Contents

non-protein-coding RNAs as regulators of development in tunicates

Cristian A. Velandia-Huerto*, Federico D. Brown, Adriaan A. Gittenberger, Peter F. Stadler and Clara I. Bermúdez-Santana

Abstract Tunicates, or urochordates, are a group of small marine organisms that are found found widely throughout the seas of the world. As most plausible sister group of the vertebrates they are of utmost importance for a comprehensive understanding of chordate evolution, hence they have served as model organisms for many aspects of the developmental biology. Current genomic analysis of tunicates indicates that their genomes evolved with a fast rate not only at the level of nucleotide substitutions but also in terms of genomic organization. The latter involves genome reduction, rearrangements, as well as the loss of some important coding and non-coding RNA (ncRNAs) elements and even entire genomic regions that are otherwise well conserved. These observations are largely based on evidence from compara-

C.A. Velandia-Huerto,
Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, Universität Leipzig, Härtelstraße 16–18, D-04107, Leipzig, Germany.
Biology Department, Universidad Nacional de Colombia, Carrera 45 # 26-85, Edif. Uriel Gutiérrez, Bogotá, Colombia, *Correspondence e-mail: cristian@bioinf.uni-leipzig.de.

F.D. Brown,
Departamento de Zoologia, Instituto Biociências, Universidade de São Paulo, Rua do Matão, Tr. 14 no. 101, São Paulo, SP, Brazil
Laboratorio de Biología del Desarrollo Evolutiva, Departamento de Ciencias Biológicas, Universidad de los Andes, Cra 1 No. 18A-12, Bogotá, Colombia, e-mail: fdbrown@usp.br. A.A. Gittenberger,
Institute of Biology, Leiden University, P.O. Box 9505, 2300 RA, Leiden, Netherlands
GiMaRIS, BioScience Park Leiden, J.H. Oortweg 21, 2333 CH, Leiden, Netherlands
Naturalis Biodiversity Center, Darwinweg 2, 2333 CR,Leiden, Netherlands, e-mail: gittenberger@gimaris.com.

P.F. Stadler,
Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, Universität Leipzig, Härtelstraße 16–18, D-04107, Leipzig, Germany, e-mail: studla@bioinf.uni-leipzig.de.

C.I. Bermúdez-Santana,
Biology Department, Universidad Nacional de Colombia, Carrera 45 # 26-85, Edif. Uriel Gutiérrez, Bogotá, Colombia, e-mail: cibermudezs@unal.edu.co.

tive genomics resulting from the analysis of well-studied gene families such as the Hox genes and their non-coding elements. In this chapter the focus lies the ncRNA complement of tunicates, with a particular emphasis on microRNAs, which have already been studied extensively for other animal clades. MicroRNAs are known as important regulators of key genes in animal development and they are intimately related to the increase morphological complexity in higher metazoans. Here we review the discovery, evolution, and genome organization of the miRNA repertoire, which has been drastically reduced and restructured in tunicates compared to the chordate ancestor. Known functions of microRNAs as regulators of development in tunicates are a central topic. For instance we consider the role of miRNAs as regulators of the muscle development and their importance in the regulation of the differential expression during the oral siphon regeneration. Beyond microRNAs, we touch upon the functions of some other ncRNAs such as Yellow Crescent RNA, moRNAs, RMST lncRNAs, or spliced-leader (SL) RNAs, which have diverse functions associated with the embryonic development, neurogenesis and mediation of mRNA stability in general.

1 Introduction

Tunicates are organisms characterized by a fast rate of genomic and developmental evolution. Some fast evolving evolutionary changes include loss of synteny, fast changes in cis-regulatory sequences, and loss of several key regulatory developmental genes [71, 24], such as several central or posterior Hox genes involved in AP patterning of metazoans [37] and Gbx involved in the establishment of the midbrain-hindbrain boundary in vertebrates [89].

The role of noncoding RNAs in tunicate development has been studied since the 1990's, dating back to the seminal work by Swalla & Jeffry on the RNAs localized in the yellow crescent or myoplasm, a cytoskeletal domain in oocytes of the ascidian *Styela clava* [77]. The yellow crescent (YC) RNA identified to be present throughout embryonic development was the first example of what his now a large and rapidly growing class of ncRNAs with important roles in growth and development in tunicates. This asymmetrically distributed ascidian RNAs were part of the set of many other RNAs known as maternally synthesized cytoplasmically localized RNAs, discovered first in oocytes of Xenopus [6].

The current state of knowledge on ncRNAs in tunicates is far from comprehensive and complete. Nevertheless, in particular microRNAs have been studied already in some detail in several tunicates species, in particular, *Oikopleura dioica*, both Ciona species[1], *Didemnum vexillum*, and *Salpa thompsoni*. These studies have revealed many losses miRNA families that are very well conserved outside the tunicates. At the same time, many gains of unique miRNAs among recently divergent

[1] In accordance with the prevalent use in the ascidian community we use the term *Ciona robusta* reflecting that "Morphological evidence that the molecularly determined *C. intestinalis* type A and type B are different species: *Ciona robusta* and *C. intestinalis*" [12].

lineages in the tunicates when compared to other groups of chordates [26, 85]. Relaxed constraints in the evolution of genomes and developmental trajectories in the tunicates may have been responsible for the plethora of reproductive strategies, morphologies, and life histories observed in the group [36].

2 miRNA families origin and evolutionary perspective

2.1 Origins and Evolution of MicroRNAs

MicroRNAs (miRNAs) have been described in almost all animals and plants as well as diverse unicellular eukaryotes. They are important post-transcriptional regulators of gene expression affecting a sizable fraction of all mRNAs [4]. Mechanistically, miRNAs depends on the presence of the evolutionarily even older RNA interference pathways [15, 74] that leads to the suppression of double-stranded RNA molecules in a cell's cytoplasm.

Throughout animals, canonical miRNAs are the processed through a well-characterized pathway. The primary precursor transcript (pri-miRNA) is transcribed by pol-II. While in most cases the pri-miRNA is a long noncoding RNA, some miRNAs are processed from protein-coding transcripts, where they are mostly derived from introns [51]. In the next step, hairpin-shaped precursors, the pre-miRNAs, are extracted while the RNA is still residing in the nucleus. These are exported into the cytoplasm [53] and then processed further into miRNA/miRNA* duplexes. In the final step the single-stranded mature miR or its complement, the miR*, is incorporated in RISC complex. Sequence complementarity of miR and mRNA ensures the targeting specificity [5]. As as a consequence, miRNAs share a set of structural characteristics, most importantly the extremely stable secondary structure of the precursor hairpin and the 2-bp overhang of miR and miR* generated by Dicer processing. These features make it possible to reliably identify miRNAs from short RNA-seq data, see e.g. [45, 25, 46].

Most animal microRNAs are among the most highly conserved genetic elements. The most stringent selection pressure acts on the mature miR sequence. This is a consequence of the fact that a single miR typically targets a large number of mRNAs. Mutations in the mature sequence thus simultaneous affect many interactions, and thus are almost always selected against. In conjunction with the stringent requirements on the secondary structure, the entire precursor is under strong stabilizing selection [68], explaining the observed high levels of sequence conservation. As a consequence, even evolutionarily distant homologs of miRNAs can be readily detected despite the short sequence length. Most efficiently, `infernal` [60] is used for this purpose, since it makes use of both sequence and structure comparison. The evolution of miRNAs can thus be traced back in time with high accuracy [34].

Like other gene families, miRNAs form paralogs [80, 33] and hence often appear as families as homologous genes. This forms the basis of the `miRBase` nomen-

clature [3]. A series of investigations into the phylogenetic distribution of miRNA families showed that miRNAs are infrequently lost at family level and thus serve as excellent phylogenetic markers [72, 31, 30, 87], although the massive restructuring of the miRNA complement of tunicates is an important exception to this rule [26].

