutionary deconstruction with massive ancestral losses of genes (*Mesp*, *Ets1/2*, *Gata4/5/6*, *Mek1/2*, *Tbx1/10*, and RA- and FGF-signaling related genes) and subfunctions (e.g. *FoxF*, *Islet*, *Ebf*, *Mrf*, *Dach* and Bmp signaling). These losses have led to the dismantling of two modules that in ascidians are related to early and late multipotent state cells involved in lineage fate determination towards first and secondary heart fields, and siphon muscle. Our results allow us to propose an evolutionary scenario, in which the evolutionary deconstruction of the cardiopharyngeal GRN has had an adaptive impact on the acceleration of the developmental cardiac program, the redesign of the cardiac architecture into an open-wide laminar structure, and the loss of siphon muscle. Altogether, these adaptations have facilitated the transition from a sessile ascidian-like ancestral condition to a pelagic free-living style connected to the innovation of the house in appendicularians.

1 INTRODUCTION

A key problem in the field of Evolutionary Developmental Biology (EvoDevo) is understanding the origin and radiation of our own phylum, the chordates (Gee, 2018; Satoh, 2016). The discovery that tunicates (a.k.a. urochordates) are the sister group of vertebrates, and therefore that the branching of cephalochordates is basal within chordates (Bourlat et al., 2006; Delsuc et al., 2006), has provided a novel view of the last common ancestor of chordates as a free-living organism, in contrast to the traditional view proposed by Garstang (1928) in which it had a sessile ascidian-like adult lifestyle (Garstang, 1928). This novel view has brought renewed interest in appendicularians, whose complete free-living style could parsimoniously represent the ancestral condition of tunicates (Berrill, 1950; Braun et al., 2020; Swalla et al., 2000) considering their most accepted position as the sister group of the remaining tunicates (Braun et al., 2020; Delsuc et al., 2018; Kocot et al., 2018; Wada, 1998). In such evolutionary scenario, sessility in tunicates would be a derived ascidian trait that could have evolved during the acquisition of their dramatic larval to adult metamorphosis, an event that is absent in appendicularians (Braun et al., 2020; Giribet, 2018). Certain developmental features of appendicularians that resemble Aplousobranchiata ascidians, however, have been argued to represent traces of sessility, which therefore would favor the hypothesis that appendicularians have evolved from a tunicate ancestor with a sessile adult lifestyle and a larval

dispersal stage similar to ascidians (Stach, 2007; Stach et al., 2008; Stach and Turbeville, 2002). Thus, the problem of the lifestyle of the last common ancestor of tunicates remains unsolved.

The evolution of the cardiopharyngeal gene regulatory network appears to be a pivotal aspect to understand the evolution of ascidian and vertebrate lifestyles (Diogo et al., 2015; Razy-Krajka and Stolfi, 2019; Stolfi et al., 2010). The hypothesis of "a new heart for a new head" about the origin of vertebrates (Diogo et al., 2015), in addition to the evolution of placode and neural crest derivatives as proposed by Gans and Northcutt (Gans and Northcutt, 1983; Horie et al., 2018; Martik et al., 2019), points to the development of a chambered heart and elaborated branchiomeric muscles as key evolutionary innovations that facilitated the transition from a peaceful filter-feeder lifestyle of ancestral olfactores to a blistering predatory lifestyle of vertebrates. The cardiopharyngeal field is the developmental domain that gives rise to the heart and branchiomeric muscles from a common pool of early cardiopharyngeal progenitors, that after a binary-stepwise process of fate choices gives rise to the first heart field (FHF), the second heart field (SHF) and to branchiomeric muscles (BM), the later including arch muscles involved in mandibular, facial, and branchial functions in the head and neck of vertebrates –reviewed in (Diogo et al., 2015). In ascidians, the pharyngeal muscles are considered homologous to vertebrates BM, and their cardiopharyngeal GRN is highly conserved in comparison to vertebrates using a homologous model of binary-stepwise fate choices, thus becoming an attractive system to study heart development at unprecedented spatiotemporal single-cell resolution (Christiaen et al., 2010; Davidson, 2007; Racioppi et al., 2019; Stolfi et al., 2010; Wang et al., 2019). Consistent with this model, the pre-cardiac master regulator *Mesp* is expressed in multipotent progenitor cells both in vertebrates and in ascidians (Devine et al., 2014; Lescroart et al., 2014; Saga et al., 1999; Satou et al., 2004). During ascidian gastrulation, the *Mesp+* precardiac cells divide asymmetrically, to give rise to the anterior tail muscles (ATM) under the influence of RA-signaling (Christiaen et al., 2008a; Nagatomo and Fujiwara, 2003) and to multipotent cardiopharyngeal progenitors (trunk ventral cells, TVCs) under the influence of FGF-signaling (Davidson et al., 2006). TVCs divide and migrate to the ventral part of the trunk where they will be specified to become first heart progenitors (FHP), second heart progenitors (SHP) and pharyngeal muscle under the influence

of FGF and BMP signaling, and the upregulation of cardiac and muscular factors including *Nk4*, *Hand1/2*, *Gata4/5/6*, *Dach* and *Tbx1/10* as some of the crucial members of the cardiopharyngeal GRN also conserved in vertebrates (Beh et al., 2007; Christiaen et al., 2010; Davidson et al., 2006; Davidson and Levine, 2003; Hutson et al., 2010; Razy-Krajka et al., 2014; Satou et al., 2004; Tirosh-Finkel et al., 2010; Tolkin and Christiaen, 2016; Wang et al., 2013).

In contrast to ascidians, as far as we know, the cardiopharyngeal GRN in appendicularians has never been studied, and therefore whether evolutionary differences between the cardiopharyngeal GRNs of appendicularians and ascidians reflect evolutionary adaptations to the active free-living and adult sessile lifestyles of these two groups of tunicates, respectively, remains unknown. To address this problem, we have performed a genome survey of the cardiopharyngeal GRN in seven appendicularian species, and studied *Oikopleura dioica* as a model to investigate heart development in this group of tunicates. Our work reveals that appendicularians have suffered an evolutionary "deconstruction" of their cardiopharyngeal GRN, including numerous losses of essential genes and cardiopharyngeal subfunctions of GRN components that are crucial in ascidians and vertebrates. The term "deconstruction", originally coined in Philosophy and later applied in disciplines such as Literature, Architecture, Fashion and Cookery, or even in Developmental Biology (Hogan, 2004) is not synonymous with destruction, but it generally refers to the process of dismantling or breaking apart elements that traditionally are combined, and whose analysis facilitate the recognition of structural modules. In the field of EvoDevo, our work shows how the evolutionary deconstruction by co-elimination of genes and subfunctions during the evolution of appendicularians highlights the modular organization of the cardiopharyngeal GRN. This deconstruction unveils "evolvable modules" that can be related to the evolution of multipotent cellular states that favored the diversification of cardiopharyngeal structures in ascidians and vertebrates, but the loss of pharyngeal muscle and the simplification of the heart in appendicularians. Our results suggest an evolutionary scenario in which the deconstruction of cardiopharyngeal GRN in appendicularians facilitated the transition from an ancestral adult ascidian-like sessile style to a pelagic free-living style based on the evolutionary innovation of the house as the filter-feeding apparatus. Overall, our

work provides an example of the "less is more" hypothesis (Olson, 1999) illustrating how the study of particular gene losses helps to better understand the evolution of adaptations of certain groups of animals (Albalat and Cañestro, 2016), and supports the notion that *O. dioica* is a successful gene loser among chordates that can be used as an attractive evolutionary knockout model to understand the impact of gene loss in the evolution and modular structure of the mechanisms of development in our own phylum (Albalat and Cañestro, 2016; Ferrández-Roldán et al., 2019).

2 MATERIALS AND METHODS

2.1 Biological material

O. dioica specimens were obtained from the Mediterranean coast of Barcelona (Catalonia, Spain). Culturing of *O. dioica* and embryos collections have been performed as previously described in Martí-Solans et al., 2015.

2.2 Genome databases searches and phylogenetic analysis

Protein sequences from the tunicate *C. robusta* and the vertebrate *Homo sapiens* were used as queries in BLASTp and tBLASTn searches in genome databases of selected species: https://blast.ncbi.nlm.nih.gov/Blast.cgi for *Branchiostoma floridae*, *Branchiostoma belcheri*, *Gallus gallus*, *Lepisosteus oculatus*, and *Latimeria chalumnae*; http://www.aniseed.cnrs.fr/ for ascidian species (*Ciona savignyi*, *Phallusia fumigata*, *Phallusia mammillata*, *Halocynthia roretzi*, *Halocynthia aurantium*, *Botryllus schlosseri*, *Botryllus leachii*, *Molgula occulta*, *Molgula oculata*, and *Molgula occidentalis*); http://oikoarrays.biology.uiowa.edu/Oiko/ for *O. dioica*, and a local blast server for six other larvacean species (*Oikopleura albicans*, *Oikopleura vanhoeffeni*, *Oikopleura longicauda*, *Mesochordaeus erythrocephalus*, *Bathochordaeus stygius*, *Fritillaria borealis*) with public genomes (Naville et al., 2019) and *Branchiostoma lanceolatum*. The orthology between cardiac genes was initially assessed by blast reciprocal best hit (BRBH) (Wall et al., 2003) and subsequently corroborated by phylogenetic analysis based on ML inferences calculated with PhyML v3.0 and an automatic substitution model (Guindon et al., 2010) using protein alignment generated by the MUSCLE (Edgar, 2004) program and reviewed manually. Accession numbers are provided in **Sup. Table 1**.

2.3 Cloning and expression analysis

O. dioica genes were PCR amplified from cDNA obtained as described in Martí-Solans et al., 2016. Then, they were cloned using the Topo TA Cloning® Kit (K4530-20, Invitrogen) to synthesize antisense digoxigenin (DIG) and fluorescein (FITC) riboprobes for whole-mount in situ hybridization (WMISH) and double fluorescent whole-mount in situ hybridization (FWMISH) (**Sup. Table 2**). The WMISH were performed as previously described (Bassham and Postlethwait, 2000; Cañestro and Postlethwait, 2007; Martí-Solans et al., 2016) with minor modifications.

For FWMISH, fixed embryos were rehydrated in PBT (PBS/0.2%Tween-20), treated with 50 mM DTT in PBT (10 min at room temperature), washed in 0.1 M triethanolamine in PBT (2 x 5 min at room temperature), treated with two successive dilutions of acetic anhydride (0.25% and 0.5%) in 0.1 M triethanolamine (10 min at room temperature), and washed in PBT (2 x 5 min at room temperature). Prehybridization and hybridization solutions were carried out as for WMISH. Prehybridization incubation lasted 2 hours at 63ºC and hybridization incubation with the two probes, at between 0,5-1 ng/μL each, was conducted overnight at 63ºC. The next day, embryos were washed in successive dilution of SSC (2 x 10 min at 65ºC of 2xSSC/0.2%Tween-20; 2 x 10 min at 65ºC of 0.2xSSC/0.2%Tween-20; 1 x 5 min at room temperature of 0.1xSSC/0.2%Tween-20) and twice with MABT (0.1 M maleic acid, 0.15 M NaCl, 0.1% Tween 20, pH 7.5) for 5 min. Blocking was performed as for WMISH and, finally, Anti-fluorescein antibody conjugated with POD (1:1000 in blocking solution) (11426346910, Roche) was added to the samples for overnight incubation at 4ºC. The day after, samples were washed in MABT (8 x 15 min at room temperature) and then in TNT (0.1 M Tris-HCl pH7, 0.15M NaCl, 0.3% TritonX-100) for 10 minutes. For the staining, embryos were incubated in TSA-tetramethylrhodamine (NEL742001KT, Perkin Elmer) for 10 min. Then, they were washed in TNT for 10 min, in PBT for 10 min, once in 2%H_2O_2/PBT (45 min at room temperature), in PBT (2 x 5 min at room temperature), and in MABT (2 x 5 min at room temperature). Then, a second blocking step was

performed followed by the addition of an anti-Digoxigenin-POD (11207733910, Roche) antibody that was incubated overnight at 4ºC. The morning after, as the previous day, the samples were washed and the coloration reaction was added, that in this case included TSA-Fluorescein system-green (NEL741E001KT, Perkin Elmer) for 1 hour 30 min. After coloration, embryos were washed in TNT (2 x 5 min at room temperature) and PBT (2 x 5 min at room temperature). Mounting was made in 80%glycerol/PBS with Hoechst-33342 1µM (Invitrogen-62249). A confocal microscopy LSM880 (Zeiss) was used for imaging of samples and FIJI (Schindelin et al., 2012) was used to compose the confocal series and adjust the brightness and contrast.

2.4 Pharmacological treatments

For FGF inhibition, animals were treated with 50 µM and 100 µM of SU5402 (SML0443, Merk) from 2-cell stage (30 minutes post fertilization, mpf) and from 32-cell stage (70 mpf), respectively, to hatchling stage (4 hours post fertilization, hpf) in darkness. For BMP inhibition, animals were treated with 10µM of LDN (SML1119, Merk) from 2-cell stage and from 32-cell stage until hatchling stage. To perform these treatments, eggs were pooled in 4mL of SSW and fertilized with 200 µL of sperm dilution (the sperm of 3 males in 5 mL of SSW). At the desired time, embryos were transferred to a 3 mm Petri dish plate with 4 mL of treatment solution at 19ºC. Control embryos were incubated in DMSO 0,2% or 0,3% (v/v) depending on the treatment. The effects of the treatments were scored by in situ hybridization. For tailbud embryos, we used cross-hybridizing *ActnM1*, *Nk4* and *Brachyury* probes (Almazán et al., 2019; Torres-Águila et al., 2018), while for hatchling embryos we used the specific *ActnM1* probe (Almazán et al., 2019) (**Sup.Table. 2**).

3 RESULTS

3.1 Developmental atlas of the heart in O. dioica

The anatomy of the heart has been thoroughly described in adults of *O. dioica* and some other appendicularian species (Fenaux, 1998; Fol, 1872; Lankester, 1882; Onuma et al., 2017; Savelieva and Temereva, 2020). Due to its chamber-less structure made of two layers (the myocardium and the pericardium) ventrally located between the left stomach and the intestine, it could be probably considered one of the simplest hearts of chordates (**Fig. 1A-A'**). The development of the heart in appendicularian embryos, to the best of our knowledge, had never been investigated and
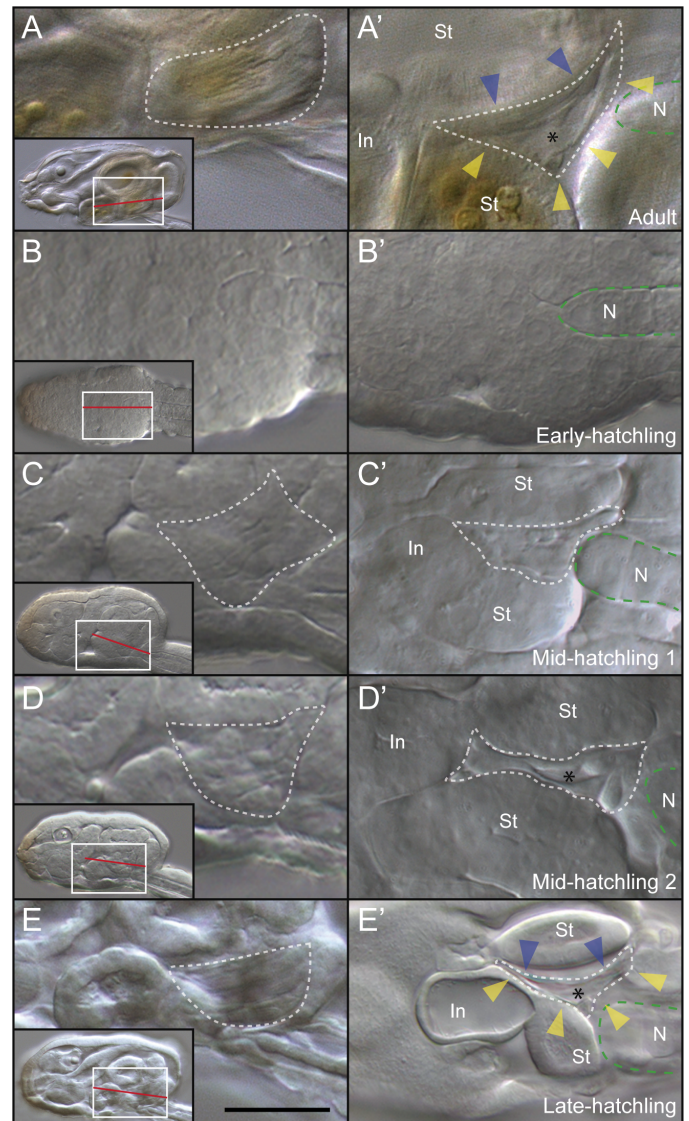
Fig 1. **Developmental atlas of heart development in O. dioica.** (**A**) Anatomy of the adult heart of *O. dioica*. The pericardium (blue arrows) beats against the stomach wall. The myocardium (yellow arrows) protects the myocardium. (**B-B'**) In early-hatchling embryos (4.5 hpf) the heart progenitors are still unrecognizable. (**C-C'**) In mid hatchling 1 embryos (6 hpf) heart progenitors are recognizable for the first time. (**D-D'**) In mid-hatchling 2 embryos (7.5 hpf) the internal cavity of the heart is already recognizable. (**E-E'**) In late-hatchling embryos (9.5 hpf) the cavities of the trunk are fully expanded and the heart has already started beating. Capital letters represent left optical sections at the level indicated by white squares. Prime letters represent ventral optical sections at the level indicated by red lines. Discontinuous white lines delimit the shape of the heart during organogenesis. Discontinuous green lines delimit the notochord. Asterisks indicate the internal cavity of the heart. Intestine (In), Notochord (N), Stomach (St). Scale bar represent 20µm in big images and 60µm in small images.

therefore remained unknown. Thus, we first performed a morphological study by DIC microscopy of *O. dioica* embryos to provide the first developmental atlas of the appendicularian heart (**Fig. 1B-E'**). Up to the early hatchling stage (4.5 hpf) (**Fig. 1B**), no morphological evidence could be distinguished to recognize the position of the cardiac precursor cells. By mid-hatchling
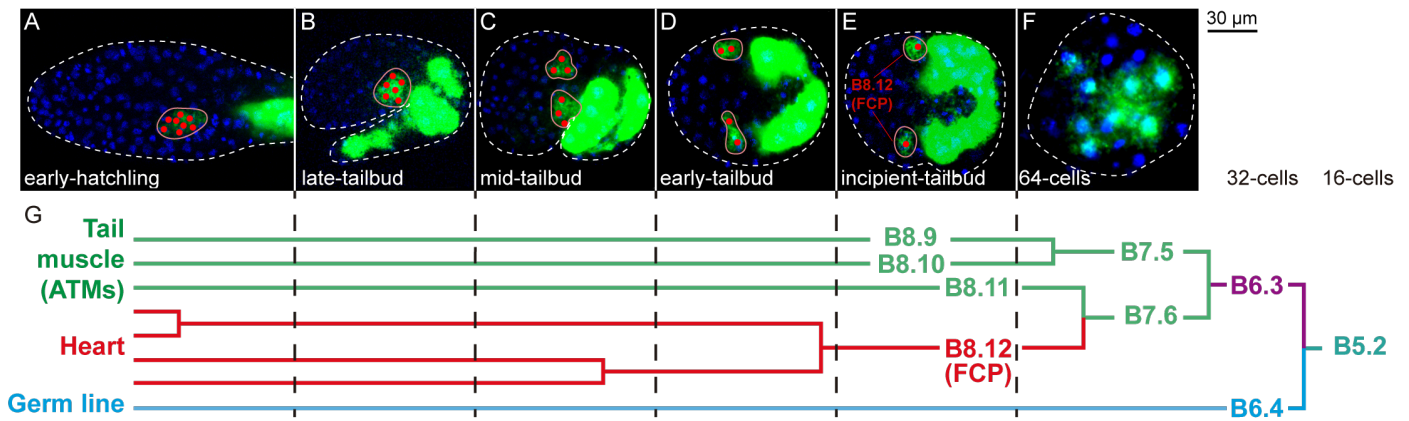
Fig 2. Integration of *ActnM1* expression with the fate map of the tail muscle, the heart, and the germline in *O. dioica*. (A-F) Dorsal sections of different embryo stages stained with *ActnM1* to detect the cardiac progenitors throughout development. Nuclei are stained with Hoechst (blue). (G) Cell lineage diagram of B5.2 blastomere that originates the three anterior muscle cells of the tail, the heart, and the germline (Modified from Stach et al., 2008). To facilitate comparisons, the nomenclature of the blastomeres is according to that of Conklin for ascidians (1905). Fates are color-coded. Close shapes encircle cardiac progenitors. White dashed lines enclose the embryo. Red dots indicate the nuclei of the cardiac precursors. FCP, first cardiac progenitor; ATM, anterior tail muscle cell.

stage (6 hpf), when the borders delineating organ primordia started to appear throughout the trunk, a striated morphology distinguished from a left-side view allowed us to recognize the heart primordium for the first time located in a ventral position close to the ventral midline anterior to the notochord and between the already apparent stomach lobes and intestine (**Fig. 1C**). By 7.5 hpf, ventral view revealed the beginning of the expansion of an internal cavity separating the prospective myocardium and pericardium (**Fig. 1D'**). By 8.5 hpf, the heart started beating suggesting that its contractile properties were already functional (**Sup. video 1**). By late-hatchling stage (9.5 hpf), the internal cavity of the heart was fully expanded, and the position of the heart had asymmetrically displaced towards the left side trunk (**Fig. 1E-E'**).

To study the development of the heart at stages at which its precursor cells could not be recognized by their morphology (i.e. early- and pre-hatchling stages), we first attempted to use the expression of *Mesp* as the preferred marker used in ascidians to trace the development of the cardiac cell lineage (Davidson, 2007). We surprisingly found, however, that *O. dioica* had lost its *Mesp* homolog (see section 1.2 below), and therefore we used *muscular Actin 1* (*ActnM1*) as an alternative cardiac marker since we had previously found that it was expressed in the heart during *O. dioica* development (Almazán et al., 2019). *ActnM1* helped us to trace back and describe the cardiac cell lineage in reverse temporal order from hatchlings to 16-cell stage embryos, and therefore to identify the early blastomeres that give rise to the heart (**Fig 2A-G**). Thus, in the ventral area of the trunk, adjacent to the tip of the notochord at early-hatchling stage, we observed one single domain of *ActnM1* expression in which we counted 8 prospective cardiac cells (**Fig. 2A**). While one single *ActnM1* expression domain was still observed at late-tailbud stage (**Fig. 2B**), two distant bilateral domains were distinguished at mid- and early-tailbud stages, in which we counted 3 and 2 prospective cardiac cells in each domain, respectively (**Fig. 2C-D**). At Incipent tailbud stage, the bilateral *ActnM1* expression domains were made of one single cell adjacently located to the first anterior tail muscle (ATM) cell (**Fig. 2E**). Integration of the expression data of *ActnM1* with the 4D-nuclear tracing (**Sup. Fig. 1**) (Stach et al., 2008) from tailbud up to the 16-cell stage revealed that cardiac cells shared lineage with the first three ATM cells (B8.9, B8.10, and B8.11; **Fig. 2G**), and allowed us to identify B8.12 blastomere as the first cardiac progenitor cell (FCP) in *O. dioica* (**Fig. 2E, G**). We noticed that the onset and temporal progression of the *ActnM1* expression occurred earlier in the left than in the right side, suggesting an asynchronous bilateral asymmetry on the development of the FCP cells. At 64-cell stage, we observed *ActnM1* expression in cells of the vegetal hemisphere compatible with B7.5 and B7.6, which are the precursor cells of the ATM and the FCP (**Fig. 2F-G**). At 32-cell stage, according to the 4D cell fate map reconstruction (**Sup. Fig. 1**) we inferred that B6.3 should be considered the first precardiac precursor that will give rise to both the heart and ATM lineages (**Fig. 2G**), but no *ActnM1* signal was distinguished over background levels at this stage. B6.3 is the sister cell of the germline precursor B6.4, both of which

61

descended from B5.2 located in the vegetal hemisphere at the 16-cell stage (**Fig. 2G**). Thus, our findings allowed us to infer that the cardiac precursors of appendicularians and ascidians shared cell lineage fate maps and origin from the same early blastomeres. Moreover, our observations in *O. dioica* suggested that the FCP was also originated from an asymmetric division in the tail-trunk interface that also gives rise to the anterior tail muscles in the same way as it occurs in ascidians, in both cases ending in a final ventral distant position of the trunk from the anterior muscle cells of the tail. These findings, therefore, provide solid ontogenetic evidence that the heart of ascidians and appendicularians were homologous.

3.2 Loss of an early multipotent state module during the deconstruction of the cardiopharyngeal GRN

To investigate the evolution of the cardiogenetic toolkit of *O. dioica*, we performed a comprehensive in silico survey of conserved components of the GRN responsible for the development of the cardiac progenitor cells in ascidians and vertebrates (Diogo et al., 2015; Razy-Krajka and Stolfi, 2019). In order to map gene losses in the context of tunicate evolution, we included in the survey the genomes of six other appendicularian species (Naville et al., 2019) as well as eleven ascidian species (Brozovic et al., 2018).

First, we started by searching the *O. dioica* homolog of *Mesp*, which is the earliest known marker of pre-cardiac progenitors in ascidians and vertebrates (Bondue and Blanpain, 2010; Davidson et al., 2005; Satou et al., 2004). Using ascidian *Mesp* and vertebrate *Mesp1* and *Mesp2* proteins as tBLASTn queries, however, none of the resulting hits returned *Mesp* in best reciprocal blast hits (BRBH), but members of other bHLH gene families such as *Math*, *Achaete scute*, *Neurogenin*, and *Hand* (**Sup. Table 3**). Analyses by tBLASTn of the other six appendicularian species also revealed the absence of *Mesp* homologs in all their genomes, which contrasted with its presence in all eleven ascidian species analyzed (**Fig. 3**). The apparent absence of *Mesp* among BLAST hits was confirmed by phylogenetic analyses, suggesting that appendicularians had lost the homolog of *Mesp* (**Fig. 3** and **Sup. Fig. 2A**). To test for the possibility of 'function shuffling' among bHLH genes upon the loss of *Mesp*, we checked for the expression of the gene that gave the most significant blast hits, but we did not observe any expression domain compatible with the position of the cardiac progenitors (**Sup. Fig. 2B-G**). These findings, therefore, suggested that the loss of *Mesp* occurred at the base of the appendicularian clade after its split from the ascidian lineage, and therefore, the activation of the cardiac pathway in appendicularians became *Mesp*-independent.

We continued our survey with the search of homologs of *Ets1/2*, the direct downstream target of *Mesp* in the TVCs of ascidians, whose phosphorylation by the FGF/MAPK pathway results in the upregulation of the primary cardiogenic transcription factors (Davidson et al., 2006). BLAST searches revealed the presence of two *Ets1/2* homologs in *O. dioica* as well as in all other appendicularians, and interestingly, our gene survey also revealed the presence of a previously unnoticed second *Ets1/2* gene in all ascidian species (**Sup. Fig. 3A**). Phylogenetic analyses revealed that *Ets1/2* was duplicated at the base of the tunicates, giving rise to two paralogs that we have named as *Ets1/2a* and *Ets1/2b*, being the former the direct target of *Mesp* involved in the specification of the cardiopharyngeal lineage in ascidians. Our phylogenetic tree, however, suggested that the two *Ets1/2* genes of appendicularians were co-orthologs to the ascidian *Ets1/2b* and that they had been generated by a gene duplication occurred within the appendicularian lineage (*Ets1/2b1* and *Ets1/2b2*), while no *Ets1/2a* homolog was present in appendicularians (**Fig. 3B**). Experiments by whole-mount in situ hybridization with *Ets1/2b1* and *Ets1/2b2* in *O. dioica* revealed no expression domains compatible with cardiac precursors but expression in the notochord, the tail muscle cells, the migratory endodermal strand, the ventral organ, and the oikoplastic epithelium (**Sup. Fig. 3B-M**). Our findings, therefore, suggested that *Ets1/2a* was lost in appendicularians, and none of the *Ets1/2b* paralogs played any similar role to *Ets1/2a* in the cardiogenic pathway in ascidians. Moreover, we also analyzed the components of the FGF/MAPK pathway, since in ascidians this is responsible for *Ets1/2a* phosphorylation in the TVCs (Davidson et al., 2006). Our results revealed the absence of many of the key components of the FGF/MAPK pathway, including the absence of the homolog of the *MEK1/2*, which in ascidians, through a kinase cascade mediated by ERK, is responsible for Ets1/2a phosphorylation in the cardiogenetic pathway (**Sup. Fig. 4A**). Whole-mount in situ hybridization revealed that ERK was expressed in the oikoplastic epithelium, but no specific signal
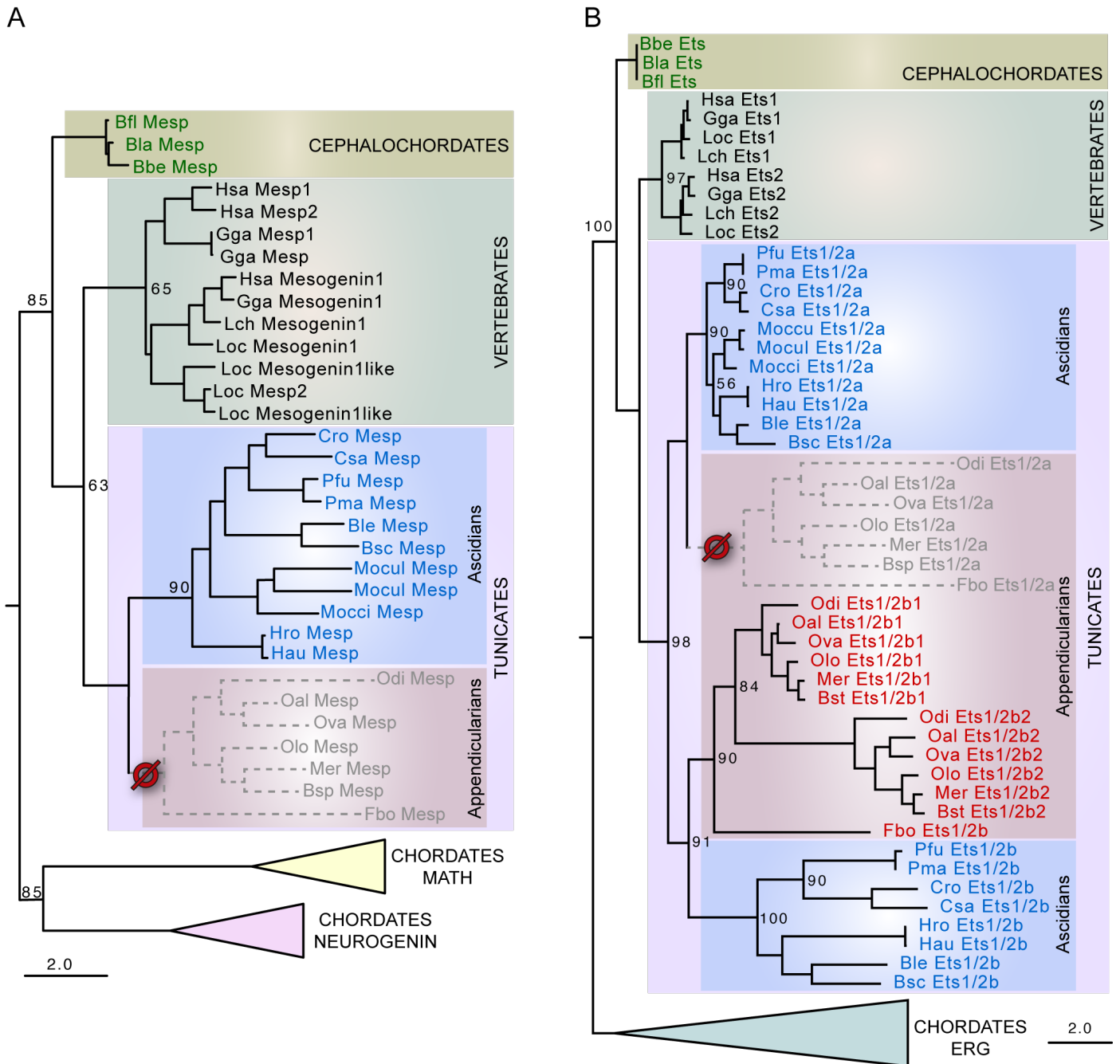
Fig. 3. ML phylogenetic trees of *Mesp* and *Ets* families revealing the loss of *Mesp* and *Ets1/2a* in appendicularians. (**A**) ML phylogenetic tree of *Mesp*, *Math* and *Neurogenin* protein families showed a high bootstrap value separating these protein families and corroborating the loss of *Mesp* in appendicularians. (**B**) ML phylogenetic tree of *Ets* and *Erg* protein families showed a high bootstrap value separating both protein families and revealed that the ancestor of tunicates duplicated *Ets1/2,* and the ancestor of appendicularians lost *Ets1/2a* but duplicated *Ets1/2b*. Scale bar indicates amino acid substitutions. Bootstrap values are shown in the nodes. Vertebrates: *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Tunicates: *Bathochordaeus sp.* (Bsp). *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Fritillaria borealis* (Fbo), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Mesochordaeus erythrocephalus* (Mer), *Molgula occidentalis* (Moocci), *Molgula occulta* (Mooccu), *Molgula oculata* (Moocul), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Cephalochordates: *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bfl).

domain was observed in the cardiac region (**Sup. Fig. 3B-H)**.

Next, we searched for *O. dioica* homologs of *Gata4/5/6*, *FoxF*, *Nk4,* and *Hand1/2* families since in ascidians phosphorylated Ets1/2a activates *FoxF* and *Gata4/5/6*, two members of the primary cardiac transcription factors (CTF) in

charge of cardiopharyngeal migration and specification, respectively (Beh et al., 2007), and other downstream members of the cardiopharyngeal kernel as *Nk4, Hand1/2,* and *Hand-r* (Davidson, 2007; Ragkousi et al., 2011; Wang et al., 2019). Blast searches and phylogenetic analyses revealed the absence of *Gata4/5/6* homologs in *O. dioica*, but the

presence of four paralogs of the closely related *Gata1/2/3* (**Sup. Fig. 5)**. The absence of *Gata4/5/6* homologs in all six appendicularians, but their presence in all ascidians, suggested that the loss of *Gata4/5/6* occurred at the base of the appendicularian lineage, which was accompanied by an expansion of the *Gata1/2/3* subfamily (**Sup. Fig. 5**). In the case of *FoxF* and *Nk4*, we did find a single copy of each one in *O. dioica* and all appendicularian species (**Sup. Fig. 6 and 7**). Finally, blast searches revealed the presence of a single *Hand1/2* homolog in *O. dioica*, as well as in all other six appendicularian species, and the presence of a second paralog *Hand-r* in all ascidian species. Phylogenetic analyses suggested that *Hand-r* was originated by an ascidian-specific duplication, and therefore its absence from appendicularians was not due to a gene loss, but to an innovation of the ascidian group (**Sup. Fig. 8**).

In an initial expression analysis by colorimetric WMISH of members of the *Gata1/2/3*, *FoxF*, *Nk4,* and *Hand1/2* gene families, we observed expression domains that were in the nearby of the presumptive cardiac region. Therefore, we performed double fluorescent in situ hybridization with *ActnM1* to test if any of these genes were expressed in the cardiac progenitors (**Fig. 4**). At the incipient tailbud stage, we observed that *Nk4* was the first CTF to be expressed in the FCP, as soon as B8.12 was originated from the division of B7.6 but, we did not observe any signal in the daughter cell B8.11 that remained positioned posteriorly as the first ATM (**Fig. 4A**). Expression of *Nk4* was also visible earlier on the left than on the right side, consistently with the asynchronous bilateral asymmetry on the development of the FCP also revealed by *ActnM1* expression (**Fig. 4A, A'**). By mid/late-tailbud stage, we observed that *Nk4* signal faded and was more difficult to detect (**Fig. 4A''**). At that stage, we observed *Hand1/2* signal for the first time in some embryos in the cardiac progenitors localized in the ventral midline of the trunk (**Fig. 4B''**). However, we did not observe *FoxF* signal in the cardiac precursors (**Fig 4 C-C'''**), but in adjacent epidermal cells that continued positive for *FoxF* to finally become restricted to the fol domain of the oikoplastic epithelium (**Fig. 4C'''**). Despite the absence of *Gata4/5/6* homolog in *O. dioica*, we tested for the possibility of function shuffling among paralogs by checking the expression of two of the four *Gata1/2/3* paralogs that seemed to have expression in the nearby of the cardiac area. Double fluorescent experiments, however,

revealed that the *Gata1/2/3b* and *Gata1/2/3d* were not co-expressed with *ActnM1* in the FCP, but in an adjacent epidermal cell (**Fig. 4D-D' and E-E'**). At late-tailbud and hatchling stages, the expression of *Gata1/2/3b* and *Gata1/2/3d* had expanded in lateral epidermal domains and anterior prospective placodal populations, respectively, but were never detected in the heart primordium (**Fig. 4D''-D''' and E''-E'''**). These results, therefore, suggest that *FoxF* and *Gata* factors did not participate in the specification or migration of the cardiac progenitors in *O. dioica*, in contrast to their roles in ascidians and vertebrates. Overall, our findings of the gene co-elimination of *Mesp*, *Ets1/2a*, *MEK1/2*, *Gata4/5/6* and RA signaling, and the loss of the cardiac subfunctions of *FoxF*, highlight a process of evolutionary deconstruction of the cardiopharyngeal gene regulatory network in *O. dioica* with the loss of what can be considered an "early multipotent" (EM) module that in ascidians is related to the early maintenance of the multipotent state of the TVCs after their split from the ATM lineage. In *O. dioica*, the cells resulting from the split of the ATM lineage rapidly activated the onset of the cardiac kernel (i.e. *Nk4* and *Hand1/2*), the reason why we have termed them first cardiac precursors (FCP), rather than TVCs as in ascidians.

3.3 Loss of a late multipotent state module during the deconstruction of the cardiopharyngeal GRN

In ascidians, the TVCs undertake a series of asymmetric cell divisions and regulatory transient states through a process of binary fate decisions that do not only give rise to first and second heart precursors, but also to the atrial siphon muscle founder cells (ASMF) (Stolfi et al., 2010). In our work, we then investigated *O. dioica* homologs of *Hand-r*, *Tbx1/10*, *Islet1* and *Ebf* (*Coe*), which in ascidians become activated in a FGF/MAPK-dependent manner to determine the trajectory towards the ASMF including the activation of *MyoD* (*Mrf*) (Stolfi et al., 2010; Wang et al., 2013, 2019). Our genome survey and phylogenetic analysis revealed that in addition to the absence of *Hand-r* (see above), none of the *Tbx* genes of *O. dioica* belonged to the *Tbx1/10* family (**Sup. Fig. 9A**). The absence of *Tbx1/10* in all examined appendicularian species suggested that this gene was lost at the base of the appendicularian clade after its split from the ascidian lineage (**Sup. Fig. 9A**). In the case of *Islet1* and *Ebf*, despite the existence of one homolog of each in *O. dioica*, our results by WMISH revealed that the two of them were expressed in the nervous system, but no
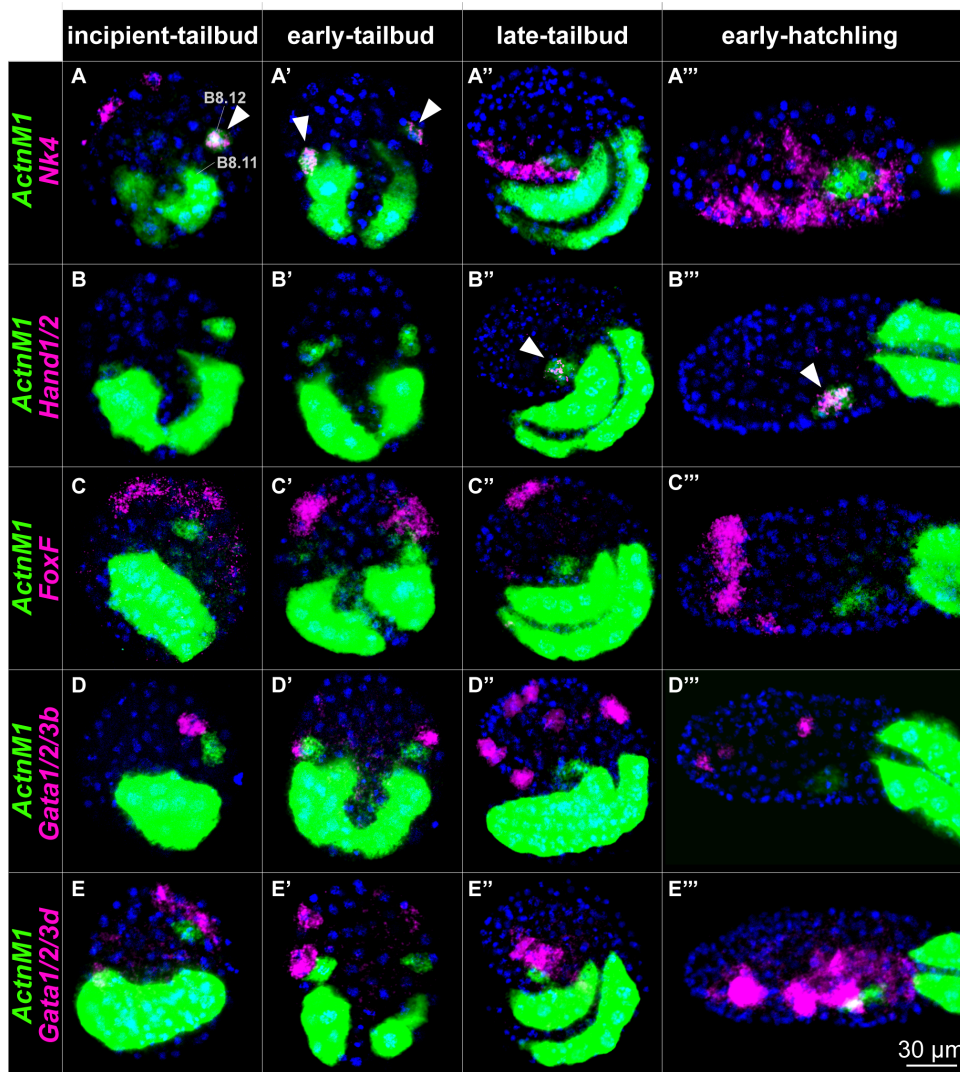
Fig. 4. Developmental expression patterns of *O. dioica ActnM1* and prospective cardiac transcription factors. Double fluorescent in situ hybridization of *ActnM1* with *Nk4*, *Hand1/2*, *FoxF*, *Gata1/2/3b*, and *Gata1/2/3d*. *Nk4* expression signal was detected in ventral epidermis and the FCP (B8.12) from the inipient-tailbud stage (**A**) until the early-tailbud (**A'**). In later stages, we only detected expression in the epidermis, but not in the cardiac precursors (**A"-A'''**). *Hand1/2* was specifically expressed in the cardiac progenitors from late-tailbud to hatchling stages (**B-B'''**). We did not detect expression of *FoxF*, *Gata1/2/3b*, nor *Gata1/2/3d* in cardiac precursors, but they were expressed in different epidermal domains (**C-E'''**). Incipient- and early-tailbud stages correspond to ventral views oriented anterior towards the top. Late-tailbud and early-hatchling stages correspond to lateral views oriented anterior towards the left and dorsal towards the top. White arrows indicate coexpression of *ActnM1* with the corresponding gene in cardiac progenitors.

expression was found that could suggest the presence of presumptive muscle cells in the gill slits homologous to the atrial muscle cells of ascidians (**Sup. Fig. 9B-L**). Finally, we have identified the *MyoD* homolog in *O. dioica* which did not appear to be expressed in muscle cells, but its expression domains were restricted to some fields of the oikoplastic epithelium during late hatchling stages (**Sup. Fig. 9M-Q**). Our results, therefore, revealed that the loss of pharyngeal muscle during the evolution of appendicularians was accompanied by the loss of *Tbx1/10* and the loss of the muscle subfunctions of *Islet1, Ebf* and *MyoD*.

To test for the presence of presumptive second heart field in *O. dioica*, we investigated the expression of the homolog of *Dach*, which in ascidians is activated by *Tbx1/10* in the absence of FGF/MAPK-signaling, and it is sufficient to determine the second heart precursors identity (Wang et al., 2019). Our genome survey revealed the presence of a single homolog of *Dach* in *O. dioica* (**Sup. table 3**). Experiments of WMISH revealed, however, that *Dach* was expressed in the nervous system, the endostyle, and the trunk epidermis, but not in the heart of *O. dioica* (**Sup. Fig. 10**). The absence of *Dach* expression in the heart could suggest that *O. dioica* might lack a homolog to the second heart field of ascidians. In addition to *Dach*, single-cell transcriptomic analysis in ascidians has revealed a total of 18 cell-specific gene markers for the first heart precursors and 7 for the second (Wang et al., 2019). Our gene survey by BRBH revealed that 12

out of those 25 cell-specific gene markers were absent in the genome of *O. dioica* (**Sup. table 3**). Using the human genome as an outgroup, the BRBH results were compatible with 4 gene losses in *O. dioica* (3 and 1 for FHP and SHP specific gene markers, respectively), and 8 ascidian specific gene duplications (6 and 2 for FHP and SHP markers, respectively). Altogether, our results highlights that during the deconstruction of the cardiopharyngeal gene regulatory network the loss of *Tbx1/10* and the loss of subfunctions of *Dach, Islet1, Ebf* and *MyoD* might represent the loss of a late module of multipotent states that in ascidians are responsible of the differentiation of the SHF and atrial muscle, structures that appeared to have been lost during the evolution of appendicularians.

Finally, to investigate how the daughter cells of the cardiac progenitors differentiated into myocardium and pericardium in hatchling stages, we analyzed the expression patterns of some of the CTFs we had found in *O. dioica* such as *Hand1/2, Rapostlin (FNBP1),* and *Nk4* as well as some structural genes that code for motor proteins such as *ActnM1, Myosin, FilaminC (FlnC),* and *Troponins T (TnnT)* (**Fig. 5A-I"**). WMISH experiments revealed that the expression of *Hand1/2,* that started after the downregulation of *Nk4,* was specifically maintained in the heart throughout all hatchling stages (**Fig. 5A-A"**). Despite the first heart beatings did not occur until late-hatchling (8,5 hpf), in the early-hatchling stage (5 hpf), in addition to *ActnM1,* we also observed the first expression signals of some motor genes such as *FilaminC* and *TnnT7* (**Fig. 5F, I**). By mid-hatchling stage (6,5 hpf), in addition to *Hand1/2,* we observed expression signals of the CTF *Raspostlin,* as well as additional motor genes such as *Myosin, TnnT1,* and *TnnT4* (**Fig. 5B', E', G', H'**). Finally, by late-hatchling stage (8 hpf), we observed a second wave of *Nk4* expression, which had been downregulated since mid-tailbud stage (**Fig. 4A"**). The ventral view of late hatchlings, in which the expansion of internal cavities allowed us to clearly differentiate the myocardium and pericardium, revealed that some of the analyzed genes were expressed both in the pericardium and myocardium (*Hand1/2, Rapostlin*), while others only in the myocardium (*Nk4, mAct1, TnnT1, TnnT4, Tnnt7,* and *FilaminC*). These results suggest that despite the loss of the early and late modules, many downstream regulators and structural genes that characterize the cardiac program in ascidians and vertebrates were conserved in *O. dioica,* and our results provide the first insights into the spatio-temporal dynamics of the expression patterns of CTF and motor-genes that will be useful for further functional investigations of heart development in appendicularians.

3.4 FGF and BMP signaling are not involved in the specification of cardiac progenitors cells in O. dioica

In ascidians, the FGF/MAPK signaling via phosphorylation of Ets1/2a is essential for the onset of the pan-cardiac program in the cardiopharyngeal precursors (Davidson et al., 2006). This induction event takes place only in the anterior pairs of B7.5 daughters turning them into TVCs, while the posterior B7.5 descendants become into ATMs (Davidson et al., 2006; Tolkin and Christiaen, 2012). The loss of *Ets1/2a* in appendicularians, the absence of *Ets1/2b* expression domains compatible with a cardiac function, and the loss of *MEK1/2* in the FGF/MAPK pathway suggested the hypothesis that the determination of the FCP could have become FGF independent in *O. dioica.* To test this hypothesis, we performed FGF inhibitory treatments with SU5402, an inhibitor of FGF receptor (FGFR). Treatments with SU5402 induced developmental alterations, whose severity depended on the concentration and duration of the treatment (**Sup. Table 4**). In treatments starting at 2-cell stage, SU5402 concentrations at 50 μM induced obvious malformations at gastrula stage, affecting the proper formation of tailbud morphologies and arresting development before hatching (**Sup. Table 4**). In treatments starting after gastrulation at 32-cell stage, however, SU5402 concentrations of 50 μM did not produce any obvious altered phenotype, and required concentrations of at least 100 μM to induce aberrant morphologies in hatchling stages (**Sup. Table 4**). WMISH experiments in embryos treated with SU5402 at 50 μM from the 2-cell stage revealed that a majority of the embryos showed aberrant or no *ActnM1* signal, suggesting that the development of mesodermal derivatives such as the muscle lineage was severely affected (**Fig. 6B-B'**). Interestingly, those few embryos that were able to achieve incipient tailbud morphologies with two recognizable rows of tail muscle cells labeled by *ActnM1,* also showed the presence of presumptive cardiac progenitors, comparably to DMSO treated control embryos (**Fig. 6B"**). In embryos treated with SU5402 at 100 μM from the 32-cell stage, in which gastrulation was not affected, WMISH with *Nk4* revealed the presence of cardiac precursors in
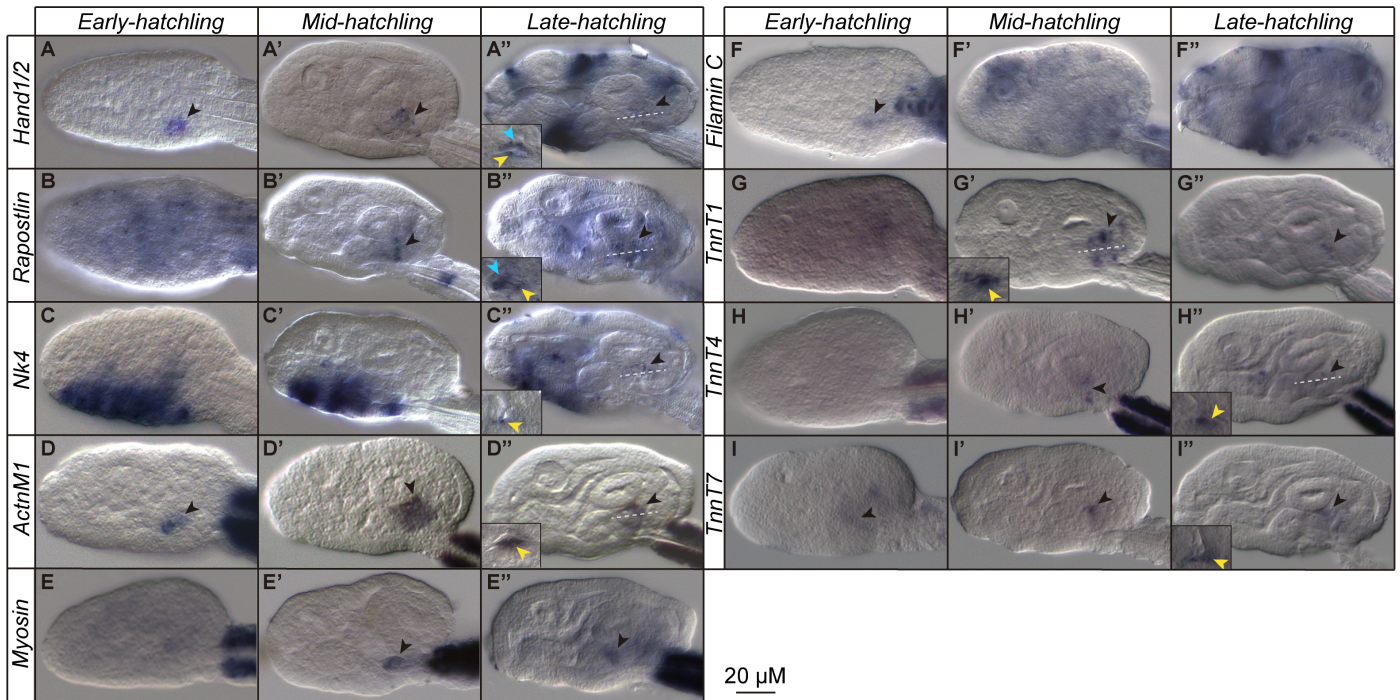
Fig. 5. **Expression atlas of cardiac genes during hatchling stages of *Oikopleura dioica* development**. Developmental expression analysis by whole mount in situ hybridization of *O. dioica* cardiac transcription factors (**A-C"**) and motor proteins (**D-I"**). Images correspond to lateral views oriented anterior to towards the left and dorsal towards the top. Detailed small images represent ventral optical section at the level indicated by white lines. Black arrows indicate expression in the heart. Blue arrows indicate expression in the pericardium. Yellow arrows indicate expression in the myocardium.

most embryos, similarly to control DMSO embryos at incipient tailbud stage (**Fig. 6D-D'**). These results suggested that despite FGF appeared to be important for early development during gastrula stage, FGF was not essential for the specification of the cardiac progenitors at incipient tailbud stage. These results together with the loss of *Mesp*, *Ets1/2a* and *MEK1/2* support the hypothesis that the specification and induction of cardiac progenitors have become FGF independent in appendicularians, in contrast to ascidians.

In ascidians, in addition to FGF, BMP signaling is also relevant for cardiac development contributing to the specification and migration of the TVCs. The migration, activated by *FoxF*, follows a gradient of BMP originated in the ventral epidermis (Christiaen et al., 2010). The increasing levels of BMP initiate the upregulation of cardiac kernel genes as *Gata4/5/6* or *Nk4*, which sustains and arrests migration, respectively (Christiaen et al., 2010). Our results showing the loss of *Gata4/5/6* and the apparent absence of expression of the migratory factor *FoxF* in the FCP of *O. dioica* suggested the hypothesis that the role of BMP in the cardiac gene regulatory network could have changed in appendicularians. To test this hypothesis, we performed BMP inhibitory treatments with LDN –a highly specific inhibitor of BMP receptors–

during different time windows (**Sup. Table 4**). WMISH experiments using *ActnM1* and *Nk4* probes in tailbud embryos that had been previously treated with 10 µM LDN from the 2-cell stage revealed no differences in the formation of the FCPs nor in the onset of the cardiac kernel marker *Nk4* in comparison to DMSO-control embryos (**Fig. 6F, H**). These results suggested that BMP was not necessary for the specification of the FCP nor the activation of the expression of the cardiac kernel as soon as the FCP split from the ATMs.

We also checked for the convergence of cardiac progenitors in the ventral part of the trunk after treating embryos with FGF and BMP inhibitors. In early hatchling DMSO control embryos, *ActnM1* revealed that cardiac progenitors were localized in a single domain near the trunk midline, while in SU5402 and LDN treated embryos cardiac precursors were bilaterally distributed far from the midline (**Fig. 6J, J", K**). Interestingly, both, BMP and FGF inhibitory treatments, also affected the elongation of the tail and its rotation relative to the trunk. Thus, while a ventral view of control embryos showed a row of muscle cells (**Fig. 6I**), most of the treated embryos showed two rows revealing that the rotation had not taken place (**Fig. 6J, J", K**). These results
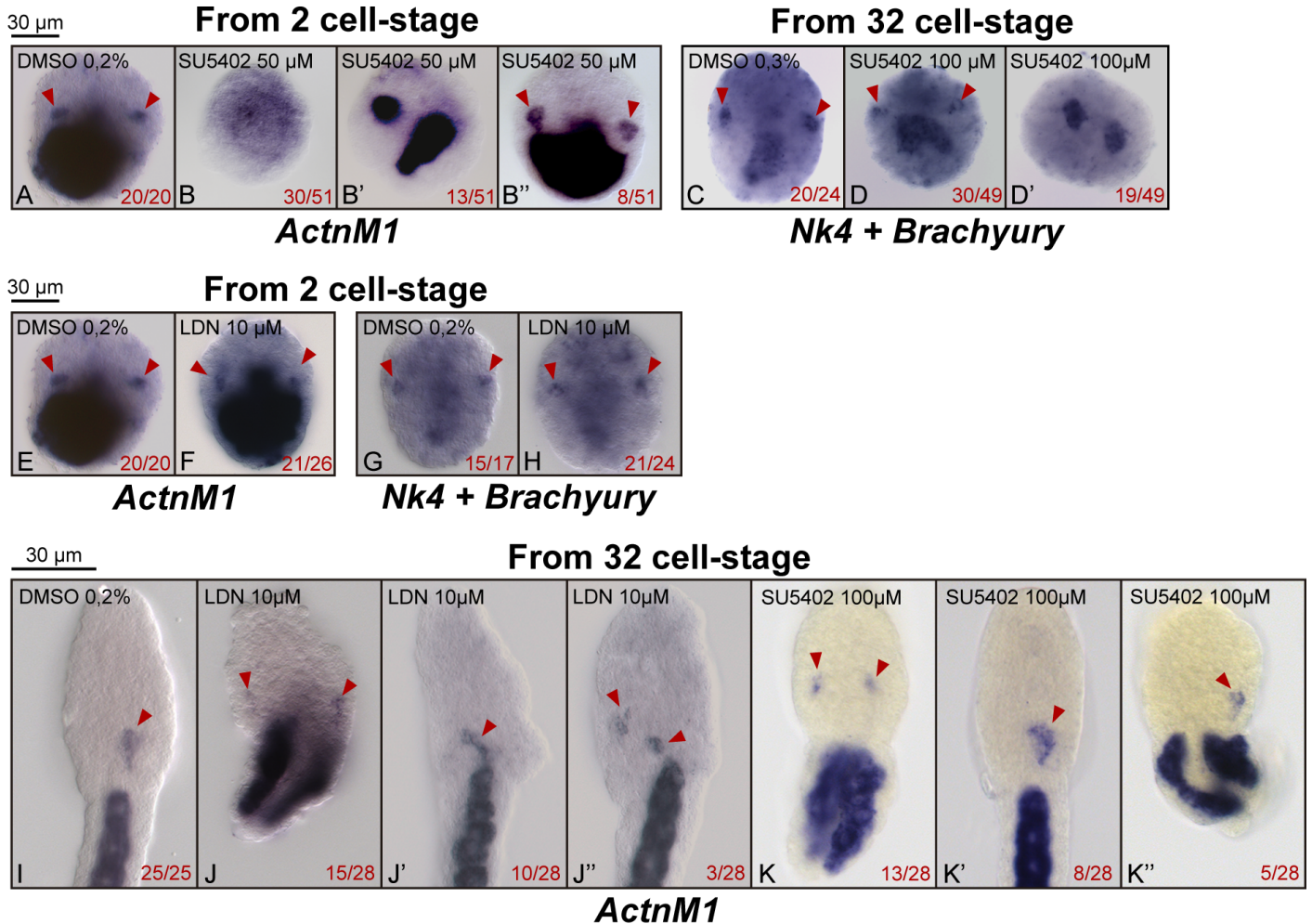
Fig. 6. Effects FGF and BMP inhibitions in the development of the heart of *O. dioica*. Whole mount in situ hybridization of *ActnM1* in DMSO-control (**A**) and treated embryos with SU5402 50μM starting at 2-cell stage (**B-B"**), in which more than a half of the embryos showed no specific signal of *ActnM1* (**B**). Whole mount in situ hybridization of *Nk4* and *Brachyury* in DMSO-control (**C**) and SU5402 100μM starting at 32-cell stage (**D-D'**), in which more than a half of the embryos showed expression of *Nk4* in the cardiac progenitors but alterations in the notochord development (**D**). Whole mount in situ hybridization of *ActnM1* in DMSO-control (**E**) and LDN 10μM starting at 2-cell stage in which most of the embryos displayed an *ActnM1* expression comparable to the DMSO control (**F**). Whole mount in situ hybridization of *Nk4* and *Brachyury* in DMSO-control (**G**) and LDN 10μM from 2-cell stage in which most of the embryos displayed *Nk4* expression in the cardiac progenitors (**H**). Whole mount in situ hybridization of *ActnM1* in DMSO-control (**I**) and treated embryos with LDN 10μM (**J-J"**) and SU5402 100μM (**K-K"**) starting at 32-cell stage, in which many embryos display separated cardiac progenitors (**J, J", K, K"**), rather than one single domain in the midline such in DMSO-control (I). Tailbud embryos were stained using the cross-hybridizing ActnM1 probe (Almazán et al., 2019). Hatchling embryos were stained with the specific ActnM1 probe (Almazán et al., 2019). Tailbud embryos images correspond to dorsal views with anterior to the top. Hatchling images represent ventral views with anterior to the top. Red arrowheads indicate cardiac precursors.

suggested that the inhibition of BMP- and FGF-signaling could affect the positioning of heart precursors towards the midline, as well as other developmental processes such as the elongation and rotation of the tail.

4 DISCUSSION

4.1 The hearts of appendicularians and ascidians are homologous

The vast morphological variability among hearts and pumping organs across metazoans has made cardiac development a hot topic in the discussion of homologies and analogies in the field of EvoDevo (Xavier-Neto et al., 2007). In the case

of chordates, numerous studies have provided strong evidence supporting that the hearts of vertebrates and ascidians are homologous despite the great morphological differences between the multichambered complex heart of vertebrates and the tubular simple heart of ascidians (Davidson, 2007). At the beginning of our study, however, the striking morphological and physiological differences between the heart of ascidians and *O. dioica* casted doubt on whether these two organs were truly homologous, or on the contrary, could be analogous pumping organs. At the morphological level, the cylindrical V-shape heart of ascidians contrasts with the chamberless flat heart of *O. dioica*. At the physiological level, while
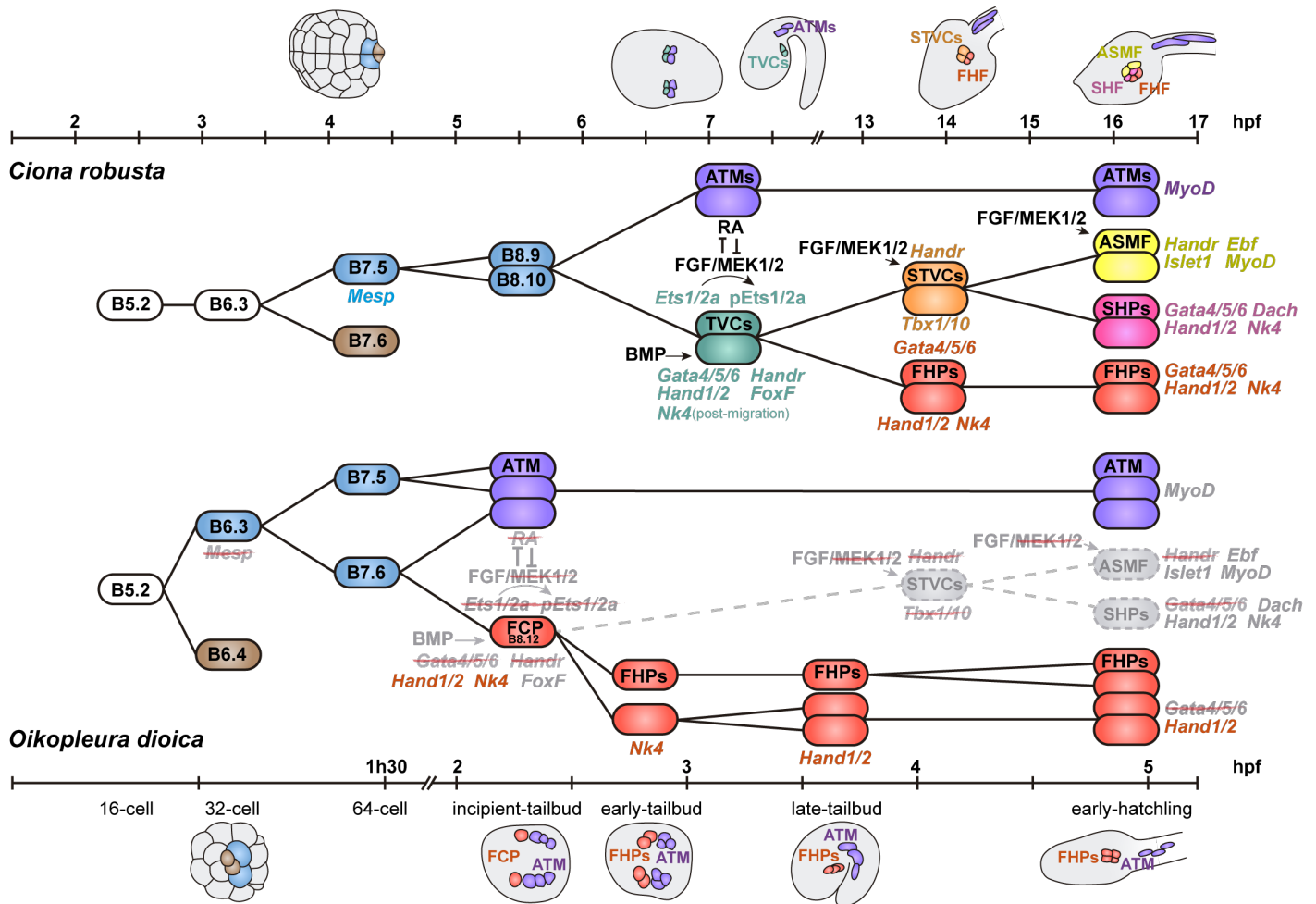
Fig. 7. Cell lineage diagrams of *C. robusta* and *O. dioica* cardiopharyngeal progenitors. The hearts of *Oikopleura* and *Ciona* share the same cell fate map and the same ontogeny with minor differences. While ascidians comprise two TVCs and two ATMs per side, *Oikopleura* only comprises one FCP and one ATM per side. Moreover, in *Oikopleura* the germline (brown), the cardiac precursors (green), the ATM (purple), and the FHPs (red) appear one cleavage earlier than in ascidians. Genes and structures that do not play a role in cardiac development in *O. dioica* are represented in grey, absent genes are strikethrough.

in ascidians the heart does not start beating until several days after the metamorphosis in juveniles, in *O. dioica* the heart starts beating against the stomach wall as soon as 8.5 hpf at late-hatchling stage. Despite these marked differences, our results provide strong evidence that both hearts are homologous by revealing that the cardiac precursor cells of *O. dioica* and *C. robusta* share the same cell fate map and the same ontogeny. In both organisms, the muscle/heart lineage descends from the B5.2 blastomere located in the posterior-vegetal hemisphere at 16-cell stage, which also gives rise to the germline lineage (**Fig. 7**). In *O. dioica*, the division of B5.2 at the 32-cell stage splits the germline and muscle/heart lineages, being B6.3 the precursor of the later. In ascidians, however, the split of the germline and muscle/heart lineages occurs one cleavage later at the 64-cell stage, being B7.5 therefore, the muscle/heart precursor equivalent to B6.3 in *O. dioica* (**Fig. 7**).

Moreover, the origin of the first cardiac precursors in *O. dioica* and *C. robusta* (FCP and TVCs, respectively) share the same ontogenic origin, since in both cases occurs through an asymmetric division at the interface between the trunk and tail, in which the anterior daughter cell will end up in the ventral part of the trunk giving rise to the heart, while the posterior daughter cell will remain in the tail giving rise to the ATMs. The common ontogenic origin of the cardiac lineage and the most anterior axial muscles can be therefore considered an ancestral synapomorphy of tunicates. The timing of the split between the cardiac and the axial muscle lineages between *O. dioica* and *C. robusta* occurs again one cleavage earlier in the appendicularian (after the division of B7.6 at the approximately 110-cell incipient-tailbud stage) than in the ascidian (after the divisions of B8.9 and B8.10 at the approximately 300-cell post-neurula stage). This timing difference is probably connected to the difference in the number of cardiac precursors

69

between *O. dioica* and ascidians, with just one single FCP in each side in *O. dioica* in contrast to the two TVCs in each side in ascidians, as well as just one single row of muscle cells in each side of the tail of appendicularians in contrast to the two rows in the tail of ascidians (Almazán et al., 2019; Nishino et al., 2000).

The "one-cleavage earlier" trend of fate decisions observed in this work is consistent with the observations already made by Delsman back in 1910 (Delsman, 1910) when describing that cell ingression during gastrulation in *O. dioica* occurred one cleavage earlier than in ascidians. This idea that fate restriction occurs earlier in *O. dioica* than in ascidians has been corroborated by the characterization of the cell lineage fate map in *O. dioica* (Nishida and Stach, 2014; Stach et al., 2008), as well as in expression analyses showing for instance that the onset of *Otx* (Cañestro and Postlethwait, 2007) and *Brachyury* (Bassham and Postlethwait, 2000) during early embryogenesis occurred "one-cleavage earlier" than in ascidians. This trend of "one-cleavage earlier" fate decision in *O. dioica* probably reflects a generalized feature of the evolution of the developmental program of appendicularians, which might have likely contributed to the acceleration of development, morphological simplification and cell number reduction of this group of tunicates.

4.2 Modular deconstruction of the cardiopharyngeal gene regulatory network

One of the most striking findings of our work is the numerous losses of genes and cardiopharyngeal subfunctions that have suffered the cardiopharyngeal GRN in appendicularians. In agreement to the modular model for the control of heart cell identity proposed by Wang et al., (2019)(Wang et al., 2019), our comparative analysis between ascidians and appendicularians allows proposing an evolutionary scenario in which we can recognize how the modular deconstruction of the cardiopharyngeal GRN might have affected the evolution of the heart and siphon muscle in appendicularians (**Fig. 7**). In ascidians, the binary fate decision between the ATM and TVC lineages occurs in the trunk-tail interface, under the antagonistic posterior influence of retinoic acid signaling promoting the fate of ATMs in the tail, and the anterior influence of FGF signaling promoting the multipotent cardiopharyngeal fate of the TVCs in the trunk (Christiaen et al., 2008b; Davidson et al., 2006; Nagatomo and Fujiwara, 2003). FGF action is mediated by the phosphorylation of Ets1/2a, which had been previously upregulated by *Mesp*

in the TVCs (Christiaen et al., 2008; Davidson et al., 2006; Nagatomo and Fujiwara, 2003). The multipotent state of the TVCs is maintained until the *FoxF*-dependent ventral migration of the TVCs, and the action of *Gata4/5/6* and BMP-signaling from the ventral epidermis that triggers the cardiogenic kernel upon the activation of *Nk4* (Beh et al., 2007; Davidson and Levine, 2003; Satou et al., 2004; Wang et al., 2019). In *O. dioica*, however, our results show that the binary fate decision between the ATM and FCP lineages occurs in a very different molecular context due to the gene losses of *Mesp* and *Ets1/2a*, the loss of *MEK1/2* affecting the FGF/MAPK signaling pathway, and the loss of the retinoic acid signaling pathway (**Fig. 7**) (Martí-Solans et al., 2016). Our results also show that, in contrast to TVCs, FCPs activate the expression of *Nk4* as soon as B8.12 blastomeres are born from the division of B7.6, suggesting that the FCP is not multipotent and has already activated the cardiogenic kernel. This is the reason why we have termed these cells as FCPs instead of TVCs as in ascidians. Our results, moreover, show that the activation of *Nk4* is independent of *Gata4/5/6* (which has been lost in appendicularians), independent of *FoxF* (which is not expressed in the FCP), and independent of FGF- and BMP-signaling (as shown by our inhibitory treatments). In the light of these results, we propose an evolutionary scenario in which all these losses highlights the deconstruction of an "early-multipotent" (EM) module of the cardiopharyngeal GRN that in ascidians contributes to the maintenance of the multipotent state in the TVCs, which in the FCP of *O. dioica* has been lost.

In ascidians, the multipotent states of TVCs and STVCs allow the differentiation of the second heart field and the atrial siphon and longitudinal muscles. In appendicularians, however, our results revealing the ancestral loss of *Tbx1/10*, the loss of FGF and BMP cardiac inductive functions, together with the losses of expression of *Dach* in the heart and *Islet1*, *Ebf* and *MyoD* in muscles of *O. dioica* highlights the deconstruction of a "late multipotent" (LM) module of the cardiopharyngeal GRN that in ascidians is responsible for the maintenance of the multipotency in the STVC and fate specification of the SHP and ASMF (**Fig. 7**). These results suggest, therefore, an evolutionary scenario in which the heart of *O. dioica* might be homologous to the FHP of ascidians, while the homologous tissues to the SHP, atrial siphon muscles, and longitudinal muscles have been lost.

4.3 Evolution of pelagic appendicularians from a sessile ascidian-like ancestor

The extensive losses affecting the cardiopharyngeal GRN suggest a parsimonious evolutionary scenario in which appendicularians might have suffered a process of evolutionary simplification from an ancestral ascidian-like condition. The modular deconstruction of the GRN is compatible with the adaptive evolution of three innovations that likely facilitated the transition from an ancestral sessile ascidian-like lifestyle to the pelagic house-based free lifestyle of appendicularians: an accelerated cardiac development, the formation of an open-wide laminar heart and the loss of siphon muscle.

The "one-cleavage earlier" trend and the deconstruction of the EM module likely facilitated the earlier activation of the cardiac developmental kernel in appendicularians than in ascidians, causing an accelerated cardiac development that resulted in the activation of the beating of the heart by 8.5 hpf in *O. dioica*, in contrast to few days after metamorphosis in ascidians. Accelerated cardiac development, therefore, could have been plausibly selected as an adaptation to the high energetic demand of appendicularians associated with their intense and constant beating of the tail to power water circulation through the house right after the inflation of the first house as soon as 10 hpf (Ferrández-Roldán et al., 2019; Nishida, 2008).

In addition to accelerated cardiac development, the profound re-modeling of the cardiac cell lineage development with the apparent loss of the SHP in appendicularians plausibly may have contributed to the transformation of an ascidian-like tubular heart into an open-wide laminar structure that beats against the stomach like in *O. dioica*. Considering that hemolymph circulation in appendicularians is not only powered by the heart, but also by the movements of the tail (Fenaux, 1998), the adaptive innovation of a laminar cardiac structure likely offered a more efficient system to pump hemolymph waves propelled by the tail movements through an open-wide structure than through the less accessible space of a tubular ascidian-like heart.

Finally, the loss of muscles to propel fluid inside the organism is likely the result of adaptive evolution during the transition from a sessile
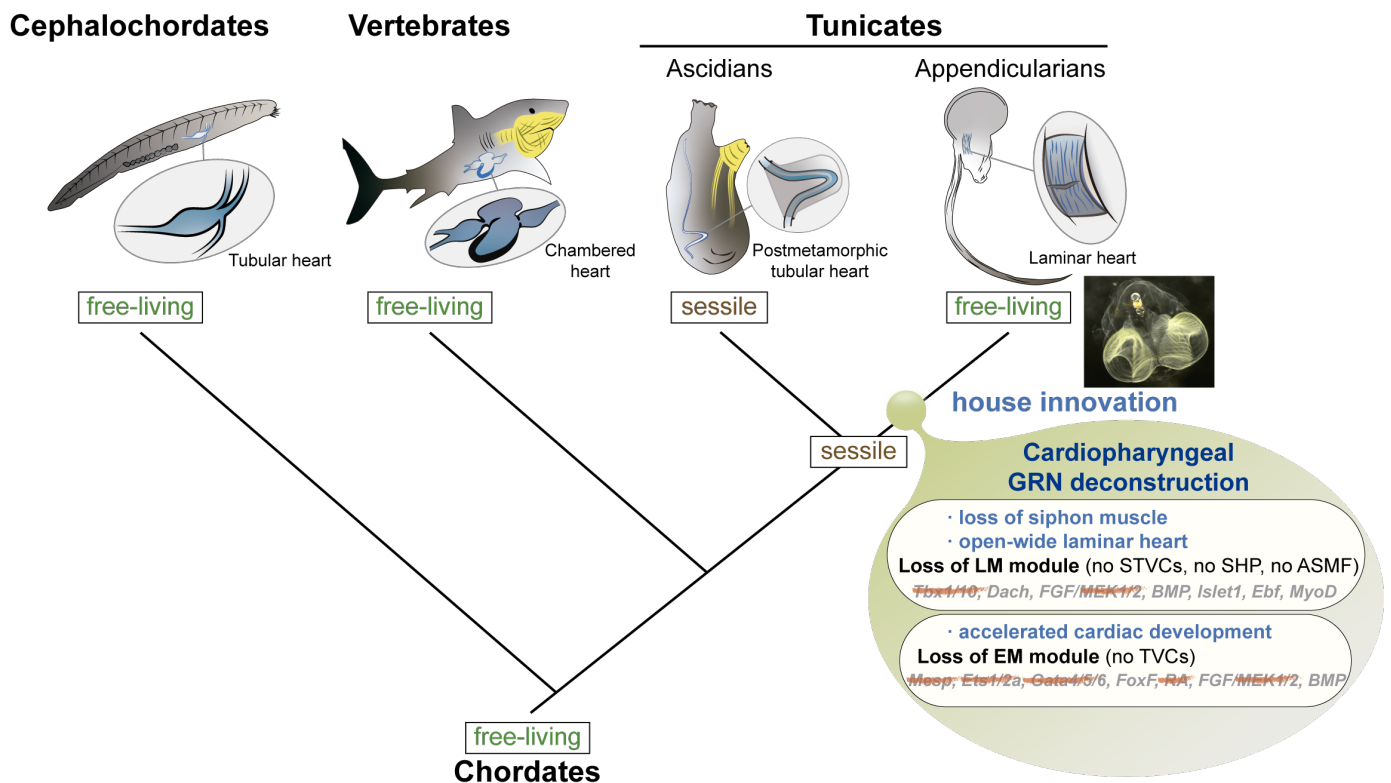
Fig. 8. The deconstruction of the cardiopharyngeal GRN in appendicularians favored their transition from an ascidian-like ancestral condition to a free-living house-based life style. The loss of the siphon muscle, the development of an open-wide laminar heart, and the acquisition of an accelerated development can be considered adaptive consequences of the deconstruction of the cardiopharyngeal GRN that facilitated the innovation of the free-living life style of appendicularians based on the house filtrating device. GRN, gene regulatory network; STVCs, second trunk ventral cells; SHP, second heart progenitors; ASMF, atrial siphon muscle field; TVCs, trunk ventral cells; LM, late multipotent; EM, early multipotent.

71

filter-feeding ascidian-like strategy to a pelagic filter-feeding strategy based on the innovation of the house in appendicularians (Mikhaleva et al., 2018). While in ascidians siphon muscles are responsible for water current throughout the organism during filter feeding, in appendicularians the water current is propelled by actively motile tail and the cilia in the gills (spiracles). Importantly, the loss of siphon muscle in appendicularians probably was not only the consequence of regressive evolution upon the innovation tail-beating propulsion of water through the house, but it plausibly facilitated the evolution of the capability of these tunicates to control the orchestrated movement of the cilia in the gills to efficiently reverse the direction of the water flow, inwards during feeding or outwards during the inflation of the house without the interference of graceless muscle contractions (Bassham et al., 2008; Conley et al., 2018).

In conclusion, our results support an evolutionary scenario in which, independently of the phylogenetic relationship of appendicularians with the rest of tunicates, the last common tunicate ancestor had a sessile ascidian-like lifestyle, and appendicularians suffered a process of evolutionary simplification in which the modular deconstruction of their cardiopharyngeal GRN facilitated the transition to a pelagic free-living style connected to the innovation of the house **(Fig. 8)**. Future comparative studies of the cardiopharyngeal GRN responsible for the development of diverse cardiac structures among different ascidian and appendicularian species will be necessary to provide further insights into the evolution of the heart in tunicates and maybe reveal data that could help to better understand the phylogenetic relationship of appendicularians with the rest of tunicates. In particular, the sequencing of the genome of heartless appendicularian species of the *Kowelevskia* genera will be of special interest (Brena et al., 2003), since it may represent a step beyond the evolutionary process of the complete dismantling of the cardiac GRN and it might help to better recognize GRN modules deconstructed during the evolution of this heartless appendicularian.

5 REFERENCES

Albalat, R., Cañestro, C., 2016. Evolution by gene loss. Nat. Rev. Genet. 17, 379–391. https://doi.org/10.1038/nrg.2016.39

Almazán, A., Ferrández-Roldán, A., Albalat, R., Cañestro, C., 2019. Developmental atlas of appendicularian Oikopleura dioica actins provides new insights into the evolution of the notochord and the cardio-paraxial muscle in chordates. Dev. Biol. 448, 260–270. https://doi.org/10.1016/j.ydbio.2018.09.003

Bassham, S., Cañestro, C., Postlethwait, J.H., 2008. Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. BMC Biol. 6, 35. https://doi.org/10.1186/1741-7007-6-35

Bassham, S., Postlethwait, J., 2000. Brachyury (T) expression in embryos of a larvacean urochordate, Oikopleura dioica, and the ancestral role of T. Dev Biol 220, 322–32.

Beh, J., Shi, W., Levine, M., Davidson, B., Christiaen, L., 2007. FoxF is essential for FGF-induced migration of heart progenitor cells in the ascidian Ciona intestinalis. Development 134, 3297–3305. https://doi.org/10.1242/dev.010140

Berrill, N.J., 1950. The Tunicata with an account of the British species. Sold by Quaritch, London.

Bondue, A., Blanpain, C., 2010. Mesp1: A key regulator of cardiovascular lineage commitment. Circ. Res. https://doi.org/10.1161/CIRCRESAHA.110.227058

Bourlat, S.J., Juliusdottir, T., Lowe, C.J., Freeman, R., Aronowicz, J., Kirschner, M., Lander, E.S., Thorndyke, M., Nakano, H., Kohn, A.B., Heyland, A., Moroz, L.L., Copley, R.R., Telford, M.J., 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444, 85–88. https://doi.org/10.1038/nature05241

Braun, K., Leubner, F., Stach, T., 2020. Phylogenetic analysis of phenotypic characters of Tunicata supports basal Appendicularia and monophyletic Ascidiacea. Cladistics 36, 259–300. https://doi.org/10.1111/cla.12405

Brena, C., Cima, F., Burighel, P., 2003. Alimentary tract of Kowalevskiidae (Appendicularia, Tunicata) and evolutionary implications. J Morphol 258, 225–238.

Brozovic, M., Dantec, C., Dardaillon, J., Dauga, D., Faure, E., Gineste, M., Louis, A., Naville, M., Nitta, K.R., Piette, J., Reeves, W., Scornavacca, C., Simion, P., Vincentelli, R., Bellec, M., Aicha, S. Ben, Fagotto, M., Guéroult-Bellone, M., Haeussler, M., Jacox, E., Lowe, E.K., Mendez, M., Roberge, A., Stolfi, A., Yokomori, R., Brown, C.T., Cambillau, C., Christiaen, L., Delsuc, F., Douzery, E., Dumollard, R., Kusakabe, T., Nakai, K., Nishida, H., Satou, Y., Swalla, B., Veeman, M., Volff, J.N., Lemaire, P., 2018. ANISEED 2017: Extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. Nucleic Acids Res. 46, D718–D725. https://doi.org/10.1093/nar/gkx1108

Cañestro, C., Postlethwait, J.H.H., 2007. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. Dev Biol 305, 522–538. https://doi.org/10.1016/j.ydbio.2007.02.032

Christiaen, L., Davidson, B., Kawashima, T., Powell, W., Nolla, H., Vranizan, K., Levine, M., 2008a. The transcription/migration interface in heart precursors of Ciona intestinalis. Science (80-.). 320, 1349–1352. https://doi.org/320/5881/1349 [pii]10.1126/science.1158170

Christiaen, L., Stolfi, A., Levine, M., 2010. BMP signaling

coordinates gene expression and cell migration during precardiac mesoderm development. Dev. Biol. 340, 179–187. https://doi.org/10.1016/j.ydbio.2009.11.006

Conley, K.R., Gemmell, B.J., Bouquet, J.M., Thompson, E.M., Sutherland, K.R., 2018. A self-cleaning biological filter: How appendicularians mechanically control particle adhesion and removal. Limnol. Oceanogr. 63, 927–938. https://doi.org/10.1002/lno.10680

Davidson, B., 2007. Ciona intestinalis as a model for cardiac development. Semin Cell Dev Biol 18, 16–26. https://doi.org/10.1016/j.semcdb.2006.12.007

Davidson, B., Levine, M., 2003. Evolutionary origins of the vertebrate heart: Specification of the cardiac lineage in Ciona intestinalis. Proc Natl Acad Sci U S A 100, 11469–11473.

Davidson, B., Shi, W., Beh, J., Christiaen, L., Levine, M., 2006. FGF signaling delineates the cardiac progenitor field in the simple chordate, Ciona intestinalis. Genes Dev 20, 2728–2738. https://doi.org/20/19/2728 [pii]10.1101/gad.1467706

Davidson, B., Shi, W., Levine, M., 2005. Uncoupling heart cell specification and migration in the simple chordate Ciona intestinalis. Development 132, 4811–4818. https://doi.org/10.1242/dev.02051

Delsman, H.C., 1910. Contributions on the ontogeny of Oikopleura dioica. Verch. Rijksinst. Onderz. Zee 3, 1–24.

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

Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.K., Turon, X., López-Legentil, S., Piette, J., Lemaire, P., Douzery, E.J.P., 2018. A phylogenomic framework and timescale for comparative studies of tunicates. BMC Biol. 16. https://doi.org/10.1186/s12915-018-0499-2

Devine, W.P., Wythe, J.D., George, M., Koshiba-Takeuchi, K., Bruneau, B.G., 2014. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. Elife 3. https://doi.org/10.7554/eLife.03848

Diogo, R., Kelly, R.G., Christiaen, L., Levine, M., Ziermann, J.M., Molnar, J.L., Noden, D.M., Tzahor, E., 2015. A new heart for a new head in vertebrate cardiopharyngeal evolution. Nature 520, 466–473. https://doi.org/10.1038/nature14435

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Fenaux, R., 1998. Anatomy and functional morphology of the Appendicularia., in: Bone, Q. (Ed.), The Biology of Pelagic Tunicates. Oxford University Press, New York, pp. 25–34.

Ferrández-Roldán, A., Martí-Solans, J., Cañestro, C., Albalat, R., 2019. Oikopleura dioica: An Emergent Chordate Model to Study the Impact of Gene Loss on the Evolution of the Mechanisms of Development, Results and Problems in Cell Differentiation. Springer Verlag. https://doi.org/10.1007/978-3-030-23459-1_4

Fol, H., 1872. Etudes sur les Appendiculaires du De troit de Messine. Mem. Soc. Phys. Hist. Nat. Geneve 21, 445.

Gans, C., Northcutt, R.G., 1983. Neural crest and the origin of vertebrates: a new head. Science (80-.). 220, 268–274.

Garstang, W., 1928. The morphology of the Tunicata, and its bearings on the phylogeny of the chrodata. Quar. J. Micr. Sci. 72, 51–186.

Gee, H., 2018. Across the Bridge: Understanding the Origin of the Vertebrates. University of Chicago Press.

Giribet, G., 2018. Phylogenomics resolves the evolutionary chronicle of our squirting closest relatives. BMC Biol. 16, 4–6. https://doi.org/10.1186/s12915-018-0517-4

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307–321. https://doi.org/syq010 [pii]10.1093/sysbio/syq010

Hogan, B., 2004. Deconstructing the genesis of animal form. Development 131, 2515 LP – 2520. https://doi.org/10.1242/dev.01192

Horie, R., Hazbun, A., Chen, K., Cao, C., Levine, M., Horie, T., 2018. Shared evolutionary origin of vertebrate neural crest and cranial placodes. Nature 560, 228–232. https://doi.org/10.1038/s41586-018-0385-7

Hutson, M.R., Zeng, X.L., Kim, A.J., Antoon, E., Harward, S., Kirby, M.L., 2010. Arterial pole progenitors interpret opposing FGF/BMP signals to proliferate or differentiate. Development 137, 3001–3011. https://doi.org/10.1242/dev.051565

Kocot, K.M., Tassia, M.G., Halanych, K.M., Swalla, B.J., 2018. Phylogenomics offers resolution of major tunicate relationships. Mol. Phylogenet. Evol. 121, 166–173. https://doi.org/10.1016/j.ympev.2018.01.005

Lankester, E.R., 1882. The vertebration of the tail of Appendiculariae. Quar. J. Micr. Sci. 22, 387–390.

Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolosse, A., Auer, H., Achouri, Y., Dubois, C., Bondue, A., Simons, B.D., Blanpain, C., 2014. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. Nat. Cell Biol. 16, 829–840. https://doi.org/10.1038/ncb3024

Martí-Solans, J., Belyaeva, O.V.O. V., Torres-Aguila, N.P.N.P., Kedishvili, N.Y.N.Y., Albalat, R., Cañestro, C., 2016. Coelimination and Survival in Gene Network Evolution: Dismantling the RA-Signaling in a Chordate. Mol. Biol. Evol. 33, 2401–2416. https://doi.org/10.1093/molbev/msw118

Martí-Solans, J., Ferrández-Roldán, A., Godoy-Marín, H., Badia-Ramentol, J., Torres-Aguila, N.P.N.P., Rodríguez-Marí, A., Bouquet, J.M.J.M., Chourrout, D., Thompson, E.M.E.M., Albalat, R., Cañestro, C., 2015. Oikopleura dioica culturing made easy: A Low-Cost facility for an emerging animal model in EvoDevo. Genesis 53, 183–193. https://doi.org/10.1002/dvg.22800

Martik, M.L., Gandhi, S., Uy, B.R., Gillis, J.A., Green, S.A., Simoes-Costa, M., Bronner, M.E., 2019. Evolution of the new head by gradual acquisition of neural crest regulatory circuits. Nature 574. https://doi.org/10.1038/s41586-019-1691-4

Mikhaleva, Y., Skinnes, R., Sumic, S., Thompson, E.M., Chourrout, D., 2018. Development of the house secreting epithelium, a major innovation of tunicate larvaceans, involves multiple homeodomain transcription factors. Dev. Biol. 443, 117–126. https://doi.org/10.1016/j.ydbio.2018.09.006

Nagatomo, K., Fujiwara, S., 2003. Expression of Raldh2, Cyp26 and Hox-1 in normal and retinoic acid-treated Ciona intestinalis embryos. Gene Expr Patterns 3, 273–277. https://doi.org/10.1016/S1567-133X(03)00051-6

Naville, M., Henriet, S., Warren, I., Sumic, S., Reeve, M., Volff, J.N., Chourrout, D., 2019. Massive Changes of Genome Size Driven by Expansions of Non-autonomous Transposable Elements. Curr. Biol. 29, 1161-1168.e6. https://doi.org/10.1016/j.cub.2019.01.080

Nishida, H., 2008. Development of the appendicularian Oikopleura dioica: culture, genome, and cell lineages.

Dev Growth Differ 50 Suppl 1, S239-56.

Nishida, H., Stach, T., 2014. Cell Lineages and Fate Maps in Tunicates: Conservation and Modification. Zoolog. Sci. 31, 645–652. https://doi.org/10.2108/zs140117

Nishino, A., Satou, Y., Morisawa, M., Satoh, N., 2000. Muscle actin genes and muscle cells in the appendicularian, Oikopleura longicauda: phylogenetic relationships among muscle tissues in the urochordates. J Exp Zool 288, 135–150.

Olson, M. V, 1999. When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet 64, 18–23.

Onuma, T.A., Isobe, M., Nishida, H., 2017. Internal and external morphology of adults of the appendicularian, Oikopleura dioica: an SEM study. Cell Tissue Res. 367, 213–227. https://doi.org/10.1007/s00441-016-2524-5

Racioppi, C., Wiechecki, K.A., Christiaen, L., 2019. Combinatorial chromatin dynamics foster accurate cardiopharyngeal fate choices. Elife 8, 1–33. https://doi.org/10.7554/eLife.49921

Ragkousi, K., Beh, J., Sweeney, S., Starobinska, E., Davidson, B., 2011. A single GATA factor plays discrete, lineage specific roles in ascidian heart development. Dev Biol 352, 154–163. https://doi.org/S0012-1606(11)00023-6 [pii]10.1016/j.ydbio.2011.01.007

Razy-Krajka, F., Lam, K., Wang, W., Stolfi, A., Joly, M., Bonneau, R., Christiaen, L., 2014. Collier/OLF/EBF-Dependent Transcriptional Dynamics Control Pharyngeal Muscle Specification from Primed Cardiopharyngeal Progenitors. Dev. Cell 29, 263–276. https://doi.org/10.1016/j.devcel.2014.04.001

Razy-Krajka, F., Stolfi, A., 2019. Regulation and evolution of muscle development in tunicates. Evodevo 10, 1–34. https://doi.org/10.1186/s13227-019-0125-6

Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J. i., Inoue, T., 1999. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. Development 126, 3437 LP – 3447.

Satoh, N., 2016. Chordate Origins and Evolution, Chordate Origins and Evolution. Academic Press, San Diego. https://doi.org/https://doi.org/10.1016/B978-0-12-802996-1.00002-X

Satou, Y., Imai, K.S., Satoh, N., 2004. The ascidian Mesp gene specifies heart precursor cells. Development 131, 2533–2541. https://doi.org/10.1242/dev.01145

Savelieva, A. V., Temereva, E.N., 2020. Ultrastructure of the body cavities in juveniles and adults of the appendicularian Oikopleura gracilis (Tunicata, Chordata) suggests how the heart may have evolved in tunicates . Invertebr. Biol. e12286. https://doi.org/10.1111/ivb.12286

Schindelin, J., Arganda-Carreras, I., Frise, E., 2012. Fiji: an open-source platform for biological image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Stach, T., 2007. Ontogeny of the appendicularian Oikopleura dioica (Tunicata, Chordata) reveals characters similar to ascidian larvae with sessile adults. Zoomorphology 126, 203–214. https://doi.org/10.1007/s00435-007-0041-5

Stach, T., Turbeville, J.M., 2002. Phylogeny of Tunicata inferred from molecular and morphological characters. Mol Phylogenet Evol 25, 408–428.

Stach, T., Winter, J., Bouquet, J.-M.M., Chourrout, D., Schnabel, R., 2008. Embryology of a planktonic tunicate reveals traces of sessility. Proc Natl Acad Sci U S A 105, 7229–7234.

Stolfi, A., Gainous, T.B., Young, J.J., Mori, A., Levine, M., Christiaen, L., Field, H., Stolfi, A., 2010. Early chordate origins of the vertebrate second heart field. Science (80-.). 565, 565–569. https://doi.org/10.1126/science.1190181

Swalla, B.J., Cameron, C.B., Corley, L.S., Garey, J.R., 2000. Urochordates are monophyletic within the deuterostomes. Syst. Biol. 49, 52–64. https://doi.org/10.1080/10635150050207384

Tirosh-Finkel, L., Zeisel, A., Brodt-Ivenshitz, M., Shamai, A., Yao, Z., Seger, R., Domany, E., Tzahor, E., 2010. BMP-mediated inhibition of FGF signaling promotes cardiomyocyte differentiation of anterior heart field progenitors. Development 137, 2989–3000. https://doi.org/10.1242/dev.051649

Tolkin, T., Christiaen, L., 2016. Rewiring of an ancestral Tbx1/10-Ebf-Mrf network for pharyngeal muscle specification in distinct embryonic lineages. Dev. 143, 3852–3862. https://doi.org/10.1242/dev.136267

Tolkin, T., Christiaen, L., 2012. Development and Evolution of the Ascidian Cardiogenic Mesoderm, in: Current Topics in Developmental Biology. Academic Press Inc., pp. 107–142. https://doi.org/10.1016/B978-0-12-387786-4.00011-7

Torres-Águila, N.P., Martí-Solans, J., Ferrández-Roldán, A., Almazán, A., Roncalli, V., D'Aniello, S., Romano, G., Palumbo, A., Albalat, R., Cañestro, C., 2018. Diatom bloom-derived biotoxins cause aberrant development and gene expression in the appendicularian chordate Oikopleura dioica. Commun. Biol. 1. https://doi.org/10.1038/s42003-018-0127-2

Wada, H., 1998. Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny. Mol. Biol. Evol. 15, 1189–1194.

Wall, D.P., Fraser, H.B., Hirsh, A.E., 2003. Detecting putative orthologs. Bioinformatics 19, 1710–1711.

Wang, W., Niu, X., Stuart, T., Jullian, E., Mauck, W.M., Kelly, R.G., Satija, R., Christiaen, L., 2019. A single-cell transcriptional roadmap for cardiopharyngeal fate diversification. Nat. Cell Biol. 21, 674–686. https://doi.org/10.1038/s41556-019-0336-z

Wang, W., Razy-Krajka, F., Siu, E., Ketcham, A., Christiaen, L., 2013. NK4 Antagonizes Tbx1/10 to Promote Cardiac versus Pharyngeal Muscle Fate in the Ascidian Second Heart Field. PLoS Biol. 11. https://doi.org/10.1371/journal.pbio.1001725

Xavier-Neto, J., Castro, R.A., Sampaio, A.C., Azambuja, A.P., Castillo, H.A., Cravo, R.M., Simões-Costa, M.S., 2007. Cardiovascular development: Towards biomedical applicability - Parallel avenues in the evolution of hearts and pumping organs. Cell. Mol. Life Sci. 64, 719–734. https://doi.org/10.1007/s00018-007-6524-1

CHAPTER II: SUPPLEMENTARY INFORMATION

Supplementary Table 1. Accession numbers of the sequences used in phylogenetic analysis.

Mesp/Math/Neurogenin

Gene name	Gene ID		Database
BbeMath6	XP_019619932.1		NCBI
BbeMesp	XP_019645972.1		NCBI
BbeNeurogenin	XP_019635448.1		NCBI
BflMath6	XP_002596084.1		NCBI
BflMesp	ABD57444.1		NCBI
BflNeurogenin	AAF81766.1		NCBI
BlaMath6	BL15626_cuf0		
BlaMesp	BL14804_evm1		
BlaNeurogenin	ACE79717.1		NCBI
BleMath	Boleac.CG.SB_v3.S636.g12639.01.p		Aniseed
BleMesp	S12	Manual annotation	Aniseed
BleNeurogenin	Boleac.CG.SB_v3.S623.g12503.01.p		Aniseed
BleNeurogenin	Boleac.CG.SB_v3.S887.g15003.01.p		Aniseed
BscMath	Boschl.CG.Botznik2013.chr5.g36966.01.t		Aniseed
BscMesp	ChrUn	Manual annotation	Aniseed
BscNeurogenin	Boschl.CG.Botznik2013.chr10.g56010.01.p		Aniseed
BspMath	SCLE01239960	Manual annotation	
BspNeurogenin	SCLE01432724	Manual annotation	
CroMath6	KH.C9.872_XP_002122506.1		Aniseed
CroMesp	KH.C3.100_NP_001071761.1		Aniseed
CroNeurogenin	KH.C6.129		Aniseed
CsaMath8	Cisavi.CG.ENS81.R132.671158-674092.07637.p		Aniseed
CsaMesp	BAD18072.1	Manual annotation	Aniseed
CsaNeurogenin	Cisavi.CG.ENS81.R532.88248-90192.05863.p		Aniseed
FboMath	SDII01115774	Manual annotation	
FboNeurogenin	SDII01071635	Manual annotation	
FboNeurogenin	SDII01061863	Manual annotation	
GgaMath8	NP_001171210.1		NCBI
GgaMesogenin	NP_990015.2		NCBI
GgaMesp	CAD12882.1		NCBI
GgaMesp1	XP_003641880.1		NCBI
GgaNeurogenin	NP_990214.1		NCBI
GgaNeurogenin	XP_003641519.2		NCBI
GgaNeurogenin	NP_990127.1		NCBI
HauMath	S239	Manual annotation	Aniseed
HauMesp	S7142	Manual annotation	Aniseed
HauNeurogenin	Haaura.CG.MTP2014.S859.g06918.01.p		Aniseed
HroMath6	CG.MTP2014.S5.g02224.01.p		Aniseed
HroMesp	Harore.CG.MTP2014.S97.g06901.01.p		Aniseed
HroNeurogenin	CG.MTP2014.S5.g08306.01.p		Aniseed
HsaMath8	NP_116216.2		NCBI
HsaMesogenin	NP_001099039.1		NCBI
HsaMesp1	NP_061140.1		NCBI
HsaMesp2	NP_001035047.1		NCBI
HsaNeurogenin	NP_076924.1		NCBI
HsaNeurogenin	NP_066279.2		NCBI
HsaNeurogenin	NP_006152.2		NCBI
LchMath8	XP_006003191.1		NCBI
LchMesogenin	XP_005997585.1		NCBI
LchNeurogenin	XP_006003536.1		NCBI

LchNeurogenin	XP_006004992.1		NCBI
LchNeurogenin	XP_006011419.1		NCBI
LocMath	XP_006629700.2		NCBI
LocMesogenin	XP_015213357.1		NCBI
LocMesogenin	XP_006629065.1		NCBI
LocMesogenin	XP_015198957.1		NCBI
LocMesp2	XP_006629064.1		NCBI
LocNeurogenin	XP_015202930.1		NCBI
LocNeurogenin	XP_015205245.1		NCBI
MerMath	SCLF01217498	Manual annotation	
MerNeurogenin	SCLF01145285	Manual annotation	
MocciMath	Moocci.CG.ELv1_2.S264788.g06238.01.p		Aniseed
MocciMesp	S145165	Manual annotation	Aniseed
MocciNeurogenin	Moocci.CG.ELv1_2.S647877.g29858.01.p		Aniseed
MoccuMath	Mooccu.CG.ELv1_2.S602351.g36667.01.p		Aniseed
MoccuNeurogenin	S438802	Manual annotation	Aniseed
MoculMath	Moocul.CG.ELv1_2.S30383.g01453.01.p		Aniseed
MoculMesp	Moocul.CG.ELv1_2.S92333.g07428.01.p		Aniseed
MoculNeurogenin	Moocul.CG.ELv1_2.S15977.g00531.01.p		Aniseed
MooccuMesp	S460719	Manual annotation	Aniseed
OalMath	SCLG01000009	Manual annotation	
OalNeurogenin	SCLG01000119.1	Manual annotation	
OdiMath6	GSOIDP00016418001		Oikobase
OdiNeurogenin	GSOIDP00008211001		Oikobase
OloMath	SCLD01059283	Manual annotation	
OloNeurogenin	SCLD01144392	Manual annotation	
OloNeurogenin	SCLD01090848	Manual annotation	
OvaMath	SCLH01000781	Manual annotation	
OvaNeurogenin	SCLH01000494.1	Manual annotation	
PfuMat	Phfumi.CG.MTP2014.S940.g03013.01.p		Aniseed
PfuMesp	S15	Manual annotation	Aniseed
PfuNeurogenin	Phfumi.CG.MTP2014.S12979.g09049.01.p		Aniseed
PmaMath	phmamm.CG.MTP2014.S497.g09451.02.p		Aniseed
PmaMesp	S234	Manual annotation	Aniseed
PmaNeurogenenin	Phmamm.CG.MTP2014.S229.g05747.01.p		Aniseed

Ets/Erg

Gene name	Gene ID		Database
BfErg	XP_002613111		NCBI
BflEts	XP_002610126.1		NCBI
BlaErg	BL14695_evm0		Aniseed
BlaEts	14693		Aniseed
BleErga	S23.g05401.01		Aniseed
BleErgb	S391.g08928.01		Aniseed
BleErgc	S74	manual annotation	Aniseed
BleEts1/2a	S422.g09525.01.p		Aniseed
BleEts1/2b	S498.g10696.01.t		Aniseed
BscErga	chr7.g26841.01.p		Aniseed
BscErgb	chrUn.g52263.01		Aniseed
BscErgc	chrUn.g09892.01.t		Aniseed
BscEts1/2a	ChrUn_p381405427	manual annotation	Aniseed
BscEts1/2b	chr5.g69093.01.p		Aniseed
BspEts1/2b2	SCLE01158590	manual annotation	
BspEts1/2b1	SCLE01307272.1	manual annotation	
CroErga	KH.C4.539.v1.A.ND1-1		Aniseed
CroErgb1	KH.C10.148.v1.A.SL6-1		Aniseed
CroErgb2	KH.C10.420.v1.A.SL2-1		Aniseed
CroErgc	KH.C2.290.v1.A.SL2-1		Aniseed
CroEts1/2a	KH.C10.113.v4.A.SL1-1		Aniseed
CroEts1/2b	KH.C11.10.v1.R.ND1-1		Aniseed
CsaErga	1436501-1439403.13220		Aniseed
CsaErgb1	172747-182062.06822		Aniseed
CsaErgb2	150990-155653.06815		Aniseed
CsaErgc	343670-348289.01847		Aniseed
CsaEts1/2a	R44.821279-826601.11000.p		Aniseed
CsaEts1/2b	R16.345322-348455.08453.p		Aniseed
FboEts1/2b	SDII01087108	manual annotation	
GgaErg	NP_989611.1		NCBI
GgaEts1	XP_015153454.1		NCBI
GgaEts2	NP_990643.1		NCBI
HauErga	S207.g03311.01.p		Aniseed
HauErgb	S806.g06729.01		Aniseed

HauErgc	S3595.g10346.01.p		Aniseed
HauEts1/2a	S287	manual annotation	Aniseed
HauEts1/2b	S120.g02312.01.p		Aniseed
HroErga	S37.g15279.01.p		Aniseed
HroErgb	S156.g00533.01		Aniseed
HroErgc	S414.g02267.01.p		Aniseed
HroEts1/2a	S361	manual annotation	Aniseed
HroEts1/2b	S7.g06548.01.p		Aniseed
HsaErg	NP_891548.1		NCBI
HsaEts1	NP_001137292.1		NCBI
HsaEts2	NP_005230.1		NCBI
LchErg	XP_006002586.1		NCBI
LchEts1	XP_005989874.1		NCBI
LchEts2	XP_006002591.1		NCBI
LocErg	XP_006627839.1		NCBI
LocEts1	XP_015192990.1		NCBI
LocEts2	XP_015197244.1		NCBI
Mer9Ets1/2b2	SCLF01704939	manual annotation	
MerEts1/2b1	SCLF01279254.1	manual annotation	
MocciErga	S511319.g19317.01		Aniseed
MocciErgb	S244705.g05536.01.t		Aniseed
MocciErgc	S79112.g01052.01		Aniseed
MocciEts1/2a	S632673.g27652.01.p		Aniseed
MoccuErga	S269753.g10395.01		Aniseed
MoccuErgb	S197076.g06857.01.p		Aniseed
MoccuErgc	S413423.g19312.01		Aniseed
MoccuEts1/2a	S711677	manual annotation	Aniseed
MocuErgc	S60328.g03798.01		Aniseed
MoculErga	S51598.g02982.01		Aniseed
MoculErgb	S97310.g08238.01.t		Aniseed
MoculEts1/2a	S99115.g08619.01.p		Aniseed
OalEts1/2b1	SCLG01000003	manual annotation	
OalEts1/2b2	SCLG01000710.1	manual annotation	
OdiErg	GSOIDP00014411001_S53	manual annotation	Oikobase
OdiEts1/2b2	GSOIDP00003683001		Oikobase
OdiEts1/2b1	GSOIDP00006792001_S212	manual annotation	Oikobase
OloEts1/2b1	SCL0D01110143	manual annotation	
OloEts1/2b2	SCLD01120505.1	manual annotation	
OvaEts1/2b1	SCLH01001059	manual annotation	
OvaEts1/2b2	SCLH01001181.1	manual annotation	
PfuErga	S2966_S20668	manual annotation	Aniseed
PfuErgb	S3340.g05634.01		Aniseed
PfuErgc	S1489.g03780.02.p		Aniseed
PfuEts1/2a	S9396	manual annotation	Aniseed
PfuEts1/2b	S357.g01840.01.p		Aniseed
PmaErga	S491.g09369.01.p		Aniseed
PmaErgb	S18.g00779.01		Aniseed
PmaErgc	S64.g02300.02.p		Aniseed
PmaEts1/2a	S1111.g13716.02.p		Aniseed
PmaEts1/2b	S484.g09279.01.p		Aniseed

Mek

Gene name	Gene ID		Database
BbeMEK12	XP_019619563.1		NCBI
BbeMEK36	XP_019644388.1		NCBI
BbeMEK4	XP_019642021.1		NCBI
BbeMEK5	XP_019624095.1		NCBI
BbeMEK7	XP_019643631.1		NCBI
Bfl_STKIN25	XP_002598737.1		NCBI
Bfl_STKIN3	XP_002589064.1		NCBI
BflMEK12	XP_002601633.1		NCBI
BflMEK36	XP_002595698.1		NCBI
BflMEK4	XP_002607255.1		NCBI
BflMEK5	XP_035666864.1		NCBI
BflMEK7	XP_002589630.1		NCBI
BlaMEK12	BL22940_cuf4		
BlaMEK36	BL19281_evm0		
BlaMEK4	BL05717_evm0		
BlaMEK5	BL03615_evm11		
BlaMEK7	BL23231_evm0		
BleMEK12	Boleac.CG.SB_v3.S442.g09834.01.p		Aniseed
BleMEK36	Boleac.CG.SB_v3.S25.g05985.01.p		Aniseed

BleMEK4	Boleac.CG.SB_v3.S88.g14953.01.p		Aniseed
BleMEK5	Boleac.CG.SB_v3.S34.g08001.01.p		Aniseed
BleMEK7	Boleac.CG.SB_v3.S100.g00402.01.p		Aniseed
BscMEK12	Boschl.CG.Botznik2013.chrUn.g28277.01.p		Aniseed
BscMEK36	chr2+chrUN	Manual annotation	Aniseed
BscMEK4	Boschl.CG.Botznik2013.chrUn.g03398.01.p+Boschl.CG.Botznik2013.	Manual annotation	Aniseed
BscMEK5	Boschl.CG.Botznik2013.chr12.g61303.01.p		Aniseed
BscMEK7	Boschl.CG.Botznik2013.chrUn.g27076.01.p		Aniseed
BspMEK36	SCLE01439045.1+SCLE01416150.1+SCLE01459439.1		
BspMEK7	SCLE01016762.1+SCLE01426643.1+SCLE01329307.1+SCLE0104694		
Cro_STKIN25	KH.C7.308.v1.A.nonSL2-1		Aniseed
Cro_STKIN3	KH.C5.341.v1.A.nonSL1-1		Aniseed
CroMEK12	KH.L147.22.v1.A.SL1-1		Aniseed
CroMEK36	KH.L172.1.v1.A.ND1-1		Aniseed
CroMEK4	KH.C8.343.v1.A.SL2-1		Aniseed
CroMEK5	KH.C11.84.v1.A.SL3-1		Aniseed
CroMEK7	KH.L22.44.v1.A.SL2-1		Aniseed
CsaMEK12	Cisavi.CG.ENS81.R0.1216963-1218983.14576.p		Aniseed
CsaMEK36	Cisavi.CG.ENS81.R905.3972-11331.06316.p		Aniseed
CsaMEK4	Cisavi.CG.ENS81.R134.411618-418460.08740.p		Aniseed
CsaMEK5	Cisavi.CG.ENS81.R16.1649016-1652382.14365.p		Aniseed
CsaMEK7	Cisavi.CG.ENS81.R58.2891242-2899577.17580.p		Aniseed
FboMEK36a	SDII01028860.1		
FboMEK36b	SDII01028860.1		
GgaMEK1	XP_015147582.1		NCBI
GgaMEK2	XP_015155404.1		NCBI
GgaMEK3	XP_025010848.1		NCBI
GgaMEK4	XP_015150723.1		NCBI
GgaMEK5	XP_015147584.1		NCBI
GgaMEK6	XP_003642396.1		NCBI
GgaMEK7	XP_025001730.1		NCBI
HauMEK12	Haaura.CG.MTP2014.S88.g01839.02.p		Aniseed
HauMEK36	Haaura.CG.MTP2014.S1027.g07428.01.p		Aniseed
HauMEK4	Haaura.CG.MTP2014.S11.g00363.01.p+Haaura.CG.MTP2014.S11.g0	Manual annotation	Aniseed
HauMEK5	Haaura.CG.MTP2014.S2757.g09855.01.p		Aniseed
HauMEK7	Haaura.CG.MTP2014.S228.g03495.01.p		Aniseed
HroMEK12	Harore.CG.MTP2014.S158.g12521.02.p		Aniseed
HroMEK36	Harore.CG.MTP2014.S42.g10988.01.p		Aniseed
HroMEK4	Harore.CG.MTP2014.S12.g12037.01.p+Harore.CG.MTP2014.S12.g1	Manual annotation	Aniseed
HroMEK5	Harore.CG.MTP2014.S150.g06112.01.p		Aniseed
HroMEK7	Harore.CG.MTP2014.S21.g00956.01.p		Aniseed
Hsa_STKIN25	NP_001258906.1		NCBI
Hsa_STKIN3	NP_006272.2		NCBI
HsaMEK1	NP_002746.1		NCBI
HsaMEK2	NP_109587.1		NCBI
HsaMEK3	NP_659731.1		NCBI
HsaMEK4	NP_003001.1		NCBI
HsaMEK5	NP_002748.1		NCBI
HsaMEK6	NP_002749.2		NCBI
HsaMEK7	NP_001284485.1		NCBI
LchMEK1	XP_005990160.1		NCBI
LchMEK2	XP_005999133.1		NCBI
LchMEK3	XP_006008481.1		NCBI
LchMEK4	XP_006011477.1		NCBI
LchMEK5	XP_006001847.1		NCBI
LchMEK6	XP_005992212.2		NCBI
LchMEK7	XP_014347656.1		NCBI
LocMEK1	XP_006628796.1		NCBI
LocMEK2	XP_006639937.1		NCBI
LocMEK3	XP_015215390.1		NCBI
LocMEK4	XP_006635184.2		NCBI
LocMEK5	XP_006628877.1		NCBI
LocMEK6	XP_006635234.1		NCBI
LocMEK7	XP_015204544.1		NCBI
MerMEK36	SCLF01578774.1+SCLF01664085.1+SCLF01097487.1+SCLF01627178		
MerMEK7	SCLF01730825.1+SCLF01141466.1+SCLF01164915.1		
MoocciMEK12	Moocci.CG.ELv1_2.S600390.g24657.01.p		Aniseed
MoocciMEK36	Moocci.CG.ELv1_2.S257470.g05970.01.p		Aniseed
MoocciMEK4	Moocci.CG.ELv1_2.S263366.g06185.01.pS301794		Aniseed
MoocciMEK5	Moocci.CG.ELv1_2.S171171.g03276.01.p		Aniseed
MoocciMEK7	Moocci.CG.ELv1_2.S646071.g29530.01.p		Aniseed
MooccuMEK12	Mooccu.CG.ELv1_2.S441521.g21440.01.p+Mooccu.CG.ELv1_2.S558	Manual annotation	Aniseed
MooccuMEK36	Mooccu.CG.ELv1_2.S540638.g31404.01.p		Aniseed
MooccuMEK4	Mooccu.CG.ELv1_2.S722400.g47409.01.p		Aniseed

MooccuMEK5	Mooccu.CG.ELv1_2.S721609.g47275.01.p		Aniseed
MooccuMEK7	Mooccu.CG.ELv1_2.S703145.g44956.01.p+S703145+S665598+S703	Manual annotation	Aniseed
MooculMEK12	Moocul.CG.ELv1_2.S91994.g07329.01.p		Aniseed
MooculMEK36	Moocul.CG.ELv1_2.S20893.g00805.01.p		Aniseed
MooculMEK4	Moocul.CG.ELv1_2.S31823.g01546.01.p		Aniseed
MooculMEK5	Moocul.CG.ELv1_2.S110643.g11702.01.p		Aniseed
MooculMEK7	Moocul.CG.ELv1_2.S80168.g05886.01.p		Aniseed
OalMEK36	SCLG01000032.1		
OalMEK7a	SCLG01000544.1		
Odi_STKIN25	GSOIDP00001006001		Oikobase
Odi_STKIN4	GSOIDP00013245001		Oikobase
OdiMEK36	GSOIDP00015053001		Oikobase
OdiMEK7	GSOIDP00010943001+Sc37_231336-231910	Manual annotation	Oikobase
OloMEK36a	SCLD01163008.1		
OloMEK36b	SCLD01098264.1		
OloMEK36c	SCLD01109975.1		
OloMEK7a	SCLD01093505.1		
OloMEK7b	SCLD01178439.1		
OvaMEK36	SCLH01000769.1		
OvaMEK7	SCLH01022748.1+SCLH01001128.1+SCLH01045676.1		
PfuMEK12	Phfumi.CG.MTP2014.S42.g00590.01.p		Aniseed
PfuMEK36	Phfumi.CG.MTP2014.S223.g01451.01.p		Aniseed
PfuMEK4	Phfumi.CG.MTP2014.S1108.g03276.01.p		Aniseed
PfuMEK5	Phfumi.CG.MTP2014.S835.g02829.01.p		Aniseed
PfuMEK7	Phfumi.CG.MTP2014.S2409.g04821.01.p		Aniseed
PmaMEK12	Phmamm.CG.MTP2014.S3266.g17598.01.p		Aniseed
PmaMEK36	Phmamm.CG.MTP2014.S195.g05134.01.p		Aniseed
PmaMEK4	Phmamm.CG.MTP2014.S12.g00494.01.p		Aniseed
PmaMEK5	Phmamm.CG.MTP2014.S469.g09085.01.p		Aniseed
PmaMEK7	Phmamm.CG.MTP2014.S179.g04836.02.p		Aniseed

Gata

Gene name	Gene ID		Database
BbeGata123	XP_019627141.1		NCBI
BbeGata456	XP_019618792.1		NCBI
BflGTF123	ACR66214.1		NCBI
BflGTF456a	ACR66216.1		NCBI
BlaGata123	AFJ79490.1		NCBI
BlaGata456	AFJ79491.1		NCBI
BleGata123	S314	manual annotation	Aniseed
BleGata456	Boleac.CG.SB_v3.S147.g02878.01.p		Aniseed
BscGata123	ChrUn_Chr10	manual annotation	Aniseed
BscGata123	Chr10	manual annotation	Aniseed
BscGata456	ChrUn	manual annotation	Aniseed
BspGata	SCLE01420081	manual annotation	
BspGata	SCLE01408957	manual annotation	
BspGata	SCLE01414813	manual annotation	
BspGata	SCLE01211222	manual annotation	
CroGata123	KH.S696.1.v1.R.ND1_1		Aniseed
CroGata456	KH.L20.1.v1.A.SL2-1		Aniseed
CsaGata123	Cisavi.CG.ENS81.R41.2121992-2129284.14153.p		Aniseed
CsaGata456	Cisavi.CG.ENS81.R77.2996926-3008818.14416.p		Aniseed
FboGata	SDII01097846	manual annotation	
FboGata	SDII01078871	manual annotation	
GgaGata1	NP_990795.1		NCBI
GgaGata2	NP_001003797.1		NCBI
GgaGata3	NP_001008444.1		NCBI
GgaGata4	NP_001280035.1		NCBI
GgaGata5	NP_990752.2		NCBI
GgaGata6	NP_990751.1		NCBI
HauGata123	Haaura.CG.MTP2014.S24.g00692.01.p_S24		Aniseed
HauGata456	Haaura.CG.MTP2014.S112.g02206.01.p		Aniseed
HroGata123	Harore.CG.MTP2014.S12.g11731.01.t_S12		Aniseed
HroGata456	Harore.CG.MTP2014.S45.g04355.01.p		Aniseed
HsaGata1	NP_002040.1		NCBI
HsaGata2	NP_001139133.1		NCBI
HsaGata3	NP_001002295.1		NCBI
HsaGata4	NP_001295022.1		NCBI
HsaGata5	NP_536721.1		NCBI
HsaGata6	NP_005248.2		NCBI
LchGata1	XP_006008598.1		NCBI
LchGata2	XP_014341892.1		NCBI

LchGata3	XP_005993937.1		NCBI
LchGata4	XP_014351854.1		NCBI
LchGata5	XP_014344220.1		NCBI
LchGata6	XP_005995770.1		NCBI
LocGata1	XP_006625356.2		NCBI
LocGata2	XP_006631129.1		NCBI
LocGata3	XP_006633355.1		NCBI
LocGata4	XP_006642996.1		NCBI
LocGata5	XP_015220493.1		NCBI
LocGata6	XP_006633923.1		NCBI
MerGata	SCLF01270895	manual annotation	
MerGata	SCLF01415120	manual annotation	
MerGata	SCLF01211403_1	manual annotation	
MerGata	SCLF01530214	manual annotation	
MerGata	SCLF01239871	manual annotation	
MocciGata123	S487156	manual annotation	
MocciGata456	S458966	manual annotation	
MoccuGata123	S723572	manual annotation	
MoccuGata456	S557110_S209818	manual annotation	
MoculGata123	S97234	manual annotation	
MoculGata456	Moocul.CG.ELv1_2.S44512.g02355.01.p		Aniseed
OalGata	SCLG01000167	manual annotation	
OalGata	SCLG010000323	manual annotation	
OdiGBP2	GSOIDP00008593001		Oikobase
OdiGBP3a	GSOIDP00000004001		Oikobase
OdiGBP3b	GSOIDP00011092001		Oikobase
OdiGTFb	GSOIDP00007921001		Oikobase
OloGata	SCLD01106159	manual annotation	
OloGata	SCLD01172774	manual annotation	
OloGata	SCLD01119148	manual annotation	
OloGata	SCLD01126560	manual annotation	
OloGata	SCLD01018184	manual annotation	
OvaGata	SCLH01000841	manual annotation	
OvaGata	SCLH01000110	manual annotation	
OvaGata	SCLH01000261	manual annotation	
OvaGata	SCLH01002033_F	manual annotation	
OvaGata	SCLH01002033_R	manual annotation	
PfuGata123	S244_S11342	manual annotation	Aniseed
PfuGata456	S6271_0_S2222	manual annotation	Aniseed
PmaGata123	S535	manual annotation	Aniseed
PmaGata456	Phmamm.CG.MTP2014.S835.g12260.01.p		Aniseed

FoxF

Gene name	Gene ID		Database
BbeFoxF			
BbeFoxQ	XP_019628968.1		NCBI
BflFoxF	CAH69695.1		NCBI
BflFoxQ	XP_035680367.1		NCBI
BlaFoxF	BL02311_evm2		
BlaFoxQ	BL00713_evm0		
BleFoxF	S21	manual annotation	Aniseed
BscFoxF	Chr9_ChrUn	manual annotation	Aniseed
BspFoxF	SCLE01409615		
CrFoxQ	KH.L44.19.v1.A.nonSL4-1		Aniseed
CroFoxF	KH.C3.170.v1.A.SL1-1		Aniseed
CsaFoxF	Cisavi.CG.ENS81.R11.1287035-1294060.15471.p		Aniseed
CsaFoxQ	Cisavi.CG.ENS81.R1.1226140-1229479.11615.p		Aniseed
FboFoxF	SDII01127328	manual annotation	
GgaFoxF1	XP_414186.5		NCBI
GgaFoxF2	XP_015137672.2		NCBI
GgaFoxQ	XP_015137671.2		NCBI
HauFoxF	Haaura.CG.MTP2014.S882.g06990.01.p		Aniseed
HroFoxF	Harore.CG.MTP2014.S48.g08804.01.p		Aniseed
HsaFoxF1	NP_001442.2		NCBI
HsaFoxF2	NP_001443.1		NCBI
HsFoxQ1	NP_150285.3		NCBI
LchFoxF1	XP_005995947.1		NCBI
LchFoxF2	XP_006001190.1		NCBI
LchFoxQ	XP_006013316.1		NCBI
LocFoxF1	XP_006641203.1		NCBI
LocFoxF2	XP_006634525.1		NCBI
LocFoxQ	XP_006634621.1		NCBI

MerFoxF	SCLF01086355	manual annotation	
MocciFox	S456097	manual annotation	Aniseed
MoccuFoxF	S685721	manual annotation	Aniseed
MoculFoxF	Moocul.CG.ELv1_2.S114506.g13471.01.p		Aniseed
OalFoxF	SCLG01000319	manual annotation	
OdFoxQ	GSOIDP00009755001		Oikobase
OdiFoxF	GSOIDP00007763001		Oikobase
OloFox	SCLD01163618	manual annotation	
OvaFoxF	SCLH01001792	manual annotation	
PfuFoxF	Phfumi.CG.MTP2014.S1012.g03122.01.p		Aniseed
PfuFoxQ	Phfumi.CG.MTP2014.S3505.g05754.01.p		
PmaFoxF	S371	manual annotation	Aniseed
PmaFoxQ	Phmamm.CG.MTP2014.S654.g10956.01.p		

Nk

Gene name	Gene ID		Database
BbeNkx2.1	XP_019648010.1		NCBI
BbeNkx2.2	XP_019647989.1		NCBI
BbeNkx4	XP_019646148.1		NCBI
BflNkx2.1	XP_002589194.1		NCBI
BflNkx2.2	XP_002589199.1		NCBI
BflNkx4	XP_002610435.1		NCBI
BlaNkx4	AWV91585.1		NCBI
Ble_nk4	S7		Aniseed
BleNk2	S959		Aniseed
Bsc_nk4	chr9		Aniseed
BscNk2	ChrUn		Aniseed
Bst_Nk4	5'long	manual annotation	
CroNK2a	KH.C10.338.v1.A.SL4		Aniseed
CroNk4	KH.C8.482.v1.A.SL2-1		Aniseed
CsaNK2	ENS81.R37.1525232		Aniseed
CsaNk4	Cisavi.CG.ENS81.R24.410826-416079.10470.p		Aniseed
FboNk2	SDII1122225	manual annotation	
FboNk2	SDII01045043	manual annotation	
GgaNkx2.1	NP_989947.1		NCBI
GgaNkx2.2	NP_001264647.1		NCBI
GgaNkx2.3	NP_990362.2		NCBI
GgaNkx2.4	XP_015138916.1		NCBI
GgaNkx2.5	NP_990495.1		NCBI
GgaNkx2.6	NP_990468.1		NCBI
GgaNkx2.8	XP_003641408.3		NCBI
Hau	S249RC_S5459	manual annotation	Aniseed
HauNk2	Haaura.CG.MTP2014.S265.g03837.01.p		Aniseed
Hro_Nk4	S41_333208-368464	manual annotation	Aniseed
HroNk2	Harore.CG.MTP2014.S132.g01719.01.p		Aniseed
HsaNk2.4	NP_149416.1		NCBI
HsaNk2.6	NP_001129743.2		NCBI
HsaNKx2.2	NP_002500.1		NCBI
HsaNkx2.3	NP_660328.2		NCBI
HsaNkx2.5	NP_004378.1		NCBI
LchNkx2.1	XP_005989231.1		NCBI
LchNkx2.2	XP_006013284.1		NCBI
LchNkx2.3	XP_006001881.1		NCBI
LchNkx2.4	XP_006008054.1		NCBI
LchNkx2.5	XP_005992571.1		NCBI
LchNkx2.6	XP_005991959.1		NCBI
LchNkx2.8	XP_005989262.1		NCBI
LocNkx2.1	XP_006632731.1		NCBI
LocNkx2.2	XP_006626197.1		NCBI
LocNkx2.3	XP_006630496.1		NCBI
LocNkx2.4	XP_015208466.1		NCBI
LocNkx2.5	XP_006631743.1		NCBI
LocNkx2.6	XP_006625550.1		NCBI
LocNkx2.8	XP_006632762.2		NCBI
Mer_Nk4	5'long	manual annotation	Aniseed
MocciNk2	S77334	manual annotation	Aniseed
MocciNk4	S332293	manual annotation	Aniseed
MoccuNk2	S690034	manual annotation	Aniseed
MoccuNk4	Mooccu.CG.ELv1_2.S657543.g40314.01.p		Aniseed
MoculNk2	S113495	manual annotation	Aniseed
MoculNk4	Moocul.CG.ELv1_2.S112088.g12247.01.p_modify		Aniseed
Oal_Nk4	5'long	manual annotation	

OalNk2	SCLG01000088	manual annotation	
Odi_Nk4	BCN_5long	manual annotation	
OdiNK2a	GSOIDP00010368001		Oikobase
OdiNK2b	GSOIDP00011992001		Oikobase
Olo_Nk4	5'long	manual annotation	
Ova_Nk4	5'long	manual annotation	
OvaNk2b	SCLH01000727	manual annotation	
PfuNk2	Phfumi.CG.MTP2014.S953.g03028.02.p		Aniseed
PfuNk4	S1495	manual annotation	Aniseed
PmaNk2	Phmamm.CG.MTP2014.S597.g10419.03.p		Aniseed
PmaNk4	185	manual annotation	Aniseed