The innovation of new miRNA families is an on-going process. Experimental surveys of the miRNA repertoir thus have reported a large number of very young and even species-specific miRNAs [7, 9]. The process was studied quantitatively in fruit flies, where innovation rate of as many as 12 new miRNA genes per million years has been estimated [52]. This is consistent with the fact that stable hairpins are abundant structural elements in random RNAs, which makes is not only possible but actually quite likely that miRNA precursors appear by chance in transcribed genomic regions [80, 13, 54]. Of course, only a tiny fraction of these fortuitously processed hairpins have a function and hence are subject to selection, and an even smaller subset is conserved over long evolutionary time scales. Detailed studies showed that evolutionarily young miRNA have comparably low expression levels. Initially, they go through a phase of relatively fast sequence evolution [50, 58], which slows down as the selective pressures from a gradual increase in the number of target site increases. A large, diverse set of targets then protects against miRNA loss [47]. The rate of gain of miRNA families that retained essentially permanently amounts to only 1 per several million years. This number is consistent with divergence of the miRNA complements between animal phyla.

Many authors have observed that overall the miRNA repertoire has been expansing throughout animal evolution in a manner that at least roughly correlates with morphological complexity [34, 72, 63, 69, 47, 31, 67, 8]. Several bursts of miRNA innovation have been observed [34, 31, 79, 35], most notably at the root of the placental mammals, the ancestor of "free-living" nematodes, or the radiation of the drosophilids. Massive morphological simplification, on the other hand, is sometimes associated with a drastic loss of miRNA families. This has been observed most prominently for tunicates [26, 22].

2.2 miRNA identification and validation

The first miRNA reported in any tunicate was let-7, which was first detected in *Ciona robusta* and *Herdmania curvata* [66]. A previous study the same year in *C. elegans* had shown that small RNA let-7 (21 nt) was required for late larval to adult developmental transition [70]. Small RNA let-7 was then shown to also be differentially expressed during the development of many distantly related animal taxa, but was not detected in Porifera, Ctenophora, Cnidaria, and Acoelomorpha, suggesting that let-7 was involved in the regulation of late temporal transitions during development or in the evolution of complex life histories in the Nephrozoa [66, 65].

The first systematic computational screen of a tunicate genome in 2005 followed the first large animal sequencing projects. Beginning with *C. robusta* and *C. savignyi*, a profile-based strategy was implemented in the ERPIN program [49].

This work detected a set of new miRNAs candidates considered as *C. robusta* specific, such as the members of the family miR-9 and miR-79 together with many other miRNA families that, as expected, were found to be homologous between both Ciona species. Among these were miR-124, miR-92, miR-98, miR-325, the miR310-313 group, and let-7. In the same year, a whole-genomic comparative approach in the urochordate lineage was performed on the species *C. robusta*, *C. savignyi*, and *O. dioica*. Using a computational screening of structured ncRNAs based upon homology between predicted precursor hairpin structures 41 miRNA candidates were detected including let-7 and other six known candidates in *C. robusta* [59]. After all, the same group in 2007 implemented a structure-based clustering approach in *C. robusta* predicted 58 miRNAs, of which only let-7, miR-7, miR-124, and miR-126 coincided with the previously annotated miRNAs [88].

Up to this point in time, efforts to map the miRNAs in urochordate lineages were mainly focused on computational approaches. With improved sequence technologies, new hybrid strategies were employed that combined computational and experimental studies to validate candidate families previously detected. For instance, the first bona fide records for *C. robusta* miRNAs were registered in mirBase only in Release 11 based on the data reported in [64]. This study searched the genomes of *C. robusta* and *C. savignyi* for conservation of the seed regions of the known mature miRNA sequences compiled in the 2006 release of miRBase release. Those miRNAs were aligned locally using the FASTA/ssearch34 program, retaining only matches with at least 90% identity. The characteristic hairpin structure and the relative positioning of the mature sequences were used as additional filters. After manual curation a set of 18 miRNAs that appeared conserved in both Cionas was reported. The expression of 14 of these families, including let-7, miR-7, and miR-126, was then confirmed by Northern blot analyses in the adult tissue of *C. robusta*.

Until 2008, most of the miRNAs annotations were concentrated in Cionas, but new annotation approaches for other species in tunicates were appearing slowly to increase then the repertory of new miRNAs families in urochordates. The first repertory of miRNAs in tunicate beyond the Ciona species was published in 2008 for the larvacean *O. dioica* [26]. At that time the Fu *et al.* were studying the temporal-spatial expression patterns of conserved miRNAs in different developmental stages of oocytes, 1-cell zygote, 2-8 cell embryons, blastulas, gastrulas, tadpoles (in different stages) and animals from 1 to 6 days from *O. dioica*. Small RNAs were isolated, amplified by RT-PCR and rapid amplification of cDNA ends (RACE) of the developmental stages, cloned and sequenced. Blast searches using the sequences of cloned small RNA libraries were used to annotate small RNAs as miRNA candidates. In further steps the recovered genomic flanking sequences each side of those mapped candidates were used as input to predicted secondary structures by mfold v3.1. This step was used to detect candidates that folds like miRNA hairpins and aimed to decrease the set of false positive potential miRNAs in *O. dioica*. Finally, for this set of potential candidates a developmental miRNA array dot blot analyses were performed to detect miRNA expression. With this approach from 3066 sequenced small RNA clones, expression was detected for only 55 miRNAs. The authors suggested that these candidates were expressed throughout the short life cycle of *O.*

dioica showing that some of them were stocked as maternal determinants prior to rapid embryonic development. The authors also identified a set of sex-specific miRNAs that appeared as male/female gonad differentiation which became apparent and was maintained throughout spermatogenesis [26]. Unexpectedly, the majority of the miRNAs loci in *O. dioica* were located in antisense orientations in the host genes, in contrast to the majority of the mammalian miRNAs known at the time.

Between the years 2009 and 2015 the majority of the studies of miRNAs in tunicates were focused into the validation of expression of computational predicted miRNAs in Cionas specially focused in *C. robusta* as model organism of tunicates or into the test of new computational approaches as miRTRAP, miRDeep2 and miRRim2 which used next-generation sequencing libraries of small RNAs derived from *C. robusta* to validate their algorithms. Then by the year 2016 the first comparative homology based search strategy let us to identify the repertory on miRNAs and other ncRNAs in the carpet sea squirt *Didemnum vexillum* with a preliminary comparative analysis of gain and losses of miRNA families on chordates which included the *Cionas*, *O. dioca* and the colonial tunicate *Botryllus schlosseri* [85]. By the same year, from the preliminary genome sequence assembled for the Southern Ocean salp, *Salpa thompsoni* (Urochordata, Thaliacea) a set of miRNAs families were detected [39] and in 2017 the prediction of miRNAs families were reported to the species *Halocynthia roretzi* [86].