Hand

Gene name	Gene ID		Database
BbeHand	XP_019625359.1		NCBI
BbeNeurogenin	XP_019635448.1		NCBI
BflHand	AEL13770.1		NCBI
BflNeurogenin	AAF81766.1		NCBI
BlaHand	AFJ79493.1		NCBI
BlaNeurogenin	ACE79717.1		NCBI
BleHand	Boleac.CG.SB_v3.S154.g03139.01.p		Aniseed
BleHandr	S96	Manual annotation	Aniseed
BleNeurogenin	Boleac.CG.SB_v3.S623.g12503.01.p		Aniseed
BleNeurogenin	Boleac.CG.SB_v3.S887.g15003.01.p		Aniseed
BscHand	ChrUn	Manual annotation	Aniseed
BscHandr	Boschl.CG.Botznik2013.chr12.g10820.01.p		Aniseed
BscNeurogenin	Boschl.CG.Botznik2013.chr10.g56010.01.p		Aniseed
BspHand	SCLE01437072	Manual annotation	
BspNeurogenin	SCLE01432724	Manual annotation	
CroHand	KH.C14.604.v1.A.SL2_1		Aniseed
CroHandr	KH.C1.1116.v1.A.SL1_1		Aniseed
CroNeurogenin	KH.C6.129		Aniseed
CsaHand	Cisavi.CG.ENS81.R49.2261498-2262994.16124.p		Aniseed
CsaHandr	Cisavi.CG.ENS81.R14.220962-222552.16137.p		Aniseed
CsaNeurogenin	Cisavi.CG.ENS81.R532.88248-90192.05863.p		Aniseed
FboHand	SDII01065849		
FboNeurogenin	SDII01071635	Manual annotation	
FboNeurogenin	SDII01061863	Manual annotation	
GgaHand1	NP_990296.1		NCBI
GgaHand2	NP_990297.2		NCBI
GgaNeurogenin	NP_990214.1		NCBI
GgaNeurogenin	XP_003641519.2		NCBI
GgaNeurogenin	NP_990127.1		NCBI
HauHand	Haaura.CG.MTP2014.S167.g02904.01.p		Aniseed
HauHandr	Haaura.CG.MTP2014.S72.g01613.01.p		Aniseed
HauNeurogenin	Haaura.CG.MTP2014.S859.g06918.01.p		Aniseed
HroHand	Harore.CG.MTP2014.S40.g09177.01.p		Aniseed
HroHandr	Harore.CG.MTP2014.S74.g11434.01.p		Aniseed
HroNeurogenin	CG.MTP2014.S5.g08306.01.p		Aniseed
HsaHand1	NP_004812.1		NCBI
HsaHand2	NP_068808.1		NCBI
HsaNeurogenin	NP_076924.1		NCBI
HsaNeurogenin	NP_066279.2		NCBI
HsaNeurogenin	NP_006152.2		NCBI
LchHand1	XP_005988374.1		NCBI
LchHand2	XP_005993037.1		NCBI
LchNeurogenin	XP_006003536.1		NCBI
LchNeurogenin	XP_006004992.1		NCBI
LchNeurogenin	XP_006011419.1		NCBI
LocHand1	XP_006632012.1		NCBI
LocHand2	XP_006630021.1		NCBI
LocNeurogenin	XP_015202930.1		NCBI
LocNeurogenin	XP_015205245.1		NCBI
MerHand	SCLF01004291_SCLF01727318	Manual annotation	
MerNeurogenin	SCLF01145285	Manual annotation	
MocciHand	S492667	Manual annotation	Aniseed
MocciHandr	Moocci.CG.ELv1_2.S624793.g26726.01.p		Aniseed
MocciNeurogenin	Moocci.CG.ELv1_2.S647877.g29858.01.p		Aniseed
MoccuHand	Mooccu.CG.ELv1_2.S546820.g32354.01.p		Aniseed
MoccuHandr	Mooccu.CG.ELv1_2.S674203.g41775.01.p		Aniseed
MoccuNeurogenin	S438802	Manual annotation	Aniseed
MoculHand	S112757	Manual annotation	Aniseed

MoculHandr	Moocul.CG.ELv1_2.S82437.g06150.01.p		Aniseed
MoculNeurogenin	Moocul.CG.ELv1_2.S15977.g00531.01.p		Aniseed
OalHand	SCLG01000563	Manual annotation	
OalNeurogenin	SCLG01000119.1	Manual annotation	
OdiHand	GSOIDP00005552001		Oikobase
OdiNeurogenin	GSOIDP00008211001		Oikobase
OloHand	SCLD01144104		
OloHand	SCLD01012489_SCLD01051243	Manual annotation	
OloHand	SCLD01062667	Manual annotation	
OloHand	SCLD01091277	Manual annotation	
OloNeurogenin	SCLD01144392	Manual annotation	
OloNeurogenin	SCLD01090848	Manual annotation	
OvaHand	SCLH01000568	Manual annotation	
OvaNeurogenin	SCLH01000494.1	Manual annotation	
PfuHand	Phfumi.CG.MTP2014.S2952.g05296.01.p		Aniseed
PfuHandr	S5260	Manual annotation	Aniseed
PfuNeurogenin	Phfumi.CG.MTP2014.S12979.g09049.01.p		Aniseed
PmaHand	Phmamm.CG.MTP2014.S2382.g16605.01.p		Aniseed
PmaHandr	Phmamm.CG.MTP2014.S4.g00171.02.p		Aniseed
PmaNeurogenenin	Phmamm.CG.MTP2014.S229.g05747.01.p		Aniseed

Tbx

Gene name	Gene ID	Database
BflBrachyury1	XP_002590391.1	NCBI
BflBrachyury2	XP_002590390.1	NCBI
BflTbr1	AAG34893.2	NCBI
BflTbx1/10	AAG34887.2	NCBI
BflTbx15	XP_002599057.1	NCBI
BflTbx2_3	XP_002598922.1	NCBI
BflTbx20	XP_002612350.1	NCBI
BflTbx4_5	ABU50779.1	NCBI
BflTbx6_16	XP_002612013.1	NCBI
BlaBrachyury	BL23557_evm0	Aniseed
BlaTbr1	BL08925	Aniseed
BlaTbx1/10	BL01988_evm1	Aniseed
BlaTbx15	BL95766_evm0	Aniseed
BlaTbx2_3	BL12502_BL02331_BL1275	Aniseed
BlaTbx20	BL16028_evm0	Aniseed
BlaTbx4_5	BL06379_evm0	Aniseed
BlaTbx6_16	BL04677_cuf1	Aniseed
BleTbx1/10	S60.g12196.01	Aniseed
BleTbx15/18/22	S749.g13814.01	Aniseed
BleTbx19/Brachyury	S38.g08688.01	Aniseed
BleTbx2/3	S553.g11558.01	Aniseed
BleTbx20	S562.g11653.01	Aniseed
BleTbx6	S256.g06109.01	Aniseed
BscTbx1/10	chrUn_g48325.01	Aniseed
BscTbx19/Brachyury	chr7.g63408.01	Aniseed
BscTbx2/3	chr7.g53212.01	Aniseed
BscTbx6	chr1.g48703.01	Aniseed
BstBrachyury/Tbx19	SCLE01228003.1	Aniseed
BstTbx15/18/22	SCLE01017494.1	Aniseed
BstTbx2/3	SCLE01446362.1	Aniseed
BstTbx2/3	SCLE01429136.1	Aniseed
BstTbx6/16/24/VegT	SCLE01434700.1	Aniseed
CroEomes/Tbr1	KH.C3.773	Aniseed
CroTbx1/10	KH.C7.628	Aniseed
CroTbx15/18/22	KH.S589.4	Aniseed
CroTbx19/Brachyury	KH.S1404.1	Aniseed
CroTbx2/3	KH.L96.87	Aniseed
CroTbx20	KH.C1.224	Aniseed
CroTbx6.1	KH.S654.1	Aniseed
CroTbx6.2	KH.S654.2	Aniseed
CroTbx6.3	KH.S654.3	Aniseed
CsaEomes/Tbr1	R107.324117-331918.03246	Aniseed
CsaTbx1/10	R125.93093-108455.01412	Aniseed
CsaTbx15/18/22	R118.128673-1340382	Aniseed
CsaTbx19/Brachyury	R388.258080-264338.03855	Aniseed
CsaTbx2/3	R90.318198-331613.16578	Aniseed
CsaTbx20	R125.93093-108455.01412	Aniseed
CsaTbx6.1	R16.5012720-5014970.20043	Aniseed
CsaTbx6.1	R16.5041500-5055627.20057	Aniseed

FboBrachyury/Tbx19	SDII01062704.1	Manual annotation
FboTbx15/18/22	SDII01116622.1	Manual annotation
FboTbx2/3	SDII01124641.1	Manual annotation
FboTbx20	SCII01131462.1	Manual annotation
GgaBrachyury	XP_015139531.2	NCBI
GgaEomes	NP_001308484.1	NCBI
GgaTbr1	XP_015145201.1	NCBI
GgaTbx1	XP_025011481.1	NCBI
GgaTbx15	XP_416537.3	NCBI
GgaTbx18	NP_989784.2	NCBI
GgaTbx19	NP_990281.1	NCBI
GgaTbx2	XP_001235321.4	NCBI
GgaTbx20	NP_989475.1	NCBI
GgaTbx21	XP_015154966.2	NCBI
GgaTbx22	NP_989437.1	NCBI
GgaTbx3	NP_001257807.1	NCBI
GgaTbx4	NP_001025708.1	NCBI
GgaTbx5	XP_015150317.1	NCBI
GgaTbx6	XP_025011193.1	NCBI
HauEomes/Tbr1	S1059.g07498.01	Aniseed
HauTbx15/18/22	S4943.g10821.01	Aniseed
HauTbx19/Brachyury	S324.g04307.01	Aniseed
HauTbx2/3	S9.g00298.01	Aniseed
HauTbx6.1	S829.g06809.03	Aniseed
HauTbx6.2	S1817.g08930.01	Aniseed
HroEomes/Tbr1	S78.g14749.01	Aniseed
HroTbx15/18/22	S61.g03659.01	Aniseed
HroTbx19/Brachyury	S42.g07916.02	Aniseed
HroTbx2/3	S120.g12743.01	Aniseed
HroTbx6.1	S933.g10799.03	Aniseed
HroTbx6.2	S461.g13082.01	Aniseed
HsaBrachyury	NP_001257413.1	NCBI
HsaEomes	NP_001265111.1	NCBI
HsaTbr1	NP_006584.1	NCBI
HsaTbx1	NP_005983.1	NCBI
HsaTbx15	NP_689593.2	NCBI
HsaTbx18	NP_001073977.1	NCBI
HsaTbx19	NP_005140.1	NCBI
HsaTbx2	NP_005985.3	NCBI
HsaTbx20	NP_001159692.1	NCBI
HsaTbx21	NP_037483.1	NCBI
HsaTbx22	NP_058650.1	NCBI
HsaTbx6	NP_004599.2	NCBI
HsTbx10	NP_005986.2	NCBI
HsTbx3	NP_005987.3	NCBI
HsTbx4	NP_001308049.1	NCBI
HsTbx5	NP_000183.2	NCBI
LchBrachyury	XP_005993410.1	NCBI
LchEomes	XP_006011925.1	NCBI
LchTbr1	XP_006010469.1	NCBI
LchTbx1	XP_014340634.1	NCBI
LchTbx15	XP_014344889.1	NCBI
LchTbx18	XP_005995114.1	NCBI
LchTbx19	XP_014346004.1	NCBI
LchTbx2	XP_005986878.1	NCBI
LchTbx20	XP_006005638.1	NCBI
LchTbx21	XP_005990126.1	NCBI
LchTbx22	XP_014347604.1	NCBI
LchTbx3	XP_006007893.1	NCBI
LchTbx4	XP_005986877.1	NCBI
LchTbx5	XP_006007890.1	NCBI
LchTbx6	XP_014344551.1	NCBI
LocBrachyury	XP_015218241.1	NCBI
LocEomes	XP_006635610.1	NCBI
LocTbr1	XP_006636611.1	NCBI
LocTbx1	XP_006640182.1	NCBI
LocTbx15	XP_006639276.1	NCBI
LocTbx18	XP_006625898.1	NCBI
LocTbx19	XP_015216992.1	NCBI
LocTbx2	XP_006640960.1	NCBI
LocTbx20	XP_006635615.1	NCBI
LocTbx21	XP_015217756.1	NCBI
LocTbx22	XP_006633032.1	NCBI
LocTbx3	XP_006640338.1	NCBI

LocTbx4	XP_015222868.1		NCBI
LocTbx5	XP_006640339.1		NCBI
LocTbx6	XP_006641063.1		NCBI
MerBrachyury/Tbx19	SCLF01152421.1	Manual annotation	
MerMod_Tbx2/3	SCLF01056641.1	Manual annotation	
MerTbx15/18/22	SCLF01251203.1	Manual annotation	
MerTbx2/3	SCLF01026372.1	Manual annotation	
MerTbx2/3	SCLF01048308.1	Manual annotation	
MerTbx6/16/24/VegT	SCLF01036633.1	Manual annotation	
MocciEomes/Tbr1	S477916.g16708.01		Aniseed
MocciTbx15/18/22	S530819.g20232.01		Aniseed
MocciTbx19/Brachyury	S634979.g27904.01		Aniseed
MocciTbx2/3	S479065.g16818.01		Aniseed
MocciTbx6	S497697.g18857.01		Aniseed
MoccuEomes/Tbr1	S716823.g46618.01		Aniseed
MoccuTbx15/18/22	S640811.g39040.01		Aniseed
MoccuTbx19/Brachyury	S705330.g45232.01		Aniseed
MoculEomes/Tbr1	33300.g01638.01		Aniseed
MoculTbx15/18/22	S109292.g11175.01		Aniseed
MoculTbx19/Brachyury	S90886.g07224.01		Aniseed
MoculTbx2/3	S99115.g08615.01		Aniseed
MoculTbx6	S69504.g04594.01		Aniseed
OalBrachyury/Tbx19	SCLG01000117.1	Manual annotation	
OalTbx15/18/22	SCLG01000201.1	Manual annotation	
OalTbx2/3	SCLG01022965.1	Manual annotation	
OalTbx2/3	SCLG01000157.1	Manual annotation	
OalTbx2/3	SCLG01000314.1	Manual annotation	
OalTbx2/3	SCLG01000339.1	Manual annotation	
OalTbx6/16/24/VegT	SCLG01000014.1	Manual annotation	
OdiBrachyury/Tbx19	GSOIDP00000279001		Oikobase
OdiTbx15/18/22	GSOIDP00014261001		Oikobase
OdiTbx15/18/22	GSOIDP00014264001		Oikobase
OdiTbx2/3	GSOIDP00004289001		Oikobase
OdiTbx2/3	GSOIDP00011582001		Oikobase
OdiTbx2/3	GSOIDP00004999001		Oikobase
OdiTbx2/3	GSOIDP00007455001		Oikobase
OdiTbx6/16/24/VegT	GSOIDP00006159001		Oikobase
OloBrachyury/Tbx19	SCLD01093191.1	Manual annotation	
OloTbx15/18/22	SCLD01119088.1	Manual annotation	
OloTbx15/18/22	SCLD01116716.1	Manual annotation	
OloTbx2/3	SCLD01098799.1	Manual annotation	
OloTbx2/3	SCLD01094440.1	Manual annotation	
OloTbx2/3	SCLD01116750.1	Manual annotation	
OloTbx2/3	SCLD01177323.1	Manual annotation	
OloTbx6/16/24/VegT	SCLD01038142.1	Manual annotation	
OvaBrachyury/Tbx19	SCLH01001620.1	Manual annotation	
OvaTbx15/18/22	SCLH01000284.1	Manual annotation	
OvaTbx2/3	SCLH01000456.1	Manual annotation	
OvaTbx2/3	SCLH01000169.1	Manual annotation	
OvaTbx2/3	SCLH01000635.1	Manual annotation	
OvaTbx2/3	SCLH01000589.1	Manual annotation	
OvaTbx6/16/24/VegT	SCLH01000086.1	Manual annotation	
PfuEomes/Tbr1	S70.g00816.01		Aniseed
PfuTbx1/10	S5861.g07133.01		Aniseed
PfuTbx19/Brachyury	S1283.g03521.01		Aniseed
PfuTbx20	S21906.g09793.01		Aniseed
PfuTbx6.1	S4584.g06449.01		Aniseed
PfuTbx6.2	S4556.g06436.01		Aniseed
PmaEomes/Tbr1	S631.g10737.02		Aniseed
PmaTbx1/10	S356.g07736.01		Aniseed
PmaTbx15/18/22	S497.g09447.01		Aniseed
PmaTbx19/Brachyury	S305.g07005.01		Aniseed
PmaTbx2/3	S17.g00705.01		Aniseed
PmaTbx20	S175.g04755.01		Aniseed
PmaTbx6.1	S1645.g15326.01		Aniseed
PmaTbx6.2	S855.g12396.01		Aniseed

Supplementary Table 2. Information about the cloning fragments and riboprobes and primers used to amplify them.

Gene	Forward	Reverse	source; length; vector; RNA-pol/digestion enzyme
OdActnM1 cross-hybridizing	5'..GTCCCGCCATGTACGTCTG..3'	5'..GCATGGGAATCGCTCGTTACCA..3'	gDNA exon 2 partial; 389bp; PCR4-TOPO; T3/NotI
OdActnM1 specific	5'..GATCGTCCACCGAAAGTGC..3'	5'..GTCAGCAACTGTTTGAATATATTG..3'	cDNA 3'UTR; 351bp; PCR4-TOPO; T7/PstI
OdBrachyury	5'..GGTTCGCACTGGATGAAACAGCC..3'	5'..TATCCGTTCTGACACCAGTCGTTC..3'	gDNA exon 3; 630bp; PCR4-TOPO; T3/NotI
OdDach	5'..GAGATGGATCCTGCGCAGC..3'	5'..GTAAGTTTAAGAATATCCGAAAATCC..3'	cDNA full lenght; 770bp; PCR4-TOPO; T7/SpeI
OdEbf (COE)	5'..GAGATCATGTGTTCCCGATGTTG..3'	5'..GTTGAGTGAAAGAAAACCTTGCT..3'	cDNA exon 5 to exon 8; 578bp; PCR4-TOPO; T3/NotI
OdERK	5'..GAAGGAGCCTACGGCATAG..3'	5'..GCTAGAATACATCCGACAGAC..3'	cDNA exon 1 to exon 5; 563bp; PCR4-TOPO; T3/NotI
OdEts1/2b1	5'..GACGGCATTGATGGATTACAGCTAT..3'	5'..GTTACTCTTCAGTCTCTGGCTC..3'	cDNA exon 3-4 to exon 7; 480bp; PCR4-TOPO; T3/NotI
OdEts1/2b2	5'..GACTCTCTTTGCCATCAATCC..3'	5'..GCCTTTTTCGTCCAGCTAATG..3'	cDNA exon 2 to exon 5; 728bp; PCR4-TOPO;T3/NotI
OdFilamin C	5'..GCTACGAGCCACAAGAGAAG..3'	5'..GCGTATTCACCAGCCTCTGTT..3'	gDNA exon 7; 604bp; PCR4-TOPO; T7/PstI
OdFoxF	5'..GGCTGGAAGAATTCCGTCCG..3'	5'..GAGCTGATTCGCATGGGCAGG..3'	cDNA exon 3 to exon 7; 639bp; PCR4-TOPO; T7/SpeI
OdGata1/2/3b	5'..GCCTCTTCTGATTCGCCATTC..3'	5'..GGAATGACTGTTTGGTGTTGG..3'	cDNA exon 1 to exon 3; 791bp; PCR4-TOPO; T3/XbaI
OdGata1/2/3d	5'..GGGCAGAATATGAAAATGTATTTT..3'	5'..GCTGACCGTCCGCTAGTC..3'	cDNA full lenght; 1086bp; PCR4-TOPO; T7/BamHI
OdHand1/2	5'..GATGGAGTTGAATTGTATTCCGATC..3'	5'..GATTCTTTTCTAATCAGATGGGCA..3'	cDNA full lenght; 462bp; PCR4-TOPO; T7/SpeI
OdIslet	5'..GGTCTCCGGTGATGAATTCCT..3'	5'..GCTTTGTCGTAGGTTTAGGCCA..3'	cDNA exon 4 to exon 8; 766bp; PCR4-TOPO; T7/PstI
OdMath6	5'..GCCGAATTCCACACAACGAAG..3'	5'..CAACGTTAGCAGGTATAGAATAG..3'	cDNA exon 1-2 to 3'UTR; 602bp; PCR4-TOPO; T7/SpeI
OdMyosin	5'..GACTCTTCAAGTGGCTCGTCG..3'	5'..GGCCAAATCCAATCCGAAATCG..3'	gDNA exon 2; 254bp; PCR4-TOPO; T3/NotI
OdNk4	5'..GACCGAAAAATTACAACTATGAGC..3'	5'..GCTGTAGCGCCGAGCTCAC..3'	cDNA exon 1 to exon 3; 649bp; PCR4-TOPO; T3/NotI
OdRapostlin	5'..GAAGATATGAGCCACTTGCCTC..3'	5'..GTTGTGATTGAAGAAGTTAGTCCAG..3'	cDNA exon 6 to exon 8; 628bp; PCR4-TOPO; T3/NotI
OdTrnnT1	5'..GACGAAGAAGATCTTGATAGC..3'	5'..GGATTCGAAGTTGCGTACTG..3'	gDNA exon 2; 218bp; PCR4-TOPO; T7/PstI
OdTrnnT4	5'..GAAATCGGATCTAAAAAGG..3'	5'..GTCATTTTCACGCACGCGCA..3'	cDNA exon 1 to exon 5; 569bp; PCR4-TOPO; T3/NotI
OdTrnnT7	5'..GCACGTCATCCCTAGAATA..3'	5'..GAAGCTTGCCAACTCGTGT..3'	cDNA 5'UTR to exon 5; 670bp;PCR4-TOPO; T7/PstI
OdMyoD	5'..GTTACAAAATGACTATGACGGAAAC..3'	5'..GGCTTCCAAGTTTCTTGACCAG..3'	cDNA full lenght; 813bp; PCR4-TOPO; T7/PstI

Supplementary Table 3. Results from the BRBHs. To obtain the whole information about the sequences obtained visit:

https://www.dropbox.com/scl/fi/ddaxdsucit7fle40npv5c/SupFig3.xlsx?dl=0&rlkey=vcmridfsd2aowc839s2umammn

#	Ciona GeneID	UniqueName	Cell type Specific Marker	Oikopleura Ids from Ciona BRBH	Human Ids from Ciona BRBH	CIONA <-> OIKOPLEURA RBBH YES or NO	Human or Oikopleura Ids from its own RBBH	CIONA <-> HUMAN RBBH YES or NO	OIKOPLEURA <-> HUMAN RBBH YES or NO	CONCLUSION
1	KH2013:KH.C1.1083	EPHA10/2/3/4/5/6/7/8	FHP	GSOIDP0001175001/GSOIDP0001762001	NP_001291465.1/NP_001275558.1	YES	NP_004431.1/NP_001363394.1	YES	YES	Present in all
2	KH2013:KH.C1.1211	LAMA1/2/3	FHP	GSOIDP00000234001/GSOIDP00006179001	NP_001073291.2/NP_000417.3	NO	GSOIDP00000234001/GSOIDP00006179001	YES	NO	Lost in O. dioica
3	KH2013:KH.C1.1161	SCN10A; SCN2A; SCN9A	FHP	GSOIDP00013476001/GSOIDP00011229001	XP_016860147.1	NO	NA/XP_016862506.1	NO	NA	Ascidian innovation
4	KH2013:KH.C11.186	JAG1/2 (= NOTCH)	FHP	GSOIDP00008711001	NP_000205.1	YES	NP_000205.1	YES	YES	Present in all
5	KH2013:KH.C12.13	FRAS1; FREM2/3	FHP	GSOIDP00016087001	NP_079350.5	YES	NP_079350.5	YES	YES	Present in all
6	KH2013:KH.C12.49	COL21A1; COL24A1; COL4A2	FHP	GSOIDP00011829001	XP_016856418.1	NO	NA	NO	NA	Ascidian innovation
7	KH2013:KH.C12.244	KH2013:KH.C12.244	FHP	GSOIDP00002326001	NP_060152.2	NO	NA	NO	NA	Ascidian innovation
8	KH2013:KH.C2.170	ODZ1/2/3/4; TENM2/3/4	FHP	GSOIDP00004698001	NP_001073946.1	YES	NP_001073946.1	YES	YES	Present in all
9	KH2013:KH.C3.815	DSE;DSEL	FHP	GSOIDP00015632001	NP_115536.2	NO	GSOIDP00004937001	YES	NO	Lost in O. dioica
10	KH2013:KH.C3.693	MYLK2/3/4	FHP	GSOIDP00014257001	NP_001295230.1	YES	NP_001295230.1	YES	YES	Present in all
11	KH2013:KH.C3.21	HMCN1; HSPG2; LAMA2	FHP	GSOIDP00008906001	NP_001278789.1	YES	NP_001278789.1	YES	YES	Present in all
12	KH2013:KH.C4.231	BX005256.5	FHP	GSOIDP00004211001	NP_002338.3	NO	NA	NO	NA	Ascidian innovation
13	KH2013:KH.C5.236	MMP17; MMP21; MMP25	FHP	GSOIDP00009547001/GSOIDP00006642001	NP_671724.1	NO	GSOIDP00009547001/GSOIDP00006642001	YES	NO	Lost in O. dioica
14	KH2013:KH.C7.207	FOXP1/2/4	FHP	GSOIDP00008786001	XP_016865723.1	YES	XP_016865723.1	YES	YES	Present in all
15	KH2013:KH.C8.738	ERG5;CYP26A1; CYP26B1; CYP26C1	FHP	GSOIDP00002670001	NP_000774.2	NO	NA	NO	NA	Ascidian innovation
16	KH2013:KH.L154.10	SMTN; SMTNL1; SMTNL2	FHP	GSOIDP00000295001	XP_011528643.1	YES	XP_011528643.1	YES	YES	Present in all
17	KH2013:KH.L96.80	KH2013:KH.L96.80_SLC5A12/8	FHP	GSOIDP00002507001/GSOIDP00016116001	NP_666018.3	NO	NA/NP_666018.3	YES	NA	Ascidian innovation
18	KH2013:KH.S423.6	SMYD1/2/3	FHP	GSOIDP00013473001	NP_001317293.1	YES	NP_001317293.1	YES	YES	Present in all
1	KH2013:KH.C1.356	MATN1/3/4 COL12A1; MATN1; MATN3	SHP	GSOIDP00008632001/GSOIDP00011558001	NP_002017.2	NO	NA	NO	NA	Ascidian innovation
2	KH2013:KH.C1.638	KH2013:KH.C1.638_C17ORF105_CFAP97D1	SHP	GSOIDP00016085001	NP_001129955.1/ XP_016876399.1	NO	NA/GSOIDP00016085001	YES	NA	Ascidian innovation
3	KH2013:KH.C10.72	ZFP318; ZNF638	SHP	GSOIDP00000039001/GSOIDP00017863001	NP_055160.2	YES	NP_055160.2	YES	YES	Present in all
4	KH2013:KH.C10.172	KIRREL1/2/3	SHP	GSOIDP00009042001/GSOIDP00006155001	NP_060710.3/XP_011541334.1	YES	NP_060710.3/NP_115499.5	YES	YES	Present in all
5	KH2013:KH.C13.183	SP9; ZNF503; ZNF703	SHP	GSOIDP00003237001/GSOIDP00014884001	NP_116161.2/NP_079345.1	NO	GSOIDP00002297001/GSOIDP00003118001	YES	NO	Lost in O. dioica
6	KH2013:KH.C14.291	DACH1/2; SKIL	SHP	GSOIDP00011938001	NP_001353641.1/NP_001132986.1	YES	GSOIDP00011938001	YES	YES	Present in all
7	KH2013:KH.C5.45	GDH1/3	SHP	GSOIDP00008070001/GSOIDP00008544001	NP_036216.2	YES	NP_005262.1	NO	NO	Urochordate innovation or vertebrate loss
1	KH.C3.100	Mesp	TVC	GSOIDP00016418001	NP_001099039.1	NO	GSOIDP00008211001/GSOIDP00002088001	YES	NO	Lost in O. dioica
2	KH.C10.113	Ets1/2	TVC	GSOIDP00006792001	NP_005229.1	YES	NP_005229.1	YES	YES	Present in all
3	KH.C3.170	FoxF	TVC	GSOIDP00007763001	NP_001442.2	YES	NP_001442.2	YES	YES	Present in all
4	KH.C1.1116	Hand-r/Hand-like	TVC	GSOIDP00005552001	NP_068808.1/XP_005268588.1	NO	GSOIDP00005552001	YES	NO	Ascidian innovation
5	KH.L20.1	GATA4/5/6	TVC	GSOIDP00000004001	NP_001295023.1	YES	GSOIDP00000004001	YES	No	Lost in O. dioica
6	KH.C8.482	NK4	TVC	GSOIDP00003812001	NP_001129743.2	YES	NP_660328.2/NP_001129743.2	YES	YES	Present in all
7	KH.C7.205	ASB	TVC	GSOIDP00003280001	XP_011535137.1	YES	XP_016867247.1/XP_011535137.1	YES	YES	Present in all
8	KH.C11.698	LGR4/5/6	TVC	GSOIDP00005966001	NP_060960.2	NO	GSOIDP00005966001	YES	NO	Lost in O. dioica
9	KH.C11.608	Irx1/2/3/5	TVC/ASM	GSOIDP00017171001/GSOIDP00013453001	XP_005256194.1	yes	XP_005256194.1	yes	YES	Present in all
10	KH.C1.129	RhoD/F	TVC	GSOIDP00003459001	NP_004031.1	NO	GSOIDP00003459001	NO	YES	Ascidian innovation
11	KH.C9.371	DDR1/2	TVC	GSOIDP00003173001	NP_001014796.1	YES	NP_001014796.1	YES	YES	Present in all
12	KH.S215.4	Lhx3	pre-TVC	GSOIDP00012755001	XP_016870657.1	YES	NP_203129.1/XP_016870657.1	YES	YES	Present in all
13	KH.I8.11	Tbx6	pre-TVC	GSOIDP00004289001/GSOIDP00004999001/GSOIDP000061 59001	NP_004599.2	YES	NP_005985.3	YES	YES	Present in all
1	KH.C13.73	FNBP1/Rapostlin	ASM	GSOIDP00007393001	NP_001350684.1	YES	XP_005251891.1/NP_001350684.1	YES	YES	Present in all
2	KH.l166.1	RGS/Gdd	ASM	GSOIDP00012995001	NP_937870.1	YES	NP_937870.1	YES	YES	Present in all
3	KH.L152.2.v1	Islet	ASM	GSOIDP00010709001	NP_002193.2	YES	NP_665804.1/NP_002193.2	YES	YES	Present in all
4	KH.C14.307	MyoD/MYF	ASM	GSOIDP00004266001	NP_005584.2	YES	NP_002460.1/NP_005584.2	YES	YES	Present in all
1	KH.C12.671	FilaminC	Structural	GSOIDP00000803001	NP_001120959.1	YES	NP_001447.2/NP_001120959.1	YES	YES	Present in all
2	KH.C4.57	TroponinT	Structural	GSOIDP00006024001	NP_001001431.1	YES	NP_001354779.1/NP_001001431.1	YES	YES	Present in all
3	KH.C1.216	Myl3/Myosin light chain 3	Structural	GSOIDP00014945001	NP_524146.1	YES	NP_524147.2/NP_524146.1	YES	YES	Present in all

COUNT:

Lost in O. dioica — **3FHP+1SHP+3TVC**

Ascidian Innovation — **6FHP+2SHP+2TVC**

Supplementary Table 4. Developmental effects of LDN and SU5402 at different concentrations and periods of incubations.

Starting time incubation	Concentration	Hatch success	Correct early-hatchling
From 2-cell stage	DMSO 0,2%	83% (n=231)	60% (n=168)
	LDN 10 µM	31% (n=72)	1% (n=2)
	SU5402 50µM	4% (n=12)	0% (n=0)
From 32-cell stage	DMSO 0,2%	86% (n=391)	68% (n=312)
	LDN 10 µM	55% (n=216)	10% (n=34)
	SU5402 50 µM	91% (n=240)	60% (n=158)
	DMSO 0,3%	94% (n=262)	85% (n=236)
	SU5402 100 µM	66% (n=116)	25% (n=45)
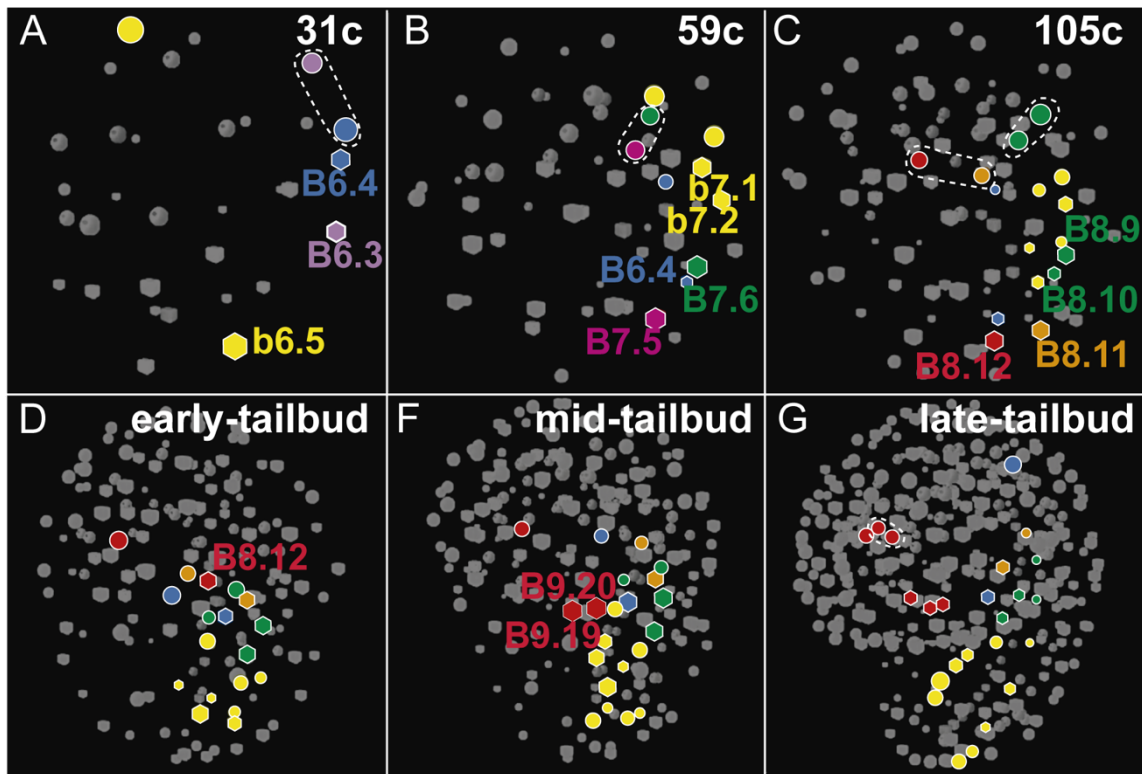
Sup. Fig. 1. **4D-reconstruction of a virtual cell tracing experiment based on nuclear position from the 30 cell stage to tailbud stages of *O. dioica* embryos** (Modified from Stach, 2008). The fate of the blastomeres are indicated in different colors: muscle and heart (purple, pink), posterior tail muscle cells (yellow), anterior tail muscle cells (orange, green), heart (red), germ-line (blue). Circles and hexagons represent blastomeres derived from right and left sides of the embryo, respectively. Dashed lines indicate twin blastomeres.

A

BflMesp_ABD57444.1
BlaMesp_BL14804_evm1_gene
BbeMesp_XP_019645972.1
HsaMesp1_NP_061140.1
HsaMesp2_NP_001035047.1
GgaMesp1_XP_003641880.1
GgaMesp_CAD12882.1
HsaMesogenin_1_NP_001099039.1
GgaMesogenin1_NP_990015.2
LchMesogenin1_XP_005997585.1
LocMesogenin1_XP_015213357.1
LocMesogenin1like_XP_006629065.1
LocMesp2_XP_006629064.1
LocMesogenin1like_XP_015198957.1
CroMesp_KH.C3.100_NP_001071761.1
CsaMesp_BAD18072.1
PfuMesp_S15
PmaMesp_S234
BleMesp_S12
BscMesp_ChrUn
MoocciMesp_S145165
MooculMesp_Moocul.CG.ELv1_2.S92333.g07428.01.p
MooccuMesp_S460719
HroMesp_Harore.CG.MTP2014.S97.g06901.01.p
HauMesp_S7142
BbeMath6_XP_019619932.1
BflMath6_XP_002596084.1
BlaMath6_BL15626_cuf0
HsaMath8_NP_116216.2
LocMath_XP_006629700.2
LchMath8_XP_006003191.1
GgaMath8_NP_001171210.1
FboMath_SDII01115774
HauMath_S239
HroMath6_CG.MTP2014.S5.g02224.01.p
CroMath6_KH.C9.872_XP_002122506.1
CsaMath8_Cisavi.CG.ENS81.R132.671158-674092.07637.p
PfuMath_Phfumi.CG.MTP2014.S940.g03013.01.p
PmaMath_phmamm.CG.MTP2014.S497.g09451.02.p
BscMath_Boschl.CG.Botznik2013.chr5.g36966.01.t
BleMath_Boleac.CG.SB_v3.S636.g12639.01.p
MoocciMath_Moocci.CG.ELv1_2.S264788.g06238.01.p
MooculMath_Moocul.CG.ELv1_2.S30383.g01453.01.p
MooccuMath_Mooccu.CG.ELv1_2.S602351.g36667.01.p
OdiMath6_GSOIDP00016418001
OloMath_SCLD01059283
MerMath_SCLF01217498
BspMath_SCLE01239960
OvaMath_SCLH01000781
OalMath_SCLG01000009
HsaNeurogenin2_NP_076924.1
GgaNeurogenin2_NP_990127.1
LchNeurogenin2_XP_006011419.1
HsaNeurogenin3_NP_066279.2
LchNeurogenin3like_XP_006004992.1
LocNeurogenin1like_XP_015202930.1
GgaNeurogenin3_XP_003641519.2
HsaNeurogenin1_NP_006152.2
GgaNeurogenin1_NP_990214.1
LchNeurogenin1_XP_006003536.1
LocNeurogenin1_XP_015205245.1
BflNeurogenin_AAF81766.1
BlaNeurogenin_ACE79717.1
BbeNeurogenin_XP_019635448.1
CroNeurogenin_KH.C6.129
CsaNeurogenin_Cisavi.CG.ENS81.R532.88248-90192.05863.p
PmaNeurogenin_Phmamm.CG.MTP2014.S229.g05747.01.p
PfuNeurogenin_Phfumi.CG.MTP2014.S12979.g09049.01.p
HauNeurogenin_Haaura.CG.MTP2014.S859.g06918.01.p
HroNeurogenin_CG.MTP2014.S5.g08306.01.p
MoocciNeurogenin_Moocci.CG.ELv1_2.S647877.g29858.01.p
MooccuNeurogenin_S438802
MooculNeurogenin_Moocul.CG.ELv1_2.S15977.g00531.01.p
BscNeurogenin_Boschl.CG.Botznik2013.chr10.g56010.01.p
BleNeurogenin_Boleac.CG.SB_v3.S623.g12503.01.p
BleNeurogenin_Boleac.CG.SB_v3.S887.g15003.01.p
OdiNeurogenin_GSOIDP00008211001
OvaNeurogenin_SCLH01000494.1
OloNeurogenin_SCLD01144392
BspNeurogenin_SCLE01432724
MerNeurogenin_SCLF01145285
OalNeurogenin_SCLG01000119.1
OloNeurogenin_SCLD01090848
FboNeurogenin_SDII01071635
FboNeurogenin_SDII01061863

2.0

90

Math

Sup. Fig. 2. (**A**) ML phylogenetic analysis of Mesp, neurogenin and math revealing the loss of *Mesp* in Appendicularians. Scale bar indicates amino acid substitutions. Bootstrap values are shown in the nodes. Vertebrates: *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Tunicates: *Bathochordaeus sp.* (Bsp). *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Fritillaria borealis* (Fbo), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Mesochordaeus erythrocephalus* (Mer), *Molgula occidentalis* (Moocci), *Molgula occulta* (Mooccu), *Molgula oculata* (Moocul), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Cephalochordates: *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bfl). Developmental expression pattern of *O. dioica Math6* homolog (**B-G**). Whole mount in situ hybridization in different stages of *O. dioica* development showing expression in the notochord in tailbud and early-hatchling embryos (red arrowheads) (**C, D**), in epidermis (blue arrowheads) (**C-F**), in the annal domain in hatchling stages (yellow arrowheads) (**D-F**), in later stages of neural system development (pink arrowheads) (**E-F**), and in later stages of digestive system development (green arrowheads) (**E-F**). Images from tailbud in advance correspond to left lateral views orientated anterior towards the left and dorsal towards the top.
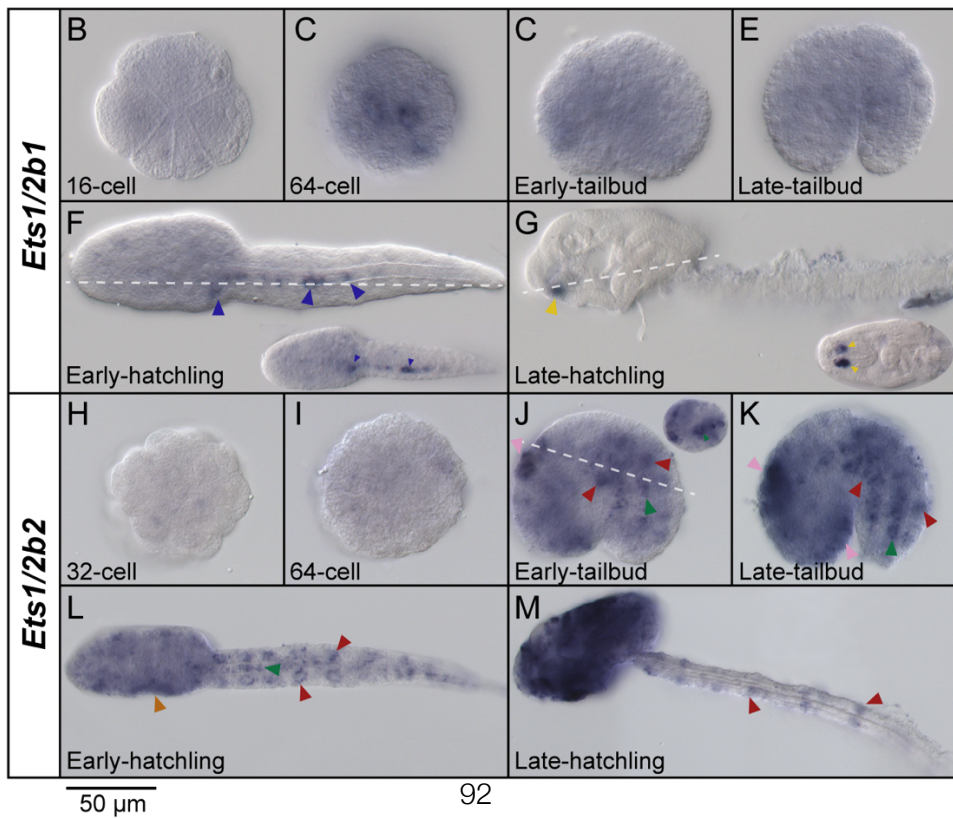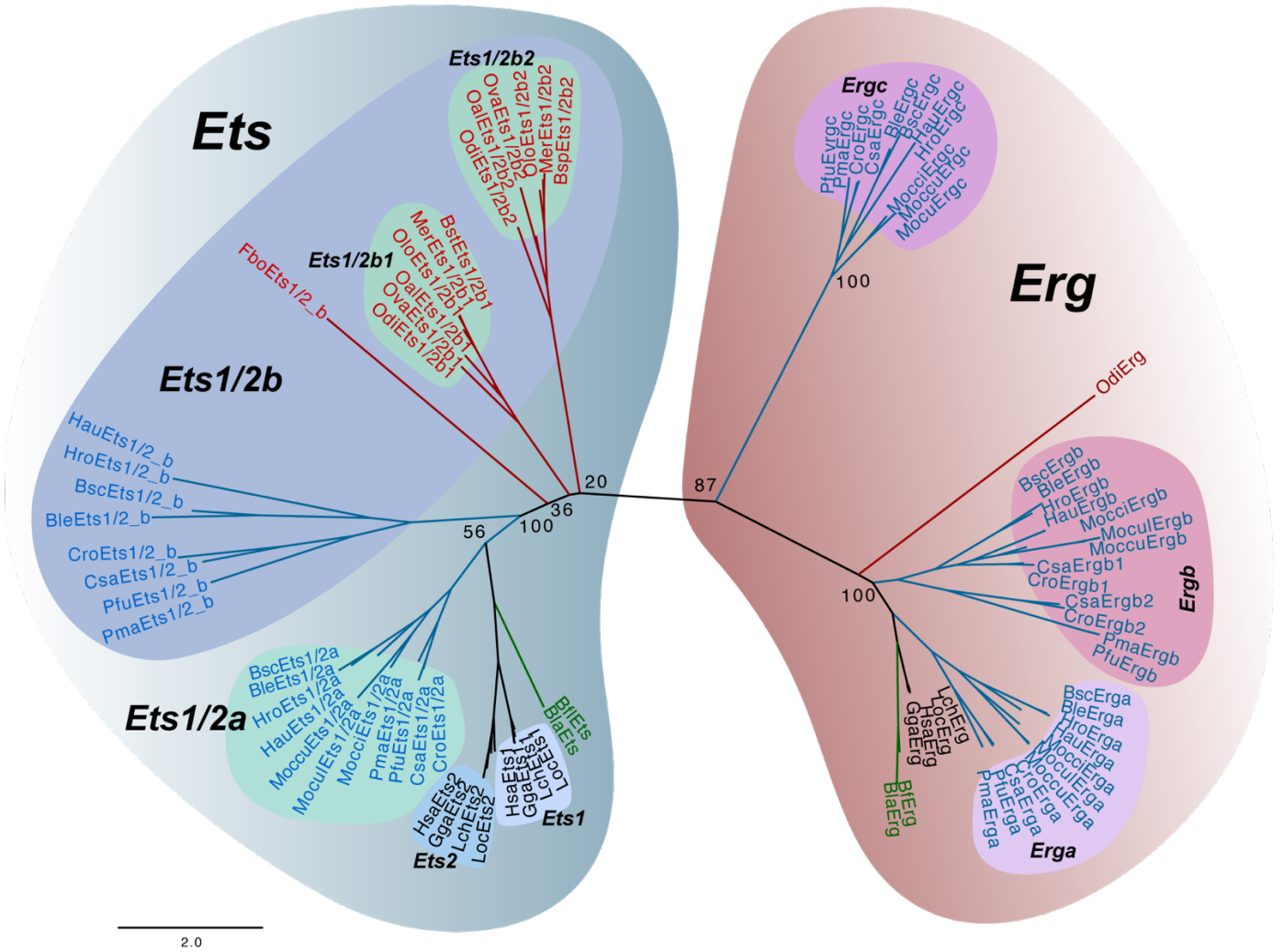
Sup. Fig. 3. Phylogenetic analysis of the *Ets* family and developmental expression patterns of *O. dioica Ets1/2b* paralogs. (**A**) Phylogenetic analysis of the *Ets* and *Erg* protein families showed a high bootstrap value separating both protein families what corroborated the existence of two Ets1/2 proteins in appendicularians. Whole mount in situ hybridization of *O. dioica Ets1/2b1* did not show any clear expression before hatchling stages (**B-E**). In early-hatchling stage *Ets1/2b1* revealed expression in the migratory endodermal strand cells (blue arrowheads) (**F**). In late-hatchling the expression signal was restricted to the buccal gland (yellow arrowheads) (**G**). *Ets1/2b2* did not show expression until tailbud stage (**H, I**). In tailbud embryos, expression signal was detected in tail muscle cells (red arrowheads), the notochord (green arrowheads) and the epidermis of the trunk (pink arrowheads) (**J, K**). In early-hatchling expression signal continued in the tail muscle and the notochord and increased in the anal domain (orange arrowheads) (**L**). In late hatchling stage, the *Ets1/2b2* expression covered the entire oikoplastic epithelium, and continued in the muscle cells of the

A

ERK
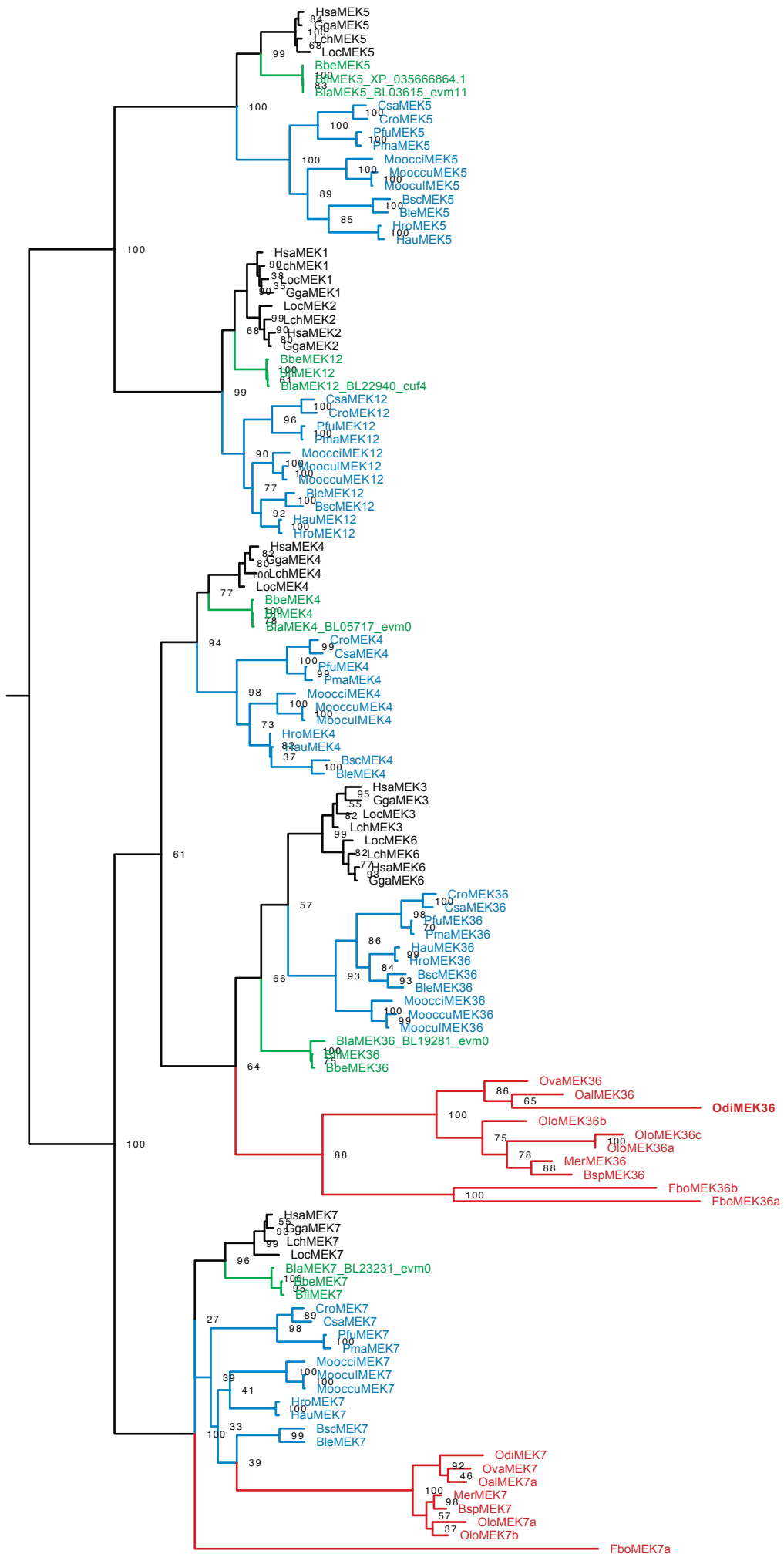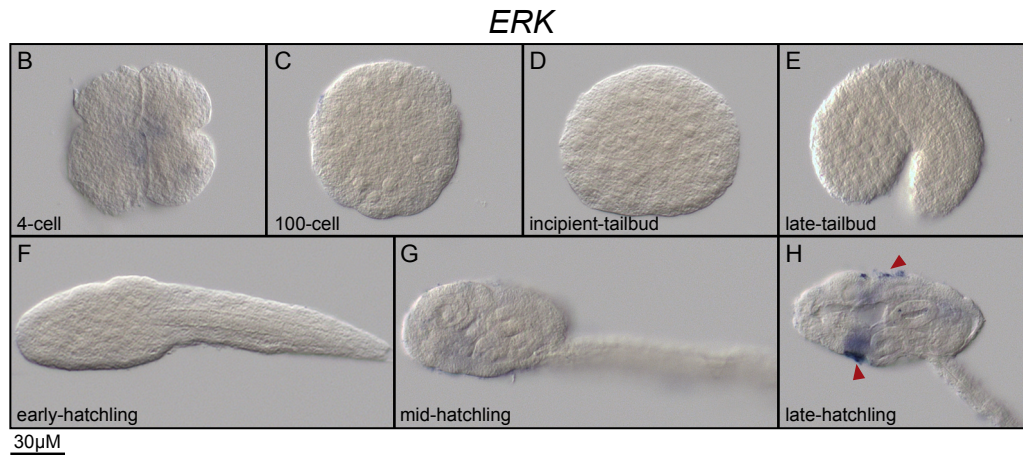
Sup. Fig. 4. (**A**) ML phylogenetic tree of the *MEK* subfamilies in chordates revealing the loss of the *MEK4*, *MEK5* and *MEK1/2* subfamilies in appendicularians, but the surviving of *MEK7* and *MEK3/6* subfamilies. Scale bar indicates amino acid substitutions. Bootstrap values are shown in the nodes. Vertebrates: *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Tunicates: *Bathochordaeus sp.* (Bsp). *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Fritillaria borealis* (Fbo), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Mesochordaeus erythrocephalus* (Mer), *Molgula occidentalis* (Moocci), *Molgula occulta* (Mooccu), *Molgula oculata* (Moocul), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Cephalochordates: *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bfl). Developmental expression pattern of O. dioica ERK homolog (**B-H**). Whole mount in situ hybridization in different stages of *O. dioica* did not detect expression of ERK in any studied stage (**B-G**) until late-hatchling stage when expression is detected in the oikoplastic epithelium (**H**). Images from tailbud in advanced correspond to left lateral views orientated anterior towards the left and dorsal towards the top. Red arrowheads indicate the expression in the oikoplastic epithelium.

BbeGata456_XP_019618792.1
BflGTF456a_ACR66216.1
BlaGata456_AFJ79491.1
0.855
0.581
HsaGata4_NP_001295022.1
GgaGata4_NP_001280035.1
LocGata4_XP_014351854.1
LocGata4_XP_006642996.1
0.978
0.783
0.27
1
HsaGata6_NP_005248.2
GgaGata6_NP_990751.1
LocGata6_XP_006633923.1
LchGata6_XP_005995770.1
0.982
0.1
0.356
0.939
HsaGata5_NP_536721.1
GgaGata5_NP_990752.2
LocGata5_XP_015220493.1
LchGata5_XP_014344220.1
0.86
0.793
0.91
0.9037
0.236
CroGata456_KH.L20.1.v1.A.SL2-1
CsaGata456_Cisavi.CG.ENS81.R77.2996926-3008818.14416.p
PmaGata456_Phmamm.CG.MTP2014.S835.g12260.01.p
PfuGata456_S6271_0_S2222
0.998
0.934
HauGata456_Haaura.CG.MTP2014.S112.g02206.01.p
HroGata456_Harore.CG.MTP2014.S45.g04355.01.p
BleGata456_Boleac.CG.SB_v3.S147.g02878.01.p
BscGata456_ChrUn_Ugene
0.998
0.699
1
0.949
MocciGata456_S458966
MoculGata456_Moocul.CG.ELv1_2.S44512.g02355.01.p
MoccuGata456_S557110_S209818
1
1
1
HsaGata1_NP_002040.1
GgaGata1_NP_990795.1
LocGata1_XP_006625356.2
LchGata1_XP_006008598.1
0.997
0.94
0.002
LchGata3_XP_005993937.1
GgaGata3_NP_001008444.1
HsaGata3_NP_001002295.1
LocGata3_XP_006633355.1
0.941
0.985
0.758
LocGata2_XP_006631129.1
LchGata2_XP_014341892.1
HsaGata2_NP_001139133.1
GgaGata2_NP_001003797.1
0.898
0.983
0.661
0.892
0.789
BbeGata123_XP_019627141.1
BlaGata123_AFJ79490.1
BflGTF123_ACR66214.1
1
0.954
1
CsaGata123_Cisavi.CG.ENS81.R41.2121992-2129284.14153.p
CroGata123_KH.S696.1.v1.R.ND1_1
PmaGata123_S535
PfuGata123_S244_S11342
1
0.874
1
BleGata123_S314
BscGata123_ChrUn_Chr10
BscGata123_Chr10
0.997
0.754
0.867
0.976
HroGata123_Harore.CG.MTP2014.S12.g11731.01.t_S12
HauGata123_Haaura.CG.MTP2014.S24.g00692.01.p_S24
0.995
MoccuGata123_S723572
MoculGata123_S97234
0.995
0.966
MocciGata123_S487156
FboGata_SDII01097846
OalGata_SCLG01000167
OloGata_SCLD01018184
OvaGata_SCLH01000110
OdiGBP3a_GSOIDP00000004001
MerGata_SCLF01239871
BspGata_SCLE01408957
0.92
0.785
0.762
0.658
0
0.958
OloGata_SCLD01172774
OvaGata_SCLH01000841
OvaGata_SCLH01000261
OvaGata_SCLH01002033_F
0.826
0.682
0.954
OloGata_SCLD01106159
OloGata_SCLD01119148
MerGata_SCLF01270895
BspGata_SCLE01420081
0.996
0
0.885
0.686
0.906
0.993
OloGata_SCLD01126560
OalGata_SCLG010000323
OvaGata_SCLH01002033_R
OdiGTFb_GSOIDP00007921001
0.727
0.902
0.913
OdiGBP3b_GSOIDP00011092001
OdiGBP2_GSOIDP00008593001
FboGata_SDII01078871
MerGata_SCLF01530214
BspGata_SCLE01414813
0.95
0
0.942
0.93
0.767
0.773
MerGata_SCLF01211403_1
MerGata_SCLF01415120
BspGata_SCLE01211222
0.672
0.764

96

2.0

Sup. Fig. 5. ML phylogenetic tree of the Gata subfamilies in chordates revealing the loss of the *Gata4/5/6* in appendicularians, but the surviving and lineage specific duplications of *Gata1/2/3* in appendicularians. Scale bar indicates amino acid substitutions. Approximate likelihood-ratio test (aLRT) values are shown in the nodes. Vertebrates: *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Tunicates: *Bathochordaeus sp*. (Bsp). *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Fritillaria borealis* (Fbo), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Mesochordaeus erythrocephalus* (Mer), *Molgula occidentalis* (Moocci), *Molgula occulta* (Mooccu)*, Molgula oculata* (Moocul), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Cephalochordates: *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bfl).

Sup. Fig. 6. ML phylogenetic tree of the FoxF subfamily in chordates reveals the presence of an ortholog of *FoxF* in appendicularians. The sister *FoxQ* subfamily was used as outgroup to root the tree. Scale bar indicates amino acid substitutions. Bootstrap values are shown in the nodes. Vertebrates: *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Tunicates: *Bathochordaeus sp.* (Bsp). *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Fritillaria borealis* (Fbo), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Mesochordaeus erythrocephalus* (Mer), *Molgula occidentalis* (Moocci), *Molgula occulta* (Mooccu), *Molgula oculata* (Moocul), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Cephalochordates:

Nk2

Nk4

Appendicularians
Ascidians
Cephalochordates
Vertebrates

2.0

Sup. Fig. 7. ML phylogenetic tree of the *Nk4* and *Nk2* subfamilies showing an ortholog of *Nk4* in *O. dioica* and two homologs of *Nk2*. Approximate likelihood-ratio test (aLRT) values are shown in the nodes. Vertebrates: *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Tunicates: *Bathochordaeus sp.* (Bsp). *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Fritillaria borealis* (Fbo), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Mesochordaeus erythrocephalus* (Mer), *Molgula occidentalis* (Moocci), *Molgula occulta* (Mooccu), *Molgula oculata* (Moocul), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Cephalochordates: *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bfl).

Sup. Fig. 8. Preliminary ML phylogenetic tree of the *Hand* and *Neurogenin* subfamilies revealed the orthology between the *Hand* of appendicularians and ascidians. Scale bar indicates amino acid substitutions. Approximate likelihood-ratio test (aLRT) values are shown in the nodes. Vertebrates: *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Tunicates: *Bathochordaeus sp.* (Bsp). *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Fritillaria borealis* (Fbo), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Mesochordaeus erythrocephalus* (Mer), *Molgula occidentalis* (Moocci), *Molgula occulta* (Mooccu), *Molgula oculata* (Moocul), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Cephalochordates: *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bfl).

Islet

B 64-cell
C early-tailbud
D late-tailbud
E early-hatchling
F mid-hatchling

Ebf

G 32-cell
H 64-cell
I early-tailbud
J late-tailbud
K mid-hatchling
L late-hatchling

MyoD

M 32-cell
N 100-cell
O mid-tailbud
P early-hatchling
Q late-hatchling

Sup. Fig. 9. Phylogenetic analysis of the *Tbx* family and developmental expression patterns of *O. dioica Islet*, *Ebf,* and MyoD homologs. (**A**) ML phylogenetic tree of the Tbx subfamilies in chordates reveals the loss of Tbx1/10 and Tbx21/Eomes/Tbr1 subfamilies in appendicularians and the ancestral loss of Tbx4/5 subfamily in tunicates. Scale bar indicates amino acid substitutions. Approximate likelihood-ratio test (aLRT) values are shown in the nodes. Vertebrates: *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Tunicates: *Bathochordaeus sp*. (Bsp). *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Fritillaria borealis* (Fbo), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Mesochordaeus erythrocephalus* (Mer), *Molgula occidentalis* (Moocci), *Molgula occulta* (Mooccu), *Molgula oculata* (Moocul), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Cephalochordates: *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bfl). Whole mount in situ hybridization of *O. dioica Islet, Ebf* and *MyoD* homologs (**B-Q**). 64-cell embryos did not showed expression of *Islet* (**B**) which was only detected in the developing nervous system from tailbud to hatchling embryos (**C-F**). *Ebf* (*COE*) did not showed expression in early stages (**G-H**). We detected expression in the nervous system from tailbud to mid-hatchling stage (**I-K**). Late-hatchling embryos showed expression in the epidermis (**L**). We did not detect expression of *MyoD* from 32-cell to hatchling embryos (**M-P**). In late-hatchling embryos MyoD was expressed in the oikoplastic epithelium (Q). Large images from tailbud in advance correspond to left lateral views oriented anterior towards the left and dorsal towards the top. Inset images
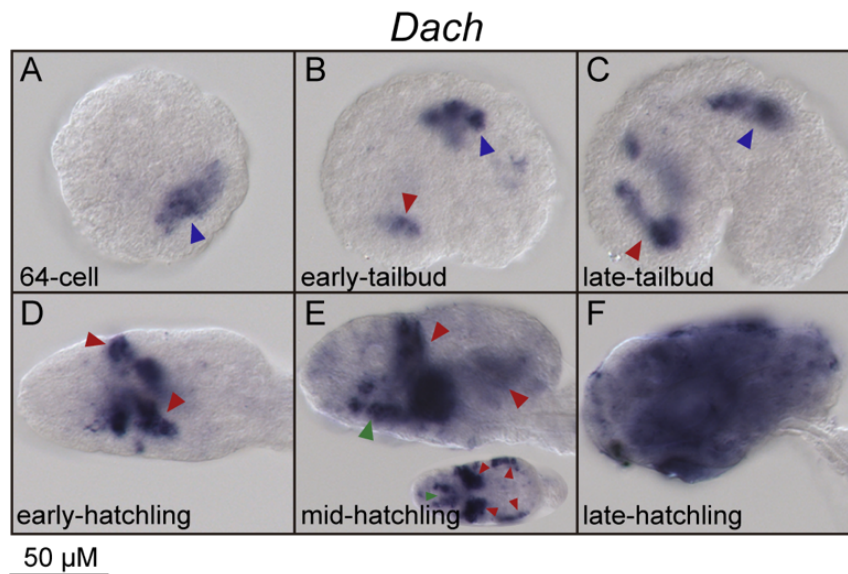
Sup. Fig. 10. Developmental expression pattern of *O. dioica Dach* homolog. Whole mount in situ hybridization of *O. dioica Dach* homolog show expression in the developing nervous system in 64-cell and tailbud stages (**A-C**). In tailbud stages, *Dach* start expressing in the trunk epidermis which is maintain until late-hatchling stages when it is expressed in the whole oikoplastic epithelium (**B-F**). In mid-hatchling stage, beside the epidermis, *Dach* expression is also detected in the endostyle (**E**). Images from tailbud in advance correspond to left lateral views oriented anterior towards the left and dorsal towards the top. Blue arrowheads indicate the nervous system. Red arrowheads indicate the epidermis. Green arrowheads indicate the endostyle.

CHAPTER III

Oikopleura dioica: AN EMERGENT CHORDATE MODEL TO STUDY THE IMPACT OF GENE LOSS ON THE EVOLUTION OF THE MECHANISMS OF DEVELOPMENT

Chapter 4
Oikopleura dioica: An Emergent Chordate Model to Study the Impact of Gene Loss on the Evolution of the Mechanisms of Development

Alfonso Ferrández-Roldán, Josep Martí-Solans, Cristian Cañestro, and Ricard Albalat

Abstract The urochordate *Oikopleura dioica* is emerging as a nonclassical animal model in the field of evolutionary developmental biology (a.k.a. evo-devo) especially attractive for investigating the impact of gene loss on the evolution of mechanisms of development. This is because this organism fulfills the requirements of an animal model (i.e., has a simple and accessible morphology, a short generation time and life span, and affordable culture in the laboratory and amenable experimental manipulation), but also because *O. dioica* occupies a key phylogenetic position to understand the diversification and origin of our own phylum, the chordates. During its evolution, *O. dioica* genome has suffered a drastic process of compaction, becoming the smallest known chordate genome, a process that has been accompanied by exacerbating amount of gene losses. Interestingly, however, despite the extensive gene losses, including entire regulatory pathways essential for the embryonic development of other chordates, *O. dioica* retains the typical chordate body plan. This unexpected situation led to the formulation of the so-called inverse paradox of evo-devo, that is, when a genetic diversity is able to maintain a phenotypic unity. This chapter reviews the biological features of *O. dioica* as a model animal, along with the current data on the evolution of its genes and genome. We pay special attention to the numerous examples of gene losses that have taken place during the evolution of this unique animal model, which is helping us to understand to which the limits of evo-devo can be pushed off.

Alfonso Ferrández-Roldán and Josep Martí-Solans contributed equally to this work.

A. Ferrández-Roldán · J. Martí-Solans · C. Cañestro · R. Albalat (✉)
Facultat de Biologia, Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Catalonia, Spain
e-mail: ralbalat@ub.edu

4.1 Introduction

In 1872, the Swiss marine biologist Hermann Fol during his stay in Messina (Sicilia, Italy) described a new planktonic species of Urochordates (a.k.a. Tunicates), the appendicularian (a.k.a. Larvacean) *Oikopleura dioica* (Fol 1872). From the beginning, *O. dioica* exhibited interesting biological features because, using Fol's own words, "Je n'ai jamais trouvé sur un meme individu les organes male et femelle. Notre espece est réellement un tunicier a sexes distincts" (I have never found on the same individual the male and female organs. Our species is really a tunicate with distinct sexes) (Fol 1872). This was a remarkable finding because until then, it was assumed that Urochordates were hermaphrodites. At the beginning of the twentieth century, the Russian embryologist Vladimir Vladimirovich Salensky (commonly known as W. Salensky) described the anatomy of different Appendicularia species (Salensky 1903, 1904, 1905). Although in 1903 Richard Benedict Goldschmidt provided a first brief description of the larva of *O. dioica* (Goldschmidt 1903), the Dutch biologist Hendricus Christoffel Delsman can be considered the father of *O. dioica* embryogenesis. He, analyzing embryos directly collected from the field, was able to laboriously reconstruct its cleavage pattern with amazing exactitude (Delsman 1910, 1912), which a century later has been confirmed by confocal microscopy and live imaging (Fujii et al. 2008; Stach et al. 2008).

During the 1960s and 1970s, the *O. dioica* anatomy was described in deep detail, paying special attention to some conspicuous structures such as the notochord or the endostyle (Olsson 1963, 1965a, b; Welsch and Storch 1969) and to the organs required for the construction and use of the "house," a filtering complex structure that traps the food (Galt 1972). Almost at the same time, the first *O. dioica* cultures in the laboratory were successfully maintained through numerous generations (Paffenhöfer 1973), and over the next 30 years, numerous ecological and biogeographical studies were reported (e.g., Galt 1978; King et al. 1980; Gorsky et al. 1984; Bedo et al. 1993; Acuña et al. 1995; Hopcroft and Roff 1995; Uye and Ichino 1995; Nakamura et al. 1997; Hopcroft et al. 1998) along with multiple morphological, anatomical, and physiological descriptions (e.g., Last 1972; Bone and Mackie 1975; Olsson 1975; Fenaux 1976, 1986; Flood and Afzelius 1978; Fredriksson and Olsson 1981, 1991; Holmberg 1982, 1984; Fredriksson et al. 1985; Bollner et al. 1986; Georges et al. 1988; Holland et al. 1988; Olsson et al. 1990; Nishino and Morisawa 1998; Lopez-Urrutia and Acuña 1999; Acuña and Kiefer 2000).

All this accumulated knowledge led *O. dioica* to the twenty-first century, when shotgun sequencing techniques (Seo et al. 2001; Denoeud et al. 2010) together with the establishment of protocols for gene expression analysis (Bassham and Postlethwait 2000; Nishino et al. 2000) paved the groundwork for *O. dioica* as a promising nonclassical model species for different biological disciplines, including comparative genomics, molecular genetics, or evolutionary developmental biology (evo-devo). Here, we will review the main biological features of *O. dioica* as chordate model at the morphological, physiological, embryonic, ecological, methodological, evolutionary, genetic, and genomic levels, paying special attention to the

fact that *O. dioica* has shrunk and compacted its genome retaining the archetypical chordate body plan. We will finish by analyzing the impact of gene loss, especially the loss of genes related to embryonic development, on the evolution of *O. dioica*.

4.2   *Oikopleura dioica* as a Nonclassical Model System for Evo-Devo

A model organism is a simplified and accessible system that represents a more complex entity, based on the notion summarized by Jacques Monod: "what is true for *E. coli* is true for the elephant." This notion reflects the evolutionary principle that all organisms share some degree of functional and genetic similarity due to common ancestry. There are several classical model species in cell and developmental biology, from non-vertebrates such as the roundworm *Caenorhabditis elegans* or the fruit fly *Drosophila melanogaster* to vertebrates such as zebrafish and mouse. The choice of a particular model organism depends, essentially, on the specific scientific question that needs to be investigated, and new scientific questions usually demand new model systems. Thus, when we became interested to understand the impact of the gene loss on the evo-devo of chordates, we had to look for a new "nonclassical" model system suitable for our studies. *O. dioica* was our first choice because it has many technical and biological advantages that made it attractive for our objectives: (1) *O. dioica* has a simple and accessible morphology, maintaining the typical chordate body plan throughout its life; (2) its generation time and life span is short with a fast, invariant, and determinative embryonic development; (3) *O. dioica* is easily accessible and affordably cultured; (4) *O. dioica* can be experimentally manipulated; (5) *O. dioica* occupies a privileged phylogenetic position within the chordates to better understand the evolution of our own phylum; and (6) its genome is small, compacted, and prone to lose genes. In the next sections, we will summarize these technical and biological advantages.

4.2.1   O. dioica *Is a Morphologically Simple Animal with a Typical Chordate Body Plan*

O. dioica has a simple but typical chordate body plan with organs and structures that are unequivocally homologous to those in vertebrates and other chordates, including a notochord anchoring muscle cells along a postanal tail, a hollow neural tube that becomes a central nervous system (CNS) organized in two big ganglia and a nerve cord, as well as a pair of gill slits and an endostyle in the pharyngeal region, which connects with the digestive tract (Fig. 4.1). The body of *O. dioica* is divided in two parts, the trunk (0.12–0.85 mm), which contains most of the organs, and the tail (1.2–3.2 mm), which contains the notochord, the tail nerve cord, and two rows of
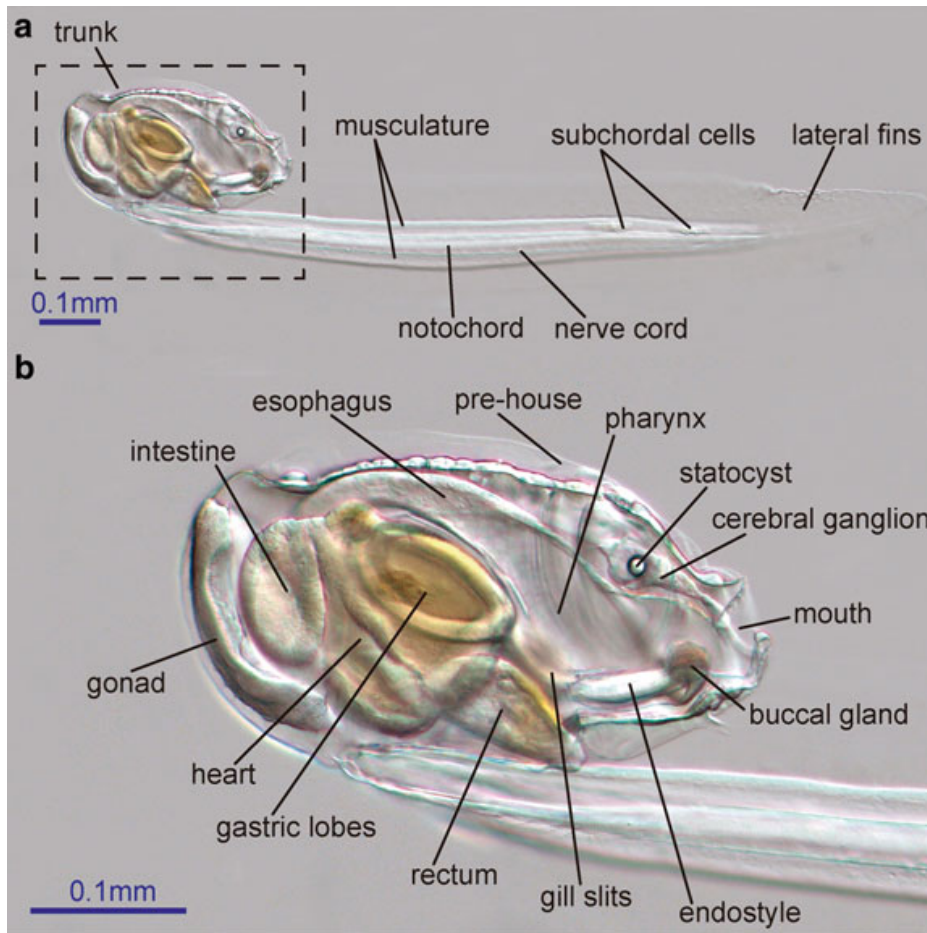
Fig. 4.1 Immature adult specimen of *O. dioica*. (**a**) Differential interference contrast (DIC) micrograph in right lateral view with anterior to the left and ventral down. (**b**) Magnification of the trunk showing the main organs and structures. Bar: 0.1 mm

paired muscle cells. The whole body is transparent so the inner structures can be visualized by simply changing the focal plane of the microscope.

4.2.1.1 Nervous System

The CNS of *O. dioica* is made by only approximately 100 cells. Despite this extremely simple structure, *O. dioica* CNS shows homology with the vertebrate forebrain, hindbrain, and spinal cord of vertebrates (but not the midbrain) (Cañestro et al. 2005). The CNS of *O. dioica* comprises a nerve cord running along the tail and two big ganglia (i.e., the cerebral ganglion in the most anterior part of the trunk and the caudal ganglion in the most proximal region of the tail) (Holmberg 1984). Both ganglia secrete gamma-aminobutyric acid (GABA) (Soviknes et al. 2005). The cerebral ganglion, proposed to be homologous to the vertebrate forebrain (Cañestro et al. 2005), is made of approximately 70 neurons and includes a statocyst, a balance sensory receptor in which a mineralized mass called statolith pushes a group of ciliary cells in response to gravity, providing therefore feedback to the animal on

change in orientation (Holmberg 1984) (Fig. 4.1). The cerebral ganglion connects with sensory cells in the mouth, the pharynx, and the ventral organ (a simple structure of about 30 cells homologous to the olfactory placode and the pituitary of vertebrates) (Bollner et al. 1986; Cañestro et al. 2005) and integrates the information received from the sensory cells. The cerebral ganglion also connects to the caudal ganglion through the trunk nerve cord, which turns to the right before entering the tail. Caudal ganglion and trunk nerve cord have been proposed to be homologous to at least part of vertebrate hindbrain (Cañestro et al. 2005). The caudal ganglion and the nerve cord that runs along the tail consist of approximately 30 neurons. Along the nerve cord, neuronal nuclei are collected in groups of two to four forming small ganglia. Small fibers arising from these ganglia innervate the symmetric musculature of the tail with cholinergic motoneurons and coordinate its movements (Galt 1972; Soviknes and Glover 2007).

4.2.1.2 Epidermis and Secreted "House"

The epidermis is a monolayered epithelium without any underlying mesodermal tissue that covers the whole body of *O. dioica* (Nishida 2008). The epidermis of the trunk, called oikoblast or oikoplastic epithelium, contains a fixed number of cells (approximately 2000) organized in different domains defined by the shape of the cells and the morphology of their nuclei and showing a complex bilateral symmetric pattern (Thompson et al. 2001; Kishi et al. 2017; Mikhaleva et al. 2018) (Fig. 4.2a). The oikoplastic epithelium secretes the so-called house, a filter-feeding device made of glycopolysaccharides, mucopolysaccharides, and cellulose (Fig. 4.2b, c) (Kimura et al. 2001; Spada et al. 2001; Thompson et al. 2001). The distribution of the epidermal cells of the trunk has a direct correlation with the different architectonic structures of the house (Fig. 4.2a) (Spada et al. 2001; Thompson et al. 2001; Kishi et al. 2017). The adult specimens of *O. dioica* live inside this mucous house (Fig. 4.2b, c) that works as a food-trapping device by filtering picoplankton particles of different sizes (from 0.1 to 50 µm) from the water stream propelled by stylish and grooving tail movements (Fenaux 1986; Thompson et al. 2001). The house is abandoned when the filters are obstructed, and a new one is immediately inflated, which happens between four and ten times a day increasing with higher water temperatures (Flood and Deibel 1998; Sato et al. 2003). This energetically exhausting strategy has been proposed as an evolutionary adaptation to poor food environments (Acuña et al. 2002).

4.2.1.3 Pharyngeal Region and Digestive Tract

The food particles caught by the house enter through the mouth situated at the anterior region of the trunk to the pharynx, thanks to the water flux generated by the spiracles, a pair of gill slits with beating cilia that generate a water stream into the body to draw the food (Fenaux 1998a) (Fig. 4.3). Within the pharynx, food particles
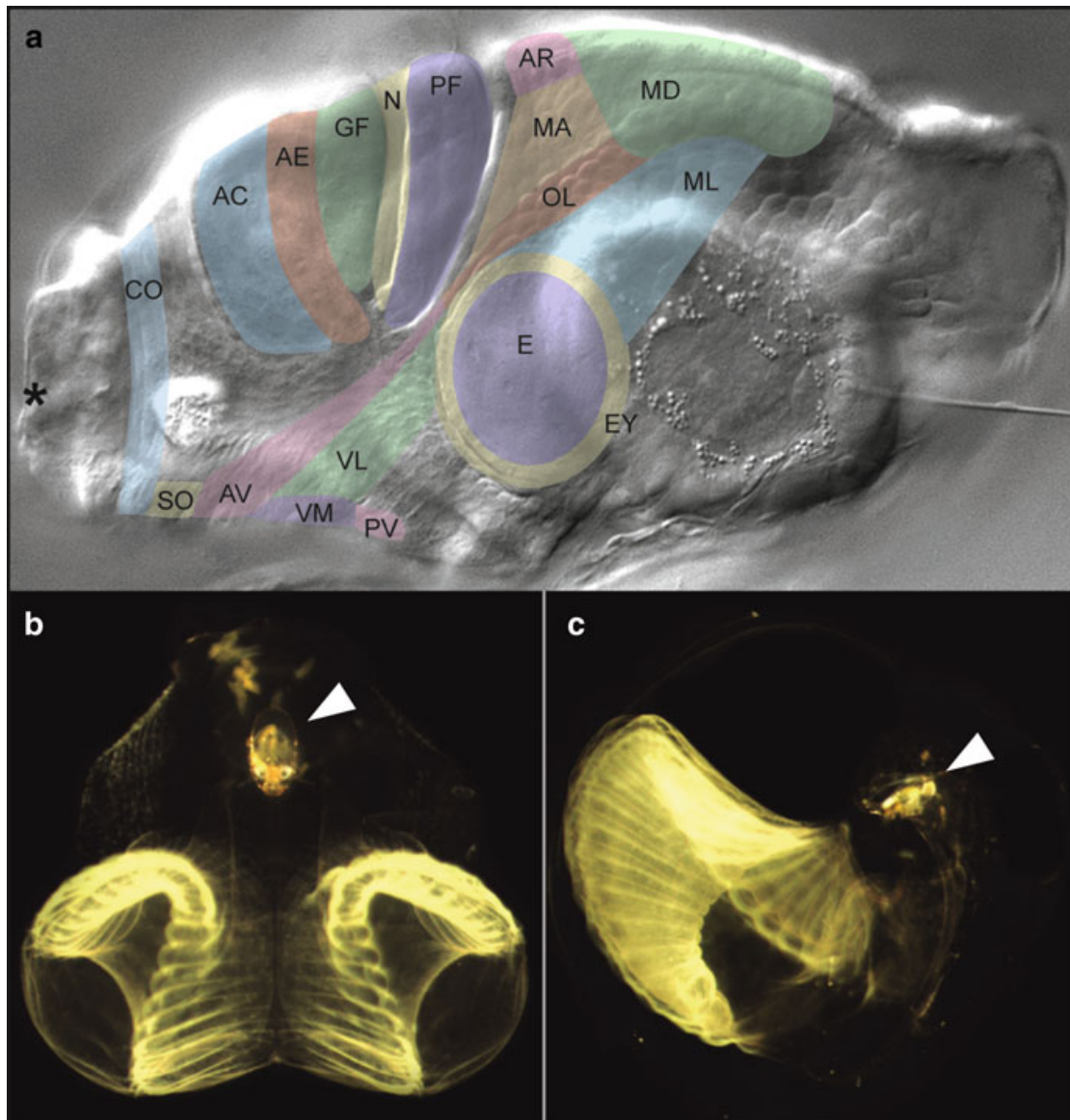
Fig. 4.2 The *O. dioica* epidermis and the secreted house. Lateral view of an *O. dioica* trunk, in which the main domains of the oikoplastic epithelium that correlate with the different architectonic structures of the house are depicted (**a**): *AC* anterior crescent, *AE* anterior elongated region, *AR* anterior rosette, *AV* anterior ventral domain, *CO* circumoral domain, *E* Eisen giant cells, *EY* Eisen's yard, *GF* giant Fol, *MA* martini, *ML* mid-lateral domain, *MD* middorsal domain, *N* nasse, *OL* oblique line, *PF* posterior fol, *PV* posterior ventral domain, *SO* sensory organ, *VL* ventrolateral domain, *VM* ventromedial domain. AC, AE, GF, N, and PF form the fol domain (Kishi et al. 2017). Anterior is to the left and ventral down. The asterisk indicates the mouth. Frontal (**b**) and lateral (**c**) views of the *O. dioica* house, which is visible thanks to the algae trapped in the structure. The trunk of the animal inside the house is also visible (arrowhead)

are trapped by the mucus that covers its walls. The mucus is secreted by the endostyle, an organ homologous to the thyroid of vertebrates, and localized on the ventral side of the pharynx (Olsson 1963; Cañestro et al. 2008). This mucus is conducted to the digestive tract by two symmetric rows of cilia at the lateral walls of the pharynx that finally meet at the superior region (Fenaux 1998a). The digestive
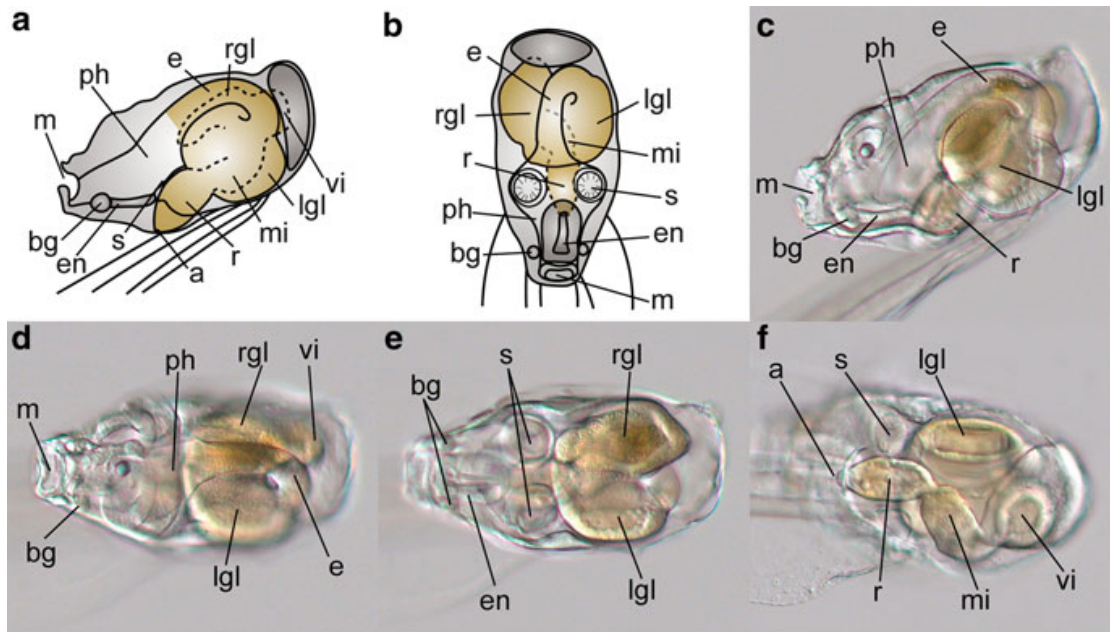
Fig. 4.3 Pharyngeal region and digestive system in the *O. dioica* trunk. Diagrams of the trunk and anatomy of the pharyngeal region and digestive tract of *O. dioica* from left (**a**) and dorsal (**b**) views. DIC micrographs of the trunk of an *O. dioica* adult specimen in left lateral view (**c**), dorsal views at two different focal planes (**d**, **e**), and ventral view (**f**). Anterior is to the left. Anus (a), buccal gland (bg), endostyle (en), esophagus (e), left gastric lobe (lgl), mid-intestine (mi), mouth (m), pharynx (ph), rectum (r), right gastric lobe (rgl), spiracles (s), and vertical intestine (vi)

tract is composed of an esophagus, two gastric lobes (right and left stomachs), and a bent intestine divided in vertical, mid, and distal intestine (or rectum) (Burighel and Brena 2001; Cima et al. 2002) (Fig. 4.3). Although the absorption of liquids, ions, and small molecules takes place all over the digestive system (Burighel and Brena 2001), different compartments have different functions. Digestive enzymatic activity, for instance, is led mainly by the left gastric lobule where different enzymes such as α-amylase, acid phosphatase, nonspecific esterase, 5′-nucleodidase, and aminopeptidase M are released (Cima et al. 2002). In contrast, accumulation of lipid droplets in ciliated cells of the right gastric lobule and the vertical and mid-intestines indicates a storage role for these compartments (Cima et al. 2002). Finally, fecal pellets are formed in the vertical intestine and transit along the intestine allowing protein absorption by the rectal granular cells (Burighel and Brena 2001; Cima et al. 2002), until reaching the anus where they are released. Functional compartmentalization is further supported by spatiotemporal differences in the onset of gene expression during the development of the digestive system (Martí-Solans et al. 2016), likely reflecting processes of functional differentiation and specialization of digestive cells.
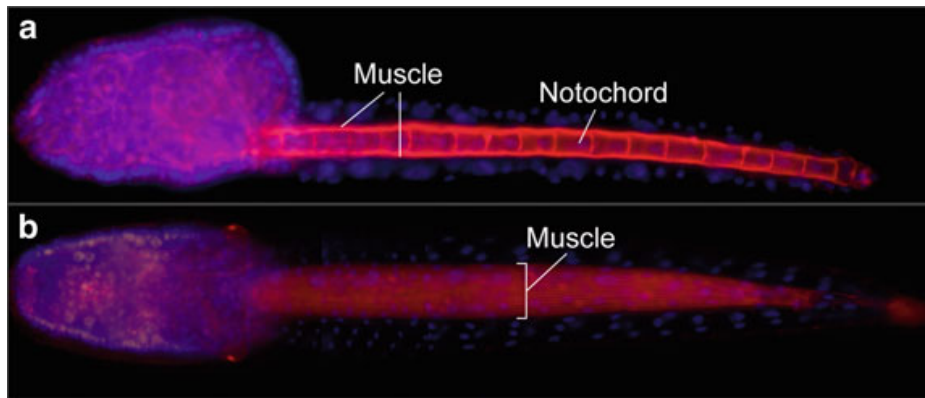
Fig. 4.4 Notochord and tail musculature in an *O. dioica* larva. These two tissues stand out over the general phalloidin staining of the actin filaments (in red), and they spread from the base of the trunk to the posterior tip. In lateral view, notochord is in the center of the tail flanked by two strips of striated muscle cells located on dorsal and ventral sides (**a**). In ventral view, sarcomeres are distinguished as repeated structures all along the strip of muscle cells (**b**). Nuclei are stained with Hoechst (in blue). Anterior is to the left

4.2.1.4 Notochord and Subchordal Cells

The notochord is a synapomorphic structure of the chordate phylum (Satoh et al. 2012) that has a dual function: providing structural support to the developing embryo and secreting patterning signals required for tissue specification and organogenesis (Cleaver and Krieg 2001). In *O. dioica*, the notochord is a row of cells or "stack of coins" that runs along the tail (Figs. 4.1 and 4.4). During larval development, notochord cells form vacuoles that eventually coalesce to form a hollow lumen (Soviknes and Glover 2008). In the adult, the notochord is a tube of cells with thin walls, whose cells secrete proteins rich in sulfur to the extracellular media to keep the stiffness of the notochord, but also flexible enough to facilitate swimming and other movements (Olsson 1965a). The genetic toolkit to build the notochord in *O. dioica* has been investigated, revealing not only similarities but also significant differences in the complement of genes employed by different chordates (Ferrier 2011; Kugler et al. 2011). The evolutionary origin of the notochord is a hot topic of discussion, and the analysis of the *O. dioica* notochord has contributed to corroborate that the urochordate noncontractile actin-expressing notochord may represent the ancestral condition in stem chordates (Almazán et al. 2019; Inoue and Satoh 2018).

At the right side of the notochord, there is a pair of conspicuous cells, the so-called subchordal cells (Fol 1872; Delsman 1910; Lohmann 1933) (Fig. 4.1). These cells migrate during *O. dioica* development from the trunk to the tip of the tail, after the endodermal strand cells enter the trunk following the same path but in opposite directions (Kishi et al. 2014). Subchordal cells together with endodermal strand cells and oral gland cells are three cell populations exhibiting long-distance migration during *O. dioica* development (Kishi et al. 2014), two of them—the subchordal and the endodermal strand cells—likely sharing the same origin (Fredriksson and Olsson 1991). The function of the subchordal cells remains elusive, although it has been proposed that they may be involved in the exchange of materials

with the body fluid, from which they pick and transform substances in a hepatic-like process (Fredriksson and Olsson 1991). These substances can either be changed into low molecular weight products merely for detoxification of harmful substances or be metabolized into new, useful molecules that are released back to the hemolymph (Fredriksson and Olsson 1991; Fenaux 1998a).

4.2.1.5 Heart and Tail Musculature

In *O. dioica*, muscle cells are only found in the heart and in the tail. They are in both cases striated muscle cells showing the characteristic repeating functional units called sarcomeres (Onuma et al. 2017; Almazán et al. 2019) (Fig. 4.4b), while the existence of smooth muscle cells in *O. dioica* has not been established so far. *O. dioica* heart is the simplest chordate heart, and assisted by the tail movement, it contributes to the circulation of the hemolymph between the ectoderm and the internal organs in an open circulatory system (Fenaux 1998a). The heart consists of a posterior ventral bag wedged between the left stomach wall and the intestine (Fig. 4.1). The internal lumen of this bag is the only coelomic space in *O. dioica*, and it is completely lined by two types of flat mesodermal tissues, the myocardium composed by muscle fibers and the nonmuscular pericardium (Stach 2009; Onuma et al. 2017). Myocardial cells are connected by cell junctions, and their cytoplasm is full of actin filaments (Stach 2009; Onuma et al. 2017; Almazán et al. 2019). This muscle-structured tissue produces peristaltic contractions that periodically reverse, causing the hemolymph to course between the myocardium and the stomach wall reaching the rest of the organs in the trunk and circulating through the tail (Fenaux 1998a).

The muscular tissue of the tail spreads posteriorly from the base of the trunk in two strips of striated cells distributed in the dorsal and ventral sides of the notochord (Nishino et al. 2000; Nishino and Satoh 2001; Almazán et al. 2019) (Fig. 4.4). This paraxial musculature, combined with the notochord, provides the ability to drive movement to the tail, which is fundamental not only for locomotion but also for *O. dioica* feeding. The tail musculature of *O. dioica* consists of only ten multinucleated striated cells on the dorsal and ventral sides of the notochord (Nishino et al. 2000; Nishino and Satoh 2001; Soviknes et al. 2007; Almazán et al. 2019). In just hatched larvae, eight mononucleated muscular cells are easily recognizable along the anterior–posterior axes of the tail. Along the larval development, two additional small muscle cells appear at the tip of the tail although their origin remains unknown (Nishino and Satoh 2001).

Structurally, the actomyosin contractile complexes of the muscle cells in *O. dioica* have not been fully characterized, but actin and myosin genes can be identified in the *O. dioica* database (Almazán et al. 2019). *O. dioica* has four muscular-actin genes that appear to be expressed only in the muscle cell lineage (Almazán et al. 2019). The four muscular actins show differences in their expression domains during embryonic development, which suggests differences in their spatio-temporal regulation (Almazán et al. 2019). Muscular actin expression can be
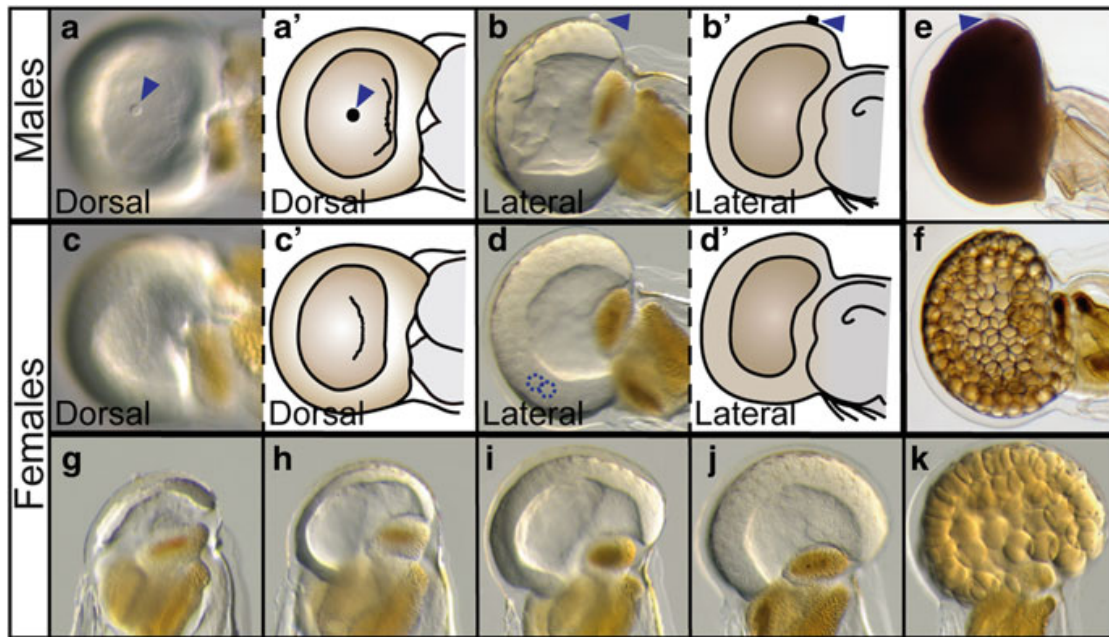
Fig. 4.5 Maturation of male and female gonads. DIC micrographs and schematic representations of dorsal (**a–a'**) and lateral (**b–b'**) views of a premature male gonad. The sperm duct appears as a small protuberance in the most dorsal part of the male gonad (arrowhead). DIC micrographs and schematic representations of dorsal (**c–c'**) and lateral (**d–d'**) views of a premature female ovary. Pro-oocytes can be distinguished inside the gonad (dashed circles). DIC micrographs of mature male (**e**) and female (**f**) gonads. Mature male gonads become brown and opaque as spermatogenesis progresses, while female gonads become yellowish and refringent until the round eggs are clearly visible (**g–k**)

detected as early as at 64-cell stage, although muscle fibers are formed later in the larvae. They are restricted to the inner part of the cell, while the epidermal side is full of mitochondria generating the energy for the tail movement (Nishino et al. 2000; Almazán et al. 2019). Comparative analyses among different chordate species, including *O. dioica*, have suggested that muscular actins originated from a cytoplasmic actin at the base of chordates, which was followed by the recruitment of the myosin motor-machinery that conferred contractile capability to muscle cells (Kusakabe et al. 1997; Almazán et al. 2019; Inoue and Satoh 2018).

4.2.1.6 Reproductive System

O. dioica is the sole dioecious species so far reported in the urochordate subphylum (Fol 1872; Nishida 2008). The gonad is localized in the posterior region of the trunk and is the only apparent character that allows us to differentiate males from females: male gonad is full of sperm and homogenously brown and dark, whereas the female one is yellowish and refringent as it is full of eggs (Fig. 4.5). The growth of the gonad is parallel to the growth of the animal, acquiring their final size and maturity in only some hours, when gametogenesis occurs (Nishino and Morisawa 1998; Ganot et al. 2007). Spermatogenesis takes place in the testicle, a syncytium with a

substantial number of identical nuclei that become individual spermatozoa (Martinucci et al. 2005; Onuma et al. 2017). The ovary has also an initial phase of syncytial nuclear proliferation. Then, it becomes a coenocyst where half of the nuclei enter meiosis, whereas the other half became highly polyploid nurse nuclei (Ganot et al. 2007). Inside the coenocyst, a ramified structure composed of filamentous actin surrounds each of the pro-oocytes that are connected to the common cytoplasm via intercellular bridges termed also ring canals (Ganot et al. 2007). During oogenesis a subset of pro-oocytes grows by transferring the cytoplasm through the ring canals while the others degenerate (Ganot et al. 2007). Typically, a mature female spawns between 100 and 400 eggs, although as a clutch manipulator, *O. dioica* females might increase or decrease the oocyte number depending on favorable or unfavorable environmental conditions, respectively (Troedsson et al. 2002). This feature can be useful to manipulate the clutch size in laboratory conditions.

4.2.2 O. dioica *Has a Short Generation Time and a Determinative Embryonic Development*

A convenient characteristic shared by many model organisms is that their generation time (i.e., the time between two consecutive generations) is short, facilitating genetic analyses through generations. The generation times of *D. melanogaster* and *C. elegans*, for instance, are of only 9 and 3 days, respectively. The generation time of *O. dioica* is also short, from 5 days at 19 °C to 10 days at 13 °C (Nishida 2008; Bouquet et al. 2009; Martí-Solans et al. 2015). After the external fertilization, the zygote starts embryonic development, which lasts 3.5–6.0 h at 19 °C and 13 °C, respectively, and terminates when a larva hatches breaking the chorion (Fig. 4.6). Next, larval development starts; it takes 6–13 h (at 19 °C–13 °C, respectively) and terminates with a 120° twist of the tail at the tailshift stage. Ten minutes later, the juvenile makes its first house and starts feeding. During the next few days, animals filtrate water and feed arriving to the 1.5–4.0 mm of body size for an adult animal. At that point, the gonads complete their maturation, and males and females spawn the gametes and, as semelparous animals, die afterwards (Fig. 4.6).

In vitro fertilization protocols (Fenaux 1976; Holland et al. 1988; Nishino and Morisawa 1998) have enabled the understanding of processes that lead the egg to become a larva. Cleavage patterns, cell lineages, and morphogenetic movements, which are especially important for the understanding of the embryology of organisms that follow an invariant and determinative embryonic development such as those of *O. dioica*, have been extensively studied (reviewed in Nishida and Stach 2014). Thus, fate maps from one cell to the tail bud stage have been established (Fig. 4.7), similarly to what has been made in other organisms as the nematode *C. elegans* or the ascidian *Ciona robusta* (formerly *C. intestinalis*) (Sulston et al. 1983; Nishida 1987; Fujii et al. 2008; Stach et al. 2008; Nishida and Stach 2014; Stach and Anselmi 2015). After fertilization, the surface of the egg becomes rough,
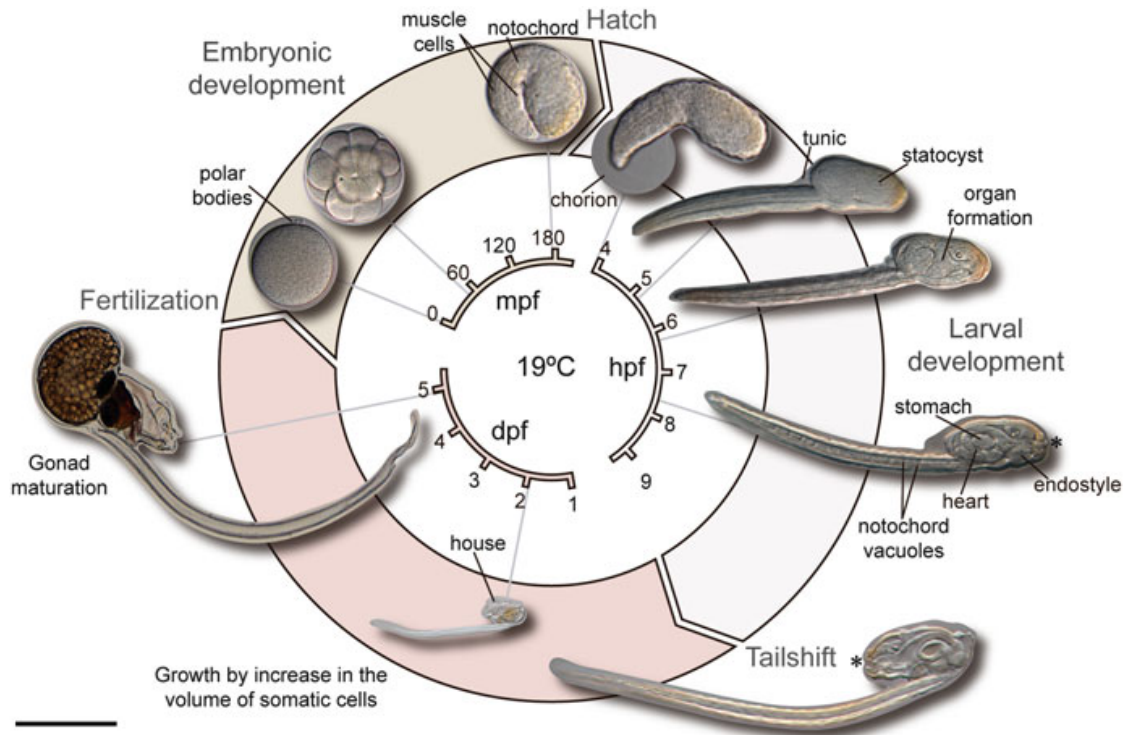
Fig. 4.6 Schematic representation of *O. dioica* life cycle. Embryonic development starts with the fertilization of the oocyte and ends when a larva hatches breaking the chorion (≈3.6 hpf at 19 °C). Larval development lasts for ≈6 h (at 19 °C) and ends when the tail of the larva changes 120° its orientation in a process called tailshift. Notice the change of the position of the mouth (asterisk) relative to the tail in juvenile animals. During the next ≈4.5 days (at 19 °C), juvenile animals feed, grow, and become mature males or females, which spawn the gametes closing the cycle. Scale bar represents 100 μm for all stages except for day 2 and mature adults (1 mm)
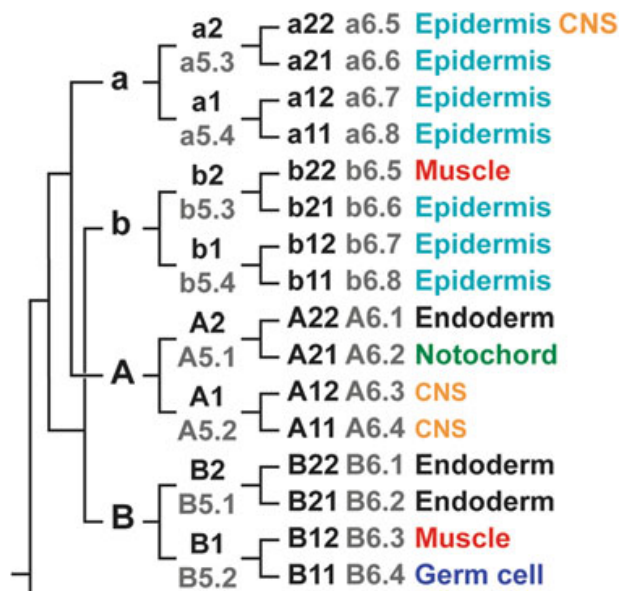
Fig. 4.7 Cell-lineage tree of *O. dioica* embryos. The tree is bilaterally symmetric, and therefore, only one-half is shown. The nomenclature agrees with that of Delsman (1910) and reviewed in Nishida (2008). The ascidian nomenclature system is shown in gray according to Stach et al. (2008) and Nishida and Stach (2014). Developmental fates of cells at 32-cell stage are indicated. CNS, central nervous system

and 9 min later, two polar bodies appear as an obvious signal of successful fertilization (Fig. 4.6) (Fenaux 1998a; Fujii et al. 2008). At 19 °C, the first division takes place after 23 min postfertilization (mpf), generating two morphologically equal cells that correspond to the right and left sides of the animal. At 32 mpf, another symmetric division takes place. This division is meridional and perpendicular to the previous one giving place to the anterior and posterior hemispheres. The third division (at 40 mpf) is asymmetric and leads to the formation of 8 cells, 4 big cells in the animal pole and 4 small cells in the vegetal pole. The morula stage appears 52 mpf, after the fourth division generates an embryo composed of 16 cells. Gastrulation starts 67 mpf, at 32-cell stage, and it consists in the ingression of the vegetal blastomeres that remain covered by the animal blastomeres (Nishida 2008). Neurulation starts at the 64-cell stage 80 mpf, when 8 cells from the anterior region are aligned in two rows of 4 cells that form a matrix that is internalized. Tail bud stage starts 135 mpf, when tail and trunk differentiate. During this stage, the tail elongates, and the embryo bents ventrally. Notochord cells align in a single row and become evident at 180 mpf. At 3.6 h postfertilization (hpf), a larva hatches, and embryogenesis terminates (Fig. 4.6) (Delsman 1910; Cañestro et al. 2005; Fujii et al. 2008; Nishida 2008; Stach et al. 2008).

Larval development is subdivided into six stages (I–VI) (Nishida 2008) that last 6 h at 19 °C, from the hatchling until the juvenile stage (Fig. 4.6). At stage I, the larva elongates and occasionally moves by tail beats. At stage II, the boundaries of the organs begin to appear, and at stage III, the organs are perceptible. At stage IV, organs are clearly recognizable, mouth opens, and the cilia of the digestive system and the ciliary rings of the gill slits start moving. Buccal glands locate at each side of the endostyle, and the heart starts beating. Notochord vacuoles fuse, and the tail flattens to form the lateral fins. At stage V, water current starts inside the larva, the lumen of all organs is continuous, and the trunk epidermis (oikoplastic epithelium) secretes house materials (pre-house). Swimming movements are vigorous. At stage VI, after a few seconds of intense movement, the tail changes its orientation 120° during a process called tailshift (Delsman 1910; Galt and Fenaux 1990; Fenaux 1998b; Nishida 2008). The tailshift event occurs when the larva reaches the juvenile stage, and it is thought comparable to metamorphosis in other urochordates (Fig. 4.6), although in *O. dioica* the definitive organs develop earlier than in ascidian species (Galt and Fenaux 1990).

After the tailshift, juvenile animals inflate the first house and start to feed. During the next days, animals filtrate water and feed at the time that they grow and their gonads mature. From day 1 to day 3, the trunk grows from 0.17 to 0.35 mm (Bouquet et al. 2009), and the gonad expands in the posterior and ventral part of the trunk to fill the whole posterior region (Fig. 4.5). On day 4, the gonad increases its size becoming wider than the trunk and becomes visible to the unaided eye. At day 5, the trunk of *O. dioica* measures almost 1 mm and the whole animal between 3 and 4 mm (Bouquet et al. 2009). At that time, the gonad is fully developed, and males and females are easily distinguishable as the gonad is full of sperm or eggs, respectively (Figs. 4.5 and 4.6).
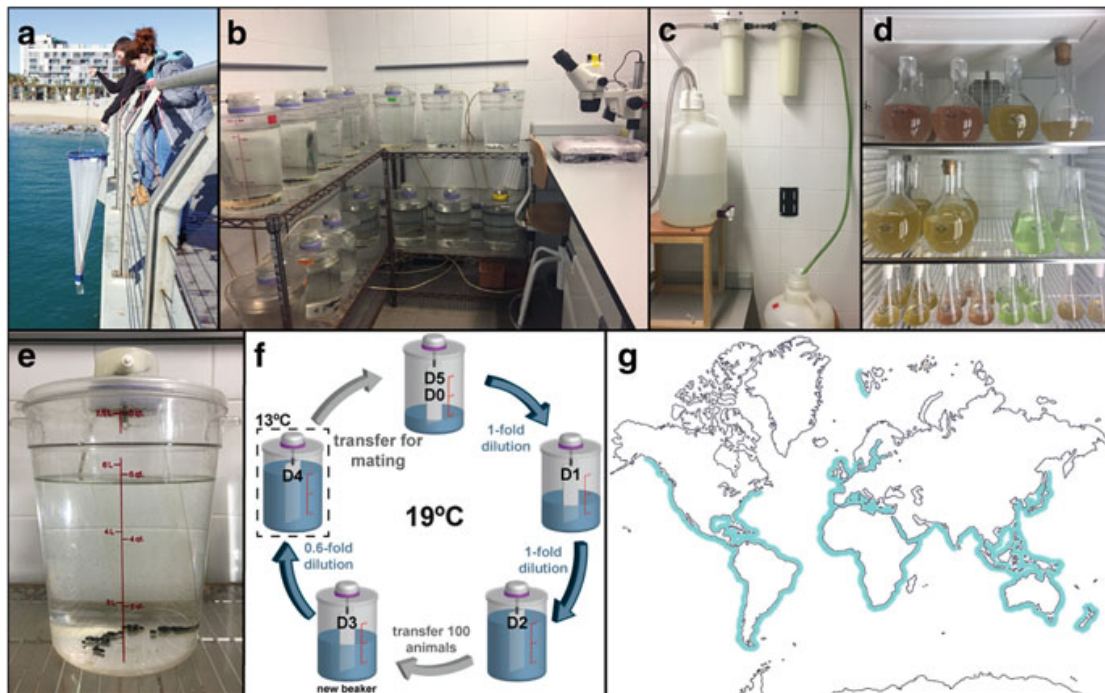
Fig. 4.8 *O. dioica* culture and geographic distribution. Animals might be collected in the coast using a plankton net or directly with a bucket (**a**) and maintained in the laboratory at 19 °C using a standard air conditioning device (**b**). Seawater is filtered through 10–0.5 μm polypropylene filters to remove major particles and organisms and sterilized using 0.2 μm filters (**c**). Three algal species and a cyanobacteria are individually cultured in a 13 °C incubator with a 12 h light/dark photoperiod and used to feed the animals (**d**). *O. dioica* animals are cultured at 19 °C in polycarbonate backers and maintained in suspension by the rotation (5–15 rpm) of a paddle driven by an electric motor mounted on the lid of the backer (**e**). A new animal culture starts by mating 20 females and 10 males close to spawn (D0) in a baker with 1.5 L of sterilized seawater (**f**, top). Next 2 days (D1 and D2), the culture is just diluted adding 1.5 and 3 L of sterilized seawater, respectively (**f**, right). At day 3 (D3), ≈100 animals are transferred to a new backer with 4 L of fresh sterilized seawater (**f**, left-down), and at day 4 (D4), cultures are diluted adding 2 L of sterilized seawater and moved to a 13 °C incubator (**f**, left-top). In the morning of day 5 (D5), most of the animals are mature and ready to set a new mating for new generation (**f**, top). For additional details, see Martí-Solans et al. (2015). Geographic distribution of the *O. dioica* species in the coast of most of the oceans (blue line) (reviewed in Fenaux et al. 1998, and supplemented with Bary 1960; Fenaux 1972; Costello and Stancyk 1983; Larson 1987; Gaughan and Potter 1994; Hwan Lee et al. 2001; Aravena and Palma 2002; Walkusz et al. 2003; Menéndez et al. 2011) (**g**)

4.2.3 O. dioica *Is Easily Accessible, with a Wide Geographic Distribution and Suitable for Lab Culturing*

4.2.3.1 Geographic Distribution and Abundance

O. dioica is a semi-cosmopolite free-swimming species present in the Atlantic, Pacific, and Indian Oceans as well as in the Mediterranean and Red Seas, but not in extreme latitude oceans as Arctic or Antarctic (Fenaux et al. 1998) (Fig. 4.8). Although *O. dioica* is a widespread species, its populations may vary in a seasonal way (Essenberg 1922; Raduan et al. 1985; Uye and Ichino 1995; Fenaux et al. 1998;

Tomita et al. 2003). In the west Mediterranean Sea, for instance, its maximum abundance is found during spring and autumn seasons (Raduan et al. 1985). *O. dioica* along with other appendicularian species are so abundant that they occupy an important trophic position in food webs. Appendicularians are the second most abundant species, after copepods, in marine mesozooplankton (Gorsky and Fenaux 1998; Captanio et al. 2008), and they graze about 10% of the ocean's primary production (Acuña et al. 2002). Acting as an important short-circuit that allows a rapid energy transfer from colloidal carbon and phytoplankton primary producers to zooplanktivorous predators such as fish larvae (Flood and Deibel 1998; Fernández et al. 2004), they contribute to at least 8% of the vertical carbon transport to the deep ocean (Davoll and Youngbluth 1990; Robison et al. 2005; Troedsson et al. 2013). Appendicularians are, indeed, a major contributor of "marine snow" (i.e., biological debris that originates from the top layers of the ocean and drifts to the seafloor) in euphotic and mesopelagic zones through the production of discarded houses and fecal pellets (Robison et al. 2005). Those particles, full of nutrients, contribute to 28–39% of total particulate organic carbon export to the deep oceans (Alldredge 2005). The ecological relevance of appendicularians is so high that human activities affecting their populations [e.g., global warming or an increase of toxins in oceans (Bouquet et al. 2018; Torres-Aguila et al. 2018)] might impact on marine food webs and in vertical carbon flux at a world scale. Therefore, the *O. dioica* populations might be valuable sentinels for monitoring marine ecosystems. In summary, *O. dioica* is an ecological relevant organism, abundant in the neritic zone in almost all marine costs, and easily accessible using ordinary plankton nets (Fig. 4.8a, g).

4.2.3.2   *O. dioica* Lab Culturing

In addition to its accessibility directly from nature, *O. dioica* cultures can be maintained in the laboratory (Fig. 4.8b, f). *O. dioica* was first cultured by Paffenhöfer in 1973, but the bases for a long-term maintenance system were developed by Fenaux and Gorsky (1985). Nowadays, to our knowledge, the procedures to culture *O. dioica* in the laboratory all year around have been published from three facilities in the world located in Norway, Japan, and Spain (our facility) (Nishida 2008; Bouquet et al. 2009; Martí-Solans et al. 2015). We culture the animals at 19 °C in 8 L polycarbonate backers containing 6 L of seawater and 10 grams of activated charcoal (see legend of Fig. 4.8 for details). Animals are maintained in suspension by the rotation of a polyvinyl carbonate paddle connected to an electric motor (5–15 rpm). We add 2 or 3 pearls of 1-hexadecanol to reduce surface tension and therefore avoiding that animals get trapped by the surface of the water. For feeding, a cocktail of three different species of algae (*Chaetoceros calcitrans*, *Isochrysis* sp., *Rhinomonas reticulata*) and a cyanobacterium (*Synechococcus* sp.) is added to the animal cultures every day. Because growing *O. dioica* specimens can trap only foods of a proper size, the amount of each alga and of the cyanobacterium in the cocktail is adjusted according to their cell sizes along the *O. dioica* life cycle. Animal cultures are a source of mature male and female

specimens from which developmentally synchronic embryos can be obtained using several in vitro fertilization protocols (Clarke et al. 2007; Martí-Solans et al. 2015; Mikhaleva et al. 2015). Embryos and larvae at different developmental stages can be thereby easily collected, fixed, and stored for experiments on demand.

4.2.4   O. dioica *Can Be Experimentally Manipulated*

4.2.4.1   Pharmacological Treatments

Pharmacological treatments are a powerful tool to perform functional studies due to its experimental simplicity, and embryos are just transferred to a solution with the drug. The fact that we can easily obtain hundreds of synchronously developing *O. dioica* embryos by in vitro fertilization allows us to perform numerous treatments at different drug concentrations in a fast, simple, and reproducible way. The small size of the *O. dioica* embryos facilitates the diffusion of the drugs to the inner tissues, and their transparency allows the observation of the phenotypic effects of the treatments by DIC microscopy in live developing specimens without the need to perform histological sections. In addition, gene expression responses might be investigated by whole-mount in situ hybridization techniques or by high-density tiling arrays. Pharmacological treatments have been performed, for instance, with developmental morphogens such as all-trans-retinoic acid (RA), which yielded a range of morphological abnormalities, from mild to severe, depending on developmental stage, duration, and concentration of RA exposure (Cañestro and Postlethwait 2007). Treatments with xenobiotic compounds such as the carcinogenic polycyclic aromatic hydrocarbon BaP or with the lipid-lowering agent clofibrate have been also assayed to investigate the transcriptional regulation of the xenobiotic defense mechanisms and to identify the *O. dioica* chemical defensome (Yadetie et al. 2012). Finally, treatments with biotoxins such as *trans*,*trans*-2,4-decadienal, a model for polyunsaturated aldehydes produced during diatom blooms, have been used to analyze the impact on marine food webs of possible future intensification of algal blooms associated with climate change (Torres-Aguila et al. 2018).

4.2.4.2   Techniques for Altering Gene Function by Morpholino, dsRNA, or dsDNA Injections

Different knockdown approaches for altering gene function have been developed in *O. dioica* by injecting morpholinos (Sagane et al. 2010), double-stranded RNA (dsRNA; RNA interference-RNAi) (Omotezako et al. 2013; Mikhaleva et al. 2015), or double-stranded DNA (dsDNA; DNA interference-DNAi) (Omotezako et al. 2015, 2017), and new techniques for genome editing based on CRISPR-Cas9 are currently being developed (Deng et al. 2018). Most techniques rely on the injection of a given molecule into the gonad of premature females, when it is still
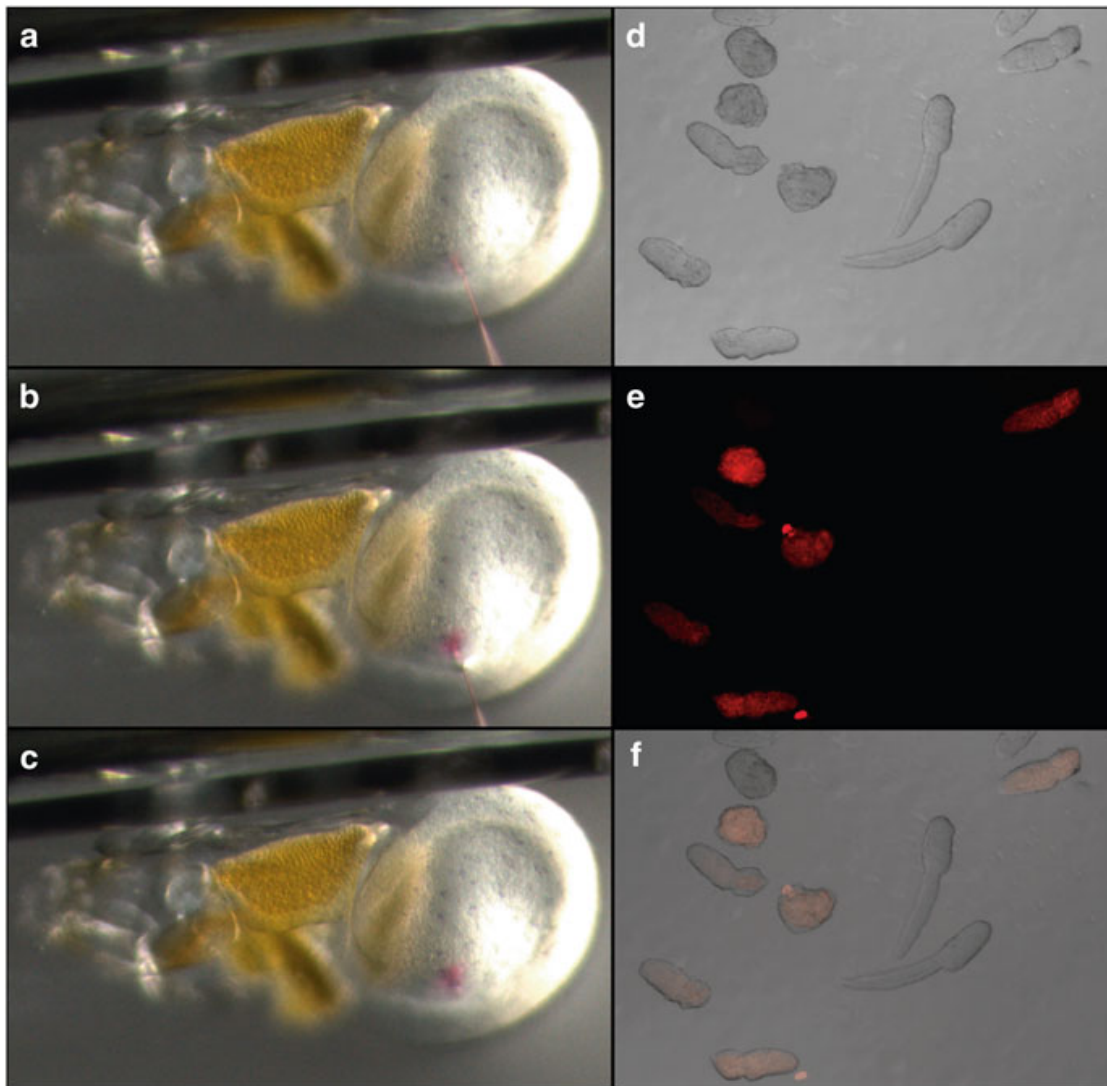
Fig. 4.9 Knocking down by dsDNA injection into the gonad of premature females. A dsDNA against the *brachyury* gene is injected into a premature female gonad (syncytium) (**a–c**), along with phenol red and *pSD64Flife-act-mCherry-mRNA*. Phenol red allows us to visualize the injected solution (red dot in panels **b** and **c**), while mCherry fluorescence highlights the embryos containing mRNA and, therefore, most likely also dsDNA after the cellularization process. In this regard, notice that malformed embryos showing the *brachyury*-dsDNA phenotype are fluorescent (compare panels **d–f**)

a syncytium of meiotic nuclei surrounded by common cytoplasm (Ganot et al. 2007). Then, the injected molecule might spread along the forming oocytes, generating tens of knockdown embryos with a single injection (Fig. 4.9). Alternatively, injections in spawned eggs are also possible (Mikhaleva et al. 2015, 2018; Deng et al. 2018), though they are more difficult and less productive.

Injection of morpholinos, synthetic molecules of approximately 25 nucleotides in length that bind to complementary sequences of RNA blocking translation or altering splicing, was the first knockdown approach in *O. dioica* (Sagane et al. 2010). Morpholinos were used to knock down a cellulose synthase gene necessary to build the house, obtaining different phenotypes related to the production of
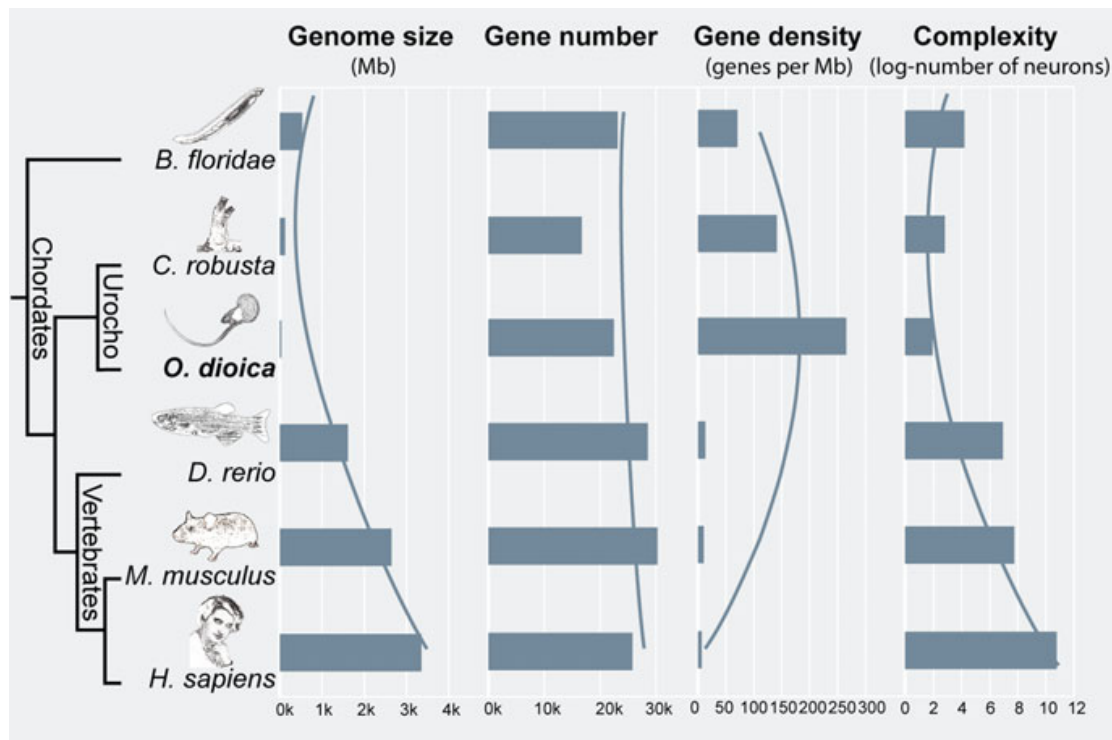
Fig. 4.10 Comparison of genome size, gene number, gene density, and biological complexity among selected chordate species arranged according to their phylogenetic relationships. Bar plots illustrate that *O. dioica* has the smallest chordate genome but has a number of genes rather similar to other chordate species, which implies that *O. dioica* gene density is the highest one. As a rough estimate of biological complexity, we plotted the number of neurons of each organism in logarithmic scale. Lines represent the overall tendency within each plot. Urocho, urochordates

cellulose fibrils (Sagane et al. 2010). In 2013, RNA interference by dsRNA injection was used to investigate the development of the tail and the function of the CNS in tail movement (Omotezako et al. 2013; Mikhaleva et al. 2015). Finally, in 2015, DNA interference by dsDNA injection, a gene silencing technique reported in plants, ciliates, and archaea, was successfully assayed in *O. dioica*, being the first case reported in any metazoan organism (Omotezako et al. 2015). Although the molecular bases for DNAi are unknown, *O. dioica* experiments suggest that dsDNA can induce sequence-specific transcription blocking and mRNA degradation.