2.2.1 High throughput studies of Ciona miRNAs

At the end of the last decade the application of next generation sequencing technologies to sequence small RNA libraries changed the common way used to detect expression of miRNAs in many organisms including the tunicates. This technology became in one of the most common approaches that supported methods like RT-PCR, microarrays or dot blotting which were previously used to validate miRNA expression in tunicates. In 2009 small RNA libraries prepared from various developmental stages including unfertilized eggs, early embryos, late embryos and adults from *C. robusta* were assayed by high-throughput sequencing of cDNA with an Illumina 1G Genome Analyzer. These sequencing data revealed 80 miRNAs families in *C. robusta*. Unexpectedly, a distinct species of small RNAs processed from the distal parts of the miRNA precursor hairpins was found to be abundantly expressed. These were termed miRNA-offset RNAs (moRs) [75]. Later on, after extracting non-coding conserved regions of whole genome alignments between *C. robusta* and *C. savigny* a set of 12 million sequences were computationally folded using RNAfold and mfold. Then after combining the following criteria: structure/sequence conservation, homology to known miRNAs, and phylogenetic footprinting the authors detected a set of 458 candidate sequences [40]. Then in order to validate those candidate, RT-PCR and PAGE were conducted to design a custom microarray. 244 of the 458 miRNA predictions were represented either in their microarray data or in the Illumina sequences from [75], leading to an estimate of about 300 miRNA genes in *C. robusta*. It is worth noting that [40] did not recover 39 previously character-

ized miRNAs. A novel computational strategy for the systematic, whole-genome identification of microRNA from high throughput sequencing information was developed in 2010 in [32] and applied to *C. robusta*. This method, miRTRAP, relies on the sequence patterns produced by mechanisms of microRNA biogenesis but also includes additional criteria regarding the prevalence and quality of small RNAs arising from the antisense strand and the neighboring loci. With that approach, nearly 400 putative microRNAs loci were detected. The miRTRAP approach is depends crucially on the depth of the small RNAs mapped to a given locus and requires a very accurate assignment of small RNA sequences on their relative positions along the hairpin, that is, miR/miR*, moR/moR* and loop [32]. A further improvement in the analysis of miRNA sequencing data became available with miRDeep2, a revised version miRDeep [25]. It was reported to identify with an accuracy of 98.6% and 99.9% canonical and non-canonical miRNAs in different species. A reanalysis of the small RNA data from [75] with miRDeep2 reported 313 known and 127 novel miRNAs in *C. robusta*. In the same year the program miRRim2 [81] was applied to the *C. robusta* genome, in which some candidates identified from the work of [32] and the several promising candidates were detected.

In 2013, [43] investigated the expression patterns of the cluster miR-1 and miR-133 in *C. robusta* and in *C. savignyi*. RT-PCR amplification of miR-1/133 precursors were performed and PCR products were subcloned and sequenced. Whole-mount in situ hybridization to detect cin-miR-1/miR-133 primary transcript was performed and LNA Northern blotting was conducted on different developmental stages.

2.2.2 High throughput miRNA searches in other Urochordates

Since 2016 new approximations has increased our knowledge about new families in other tunicates thanks to the sequence of new urochordate genomes of the species *D. vexillum*, *S. thompsoni* and *H. roretzi*. For the draft genome sequence from *D. vexillum* an homology-based computational approach was applied [85]. Blast and HMMer searches were performed with annotated small ncRNAs sequences from metazoans and hidden markov models from RFAM[2] to obtain the sort of candidates at sequence level. Structural alignments of those sequences were performed by infernal [61], using metazoan-specific covariance models to annotate the small ncRNAs collection, which 57 families and 100 loci of miRNAs were found.

For the preliminary assembled of the genome sequence for the Southern Ocean salp *S. thompsoni* [39] were small RNA libraries constructed to be sequenced on an Illumina Hiseq 2000. After filtering data sets to 18-24 nt for miRNA and 28-32 nt for piRNA, the reads were aligned to *S. thompsoni* genome and miRNA gene folding predictions were performed using RNAfold. In this initial survey of small RNAs, were revealed the presence of known, conserved miRNAs, as well as novel miRNA genes and mature miRNA signatures for varying developmental stages. Then in 2017, the prediction of 319 miRNAs candidates in *H. roretzi* were obtained

[2] http://rfam.xfam.org/

through three complementary methods. The experimental validation suggested that more than half of these candidate miRNAs are expressed during embryogenesis. The expression of some of the predicted miRNAs were validated by RT-PCR using embryonic RNA. In this approach *C. robusta* small RNA-Seq reads derived from *C. robusta* [75] (previously known as *C. intestinalis* today reclassified) was used to identify conserved miRNAs in *H. roretzi* [86].

Current repertoire of miRNAs (Figure 8) is based on the final matrix of miRNAs families from [35] and complemented by homology methods developed on **??**, specifically for *S. thompsoni* and *H. roretzi* blast searches with structural alignments where applied on the reported candidates in [39] and [86].

2.3 miRNA in clusters

One of the most interesting aspects about the patterns of genomic locations of miRNAs is to known whether those loci are randomly distributed throughout the genome as single copies or if they are arranged on consecutive locations or in tandem copies clustered to be expressed from polycistronic primary precursors or to be transcribed independently [79]. Interestingly in *O. dioca* miRNAs are located in the antisense orientations of protein-coding gene and immediately downstream of its corresponding 3‘UTR region or even more in the sense strand of introns [26]. Nevertheless, after those conspicuous distributions some clusters have been also identified in *O. dioca*. For instance four miRNAs, miR-1490a, miR-1493, miR-1497d, and miR-1504, are reported by [26] to be presented as two copies, and miR-1497d-1 and miR-1497d-2 are included in the large miR-1497 cluster. See the current structure of this cluster in Table 1 although only one copy for the miR-1497 has been reported for *C. robusta* located in an intergenic region [26, 32] and one in *C. savigny* overlapped in an intron [26]. By testing real time PCR co-expression of some miRNAs, their host and adjacent genes in *O. dioca* by [26] it was discovered for the case of the cluster miR-1487/miR-1488 a not clear positive or negative correlation with the expression of its anti-sense hosting gene. In males this cluster expression was not associated with the expression of its adjacent ABCA3 gene by the same authors.

Table 1: Details of biggest miRNA cluster for chordate species.

Specie	**Chr**	**Start**	**End**	**Size (Mb)**	**No.**	**miRNAs detail**
B. floridae	Bf_V2_118	216744	220351	3607	5	bfl-mir-4869, bfl-mir-4857, bfl-mir-4862, bfl-mir-4856b, bfl-mir-4856a

O. dioica	scaffold_3	2222857	2223714	857	6	odi-mir-1497e, odi-mir-1497d-2, odi-mir-1497d-1, odi-mir-1497c, odi-mir-1497b, odi-mir-1497a
B. schlosseri	chrUn	40003	41320	1317	2	mir-233, mir-10
C. robusta	7	4153284	4156782	3498	23	cin-mir-4006d, cin-mir-4006c, cin-mir-4001b-2, cin-mir-4000i, cin-mir-4006g, cin-mir-4001e, cin-mir-4001d, cin-mir-4000g, cin-mir-4006f, cin-mir-4006b, cin-mir-4001b-1, cin-mir-4000c, cin-mir-4006e, cin-mir-4000b-2, cin-mir-4001a-1, cin-mir-4000b-1, cin-mir-4002, cin-mir-4000d, cin-mir-4001h, cin-mir-4000a-2, cin-mir-4006a-2, cin-mir-4006a-3, cin-mir-4006a-1
C. savignyi	reftig_16	3924783	3925336	553	3	csa-mir-216b, csa-mir-216a, csa-mir-217
C. savignyi	reftig_1	1335375	1336487	1112	3	csa-mir-92b, csa-mir-92c, csa-mir-92a

D. rerio	4	28738556	28754891	16335	60	dre-mir-430a-18, dre-mir-430c-18, dre-mir-430b-4, dre-mir-430a-15, dre-mir-430c-18, dre-mir-430b-5, dre-mir-430a-10, dre-mir-430c-18, dre-mir-430b-5, dre-mir-430a-15, dre-mir-430c-18, dre-mir-430b-3, dre-mir-430a-10, dre-mir-430c-18, dre-mir-430b-8, dre-mir-430a-15, dre-mir-430c-18, dre-mir-430b-5, dre-mir-430a-17, miR-430, dre-mir-430b-20, dre-mir-430a-10, dre-mir-430c-18, dre-mir-430b-5, dre-mir-430i-3, dre-mir-430c-18, dre-mir-430b-3, dre-mir-430a-10, dre-mir-430c-18, dre-mir-430b-8, dre-mir-430a-11, dre-mir-430c-18, dre-mir-430b-5, dre-mir-430i-3, dre-mir-430c-18, dre-mir-430b-19, dre-mir-430a-10, dre-mir-430c-18, dre-mir-430b-5, dre-mir-430a-17, miR-430, dre-mir-430b-20, dre-mir-430a-10, dre-mir-430c-18, dre-mir-430b-5, dre-mir-430i-3, dre-mir-430c-18, dre-mir-430b-19, dre-mir-430a-10, dre-mir-430c-18, dre-mir-430b-5, dre-mir-430a-15, dre-mir-430c-18, dre-mir-430b-3, dre-mir-430a-10, dre-mir-430c-18, dre-mir-430b-8, dre-mir-430a-15, dre-mir-430c-18, dre-mir-430b-5