4.2.5 *O. dioica Occupies a Privileged Phylogenetic Position Within the Chordate Phylum*

The urochordates (or tunicates), group to which the appendicularian *O. dioica* belongs, are the sister subphylum of the vertebrates, which together with basally branching cephalochordates form the chordate phylum (but see Satoh et al. (2014) for the chordate superphylum hypothesis) (Fig. 4.10). Urochordates are an extremely diverse taxonomic group that appears to have undergone a rapid evolution and

speciation, spreading over many marine habitats, from shallow waters to deep sea (Holland 2016). Urochordates are classically divided into three classes, Appendicularia (larvaceans, ≈70 species including *O. dioica*), Ascidiacea (sea squirts, ≈3000 species), and Thaliacea (salps, doliolids, and pyrosomes, ≈100 species), although it has been recently proposed that Thaliacea species are nested within the Ascidiacea class (Delsuc et al. 2018; Kocot et al. 2018). In any case, Appendicularians—also known as Larvaceans because they maintain the larval morphology, including the chordate synapomorphies, throughout the entirely life cycle—appear to be the sister clade of all other urochordates, representing a branch of the subphylum that split from all other urochordates 450 million years ago (Delsuc et al. 2018). Appendicularians are divided into three families. The Fritillariidae family that includes 3 genera and 30 species, the Kowalevskiidae family, the smallest one with just 1 genus and 2 species, and the Oikopleuridae family, the most diverse group with 11 genera and 37 species (Bone 1998).

O. dioica occupies a privileged phylogenetic position within urochordates for evo-devo studies because it can be morphologically, functionally, genetically, and genomically compared with many other urochordate species (more than 10 ascidians have been totally or partially sequenced; http://www.aniseed.cnrs.fr/; http://octopus.obs-vlfr.fr/public/botryllus/blastbotryllus.php), including well-studied ascidian species of the *Ciona* or *Halocynthia* genera (Corbo et al. 2001; Dehal et al. 2002; Nishida 2002; Cañestro et al. 2003; Satoh 2003; Satoh et al. 2003; Brozovic et al. 2016, 2017). In addition, from a more general perspective, comparisons of *O. dioica* and ascidians with diverse vertebrate species (there are more than 100 vertebrate genome projects) are relevant for detecting subphylum-specific traits, which might be classified as evolutionary innovations or losses depending on their presence in the external cephalochordate subphylum (there are three *Branchiostoma* genome projects and one *Asymmetron* RNA-Seq project).

4.2.6 O. dioica *Has a Small and Compacted Genome*

With only 70 Mb, *O. dioica* genome is the smallest chordate genome and one of the smallest genomes in metazoans (Seo et al. 2001; Denoeud et al. 2010; Danks et al. 2013), only surpassed by the small genomes of some parasite animals (Burke et al. 2015; Chang et al. 2015; Mikhailov et al. 2016) (Table 4.1) (Fig. 4.10). The small size of the *O. dioica* genome is the result of a process of compaction mostly due to three factors: (1) a reduction of the length of intergenic regions, partly because of numerous operons; (2) a reduction of the abundance of transposable elements; and (3) the reduction of the size of introns (Denoeud et al. 2010; Berna and Alvarez-Valin 2014; reviewed in Chavali et al. 2011). As a result of the genome compaction, the gene density in *O. dioica* genome (1 gene per 3.9 kb) has become very high, 3 times higher than in ascidian *C. robusta*, 6 times higher than in cephalochordate amphioxus, and 35 times higher than in human genome (Table 4.1) (Fig. 4.10).

Table 4.1 Main features of *O. dioica* genes and genome (Volff et al. 2004; Denoeud et al. 2010; Chalopin et al. 2015)

Genome	
Size	70 Mb
Size of intergenic regions	53% of intergenic distances <1 kb
Population mutation rate	$\theta = 4N_e\mu = 0.0220$
Conserved synteny	Negligible
Diversity of transposable elements	4 class I retroelements + 3 class II DNA transposons
Abundance of transposable elements	10–20% genome (at most few hundreds of each TE)
Genes	
Number	18,020 predicted genes
Density	1 gene per 3.9 kb (257 genes per Mb)
Genes in operons	27% (1800 operons)
Number of introns per Gene	4.1
Intron size	<50 nt in 62% of the genes; >1 kb in 2.4% of the genes
Intron position	76% intron positions specie-specific
Noncanonical splicing sites	In 12% introns (9% GA-AG, 2.7% GC-AG, 0.7% GG-AG)

4.2.6.1　Intergenic Regions and Operons

O. dioica genes are densely packed, with 53% of their intergenic sequences smaller than 1 kb (Table 4.1). Contributing to the intragenic reduction, many *O. dioica* genes are organized in polycistronic transcription units (operons). The majority of operons are bicistronic (60%), but a substantial proportion contains three or more genes, and some have up to 11 genes. Operon organization reduces the intergenic space because it reduces the DNA segments containing transcription initiation and regulation signals. The fact that 27% of *O. dioica* genes are predicted to be organized in around 1800 operons (Denoeud et al. 2010), together with the observation that most genes have relatively small intergenic spaces, may favor the genome compaction of this species (Table 4.1).

4.2.6.2　Transposable Elements

Transposable elements (TEs or transposons) are main components of eukaryote genomes that can replicate and change to new locations. *O. dioica* genome is relatively poor in TEs, in terms of quantity and diversity, and most of the so-called pan-animal transposon families, many of them even present in *C. robusta* (Cañestro and Albalat 2012), are absent (Denoeud et al. 2010). *O. dioica* genome contains TEs of only six superfamilies—*Ty3/Gypsy*, *DIRS1*, and *Penelope*-like retrotransposons (Volff et al. 2004) and *Tc-Mariner*, *PyggyBac*, and *Maverik* DNA transposons (Chalopin et al. 2015)—plus a new family of non-LTR retrotransposon named *Odin* (Volff et al. 2004) (Table 4.1). Although the copy number of each TE is low

(at most few hundreds), LTR retrotransposons appear to be the most abundant (>60%), while non-LTR retrotransposons and DNA transposons represent less than 20% each. The extent of TEs on the *O. dioica* genome size has been estimated about 10–20% (Chalopin et al. 2015). In summary, a massive purge of pan-animal TEs has occurred in *O. dioica* genome probably associated with an intense process of genome compaction. Some TEs show a low level of sequence corruption, suggesting a rather recent activity, but their low copy number and uneven genome distribution indicate that TE expansion in *O. dioica* is under tight genetic control (Denoeud et al. 2010).

4.2.6.3   Introns

Another factor contributing to *O. dioica* genome compaction is the small size of the introns of the majority of the genes (Table 4.1). Despite *O. dioica* having a typical number of introns per gene (4.1) as in other vertebrates, it has reduced the size of the introns to less than 50 nucleotides in 62% of the genes, and only 2.4% of the introns being larger than 1 kb (Seo et al. 2001; Denoeud et al. 2010). *O. dioica* shows a high intron turnover: 76% of the introns are in positions unique to *O. dioica* (newly acquired introns) and only 17% are in ancestral positions (7% remain unclassified) (Denoeud et al. 2010) (Table 4.1). Several mechanisms have been proposed for the gain and loss of introns in *O. dioica* (Denoeud et al. 2010), but since newly acquired introns tend to be shorter than old introns, intron turnover might have favored genome compaction. Surprisingly, canonical splicing signal (GT-AG intron boundaries) is not observed in about 12% of introns (1% in other species), which mostly (9%) have noncanonical GA-AG (GC-AG, 2.7%, or GG-AG, 0.7%) sequences (Edvardsen et al. 2004; Denoeud et al. 2010). Because *O. dioica* lacks the minor spliceosome and only has one type of each spliceosomal component (Denoeud et al. 2010), a single major spliceosome, permissive in terms of splicing signals, would take care of all splicing processes. It has been proposed that such permissive splicing could favor the intron turnover (Denoeud et al. 2010) and, hence, the compaction of the *O. dioica* genome.

4.3   Evolution of *O. dioica* Genes

The *O. dioica* genome project did not only reveal that this species has an extremely compacted genome with small intergenic regions, tiny introns, and few TEs, but that its genes have suffered many evolutionary changes affecting the conservation of the sequence, organization, and number of genes.

4.3.1 Gene Sequence Evolution: High Evolutionary Rates

The sequencing of the genome confirmed that *O. dioica* is a very fast evolver and its genes always show surprisingly long branches in phylogenetic tree reconstructions (Edvardsen et al. 2005; Denoeud et al. 2010). This high rate of sequence evolution appears to be the result of two independent evolutionary forces acting at different levels and with different strength.

First, affecting at a global genome scale with a moderate strength, we found that most genes show a high evolutionary rate (Table 4.1) as a result of a combination of a high population mutation rate and a reduced negative (purifying) selection. An estimate of *O. dioica* population mutation rate ($\theta = 4N_e$, $\mu = 0.0220$) shows it is high, which is consistent with a large effective population size (N_e) and/or a high mutation rate per generation (μ) (Denoeud et al. 2010). In addition, the short generation time of this species implies that the effective mutation rate per year is substantially increased. A reduced negative selection for *O. dioica* has been proposed based on the homogeneity of its amino acid rates that is a symptom of relaxation of selective—structural and/or functional—constraints (Berna et al. 2012; Berna and Alvarez-Valin 2014). Relaxation of selective constraints associated with rapid rates of sequence evolution has been actually proposed for other fast-evolving animal species such as *Ciona robusta* and *Caenorhabditis elegans* (Holland and Gibson-Brown 2003).

Second, some particular genes have even a higher evolutionary rate due to positive selection that contributes to the adaptation to new environments or to new challenges (Berna and Alvarez-Valin 2014). Many of these genes with this higher evolutionary rate have been described to be involved in regulatory and developmental functions (Berna et al. 2012; Berna and Alvarez-Valin 2014).

4.3.2 Evolution of Gene and Genome Organization

4.3.2.1 Negligible Synteny Conservation

Conserved synteny describes the colocalization of homologous genes on homologous chromosomic regions of different species. *O. dioica* genome appears to have suffered numerous chromosomal rearrangements during evolution, and thereby, it lacks any chromosomal synteny conservation with other animal genomes (Denoeud et al. 2010). Conservation of local synteny is also almost negligible since local gene order is indistinguishable from random for distances smaller than 30 genes, and a low level of conserved synteny is only detectable at a wider distance span (Denoeud et al. 2010; Irimia et al. 2012). These observations suggest that constraints that maintain gene order in metazoans may actually be relaxed in *O. dioica*.

4.3.2.2   Disintegration of *Hox* Cluster

Another significant feature of *O. dioica* genome is the disintegration of the cluster of *Hox* genes, which is generally well conserved in all bilaterians. *Hox* genes are a subset of homeobox genes involved in establishing morphological identities along the anterior–posterior axis, and although the biological significance of *Hox* clustering remains unclear, functional and structural explanations leading to a spatiotemporal coordinated transcriptional regulation of *Hox* genes have been proposed (Kmita and Duboule 2003). In many species, the position of the *Hox* genes in the cluster correlates with their temporal and spatial sequential expression along the anterior–posterior axis (reviewed in Duboule 2007). *O. dioica* does not have a *Hox* cluster since its complement of nine *Hox* genes—three anterior *Hox* genes (*Hox1*, *Hox2*, and *Hox4*) and six posterior genes (*Hox9A*, *Hox9B*, *Hox10*, *Hox11*, *Hox12*, and *Hox13*)—is totally dispersed in its genome (Seo et al. 2004). The disorganization of the *Hox* cluster in *O. dioica* has been related to the lineage-driven mode of development of *O. dioica*, in which *Hox* genes would contribute to tissue specification with separated domains of *Hox* expressions, rather than to axial patterning with overlapping *Hox* expressions (Seo et al. 2004).

4.3.2.3   Evolution of Operon Organization

A third important genomic feature in the organization of *O. dioica* genes is their grouping in operons (Table 4.1). The operons do not only contribute to the compaction of the genome (see above) but might also serve to group genes that have to be either efficiently coregulated—repressing or activating them as a group—or ubiquitously expressed with a low degree of transcriptional regulation (Blumenthal 2004). *O. dioica* has around 1800 operons transcribed as polycistronic pre-mRNAs, most of them processed to mature monocistronic mRNAs via spliced-leader RNA (SL RNA) *trans* splicing (Ganot et al. 2004; Denoeud et al. 2010). Functional annotation of the gene set in operons shows that they are significantly enriched for genes involved in housekeeping functions or general metabolic processes such as RNA, protein, DNA, lipid, and carbohydrate processing and transport. Genes involved in developmental processes such as morphogenesis and organogenesis are, in contrast, significantly underrepresented in the operon gene set (Denoeud et al. 2010).

4.3.3   Gene Number Evolution

Despite its reduced size, *O. dioica* genome contains 18,020 predicted genes, a similar number to other urochordates (e.g., *C. robusta* ≈15,300 genes) and only slightly below other chordates such as the cephalochordate *Branchiostoma floridae*

($\approx$22,000 genes) or the vertebrate *Fugu rubripes* ($\approx$18,300 genes) (Table 4.1) (Fig. 4.10). The current number of genes is the result of the balance between gene gains and losses impacting a certain set of ancestral genes, and because *O. dioica* appears to be prone to lose genes (see next section), it has to be also prone to gain them.

Gene duplications have been proposed to be a major driving force for gene gains (Ohno 1970; Cañestro et al. 2013), and *O. dioica* seems to have retained many lineage-specific duplicates. For instance, among homeobox genes, *O. dioica* shows a high incidence of retention of lineage-specific duplicates of genes such as *Irx*, *Not*, and *Pax3/7* (Edvardsen et al. 2005). Additional examples of such duplicates are found in g-type lysozyme genes (Nilsen et al. 2003), caspase genes (Weill et al. 2005), metallothionein genes (Calatayud et al. 2018), notochord *Noto15* and *Noto9* genes (Kugler et al. 2011), *RdhE2* and *Cco* genes (Martí-Solans et al. 2016), and actin as well as other muscle structural genes (Almazán et al. 2019; Inoue and Satoh 2018), among many others predicted from the genome project analysis (Denoeud et al. 2010). Lineage-specific duplicates might contribute to both general evolutionary adaptations of the organisms or to innovations associated to the unique biology of the lineage. For example, homeobox genes expressed—and possibly patterning—in the oikoplastic epithelium, which represents a significant novelty of appendicularians for secreting the mucous house, mostly belong to the duplicated groups (Denoeud et al. 2010; Mikhaleva et al. 2018), while gene families with cell adhesion roles are overrepresented in *O. dioica*, most likely because of the extensive and assorted interactions required for building the house (Chavali et al. 2011).

4.4 Evolution by Gene Loss: *O. dioica* as a Model System for Evo-Devo Studies

Urochordate genomes (e.g., *C. robusta*) appear to have a "liberal" evolutionary pattern of gene loss (Dehal et al. 2002; Holland and Gibson-Brown 2003; Hughes and Friedman 2005; Somorjai et al. 2018), which contrasts with the "conservative" pattern of cephalochordates and vertebrates (Somorjai et al. 2018). *O. dioica* appears to have pushed this urochordate trend to its limits (Ferrier 2011) by having lost many genes or entire genetic pathways (Table 4.2). In this section, we review these losses grouped into three categories: Sect. 4.4.1, losses of genes involved in general characteristics of gene/genomic structure and expression, Sect. 4.4.2, losses of genes related to particular cellular and physiological functions, and Sect. 4.4.3, losses of genes essential for embryonic development in chordates.

Table 4.2 Absent genes[a] in *O. dioica* genome

Category	Genes	References
cNHEJ repair system	*Xrcc5* (*Ku80*), *Xrcc6* (*Ku70*), *Xrcc4*, *Lig4*, *NHEJ1* (*Xlf*), *DNA-PKc*, *Dclre1c*	Denoeud et al. (2010)
Epigenetic machinery	*Dnmt1*, *Dnmt3*, *Mbd4/MeCP2*, *Gcn5/PcaF*, *Hat1*, *Mll1*, *Ash1*, *Rnf1*, *Suz12*, *Pcgf*, *Scmh1*, *Kdm2b*	Albalat et al. (2012), Navratilova et al. (2017)
Spliceosome machinery	snRNA U11, snRNA U12, snRNA U4atac, snRNA U6atac	Denoeud et al. (2010)
Caspase family	*Csp1/4/5*, *Csp6*, *Csp2/9*, *Csp8/10*	Weill et al. (2005)
Immune system	*NLRs*, *RLHs*, *MyD88-like*, *Sarm1-like*, *Tirap-like*, *Ticamp2-like*	Denoeud et al. (2010)
Defensome system	*AhR*, *AhRR*, *Nr1C* (*Ppar*), CYP1 genes	Yadetie et al. (2012)
Peroxins	*Pex1*, *Pex2*, *Pex3*, *Pex5*, *Pex6*, *Pex7*, *Pex10*, *Pex11*, *Pex12*, *Pex13*, *Pex14*, *Pex16*, *Pex19*, *Pex26*	Zarsky and Tachezy (2015)
Retinoic acid signaling	*Rdh10*, *Rdh16*, *Bco1*, *Aldh1a*, *Cyp26*, *RAR*, *PPAR*	Cañestro et al. (2006), Cañestro and Postlethwait (2007), Martí-Solans et al. (2016)
Homeobox genes	*Hox3*, *Hox5*, *Hox6*, *Hox7*, *Hox8*, *Gbx*, *Nk3*, *Nk6*, *TGIF*, *POU VI*, *Lhx6/7*, *Vax*, *Cux*, *SATB*, *ZFH1*, *Sax*, *Xlox*, *Mox*, *Hlx*, *Bsh*, *Chox10*, *Otp*, *Prx*, *Goosecoid*, *Prox*, *Tlx*	Seo et al. (2004), Edvardsen et al. (2005)
Sox genes	*SoxC*, *SoxE*, *SoxF*, *SoxH*	Heenan et al. (2016)
miRNA	miR-9, miR-29, miR-33, miR-34, miR-96, miR-126, miR-133, miR-135, miR-153, miR-182, miR-183, miR-184, miR-196, miR-200, miR-216, miR-217, miR-218, miR-367	Wang et al. (2017)
Notochord genes	*Entactin*, *Fibrinogen-like*, *Multidom*, *Myomegalin*, *Noto1*, Noto2, *Noto3*, *Noto4*, *Noto5*, *Noto6*, *Noto7*, *Noto8*, *Noto11*, *Noto12*, *Noto13*, *Noto14*, *Noto16*, *Perlecan*, *Ptp*, *Slc*, *Swipi*, *Tropomyosin-like*, *Tune*, *Ube2*	Kugler et al. (2011)
General repair system	*Polb*, *Apex2*, *Lig3*, *Msh3*, *Atm*, *Chek2*, *Aptx*, *Nbn*, *Rad52*	Denoeud et al. (2010)
Apoptotic genes	*Bax*, *Bak*, *BCL-X$_L$*	Robinson et al. (2012)
Cyclins and CDK	*Cyclin J*, *Cyclin G*, *Cyclin F*, *Cdk14/15*	Campsteijn et al. (2012)
Rab GTPases	*Rab4*, *Rab7L1*, *Rab9*, *Rab15*, *Rab19/43*, *Rab20*, *Rab21*, *Rab22*, *Rab24*, *Rab26/37*, *Rab28*, *Rab30*, *Rab32LO*, *Rab40*, *Ift27*, *Rasef*, *EFcab44/Rab44*, *RabX1*, *RabX4*, *RabX6*	Coppola et al. (2019)

[a]Genes do not found in *O. dioica*, and, thereby, likely lost during the evolution of the *O. dioica* lineage

4.4.1 Loss of Genes Involved in Gene/Genomic Structure and Expression

4.4.1.1 Loss of Non-homologous End Joining Repair Genes

All organisms have the ability to repair the double-strand breaks (DSBs) that normally happen in DNA due to numerous external and internal factors. One fundamental mechanism present in all eukaryotes to repair these breaks is the canonical non-homologous end joining (cNHEJ) repair system. Unexpectedly, *O. dioica* (Denoeud et al. 2010) and other six species of appendicularians (Deng et al. 2018; Ferrier and Sogabe 2018) do not have the cNHEJ machinery (Table 4.2), meaning this repair system became dispensable during the evolution of this chordate lineage. In fact, *O. dioica* seems to use the alternative NHEJ (aNHEJ) system (a.k.a. alternative end joining, aEJ, Pannunzio et al. 2018) as a compensatory system that overcomes DSBs, although the existence of another so far undescribed pathway cannot be excluded (Deng et al. 2018). Because the aNHEJ system seems to be less faithful than cNHEJ (it often involves deletion of some intervening nucleotides and commonly leads to chromosome rearrangements, for example, translocations, (Deriano and Roth 2013)), the loss of the cNHEJ system might have had important consequences in the genome architecture. It may have contributed, for instance, to the compaction of the *O. dioica* genome by either favoring the deletions caused by the aNHEJ repair system or by restricting replication of autonomous retrotransposons that uses the cNHEJ system for their propagation (Deng et al. 2018). In addition, this loss may have also contributed to the reorganization of the genome through accumulated rearrangements, leading to negligible synteny conservation (Deng et al. 2018) that may be associated with modifications of the mechanisms of gene regulation, changing from long-range mechanisms acting on topologically associated genomic blocks that include several genes to short-range and gene-specific systems for gene regulation (Cañestro et al. 2007; Ferrier and Sogabe 2018).

4.4.1.2 Loss of Genes of the Epigenetic Machinery

Histone modifications and DNA methylation are epigenetic marks mainly associated with regulation of gene expression, replication, DNA repair, recombination, chromosome segregation, and other meiotic and mitotic processes. Histone proteins, responsible of packing DNA into nucleosomes, provide multiple sites for posttranslational modifications by evolutionarily conserved histone modifiers that establish the so-called histone code. Among the proteins that modify histones, *O. dioica* has lost several genes for histone acetyltransferases such as *Gcn5/Pcaf* and *Hat1* (Navratilova et al. 2017) (Table 4.2). In addition, homologs of core components of the canonical Polycomb complexes (*Rnf1*, *Suz12*, *Pcgf*, and *Scmh1*) and genes for several proteins of the trithorax group (*Mll1*, *Ash1*, and *Kdm2b*) have neither been

found in *O. dioica*'s genome (Navratilova et al. 2017). Regarding DNA methylation, *O. dioica* has lost the two main DNA methyltransferases *Dnmt1* and *Dnmt3*, and one of the two *Mbd* genes, *Mbd4/MeCP2* (Cañestro et al. 2007; Albalat et al. 2012) (Table 4.2). Overall, it is through that the loss of components of the epigenetic machinery might have favored changes in the genome architecture and modifications of the mechanisms of gene regulation, leading to compacted regulatory spaces, reduced chromatin state domain widths, evolution of operons, and loss of synteny and dispersion of the Hox cluster (Cañestro et al. 2007; Navratilova et al. 2017).

4.4.1.3   Loss of Minor Spliceosome Genes

While the major spliceosome containing the snRNAs U1, U2, U4, U5, and U6 removes introns with canonical GT-AG boundaries, the minor spliceosome, which contains the snRNAs U11, U12, U4atac, U5, and U6atac, mainly acts on noncanonical AT-AC boundaries (Patel and Steitz 2003; Sheth et al. 2006). Both major and minor spliceosomes are present across most eukaryotic lineages tracing them back to the origin of eukaryote evolution. Although *O. dioica* has many noncanonical introns (see above), no AT-AC introns have been detected in cDNA resources (Denoeud et al. 2010), which may explain that none of the minor splicesomal snRNAs are found in *O. dioica* (Table 4.2). It seems therefore that in the absence of AT-AC introns, the components of the minor spliceosome became dispensable and lost during *O. dioica* evolution and at the same time that the major spliceosome became permissive and able to remove all—canonical and noncanonical—*O. dioica* introns, although with different efficiency (Denoeud et al. 2010). Whether these differences in splicing efficiency could reflect specific usage of noncanonical introns for gene expression regulation and what evolutionary impact these changes had needs further investigation.

4.4.2   Loss of Genes for Cellular and Physiological Functions

4.4.2.1   Loss of Genes of the Caspase Family

Caspases are a family of protease enzymes that play an important role not only in apoptosis but also in maturation of different immunity system proteins and in the control of the proliferation and differentiation of specific cell types (Weill et al. 2005). Caspases are present in all kingdoms (Uren et al. 2000), but their number has a taxon-dependent diversity ranging from 3 members in *C. elegans* to 10–13 in vertebrates and 17 in *C. robusta* (Weill et al. 2005). *O. dioica* has only three caspase genes deriving from a single founder distantly related to *Caspase 3/7* (Weill et al. 2005) (Table 4.2). *O. dioica* seems, therefore, to have lost all of the components of the caspase family except one. The reduced complexity of the caspase family might be associated with a low cell number at the adult stage, the absence of a major

metamorphosis event, and a minimized immune system in this species (Weill et al. 2005).

4.4.2.2 Loss of Genes of the Immune System

As a marine organism, *O. dioica* should have an innate immune system prepared to fight against viruses and bacteria present in high amounts in seawater, but its short life span along with the absence of hemolymph cells—hemocytes or macrophages—points to a rapid immune system based on transcriptional upregulation response or on constitutively expressed effectors, rather than to a more slow system reliant on cell proliferation triggered by immune receptors (Denoeud et al. 2010). In agreement, *O. dioica* has lost almost all genes with domains corresponding to typical immune receptors or immune effectors and lacks homologous genes to interleukins or cytokines involved in the immunity response of more complex chordates (Denoeud et al. 2010) (Table 4.2). Interestingly, it has been proposed that a simplified immune system may be compensated by the antibacterial function of some oikosins with phospholipase A2 domains, which hydrolyze glycerophospholipids present in bacterial cell walls (Hosp et al. 2012). In summary, *O. dioica* seems to have a highly derived and simplified strategy of defense that together with its high fertility, short life cycle, and high polymorphism contributes to the survival of the *O. dioica* populations (Denoeud et al. 2010).

4.4.2.3 Loss of *Cyp* Genes for Xenobiotic Defensome Systems

The first step for elimination or inactivation of xenobiotic compounds often involves the oxidative modification of the toxic chemicals, mostly performed by enzymes of the cytochrome P450 (CYP) family, some of which are also required for embryonic development (Goldstone et al. 2006). Due to its variable functions, Cyp enzymes are quite abundant in animal genomes—up to 120 *Cyp* genes in sea urchin (Goldstone et al. 2006), 94 in zebrafish (Goldstone et al. 2010), or 236 in amphioxus (Nelson et al. 2013). *O. dioica*, however, has lost many *Cyp* genes, and with only 23 *Cyp* genes, this species has the smallest *Cyp* repertoire among sequenced metazoan genomes (Yadetie et al. 2012). *O. dioica* lacks, for instance, the CYP1 family genes, which play a central role in the metabolism of environmental toxicants (Table 4.2). *O. dioica* also lacks the aryl hydrocarbon receptor (*AhR*) gene (and its repressor *AhRR*) (Yadetie et al. 2012), which is the transcriptional regulator of CYP1 family genes and a major xenobiotic sensing receptor activated by pollutants. Overall, these data suggest the absence of an AhR-mediated xenobiotic biotransformation signaling pathway in *O. dioica* and suggest the evolution of alternative mechanisms of response to environmental xenobiotic compounds (Yadetie et al. 2012).

4.4.2.4 Loss of Peroxin Genes and Absence of Peroxisomes

Peroxisomes are single membrane-bound organelles with important functions in detoxification of reactive oxygen species, long-chain fatty acid beta-oxidation, plasmalogen synthesis, amino acid degradation, and purine metabolism. Peroxisome biosynthesis and protein import is mediated by a group of 13 highly conserved eukaryotic proteins called peroxins (Gabaldon et al. 2006; Schluter et al. 2006; Zarsky and Tachezy 2015). Peroxisomes are ubiquitous in eukaryotes, and only some groups of anaerobic protist and parasitic helminths lack peroxisomes (Schluter et al. 2006; Gabaldon and Capella-Gutierrez 2010). It is therefore extraordinary that *O. dioica* has lost all peroxin genes (Table 4.2), becoming the only known aerobic non-parasitic organism that does not have peroxisomes (Zarsky and Tachezy 2015). The evolutionary and physiological conditions in which *O. dioica* was able to lose the peroxisomes are still a matter of debate as this organism has high oxygen consumption and oxidizes fatty acids for ATP synthesis, unlike the anaerobic parasitic lineages missing these organelles (Zarsky and Tachezy 2015). It has been proposed that the loss of peroxisomes might be evolutionary adaptive, since it has been associated with reduced genomes and traits of r-reproductive strategies such as high fecundity, early maturation, and simplified ontogenesis, which *O. dioica* shares with parasite organisms as well as with some selective advantages by rendering organisms resistant to xenobiotics that become activated in the peroxisomal lumen by redox reactions (Zarsky and Tachezy 2015).

4.4.3 Loss of Genes Essential for Embryonic Development in Chordates

4.4.3.1 Loss of the RA Genetic Machinery

All-*trans*-RA is a vitamin A-derived compound that acts as a crucial signaling system involved in the differentiation and outgrowth of neurons in many metazoans (reviewed in Albalat 2009), which was adopted for Hox-controlled anterior–posterior patterning in the chordate phylum (Handberg-Thorsager et al. 2018). Despite the crucial role of RA in chordate axial patterning, it has been shown that *O. dioica* has lost most of the genes for RA production, degradation, and signaling (i.e., the RA genetic machinery, including *Rdh10*, *Aldh1a*, *Cyp26*, and *RAR* genes) (Cañestro et al. 2006, 2007; Cañestro and Postlethwait 2007; Martí-Solans et al. 2016) (Table 4.2) (Fig. 4.11). In the absence of mutational robustness (i.e., alternative pathways) capable of compensating them, these gene losses were accompanied by the loss of the RA signaling in this species (Martí-Solans et al. 2016).

The loss of the RA genetic machinery in *O. dioica* probably took place in the context of regressive evolution associated with the disintegration of the Hox cluster that, as mentioned in Sect. 4.3.2, has been related with the shift to a determinative mode of development. This shift would have released the restrictions to maintain the
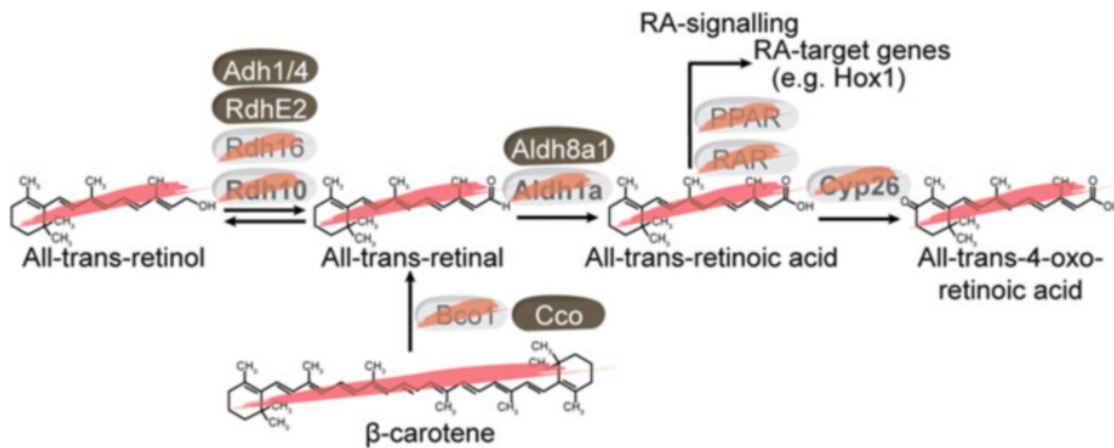
Fig. 4.11 The loss of the retinoic acid genetic machinery in *O. dioica* exemplifies how functionally linked genes are co-eliminated during evolution. In the absence of mutational robustness capable of compensating the loss, the co-elimination was accompanied by the loss of the RA signaling. *O. dioica* has lost genes for the Rdh10, Rdh16, Bco1, Aldh1a, and Cyp26 enzymes, as well as genes for the RAR and PPAR nuclear receptors (strikethrough in lighter boxes). The surviving genes (i.e., *Adh1/4*, *RdhE2*, *Aldh8a1*, and *Cco* in dark boxes) (Cañestro et al. 2010; Martí-Solans et al. 2016) do not constitute an alternative pathway for RA synthesis because neither RA nor RA precursors have been detected at concentrations that are likely to play any role in developmental or physiological processes of this species (strikethrough)

integrity of the Hox cluster (Seo et al. 2004) and led extracellular signals such as RA that establish embryonic coordinates and regulate the expression of *Hox* genes by gradually increasing the portion of the cluster exposed to transcription machinery over time to become dispensable and eventually lost (Cañestro and Postlethwait 2007; Cañestro et al. 2007).

4.4.3.2   Loss of Homeobox Gene Families

Homeobox genes encode transcription factors involved in many developmental processes in eukaryotes. Homeobox genes are classified into 11 classes and over 100 gene families/groups (Holland et al. 2007), some of them vertebrate innovations. *O. dioica* has lost 36 homeobox groups, 12 of which probably lost during the early urochordate evolution and 14 specifically lost in the *O. dioica* lineage (Edvardsen et al. 2005) (Table 4.2). Noteworthy, among the homeobox families, *O. dioica* has lost the anterior *Hox3* and all the central *Hox* genes of the cluster (Seo et al. 2004). In parallel with the losses, homeobox genes have also duplicated during the evolution of *O. dioica* lineage (Edvardsen et al. 2005). Interpreting these gains and losses in terms of the evolution of developmental mechanisms remains difficult until the function of *O. dioica* homeobox genes is revealed. It has been proposed that family losses may be the result of a relaxed selective pressure to conserve the full set of homeobox families in the *O. dioica* (or in the urochordate) lineage due to a simplification of its body plan in comparison with that of the chordate ancestor (Edvardsen et al. 2005). On the other hand, the preservation of duplicates in some

families might be the result of a re-diversification of homeobox genes after major group losses (Edvardsen et al. 2005).

4.4.3.3   Loss of *Sox* Genes

Sox (Sry-type HMG box) genes are a family of transcription factors defined by a conserved sequence called the high-mobility group (HMG) box. *Sox* genes are involved in a number of essential functions during embryonic development, including sex determination and neuronal development (Wegner 2010). Ten different Sox groups have been described (A to J), some of them (B, C, D, E, F) conserved in almost all metazoans, and some groups being specific of certain lineages (A, G, and I of vertebrates, H of chordates, and J of nematodes) (Bowles et al. 2000; Heenan et al. 2016). *O. dioica*, however, has only four *Sox* genes, two of them belonging to the *SoxD* group and the other two to the *SoxB* group (Heenan et al. 2016; Torres-Aguila et al. 2018). These results imply that *O. dioica* has lost three ancient metazoan Sox groups (C, E, F) and the chordate Sox group (H) (Heenan et al. 2016) (Table 4.2). As in other examples reviewed above, the evolutionary impact of the gains and losses in the Sox family on the developmental mechanisms remains difficult to interpret since the expression of only two SoxB genes has been described (Torres-Aguila et al. 2018).

4.4.3.4   Loss of miRNAs

MicroRNAs (miRNAs) are noncoding RNAs of about 22 nucleotides involved in the regulation of different biological processes, including embryo development, cell differentiation, and growth. miRNA innovation has been correlated with increased developmental complexity during animal evolution (Hertel et al. 2006). Analysis of the miRNA repertoire of *O. dioica* has shown that this species has lost (or derived to the point they cannot be recognized anymore) at least eighteen highly conserved bilaterian miRNA families (Fu et al. 2008; Wang et al. 2017) (Table 4.2). On the other hand, at least 29 new miRNA families would have appeared in appendicularians (Fu et al. 2008), suggesting a profound reorganization of the miRNA repertoire due to recurrent events of gene losses and gains during *O. dioica* evolution. This scenario is consistent with the general notion that animal miRNAs are poorly conserved between distant taxa (Wang et al. 2017) and, therefore, that changes in miRNA repertoires have been important in shaping animal evolution (Fu et al. 2008). It is thought that the reorganization of *O. dioica* miRNA repertoire might have impacted on the temporal robustness of the rapid developmental program of this urochordate lineage (Fu et al. 2008), which would be in agreement with the idea that modifications in the miRNA repertoires have been important in adapting radically different life-history strategies from a common larval body plan (Fu et al. 2008).

4.4.3.5 Loss of Notochord Genes

The notochord, one of defining features of the chordate phylum, is a stiff rod of tissue located ventral to the neural tube that provides a rigid, but still bendy, structure for muscle attachment, as well as an important source of embryonic developmental signals (Stemple 2005). Actually, developing chordate embryos require a notochord as an organizer for secreting signals that pattern several organs such as the somites, heart, or pancreas. Comparison of a set of notochord genes known to be targets of the notochord-specific Brachyury transcription factor revealed that from 50 notochord genes of *C. robusta* used as a reference, 24 are absent in *O. dioica* (likely lost, e.g., *Noto2*), and 15 are not expressed (8 genes, e.g., *Asak*) or expressed at low levels (7 genes, e.g., *Noto10*) in the notochord (Kugler et al. 2011) (Table 4.2). Taken together, these results suggest a considerable divergence in the genetic toolkit for notochord development in *O. dioica*, which would have been impacted by events of gene loss or by the loss of the notochord function of some genes. Considering the morphological similarities between urochordate notochords, the divergence in the genetic toolkits used to develop them is surprising.

4.4.4 O. dioica, *Gene Loss, and the Inverse Paradox*

The original paradox in the field of evo-devo arose by the discovery that similar genetic toolkits were able to build a wide variety of morphologies in disparate animals, implying that the same genes (genetic unity) were used to build such different forms (phenotypic diversity) (Jacob 1977). The discovery of the pervasiveness of gene loss along evolution (Albalat and Cañestro 2016) affecting relevant developmental genes led to the formulation of the so-called inverse paradox of evo-devo. This hypothesis proposes that organisms might develop fundamentally similar morphologies (phenotypic unity) despite important differences in their genetic toolkits (genetic diversity) (Cañestro et al. 2007) (Fig. 4.12). Because differences between the genetic toolkits are often a consequence of the loss of some of their crucial genes, the study of *O. dioica* has become fundamental to build the new conceptual framework as this species has been able to maintain a phylotypic chordate-body plan after having lost many genes thought to be crucial for the archetypal chordate development (see Sect. 4.4.3). In this regard, the characterization of the *O. dioica* "lossosome" (the complete catalogue of gene losses in a phylogenetic context; Cañestro and Roncalli 2018) and the analysis of its functional consequences should provide the framework to investigate how gene loss might have been an important evolutionary force for generating differences in the developmental genetic toolkits of chordates and, thereby, to understand the evolution of the mechanisms of development of our own phylum.
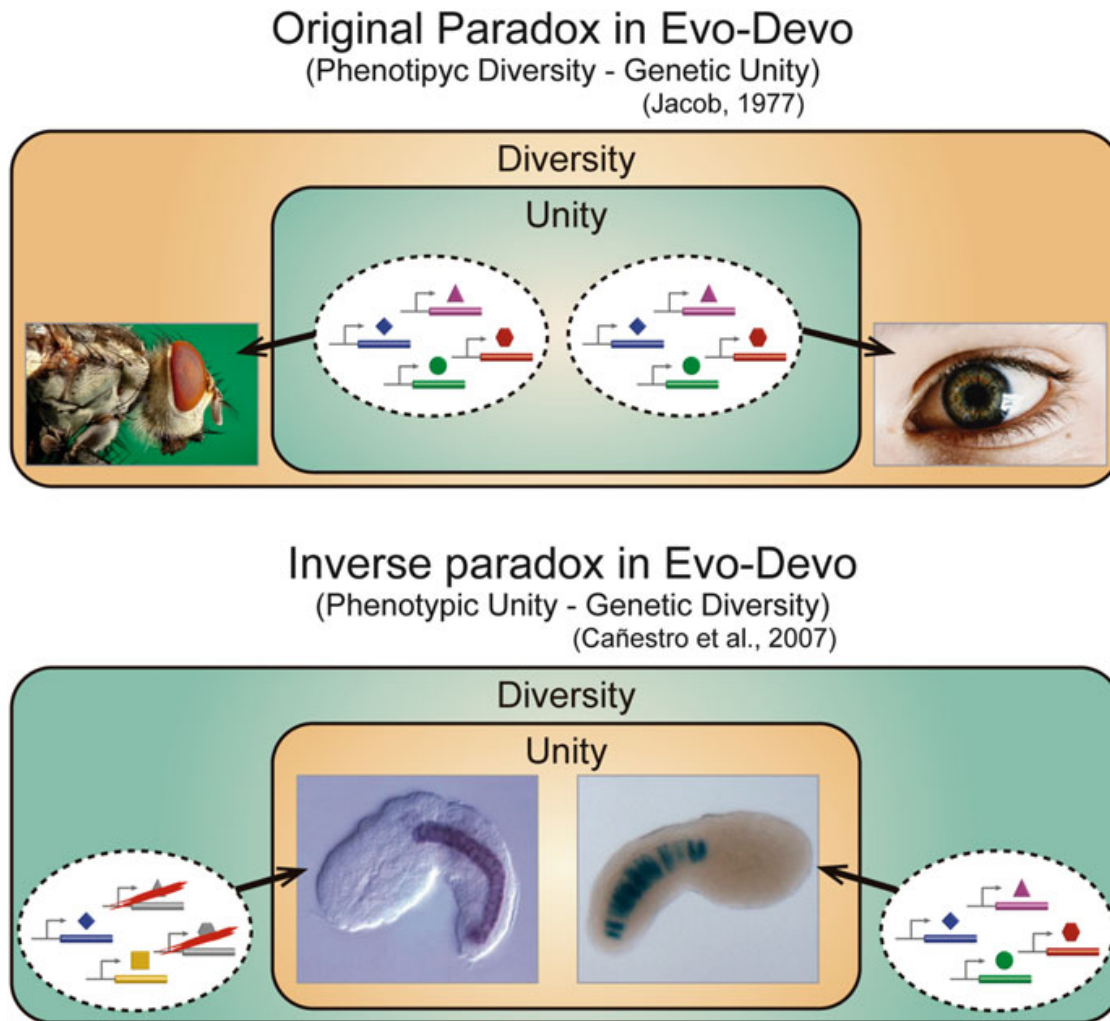
Fig. 4.12 In contrast to the original paradox, the inverse paradox in evo-devo proposes that organisms might develop fundamentally similar morphologies (phenotypic unity) despite important differences in their genetic toolkits (genetic diversity). Because differences in the genetic toolkits often are the consequence of the gene losses, the study of *O. dioica* has been fundamental to build this new conceptual framework

4.5   Future Directions

The tiny planktonic chordate *Oikopleura dioica* appears as an emerging nonclassical animal model, not only attractive to better understand the radiation of our own phylum in the field of evo-devo but also attractive in many other biological fields, such as ecology, environmental toxicology, biomedicine, or basic research. In ecology and toxicology, for instance, the relevance of *O. dioica* and other larvacean species in the marine food webs makes the study of its defense and detoxification mechanisms significant for understanding the impact of industrial pollutants or the effect of the climate change on marine environments, trophic webs, and ocean production. Monitoring *O. dioica* populations has been proposed to be used as valuable sentinels for following the health of marine ecosystems and the expression of its defensome genes as molecular biosensors that marine biologists could use to

monitor stress situations of natural populations. For biomedical applications, the chordate condition of *O. dioica* makes it closer to humans than classical model animals such as fruit flies or worms, but genetically and functionally more tractable than classical vertebrate models as mouse or zebrafish. The homology of organs, tissues, and structures between *O. dioica* and vertebrates (i.e., human) is unquestionable and therefore a good proxy to understand the genetic bases of many human disorders. In addition, the genetic, functional, and structural similarities of *O. dioica* and vertebrates make this species a promising model system for pharmacological screenings and other functional tests by amenable knockdown systems such as RNAi or DNAi approaches. Moreover, the relative small and compact genome of *O. dioica* makes it a good system for basic research in genetics. In the field of gene regulation analysis, for example, its small intergenic regions should facilitate the identification of *cis*-regulatory elements that control gene transcription. And in the field of evolutionary biology, the high mutation rate of *O. dioica*, its propensity to lose genes, and its miniature genome that can be affordably sequenced in many individuals should facilitate studies of population genomics for better understanding evolutionary forces in natural populations, searching for adaptive interpopulation differences, or investigating microevolution processes. Thus, *O. dioica* research has a brilliant present, but an even more promising future.

References

Acuña JL, Kiefer M (2000) Functional response of the appendicularian *Oikopleura dioica*. Limnol Oceanogr 45:608–618

Acuña JL, Bedo AW, Harris RP, Anadón R (1995) The seasonal succession of appendicularians (Tunicata: Appendicularia) off Plymouth. J Mar Biol Assoc UK 75:755–758

Acuña JL, Deibel D, Saunders P, Booth B, Hatfield E, Klein B, Mei ZP, Rivkin R (2002) Phytoplankton ingestion by appendicularians in the North Water. Deep Sea Res Part II Top Stud Oceanogr 49:5101–5115

Albalat R (2009) The retinoic acid machinery in invertebrates: ancestral elements and vertebrate innovations. Mol Cell Endocrinol 313:23–35

Albalat R, Cañestro C (2016) Evolution by gene loss. Nat Rev Genet 17:379–391

Albalat R, Martí-Solans J, Cañestro C (2012) DNA methylation in amphioxus: from ancestral functions to new roles in vertebrates. Brief Funct Genomics 11:142–155

Alldredge AL (2005) The contribution of discarded appendicularian houses to the flux of particulate organic carbon from oceanic surface waters. In: Gorsky G, Youngbluth MJ, Deibel D (eds) Response of marine ecosystems to global change: ecological impact of appendicularians. Éditions Scientifiques, Paris

Almazán A, Ferrández-Roldán A, Albalat R, Cañestro C (2019) Developmental atlas of appendicularian *Oikopleura dioica* actins provides new insights into the evolution of the notochord and the cardio-paraxial muscle in chordates. Dev Biol 448:260–270

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

Bary BM (1960) Notes on ecology, distribution, and systematics of Pelagic Tunicata from New Zealand. Pac Sci 14(2):101–121

Bassham S, Postlethwait J (2000) Brachyury (T) expression in embryos of a larvacean urochordate, *Oikopleura dioica*, and the ancestral role of T. Dev Biol 220:322–332

Bedo AW, Acuña JL, Robins D, Harris RP (1993) Grazing in the micron and the sub-micron particle size range: the case of *Oikopleura dioica* (Appendicularia). Bull Mar Sci 52:2–14

Berna L, Alvarez-Valin F (2014) Evolutionary genomics of fast evolving tunicates. Genome Biol Evol 6:1724–1738

Berna L, D'onofrio G, Alvarez-Valin F (2012) Peculiar patterns of amino acid substitution and conservation in the fast evolving tunicate *Oikopleura dioica*. Mol Phylogenet Evol 62:708–717

Blumenthal T (2004) Operons in eukaryotes. Brief Funct Genomic Proteomic 3:199–211

Bollner T, Holmberg K, Olsson R (1986) A rostral sensory mechanism in *Oikopleura dioica* (Appendicularia). Acta Zool (Stockholm) 67:235–241

Bone Q (1998) The biology of pelagic tunicates. Oxford University Press, New York

Bone Q, Mackie GO (1975) Skin impulses and locomotion in *Oikopleura* (tunicata: larvacea). Biol Bull 149:267–286

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

Bouquet JM, Troedsson C, Novac A, Reeve M, Lechtenborger AK, Massart W, Skaar KS, Aasjord A, Dupont S, Thompson EM (2018) Increased fitness of a key appendicularian zooplankton species under warmer, acidified seawater conditions. PLoS One 13:e0190625

Bowles J, Schepers G, Koopman P (2000) Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev Biol 227:239–255

Brozovic M, Martin C, Dantec C, Dauga D, Mendez M, Simion P, Percher M, Laporte B, Scornavacca C, Di Gregorio A, Fujiwara S, Gineste M, Lowe EK, Piette J, Racioppi C, Ristoratore F, Sasakura Y, Takatori N, Brown TC, Delsuc F, Douzery E, Gissi C, Mcdougall A, Nishida H, Sawada H, Swalla BJ, Yasuo H, Lemaire P (2016) ANISEED 2015: a digital framework for the comparative developmental biology of ascidians. Nucleic Acids Res 44:D808–D818

Brozovic M, Dantec C, Dardaillon J, Dauga D, Faure E, Gineste M, Louis A, Naville M, Nitta KR, Piette J, Reeves W, Scornavacca C, Simion P, Vincentelli R, Bellec M, Aicha SB, Fagotto M, Gueroult-Bellone M, Haeussler M, Jacox E, Lowe EK, Mendez M, Roberge A, Stolfi A, Yokomori R, Brown CT, Cambillau C, Christiaen L, Delsuc F, Douzery E, Dumollard R, Kusakabe T, Nakai K, Nishida H, Satou Y, Swalla B, Veeman M, Volff JN, Lemaire P (2017) ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. Nucleic Acids Res 46:D718–D725

Burighel P, Brena C (2001) Gut ultrastructure of the appendicularian *Oikopleura dioica* (Tunicata). Invertebr Biol 120:278–293

Burke M, Scholl EH, Bird DM, Schaff JE, Colman SD, Crowell R, Diener S, Gordon O, Graham S, Wang X, Windham E, Wright GM, Opperman CH (2015) The plant parasite *Pratylenchus coffeae* carries a minimal nematode genome. Nematology 17:621

Calatayud S, Garcia-Risco M, Rojas NS, Espinosa-Sanchez L, Artime S, Palacios O, Cañestro C, Albalat R (2018) Metallothioneins of the urochordate *Oikopleura dioica* have Cys-rich tandem repeats, large size and cadmium-binding preference. Metallomics 10:1585–1594

Campsteijn C, Ovrebo JI, Karlsen BO, Thompson EM (2012) Expansion of cyclin D and CDK1 paralogs in *Oikopleura dioica*, a chordate employing diverse cell cycle variants. Mol Biol Evol 29:487–502

Cañestro C, Albalat R (2012) Transposon diversity is higher in amphioxus than in vertebrates: functional and evolutionary inferences. Brief Funct Genomics 11:131–141

Cañestro C, Postlethwait JH (2007) Development of a chordate anterior-posterior axis without classical retinoic acid signaling. Dev Biol 305:522–538

Cañestro C, Roncalli V (2018) Gene losses did not stop the evolution of big brains. elife 7:e41912

Cañestro C, Bassham S, Postlethwait JH (2003) Seeing chordate evolution through the *Ciona* genome sequence. Genome Biol 4:208–211

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

Cañestro C, Postlethwait JH, Gonzalez-Duarte R, Albalat R (2006) Is retinoic acid genetic machinery a chordate innovation? Evol Dev 8:394–406

Cañestro C, Yokoi H, Postlethwait JH (2007) Evolutionary developmental biology and genomics. Nat Rev Genet 8:932–942

Cañestro C, Bassham S, Postlethwait JH (2008) Evolution of the thyroid: Anterior-posterior regionalization of the *Oikopleura* endostyle revealed by Otx, Pax 2/5/8, and Hox1 expression. Dev Dyn 237:1490–1499

Cañestro C, Albalat R, Postlethwait JH (2010) *Oikopleura dioica* alcohol dehydrogenase class 3 provides new insights into the evolution of retinoic acid synthesis in chordates. Zool Sci 27:128–133

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

Captanio FL, Curelovich J, Tresguerres M, Negri RM, Viñas MD, Esnal GB (2008) Seasonal cycle of appendicularians at a coastal station (38°28′S, 57°41′W) of the SW Atlantic Ocean. Bull Mar Sci 82:171–184

Chalopin D, Naville M, Plard F, Galiana D, Volff JN (2015) Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol 7:567–580

Chang ES, Neuhof M, Rubinstein ND, Diamant A, Philippe H, Huchon D, Cartwright P (2015) Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. Proc Natl Acad Sci USA 112:14912–14917

Chavali S, Morais DA, Gough J, Babu MM (2011) Evolution of eukaryotic genome architecture: insights from the study of a rapidly evolving metazoan, *Oikopleura dioica*: non-adaptive forces such as elevated mutation rates may influence the evolution of genome architecture. BioEssays 33:592–601

Cima F, Brena C, Burighel P (2002) Multifarious activities of gut epithelium in an appendicularian (*Oikopleura dioica*: Tunicata). Mar Biol 141:479–490

Clarke T, Bouquet JM, Fu X, Kallesoe T, Schmid M, Thompson EM (2007) Rapidly evolving lamins in a chordate, *Oikopleura dioica*, with unusual nuclear architecture. Gene 396:159–169

Cleaver O, Krieg PA (2001) Notochord patterning of the endoderm. Dev Biol 234:1–12

Coppola U, Ristoratore F, Albalat R, D'aniello S (2019) The evolutionary landscape of the Rab family in chordates. Cell Mol Life Sci. https://doi.org/10.1007/s00018-019-03103-7

Corbo JC, Di Gregorio A, Levine M (2001) The ascidian as a model organism in developmental and evolutionary biology. Cell 106:535–538

Costello J, Stancyk SE (1983) Tidal influence upon appendicularian abundance in North Inlet estuary, South Carolina. J Plankton Res 5:263–277

Danks G, Campsteijn C, Parida M, Butcher S, Doddapaneni H, Fu B, Petrin R, Metpally R, Lenhard B, Wincker P, Chourrout D, Thompson EM, Manak JR (2013) OikoBase: a genomics and developmental transcriptomics resource for the urochordate *Oikopleura dioica*. Nucleic Acids Res 41:D845–D853

Davoll P, Youngbluth M (1990) Heterotrophic activity on appendicularian (Tunicata: Appendicularia) houses in mesopelagic regions and their potential contribution to particle flux. Deep-Sea Res 37:285–294

Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomoso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KEM, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang H-G, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, Desantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee B-I, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-I T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS (2002) The draft genome of *Ciona intestinalis:* insights into chordate and vertebrate origins. Science 298:2157–2167

Delsman HC (1910) Beiträge zur Entwicklungsgeschichte von *Oikopleura dioica*. Verh Rijksinst Onderz Zee 3:1–24

Delsman HC (1912) Weitere beobachtungen über die entwicklung von *Oikopleura dioica*. Tijdschr ned dierk Ver 12:199–206

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

Deng W, Henriet S, Chourrout D (2018) Prevalence of mutation-prone microhomology-mediated end joining in a chordate lacking the c-NHEJ DNA repair pathway. Curr Biol 28(3337–3341): e3334

Denoeud F, Henriet S, Mungpakdee S, Aury JM, Da Silva C, Brinkmann H, Mikhaleva J, Olsen LC, Jubin C, Cañestro C, Bouquet JM, Danks G, Poulain J, Campsteijn C, Adamski M, Cross I, Yadetie F, Muffato M, Louis A, Butcher S, Tsagkogeorga G, Konrad A, Singh S, Jensen MF, Cong EH, Eikeseth-Otteraa H, Noel B, Anthouard V, Porcel BM, Kachouri-Lafond R, Nishino A, Ugolini M, Chourrout P, Nishida H, Aasland R, Huzurbazar S, Westhof E, Delsuc F, Lehrach H, Reinhardt R, Weissenbach J, Roy SW, Artiguenave F, Postlethwait JH, Manak JR, Thompson EM, Jaillon O, Du Pasquier L, Boudinot P, Liberles DA, Volff JN, Philippe H, Lenhard B, Roest Crollius H, Wincker P, Chourrout D (2010) Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science 330:1381–1385

Deriano L, Roth DB (2013) Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Annu Rev Genet 47:433–455

Duboule D (2007) The rise and fall of Hox gene clusters. Development 134:2549–2560

Edvardsen RB, Lerat E, Maeland AD, Flat M, Tewari R, Jensen MF, Lehrach H, Reinhardt R, Seo HC, Chourrout D (2004) Hypervariable and highly divergent intron-exon organizations in the chordate *Oikopleura dioica*. J Mol Evol 59:448–457

Edvardsen RB, Seo HC, Jensen MF, Mialon A, Mikhaleva J, Bjordal M, Cartry J, Reinhardt R, Weissenbach J, Wincker P, Chourrout D (2005) Remodelling of the homeobox gene complement in the tunicate *Oikopleura dioica*. Curr Biol 15:R12–R13

Essenberg CE (1922) The seasonal distribution of the Appendicularia in the region of San Diego, California. Ecology 3:55–64

Fenaux R (1972) A historical survey of the appendicularians from the area covered by the IIOE. Mar Biol 16:230–235

Fenaux R (1976) Cycle vital d'un appendiculaire: *Oikopleura dioica* Fol, 1872. Ann Inst Océanogr Paris 52:89–101

Fenaux R (1986) The house of *Oikopleura dioica* (Tunicata, Appendicularia): structure and functions. Zoomorphology 106:224–231

Fenaux R (1998a) Anatomy and functional morphology of the Appendicularia. In: Bone Q (ed) The biology of pelagic tunicates. Oxford University Press, New York

Fenaux R (1998b) Life history of the Appendicularia. In: Bone Q (ed) The biology of pelagic tunicates. Oxford University Press, Oxford

Fenaux R, Gorsky G (1985) Nouvelle technique d'élevage des appendiculaires. Rapp Comm Int Mer Médit 29:291–292

Fenaux R, Bone Q, Deibel D (1998) Appendicularian distribution and zoogeography. In: Bone Q (ed) The biology of pelagic tunicates. Oxford University Press, Oxford

Fernández D, López-Urrutia A, Fernández A, Acuña JL, Harris R (2004) Retention efficiency of 0.2 to 6 μm particles by the appendicularians *Oikopleura dioica* and *Fritillaria borealis*. Mar Ecol Prog Ser 266:89–101

Ferrier DE (2011) Tunicates push the limits of animal evo-devo. BMC Biol 9:3

Ferrier DEK, Sogabe S (2018) Genome biology: unconventional DNA repair in an extreme genome. Curr Biol 28:R1208–R1210

Flood PR, Afzelius BA (1978) The spermatozoon of *Oikopleura dioica* Fol (Larvacea, Tunicata). Cell Tissue Res 191:27–37

Flood PR, Deibel D (1998) The appendicularian house. In: Bone Q (ed) The biology of pelagic tunicates. Oxford University Press, Oxford

Fol H (1872) Etudes sur les Appendiculaires du Détroit de Messine. Mem Soc Physique Hist Nat Geneve 21:445

Fredriksson G, Olsson R (1981) The oral gland cells of *Oikopleura dioica* (Tunicata Appendicularia). Acta Zool (Stockholm) 62:195–200

Fredriksson G, Olsson R (1991) The subchordal cells of *Oikopleura dioica* and *Oikopleura albicans* (Appendicularia, Chrodata). Acta Zool (Stockholm) 72:251–256

Fredriksson G, Ofverholm T, Ericson LE (1985) Ultrastructural demonstration of iodine binding and peroxidase activity in the endostyle of *Oikopleura dioica* (Appendicularia). Gen Comp Endocrinol 58:319–327

Fu X, Adamski M, Thompson EM (2008) Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*. Mol Biol Evol 25:1067–1080

Fujii S, Nishio T, Nishida H (2008) Cleavage pattern, gastrulation, and neurulation in the appendicularian, *Oikopleura dioica*. Dev Genes Evol 218:69–79

Gabaldon T, Capella-Gutierrez S (2010) Lack of phylogenetic support for a supposed actinobacterial origin of peroxisomes. Gene 465:61–65

Gabaldon T, Snel B, Van Zimmeren F, Hemrika W, Tabak H, Huynen MA (2006) Origin and evolution of the peroxisomal proteome. Biol Direct 1:8

Galt CP (1972) Development of *Oikopleura dioica* (Urochordata: Larvacea): ontogeny of behavior and of organ systems related to construction and use of the house. PhD, University of Washington, Seattle

Galt C (1978) Bioluminescence: dual mechanism in a planktonic tunicate produces brilliant surface display. Science 200:70–72

Galt CP, Fenaux R (1990) Urochordata larvacea. In: Adiyodi KG, Adiyodi RG (eds) Reproductive biology of invertebrates. Oxford and IBH, New Delhi

Ganot P, Kallesoe T, Reinhardt R, Chourrout D, Thompson EM (2004) Spliced-leader RNA trans splicing in a chordate, *Oikopleura dioica*, with a compact genome. Mol Cell Biol 24:7795–7805

Ganot P, Bouquet JM, Kallesoe T, Thompson EM (2007) The Oikopleura coenocyst, a unique chordate germ cell permitting rapid, extensive modulation of oocyte production. Dev Biol 302:591–600

Gaughan D, Potter IC (1994) Relative abundance and seasonal changes in the macrozooplankton of the lower estuary in South-Western Australia. Rec West Aust Mus 16:461–474

Georges D, Holmberg K, Olsson R (1988) The ventral midbrain cells in *Oikopleura dioica* (Appendicularia). Acta Embryol Morphol Exp 9:39–47

Goldschmidt R (1903) Notiz über die Entwickekung der Appendicularien-Biologisches Centralblatt, Band

Goldstone JV, Hamdoun A, Cole BJ, Howard-Ashby M, Nebert DW, Scally M, Dean M, Epel D, Hahn ME, Stegeman JJ (2006) The chemical defensome: environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. Dev Biol 300:366–384

Goldstone JV, Mcarthur AG, Kubota A, Zanette J, Parente T, Jonsson ME, Nelson DR, Stegeman JJ (2010) Identification and developmental expression of the full complement of Cytochrome P450 genes in Zebrafish. BMC Genomics 11:643

Gorsky G, Fenaux R (1998) The role of Appendicularia in marine food webs. In: Bone Q (ed) The biology of pelagic tunicates. Oxford University Press, Oxford

Gorsky G, Fisher N, Fowler S (1984) Biogenic debris from the pelagic tunicate *Oikopleura dioica* and its role in the vertical transport of a transuranium element. Est Coast Shelf Sci 18:13–23

Handberg-Thorsager M, Gutierrez-Mazariegos J, Arold ST, Kumar Nadendla E, Bertucci PY, Germain P, Tomancak P, Pierzchalski K, Jones JW, Albalat R, Kane MA, Bourguet W, Laudet V, Arendt D, Schubert M (2018) The ancestral retinoic acid receptor was a low-affinity sensor triggering neuronal differentiation. Sci Adv 4:eaao1261

Heenan P, Zondag L, Wilson MJ (2016) Evolution of the Sox gene family within the chordate phylum. Gene 575:385–392

Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF, Students of Bioinformatics Computer Labs A (2006) The expansion of the metazoan microRNA repertoire. BMC Genomics 7:25

Holland LZ (2016) Tunicates. Curr Biol 26:R146–R152

Holland LZ, Gibson-Brown J (2003) The *Ciona intestinalis* genome: when the constraints are off. BioEssays 25:529–532

Holland L, Gorsky G, Fenaux R (1988) Fertilization in *Oikopleura dioica* (Tunicata, Appendicularia): acrosome reaction, cortical reaction and sperm-egg fusion. Zoomorphology 108:229–243

Holland PW, Booth HA, Bruford EA (2007) Classification and nomenclature of all human homeobox genes. BMC Biol 5:47

Holmberg K (1982) The ciliated brain duct of *Oikopleura dioica* (Tunicata, Appendicularia). Acta Zool 63:101–109

Holmberg K (1984) A transmission electron microscopic investigation of the sensory vesicle in the brain of *Oikopleura dioica* (Appendicularia). Zoomorphology 104:298–303

Hopcroft RR, Roff JC (1995) Zooplankton growth rates: extraordinary production by the larvacean *Oikopleura dioica* in tropical waters. J Plankton Res 17:205–220

Hopcroft RR, Roff JC, Bouman HA (1998) Zooplankton growth rates: the larvaceans *Appendicularia*, *Fritillaria* and *Oikopleura* in tropical waters. J Plankton Res 20:539–555

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

Hughes AL, Friedman R (2005) Loss of ancestral genes in the genomic evolution of *Ciona intestinalis*. Evol Dev 7:196–200

Hwan Lee J, Chae J, Kim W-R, Won Jung S, Man Kim J (2001) Seasonal variation of phytoplankton and zooplankton communities in the Coastal Waters off Tongyeong in Korea. Ocean Polar Res 23:245–253

Inoue J, Satoh N (2018) Deuterostome genomics: lineage-specific protein expansions that enabled chordate muscle evolution. Mol Biol Evol 35:914–924

Irimia M, Tena JJ, Alexis MS, Fernandez-Minan A, Maeso I, Bogdanovic O, De La Calle-Mustienes E, Roy SW, Gomez-Skarmeta JL, Fraser HB (2012) Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res 22:2356–2367

Jacob F (1977) Evolution and tinkering. Science 196:1161–1166

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

King KR, Hollibaugh JT, Azam F (1980) Predator-prey interactions between the larvacean *Oikopleura dioica* and bacterioplankton in enclosed water columns. Mar Biol 56:49–57

Kishi K, Onuma TA, Nishida H (2014) Long-distance cell migration during larval development in the appendicularian, *Oikopleura dioica*. Dev Biol 395:299–306

Kishi K, Hayashi M, Onuma TA, Nishida H (2017) Patterning and morphogenesis of the intricate but stereotyped oikoplastic epidermis of the appendicularian, *Oikopleura dioica*. Dev Biol 428:245–257

Kmita M, Duboule D (2003) Organizing axes in time and space; 25 years of colinear tinkering. Science 301:331–333

Kocot KM, Tassia MG, Halanych KM, Swalla BJ (2018) Phylogenomics offers resolution of major tunicate relationships. Mol Phylogenet Evol 121:166–173

Kugler JE, Kerner P, Bouquet JM, Jiang D, Di Gregorio A (2011) Evolutionary changes in the notochord genetic toolkit: a comparative analysis of notochord genes in the ascidian *Ciona* and the larvacean *Oikopleura*. BMC Evol Biol 11:21

Kusakabe T, Araki I, Satoh N, Jeffery WR (1997) Evolution of chordate actin genes: evidence from genomic organization and amino acid sequences. J Mol Evol 44:289–298

Larson RJ (1987) Daily ration and predation by medusae and ctenophores in Saanich Inlet, B.C., Canada. Neth J Sea Res 21:35–44

Last JM (1972) Egg development, fecundity and growth of *Oikopleura Dioica* Fol in the North Sea. ICES J Mar Sci 34:232–237

Lohmann H (1933) Erste Klasse der Tunicaten: Appendiculariae. In: Kükenthal W, Krumbach T (eds) Handbuch der Zoologie. Walter De Gruyter, Berlin

Lopez-Urrutia A, Acuña JL (1999) Gut throughput dynamics in the appendicularian *Oikopleura dioica*. Mar Ecol Prog Ser 191:195–205

Martinucci G, Brena C, Cima F, Burighel P (2005) Synchronous spermatogenesis in appendicularians. In: Gorsky G, Youngbluth M, Deibel D (eds) Response of marine ecosystems to global changes: ecological impact of appendicularians. Éditions des Archives Contempoaines, Paris

Martí-Solans J, Ferrández-Roldán A, Godoy-Marín H, Badia-Ramentol J, Torres-Aguila NP, Rodríguez-Marí A, Bouquet JM, Chourrout D, Thompson EM, Albalat R, Cañestro C (2015) *Oikopleura dioica* culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. Genesis 53:183–193

Martí-Solans J, Belyaeva OV, Torres-Aguila NP, Kedishvili NY, Albalat R, Cañestro C (2016) Coelimination and survival in gene network evolution: dismantling the RA-signaling in a chordate. Mol Biol Evol 33:2401–2416

Menéndez M, Biancalana F, Berasategui A, Fernández Severini MD, Hoffmeyer MS, Esteves JL (2011) Mesozooplankton composition and spatial distribution, Nuevo Gulf, Patagonia, Argentina. Check List 7:101–107

Mikhailov KV, Slyusarev GS, Nikitin MA, Logacheva MD, Penin AA, Aleoshin VV, Panchin YV (2016) The genome of *Intoshia linei* affirms orthonectids as highly simplified spiralians. Curr Biol 26:1768–1774

Mikhaleva Y, Kreneisz O, Olsen LC, Glover JC, 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

Mikhaleva Y, Skinnes R, Sumic S, Thompson EM, Chourrout D (2018) Development of the house secreting epithelium, a major innovation of tunicate larvaceans, involves multiple homeodomain transcription factors. Dev Biol 443:117–126

Nakamura Y, Suzuki K, Suzuki S, Hiromi J (1997) Production of *Oikopleura dioica* (Appendicularia) following a picoplankton 'bloom' in a eutrophic coastal area. J Plankton Res 19:113–124

Navratilova P, Danks GB, Long A, Butcher S, Manak JR, Thompson EM (2017) Sex-specific chromatin landscapes in an ultra-compact chordate genome. Epigenetics Chromatin 10:3

Nelson DR, Goldstone JV, Stegeman JJ (2013) The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. Philos Trans R Soc Lond Ser B Biol Sci 368:20120474

Nilsen IW, Myrnes B, Edvardsen RB, Chourrout D (2003) Urochordates carry multiple genes for goose-type lysozyme and no genes for chicken- or invertebrate-type lysozymes. Cell Mol Life Sci 60:2210–2218

Nishida H (1987) Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III Up to the tissue restricted stage. Dev Biol 121:526–541

Nishida H (2002) Patterning the marginal zone of early ascidian embryos: localized maternal mRNA and inductive interactions. BioEssays 24:613–624

Nishida H (2008) Development of the appendicularian *Oikopleura dioica*: culture, genome, and cell lineages. Develop Growth Differ 50(Suppl 1):S239–S256

Nishida H, Stach T (2014) Cell lineages and fate maps in tunicates: conservation and modification. Zool Sci 31:645–652

Nishino A, Morisawa M (1998) Rapid oocyte growth and artificial fertilization of the Larvaceans *Oikopleura dioica* and *Oikopleura longicauda*. Zool Sci 15:723–727

Nishino A, Satoh N (2001) The simple tail of chordates: phylogenetic significance of appendicularians. Genesis 29:36–45

Nishino A, Satou Y, Morisawa M, Satoh N (2000) Muscle actin genes and muscle cells in the appendicularian, *Oikopleura longicauda*: phylogenetic relationships among muscle tissues in the urochordates. J Exp Zool 288:135–150

Ohno S (1970) Evolution by gene duplication. Springer, New York

Olsson R (1963) Endostyles and endostylar secretions: a comparative histochemical study. Acta Zool (Stockholm) 44:299–329

Olsson R (1965a) Comparative morphology and physiology of the *Oikopleura* notochord. Isr J Zool 14:213–220

Olsson R (1965b) The cytology of the endostyle of *Oikopleura dioica*. Ann NY Acad Sci 118:1038–1051

Olsson R (1975) Primitive coronet cells in the brain of *Oikopleura* (Appendicularia, Tunicata). Acta Zool (Stockholm) 56:155–161

Olsson R, Holmberg K, Lilliemark Y (1990) Fine structure of the brain and brain nerves of *Oikopleura dioica* (Urochordata, Appendicularia). Zool Morphol 110:1–7

Omotezako T, Nishino A, Onuma TA, Nishida H (2013) RNA interference in the appendicularian *Oikopleura dioica* reveals the function of the Brachyury gene. Dev Genes Evol 223:261–267

Omotezako T, Onuma TA, Nishida H (2015) DNA interference: DNA-induced gene silencing in the appendicularian *Oikopleura dioica*. Proc Biol Sci 282:20150435

Omotezako T, Matsuo M, Onuma TA, Nishida H (2017) DNA interference-mediated screening of maternal factors in the chordate *Oikopleura dioica*. Sci Rep 7:44226

Onuma TA, Isobe M, Nishida H (2017) Internal and external morphology of adults of the appendicularian, *Oikopleura dioica*: an SEM study. Cell Tissue Res 367:213–227

Paffenhöfer G-A (1973) The cultivation of an appendicularian through numerous generations. Mar Biol 22:183–185

Pannunzio NR, Watanabe G, Lieber MR (2018) Nonhomologous DNA end-joining for repair of DNA double-strand breaks. J Biol Chem 293:10512–10523

Patel AA, Steitz JA (2003) Splicing double: insights from the second spliceosome. Nat Rev Mol Cell Biol 4:960–970

Raduan A, Blanco C, Soler E, Del Rio JG, Raga JA (1985) The seasonal distribution of the Appendicularia in the Bay of Cullera, Spain. Rapp Comm Int Mer Médit 29:293–294

Robinson AJ, Kunji ER, Gross A (2012) Mitochondrial carrier homolog 2 (MTCH2): the recruitment and evolution of a mitochondrial carrier protein to a critical player in apoptosis. Exp Cell Res 318:1316–1323

Robison BH, Reisenbichler KR, Sherlock RE (2005) Giant larvacean houses: rapid carbon transport to the deep sea floor. Science 308:1609–1611

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

Salensky W (1903) Etudes anatomiques sur les Appendiculaires. I *Oikopleura vanhoeffeni* Lohmann. Mem Acad Sci St Petesbourg Ser 13:1–44

Salensky W (1904) Etudes anatomiques sur les Appendiculaires. II *Oikopleura refescens*. Fol Mem Acad Sci St Petesbourg Ser 15:1–54

Salensky W (1905) Zur Morphologie der Cardialorgane der Appendicularien. Sixième Congrès International de Zoologie, Berne. Kandiget fils. Genève, 381–383

Sato R, Tanaka Y, 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 4:285–295

Satoh N, Satou Y, Davidson B, Levine M (2003) *Ciona intestinalis*: an emerging model for whole-genome analyses. Trends Genet 19:376–381

Satoh N, Tagawa K, Takahashi H (2012) How was the notochord born? Evol Dev 14:56–75

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

Schluter A, Fourcade S, Ripp R, Mandel JL, Poch O, Pujol A (2006) The evolutionary origin of peroxisomes: an ER-peroxisome connection. Mol Biol Evol 23:838–845

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

Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D (2004) Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. Nature 431:67–71

Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R (2006) Comprehensive splice-site analysis using comparative genomics. Nucleic Acids Res 34:3955–3967

Somorjai IML, Martí-Solans J, Diaz-Gracia M, Nishida H, Imai KS, Escriva H, Cañestro C, Albalat R (2018) Wnt evolution and function shuffling in liberal and conservative chordate genomes. Genome Biol 19:98

Soviknes AM, Glover JC (2007) Spatiotemporal patterns of neurogenesis in the appendicularian *Oikopleura dioica*. Dev Biol 311:264–275

Soviknes AM, Glover JC (2008) Continued growth and cell proliferation into adulthood in the notochord of the appendicularian *Oikopleura dioica*. Biol Bull 214:17–28

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

Soviknes AM, Chourrout D, Glover JC (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, Thompson EM (2001) Molecular patterning of the oikoplastic epithelium of the larvacean tunicate *Oikopleura dioica*. J Biol Chem 276:20624–20632

Stach TG (2009) Anatomy of the trunk mesoderm in tunicates: homology considerations and phylogenetic interpretation. Zoomorphology 128:97–109

Stach T, Anselmi C (2015) High-precision morphology: bifocal 4D-microscopy enables the comparison of detailed cell lineages of two chordate species separated for more than 525 million years. BMC Biol 13:113

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

Stemple DL (2005) Structure and function of the notochord: an essential organ for chordate development. Development 132:2503–2512

Sulston JE, Schierenberg E, White JG, Thomson JN (1983) The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev Biol 100:64–119

Thompson EM, Kallesøe T, Spada F (2001) Diverse genes expressed in distinct regions of the trunk epithelium define a monolayer cellular template for construction of the oikopleurid house. Dev Biol 238:260–273

Tomita M, Shiga N, Ikeda T (2003) Seasonal occurrence and vertical distribution of appendicularians in Toyama Bay, southern Japan Sea. J Plankton Res 25:579–589

Torres-Aguila NP, Martí-Solans J, Ferrandez-Roldan A, Almazan A, Roncalli V, D'aniello S, Romano G, Palumbo A, Albalat R, Cañestro C (2018) Diatom bloom-derived biotoxins cause aberrant development and gene expression in the appendicularian chordate *Oikopleura dioica*. Commun Biol 1:121

Troedsson C, Bouquet J-M, Aksnes DL, Thompson EM (2002) Resource allocation between somatic growth and reproductive output in the pelagic chordate *Oikopleura dioica* allows opportunistic response to nutritional variation. Mar Ecol Prog Ser 243:83–91

Troedsson C, Bouquet J-M, Lobon C, Novac A, Nejstgaard J, Dupont S, Bosak S, Jakobsen H, Romanova N, Pankoke L, Isla A, Dutz JR, Sazhin A, Thompson E (2013) Effects of ocean acidification, temperature and nutrient regimes on the appendicularian *Oikopleura dioica*: a mesocosm study. Mar Biol 160:2175–2187

Uren AG, O'rourke K, Aravind LA, Pisabarro MT, Seshagiri S, Koonin EV, Dixit VM (2000) Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. Mol Cell 6:961–967

Uye S-I, Ichino S (1995) Seasonal variations in abundance, size composition, biomass and production rate of *Oikopleura dioica* (Fol) (Tunicata: Appendicularia) in a temperate eutrophic inlet. J Exp Mar Biol Ecol 189:1–11

Volff JN, Lehrach H, Reinhardt R, Chourrout D (2004) Retroelement dynamics and a novel type of chordate retrovirus-like elements in the miniature genome of the tunicate *Oikopleura dioica*. Mol Biol Evol 21:2022–2033

Walkusz W, Storemark K, Skau T, Gannefors C, Lundberg M (2003) Zooplankton community structure; a comparison of fjords, open water and ice stations in the Svalbard area. Pol Polar Res 24:149–165

Wang K, Dantec C, Lemaire P, Onuma TA, Nishida H (2017) Genome-wide survey of miRNAs and their evolutionary history in the ascidian, *Halocynthia roretzi*. BMC Genomics 18:314

Wegner M (2010) All purpose Sox: the many roles of Sox proteins in gene expression. Int J Biochem Cell Biol 42:381–390

Weill M, Philips A, Chourrout D, Fort P (2005) The caspase family in urochordates: distinct evolutionary fates in ascidians and larvaceans. Biol Cell 97:857–866

Welsch U, Storch V (1969) On the fine structure of the chorda dorsalis in lower chordata. (*Dendrodoa grossularia* (v. Beneden) and *Oikopleura dioica* Fol). Z Zellforsch Mikrosk Anat 93:547–549

Yadetie F, Butcher S, Forde HE, Campsteijn C, Bouquet JM, Karlsen OA, Denoeud F, Metpally R, Thompson EM, Manak JR, Goksoyr A, Chourrout D (2012) Conservation and divergence of chemical defense system in the tunicate *Oikopleura dioica* revealed by genome wide response to two xenobiotics. BMC Genomics 13:55

Zarsky V, Tachezy J (2015) Evolutionary loss of peroxisomes-not limited to parasites. Biol Direct 10:74

CHAPTER IV

Diatom bloom-derived biotoxins cause aberrant development and gene expression in the appendicularian chordate *Oikopleura dioica*

Diatom bloom-derived biotoxins cause aberrant development and gene expression in the appendicularian chordate *Oikopleura dioica*

Nuria P. Torres-Águila[1], Josep Martí-Solans[1], Alfonso Ferrández-Roldán[1], Alba Almazán[1], Vittoria Roncalli[1], Salvatore D'Aniello[2], Giovanna Romano [iD][3], Anna Palumbo[2], Ricard Albalat [iD][1] & Cristian Cañestro [iD][1]

Investigating environmental hazards than could affect appendicularians is of prime ecological interest because they are among the most abundant components of the mesozooplankton. This work shows that embryo development of the appendicularian *Oikopleura dioica* is compromised by diatom bloom-derived biotoxins, even at concentrations in the same range as those measured after blooms. Developmental gene expression analysis of biotoxin-treated embryos uncovers an aberrant golf ball-like phenotype affecting morphogenesis, midline convergence, and tail elongation. Biotoxins induce a rapid upregulation of defensome genes, and considerable delay and silencing of zygotic transcription of developmental genes. Upon a possible future intensification of blooms associated with ocean warming and acidification, our work puts an alert on the potential impact that an increase of biotoxins may have on marine food webs, and points to defensome genes as molecular biosensors that marine ecologists could use to monitor the genetic stress of natural populations exposed to microalgal blooms.

[1] Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona. Av. Diagonal 643, 08028 Barcelona, Catalonia, Spain. [2] Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Villa Comunale 80121, Napoli, Italy. [3] Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy. Correspondence and requests for materials should be addressed to R.A. (email: ralbalat@ub.edu) or to C.C. (email: canestro@ub.edu)

Climate change impact on ecosystems is complex and often with no direct relationship. Nevertheless, it is conceivable that pressures such as ocean warming and acidification may potentially intensify the severity and frequency of harmful algal blooms, globally influencing marine planktonic systems[1]. Harmful algal blooms include diatoms, which are among the most important photoautotrophic organisms that drive marine food web dynamics[2]. Diatoms can produce different biotoxins, some of which may act as a defense mechanism against grazers[3–6]. Among diatom-derived biotoxins, oxylipins are of prime interest because of their negative effects on the reproduction of copepods[4], the main grazers of these algae and one of the most abundant components of the mesozooplankton[7]. Oxylipins, which include polyunsaturated aldehydes (PUAs), are secondary metabolic end-products of a lipoxygenase/hydroperoxide lyase pathway that are toxic when released to the environment[8]. During algal blooms, cell membranes are broken by cellular friction, massive grazing or senescence at the end of the bloom, generating oxylipin-rich microzones that can alter the biology of neighboring organisms[9].

Oxylipin toxicity does not only affect copepods, but recent studies have shown that several marine species are compromised, including sea urchins[9–13], ascidian urochordates[14–16], as well as mollusks and annelids[10,17,18]. Despite the fact that analyses of different organisms have shown that PUA's toxicity can affect a wide range of physiological processes, including oocyte maturation, sperm motility, fertilization, cell proliferation, embryogenesis, hatching, metamorphosis and apoptosis (reviewed in the ref. [19]), the molecular bases of PUA's toxicity remain often unclear. Studies in copepods and ascidians have revealed, for instance, that PUAs affect the expression of stress response genes (i.e., defensome[20]) associated with the metabolism of the glutathione system (e.g., *Gclm*, *Ggt*, and *Gst*)[15,16,21,22]. The expression of defensome genes related to aldehyde detoxification derived from lipid peroxidation (e.g., *Aldh2* and *Aldh8*) are also altered by PUAs, suggesting that these compounds may also induce oxidative stress[21,23,24].

While the physiological effects of PUA's toxicity have been investigated, the developmental genetic mechanisms affected by PUAs causing embryo malformations remain largely unknown. In sea urchins and ascidians, expression analyses by qRT-PCR of developmental genes related to embryo patterning (e.g., *Hox*, *Parahox*) or signaling pathways (e.g., *Wnt* and *nitric oxide/Erk*) have revealed differences in the expression levels[15,16,24]. The mechanism by which PUAs induced those differences remains uncertain. Moreover, to our knowledge, there is no data describing what developmental processes and germ layers are altered by PUAs.

In this work, we study if PUAs affect the embryonic development of the appendicularian urochordate specie *Oikopleura dioica* for two main reasons. First, from an ecological perspective, appendicularians (a.k.a. larvaceans) are relevant because they are cosmopolitan pelagic filter-feeding organisms, considered the second most abundant after copepods in marine mesozooplankton[25,26]. Appendicularians graze about 10% of the ocean's primary production[27]. The appendicularian capability of trapping a wide range of particle sizes thanks to their unique filter-feeding apparatus house makes them to occupy an important trophic position in food webs. Appendicularians act as an important short-circuit that allows a rapid energy transfer from colloidal carbon and phytoplankton primary producers to zooplanktivorous predators such as fish larvae[28,29], contributing at the same time to at least 8% of the vertical carbon transport to the deep ocean[30–32]. Therefore, the study of PUAs toxicity on the biology of appendicularians is fundamental from an ecological perspective to better understand the potential effect of increased

amount of oxylipins on marine food webs and carbon cycle, upon potential intensification of harmful diatom blooms in the context of ocean warming and acidification.

Second, from an evolutionary and developmental biology (evo-devo) perspective, the chordate *O. dioica* is an attractive animal model because it has undergone a process of genetic and morphological simplification during the evolution of urochordates[33–36]. The low genetic redundancy of *O. dioica* genome, in comparison with the twice-duplicated vertebrate genome (reviewed in the ref. [37]), together with the extraordinary amount of gene losses[36], make that its functional redundancy is lower than in other animal species (e.g., vertebrates). Some gene losses have led this organism to be considered as an evolutionary knockout that can facilitate the dissection of complex genetic networks[36,38–41]. *O. dioica* has lost, for instance, the retinaldehyde dehydrogenase *Aldh1a* and most of the components of the metabolic pathway of retinoic acid (RA)[42], a signaling molecule that in all other chordates plays a fundamental role in the regulation of embryo development, adult organ homeostasis and gametogenesis[43–45]. The loss of the *Aldh1a* in *O. dioica* is particularly relevant for the present work because it has been proposed that PUAs such as *trans,trans*-2,4-decadienal (DD) can compete with retinaldehyde for the substrate binding site of the Aldh1a[46]. This finding suggested the hypothesis that the teratogenicity of DD could be due to its interference with RA-synthesis, impairing thereby normal RA signaling during embryo development. Thus, our work aims to study *O. dioica* as an *Aldh1a* evolutionary knockout to test this hypothesis, because DD should produce minor or no alterations on *O. dioica* embryo development if RA signaling is the main developmental pathway targeted by DD. On the contrary, if DD severely affects the development of *O. dioica*, it would imply that other developmental mechanisms, distinct from RA signaling, are affected by DD exposure.

This work reveals that oxylipins can induce aberrant development and gene expression of appendicularians, even at concentrations in the same range than those measured after blooms, and puts an alert on the potential impact that an increase of biotoxins may have on marine food webs.

Results

Dose-dependent effects of DD on embryo development of *O. dioica*. To investigate the possible teratogenic effects and dose-response of DD on *O. dioica* embryo development, we analyzed, at the time of hatching, embryos derived from eggs treated with DD at different concentrations (from 0.05 to 2.5 μg mL^{-1}, this is from 0.33 to 16.42 μM). In contrast to DMSO-control conditions, DD-treatments altered embryo development in a dose-dependent manner, in which the severity of the phenotypic aberrations and the proportion of affected embryos depended on the concentration of the DD (Fig. 1). We classified the aberrant phenotypes in three categories: abnormal hatchlings, pre-tailbud arrested embryos with golf ball morphology, and 1-cell arrested zygotes (Fig. 1a–d). According to the differences in the proportions of aberrant phenotypes (statistical analyses are provided in Supplementary Fig. 1), we classified DD concentrations in four categories (Fig. 1e): innocuous (≤0.05 μg mL^{-1}), in which no significant differences were observed in comparison with DMSO-controls (*P*-value 0.998); mild (from 0.075 to 0.125 μg mL^{-1}), in which most embryos did hatch, but the presence of aberrant morphologies with shorten or kinked tails was higher than in the DMSO-control condition; moderate (from 0.25 to 2.0 μg mL^{-1}), in which most animals did not hatch, remaining arrested in a pre-tailbud stage with the appearance of a golf ball; severe
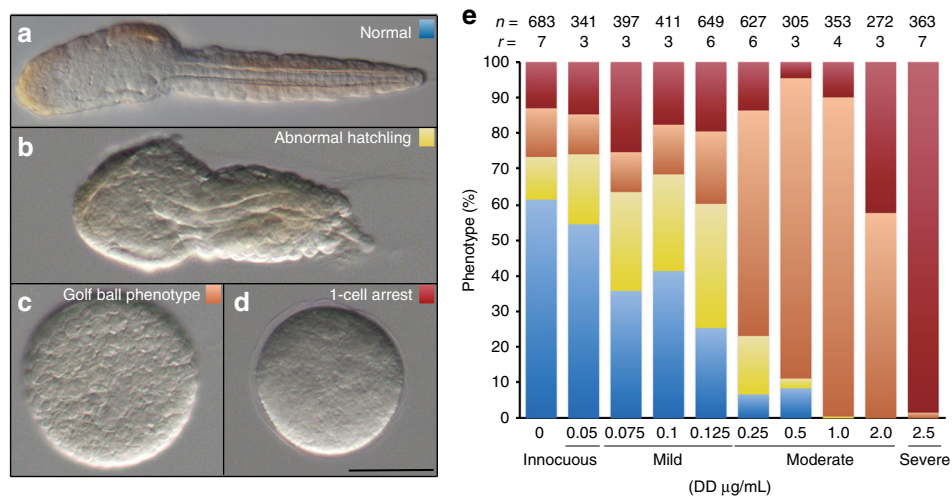
Fig. 1 Effects and dose-response of DD on Oikopleura dioica embryo development. After 4.5 h of DD treatment, embryo phenotypes in A-D could be classified in four categories: **a** normal hatchling (blue), **b** aberrant hatchling (yellow), **c** Pre-tailbud arrest with a golf ball phenotype (salmon), and **d** 1-cell arrest (cherry). Scale bar 50 μm. **e** The proportion of each aberrant phenotype at different DD concentrations allowed us to characterize the dose response and to define four ranges of concentrations depending on their severity: innocuous, mild, moderate and severe. Statistical analyses and individual data plotting are provided in Supplementary Fig. 1. Colors correspond to phenotypes in **a**. Number of analyzed embryos (n) and number replicates (r) are indicated on top of each treatment
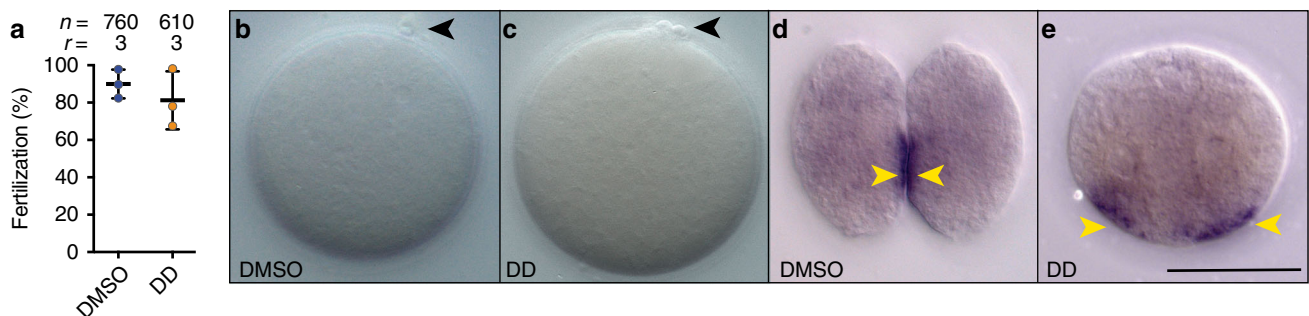
Fig. 2 Characterization of 1-cell arrest phenotype caused by severe DD concentrations (2.5 μg mL^{-1}). **a** Percentage of fertilized embryos assessed by the formation of polar bodies do not show significant differences (P-value = 0.432) between DMSO-control (blue circles in the plot) (**b**) and DD-treated embryos (orange circles in the plot) (**c**), in which moreover no morphological differences in the presence of internal granules nor surface rugosity are observed. Maternal *Wnt11* signal appeared asymmetrically distributed, mostly accumulated near the cell membrane (yellow arrowheads) that will form the prospective posterior vegetal pole both in DMSO-control (**d**) and DD-treated embryos despite the absence of the first cleavage (**e**), suggesting that DD does not alter the cellular architecture of the oocyte that drives primary axial polarity. Number of analyzed embryos (n) and number replicates (r) are indicated on top of each treatment. Reported means and standard deviations are represented in the plot with horizontal line and error bars, respectively. Scale bar 50 μm

(≥2.5 μg mL^{-1}), in which most zygotes remained at 1-cell stage with no cleaving.

One-cell arrest embryos by severe DD concentrations. To test if the absence of cell divisions at severe concentrations could be due to fertilization impairment, we scored for the formation of polar bodies after fertilization in the presence or absence of DD in 760 and 610 embryos, respectively (Fig. 2a–c). Results showed no significant difference in the number of oocytes with polar bodies between DMSO-controls and DD-treatments (ANOVA test, P-value = 0.432), suggesting therefore that DD was not impairing oocyte fertilization. Careful inspection of oocytes by differential interference contrast (DIC) microscopy did not reveal any evident morphological difference, neither in the shape and size of the internal granules of the oocytes, nor in the changes of membrane surface rugosity that normally occurs during fertilization (Fig. 2b, c). To investigate if the internal organization of the cytoplasm of oocytes could be affected by

DD, we performed whole-mount in situ hybridizations to detect maternal *Wnt11* transcripts (Fig. 2d, e). In DMSO-control two-cell stage embryos at 30-min post-fertilization (mpf), maternal *Wnt11* signal appeared asymmetrically distributed, mostly accumulated near the cell membrane that will form the prospective posterior vegetal pole of the embryo (yellow arrowheads in Fig. 2d, and unpublished data). In DD treated embryos at 30-mpf, maternal *Wnt11* signal also appeared asymmetrically distributed near the membranes of the putative posterior vegetal pole, despite the absence of embryo cleavage, suggesting no changes of cellular architecture that drives the primary axial polarity (Fig. 2e).

Golf ball pre-tailbud arrested embryos. At moderate DD concentrations (e.g., 1 μg mL^{-1}), the majority of embryos proceeded normally with embryo cleavage, but most of them never developed a normal tailbud morphology (Figs. 1c, 3a). Comparison between the development of DMSO-control and DD-treated

157
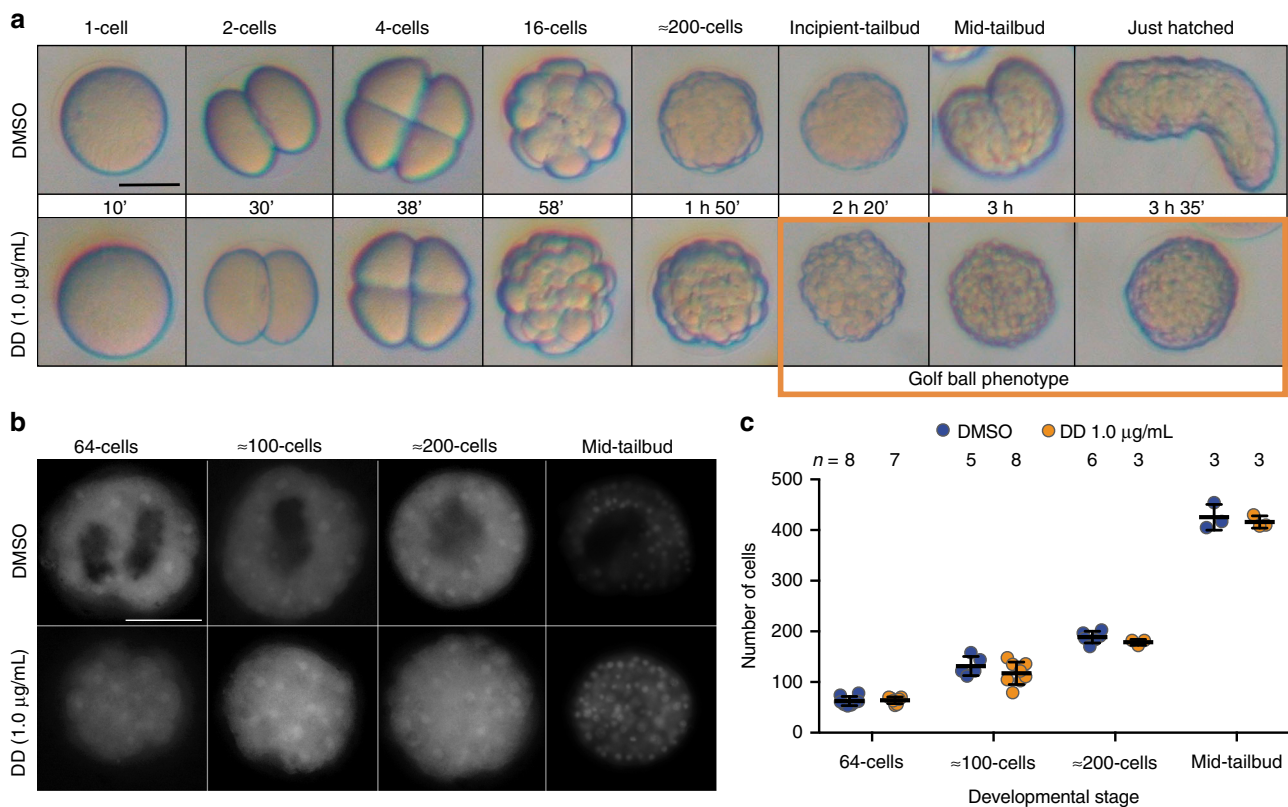
Fig. 3 Characterization of the golf ball morphology at the pre-tailbud arrest stage caused by moderate DD concentrations (1 µg mL^{-1}). **a** Developmental progression of DD-treated embryos does not show any obvious abnormality until 3 hpf, when morphogenesis starts during the formation of tailbud stages in DMSO-control embryos, but a golf ball morphology appears in an arrested pre-tailbud stage of DD-treated embryos (orange square), in which no trunk nor tail can be distinguished. **b** Nuclear staining with HOECHST reveals no significant differences in the number of cells between DMSO-control and DD-treated embryos at 64-cell, 100-cell, 200-cell, and mid-tailbud stage (ns, P-values 0.725, 0.262, 0.206, respectively). Images in **b** correspond to embryos in which *Bra* expression had been detected by whole-mount in situ hybridization, to test whether cell division had not slowed down not even in embryos that showed delay in the expression onset as described in Fig. 4. (**c**), suggesting that DD does not induce a general slow-down or blockage of development by the pre-tailbud arrest that produces the golf ball morphology. Reported means and standard deviations are represented in the plot with horizontal line and error bars, respectively. Scale bars 50 µm

embryos revealed no obvious differences up to the incipient-tailbud stage about 3 h post-fertilization (hpf). At that stage, however, while control embryos showed a clear indentation at the initial demarcation of the trunk and tail, DD-treated embryos did not show that indentation, and their morphology looked like a golf ball (Fig. 3a). During tailbud stages, while control embryos continued with the progressive differentiation of their trunk and tail until the hatch, DD-treated embryos maintained the golf ball morphology in an apparently arrested pre-tailbud stage, in which no trunk nor tail could be distinguished (Figs. 1c, 3a).

To test if the failure to proceed to tailbud stages was due to the arrest of cell divisions, we counted the number of nuclei in control and DD-treated embryos, (Fig. 3b), we found no significant differences (P-values > 0.05, Fig. 3c), suggesting that the absence of trunk and tail differentiation, and the lack of obvious internal structures were not due to a slow-down or blockage of development caused by a deceleration or arrest of nuclear divisions, but to the failure of the process of morphogenesis that normally start at incipient-tailbud stage.

DD causes zygotic expression delay of developmental genes. The fact that *O. dioica* has lost RA signaling[42] allowed us to discard that these abnormalities were due to alterations of this signaling pathway. To uncover what genetic mechanisms could be affected by DD resulting in a failure of morphogenesis and

organogenesis, we performed expression analyses by whole-mount in situ hybridization with eight specific gene markers across different embryonic germ layers comparing six developmental stages between DMSO-control and DD-treated embryos (Fig. 4a). With the exception of *Brachyury* (*Bra*)[47,48], *Actin* (*Act*)[49], and *Nkx2.3/5/6* (ID comp20628 in[50]), all other *O. dioica* genes used in this work have not been described yet, and here they have been used as makers of derivatives of the three germ cell layers (Fig. 4b).

Expression analyses by whole-mount in situ hybridization revealed that the main alteration produced by DD exposure was that most embryos showed a systematic delay in the expression onset of all developmental genes analyzed. For instance, the first signal corresponding to the expression onset of *Bra*, *Act*, *Tis11a*, and *SoxBa* was delayed from 32-cell to at least 64-cell stage, the zygotic expression onset of *Wnt11* was delayed from 64-cell to at least to 200-cell stage, and the expression onset of *SoxBb* was delayed from 200-cell stage to an embryo with a golf ball appearance equivalent to a mid-tailbud stage in the control condition (Fig. 4a). This delay was maintained during development since expression domains at later developmental stages in DD-treated embryos reminded those at earlier stages of control embryos (Fig. 4a). The fact that the total number of cells at each stage did not significantly differ between control and DD-treated embryos (Fig. 3b, c) suggested that the delay of expression could not be explained by an overall developmental slowdown due to a
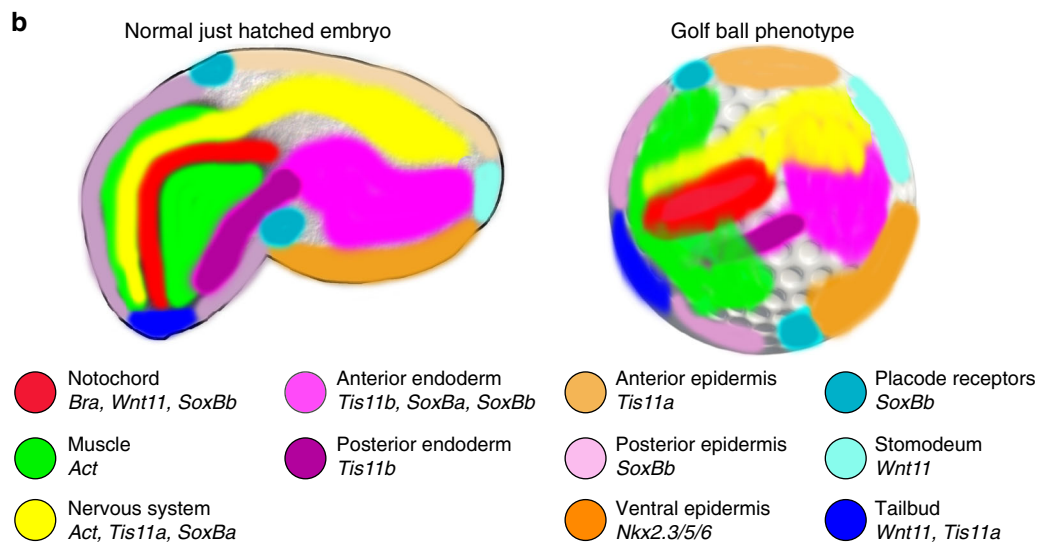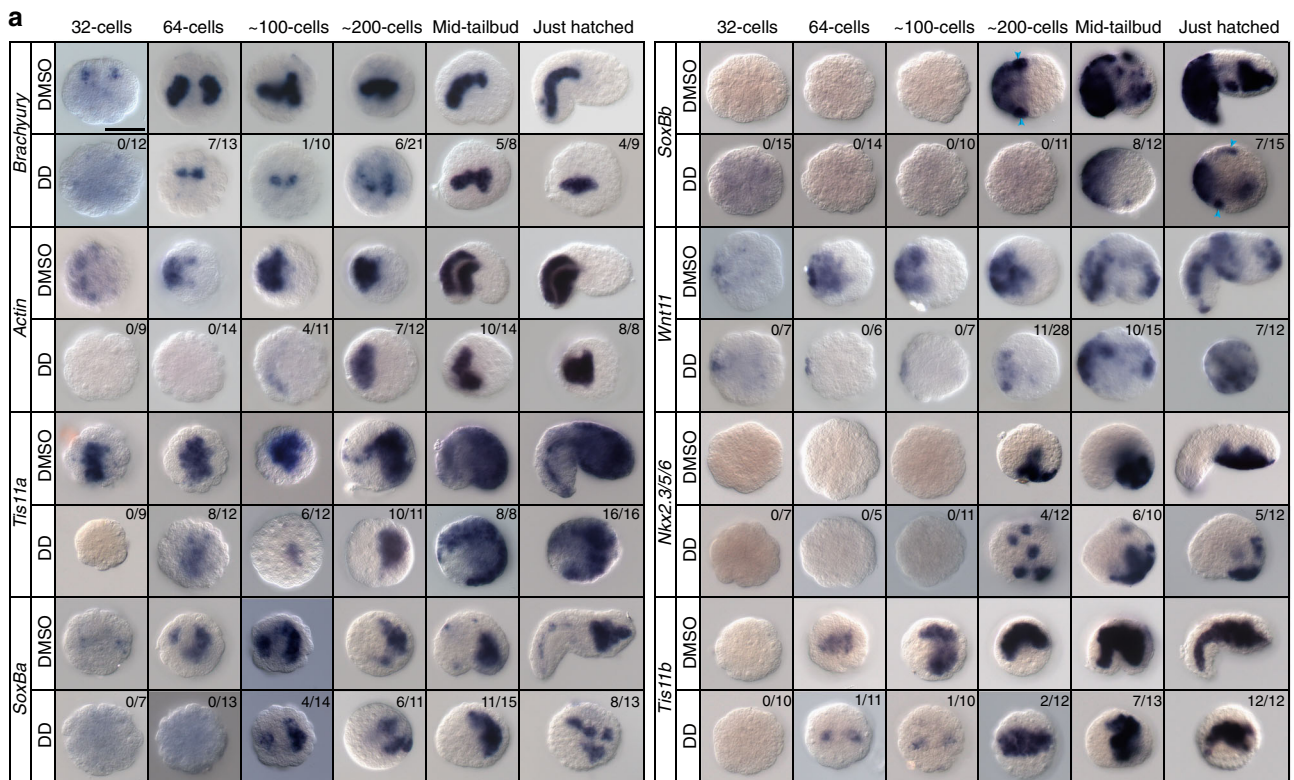
Fig. 4 Alteration of developmental gene expression patterns by DD causing the golf ball morphology. **a** Whole mount in situ hybridizations on DMSO-control embryos and DD-treated embryos at 1 µg mL^{-1} was performed throughout different developmental stages until hatch (i.e., 32-cell, 64-cell, 100-cell, 200-cell, mid-tailbud stage and just hatched) using markers for derivatives of all three germ cell layers: mesoderm (notochord labeled by *Bra, Wnt11* and *SoxBb*, and muscle labeled by *Act*), endoderm (entire endoderm labeled by *Tis11b*, only endoderm of the trunk labeled by *SoxBa* and *SoxBb*), and ectoderm (epidermis of the tail labeled by *SoxBb*, epidermis of the trunk labeled by *Tis11a*, and ventral epidermis of the trunk by *Nkx2.3/5/6*; neural cells by *Tis11a* and *SoxBa*, and presumptive placodal regions by *Wnt11* and *SoxBb*; blue arrowheads label precursor cell of Lagerhand receptors, visible from a dorsal view). Numbers of scored embryos with signal are shown in the top-right corner of each panel. The rest of the scored embryos showed no signals. All DMSO-control embryos presented signal as shown. When possible, lateral views of embryos have been oriented anterior and dorsal to right and top, respectively. Scale bar 50 µm. **b** Colored schematic representation of the developmental gene expression domains in embryos with normal or golf ball phenotype as shown in **a**. Despite the failure to begin the process of morphogenesis of DD-treated embryos with golf ball morphology, at the time in which DMSO-control embryos have hatched, developmental gene markers show that the initial cell fate determination of derivatives of all three germ cell layers is not affected, and embryos show a correct AP and DV axial patterning. The main developmental processes that seem to be mostly affected are the failure of the formation of the indentation that separates trunk and tail, a delayed midline convergence, and the absence of tail elongation

159

decrease of the rate of cell divisions, but rather to alterations in the regulatory mechanisms of the transcription of these genes.

When we scored the number of DD-treated embryos that showed or not expression signal in each stage, we found that the delay of the expression onset could reach extreme situations of detecting no signal at all in a majority of DD-treated embryos at stages in which gene expression was already detected in 100% of their control counterparts (Fig. 4a). Remarkably, while the proportion of DD-treated embryos with no expression between 32-cell and 200-cell stages was high (73%, 210 out of 288), this proportion decreased in golf ball embryos equivalent to tailbud stages (31%, 58 out of 190). To determine whether the absence of signal was due to RNA degradation or repression of zygotic transcription, we focused on the expression signal of *Wnt11* gene, which has both maternal and zygotic components. Results revealed that while signal from the maternal component was observed up to the 64-cell stage in control embryos, time at which zygotic transcription started, in DD-treated embryos, however, only signal from the maternal component was observed and no signal from zygotic expression of *Wnt11* was detected until the 200-cell stage (Fig. 4a). This finding suggested that the absence of expression signal in DD-treated embryos was not due to transcript degradation, but it was consistent with a considerable delay and silencing of zygotic transcription of developmental genes, which was more conspicuous at earlier stages up to 200-cell stage than at later stages equivalent to tailbud in control embryos.

DD alters midline convergence and posterior elongation. To test if the delay of the expression of developmental genes caused by DD was also accompanied by a delay of the onset of cell invagination during gastrulation, we analyzed *Bra* expression, which in agreement with previous reports[47] labeled one pair of blastomeres (A21) at the surface in embryos at 32-cell stage, and two pair of cells that already occupied an internal position (A211 and A212) after invagination as gastrulation proceeded at 64-cell stage in control embryos (Fig. 4a). In DD-treated embryos, however, no signal was detected at 32-cells stage, and the first *Bra* signal was observed in only one pair of internal cells rather than two pairs as expected at the 64-cell stage (Fig. 4a). The fact the first two cells expressing *Bra* were already located in an internal position of the embryo (rather than the surface) suggested that DD did not delay the process of gastrulation, but simply altered the time at which *Bra* started to be expressed.

Analyses of the temporal dynamics of markers such *Bra* and *Act* revealed that the process of midline convergence seemed to be clearly delayed in DD-treated embryos. For instance, while the processes of notochord intercalation and muscle midline bilateral-convergence were completed by 2 hpf at 200-cell stage in control embryos (Fig. 4a), in DD-treated embryos these processes were not completed up to 2 h later by the time at which normal animals were already hatching (Fig. 4b). Moreover, the length of the *Bra* and *Act* expression domains appeared to be much shorter in DD-treated embryos than control ones, suggesting that not only the midline convergence was affected, but also posterior elongation was impaired.

DD alters morphogenesis but no axial patterning or cell fate. Analysis of expression of eight developmental markers in DD-treated embryos with golf ball phenotypes revealed that the setup of the embryonic axes had been already established, and the fate determination of tissues derived from the three germ layers did not appear to be affected. For instance, the setup of the anterior-posterior (AP) axis in embryos with golf ball morphology was revealed by the presence of *Bra*, *Act*, and *SoxBb* expression

domains restricted to the presumptive posterior half of the embryo, (labeling the mesodermal precursor of the notochord, muscle cells and the ectoderm of the tail, respectively, Fig. 4a, b), as well as *SoxBa*, *SoxBb*, *Tis11a*, and *Nkx2.3/5/6* expression domains restricted to the presumptive anterior half of the embryo (labeling the endoderm, and the ectoderm of the trunk Fig. 4a, b). The setup of the dorso-ventral (DV) axis was revealed by the restricted expression of *Nkx2.3/5/6* to the presumptive ventral epidermal domain of the trunk (Fig. 4a, b). In addition to the endodermal, mesodermal and ectodermal precursors labeled with the aforementioned markers in DD-treated embryos, the presence of ectodermal *SoxBb* and *Wnt11* expression domains in precursor cells of the Langerhans receptors in the tail-trunk transition area and in the most anterior part of the embryo (Fig. 4a, b) suggested that the cell fate commitment of placode and neural cell precursors was not affected by DD. Thus, despite that DD seems to block morphogenesis and the formation of the trunk and tail in embryos with a golf ball phenotype, the setup of the embryonic axes, as well as the fate determination of tissues derived from the three germ layers, seemed not to be altered by DD.

Developmental genetic response of the defensome against DD. Previous works on sea urchins had revealed the existence of a defensome, a set of stress response genes that activates their expression during development to protect embryos upon environmental stress[20]. To test if DD upregulates defensome genes during *O. dioica* development, we have analyzed the expression pattern of four defensome genes—three previously described dehydrogenase genes *Adh3*[51], *Aldh2*[52], and *Aldh8*[42], and a here newly described glutathione-related gene (Glutamate-cysteine ligase modifier subunit; *Gclm*)—in six developmental stages from eight-cell stage to late hatchling of DMSO-control and DD-treated embryos at mild concentrations (e.g., 0.1 μg mL^{-1}) (Fig. 5). In control embryos, the expression signal of all four genes was almost negligible at early developmental stages, and it was only detected in mid/late hatchling stages, in which *Gclm* and *Adh3* showed obvious expression domains restricted to digestive compartments (e.g., mouth, pharynx, esophagus, and stomach lobes), and *Aldh2* and *Aldh8* showed either no signal or a weak and scattered signal (Fig. 5). In DD-treated embryos, however, the expression of all four genes was upregulated since very early developmental stages, showing broader and stronger staining than control embryos (Fig. 5). The obvious up-regulation of these four genes in response to DD, therefore, suggested that DD was able to trigger a genetic response of members of the defensome involved in aldehyde detoxification, and especially members that respond to alterations of the metabolism of the glutathione system such as *Gclm*, which was the gene that showed a stronger level of upregulation in response to DD (Fig. 5). Moreover, the fact that the expression of defensome genes was already noticeably upregulated by eight-cell stage, a time at which normal zygotic expression of most genes had not started yet, suggested that these genes had an immediate response capability upon DD challenge (Fig. 5).

Effects of diatom extracts on *O. dioica* embryogenesis. To test if similar developmental abnormalities induced by DD on *O. dioica* could be also caused by natural PUAs or other oxylipins produced by diatoms frequently involved in algal blooms, we tested the effect of the extracts from different microalga species on *O. dioica* embryogenesis. These species included *Skeletonema marinoi* (*Sm*), which has been characterized by its high content of the PUAs heptadienal and octadienal and the lack of decadienal[5,53]; *Chaetoceros affinis* (*Ca*), which produces other oxylipins different from PUAs[8]; *Prorocentrum minimum* (*Pm*), a
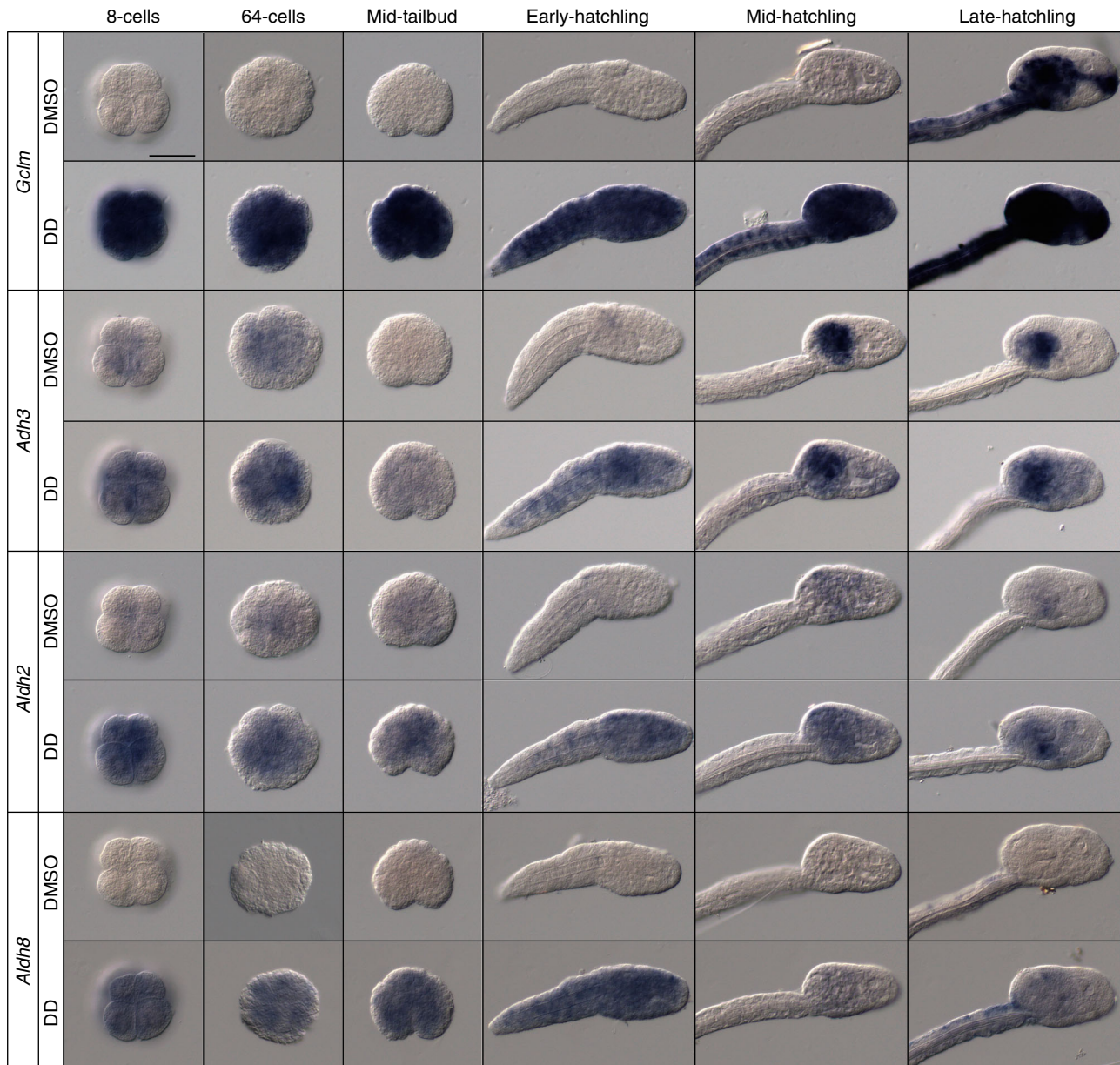
Fig. 5 Defensome gene response to DD during embryo development. Whole mount in situ hybridizations on DMSO-control embryos and DD-treated embryos at $0.1\,\mu g\,mL^{-1}$ was performed throughout different developmental stages until hatch (i.e., 8-cell, 64-cell, and mid-tailbud stages and early-, mid- and late-hatchlings). Expression analysis of *Gclm, Adh3*, *Aldh2*, and *Aldh8* reveals that DD activates a rapid response of the defensome during *O. dioica* development. The number of scored embryos in each stage after DD treatment was between 5 and 10, and in 100% of the cases showed the same expression pattern. Scale bar 50 µm

dinoflagellate species used as a negative control that does not produce oxylipins[8]; and *Chaetoceros calcitrans* (*Cc*), a diatom included in the diet used to feed *O. dioica* in our animal facility[41], but for which the capability to produce PUAs was unknown. First, we studied the dose-response of *O. dioica* development at different concentrations of *S. marinoi* extracts (from 5 to $100\,\mu g\,mL^{-1}$). Results revealed that the most concentrated extracts reproduced the same phenotypes observed during DD treatments (Supplementary Fig. 2). Our results from treatments with *S. marinoi* extracts at $100\,\mu g\,mL^{-1}$—which according to previous studies[53] contained PUAs at a concentration of approximately $2.4\,\mu g\,mL^{-1}$ (20.3 µM assuming an average MW of 117.17 for heptadienal and octadienal most abundant PUAs described in *S. marinoi*)—revealed that the golf ball morphology of pre-tailbud arrested embryos was the most abundant

phenotype, suggesting that the effects of the PUAs in $100\,\mu g\,mL^{-1}$ *S. marinoi* PUAs were comparable with the moderate, but not severe, condition observed in DD-treatments (Fig. 1 and Supplementary Fig. 2). These results were consistent with previous observations showing that higher concentrations of heptadienal and octadienal, the main PUAs of *S. marinoi*, were required to cause similar teratogenic effects to those of DD during sea urchin embryonic development[12].

To test whether treatments with *S. marinoi* extracts were also capable to trigger a defensome response as observed with DD, we performed in situ hybridization experiments with *Gclm*. Results revealed that *Gclm* expression was upregulated even after treatments with extracts at concentrations of only $10\,\mu g\,mL^{-1}$, in which no significant difference was observed in the proportion of aberrant phenotypes between Sm-extract treated embryos and
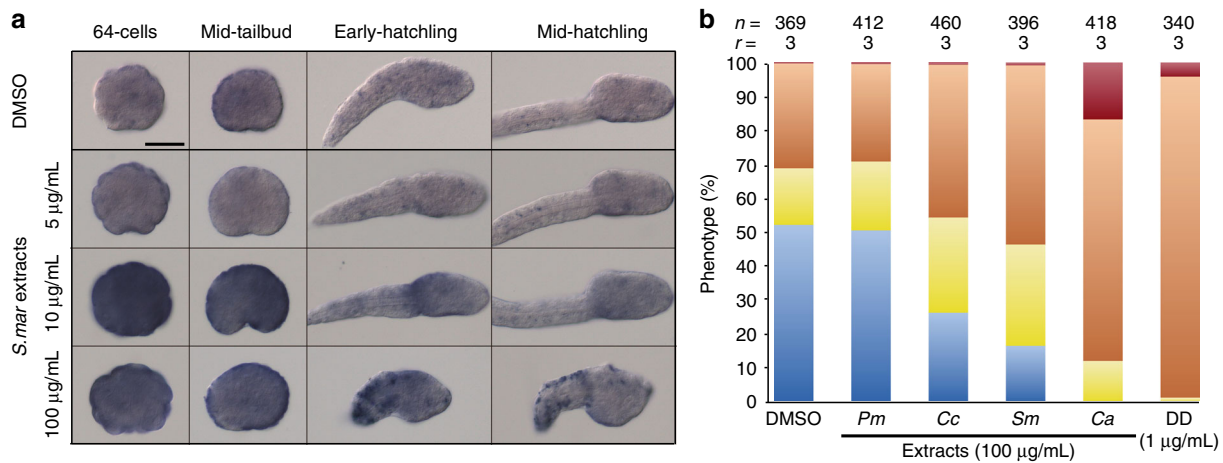
Fig. 6 Effects and defensome response of *Oikopleura dioica* embryos treated with extracts of different diatom species. **a** Whole mount in situ hybridizations with *Gclm* on DMSO-control embryos and embryos treated with extracts of *S. marinoi* at different concentrations suggest that other PUAs than DD (e.g., heptadienal and octadienal) also activates a quick response of the defensome system. Scale bar 50 μm. **b** Embryo phenotypes observed after treatments with extracts of *S. marinoi* (Sm), *C. affinis* (Ca), *C. calcitrans* (Cc) corresponds to the same observed after DD-treatment (Fig. 1). Extract of *P. minimum* (Pm), a dinoflagellate that does not produce aldehydes, was used as additional control. Statistical analyses and individual data plotting are provided in Supplementary Fig. 2

the DMSO-control condition (*P*-value = 0.261; Fig. 6a and Supplementary Fig. 2). We concluded therefore that potential secondary metabolites produced by *S. marinoi* were also capable to activate the expression of the defensome.

Finally, we tested whether *O. dioica* development was also affected by extracts of other microalgae that produce other secondary metabolites different from those of *S. marinoi*, using diatom extracts of *C. affinis* and *C. calcitrans*, as well as the dinoflagellate *P. minimum* as negative control. In all cases, with the exception of the negative control *P. minimum*, extracts affected *O. dioica* development giving rise to the same phenotypes as observed with DD and *S. marinoi* extracts (Fig. 6b and Supplementary Fig. 2). Results revealed that the proportion of normal embryos in *C. affinis* extract treatments seemed even lower than those observed with *S. marinoi* extracts, suggesting that *O. dioica* development could be even more sensitive to the toxicity of other oxylipins different than PUAs (Fig. 6b). In the case of *C. calcitrans*, a diatom whose production of PUAs or other oxylipins had not been investigated to our knowledge, we observed that embryo development was similarly affected by treatments with *S. marinoi* extracts (Fig. 6b). To our knowledge this is the first finding suggesting that *C. calcitrans* is able to produce toxic secondary metabolites, which indeed is consistent with personal observations during the feeding strategy in our *O. dioica* facility in which excess of this alga in the diet made *O. dioica* cultures to decline[41].

Discussion

Understanding the potential impact that global climate change may have on oceans' health, and particularly how the potential increase of harmful algal blooms may impact marine trophic webs is one of the main challenges that our Society is currently facing[1] (see focus areas and agenda of UNESCO-IOC). In this context, results from this work reveal that biotoxins from the oxylipin family that are abundantly released at the end of diatom blooms can compromise embryonic development of appendicularian species, and therefore their reproduction. This finding is ecologically relevant because appendicularians together with copepods, which are also affected by these biotoxins[4], represent the two most abundant components of the mesozooplankton

around the globe, playing a crucial role in the global dynamics of marine food webs[25]. Our analysis of the development of *O. dioica* embryos exposed to DD, which is the most investigated PUA among those synthesized by some bloom-forming diatoms species[3,54], reveals a characteristic set of embryonic aberrant morphologies with frequencies that follow a defined dose-response pattern (Fig. 1). Our work, furthermore, reveals that the same set of embryonic aberrant morphologies is also induced by treatments with extracts from diatom species that produce high levels of heptadienal and octadienal (*S. marinoi*[5,53]), or other oxylipins different from PUAs (*C. affinis*[8]). Also extracts from species such as *C. calcitrans*, whose oxylipin content is still unknown, affect *O. dioica* embryogenesis.

At present, our understanding of how PUAs toxicity released after blooms affect marine organisms in their natural habitats remains uncertain. Previous studies have quantified dissolved PUAs after blooms events in a range from 7.61×10^{-6} to 0.02 μg mL^{-1}[55], questioning laboratory experiments in which the concentrations of PUAs are of an order of magnitude higher than those described in the field[5,54]. Although in our study the most dramatic embryonic malformations mostly appear at concentrations of orders of magnitude higher than in the field (i.e. one-cell arrest at > 2.5 μg mL^{-1} or golf ball morphology at >0.25 μg mL^{-1}), we want to remark that we observe lethal embryonic abnormalities at concentrations as low as 0.075 μg mL^{-1}, which is in the same order of the maximum value measured in the field after diatom blooms (e.g., 0.02 μg mL^{-1} after a bloom of *S. marinoi* in the northwestern Adriatic Sea[55]). Thus, despite we do not see apparent abnormalities in embryos exposed during 4hpf to DD at concentrations lower than 0.05 μg mL^{-1}, we cannot exclude that long or repeated exposure of appendicularians to peaks of PUAs released after blooms (see ref.[54,56]), or even synergistic effect of simultaneous exposure to multiple PUAs may have detrimental effects on appendicularians' biology in natural habitats. This possibility is especially important considering the outstanding ability of appendicularians to concentrate microalgae. Moreover, our finding that longer exposures before fertilization, which facilitate eggs to accumulate DD through the chorion (Supplementary Fig. 3), increase the sensitivity of appendicularian development to DD toxicity makes plausible to suggest that adult animals exposed to DD throughout generations may accumulate this biotoxin in their gonads and become, therefore, more

vulnerable to its detrimental effects. This possibility might explain the observation that overfeeding *O. dioica* cultures with the diatom *C. calcitrans* appears to lead culture declining[41]. Considering the ecological relevance of appendicularians in marine food webs and ocean carbon cycle, future experiments are urged to better understand the threat that oxylipins and other biotoxins can cause on the biology of appendicularians in natural environments. For instance, longer exposure regimes of *O. dioica* adults at low DD doses or overfeeding with excess of oxylipin-producing diatoms will be necessary to further investigate the vulnerability of appendicularians to the accumulation of these biotoxins throughout multiple generations.

Finally, our finding that *Gclm* and other defensome genes show highly sensitive expression responses to low concentrations of PUAs and other oxylipins (Figs. 5, 6) opens the possibility to use *Gclm* and other defensome markers as biosensors to test for genetic stress of natural appendicularian populations that may have been exposed to algal blooms in their habitats. Interestingly, increased expression of *Gclm* has been also reported in *Ciona intestinalis* larvae treated with DD[16], suggesting that the use of *Gclm* expression as a biosensor can be used in other urochordates, and possibly other organisms, to monitor ecosystems with different exposures to blooms, thus providing a better understanding and a prediction of the potential impact of harmful algal blooms associated for instance to ocean warming and acidification.

Despite PUAs and other oxylipins produced by diatoms can interfere with a wide spectrum of physiological processes in many marine animals[9–18], the genetic pathways altered by these biotoxins remain unclear, especially during embryo development. The fact that PUAs such as DD had been described to be able to compete with retinaldehyde for the substrate binding site of the Aldh1a enzymes had led us to hypothesize that genetic alterations due to a disruption of RA signaling could be a potential explanation for the physiological and developmental alterations caused by DD[46]. Although many studies have reported the teratogenic effect of DD on different marine organisms[11,16,57–61], the direct effect of DD on RA signaling pathway has never been investigated in vivo, and whether changes detected by RT-PCR of the expression levels of some typical RA-target genes such *Hox1*, *Hox12*, and *Cdx* in ascidians was due to RA or some other signaling pathway remained unknown[15]. Our study, taking advantage that *O. dioica* is a natural evolutionary knockout of RA signaling[42], demonstrates for the first time that DD can cause dramatic developmental alterations independently of RA signaling.