L. chalumnae	JH126646.1	1529355	1882777	353422	7	mir-233, mir-233, mir-672, mir-233	mir-233, mir-598, MIR535,

In *C. robusta* some miRNAs are also located in introns and a small class of miRNAs are found to be deriving from mature mRNAs encoded within exons or UTR sequences [32] in contrast to the location of the loci in antisense orientations of protein-coding gene as seen in *O. dioca* but this antisense orientation is reported for some miRNAs loci which express antisense miRs derived from miRNA loci as antisense products and antisense moR products as the miR-2246. Only 44 loci appeared to be expressed as antisense products from the 300 miRNA loci predicted in 2010 by [32]. In Cionas have been also detected miRNAs organized in clusters, for example in *C. robusta* a putative cluster was detected by [40] using microarray analysis that shows a similar loci organization to the cluster let-7/miR-125/miR-100 observed in Drosophila. The miR-1473 was later classified as the orthologue of miR-100 in the analysis derived from the comparison of the evolution of this cluster conducted by [29]. The authors suggested that mir-100, mir125 and let7 are clustered in most of the bilaterian genomes including as 1473 as orthologue of mir-100.

Current analysis of this cluster shows that the distribution of miRNAs families on this let-7 cluster are distributed in all the studied chordate species. In vertebrate species like (*D. rerio* and *L. chalumnae*) exists more than one let-7 cluster, extending the loci definition which is not restricted only for one element but for a cluster of many locus with different length distributions. It is important to see that let-7 is organized sometimes with another let-7 locus or with another miRNA's loci families. The distribution of this cluster reported on amphioxus is composed by 2 let-7 and 3 mir-10 (1 bfl-mir-100, 1 bfl-mir-125a and 1 bfl-mir-125b), this cluster architecture almost conserved on vertebrates that apparently inverted the order and split the relation between let-7 and mir-10, creating two different cluster order groups: let-7 + mir-10 and let-7 + other families. In this way, tunicates reported the latter group, not including mir-10 on the cluster but including mir-233, mir-1473 or mir-125.

A second miRNA cluster consisting of the miR-182 and miR-183 was also detected in *C. robusta* in 2010 by [40] which is in the current predictions is reported another member locus the miR-96 organized in the middle of those loci as is shown in the plot 7. Here the authors also found five additional paralogs of let-7 within a 1-kb stretch, but it is important to know that those elements had been identified on chromosome 4q on Ensembl release 54 version, at the current version only two of those elements have been identified by homology approaches (Figure 1).

The cluster miR-1/miR-133, expressed specifically in Cionas muscle tissues was also reported by [43]. The authors reported that one copy of this cluster is presented in both Cionas. As is shown in the plot 3 a copy is also presented in *L. chalumnae*. In 2012 a new cluster was proposed in *C. robusta* by [81] located on the chromosome 10q and composed by the mir-4054 locus and the mir-4091. In the current distribution of this cluster a new annotated family the mir-4008 with three paral-

ogous is located on the middle of those loci. This current distribution in shown of Table 2 whose loci were validated by [64], [26], [32], and [81]. As was mentioned by [32] is not very common to find related miRNAs organized in clusters composed by closely related families that differ in just a single nucleotide in the seed sequence as was found on the cluster composed by nine Ci-mir-2200, seven Ci-mir-2201 and nine Ci-mir2203 which were previously reported under that putative names. They also found a second large cluster composed of 11 miRNAs that gather into 4 paralogous families three of Ci-mir-2200, three Ci-mir-2201, four Ci-mir-2204 and two Ci-2217. Current distribution of miRNAs families in *C. robusta* and curated annotations indicate than in other regions of the chromosome 7q are also organized miRNAs in tandem copies of families. For instance the big cluster presumably re-named from [32] today is known to be built by the families miR4000, miR-4001, miR4002, and miR4006 located on chromosome 7q (Table 2). Another cluster is also located on the same chromosome composed by the families miR-4003, miR4005 and miR4077 in Table 2. Some other cluster are also found on the chromosome 1a, 10q and 3p. See this structure on Table 2, most of them validated by [32].

Some other clusters shared between both Cionas are the cluster 92, 124 and 200 validated by [64], [26], [32] which the structure is seen on Table 2.

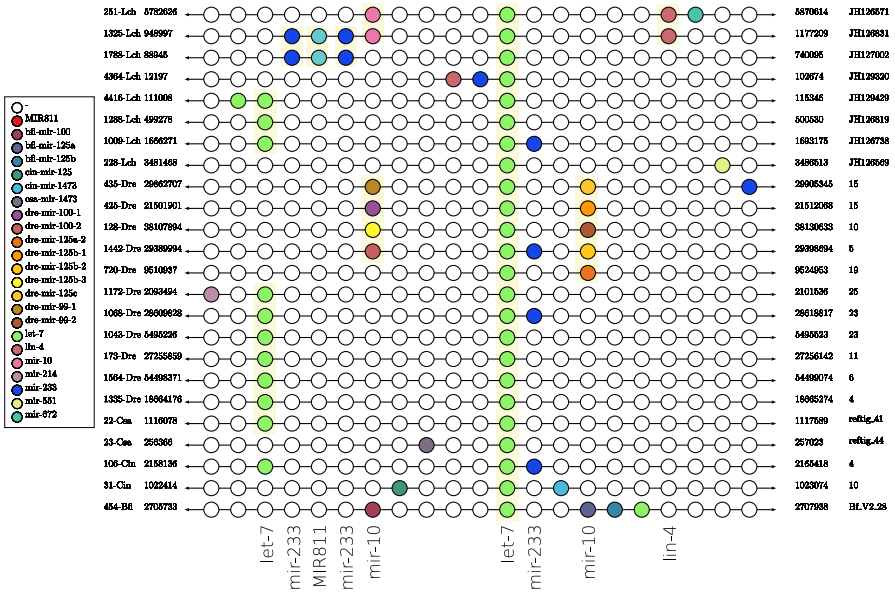

Fig. 1 Multiple alignment of let-7 clusters. Specific names from annotations and homology predictions are described at the legend. Names from miRBase families are reported at the bottom of the aligned elements.

Table 2: Reported clusters on literature. Bold text represent those miRNAs elements that are currently annotated and validated, but could not possible to detect by homology strategies. **Ciro**: *C. robusta*, **Cisa**: *C. savignyi* and **Oidi**: *O. dioica*

Specie	Chr	Start	End	miRNAs	Comments	Source DB	Ref.
Ciro	4q	2082260	2083286	**cin-let-7a-1, cin-let-7f, cin-let-7b, cin-let-7c, cin-let-7a-2**	Reported on miRBAse and annotated on Ensembl	miRBase	[32], [26]
Cisa	reftig_41	1114139	1117597	**csa-let-7c-1, csa-let-7b, csa-let-7a, csa-let-7c-2**	Reported on miRBase and does not detected by homology strategies.	miRBase	[26]
Ciro	10q	3226200	3228884	**cin-mir-34**, cin-mir-4091, cin-mir-4008a, cin-mir-4008c, cin-mir-4008b, cin-mir-4054	NA	miRBase, Homology	[64], [26], [32], [81]
Ciro	7q	4828431	4835967	cin-mir-4077b, cin-mir-4003b, cin-mir-4005b, cin-mir-4077d, cin-mir-4003a-1, cin-mir-4003c, cin-mir-4077a, cin-mir-4003a-4, cin-mir-4003d	NA	miRBase, Homology	[32]
Ciro	3q	567478	571031	cin-mir-4001f, cin-mir-4000e, cin-mir-4001c, cin-mir-1502d, cin-mir-4018a, cin-mir-4019, cin-mir-1502b, cin-mir-1502a, cin-mir-4007, cin-mir-4000f, cin-mir-4001i, cin-mir-4018b, cin-mir-1502c	Inclusion of cin-mir-4019 and cin-mir-4007	miRBase, Homology	[32]
Ciro	HT000037.1	4884	5250	**cin-mir-367**, **cin-mir-4009c**, **cin-mir-4009b**, cin-mir-367, cin-mir-4009c, cin-mir-4009a, cin-mir-4009b	Non-highlighted names could be found in the current genome at HT000037.1 scaffold	miRBase, Homology	[32]
Ciro	10q	3226200	3228884	**cin-mir-34**, cin-mir-4091, cin-mir-4008a, cin-mir-4008c, cin-mir-4008b, cin-mir-4054	NA	miRBase, Homology	[32]
Ciro	3q	884615	885508	cin-mir-92a, cin-mir-92d, cin-mir-92c	NA	miRBase, Homology	[32]
Cisa	reftig_1	1335375	1336487	csa-mir-92b, csa-mir-92c, csa-mir-92a	NA	miRBase, Homology	[26]

Oidi	scaffold_1	3086369	3086586	odi-mir-92b, odi-mir-92a	NA	Homology	[26]
Ciro	7q	4969691	4969912	cin-mir-124-1, cin-mir-124-2	NA	miRBase, Homology	[32], [26]
Cisa	reftig_262	49392	49620	csa-mir-124-1, csa-mir-124-2	NA	miRBase, Homology	[26]
Ciro	HT000325.1	8331	8778	cin-mir-200, **cin-mir-3575**, cin-mir-141, **cin-mir-5611**	NA	miRBase, Homology	[64], [26], [32], [25]
Cisa	reftig_613	31353	31949	csa-mir-200, csa-mir-141	NA	miRBase, Homology	[26]
Ciro	7q	4153284	4156782	cin-mir-4006d, cin-mir-4006c, cin-mir-4001b-2, cin-mir-4000i, cin-mir-4006g, cin-mir-4001e, cin-mir-4001d, cin-mir-4000g, cin-mir-4006f, cin-mir-4000h*, cin-mir-4006b, cin-mir-4001b-1, cin-mir-4006e, cin-mir-4001a-1, **cin-mir-4001a-2**, cin-mir-4002, cin-mir-4001h, cin-mir-4000a-2, cin-mir-4006a-2, cin-mir-4006a-3, cin-mir-4006a-1, **cin-mir-4006e**, **cin-mir-4001a-1**, **cin-mir-4006b**, cin-mir-4000c, **cin-mir-4006e**, cin-mir-4000b-2*, **cin-mir-4001a-1**, cin-mir-4000b-1*, **cin-mir-4001a-2**, **cin-mir-4002**, cin-mir-4000d*, **cin-mir-4001h**, **cin-mir-4006a-3**, **cin-mir-4006a-1**, **cin-mir-4000a-1**	Elements marked with * are identified by homology strategies at the same cluster, but in another order reported by miRBase.	miRBase, Homology	[32]

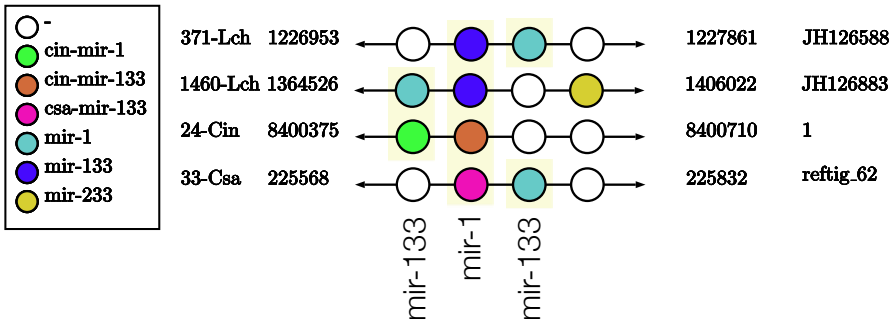

Fig. 2 mir-1/mir-133

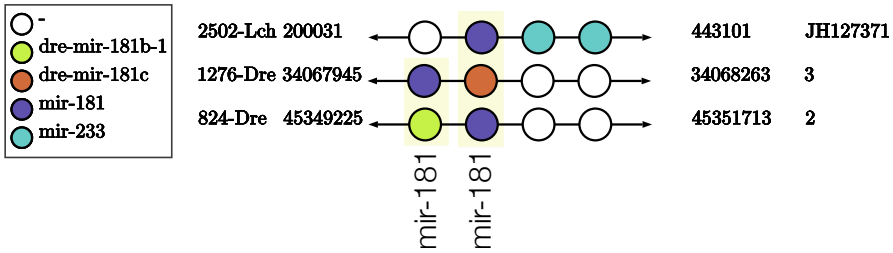

Fig. 3 mir-181

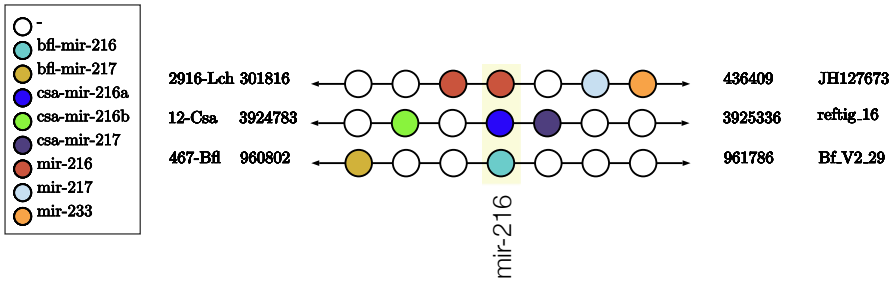

Fig. 4 mir-216/mir-217

3 miRNAs and its rol in development

3.1 miRNAs discovery and development

Both MicroRNAs (miRNAs or miRs) as well as MicroRNA offset RNA (moRNAs or moRs) are developmentally regulated as shown during *C. robusta* development [75]. In spite of the considerably higher abundance of miRs and miRs* in cells than their corresponding abundance of moRs, all three small RNA types have been shown to have regulatory roles for gene expression. Although a vast majority of miRNAs remain to be studied, there are already many cases of well-studied miRNAs (in-