To our knowledge, this work is the first general attempt to fill the gap between the teratogenic phenotypes and the developmental genetic responses induced by oxylipins by expression analyses of whole-mount in situ hybridization. Our results show which developmental processes are altered and which ones are not affected by DD. Indeed, early developmental processes, such as cleaving pattern, gastrulation, and cell division rate do not seem to be affected by the exposure to moderate concentrations of this biotoxin. DD-treated embryos, however, fail to begin the process of morphogenesis, and develop a golf ball morphology in which the indentation that normally separates the trunk from the tail at incipient-tailbud stage never appears. Molecular characterization of DD-treated embryos reveals a systematic delay of the onset of the expression of developmental genes, completely abolishing in many cases zygotic expression until late developmental stages. The fact that defensome genes highly upregulate their expression as a quick response to the presence of DD, even as early as at the eight-cell stage, when most genes have not yet started normal zygotic transcription, indicates that down-regulation of developmental genes does not result from a general transcriptional repression. On the contrary, it points to the

activation of a silencing mechanism for developmental genes, especially during early stages, that could be part of the embryonic response to environmental challenges.

Despite the absence of morphogenesis, the expression of specific cell markers in golf ball reveals a correct cell fate differentiation of the main derivatives of the ectodermal, mesodermal and endodermal germ layers, as well a correct establishment of AP and DV axes that confer a correct spatial regionalization in DD-treated embryos (Fig. 4b). Analyses of the position and size of expression domains of developmental markers show that midline convergence and tail elongation are notably affected in DD-treated embryos, suggesting that genes regulating these two developmental processes may be linked to the formation of the golf ball morphology. Future transcriptomic profiling by RNAseq and functional approaches will be necessary to uncover the entire set of defensome genes that respond to DD exposure, as well as the developmental gene pathways altered by DD which are responsible of the failure of morphogenesis in *O. dioica*.

Methods

Animal culture and embryo collection. *Oikopleura dioica* embryos were obtained by in vitro fertilization from adults cultured in our animal facility at the University of Barcelona as described in the ref.[41]. Briefly, animals were originally collected in the Mediterranean coast of Barcelona (Catalonia, Spain), and kept in the facility with a life cycle of five days at 19 °C. Animals are cultured in sterile seawater (sSW) filtered by 0.22 μm (VacuCap PF Filters 4622, Pall Corporation), and fed with four microalgae (*Isochrysis sp.*, *C. calcitrans*, *R. reticulata* and *Synechococcus sp*). Mature females and males were collected at day 5, and after natural spawning, eggs were in vitro fertilized, and raised in petri dishes in sterile sea water until the desired stage.

Decadienal treatments. Previous studies about the effect of *trans,trans*-2,4-decadienal (DD, Sigma-Aldrich®, W313505) on ascidians embryos had shown that the chorion could act as a barrier that slow down the access of DD to embryos[14,15]. Since our experience had shown that dechorionation is not possible in *O. dioica* embryos because it lethally affects development, to test if the chorion of *O. dioica* could also be a barrier against DD, we scored the hatching success of embryos that had been pre-incubated in DD previous to fertilization during different times (Supplementary Fig. 3). Thus, we empirically determined that despite the chorion of *O. dioica* also act as a barrier for exogenous DD, 10 min (min) of pre-incubation with DD before fertilization was enough to overcome the barrier. We therefore implemented a 10-min egg pre-incubation as the standard procedure in all our DD treatment experiments.

For DD treatments, pools of *O. dioica* eggs were incubated before fertilization and during development with nine increasing concentrations of DD 0.05, 0.075, 0.1, 0.125, 0.25, 0.5, 1.0, 2.0, and 2.5 μg mL⁻¹ (0.33, 0.49, 0.66, 0.82, 1.64, 3.28, 6.57, 13.14, and 16.42 μM, respectively). Stock solution of 15 mg mL⁻¹ of DD was prepared in dimethyl sulfoxide (DMSO), avoiding light exposure. To minimize the toxic effect of DMSO in the development of *O. dioica*, the stock solution was diluted 1:500 (30 μg mL⁻¹; 0.2% DMSO) in sSW. This diluted stock was used to create each incubation solution containing the corresponding concentration of DD.

Eggs were pre-incubated for 10 min before fertilization into 900 μL of DD incubation solution. Then eggs were fertilized by adding 100 μL of sperm dilution (the sperm of two males diluted into 1 mL of sSW). Five minutes after the in vitro fertilization, fertilized eggs were transferred into a 50 mm-diameter-glass dish with 3 mL of the corresponding incubation solution. Control embryos were incubated in DMSO at 0.02%(v/v) without DD, which corresponded to the experimental condition at the maximum concentration of 2.5 μg mL⁻¹ of DD. In experiments to obtain treated embryos for in situ hybridization, we added DMSO to have the same final concentration of 0.02%(v/v) in all experimental conditions. Embryos were cultured at 19 °C until hatching (≈4 hpf), time at which the effects of DD were scored. At least three replicates (r) containing hundreds of embryos (n) were analyzed for each condition.

Expression analyses and nuclear staining. Developmental gene markers in *O. dioica* were identified by in silico screening of the Oikobase[62], using *Ciona intestinalis* proteins as starting queries for tBLASTn searches against EST and genomic databases. The orthology of proteins was deduced by reciprocal best BLAST search against the *Ciona* genome (KH from Aniseed database). Whole-mount in situ hybridizations using probes of the genes of interest, which were generated by PCR amplification of cDNA with gene-specific primers (Supplementary Table 1), were performed as described in[42]. Embryos were fixed in 4% paraformaldehyde in fix buffer (0.1 M MOPS, 0.5 M NaCl, 2 mM MgSO₄, 1 mM EGTA) for 1 h at room temperature, dehydrated in 70% ethanol and stored at −20 °C. Fixed embryos were dechorionated manually in petri dishes with poly-lysine as described in the ref.[63].

For nuclear staining, fixed whole embryos were rehydrated by washing in PBST during 10 min and were incubated in staining solution (1 µM Hoechst-33342 Invitrogen-62249 in PBST) for 5–10 min. ProLong Live Antifade Reagent (Invitrogen-P36975, dilution 1:100) was added to prevent photobleaching during fluorescence image capture. Nuclei were manually counted by analyzing the images of 10–15 optical sections for each embryo.

Diatom extracts and treatments. Cultures of one liter of the diatoms *Skeletonema marinoi*, *Chaetoceros affinis*, *Chaetoceros calcitrans* and the dinoflagellate *Prorocentrum minimum* were grown up to stationary state to obtain extracts as described in the ref. [53] with minor modifications: microalgal pellets were frozen in liquid nitrogen and stored at −80 °C until cell lysis was performed by sonication. Stock solutions of 500 mg mL^{-1} of each extract were prepared in DMSO, avoiding light exposure. To minimize the toxic effect of DMSO in the embryo development of *O. dioica*, the stock solution was diluted 1:100 in sSW, which was used to create three incubation solutions at 5, 10, and 100 µg mL^{-1}. Treatments were performed as described for DD. Considering the total number of cells of *S. marinoi* calculated from their concentration (i.e., 1 L at 1,650,000 cells mL^{-1}) we perform an approximate calculation of how much PUAs could be present in the extract according to Barreiro et al.[53], in which it has been reported that the total carbon content per cell is 20.7 pg C cell^{-1} and PUAs concentration is 8.5 µg PUAs mg C^{-1}. Dry weight of total organic extracts were weighted in order that treatments could be referred in µg mL^{-1} (i.e., *S. marinoi* culture rendered 12.2×10^3 µg of extract).

Statistical analysis. Statistical analysis was performed using RStudio Version 0.98.507 for Mac OS X. The results were reported as means ± SD (standard deviation) and analyzed by Tukey HSD (honest significant difference) test for multiple comparisons between control and treatments and between treatments, and by ANOVA the pairwise comparisons. Prism GraphPad software (v. 7.0c) was used to generate plots showing the frequencies of the phenotypes in each replicate after treatments with DD and algal extracts.

Data availability. All data generated or analyzed during this study are included in this published article and its supplementary information. The identification numbers of the sequences used to generate the probes for the in situ hybridization analyses listed in Supplementary Table 1 (GSOIDG00000279001, GSOIDG00000756001, GSOIDG00015222001, GSOIDG00010386001, GSOIDG00013526001, GSOIDG00011688001, GSOIDG00003812001, GSOIDG00017080001, GSOIDG00006303001, GSOIDG00000110001, GSOIDG00002220001, GSOIDG00021101001) correspond to the publicly available database Oikobase: https://www.oikoarrays.biology.uiowa.edu/Oiko/

References

1. Wells, M. L. et al. Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful Algae* **49**, 68–93 (2015).
2. Falkowski, P. G. The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynth. Res.* **39**, 235–258 (1994).
3. Miralto, A. et al. Embryonic development in invertebrates is arrested by inhibitory compounds in diatoms. *Mar. Biotechnol. (NY)* **1**, 401–402 (1999).
4. Ianora, A. et al. Aldehyde suppression of copepod recruitment in blooms of a ubiquitous planktonic diatom. *Nature* **429**, 403–407 (2004).
5. Wichard, T. et al. Survey of the chemical defence potential of diatoms: screening of fifty one species for alpha,beta,gamma,delta-unsaturated aldehydes. *J. Chem. Ecol.* **31**, 949–958 (2005).
6. Ianora, A. & Miralto, A. Toxigenic effects of diatoms on grazers, phytoplankton and other microbes: a review. *Ecotoxicology* **19**, 493–511 (2010).
7. Longhurst, A. R. The structure and evolution of plankton communities. *Progress. Oceanogr.* **15**, 1–35 (1985).
8. Fontana, A. et al. Chemistry of oxylipin pathways in marine diatoms. *Pure Appl. Chem.* **79**, 481–490 (2007).
9. Romano, G., Russo, G. L., Buttino, I., Ianora, A. & Miralto, A. A marine diatom-derived aldehyde induces apoptosis in copepod and sea urchin embryos. *J. Exp. Biol.* **206**, 3487–3494 (2003).
10. Caldwell, G. S., Olive, P. J. & Bentley, M. G. Inhibition of embryonic development and fertilization in broadcast spawning marine invertebrates by water soluble diatom extracts and the diatom toxin 2-trans,4-trans decadienal. *Aquat. Toxicol.* **60**, 123–137 (2002).
11. Romano, G., Miralto, A. & Ianora, A. Teratogenic effects of diatom metabolites on sea urchin Paracentrotus lividus embryos. *Mar. Drugs* **8**, 950–967 (2010).
12. Varrella, S. et al. Molecular response to toxic diatom-derived aldehydes in the sea urchin Paracentrotus lividus. *Mar. Drugs* **12**, 2089–2113 (2014).
13. Hansen, E., Even, Y. & Geneviere, A. M. The alpha, beta, gamma, delta-unsaturated aldehyde 2-trans-4-trans-decadienal disturbs DNA replication and mitotic events in early sea urchin embryos. *Toxicol. Sci.* **81**, 190–197 (2004).
14. Tosti, E. et al. Bioactive aldehydes from diatoms block the fertilization current in ascidian oocytes. *Mol. Reprod. Dev.* **66**, 72–80 (2003).
15. Lettieri, A., Esposito, R., Ianora, A. & Spagnuolo, A. Ciona intestinalis as a marine model system to study some key developmental genes targeted by the diatom-derived aldehyde decadienal. *Mar. Drugs* **13**, 1451–1465 (2015).
16. Castellano, I., Ercolesi, E., Romano, G., Ianora, A. & Palumbo, A. The diatom-derived aldehyde decadienal affects life cycle transition in the ascidian Ciona intestinalis through nitric oxide/ERK signalling. *Open Biol.* **5**, 140182 (2015).
17. Adolph, S. et al. Cytotoxicity of diatom-derived oxylipins in organisms belonging to different phyla. *J. Exp. Biol.* **207**, 2935–2946 (2004).
18. Caldwell, G. S., Lewis, C., Pickavance, G., Taylor, R. L. & Bentley, M. G. Exposure to copper and a cytotoxic polyunsaturated aldehyde induces reproductive failure in the marine polychaete Nereis virens (Sars). *Aquat. Toxicol.* **104**, 126–134 (2011).
19. Caldwell, G. S. The influence of bioactive oxylipins from marine diatoms on invertebrate reproduction and development. *Mar. Drugs* **7**, 367–400 (2009).
20. Goldstone, J. V. et al. The chemical defensome: environmental sensing and response genes in the Strongylocentrotus purpuratus genome. *Dev. Biol.* **300**, 366–384 (2006).
21. Lauritano, C. et al. Molecular evidence of the toxic effects of diatom diets on gene expression patterns in copepods. *PLoS ONE* **6**, e26850 (2011).
22. Marrone, V. et al. Defensome against toxic diatom aldehydes in the sea urchin Paracentrotus lividus. *PLoS ONE* **7**, e31750 (2012).
23. Lauritano, C., Procaccini, G. & Ianora, A. Gene expression patterns and stress response in marine copepods. *Mar. Environ. Res.* **76**, 22–31 (2012).
24. Carotenuto, Y. et al. Insights into the transcriptome of the marine copepod Calanus helgolandicus feeding on the oxylipin-producing diatom Skeletonema marinoi. *Harmful Algae* **31**, 153–162 (2014).
25. Gorsky, G. & Fenaux, R. in *The Biology of Pelagic Tunicates* (ed. Bone, Q.) (Oxford University Press, Oxford, 1998).
26. Capitanio, F. L. et al. Seasonal cycle of Appendicularians at a coastal station (38°28´S, 57°41´W) of the SW Atlantic Ocean. *Bull. Mar. Sci.* **82**, 171–184 (2008).
27. Acuña, J. L. et al. Phytoplankton ingestion by appendicularians in the North Water. *Deep Sea Res. Part II* **49**, 5101–5115 (2002).
28. Flood, P. R. & Deibel, D. in *The Biology of Pelagic Tunicates* (ed. Bone, Q.) 105–137 (Oxford University Press, Oxford, 1998).
29. Diego, F. Ân, Ãƒ Ângel, L. Âp-U., Antonio, F. Ân, JosÃƒÂ© Luis, A. Âa & Roger, H. Retention efficiency of 0.2 to 6 µm particles by the appendicularians Oikopleura dioica and Fritillaria borealis. *Mar. Ecol. Prog. Ser.* **266**, 89–101 (2004).
30. Davoll, P. & Youngbluth, M. Heterotrophic activity on appendicularian (Tunicata: Appendicularia) houses in mesopelagic regions and their potential contribution to particle flux. *Deep Sea Res.* **37**, 285–294 (1990).
31. Robison, B. H., Reisenbichler, K. R. & Sherlock, R. E. Giant larvacean houses: rapid carbon transport to the deep sea floor. *Science* **308**, 1609–1611 (2005).
32. Troedsson, C. et al. Effects of ocean acidification, temperature and nutrient regimes on the appendicularian Oikopleura dioica: a mesocosm study. *Mar. Biol.* **160**, 2175–2187 (2013).
33. Seo, H. C. et al. Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. *Nature* **431**, 67–71 (2004).
34. Cañestro, C., Yokoi, H. & Postlethwait, J. H. Evolutionary developmental biology and genomics. *Nat. Rev. Genet.* **8**, 932–942 (2007).
35. Nishida, H. Development of the appendicularian Oikopleura dioica: culture, genome, and cell lineages. *Dev. Growth Differ.* **50**, S239–S256 (2008).
36. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).
37. Cañestro, C., Albalat, R., Irimia, M. & Garcia-Fernandez, J. Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Semin. Cell Dev. Biol.* **24**, 83–94 (2013).
38. Edvardsen, R. B. et al. Remodelling of the homeobox gene complement in the tunicate Oikopleura dioica. *Curr. Biol.* **15**, R12–R13 (2005).
39. Cañestro, C., Postlethwait, J. H., Gonzàlez-Duarte, R. & Albalat, R. Is retinoic acid genetic machinery a chordate innovation? *Evol. Dev.* **8**, 394–406 (2006).
40. Denoeud, F. et al. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* **330**, 1381–1385 (2010).
41. Marti-Solans, J. et al. Oikopleura dioica culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. *Genesis* **53**, 183–193 (2015).
42. Marti-Solans, J. et al. Coelimination and survival in gene network evolution: dismantling the RA-signaling in a chordate. *Mol. Biol. Evol.* **33**, 2401–2416 (2016).
43. Cunningham, T. J. & Duester, G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat. Rev. Mol. Cell Biol.* **16**, 110–123 (2015).

44. Rhinn, M. & Dolle, P. Retinoic acid signalling during development. *Development* **139**, 843–858 (2012).

45. Duester, G. Retinoid signaling in control of progenitor cell differentiation during mouse development. *Semin. Cell Dev. Biol.* **24**, 694–700 (2013).

46. Bchini, R., Vasiliou, V., Branlant, G., Talfournier, F. & Rahuel-Clermont, S. Retinoic acid biosynthesis catalyzed by retinal dehydrogenases relies on a rate-limiting conformational transition associated with substrate recognition. *Chem. Biol. Interact.* **202**, 78–84 (2013).

47. Bassham, S. & Postlethwait, J. Brachyury (T) expression in embryos of a larvacean urochordate, Oikopleura dioica, and the ancestral role of T. *Dev. Biol.* **220**, 322–332 (2000).

48. Nishino, A., Satou, Y., Morisawa, M. & Satoh, N. Brachyury (T) gene expression and notochord development in *Oikopleura longicauda* (Appendicularia, Urochordata). *Dev. Genes Evol.* **211**, 219–231 (2001).

49. Nishino, A., Satou, Y., Morisawa, M. & Satoh, N. Muscle actin genes and muscle cells in the appendicularian, Oikopleura longicauda: phylogenetic relationships among muscle tissues in the urochordates. *J. Exp. Zool.* **288**, 135–150 (2000).

50. Onuma, T. A., Matsuo, M. & Nishida, H. Modified whole-mount in situ hybridisation and immunohistochemistry protocols without removal of the vitelline membrane in the appendicularian Oikopleura dioica. *Dev. Genes Evol.* **227**, 367–374 (2017).

51. Cañestro, C., Albalat, R. & Postlethwait, J. H. Oikopleura dioica alcohol dehydrogenase class 3 provides new insights into the evolution of retinoic acid synthesis in chordates. *Zool. Sci.* **27**, 128–133 (2010).

52. Cañestro, C. & Postlethwait, J. H. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev. Biol.* **305**, 522–538 (2007).

53. Barreiro, A. et al. Diatom induction of reproductive failure in copepods: the effect of PUAs versus non volatile oxylipins. *J. Exp. Mar. Biol. Ecol.* **401**, 13–19 (2011).

54. Bartual, A. et al. Polyunsaturated aldehydes from large phytoplankton of the Atlantic Ocean surface (42 degrees n to 33 degrees s). *Mar. Drugs* **12**, 682–699 (2014).

55. Vidoudez, C., Casotti, R., Bastianini, M. & Pohnert, G. Quantification of dissolved and particulate polyunsaturated aldehydes in the Adriatic sea. *Mar. Drugs* **9**, 500–513 (2011).

56. Ribalet, F. et al. Phytoplankton cell lysis associated with polyunsaturated aldehyde release in the Northern Adriatic Sea. *PLoS ONE* **9**, e85947 (2014).

57. Romano, G., Costantini, M., Buttino, I., Ianora, A. & Palumbo, A. Nitric oxide mediates the stress response induced by diatom aldehydes in the sea urchin Paracentrotus lividus. *PLoS ONE* **6**, e25980 (2011).

58. Ianora, A. et al. Impact of the diatom oxylipin 15S-HEPE on the reproductive success of the copepod Temora stylifera. *Hydrobiologia* **666**, 265–275 (2011).

59. Ka, S. et al. Impact of the diatom-derived polyunsaturated aldehyde 2-trans,4-trans decadienal on the feeding, survivorship and reproductive success of the calanoid copepod Temora stylifera. *Mar. Environ. Res.* **93**, 31–37 (2014).

60. Caldwell, G. S., Lewis, C., Olive, P. J. & Bentley, M. G. Exposure to 2,4-decadienal negatively impacts upon marine invertebrate larval fitness. *Mar. Environ. Res.* **59**, 405–417 (2005).

61. Lewis, C., Caldwell, G. S., Bentley, M. G. & Olive, P. J. W. Effects of a bioactive diatom-derived aldehyde on developmental stability in Nereis virens (Sars) larvae: an analysis using fluctuating asymmetry. *J. Exp. Mar. Biol. Ecol.* **304**, 1–16 (2004).

62. Danks, G. et al. OikoBase: a genomics and developmental transcriptomics resource for the urochordate Oikopleura dioica. *Nucleic Acids Res.* **41**, D845–D853 (2013).

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

Acknowledgements

Author contributions

N.T-A. carried out treatments, whole-mount in situ hybridizations, phenotype characterizations. N.T-A. and V.R. performed data plotting and statistical analyses. N.T-A. and J.M-S. made microalgal extractions under the supervision of A.P., G.R. and S.D. N.T-A., J.M-S., A.F-R. and A.A. contributed with cloning and expression analyses of developmental and defensome markers. N.T-A. and C.C. interpret the data and made the figures. N.T-A., R.A., and C.C. wrote the MS with input from A.P., G.R. and S.D. C.C. conceptualized the project. R.A. and C.C. supervised the project. All authors commented on the manuscript and agreed to its final version.

Additional information

Supplementary information accompanies this paper at https://doi.org/10.1038/s42003-018-0127-2.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

CHAPTER IV: SUPPLEMENTARY INFORMATION
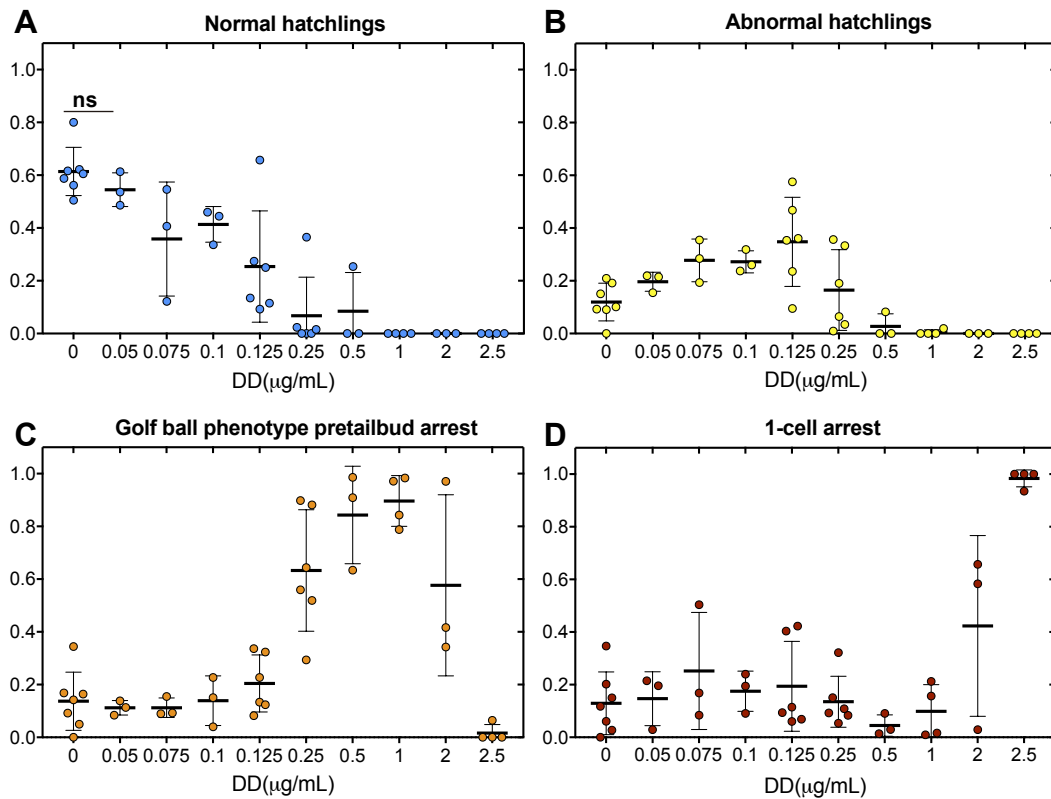
Supplementary Figure 1. Frequency of *Oikopleura dioica* embryo phenotypes after treatments with DD at different concentrations. **(A)** normal hatchling, **(B)** aberrant hatchling, **(C)** pre-tailbud arrest with a "golf-ball" phenotype, and **(D)** 1-cell arrest. For each condition, circles represent the frequency of the phenotype in each replicate individually analyzed (see **Fig 1E**). Reported means and standard deviations are represented with horizontal line and error bars, respectively. DD concentration at 0.05 μg/mL was considered innocuous, since no significant difference was observed in the amount of normal hatchlings in comparison with the DMSO-control condition (ns, P-value 0.99843). At higher concentrations than 0.05 μg/mL we started to observe severe abnormalities that normally we do not observe in DMSO-control conditions (see **Fig 1B**; P-values 0.1522 and <0.001, for 0.075 and 0.125 μg/mL, respectively).
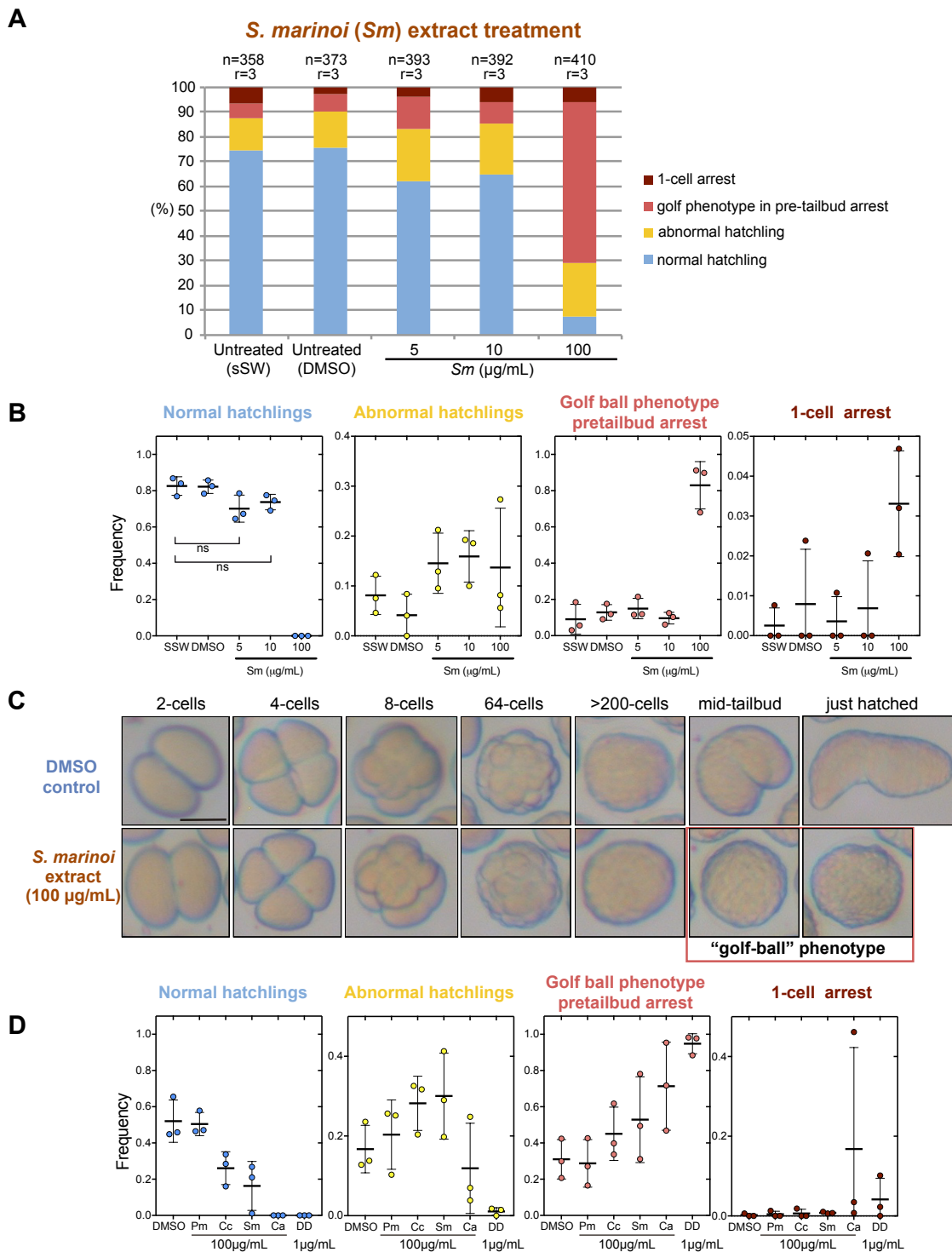
Supplementary Figure 2. Effects and dose-response of *Skeletonema marinoi* extracts on *Oikopleura dioica* embryo development. (A) After 4.5 hours of treatment with extracts of *S. marinoi* (*Sm*) at different concentrations (5, 10 and 100 μg/mL, which may correspond to PUAs approximate concentrations of 1, 2 and 20 μM, respectively), abnormal embryo phenotypes were the same as those obtained after DD-treatments (see **Fig 1**: normal hatchling (blue), aberrant hatchling (yellow), Pre-tailbud arrest with a golf ball phenotype (salmon), and 1-cell arrest (cherry)). The frequency of each abnormal phenotype depended on the extract concentration of the treatment. Number of analyzed embryos (n) and number replicates (r) are indicated on top of each treatment. **(B)** Plots of phenotype frequencies throughout treatments at different concentrations of *S. marinoi*. For each condition, circles represent the frequency of the phenotype in each replicate individually analyzed in (**A**). Reported means and standard deviations are represented with horizontal line and error bars, respectively. No significant differences were observed between embryos developed in sSW or DMSO (P-value 0.9996). Extracts at 5 and 10 μg/mL did not show significant differences with DMSO-control (ns, P-value 0.0573 and 0.2301 for normal hatchling phenotype, respectively). **(C)** The developmental progression of embryos treated with *S. marinoi* extract at 100 μg/mL reproduced the same alterations observed with DD at moderate concentrations (see **Fig 2**). Thus, no obvious abnormalities are observed until 3 hpf, time at which morphogenesis starts during the formation of tailbud stages in DMSO-control embryos, but a golf ball phenotype appears in arrested pre-tailbud stage of treated embryo. Scale bar 50 μm. **(D)** Plots of phenotype frequencies throughout treatments with different microalgal extracts at 100 μg/mL – *S. marinoi* (*Sm*), *C. affinis* (*Ca*), *C. calcitrans* (*Cc*) and *P. minimum* (*Pm*), and DD at 1 μg/mL. For each condition, circles represent the frequency of the phenotype in each replicate individually analyzed in **Fig 6B**. Reported means and standard deviations are represented with horizontal line and error bars, respectively.

167
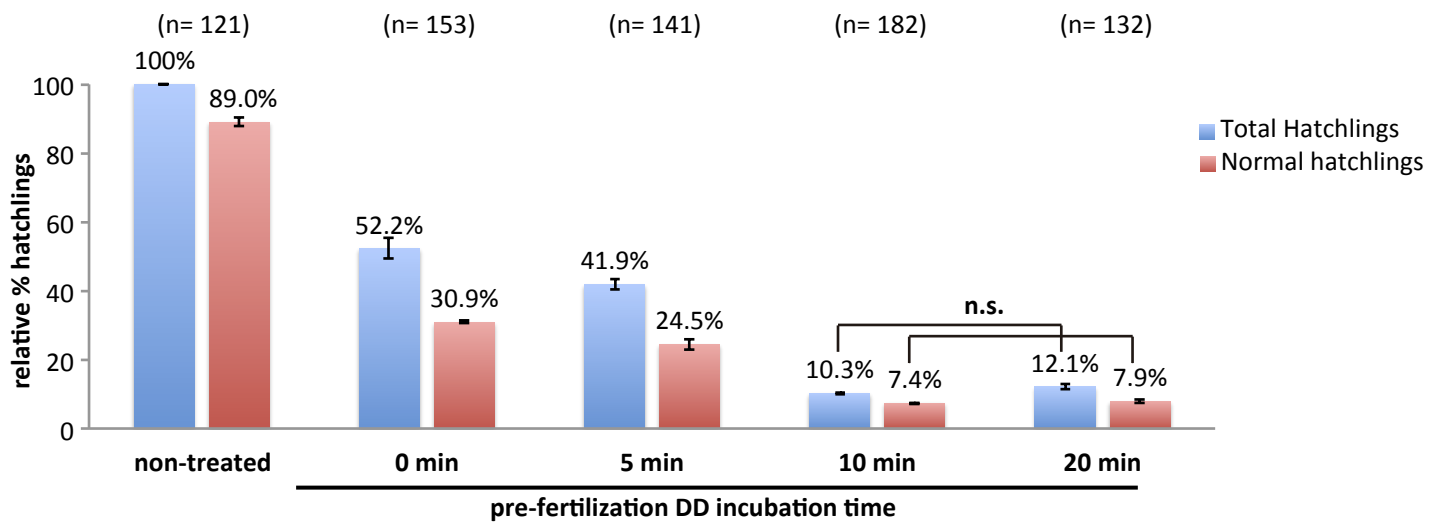
Supplementary Figure 3. Determination of the requirement of pre-incubation time of oocytes during DD treatments. Using a DD concentration of 0.28 *µg/mL* (1,84 *µM)*, oocytes were pre-incubated with DD during 0, 5, 10 and 20 minutes before fertilization, in at least 2 replicates, and at least 100 embryos analyzed (n) in each condition. After fertilization and continuous DD-treatment, we scored the percentage of animals that hatched (blue), and from those what proportion had a normal morphology (red). Error bars correspond to standard deviations. Results showed that the relative abundance of animals that hatched and those that had normal morphology in relation to non-treated embryos (negative control) was significantly lower when pre-incubations were shorter than 10 minutes. No significant (n.s.) decrease of hatch or normal morphologies was observed with longer pre-incubation times (P-values 1.0 and 0.99992, respectively). We therefore always used pre-incubation times of 10 minutes as the standard procedure in our treatment experiments.

Supplementary Table 1. Developmental and defensome gene markers.

Oikobase IDs (http://oikoarrays.biology.uiowa.edu/Oiko), names and sequences.

Gene	Gene ID	Primer name	Sequence (5'- 3')
Brachyury	GSOIDG00000279001	OdiT_F3	GGTTCGCACTGGATGAAACAGCC
		OdiT_R3	TATCCGTTCTGACACCAGTCGTTC
Actin	GSOIDG00000756001	Actin_5'_F	GTCCCCGCCATGTACGTCTG
		Actin_3'_R	GCATCGGAATCGCTCGTTACCA
Tis11a	GSOIDG00015222001	ZFC3Ha_e5_F	GGGTACTGCCCATATGGCG
		ZFC3Ha_e7_R	GCTCGAAGTTGGGCAGCTG
SoxBa	GSOIDG00010386001	SoxBa_e6_F	GCAGAAGTACCCAGCAAGGA
		SoxBa_e6_R	GTGACCACTTTCGGGCTTGT
SoxBb	GSOIDG00013526001	SoxBb_e4_F	GTTGTCGCTGGAAATGGCGA
		SoxBb_e8_R	CTCGACACGGACGCTCTGAT
Wnt11	GSOIDG00011688001	Wnt_s4_atgF	ATGAAGATTTCAGTCACCCTTTTCTCTG
		Wnt_s4_stopR	GTTATTTGCATATATGAGTGACAGTCG
Nkx2.3/5/6	GSOIDG00003812001	Nkx_5'start_F	GACCGAAAAATTACAACTATGAGC
		Nkx_ex_3_R	GCTGTAGCGCCGAGCTCAC
Tis11b	GSOIDG00017080001	ZFC3Hb_e2_F	GGCCAAATGAACGACGAAATCG
		ZFC3Hb_e4_R	GCACTCGGAGAGCGAGAG
Glcm	GSOIDG00006303001	Glcm_e2_F	GCAATAATTATCCAAGATGCCATG
		Glcm_e5_R	GTTCAGTCCTGCAAAGTATCC
Adh3	GSOIDG00000110001	Adh3_F1	CGTCGGTAAAGTGATCACGTGCA
		Adh3_R1	GCGCCCTGTGTAGCTCGGAC
Aldh2	GSOIDG00002220001	Aldh2_F	TGGAACTTCCCTCTCCTCATGCA
		Aldh2_R	TTATTTGGCGTATTGAGGAAGTTTCAT
Aldh8a1	GSOIDG00021101001	Aldh8_F	GATTTAAACAAAAAATGGAGCCGATTG
		Aldh8_R	TTTATTGCTTGCTAATTGTATTAGCTTGAAG

DISCUSSION

1 EVOLUTIONARY STUDY OF GENE LOSS

The overall objective of this PhD project has been to better understand the impact of gene loss on the evolution and the generation of biodiversity, in particular, on the evolution of the mechanisms of development in the field of EvoDevo and comparative genomics in chordates.

In the past, the influential work by Susumo Ohno proposed gene gains by gene duplication as one of the most important evolutionary driving forces (Ohno, 1970). However, upon sudden changes of selective pressures on a population, it has been shown that loss of function mutations are more likely to happen than gene duplications, suggesting therefore that gene loss could also be a major driving force of evolution (Olson, 1999). The fixation of null-mutations has traditionally been considered selectively neutral in a context of gene redundancy after an event of gene duplication or in a context of regressive evolution. However, many studies in bacteria and yeast provide solid evidence that gene loss can be a major adaptive evolutionary force specially relevant when populations are exposed to sudden changes of environmental conditions (rev. in Albalat and Cañestro,

2016). The pervasiveness of gene loss revealed by the bloom of genomics, therefore, suggests that gene losses can play a major adaptive role during the evolution and divergence of different species and major taxonomical groups (Albalat and Cañestro, 2016; De Robertis, 2008). Despite the many examples of gene losses in animal species (reviewed in Albalat and Cañestro, 2016), animals have always been less suitable for comprehensive studies than bacteria or yeast. Therefore, our overall understanding of the impact of gene loss in the context of animal EvoDevo is still scarce.

In this thesis, as a case study, we have focused on how gene losses have impacted the evolution of the cardiopharyngeal gene regulatory network (GRN) in *Oikopleura dioica*, a successful gene loser (Ferrández-Roldán et al., 2019). One of the major problems we have faced during this project is that the inference of a gene loss is based on the negative evidence of the presence of a gene in the genome database of the analyzed species. Genome surveys are fundamentally based on BLAST searches, and therefore the first problem to address has been to test if any of the first hits were indeed homologous or not to the queried gene. Establishing gene

homologies is not trivial, and it requires distinguishing orthologs, originated by speciation, from paralogs, originated by gene duplication, or even a combination of both that would give rise to co-orthologs and pro-orthologs. Moreover, many evolutionary processes can obscure the recognition of true relationships of homology, especially if gene duplications and losses are both simultaneously involved in the evolution of a gene family. For instance, complementary gene loss of two paralogs in two different lineages could result in apparent orthologs, while indeed they are paralogs. Even genes from two different families could appear as orthologs if they share a conserved domain common to an ancestral gene superfamily. To overcome these problems and to try to distinguish different types of homologies between the query and the blast results, it has been necessary to combine best reciprocal blast hits (BRBH) (Fig. D1), using an outgroup species outside the targeted group (e.g. tunicates and vertebrates) with an exhaustive phylogenetic reconstruction of the gene family of interest using close related gene families as outgroups.

A second problem of the study of gene losses is the quality and completeness of genome databases. Low-quality assemblies or shallow databases in which part of the genome might not have been assembled could lead to mistakenly infer false gene losses. To overcome this problem, it is recommendable to combine searches of both independent genome and transcriptomic databases, as well as databases of phylogenetically related species that may share the same ancestral gene loss. A wide sampling of species, moreover, is also useful to overcome the limitation due to sequence divergence, specially to recognize variable regions outside conserved domains.

In our genome survey to find the homologs of *O. dioica* to the members of cardiopharyngeal GRN (**Chapter 1 and 2**), we initially did BLAST searches in the reference genome made from specimens from Norway (http://oikoarrays.biology. uiowa.edu/Oiko/). In cases in which we did not find any significant match or the BRBH results between *O. dioica, C. robusta,* and *H. sapiens* or the phylogenetic analyses suggested the possibility of gene loss, we performed additional searches in other
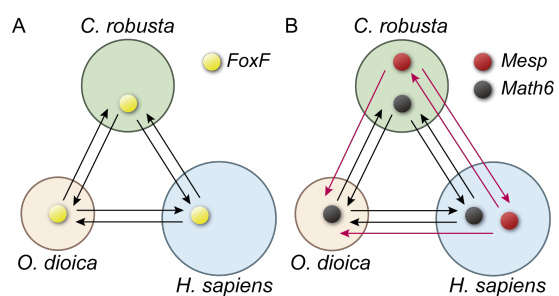
Fig. D1. **Example of the BRBH done between *C. robusta*, *O. dioica* and *H. sapiens***. (**A**) BRBH of *FoxF* (yellow ball), in which we identified orthologs in all three species. (**B**) BRBH of *Mesp* (red ball) that shows orthologs in *C. robusta* and *H. sapiens* but its loss in *O. dioica* because the best hit in the latter is *Math6* (black ball) that maintains orthologs in all three species. Arrows represent the searches between genomes. Circles represent the genomes of the different species.

databases such as transcriptomic data from Norway, as well as newly assembled genomes and transcriptomes generated by our team and collaborators in Luscombe's lab from specimens collected in Barcelona, Okinawa, Oregon, and Osaka. Moreover, during the course of my thesis, the genomes of six other appendicularian species were sequenced (Naville et al., 2019) and 11 ascidian genomes were available in Aniseed (https://www.aniseed.cnrs.fr). BLAST searches in these new genomes corroborated the gene losses we had originally found in *O. dioica*, and they extended the absence of the genes to all other appendicularian species, at the same time that confirmed the presence of all cardiopharyngeal genes in all analyzed ascidian species, validating our methodological approach of our genome survey. These results provided therefore robust evidence supporting gene losses and allowed us to conclude that in most cases the losses had occurred at the base of the appendicularian lineage after their split from ascidians and before the radiation of appendicularian species.

Phylogenetic analyses in tunicates were difficult in general due to their high evolutionary rates that generated long branches as well as the artifactual phenomena of long-branch attraction, which in some cases resulted in low support of some nodes of the trees (Felsenstein, 1978; Hendy and Penny,

1989). The wide sampling of tunicate species (7 appendicularians and 11 ascidians) not only facilitated the identification of divergent gene regions and the reannotation of errors in the reference genome of *O. dioica*, but it also rendered nodes with a higher support. Including three species of cephalochordates in the phylogenetic analyses, as well as slow evolving vertebrates such as the coelacanth (*L. chalumnae*) and the spotted gar (*L. oculatus*), besides *H. sapiens*, also improved the support values in the phylogenetic inferences. When needed, we also surveyed and included genes from the nearest gene families, which could be ancestral paralogous families or other families within the same superfamily, to provide a wider phylogenetic perspective and evaluate potential cases of long-branch attraction artifacts. Interestingly, the phylogenetic analyses revealed some cases in which gene losses had been accompanied by the duplication of paralogs (e.g., the loss of *Gata4/5/6* was accompanied by the duplications of the *Gata1/2/3* up to four paralogs *a-d* in appendicularians, or the loss of *Ets1/2a* was accompanied by the duplication of *Ets1/2b* into *Ets1/2b1* and *Ets1/2b2* in appendicularians). These findings added additional complexity to our study of gene losses because we had to test for the possibility that the loss of the gene could be accompanied by a function shuffling

among paralogs of related gene families (Cañestro et al., 2009). In our case, however, we did not detect any case of functions shuffling that could have allowed another related gene to provide the cardiopharyngeal function of the lost gene (**Chapter 2**).

The identification of so many gene losses in the cardiopharyngeal GRN made us wonder if these losses were not stochastic, but they could be the result of a process of gene co-elimination (Aravind et al., 2000; Koonin et al., 2004; Martí-Solans et al., 2016). Co-elimination can be the result of the dismantling of a functionally linked pathway within a gene network, except for those occupying 'hubs' of the network due to their pleiotropic roles involved in other functions. In such cases, pleiotropic 'hub' genes are not lost, but they lose their subfunctions related to the co-eliminated pathway, becoming less pleiotropic. In our case, the detection of so many losses of cardiopharyngeal genes (e.g., *Mesp*, *Ets1/2a*, *MEK1/2*, *Gata4/5/6*, *Tbx1/10*, *Raldh1a*, *Cyp26*, *RAR*) and cardiopharyngeal subfunctions (*FoxF*, *Dach*, *MyoD*, *Ebf*, and FGF- and BMP-signaling) prompted us the hypothesis that we had discovered a phenomenon of gene co-elimination (**Chapter 2**). During this PhD, we have shown that the study of gene losses can be useful to recognize functionally linked genes and modules within the cardiopharyngeal GRN as well as

to understand how the dismantling (or deconstruction as I discuss in the next section) of the GRN has impacted on the evolution of appendicularians and the biodiversification of tunicates.

2 THE STUDY OF EVENTS OF EVOLUTIONARY DECONSTRUCTION ALLOWS US TO UNDERSTAND GRN ARCHITECTURE AND TO IDENTIFY FUNCTIONAL MODULES

Deconstruction is not synonymous with destruction. Deconstruction is a concept that was first coined by the French philosopher Jacques Derrida arguing that the different meanings of a text (or body of knowledge) can be discovered by dissecting or breaking down the structure into pieces. This way, one can recognize "binary" and "hierarchical" relationships, and therefore to infer how each part or "module" contributes to the construction of the whole. Since the 60s, this term has inspired a wide variety of humanity disciplines including, law, historiography, sociolinguistics, or anthropology. Even the idea behind this concept has been adopted by architecture, in which deconstructivism current designed buildings giving the impression of fragmentation (**Fig. D2 A**); or by cuisine, in which chefs like Ferran Adrià pioneered a new style by boiling down a traditional dish to the core parts to then rebuild it again in a physical way unlike the original but preserving its flavors (**Fig. D2 B**); or even by Woody Allen in his acclaimed movie *Deconstructing Harry* (1997), in which the plot tells the story of the main character in broken events describing

Fig. D2. (**A**) The Guggenheim building in Bilbao an example of deconstructivism (https://www.Guggenheim-bilbao.eus/el-edificio). (**B**) Deconstruction of a potato omelet (https://macrecetas.wordpress.com/2013/12/23/tortilla-de-patata-deconstruida/). (**C**) Theatrical release poster of the film deconstructing Harry.

different episodes of his life to become alienated from himself (**Fig. D2 C**). Outside humanities, in the field of developmental biology, the concept of deconstruction has been also used (Hogan, 2004), becoming for instance the focus of attention in a Keystone conference on vertebrate organogenesis in Santa Fe (February 2004), in which under the topic of "Deconstructing the genesis of animal form" each session was subdividing different anatomical systems. In this conference, the keynote speaker Mark Krasnow (Stanford University) and others argued that the only way to identify all the genes controlling the development of a complex organ is to break the process down into simpler events, evoking the concept of

deconstructing embryonic development into simple morphometric modules. In this thesis, I will extend the use of the concept of deconstruction to the field of EvoDevo to describe the process of the dismantling of modules of a GRN by the co-elimination of genes and subfunctions intimately related to a particular developmental function that has profoundly changed or even lost during the evolution of a particular lineage. We propose, moreover, that the study of evolutionary deconstructions can be useful to recognize modules of GRN related to specific developmental functions that might have diverged during the evolution of different lineages.

Our results characterizing the deconstruction of the cardiopharyngeal GRN in appendicularians have helped us to recognize two different genetic modules implicated in the cardiopharyngeal GRN of tunicates: the early multipotent (EM) module and the late multipotent (LM) module (Fig. 7 in **Chapter 2**). In appendicularians, the clear pattern of gene co-elimination (*Raldh1a*, *Cyp26*, *RAR*, *MEK1/2*, *Mesp*, *Ets1/2a*, *Gata4/5/6*) affecting the early cardiopharyngeal GRN helped us to define the EM module (Fig. 7 in **Chapter 2**). In ascidians, the EM module is responsible for the separation of ATM from the trunk ventral cells (TVCs) that are maintained as multipotent cardiopharyngeal state and finally migrate to the ventral part of the trunk (**Fig. D3**). The
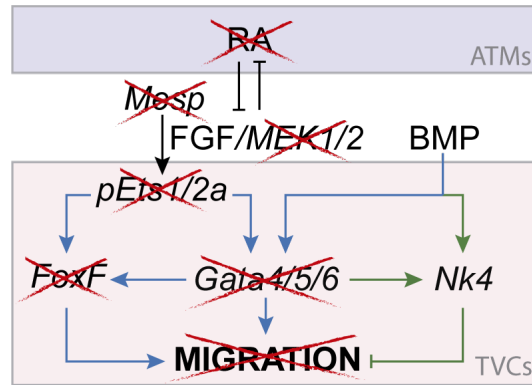
Fig. D3 Schematic representation of the co-elimination of the Early Multipotent module in *O. dioica*. In *C. robusta*, the antagonism between RA and FGF determine the fate of B7.5 granddaughters. At the beginning of migration of the TVCs, *Ets1/2a* and low levels of BMP activate the expression of *Gata4/5/6* that together with *FoxF* prompt migration. At the end of migration, *Gata4/5/6* and high levels of BMP activates *Nk4* that inhibits migration. The genes lost in *O. dioica* or not expressed in cardiac precursors (*FoxF*) are crossed out. Blue arrows represent interactions at the beginning of the migration. Green arrows indicate interactions at the end of migration. TVCs, trunk ventral cells; ATMs, anterior tail muscle cells.

EM module functions at the interface of tail and trunk, under the posterior influence of RA on the differentiation of the axial muscle fate, and the anterior influence of the FGF/MEK1/2 signaling that phosphorylates Ets1/2a, previously upregulated by *Mesp* in the TVCs (Christiaen et al., 2008; Davidson et al., 2006; Nagatomo and Fujiwara, 2003). This phosphorylation starts a *Gata4/5/6* and *FoxF*-dependent ventral migration of the TVCs, which follows a gradient of BMP originated in the trunk ventral epidermis and culminates with the activation of *Nk4* and TVCs divisions (Beh et al., 2007; Bernadskaya et al., 2019; Christiaen et al., 2010; Ragkousi et al., 2011). In *O. dioica*, this co-elimination

pattern, together with the loss of the *FoxF* subfunction in cardiac development, the early activation of *Nk4* in the newborn first cardiac precursors (FCPs) close to the tail muscle cells instead of the ventral epidermis, and the FCPs divisions before meeting the ventral region of the trunk suggest alternative mechanisms to induce cardiac development and to position cardiac progenitors in the ventral region of the trunk (**Fig. D3**). First, regarding the induction mechanism in charge of *Nk4* activation in *O. dioica*, our inhibitory results suggest that neither FGF nor BMP signaling pathways are involved, which opens two scenarios: either there are other signaling pathways as Wnt or Notch implicated as happening in other chordates (Krieg and Warkman, 2015), or FCPs have become self-autonomous for *Nk4* activation. Second, regarding an alternative mechanism for cardiac precursors migration, due to the small distance between the anterior region of the tail and the ventral region of the trunk in such a small organism, cardiac precursors might cover that small distance by cellular rearrangements during cell divisions in a process similar to gastrulation (Stach et al., 2008). This hypothesis would agree with the results showed by the FGF and the BMP inhibitory treatments. In most of the cases in which the cardiac progenitors were bilaterally distributed, the tail elongation and its rotation relative to the trunk were also affected. All these phenotypes could indicate alterations in the morphogenesis of the embryo that could affect the normal cell divisions and, therefore, the rearrangements of the cardiac progenitors towards the ventral midline.

In addition to the EM module, we have also recognized a LM module characterized by the loss of *Tbx1/10* and the loss of multiple cardiopharyngeal subfunctions of pleiotropic genes that may have been preserved likely because they may be involved in other functions, and therefore occupying hubs of intersections with other gene pathways (Fig. 7 in **Chapter 2**). In ascidians, *Tbx1/10* is expressed in the second trunk ventral cells (STVC) that are the multipotent daughters of the TVCs. *Tbx1/10+* STVCs originate second heart precursors (SHPs) expressing the key regulator *Dach*, and the atrial siphon muscle field (ASMF) expressing the pharyngeal muscle markers *MyoD*, *Ebf*, and *Islet1* (Hirano and Nishida, 1997; Stolfi et al., 2010; Wang et al., 2019). Our results in appendicularians suggest that the loss of *Tbx1/10* correlates to the loss of the atrial siphon and longitudinal muscles, which contrast to the conservation of *Dach*, *MyoD*, *Ebf*, and *Islet1.* Expression analysis of these genes in *O. dioica* did not show expression in the cardiac nor in muscle domains, but in other organ primordia as the nervous system and the oikoplastic

epithelium, suggesting that their conservation was due to their pleiotropic roles.

In conclusion, our study provides a comprehensive catalogue of the extensive losses of genes and subfunctions that have contributed to the evolutionary deconstruction of the cardiopharyngeal GRN in appendicularians. This catalogue has provided us a unique opportunity to try to understand how these losses might have affected the diversification of appendicularians and ascidian lifestyles, and in special, how these losses might have impacted the evolution of heart in both lineages until adopting such divergent morphologies.

A future challenge to better understand the development of the heart of *O. dioica* will be to identify the inductive mechanisms to activate *Nk4* transcription, the mechanism to relocate the cardiac progenitors from the tail/trunk interface to the ventral region of the trunk, new genes involved in cardiogenesis in this organism, or new insights into the genetic base of human cardiomyopathies. To address these issues, we will need to perform functional analysis by knockdown, knockout, or gene reporter assays. All these techniques require the introduction of a molecule inside the oocyte, and in *O. dioica* two different approaches have been described, gonad
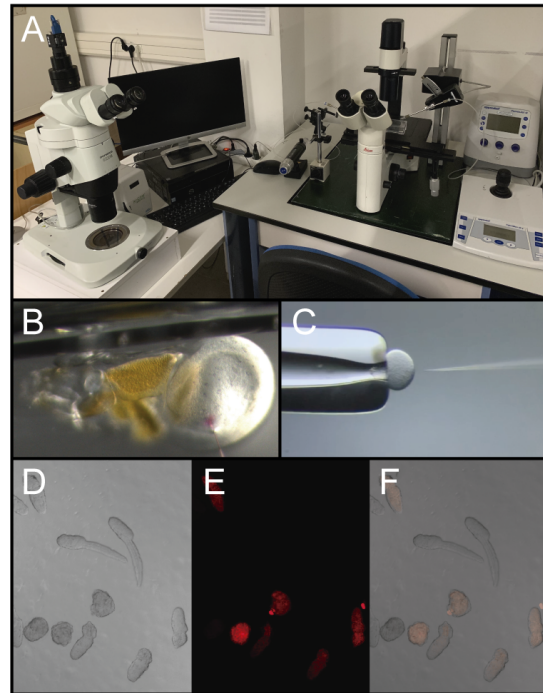
Fig. D4. (**A**) Injection facility for *O. dioica* animals in the University of Barcelona. (**B**) Injection of the gonad of a premature female of *O. dioica*, phenol red allows us to visualize the injected solution (red dot in panel B). (**C**) Recently spawned oocyte of *O. dioica* immobilized by a holding needle just before being injected. (**D-F**) Knockdown embryos by dsDNA injection into the gonad of a premature female. (**D**) Optical image of knockdown embryos. (**E**) Fluorescent image of knockdown embryos, mCherry fluorescence highlights embryos containing mRNA and, therefore, most likely also dsDNA. In this regard, notice that malformed embryos showing the brachyury-dsDNA phenotype are fluorescent (**F**).

and oocyte microinjection (Mikhaleva et al., 2018; Omotezako et al., 2017).

During my thesis, besides developing the fluorescent in situ hybridization in *O. dioica* we have also developed the gonad and oocyte microinjection techniques in our facility (Fig. D4 A). Along with its maturation, the female gonad of *O. dioica* goes through a syncytial phase in which meiotic nuclei are surrounded by common cytoplasm (Ganot et al., 2007). Injected molecules in

this maturing phase spread along the forming oocytes generating tens of altered embryos with a single injection (**Fig. D4 B**). After instigating this technique from scratch, we were able to obtain *Brachyury* knockdown embryos by DNA interference (DNAi) that showed malformations in the tail (**Fig. D4 D-F**). To correlate the severity of the phenotype with the amount of injected solution, we co-injected the dsDNA with mRNA-H2-mCherry, which is translated in vivo into a protein fusion of mCherry with the *O. dioica* histone H2 (**Chapter 3**). However, the wide variability in the number of knockdown embryos per clutch, the indeterminate concentration of the injected product on each embryo, and the difficulty in detecting mCherry fluorescence made us try oocyte microinjection.

Oocyte injection is a much more sophisticated technique that required new equipment and experience, but it solved the problems derived from gonad injection (**Fig. D4 C**). Oocyte microinjection provided more accurate traceability of the amount of the injected product in correlation to the severity of the phenotype moreover it made it easier to identify correct injected embryos by the co-injection of a fluorescent dextran brighter than fusion proteins. Both gonad and egg microinjection techniques are now ready to operate in our lab and open the possibility to perform genetic functional analyses by the generation of knockdowns by RNA interference (RNAi) or DNAi

(Mikhaleva et al., 2018; Omotezako et al., 2017), knockouts by CRISPR-Cas9 (Deng et al., 2018), or gene reporter assays.

3 THE HEART OF *O. DIOICA* AND ASCIDIANS ARE HOMOLOGOUS

At the beginning of this PhD project, we had no information about the development of the heart in appendicularians. The hearts of ascidians and appendicularians maintain a distinct anatomical organization with a tubular close structure in the case of ascidians and a flat open structure in the case of appendicularians. Moreover, they also present enormous temporal differences in their developmental times. While the heart of ascidians takes many days after metamorphosis until it becomes functional, the heart in *O. dioica* is functional and starts beating as soon as 8.5 hours post-fertilization. Therefore, the first challenge we had to face was to test if the two hearts were truly homologous organs, or on the contrary, they could be analogous pumping organs that have evolved independently in both lineages. Indeed, the study of the evolution of the vast variety of hearts and pumping organs across metazoans has made this organ to be a hot topic in the discussion of homologies and analogies in the field of EvoDevo (Xavier-Neto et al., 2007).

The historical concept of homology formulated by Mayr, 1970 and reformulated by Bock, 1973 says that "Features in two or more organisms are homologous if they stem phylogenetically from the same feature in the immediate common ancestor of these organisms." According to this definition, all pumping organs could be considered homologous if the protostome-deuterostome ancestor would have already had some kind of pumping organ from which present organs would have evolved from. The absence of pumping organs in many animal phyla would therefore indicate that hearts were secondary loss during metazoan evolution, and also open the possibility that they could have been independently innovated in the different phyla. Other researches, however, especially those interested in developmental biology rather than in taxonomy, conceived homology as a relation between traits that share the same developmental causes or generative mechanisms (e.g. traits that share the same genetic bases) (rev. in Minelli and Fusco, 2013). According to this definition, all the hearts would be homologous since, from arthropods to vertebrates, all use the same ancestral genetic network (*Nk4*, *Gata*, *Tbx*, *Hand*) (Olson, 2006). However, many studies have shown that the relationship between genotype and phenotype is so plastic that homologies at the gene level are not necessarily linked to homologies at the anatomical or evolutionary level. For example, Notch is used in the development of the wing imaginal disc of *Drosophila* as well as neural cells in chordates (Baguna and Garcia-Fernandez, 2003). Or for
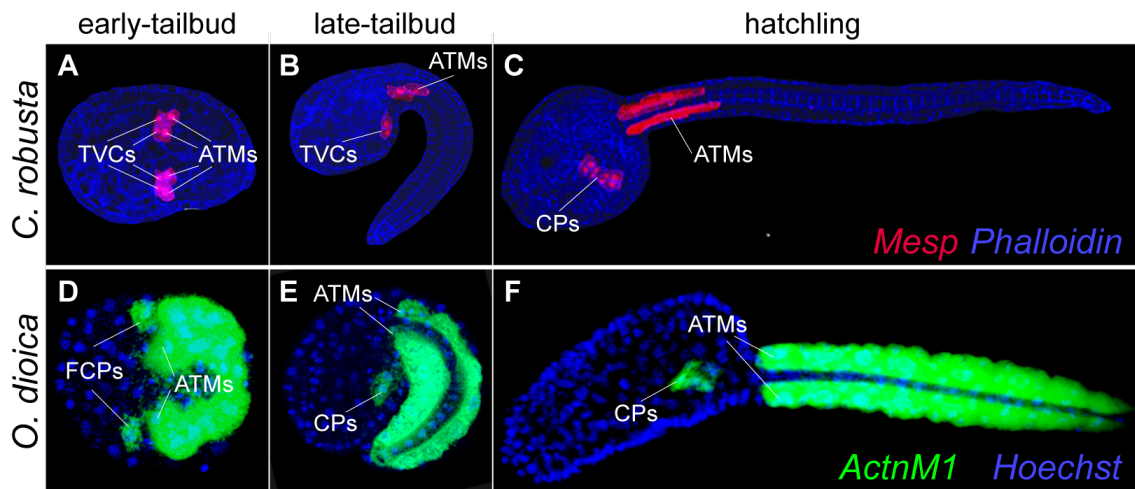
Fig. D5. **Cardiopharyngeal precursors in *O. dioica* and *C. robusta* have the same ontogeny and behavior**. Comparison between confocal micrographs of *C. robusta* (**A-C**) and *O. dioica* (**D-F**) show cardiac progenitors and anterior tail muscle cells (ATMs) together in the midline of the embryo at early-tailbud stage (**A, D**). At late-tailbud stage, ATMs are maintained in the anterior region of the tail while cardiac precursors have moved to the ventral region of the trunk (**B, E**) where they rest in hatchling stages (**C, F**). In red, cells expressing fluorescent proteins under the control of *Mesp* enhancer. In green, in situ hybridization of *ActnM1*. In *C. robusta* phalloidin in blue stains actin filaments, in *O. dioica* Hoechst in blue stain nuclei. Early-tailbud embryos are dorsal views. Late-tailbud and hatchling embryos are lateral views. Anterior is always to the left. CPs, cardiac progenitors; FCPs, first cardiac progenitors; TVCs, trunk ventral cells.

instance, the *Otx/Pax/Hox1* genetic cassette has been independently recruited to provide anterior-posterior regionalization in the central nervous system, the digestive system, or the thyroid (Cañestro et al., 2008). Moreover, the similar design between pumping organs of distant animal species (tubular hearts of arthropods and ascidians or chambered hearts of cephalopods and vertebrates) suggests that the hearts were independently created by convergence, as happen with the wings from insects and birds. All these theories were grouped in a new hypothesis proposed by Xavier-Neto et al., 2007. This hypothesis asserts that the prototype of all circulatory pumps presented in the protostome and deuterostome common ancestor may have been an organized layer

of contractile cells presumably lined to the external wall of primitive vessels that may have evolved to acquire different designs in the different phyla (Xavier-Neto et al., 2007). In the case of chordates, numerous studies have provided strong evidence supporting that the hearts of vertebrates and ascidians are homologous despite the great morphological differences between the multichambered complex heart of vertebrates and the tubular simple heart of ascidians (Davidson, 2007).

To know if the hearts of appendicularians and ascidian are homologous, we have performed an ontogenetic analysis of the heart development of *O. dioica* following the behavior of cardiac precursors. In ascidians, cardiopharyngeal progenitors are easy to trace using the expression of

Mesp (Davidson, 2007). The loss of this gene in appendicularians made us look for other markers for cardiac precursors in this organism. We tried using actins due to previous analyses with *Oikopleura longicauda* and *C. robusta* had described actin expression in cardiac precursors (Nishino et al., 2000; https://www.aniseed.cnrs.fr/). Our genomic survey and phylogenetic analysis, together with the intron position and the Diagnostic-Actin-Values (DAVs) analyses, identified seven actin genes in *O. dioica*, three cytoplasmic and four muscular, originated by independent gene duplications during the evolution of the appendicularian lineage. The expression of the seven actin genes revealed a highly dynamic tissue-specific expression pattern of the different actin genes, but only the *ActnM1* was expressed in the heart (**Chapter 1**).

Following *ActnM1* expression throughout the development of *O. dioica*, we detected that the FCPs share the same ontogenetic origin and behavior than the TVCs of ascidians. They originate from an asymmetric division at the interface between the trunk and the tail and finish their development in the ventral region of the trunk (**Fig. D5**). Integration of the expression data of *ActnM1* with the 4D-nuclear tracing from 16-cell stage up to tailbud stage (Stach et al., 2008) revealed that the development of the heart in *O.*

dioica also shares the same cell lineage fate map as ascidians. However, the split between the cardiac and the axial muscle lineages happens one cleavage earlier in *O. dioica*, probably related to the fact that *O. dioica* only has a cardiac precursor on each side against the two present in *C. robusta* (**Chapter 2**).

Therefore, although the hearts of ascidians and *O. dioica* are anatomically so different, our results reveal several criteria that provide solid evidence that both pumping organs are homologous: (1) they share the same pre-cardiac lineage fate map, in both cases coming from the same blastomere B5.2; (2) the appendicularian FCP and their ascidian counterparts TVCs both originate from an asymmetric division at the same position in the interface of the trunk and tail at the time of the split between the lineages towards axial tail muscle or cardiac fate; (3) the final position of the cardiac primordium in the ventral part of the trunk at tailbud stage is shared by appendicularians and ascidians; (4) the cardiac primordium express homologous cardiogenic genes such as *Nk4* and *Hand1/2*; (5) hearts of both appendicularians and ascidians develop into a similar double-layered structure with a myocardium and a pericardium. Therefore, despite the many losses of cardiopharyngeal genes and subfunctions, and despite the great architectural differences of the adult heart of appendicularians and ascidians, our

results provide strong evidence that both organs are homologous, and discard the possibility that they could have independently evolved by a process of convergence that could make them to be analogous. Indeed, the findings of this thesis provide a paradigmatic case of the "inverse-paradox" of EvoDevo (Cañestro et al., 2007) how two homologous structures can be developed with important differences in their genetic toolkit, in this case due to massive losses of genes and subfunctions.

4 FREE-LIVING APPENDICULARIANS REPRESENT DERIVED FORMS OF THE ANCESTRAL SESSILE TUNICATE

Once demonstrated that hearts of appendicularians and ascidians were homologous, and once our results had revealed the massive losses of cardiopharyngeal genes and subfunctions, the major challenge of this PhD project was to infer the impact of the deconstruction of the cardiopharyngeal GRN in the evolution of appendicularians. Considering the relevance that cardiopharyngeal structures on the evolution of free-living or sessile styles, the findings of this project has allowed us to propose an evolutionary scenario (for the sake of clarity, Fig. D6 reproduces here Fig. 8 from **Chapter 2**) to better understand the adaptive potential of the deconstruction of the cardiopharyngeal GRN in appendicularians, paying special attention to infer the condition of the last common ancestor (LCA) of tunicates.

The traditional view of chordate evolution proposed by Garstang (1928) suggested that the LCA of chordates had a sessile ascidian-like adult lifestyle and free-living dispersal larva, from which the fully motile style was later acquired during the evolution of cephalochordates and vertebrates by paedomorphosis (Garstang, 1928). The discovery that tunicates (a.k.a. urochordates) are the sister group of
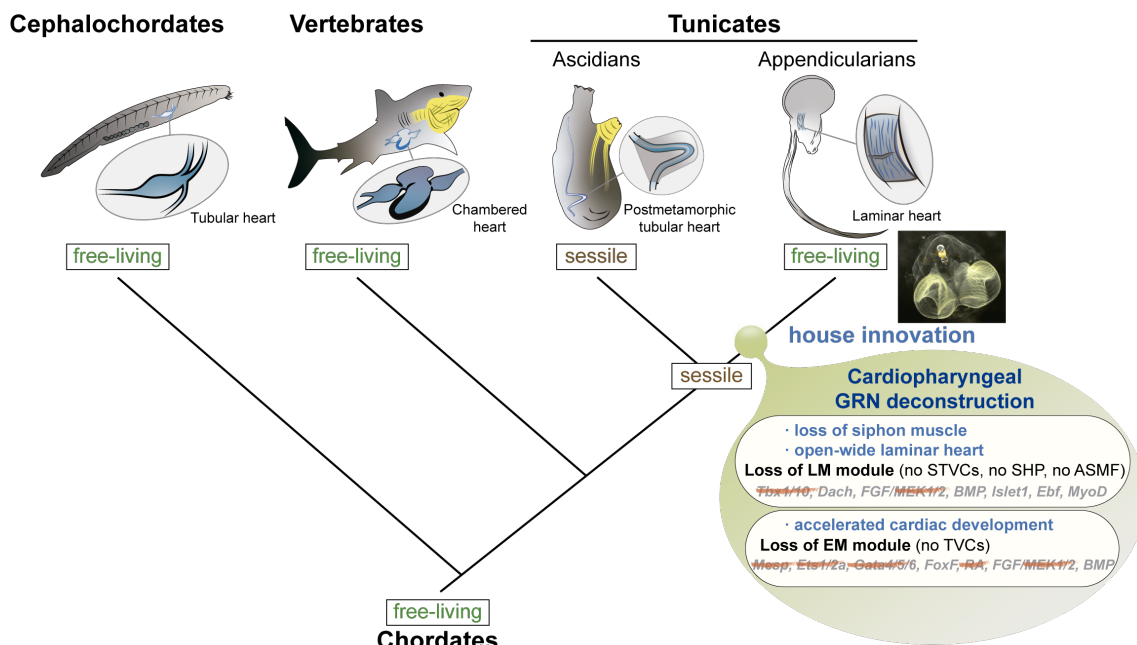
Fig. D6. The deconstruction of the cardiopharyngeal GRN in the ancestor of appendicularians favored their transition to a free-living life style. According to the deconstruction of the cardiopharyngeal GRN, the free-living life style of appendicularians represent a derived condition instigated by the innovation of the house, the loss of the siphon muscle, the development of an open-wide laminar heart, and the acquisition of an accelerated development. GRN, gene regulatory network; STVCs, second trunk ventral cells; SHP, second heart progenitors; ASMF, atrial siphon muscle field; TVCs, trunk ventral cells; LM, late multipotent; EM, early multipotent.

vertebrates, and therefore that the cephalochordate branching is basal within chordates (Bourlat et al., 2006; Delsuc et al., 2006), has radically changed our understanding of the evolution of our own phylum by providing a novel perspective in which the last common ancestor of chordates was a free-living organism. This novel view has brought renewed interest in appendicularians, whose complete free-living style could parsimoniously represent the ancestral condition of tunicates (Berrill, 1950; Braun et al., 2020; Swalla et al., 2000) considering their most accepted position as the sister group of the remaining tunicates (Braun et al., 2020; Delsuc et al., 2018; Kocot et al., 2018; Wada, 1998). Under the assumption that the LCA of tunicates was free-living, the sessility of ascidians should be considered a derived trait within tunicates that could have evolved with the innovation of the drastic metamorphosis that characterizes ascidians, but absent in appendicularians (Braun et al., 2020; Delsuc et al., 2018; Kocot et al., 2018). The results from this project, however, are in conflict with the assumption that the LCA of tunicates was free-living and provide evidence supporting that the LCA of tunicates had a sessile ascidian-like style and that the deconstruction of the cardiopharyngeal GRN can be interpreted as an evolutionary adaptation that may have contributed to the transition from an ancestral sessile to the innovation of a house-based free-living lifestyle. Particularly, the loss of the EM and LM during the deconstruction of the cardiopharyngeal GRN can be linked to: 1) an adaptive acceleration of the heart developmental program; 2) an adaptive transition from a close tubular architecture to an open wide laminar structure; and 3) the loss of the atrial and longitudinal muscle related with water circulation in ascidians that are replaced by action of tail beating and ciliary gills in appendicularians (**Fig. D6**).

ADAPTIVE ACCELERATION OF CARDIAC DEVELOPMENT UNDER HIGH ENERGETIC DEMAND OF JUVENILES

The loss of the EM and LM modules in appendicularians could correlate to the early activation of the cardiac markers *Nk4* and *Hand1/2* just after FCPs separate from the ATMs. In fact, the onset of *Hand1/2* could be regulated by *Nk4* as it takes place some minutes later than *Nk4*. Upcoming functional analysis will accept or refute this hypothesis. The early activation of the cardiac kernel would have been decisive to make a functional heart in just 8.5 hours post-fertilization, which could have been adaptive by favoring an early distribution of the hemolymph in juveniles before making its first house and start beating actively the tail, which supposes a high energetic demand.

ADAPTIVE OPEN LAMINAR HEART FACILITATES THE CIRCULATION OF THE HEMOLYMPH BY TAILBEATING

The loss of the LM module has been accompanied by the loss of the SHPs as we did not detect the SHPs marker *Dach* (Wang et al., 2019) expressed in cardiac precursors in *O. dioica*. Therefore, our results suggest that the cardiac cells in the heart of *O. dioica* might be homologous to the ascidians FHPs. The loss of the SHPs could be correlated to a reduction in the number of cells that finally comprise the heart, which could have favored the transition from a tubular to an open laminar shape. This open cardiac architecture could have been adaptive due to it would facilitate the circulation of the hemolymph, also propelled by the tail movements, to go through the heart.

THE PRESERVATION OF THE TAIL CORRELATES TO THE LOSS OF THE SIPHON MUSCLES

Besides affecting the SHPs, the loss of the LM module can be also correlated to the loss of the atrial and longitudinal siphon muscles present in the trunk of ascidians, but absent in the trunk of *O. dioica*, where the heart is the only muscle structure present in the trunk. As siphons are used in ascidians to propel the water inside the organism, we correlate this loss to the preservation of the tail in the adult phase. During the transition from a sessile to a planktonic lifestyle, the conservation of the

tail would have been positively selected becoming the system to propel water circulation, scenario in which muscles of the siphons could have disappeared by regressive evolution. Consistent with this view, the absence of body-wall actins, that would have been present in the tunicate ancestor, would be related to the loss of the muscles of the adult sessile form. Importantly, the loss of siphon muscle in appendicularians probably might not have been only the consequence of regressive evolution upon the innovation of the house, but it plausibly facilitated the evolution of the capability of appendicularians to control the orchestrated movement of the cilia in the gills to efficiently reverse the direction of the water flow, inwards during feeding or outwards during the inflation of the house without the interference of graceless muscle contractions (Bassham et al., 2008; Conley et al., 2018).

In conclusion, our results suggest an evolutionary scenario in which the deconstruction of the cardiopharyngeal GRN in appendicularians could have been an adaptive process that facilitated the innovation of a free-living style from an ancestral ascidian-like style and makes very unlikely that ascidians had evolved from a free-living ancestor similar to appendicularians. This evolutionary scenario is consistent with the observation that appendicularians maintain a U-shape intestine typical of sessile organisms

(Stach, 2007; Williams, 1996), which in past comparative morphological studies had been used to classify appendicularians as the sister group of aplousobranchia ascidians, also favored by the observation that appendicularians and aplousobranch larvae share the common derived character of the rotation of the swimming tail through 90 degrees to the left (Stach et al., 2008). Our conclusions that the LCA of tunicates was sessile is consistent with the interpretation of our findings independently of the phylogenetic position of appendicularians, either as basally divergent as the sister group of all other ascidians, or in the case that they were related to aplousobranch ascidians.

5 THE CONCEPT OF EVOLUTIONARY KNOCKOUT MODEL: THE PARADIGMATIC CASE OF THE LOSS OF THE RETINOIC ACID IN *O. DIOICA*

The recent bloom of sequenced genomes has discovered that gene loss is not occasional or insignificant, but a process that is abundant in all life kingdoms. The considerable research interest generated by gene loss encouraged us to propose the novel concept of "evolutionary knockout" (eKO) model. This concept, in contrast to "genetically-engineered knockout" models, consists of species that during their evolution have naturally lost genes. Therefore, they can be investigated as model systems to better understand, the evolution of GRNs, mechanisms of embryo development, or any physiological adaptation in the absence of any given gene of interest. There are already notorious examples of projects taking advantages of eKO models to better understand, for instance, the genetic bases of human diseases, such as the study of icefish species that have lost the development of bones as an adaption to artic low temperatures as a model to better understand osteoporosis or the study of Mexican cavefish that has lost vision and pigmentation as a model to better understand human forms of blindness and albinism (Albertson 2008, Maher 2009).

The great advantage of using eKO models rather than genetically-engineered knockout models is that while the latter normally generates large phenotypes or perturbs the normal development, eKO organisms integrate the gene loss into their established developmental and physiological systems allowing us to study the increasing genetic variation related to that phenotype.

During my PhD, our study of gene losses affecting the cardiopharyngeal GRN has contributed to extend previous work to use *O. dioica* as an eKO (**Chapter 3**). With a small compact genome full of gene losses, *O. dioica* has become an attractive model to study the impact of gene losses on animal EvoDevo. In particular, *O. dioica* has lost all the key genes involved in the classical retinoic acid (RA) synthesis pathway, and it does not have any alternative pathway to synthesize this developmental morphogen (Cañestro et al., 2006; Cañestro and Postlethwait, 2007; Martí-Solans et al., 2016). Therefore, the loss of RA makes *O. dioica* a unique case of an eKO model for RA signaling among chordates which has provided an attractive system to investigate the impact of gene loss on the evolution of the development of structures that require RA in other chordates as the development of the cardiopharyngeal field in appendicularians, deeply studied in this thesis. Moreover, during the course of this PhD project, we

have taken advantage of the use of *O. dioica* eKO for RA signaling to study genetic responses alternative to RA pathways that could be affected by environmental threats (**Chapter 4**). Polyunsaturated aldehydes (PUAs) and oxylipins are two environmental xenobiotic compounds produced naturally by diatom-blooms, which due to climate change and ocean warming are becoming alarmingly abundant. Many studies have suggested that the teratogenic effects of these compounds shown in a wide variety of marine organisms could be due to interference with RA-synthesis since PUAs can compete with retinaldehyde for the substrate-binding site of the Aldh1a (Bchini et al., 2013). To test this hypothesis, we have taken advantage of *O. dioica* to check if those toxic compounds did indeed not affect *O. dioica* development due to its condition of eKO for RA, or on the contrary, to investigate for other alternative genetic pathways that could mediate the response to the environmental threat. *O. dioica* embryos treated with PUAs showed arrested phenotypes in pre-tailbud stage that looked like a golf ball. Moreover, those golf ball embryos manifested an expression delay of developmental genes and an increased expression of defensome related genes demonstrating that those biotoxins could cause dramatic developmental alterations independently of RA. Consequently, PUAs can alter other

genetic pathways in addition to the RA synthesis pathway previously described.

The work of this thesis, therefore, has helped to further extend the study of *O. dioica* as a successful gene loser*,* in particular, revealing this organism as a useful eKO model that has lost the genetic modules related to multipotent cellular states in the cardiopharyngeal GRN. Future functional investigations in *O. dioica* could help to identify the minimal genetic toolkit for the development of a heart in chordates, and thanks to its simplicity could also provide an attractive model to better understanding heart developmental aspects that are difficult to study in complex models, such as the differentiation between myocardium and pericardium, or to investigate the genetic bases of some human cardiomyopathies.

CONCLUSIONS

1. The massive loss of cardiopharyngeal genes (*Mesp, Ets1/2a, Gata4/5/6, Mek1/2, Raldh1a, Cyp26, RAR*) and subfunctions (*FoxF, Isl, Ebf, Mrf, Dach,* FGF signaling, BMP signaling) reveal the evolutionary deconstruction of the cardiopharyngeal gene regulatory network (GRN) in appendicularians.

2. The development of the heart in appendicularians became independent of *Mesp,* the cardiopharyngeal master gene in ascidians and vertebrates.

3. The study of the evolutionary deconstruction of the cardiopharyngeal GRN in appendicularians reveals the dismantling of two developmental modules related to the early and late multipotent cardiopharyngeal states.

4. The dismantling of the early module, including the loss of FGF and RA signaling and the loss of *Mesp* and *Ets1/2*, led to the loss of the multipotential state observed in the trunk ventral cells (TVCs) of ascidians, implying the early activation of the cardiac kernel as revealed by *Nk4* first, followed by *Hand1/2* in the first cardiac progenitors (FCP) of appendicularians by the incipient- and mid-tailbud stage, respectively.

5. The early onset of *Nk4*, the loss of *Gata4/5/6*, the loss of *FoxF* subfunction, and the lack of BMP activity in the cardiac progenitors of *O. dioica* indicate a change in the developmental mechanism to modify the final position of FCPs compared to the migratory behavior observed in the TVCs of ascidians.

6. The loss of *Tbx1/10* and the cardiopharyngeal subfunctions of many genes (*Dach, Islet1, Ebf,* and *MyoD*) reveal de dismantling of the late multipotent (LM) module which in ascidians is responsible for the specification of SHPs and pharyngeal muscle.

7. *ActnM1* is the only actin gene expressed in cardiac progenitors, which allows us to trace back cardiac development until the 64-cell stage showing the same ontogenetic process as in ascidians and enabling us to conclude that both hearts are homologous.

8. The evolutionary impact of the loss of the early multipotent (EM) module led to the acceleration of the onset of the cardiac program.

9. The evolutionary impact of the loss of the LM module plausibly led to the remodeling of the cardiac structure from tubular to laminar and to the loss of the pharyngeal muscles.

10. The evolutionary impact of cardiopharyngeal gene losses was adaptive for the innovation of a free-living house-based lifestyle favoring the development of an active and open heart that facilitated the hemolymph circulation and favoring the loss of the

pharyngeal muscle that allowed the water flux reversal during feeding or the inflation of the house.

11. Independently of the phylogenetic relationship of the appendicularians and the rest of tunicates, our results suggest that the last common ancestor of tunicates had a sessile ascidian-like lifestyle, being the evolutionary deconstruction of the cardiopharyngeal GRN pivotal for the adaptation to a free-living lifestyle based on the innovation of the house.

12. Our results related to the deconstruction of the cardiopharyngeal GRN and the effect of toxic compounds in *O. dioica* development have contributed to developing the eKO model concept with *O. dioica* as a paradigmatic example of the advantages of using species that along their evolution has suffered extensive gene losses.

BIBLIOGRAPHY

Acuna, J.L., Deibel, D., Saunders, P., Booth, B., Hatfield, E., Klein, B., Mei, Z.P., Rivkin, R., 2002. Phytoplankton ingestion by appendicularians in the North Water. Deep. Res. PART II-TOPICAL Stud. Oceanogr. 49, 5101–5115.

Albalat, R., Cañestro, C., 2016. Evolution by gene loss. Nat. Rev. Genet. 17, 379–391. https://doi.org/10.1038/nrg.2016.39

Albalat, R., Marti-Solans, J., Cañestro, C., Martí-Solans, J., Cañestro, C., Marti-Solans, J., Cañestro, C., 2012. DNA methylation in amphioxus: from ancestral functions to new roles in vertebrates. Br. Funct Genomics 11, 142–155. https://doi.org/els009 [pii]10.1093/bfgp/els009

Aravind, L., Watanabe, H., Lipman, D.J., Koonin, E. V, 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proc Natl Acad Sci U S A 97, 11319–11324. https://doi.org/10.1073/pnas.200346997

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H., 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2, 2006 0008. https://doi.org/msb4100050 [pii]10.1038/msb4100050

Baguna, J., Garcia-Fernandez, J., 2003. Evo-Devo: the long and winding road. Int J Dev Biol 47, 705–713.

Bassham, S., Cañestro, C., Postlethwait, J.H., 2008. Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. BMC Biol. 6, 35. https://doi.org/10.1186/1741-7007-6-35

Bchini, R., Vasiliou, V., Branlant, G., Talfournier, F., Rahuel-Clermont, S., 2013. Retinoic acid biosynthesis catalyzed by retinal dehydrogenases relies on a rate-limiting conformational transition associated with substrate recognition. Chem Biol Interact 202, 78–84. https://doi.org/10.1016/j.cbi.2012.11.019

Beh, J., Shi, W., Levine, M., Davidson, B., Christiaen, L., 2007. FoxF is essential for FGF-induced migration of heart progenitor cells in the ascidian Ciona intestinalis. Development 134, 3297–3305. https://doi.org/10.1242/dev.010140

Berna, L., Alvarez-Valin, F., 2014. Evolutionary genomics of fast evolving tunicates. Genome Biol. Evol. 6, 1724–1738. https://doi.org/10.1093/gbe/evu122

Bernadskaya, Y.Y., Brahmbhatt, S., Gline, S.E., Wang, W., Christiaen, L., 2019. Discoidin-domain receptor coordinates cell-matrix adhesion and collective polarity in migratory cardiopharyngeal progenitors. Nat. Commun. 10. https://doi.org/10.1038/s41467-018-07976-3

Bernard, F.H., Hinman, V.F., Degnan, B.M., 1998. Retinoic acid disrupts anterior ectodermal and endodermal development in ascidian larvae and postlarvae. Dev. Genes Evol. 208, 336–345.

Berrill, N.J., 1950. The Tunicata with an account of the British species. Sold by Quaritch, London.

Berrill, N.Y., 1950. The tunicata. London, England.

Bishopric, N.H., 2005. Evolution of the heart from bacteria to man, in: Annals of the New York Academy of Sciences. New York Academy of Sciences, pp. 13–29. https://doi.org/10.1196/annals.1341.002

Blomme, T., Vandepoele, K., De Bodt, S., Simillion, C., Maere, S., Van de Peer, Y., 2006. The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol 7, R43. https://doi.org/10.1186/gb-2006-7-5-r43

Bock, W.J., 1973. Philosophical foundations of classical evolutionary classification. Syst. Zool. 22. https://doi.org/10.2307/2412945

Bollner, T., Holmberg, K., Olsson, R., 1986. A rostral sensory mechanism in Oikopleura dioica appendicularian. Acta Zool. 67, 235–241.

Bourlat, S.J., Juliusdottir, T., Lowe, C.J., Freeman, R., Aronowicz, J., Kirschner, M., Lander, E.S., Thorndyke, M., Nakano, H., Kohn, A.B., Heyland, A., Moroz, L.L., Copley, R.R., Telford, M.J., 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444, 85–88. https://doi.org/10.1038/nature05241

Braun, K., Leubner, F., Stach, T., 2020. Phylogenetic analysis of phenotypic characters of Tunicata supports basal Appendicularia and monophyletic Ascidiacea. Cladistics 36, 259–300. https://doi.org/10.1111/cla.12405

Burighel, P., Brena, C., Martinucci, G.B., Cima, F., 2001. Gut ultrastructure of the appendicularian Oikopleura dioica (Tunicata). Invertebr. Biol. 120, 278–293. https://doi.org/10.1111/j.1744-7410.2001.tb00038.x

Cañestro, C., Albalat, R., 2012. Transposon diversity is higher in amphioxus than in vertebrates: functional and evolutionary inferences. Br. Funct Genomics 11, 131–141. https://doi.org/10.1093/bfgp/els010

Cañestro, C., Bassham, S., Postlethwait, J.H., 2005. Development of the central nervous system in the larvacean Oikopleura dioica and the evolution of the chordate brain. Dev. Biol. 285, 298–315. https://doi.org/10.1016/j.ydbio.2005.06.039

Cañestro, C., Bassham, S., Postlethwait, J.H.H.J.H., 2008. Evolution of the thyroid: Anterior-posterior regionalization of the Oikopleura endostyle revealed by Otx, Pax2/5/8, and Hox1 expression. Dev. Dyn. 237, 1490–9. https://doi.org/10.1002/dvdy.21525

Cañestro, C., Catchen, J.M.J.M.M.J.M., Rodríguez-Marí, A., Yokoi, H., Postlethwait, J.H.H.J.H., 2009. Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. PLoS Genet. 5, e1000496. https://doi.org/10.1371/journal.pgen.1000496

Cañestro, C., Postlethwait, J.H.H., 2007. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. Dev Biol 305, 522–538. https://doi.org/10.1016/j.ydbio.2007.02.032

Cañestro, C., Postlethwait, J.H.H.J.H., Gonzàlez-Duarte, R., Albalat, R., 2006. Is retinoic acid genetic machinery a chordate innovation? Evol. Dev. 8, 394–406. https://doi.org/10.1111/j.1525-142X.2006.00113.x

Cañestro, C., Yokoi, H., Postlethwait, J.H.J., Canestro, C., Yokoi, H., Postlethwait, J.H.J., Cañestro, C., Yokoi, H., Postlethwait, J.H.J., 2007. Evolutionary developmental biology and genomics. Nat. Rev. Genet. 8, 932–42. https://doi.org/10.1038/nrg2226

Capitanio, F.L., Curelovich, J., Tresguerres, M., Negri, R.M., Viñas, M.D., Esnal, G.B., 2008. Seasonal cycle of appendicularians at a

coastal station (38°28'S, 57°41'W) of the SW Atlantic Ocean. Bull. Mar. Sci. 82.

Carroll, S.B., 1995. Homeotic genes and the evolution of arthropods and chordates. Nature 376, 479–485.

Chalopin, D., Naville, M., Plard, F., Galiana, D., Volff, J.N., 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol. Evol. 7, 567–580. https://doi.org/10.1093/gbe/evv005

Chavali, S., Morais, D.A., Gough, J., Babu, M.M., 2011. Evolution of eukaryotic genome architecture: Insights from the study of a rapidly evolving metazoan, Oikopleura dioica: Non-adaptive forces such as elevated mutation rates may influence the evolution of genome architecture. Bioessays 33, 592–601. https://doi.org/10.1002/bies.201100034

Christen, R., Braconnot, J.-C., 1998. Molecular phylogeny of tunicates. A preliminary study using 28S ribosomal RNA partial sequences: important implications in terms of evolution and ecology., in: Bone, Q. (Ed.), The Biology of Pelagic Tunicates. Oxford University Press, New York, pp. 265–271.

Christiaen, L., Davidson, B., Kawashima, T., Powell, W., Nolla, H., Vranizan, K., Levine, M., 2008. The transcription/migration interface in heart precursors of Ciona intestinalis. Science (80-.). 320, 1349–1352. https://doi.org/320/5881/1349 [pii]10.1126/science.1158170

Christiaen, L., Stolfi, A., Levine, M., 2010. BMP signaling coordinates gene expression and cell migration during precardiac mesoderm development. Dev. Biol. 340, 179–187. https://doi.org/10.1016/j.ydbio.2009.11.006

Cima, F., Brena, C., Burighel, P., 2002. Multifarious activities of gut epithelium in an appendicularian (Oikopleura dioica : Tunicata). MAR BIOL 141, 479–490.

Conley, K.R., Gemmell, B.J., Bouquet, J.M., Thompson, E.M., Sutherland, K.R., 2018. A self-cleaning biological filter: How appendicularians mechanically control particle adhesion and removal. Limnol. Oceanogr. 63, 927–938. https://doi.org/10.1002/lno.10680

Coppola, U., Ristoratore, F., Albalat, R., D'Aniello, S., 2019. The evolutionary landscape of the Rab family in chordates. Cell. Mol. Life Sci. 76, 4117–4130. https://doi.org/10.1007/s00018-019-03103-7

Danks, G., Campsteijn, C., Parida, M., Butcher, S., Doddapaneni, H., Fu, B., Petrin, R., Metpally, R., Lenhard, B., Wincker, P., Chourrout, D., Thompson, E.M., Manak, J.R., 2013. OikoBase: A genomics and developmental transcriptomics resource for the urochordate Oikopleura dioica. Nucleic Acids Res. 41, 1–9. https://doi.org/10.1093/nar/gks1159

Darwin, C., 1876. The Origin of Species by Means of Natural Selection; or the Preservation of Favoured Races in the Struggle for Life, 6th edn. ed.

Davidson, B., 2007. Ciona intestinalis as a model for cardiac development. Semin Cell Dev Biol 18, 16–26. https://doi.org/10.1016/j.semcdb.2006.12.007

Davidson, B., Shi, W., Beh, J., Christiaen, L., Levine, M., 2006. FGF signaling delineates the cardiac progenitor field in the simple chordate, Ciona intestinalis. Genes Dev 20, 2728–2738. https://doi.org/20/19/2728 [pii]10.1101/gad.1467706

Davidson, B., Shi, W., Levine, M., 2005. Uncoupling heart cell specification and migration in the simple chordate Ciona intestinalis. Development 132, 4811–4818.

https://doi.org/10.1242/dev.02051

Davidson, E.H., Erwin, D.H., 2006. Gene regulatory networks and the evolution of animal body plans. Science (80-.). 311, 796–800.

Davoll, P.J., Youngbluth, M.J., 1990. Heterotrophic activity on appendicularian (Tunicata: Appendicularia) houses in mesopelagic regions and their potential contribution to particle flux. Deep Sea Res. Part A, Oceanogr. Res. Pap. 37, 285–294. https://doi.org/10.1016/0198-0149(90)90128-I

De Robertis, E.M., 2008. Evo-devo: variations on ancestral themes. Cell 132, 185–195. https://doi.org/S0092-8674(08)00050-0 [pii]10.1016/j.cell.2008.01.003

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

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

T., Spagnuolo, A., Stainier, D., Suzuki, M.M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Douglas Hyatt, P., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine, † Michael, Satoh, † Nori, Rokhsar, D.S., 2002. The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. Science (80-.). 298, 2157–2167. https://doi.org/10.1126/science.1080049

Delsman, H.C., 1910. Contributions on the ontogeny of Oikopleura dioica. Verch. Rijksinst. Onderz. Zee 3, 1–24.

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

Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.K., Turon, X., López-Legentil, S., Piette, J., Lemaire, P., Douzery, E.J.P., 2018. A phylogenomic framework and timescale for comparative studies of tunicates. BMC Biol. 16. https://doi.org/10.1186/s12915-018-0499-2

Deng, W., Henriet, S., Chourrout, D., 2018. Prevalence of Mutation-Prone Microhomology-Mediated End Joining in a Chordate Lacking the c-NHEJ DNA Repair Pathway. Curr. Biol. 28, 3337-3341.e4. https://doi.org/10.1016/j.cub.2018.08.048

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J.-M.M., Da Silva, C., Brinkmann, H., Mikhaleva, J., Olsen, L.C., Jubin, C., Cañestro, C., Bouquet, J.-M.M., Danks, G., Poulain, J., Campsteijn, C., Adamski, M., Cross, I., Yadetie, F., Muffato, M., Louis, A., Butcher, S., Tsagkogeorga, G., Konrad, A., Singh, S., Jensen, M.F., Cong, E.H., Eikeseth-Otteraa, H., Noel, B., Anthouard, V., Porcel, B.M., Kachouri-Lafond, R., Nishino, A., Ugolini, M.,

Chourrout, P., Nishida, H., Aasland, R., Huzurbazar, S., Westhof, E., Delsuc, F., Lehrach, H., Reinhardt, R., Weissenbach, J., Roy, S.W., Artiguenave, F., Postlethwait, J.H., Manak, J.R., Thompson, E.M., Jaillon, O., Du Pasquier, L., Boudinot, P., Liberles, D.A., Volff, J.-N.N., Philippe, H., Lenhard, B., Roest Crollius, H., Wincker, P., Chourrout, D., Crollius, H.R., Wincker, P., Chourrout, D., Roest Crollius, H., Wincker, P., Chourrout, D., 2010. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science (80-.). 330, 1381–1385. https://doi.org/10.1126/science.1194167

Devine, W.P., Wythe, J.D., George, M., Koshiba-Takeuchi, K., Bruneau, B.G., 2014. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. Elife 3. https://doi.org/10.7554/eLife.03848

Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., Couto, A., Marra, V., Keleman, K., Dickson, B.J., 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448, 151–156. https://doi.org/nature05954 [pii]10.1038/nature05954

Diogo, R., Kelly, R.G., Christiaen, L., Levine, M., Ziermann, J.M., Molnar, J.L., Noden, D.M., Tzahor, E., 2015. A new heart for a new head in vertebrate cardiopharyngeal evolution. Nature 520, 466–473. https://doi.org/10.1038/nature14435

Drouin, G., Godin, J.R., Page, B., 2011. The genetics of vitamin C loss in vertebrates. Curr Genomics 12, 371–378. https://doi.org/10.2174/1389202117964297 36CG-12-371 [pii]

Edvardsen, R.B., Seo, H.C., Jensen, M.F., Mialon, A., Mikhaleva, J., Bjordal, M., Cartry, J.,

Reinhardt, R., Weissenbach, J., Wincker, P., Chourrout, D., 2005. Remodelling of the homeobox gene complement in the tunicate Oikopleura dioica. Curr Biol 15, R12-3. https://doi.org/10.1016/j.cub.2004.12.010

Evans, S.M., Yan, W., Murillo, M.P., Ponce, J., Papalopulu, N., 1995. tinman, a Drosophila homeobox gene required for heart and visceral mesoderm specification, may be represented by a family of genes in vertebrates: XNkx-2.3, a second vertebrate homologue of tinman. Development 121, 3889–99.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Fenaux, R, 1998. Life history of the Appendicularia, in: Bone, Q. (Ed.), The Biology of Pelagic Tunicates. Oxford University Press, Oxford, pp. 151–159.

Fenaux, R., 1998. Anatomy and functional morphology of the Appendicularia., in: Bone, Q. (Ed.), The Biology of Pelagic Tunicates. Oxford University Press, New York, pp. 25–34.

Fernández, R., Gabaldón, T., 2020. Gene gain and loss across the metazoan tree of life. Nat. Ecol. Evol. 4. https://doi.org/10.1038/s41559-019-1069-x

Ferrández-Roldán, A., Martí-Solans, J., Cañestro, C., Albalat, R., 2019. Oikopleura dioica: An Emergent Chordate Model to Study the Impact of Gene Loss on the Evolution of the Mechanisms of Development, Results and Problems in Cell Differentiation. Springer Verlag. https://doi.org/10.1007/978-3-030-23459-1_4

Ferrier, D.E.K., 2011. Tunicates push the limits of animal evo-devo. BMC Biol. 9. https://doi.org/10.1186/1741-7007-9-3

Fol, H., 1872. Etudes sur les Appendiculaires du De troit de Messine. Mem. Soc. Phys. Hist. Nat.

Geneve 21, 445.

Force, A., Lynch, M., Pickett, B., Amores, A., Yan, Y.-L., Postlethwait, J., 1999. Preservation of Duplicate Genes by Complementary, Degenerative Mutations.

Fu, X., Adamski, M., Thompson, E.M., 2008. Altered miRNA repertoire in the simplified chordate, Oikopleura dioica. Mol. Biol. Evol. 25, 1067–1080. https://doi.org/10.1093/molbev/msn060

Fujii, S., Nishio, T., Nishida, H., 2008. Cleavage pattern, gastrulation, and neurulation in the appendicularian, Oikopleura dioica. Dev Genes Evol 218, 69–79.

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

Galt, C.P., Fenaux, R., 1990. Urochordata larvacea., in: Adiyodi, K.G., Adiyodi, R.G. (Eds.), Reproductive Biology of Invertebrates. Oxford & IBH Publishing, New Delhi, pp. 471–500.

Ganot, P., Bouquet, J.M., Kallesoe, T., Thompson, E.M., 2007. The Oikopleura coenocyst, a unique chordate germ cell permitting rapid, extensive modulation of oocyte production. Dev Biol 302, 591–600.

Garcia-Fernàndez, J., 2005. The genesis and evolution of homeobox gene clusters. Nat Rev Genet 6, 881–892. https://doi.org/10.1038/nrg1723

Garstang, W., 1928. The morphology of the Tunicata, and its bearings on the phylogeny of the chrodata. Quar. J. Micr. Sci. 72, 51–186.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A.P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J.H., Hempel, S., Herman, Z., Jaramillo, D.F., Kelly, D.E., Kelly, S.L., Kotter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S.L., Revuelta, J.L., Roberts, C.J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K., Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C.Y., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W., Johnston, M., 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature 418, 387–391. https://doi.org/10.1038/nature00935nature00935 [pii]

Goldstone, J. V, Hamdoun, A., Cole, B.J., Howard-Ashby, M., Nebert, D.W., Scally, M., Dean, M., Epel, D., Hahn, M.E., Stegeman, J.J., 2006. The chemical defensome: environmental sensing and response genes in the Strongylocentrotus purpuratus genome. Dev Biol 300, 366–384.

Gorsky, G., Fenaux, R., 1998. The role of Appendicularia in marine food webs., in: Bone, Q. (Ed.), The Biology of Pelagic Tunicates. Oxford University Press, Oxford.

Guijarro-Clarke, C., Holland, P.W.H., Paps, J., 2020. Widespread patterns of gene loss in the evolution of the animal kingdom. Nat. Ecol. Evol. 4, 519–523. https://doi.org/10.1038/s41559-020-1129-2

Halder, G., Callaerts, P., Gehring, W.J., 1995. New perspectives on eye evolution. Curr Opin Genet Dev 5, 602–609.

Heenan, P., Zondag, L., Wilson, M.J., 2016. Evolution of the Sox gene family within the

chordate phylum. Gene 575. https://doi.org/10.1016/j.gene.2015.09.013

Hendy, M.D., Penny, D., 1989. A Framework for the Quantitative Study of Evolutionary Trees. Syst. Zool. 38, 297. https://doi.org/10.2307/2992396

Hirakow, R., Kajita, N., 1994. Electron microscopic study of the development of amphioxus, Branchiostoma belcheri tsingtauense: the neurula and larva. Kaibogaku Zasshi 69, 1-13.

Hirano, T., Nishida, H., 1997. Developmental fates of larval tissues after metamorphosis in ascidian Halocynthia roretzi. I. Origin of mesodermal tissues of the juvenile. Dev. Biol. 192, 199–210. https://doi.org/10.1006/dbio.1997.8772

Hogan, B., 2004. Deconstructing the genesis of animal form. Development 131, 2515 LP – 2520. https://doi.org/10.1242/dev.01192

Holland, L.Z., 2016. Tunicates. Curr. Biol. 26, R146–R152. https://doi.org/10.1016/j.cub.2015.12.024

Holland, L.Z., Gibson-Brown, J., 2003. The Ciona intestinalis genome: when the constraints are off. Bioessays 25, 529–532.

Holland, N.D., Venkatesh, T. V., Holland, L.Z., Jacobs, D.K., Bodmer, R., 2003. AmphiNk2-tin, an amphioxus homeobox gene expressed in myocardial progenitors: Insights into evolution of the vertebrate heart. Dev. Biol. 255. https://doi.org/10.1016/S0012-1606(02)00050-7

Holland, P.W.H., Marlétaz, F., Maeso, I., Dunwell, T.L., Paps, J., 2017. New genes from old: Asymmetric divergence of gene duplicates and the evolution of development. Philos. Trans. R. Soc. B Biol. Sci. https://doi.org/10.1098/rstb.2015.0480

Holmberg, K., 1984. A transmission electron microscopic investigation of the sensory vesicle in the brain of Oikopleura dioica (Appendicularia). Zoomorphology 104, 298–303. https://doi.org/10.1007/BF00312011

Hosp, J., Sagane, Y., Danks, G., Thompson, E.M., 2012. The evolving proteome of a complex extracellular matrix, the Oikopleura house. PLoS One 7. https://doi.org/10.1371/journal.pone.004017 2

Howes, R.E., Patil, A.P., Piel, F.B., Nyangiri, O.A., Kabaria, C.W., Gething, P.W., Zimmerman, P.A., Barnadas, C., Beall, C.M., Gebremedhin, A., Ménard, D., Williams, T.N., Weatherall, D.J., Hay, S.I., 2011. The global distribution of the Duffy blood group. Nat. Commun. 2. https://doi.org/10.1038/ncomms1265

Hughes, A.L., Friedman, R., 2005. Gene duplication and the properties of biological networks. J Mol Evol 61, 758–764.

Jacob, F., 1977. Evolution and tinkering. Science (80-.). 196, 1161–1166.

Jeffery, W.R., 2009. Regressive Evolution in Astyanax Cavefish . Annu. Rev. Genet. 43, 25–47. https://doi.org/10.1146/annurev-genet-102108-134216

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P., Ahringer, J., 2003. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature 421, 231–237. https://doi.org/10.1038/nature01278nature01 278 [pii]

Kelly, R.G., Evans, S.M., 2010. Chapter 2.2 - The Second Heart Field, in: Heart Development and Regeneration. pp. 143–169.

Kim, D.U., Hayles, J., Kim, D., Wood, V., Park, H.O., Won, M., Yoo, H.S., Duhig, T., Nam, M., Palmer, G., Han, S., Jeffery, L., Baek, S.T.,

Lee, H., Shim, Y.S., Lee, M., Kim, L., Heo, K.S., Noh, E.J., Lee, A.R., Jang, Y.J., Chung, K.S., Choi, S.J., Park, J.Y., Park, Y., Kim, H.M., Park, S.K., Park, H.J., Kang, E.J., Kim, H.B., Kang, H.S., Park, H.M., Kim, K., Song, K., Song, K.B., Nurse, P., Hoe, K.L., 2010. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat Biotechnol 28, 617–623. https://doi.org/nbt.1628 [pii]10.1038/nbt.1628

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

King, M.C., Wilson, A.C., 1975. Evolution at two levels in humans and chimpanzees. Science (80-.). 188, 107–116.

Kishi, K., Hayashi, M., Onuma, T.A., Nishida, H., 2017. Patterning and morphogenesis of the intricate but stereotyped oikoplastic epidermis of the appendicularian, Oikopleura dioica. Dev. Biol. 428. https://doi.org/10.1016/j.ydbio.2017.06.008

Kitajima, S., Takagi, A., Inoue, T., Saga, Y., 2000. MesP1 and MesP2 are essential for the development of cardiac mesoderm. Development 127.

Kocot, K.M., Tassia, M.G., Halanych, K.M., Swalla, B.J., 2018. Phylogenomics offers resolution of major tunicate relationships. Mol. Phylogenet. Evol. 121, 166–173. https://doi.org/10.1016/j.ympev.2018.01.005

Koonin, E. V, Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, S., Rogozin, I.B., Smirnov, S., Sorokin, A. V, Sverdlov, A. V, Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A., 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes, Koonin et al.

Korona, R., 2011. Gene dispensability. Curr Opin Biotechnol 22, 547–551. https://doi.org/S0958-1669(11)00080-2 [pii]10.1016/j.copbio.2011.04.017

Kortschak, R.D., Samuel, G., Saint, R., Miller, D.J., 2003. EST Analysis of the Cnidarian Acropora millepora Reveals Extensive Gene Loss and Rapid Sequence Divergence in the Model Invertebrates. Curr. Biol. 13, 2190–2195. https://doi.org/10.1016/j.cub.2003.11.030

Koskiniemi, S., Sun, S., Berg, O.G., Andersson, D.I., 2012. Selection-driven gene loss in bacteria. PLoS Genet 8, e1002787. https://doi.org/10.1371/journal.pgen.100278 7PGENETICS-D-12-00783 [pii]

Kowalevsky A., 1866. Entwickelungsgeschichte der einfachen Ascidien. Mém Acad Imp Sci St-pétersbg. (Sér VII) 10, 1–19.

Krieg, P.A., Warkman, A.S., 2015. Early Heart Development, in: Principles of Developmental Genetics: Second Edition. Elsevier Inc., pp. 407–420. https://doi.org/10.1016/B978-0-12-405945-0.00023-5

Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolosse, A., Auer, H., Achouri, Y., Dubois, C., Bondue, A., Simons, B.D., Blanpain, C., 2014. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. Nat. Cell Biol. 16, 829–840. https://doi.org/10.1038/ncb3024

Lescroart, F., Kelly, R.G., Le Garrec, J.F., Nicolas, J.F., Meilhac, S.M., Buckingham, M., 2010. Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. Development 137, 3269–3279. https://doi.org/10.1242/dev.050674

Lewis, E.B., 1978. A gene complex controlling

segmentation in Drosophila. Nature 276, 565–570.

Lints, T.J., Parsons, L.M., Hartley, L., Lyons, I., Harvey, R.P., 1993. Nkx-2.5: A novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. Development 119.

Lynch, M., 2002. Gene duplication and evolution. Science (80-.). 297, 945–947.

Makino, T., McLysaght, A., 2012. Positionally biased gene loss after whole genome duplication: Evidence from human, yeast, and plant. Genome Res 22, 2427–2435. https://doi.org/gr.131953.111 [pii]10.1101/gr.131953.111

Martí-Solans, J., Belyaeva, O.V.O. V., Torres-Aguila, N.P.N.P., Kedishvili, N.Y.N.Y., Albalat, R., Cañestro, C., 2016. Coelimination and Survival in Gene Network Evolution: Dismantling the RA-Signaling in a Chordate. Mol. Biol. Evol. 33, 2401–2416. https://doi.org/10.1093/molbev/msw118

Martinucci, G., Brena, C., Cima, F., Burighel, P., 2005. Synchronous spermatogenesis in appendicularians, in: Gorsky, G., Youngbluth, M., Deibel, D. (Eds.), Response of Marine Ecosystems to Global Changes: Ecological Impact of Appendicularians. Éditions des Archives Contempoaines, Paris.

Mayr, W., 1970. Principles of Systematic Zoology, Science. McGraw-Hill, New York. https://doi.org/10.1126/science.167.3924.1477

Mikhaleva, Y., Skinnes, R., Sumic, S., Thompson, E.M., Chourrout, D., 2018. Development of the house secreting epithelium, a major innovation of tunicate larvaceans, involves multiple homeodomain transcription factors. Dev. Biol. 443, 117–126. https://doi.org/10.1016/j.ydbio.2018.09.006

Miller, D.J., Ball, E.E., Technau, U., 2005. Cnidarians and ancestral genetic complexity in the animal kingdom. Trends Genet. https://doi.org/10.1016/j.tig.2005.08.002

Minelli, A., Fusco, G., 2013. Homology. https://doi.org/10.1007/978-94-007-6537-5_15

Moran, D., Softley, R., Warrant, E.J., 2015. The energetic cost of vision and the evolution of eyeless Mexican cavefish. Sci. Adv. 1. https://doi.org/10.1126/sciadv.1500363

Nagatomo, K., Fujiwara, S., 2003. Expression of Raldh2, Cyp26 and Hox-1 in normal and retinoic acid-treated Ciona intestinalis embryos. Gene Expr Patterns 3, 273–277. https://doi.org/10.1016/S1567-133X(03)00051-6

Naville, M., Henriet, S., Warren, I., Sumic, S., Reeve, M., Volff, J.N., Chourrout, D., 2019. Massive Changes of Genome Size Driven by Expansions of Non-autonomous Transposable Elements. Curr. Biol. 29, 1161-1168.e6. https://doi.org/10.1016/j.cub.2019.01.080

Navratilova, P., Danks, G.B., Long, A., Butcher, S., Manak, J.R., Thompson, E.M., 2017. Sex-specific chromatin landscapes in an ultra-compact chordate genome. Epigenetics and Chromatin 10. https://doi.org/10.1186/s13072-016-0110-4

Nelson-Sathi, S., Sousa, F.L., Roettger, M., Lozada-Chávez, N., Thiergart, T., Janssen, A., Bryant, D., Landan, G., Schönheit, P., Siebers, B., McInerney, J.O., Martin, W.F., 2015. Origins of major archaeal clades correspond to gene acquisitions from bacteria. Nature 517. https://doi.org/10.1038/nature13805

Nishida, H., 2008. Development of the appendicularian Oikopleura dioica: culture, genome, and cell lineages. Dev Growth Differ 50 Suppl 1, S239-56.

Nishino, A., Morisawa, M., 1998. Rapid Oocyte Growth and Artificial Fertilization of the Larvaceans Oikopleura dioica and Oikopleura longicauda. Zoolog. Sci. 15. https://doi.org/10.2108/zsj.15.723

Nishino, A., Satoh, N., 2001. The simple tail of chordates: phylogenetic significance of appendicularians. Genesis 29, 36–45.

Nishino, A., Satou, Y., Morisawa, M., Satoh, N., 2000. Muscle actin genes and muscle cells in the appendicularian, Oikopleura longicauda: phylogenetic relationships among muscle tissues in the urochordates. J Exp Zool 288, 135–150.

Novembre, J., Galvani, A.P., Slatkin, M., 2005. The geographic spread of the CCR5 Δ32 HIV-resistance allele. PLoS Biol. 3, 1954–1962. https://doi.org/10.1371/journal.pbio.0030339

Nüsslein-volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in drosophila. Nature 287. https://doi.org/10.1038/287795a0

Ohno, S., 1970. Evolution by gene duplication. Springer-Verlag, Heidelberg.

Olson, E.N., 2006. Gene Regulatory Networks in the Evolution and Development of the Heart. Science (80-.). 313, 1922–1927. https://doi.org/10.1126/science.1132292

Olson, M. V., Varki, A., 2003. Sequencing the chimpanzee genome: Insights into human evolution and disease. Nat. Rev. Genet. https://doi.org/10.1038/nrg981

Olson, M. V, 1999. When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet 64, 18–23.

Olsson, R., 1965. Comparative morphology and physiology of the Oikopleura notochord. Isr J Zool 14, 213–220.

Olsson, R., 1963. Endostyles and endostylar secretions: a comparative histochemical study. Acta Zool. 44, 299–329.

Omotezako, T., Matsuo, M., Onuma, T.A., Nishida, H., 2017. DNA interference-mediated screening of maternal factors in the chordate Oikopleura dioica. Sci. Rep. 7, 1–10. https://doi.org/10.1038/srep44226

Onuma, T.A., Isobe, M., Nishida, H., 2017. Internal and external morphology of adults of the appendicularian, Oikopleura dioica: an SEM study. Cell Tissue Res. 367, 213–227. https://doi.org/10.1007/s00441-016-2524-5

Panopoulou, G.D., Clark, M.D., Holland, L.Z., Lehrach, H., Holland, N.D., 1998. AmphiBMP2/4, an amphioxus bone morphogenetic protein closely related to Drosophila decapentaplegic and vertebrate BMP2 and BMP4: insights into evolution of dorsoventral axis specification. Dev Dyn 213, 130–9.

Papp, B., Notebaart, R.A., Pal, C., Pál, C., 2011. Systems-biology approaches for predicting genomic evolution. Nat Rev Genet 12, 591–602. https://doi.org/nrg3033 [pii]10.1038/nrg3033

Pascual-Anaya, J., Albuixech-Crespo, B., Somorjai, I.M.L., Carmona, R., Oisi, Y., Álvarez, S., Kuratani, S., Muñoz-Chápuli, R., Garcia-Fernàndez, J., Alvarez, S., Kuratani, S., Munoz-Chapuli, R., Garcia-Fernandez, J., 2013. The evolutionary origins of chordate hematopoiesis and vertebrate endothelia. Dev. Biol. 375, 182–192. https://doi.org/10.1016/j.ydbio.2012.11.015

Pastor-Satorras, R., Smith, E., Solé, R. V., 2003. Evolving protein interaction networks through gene duplication. J. Theor. Biol. 222. https://doi.org/10.1016/S0022-5193(03)00028-6

Perl, E., Waxman, J.S., 2019. Reiterative

mechanisms of retinoic acid signaling during vertebrate heart development. J. Dev. Biol. https://doi.org/10.3390/JDB7020011

Poelmann, R.E., Gittenberger-de Groot, A.C., 2019. Development and evolution of the metazoan heart. Dev. Dyn. 248, 634–656. https://doi.org/10.1002/dvdy.45

Puigbò, P., Lobkovsky, A.E., Kristensen, D.M., Wolf, Y.I., Koonin, E. V., 2014. Genomes in turmoil: Quantification of genome dynamics in prokaryote supergenomes. BMC Med. 12. https://doi.org/10.1186/s12915-014-0066-4

Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V, Jurka, J., Genikhovich, G., Grigoriev, I. V, Lucas, S.M., Steele, R.E., Finnerty, J.R., Technau, U., Martindale, M.Q., Rokhsar, D.S., 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science (80-.). 317, 86–94.

Quiring, R., Walldorf, U., Kloter, U., Gehring, W.J., 1994. Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans. Science (80-.). 265, 785–789.

Ragkousi, K., Beh, J., Sweeney, S., Starobinska, E., Davidson, B., 2011. A single GATA factor plays discrete, lineage specific roles in ascidian heart development. Dev Biol 352, 154–163. https://doi.org/S0012-1606(11)00023-6 [pii]10.1016/j.ydbio.2011.01.007

Razy-Krajka, F., Stolfi, A., 2019. Regulation and evolution of muscle development in tunicates. Evodevo 10, 1–34. https://doi.org/10.1186/s13227-019-0125-6

Reifers, F., Walsh, E.C., Léger, S., Stainier, D.Y.R., Brand, M., 2000. Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (fgf8/acerebellar). Development 127,

225–235.

Robison, B.H., Reisenbichler, K.R., Sherlock, R.E., 2005. Ocean science: Giant larvacean houses: Rapid carbon transport to the deep sea floor. Science (80-.). 308, 1609–1611. https://doi.org/10.1126/science.1109104

Ronshaugen, M., Mcginnis, N., Mcginnis, W., 2002. Hox protein mutation and macroevolution of the insect body plan.

Saga, Y., Kitajima, S., Miyagawa-Tomita, S., 2000. Mesp1 expression is the earliest sign of cardiovascular development. Trends Cardiovasc. Med. 10, 345–52. https://doi.org/10.1016/s1050-1738(01)00069-x

Salensky. W, 1903. Etudes anatomiques sur les Appendiculaires. I. Oikopleura vanhieffeni Lohmann. Mem. Acad. Sci. St. Petesbourg. Ser. 13, 1–44.

Salensky, W., 1904. Etudes anatomiques sur les Appendiculaires. II. Oikopleura refescens Fol. Mem. Acad. Sci. St. Petesbourg. Ser. 15, 1–54.

Sambasivan, R., Kuratani, S., Tajbakhsh, S., 2011. An eye on the head: The development and evolution of craniofacial muscles. Development. https://doi.org/10.1242/dev.040972

Satou, Y., Imai, K.S., Satoh, N., 2004. The ascidian Mesp gene specifies heart precursor cells. Development 131, 2533–2541. https://doi.org/10.1242/dev.01145

Seo, H.-C.C., Kube, M., Edvardsen, R.B., Jensen, M.F., Beck, A., Spriet, E., Gorsky, G., Thompson, E.M., Lehrach, H., Reinhardt, R., Chourrout, D., 2001. Miniature genome in the marine chordate Oikopleura dioica. Science (80-.). 294, 2506.

Seo, H.C., Edvardsen, R.B., Maeland, A.D., Bjordal,

M., Jensen, M.F., Hansen, A., Flaat, M., Weissenbach, J., Lehrach, H., Wincker, P., Reinhardt, R., Chourrout, D., 2004. Hox cluster disintegration with persistent anteroposterior order of expression in oikopleura dioica. Nature 431, 67–71. https://doi.org/10.1038/nature02709

Simakov, O., Marletaz, F., Cho, S.-J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.-H., Larsson, T., Lv, J., Arendt, D., Savage, R., Osoegawa, K., De Jong, P., Grimwood, J., Chapman, J.A., Shapiro, H., Aerts, A., Otillar, R.P., Terry, A.Y., Boore, J.L., Grigoriev, I. V, Lindberg, D.R., Seaver, E.C., Weisblat, D.A., Putnam, N.H., Rokhsar, D.S., 2013. Insights into bilaterian evolution from three spiralian genomes. https://doi.org/10.1038/nature11696

Simões-Costa, M.S., Vasconcelos, M., Sampaio, A.C., Cravo, R.M., Linhares, V.L., Hochgreb, T., Yan, C.Y.I., Davidson, B., Xavier-Neto, J., Simoes-Costa, M.S., Vasconcelos, M., Sampaio, A.C., Cravo, R.M., Linhares, V.L., Hochgreb, T., Yan, C.Y.I., Davidson, B., Xavier-Neto, J., 2005. The evolutionary origin of cardiac chambers. Dev. Biol. 277, 1–15. https://doi.org/10.1016/j.ydbio.2004.09.026

Somorjai, I.M.L., Martí-Solans, J., Diaz-Gracia, M., Nishida, H., Imai, K.S., Escrivà, H., Cañestro, C., Albalat, R., 2018. Wnt evolution and function shuffling in liberal and conservative chordate genomes. Genome Biol. 19. https://doi.org/10.1186/s13059-018-1468-3

Soviknes, A.M., Chourrout, D., 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. https://doi.org/10.1002/cne.21376

Spada, F., Steen, H., Troedsson, C., Kallesoe, T., Spriet, E., Mann, M., Thompson, E.M., 2001.

Molecular patterning of the oikoplastic epithelium of the larvacean tunicate Oikopleura dioica. J Biol Chem 276, 20624–20632.

Stach, T., 2007. Ontogeny of the appendicularian Oikopleura dioica (Tunicata, Chordata) reveals characters similar to ascidian larvae with sessile adults. Zoomorphology 126, 203–214. https://doi.org/10.1007/s00435-007-0041-5

Stach, T., Turbeville, J.M., 2002. Phylogeny of Tunicata inferred from molecular and morphological characters. Mol Phylogenet Evol 25, 408–428.

Stach, T., Winter, J., Bouquet, J.-M.M., Chourrout, D., Schnabel, R., 2008. Embryology of a planktonic tunicate reveals traces of sessility. Proc Natl Acad Sci U S A 105, 7229–7234.

Stolfi, Alberto, Gainous, T.B., Young, J.J., Mori, A., Levine, M., Christiaen, L., Field, H., Stolfi, A., 2010. Early chordate origins of the vertebrate second heart field. Science (80-.). 565, 565–569. https://doi.org/10.1126/science.1190181

Stolfi, A, Gainous, T.B., Young, J.J., Mori, A., Levine, M., Christiaen, L., Field, H., Stolfi, Alberto, 2010. Early chordate origins of the vertebrate second heart field. Science (80-.). 565, 565–569. https://doi.org/10.1126/science.1190181

Suga, H., Chen, Z., De Mendoza, A., Sebé-Pedrós, A., Brown, M.W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N., Torruella, G., Derelle, R., Manning, G., Lang, B.F., Russ, C., Haas, B.J., Roger, A.J., Nusbaum, C., Ruiz-Trillo, I., 2013. The Capsaspora genome reveals a complex unicellular prehistory of animals. Nat. Commun. 4. https://doi.org/10.1038/ncomms3325

Swalla, B.J., Cameron, C.B., Corley, L.S., Garey, J.R., 2000. Urochordates are monophyletic within the deuterostomes. Syst. Biol. 49, 52–

64. https://doi.org/10.1080/10635150050207384

Szathmáry, E., Jordán, F., Pál, C., 2001. Can genes explain biological complexity? Science (80-.). https://doi.org/10.1126/science.1060852

Technau, U., Rudd, S., Maxwell, P., Gordon, P.M.K., Saina, M., Grasso, L.C., Hayward, D.C., Sensen, C.W., Saint, R., Holstein, T.W., Ball, E.E., Miller, D.J., n.d. Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians.

Teichmann, S.A., Babu, M.M., 2004. Gene regulatory network growth by duplication. Nat. Genet. 36. https://doi.org/10.1038/ng1340

Thompson, E.M., Kallesøe, T., Spada, F., 2001. Diverse genes expressed in distinct regions of the trunk epithelium define a monolayer cellular template for construction of the oikopleurid house. Dev. Biol. 238, 260–273.

Tolkin, T., Christiaen, L., 2012. Development and Evolution of the Ascidian Cardiogenic Mesoderm, in: Current Topics in Developmental Biology. Academic Press Inc., pp. 107–142. https://doi.org/10.1016/B978-0-12-387786-4.00011-7

Tonissen, K.F., Drysdale, T.A., Lints, T.J., Harvey, R.P., Krieg, P.A., 1994. XNkx-2.5, a Xenopus gene related to Nkx-2.5 and tinman: Evidence for a conserved role in cardiac development. Dev. Biol. 162. https://doi.org/10.1006/dbio.1994.1089

Torres-Águila, N.P., Martí-Solans, J., Ferrández-Roldán, A., Almazán, A., Roncalli, V., D'Aniello, S., Romano, G., Palumbo, A., Albalat, R., Cañestro, C., 2018. Diatom bloom-derived biotoxins cause aberrant development and gene expression in the appendicularian chordate Oikopleura dioica. Commun. Biol. 1. https://doi.org/10.1038/s42003-018-0127-2

Troedsson, C., Bouquet, J.-M., Lobon, C., Novac, A., Nejstgaard, J., Dupont, S., Bosak, S., Jakobsen, H., Romanova, N., Pankoke, L., Isla, A., Dutz, J., Sazhin, A., Thompson, E., 2013. Effects of ocean acidification, temperature and nutrient regimes on the appendicularian Oikopleura dioica: a mesocosm study. Mar. Biol. 160, 2175–2187. https://doi.org/10.1007/s00227-012-2137-9

Volff, J.N., Lehrach, H., Reinhardt, R., Chourrout, D., 2004. Retroelement Dynamics and a Novel Type of Chordate Retrovirus-like Element in the Miniature Genome of the Tunicate Oikopleura dioica. Mol Biol Evol 21, 2022–2033. https://doi.org/10.1093/molbev/msh207

Wada, H., 1998. Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny. Mol. Biol. Evol. 15, 1189–1194.

Wang, K., Dantec, C., Lemaire, P., Onuma, T.A., Nishida, H., 2017. Genome-wide survey of miRNAs and their evolutionary history in the ascidian, Halocynthia roretzi. BMC Genomics 18. https://doi.org/10.1186/s12864-017-3707-5

Wang, W., Niu, X., Stuart, T., Jullian, E., Mauck, W.M., Kelly, R.G., Satija, R., Christiaen, L., 2019. A single-cell transcriptional roadmap for cardiopharyngeal fate diversification. Nat. Cell Biol. 21, 674–686. https://doi.org/10.1038/s41556-019-0336-z

Wang, W., Razy-Krajka, F., Siu, E., Ketcham, A., Christiaen, L., 2013. NK4 Antagonizes Tbx1/10 to Promote Cardiac versus Pharyngeal Muscle Fate in the Ascidian Second Heart Field. PLoS Biol. 11. https://doi.org/10.1371/journal.pbio.1001725

Weill, M., Philips, A., Chourrout, D., Fort, P., 2005. The caspase family in urochordates: distinct evolutionary fates in ascidians and larvaceans.

Biol Cell 97, 857–866. https://doi.org/BC20050018 [pii]10.1042/BC20050018

Wijesena, N., Simmons, D.K., Martindale, M.Q., 2017. Antagonistic BMP-cWNT signaling in the cnidarian Nematostella vectensis reveals insight into the evolution of mesoderm. Proc. Natl. Acad. Sci. U. S. A. 114, E5608–E5615. https://doi.org/10.1073/pnas.1701607114

Williams, J.B., 1996. Sessile lifestyle and origin of chordates. New Zeal. J. Zool. 23. https://doi.org/10.1080/03014223.1996.951 8072

Xavier-Neto, J., Castro, R.A., Sampaio, A.C., Azambuja, A.P., Castillo, H.A., Cravo, R.M., Simões-Costa, M.S., 2007. Cardiovascular development: Towards biomedical applicability - Parallel avenues in the evolution of hearts and pumping organs. Cell. Mol. Life Sci. 64, 719–734. https://doi.org/10.1007/s00018-007-6524-1

Xavier-Neto, J., Davidson, B., Simoes-Costa, M.S., Castro, R.A., Castillo, H.A., Sampaio, A.C.,

Azambuja, A.P., 2010. Evolutionary Origins of Hearts, in: Heart Development and Regeneration. Elsevier Inc., pp. 3–45. https://doi.org/10.1016/B978-0-12-381332-9.00001-3

Yadetie, F., Butcher, S., Forde, H.E., Campsteijn, C., Bouquet, J.M., Karlsen, O.A., Denoeud, F., Metpally, R., Thompson, E.M., Manak, J.R., Goksoyr, A., Chourrout, D., 2012. Conservation and divergence of chemical defense system in the tunicate Oikopleura dioica revealed by genome wide response to two xenobiotics. BMC Genomics 13, 55. https://doi.org/1471-2164-13-55 [pii]10.1186/1471-2164-13-55

Žárský, V., Tachezy, J., 2015. Evolutionary loss of peroxisomes - not limited to parasites. Biol. Direct 10. https://doi.org/10.1186/s13062-015-0101-6

Zufall, R.A., Rausher, M.D., 2004. Genetic changes associated with floral adaptation restrict future evolutionary potential. Nature 428, 847–850. https://doi.org/10.1038/nature02489nature02 489 [pii]