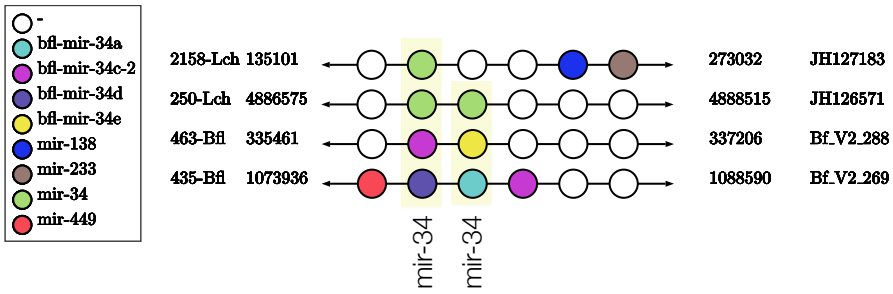

Fig. 5 mir-34

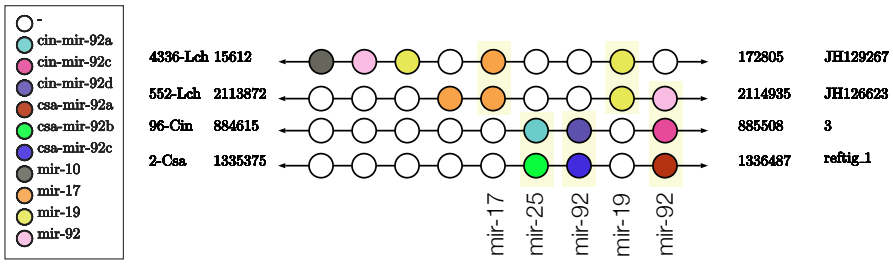

Fig. 6 mir-92

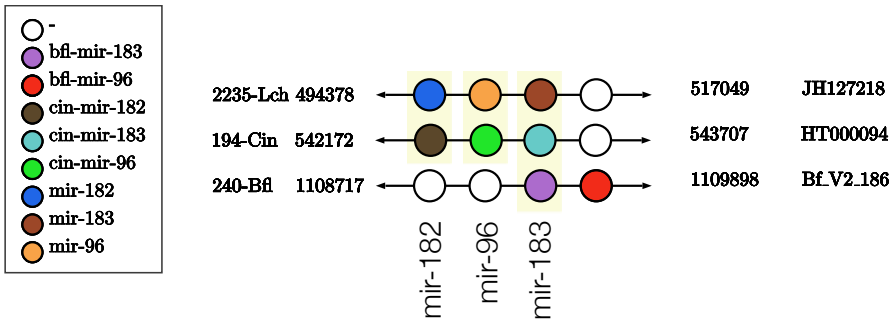

Fig. 7 mir-182/mir-96/mir-183

cluding many that are mentioned in this chapter that have been studied in tunicates) that are known to target mRNAs, modulate their levels of expression, and affect developmental processes both in plants and animals [91]. Only recently two studies demonstrated for the first time that two moRs (viral moR-rR1-3-5p and moR-21) could also modulate gene expression, and were not merely the byproduct of miRNA biogenesis [83, 90].

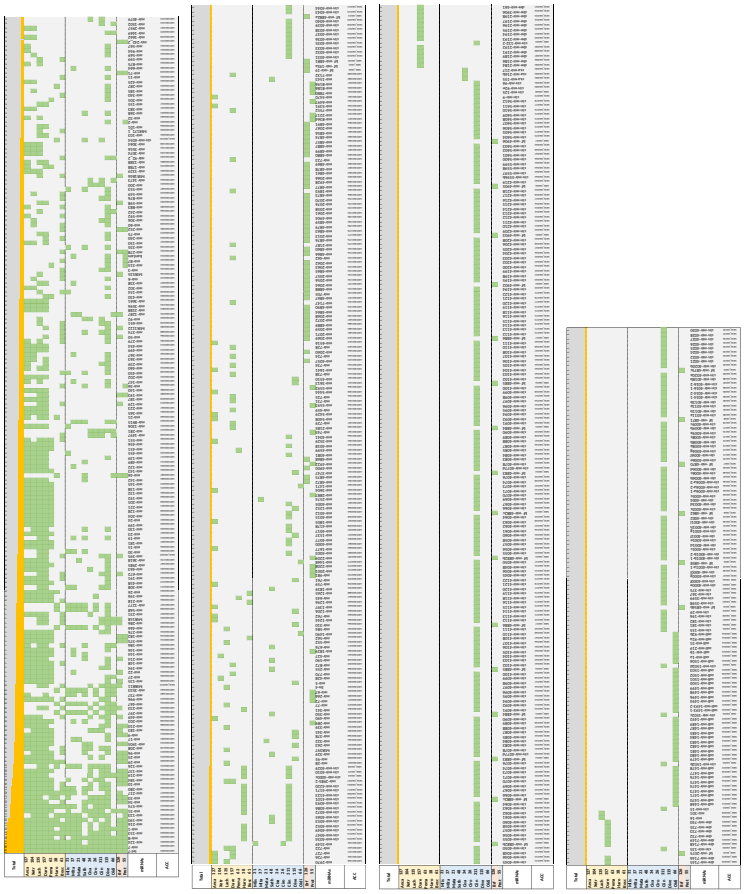

Fig. 8 Absence/Presence Matrix of miRNAs families along Bilaterian species. **Prot**: Protostomata, **Brfl**: *B. floridae*, **Oidi**: *O. dioica*, **Dvex**: *D. vexillum*, **Ciin**: *C. robusta*, **Cisa**: *C. savignyi*, **Ciro**: *C. robusta*, **Sath**: *S. thompsoni*, **Mata**: *M. oculata*, **Mlta**: *M. occulta*, **Mlis**: *M. occidentalis*, **Bosc**: *B. schlosseri*, **Haro**: *H. roretzi*, **Pema**: *P. marinus*, **Dare**: *D. rerio*, **Lach**: *L. chalumnae*, **Xetr**: *X. tropicalis* and **Anca**: *A. carolinensis*.

3.2 Neuronal fate determination and regulation by miR-124

The miRNA miR-124 is expressed in the nervous system of many animals, including Drosophila [1], *C. elegans* [20] and humans [73]. As was first observed by in vitro studies of mouse brain cells, low expression of miR-124 was related to neural stem cell maintenance, whereas high expression of miR-124 induced the differentiation of neuronal cell types [19]. A regulative role of miR-124 in non-neural vs. neural fate decisions was further investigated by embryonic experiments in vivo [17] and by theoretical and in silico modeling analyses in *C. robusta* [16]. These studies showed that miR-124 promotes nervous system development by feedback interac-

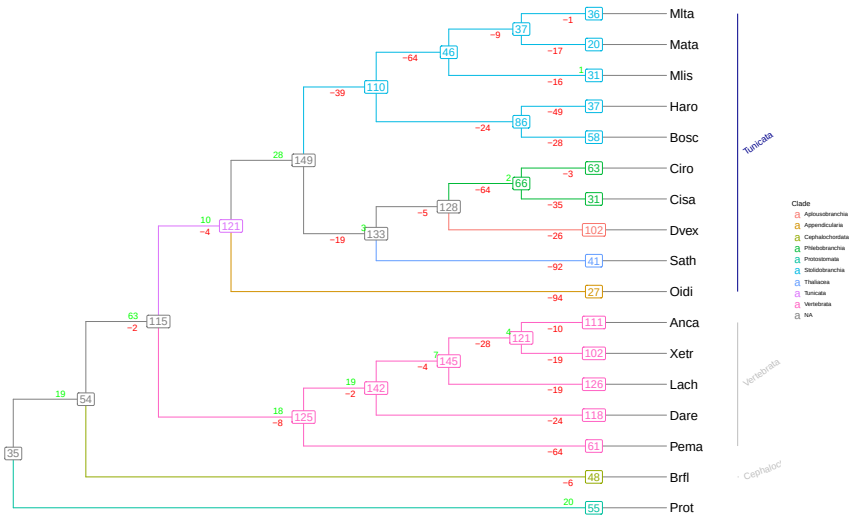

Fig. 9 Dollo parsymony of miRNAs families distribution in some chordates genomes. **Prot**: Protostomata, **Brfl**: *B. floridae*, **Oidi**: *O. dioica*, **Dvex**: *D. vexillum*, **Ciro**: *C. robusta*, **Cisa**: *C. savignyi*, **Sath**: *S. thompsoni*, **Mata**: *M. oculata*, **Mlta**: *M. occulta*, **Mlis**: *M. occidentalis*, **Bosc**: *B. schlosseri*, **Haro**: *H. roretzi*, **Pema**: *P. marinus*, **Dare**: *D. rerio*, **Lach**: *L. chalumnae*, **Xetr**: *X. tropicalis* and **Anca**: *A. carolinensis*. The phylogenetic distribution of this species were obtained from [23, 42].

tions with Notch signaling. During nervous system development of *C. robusta*, cells in the dorsal and ventral midline epidermis of the teilbud embryo either take an epidermal sensory neuron (ESN) or peripheral nervous system (PNS) fate, a decision mediated by lateral inhibition using a classical model of feedback loop regulation of Notch-Delta signaling in neighboring cells [21, 16]. Cells that take an ESN fate showed low expression of miR-124 presumably by Notch inhibition, whereas cells that take a PNS fate expressed high levels of miR-124, which in the latter case it was shown to target and repress non-neuronal genes (e.g. neuronal repressors SCP1 and PTBP1) downstream of Notch signaling [17]. In addition, expression of miR-124 in larval epidermal cells was sufficient for ectopic neural specification, which resembled mis-expression experiments using Pou4, an important transcription factor for sensory neuron specification [17, 38]. Whereas miR-124 targeting to SCP1 is thought to have evolved in the vertebrates+tunicates, miR targeting to PTBP1 may be conserved among bilaterians except for ecdysozoans [17] suggesting that the miRNA regulatory logic in lateral inhibition models of Notch-Delta signaling may have broader implications in other organisms yet to be studied [16]. The research team also showed that miR-124 acted at the gastrula stage and targeted other non-neural genes such as muscle determinant Macho-1 and notochord determinant Brachyury to allow for ectodermal fate specification [17].

3.2.1 Muscle development and the polycistronic miR-1/miR-133 cluster

A well-studied case of miRNA regulation in muscle development is the miR-1/miR-133 polycistronic cluster. Whereas miR-1 promotes differentiation of muscle, miR-133 promotes proliferation of muscle precursors [18]. In the chordates, these two miRNAs are encoded in an antisense direction in a relatively close localization (3-11 kb apart) within the gene *mind bomb 1* (MIB1), and transcribed as a single primary (i.e. polycistronic) transcript. Except for Drosophila and ambulacrarians (i.e. echinoderms and hemichordates), a close proximity of these two miRNAs has been documented in most animal taxa suggesting some form of functional regulatory constraint of a condensed miR-1/miR-133 cluster for the bilaterians [14]. During *C. robusta* development, the polycistronic transcription can be detected in the nuclei of presumptive tail muscle cells from the gastrula stage onward, and its transcription is regulated by an 850 bp sequence upstream of the transcript start site [43]. Differential expression of the two miRNAs in muscle tissues was only detected in the adult, where body wall muscle expressed similar levels of miR-1 and miR-133 and heart muscle expressed significantly higher levels of miR-1 [43].

3.2.2 miRNA expression during oral siphon (OS) regeneration

Three stages of regeneration have been proposed that reconstruct main events of regeneration that match expected expression profiles in the corresponding timeframes [41]. The three phases correspond to: i. wound Healing, ii. transition, and iii. re-development. Using miRNA-mRNA transcriptional profiling using a correlation network, differential expression of mRNAs was correlated to miRNA profiles during the three regeneration windows mentioned above in *C. robusta* oral siphon regeneration [76]. In the first phase, i.e. wound healing, miRNA target clusters of miR 4178b-5p and miR 4_20211 were found to be correlated to the differential expression of genes involved in the following GO term functional classifications: immune response, stress response and apoptosis. In the second phase, i.e. transition, miR 4008c-5p, miR 4123-5p, miR 4178-5p, miR 2_15911, miR 4_20211, and miR 11_7539 were correlated and known to target Wnt, TGFb and MAPK pathway genes that may be regulating the proliferative state characteristic of this particular timeframe. In the third phase, i.e. re-development, miR4008c-5p, miR 10_4533, and miR 11_6940 known to target ECM peptidase inhibitors are correlated with the characteristic extracellular matrix remodeling that occurs at the final phase of regeneration and which resembles the original developmental processes. In contrast other miRs were found expressed throughout the regenerative process. MiRNA miR 10_4533 known to target IGF and IGFb was found expressed presumably regulating the proliferation of progenitors. Also miR-9 was found expressed throughout regeneration and is known to be essential for neural development and function, presumably by targeting and regulating genes involved in cytoskeleton and cell cycle functions [27, 56], instead of targeting Notch or Hes-1 [76].

3.2.3 miRNA expression during *O. dioica* development

A most thorough study of the miRNA repertoire expressed during development has been published for the larvacean *O. dioca* [26]. Using a miRNA array approach with 55 candidate miRNAs and 10 developmental stages for analyses, some general patterns of miRNA occurrence emerged. MicroRNAs were expressed throughout the life cycle of the animal, and were deposited in eggs as maternal determinants for early zygotes. Expression of zygotic miRNAs, such as miR-1487 and miR-1488, was observed starting on the blastula stage (1.5h post fertilization). Most miRNAs analyzed showed developmental regulation (for specific miRNAs that were differentially expressed at each stage see [26]), except for some such as miR-1497 that was expressed throughout all stages [26]. From this study, the first sex specific miRNAs were revealed: miR-1478 was expressed day 6 females in the oocytes, whereas miR-1487/88 were expressed in day 6 males. Interestingly, the compact genomes of *O. dioica* showed one single copy of most miRNA loci, except for miR-1490a, miR-1493, miR-1497d, and miR-1504 that were in two copies [26].

4 Other ncRNAs associated to development

4.1 Yellow Crescent RNA

Yellow crescent RNA, i.e. YC RNA, concerns an about 1.2 kb long polyadenylated RNA, which can be present throughout the embryonic development of ascidians [77]. Its name refers to the fact that in situ hybridization confirmed that YC RNA is localized in the yellow crescent region of one-cell zygotes. The YC transcripts are actually already found in the cortex of unfertilized eggs, segregating with the myoplasm to the yellow crescent after fertilization [77]. Subsequently most YC transcripts enter the primary muscle cell lineage after cleavage and are also present in the secondary muscle cell lineage [77]. YC RNA was first discovered in the club tunicate Styela clava [77]. As the presence of the 1.2-kb RNA in oocytes and early cleaving embryos indicates that it is a maternal transcript, YC RNA is considered to be a maternal RNA [77]. It is associated with the cytoskeleton and segregates to the muscle cells during ascidian embryogenesis. Although the YC ORF encodes for a putative polypeptide of 49 amino acids, this protein is relatively small and does not show any significant homology to any known proteins. As the YC RNA shows various features indicating that it actually functions as an RNA rather than as a protein coding molecule, it is considered to be a noncoding RNA that may play an important role in growth and development [77].

4.2 MicroRNA-offset RNAs

MicroRNA-offset RNAs, i.e. moRNAs, concern about 20 nucleotides long RNAs that lie adjacent to pre-miRNAs. They can originate from both ends of these pre-miRNAs, although prevalently they are derived from the 5' arm [11]. During a study focused on identifying miRNAs in the simple chordate *C. robusta* moRNAs were first discovered [75]. Unexpectedly, half of the *C. robusta* miRNA loci that were detected in this study turned out to encode for previously uncharacterized small RNAs, in addition to conventional miRNA and miRNA* products. This new class of RNAs was hereafter referred to as 'moRNAs', for miRNA-offset RNAs. It became clear that these moRNAs are probably produced by RNAse II-like processing and are observed, like miRNAs, at specific developmental stages [75]. These results and subsequent studies gave rise to the hypothesis that moRNAs concern a new class of functional regulators whose qualitative alteration and/or expression dysregulation might even impact human diseases [11]. Evidence supporting this hypothesis is still fragmentary however. After the discovery in *Ciona*, moRNAs were also found in human cells by deep sequencing analysis. Hereby it was reported that moRNAs from 78 genomic loci were weakly expressed in the prefrontal cortex [44]. Additional indications that moRNA have a distinct function include the fact that some moRNAs are as conserved as miRNAs and are in fact conserved across species to an extent that correlated with expression level [75]. The expression level of certain moRNAs can even be greater than for their corresponding miRNA [82]. Finally, it can be argued [11] that it is likely that moRNAs might represent a functional class of miRNA-related agents as moRNAs are prevalently produced by the 5' arm of the precursor, independent of which arm produces the most expressed mature miRNA [44, 82]. What functions moRNAs may have, varies. For example, moRNA expression was recorded in solid tumours, together with other small RNAs [57]. In addition the fact that an 18-fold enrichment of moRNAs was observed in the nucleus [78] indicates that at least some moRNAs may have functions related to nuclear processes [11]. Recently, an specific class of moRNAs (moRNA-21) has been associated with post-transcriptional gene regulation, proliferation of vascular smooth muscle cells (VSMC) and mediated gene down-regulation in a process mediated by Ago2 [90]. Although these studies do provide good indications, the potential functional roles that moRNAs can play, remain still largely unknown.

4.3 Long Noncoding RNA RMST

Long noncoding RNAs, i.e. lncRNAs, are abundantly found within mammalian transcriptomes. One of the known groups of lncRNAs, includes the rhabdomyosarcoma 2-associated transcript (RMST), which is indispensable for neurogenesis [62]. Human RMST was shown as being responsible for the modulation of neurogenesis as its expression is regulated by the transcriptional repressor REST while it increases during neuronal differentiation [62]. Hereby it was found that RMST is actually

necessary for the binding of SOX2 to promoter regions of neurogenic transcription factors. SOX2, a transcription factor known to regulate neural fate, in combination with RMST were actually found to coregulate a large pool of downstream genes implicated in neurogenesis, i.e. more than 1 000 genes were differentially expressed upon RMST knockdown [62]. These results illustrated the role of RMST as a transcriptional coregulator of SOX2 and a key player in the regulation of neural stem cell fate [62]. A further confirmation of the importance of RMST came with the discovery of a homologue of this lncRNA in the simple chordate *D. vexillum*, i.e. the carpet sea-squirt [85]. While homologues of "human" lncRNAs are rarely found across all chordates due to their low levels of sequence conservation, a plausible homolog of RMST 9, the conserved region 9 of the Rhabdomyosarcoma 2 associated transcript known for its interaction with SOX2, was found in *D. vexillum*. Subsequently putative homologs were also found in the genomes of the ascidians *C. robusta*, *C. savignyi* and *B. schlosseri* and the Florida lancelet *B. floridae*, illustrating that RMST lncRNA are thus conserved across chordates, making them one of the best conserved lncRNAs known to date [85].

4.4 Splices-leader RNA

mRNA 5'leader trans-splicing is a mode of gene expression in which the 5' end of a pre-mRNA is discarded and replaced by the 5' segment of a spliced leader (SL) RNA [84]. Spliced-Leader RNAs, i.e. SL RNAs, hereby consist of a 5' exon and a 3' intron with a conserved consensus 5' splice donor site at the exon-intron boundary [28]. SL RNA trans splicing has not only been described for euglenoids, kinetoplastids, cnidarians, nematodes, and Platyhelminthes [28], but also for deuterostomes like the simple chordate *C. robusta* [84] and the appendicularian *O. dioica* [28]. Hereby *O. dioica* was shown to not only trans-splice SL RNAs to mRNAs, as does *C. robusta*, but also to use trans splicing in resolving polycistronic transcripts [28]. During trans splicing, the capped SL RNA exon moiety is covalently linked to the 5' ends of mRNAs, forming a leader sequence ranging from 16 nt in *C. robusta* to 41 nt in trypanosomatids [28]. The role of SL trans-splicing is still unknown in many cases. SL trans-splicing may potentially having functions varying from the mediation of mRNA stability or translatability [55] and the resolution of polycistronic pre-mRNAs [2, 10], to the production of functional mRNAs from RNA polymerase I transcripts [48].

5 Acknowledgments

This work and the computational analysis were partially supported by the equipment donation from the German Academic Exchange Service-DAAD to the Faculty of Science at the Universidad Nacional de Colombia. Also, by the computational labo-

ratory from Bioinformatics Group at Department of Computer Science and Interdisciplinary Center for Bioinformatic at the Leipzig University. The comparative analysis was partially supported by Colciencias (project no. 110165843196, contract 571-2014). CAVH acknowledges the support by DAAD scholarship: Forschungsstipendien-Promotionen in Deutschland, 2017/18 (Bewerbung 57299294). CIBS acknowledges Universidad Nacional de Colombia for the time granted to write this chapter.

References

1. A. Aziz Aboobaker, Pavel Tomancak, Nipam Patel, Gerald M. Rubin, and Eric C. Lai. Drosophila microrna exhibit diverse spatial expression patterns during embryonic development. *Proceedings of the National Academy of Sciences*, 102(50):18017–18022, 2005.
2. Nina Agabian. Trans splicing of nuclear pre-mrnas. *Cell*, 61(7):1157 – 1160, 1990.
3. V. Ambros, B. Bartel, D. P. Bartel, C. B. Burge, J. C. Carrington, X. Chen, G. Dreyfuss, S. R. Eddy, S. Griffiths-Jones, M. Marshall, M. Matzke, G. Ruvkun, and T. Tuschl. A uniform system for microrna annotation. *RNA*, 9:277–279, 2003.
4. Stefan L. Ameres and Phillip D. Zamore. Diversifying microRNA sequence and function. *Nature Rev. Mol. Cell Biol.*, 14:475–488, 2013.
5. David P. Bartel. MicroRNA target recognition and regulatory functions. *Cell*, 136:215–233, 2009.
6. Arash Bashirullah, Ramona L Cooperstock, and Howard D Lipshitz. RNA localization. *Annu. Rev. Biochem*, 67:335–394, 1998.
7. Isaac Bentwich, Amir Avniel Avniel, Yael Karov, Ranit Aharonov, Shlomit Gilad, Omer Barad, Adi Barzilai, Paz Einat, Uri Einav, Eti Meiri, Eilon Sharon, Yael Spector, and Zvi Bentwich. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.*, 37:766–770, 2005.
8. E Berezikov. Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet.*, 12:846–860, 2011.
9. E Berezikov, F Thuemmler, L W van Laake, I Kondova, R Bontrop, E Cuppen, and R H Plasterk. Diversity of microRNAs in human and chimpanzee brain. *Nat Genet.*, 38:1375–1377, 2006.
10. Thomas Blumenthal. Trans-splicing and polycistronic transcription in caenohabditis elegans. *Trends in Genetics*, 11(4):132 – 136, 1995.
11. Stefania Bortoluzzi, Marta Biasiolo, and Andrea Bisognin. Microrna-offset rnas (mornas): by-product spectators or functional players? *Trends in Molecular Medicine*, 17(9):473–474, 2011.
12. Riccardo Brunetti, Carmela Gissi, Roberta Pennati, Federico Caicci, Fabio Gasparini, and Lucia Manni. Morphological evidence that the molecularly determined ciona intestinalis type a and type b are different species: Ciona robusta and ciona intestinalis. *Journal of Zoological Systematics and Evolutionary Research*, 53(3):186–193, 2015.
13. F Campo-Paysaa, M Sémon, R A Cameron, K J Peterson, and M Schubert. MicroRNA complements in deuterostomes: origin and evolution of microRNAs. *Evol Dev*, 13:15–27, 2011.
14. Florent Campo-Paysaa, Marie Sémon, R. Andrew Cameron, Kevin J. Peterson, and Michael Schubert. microrna complements in deuterostomes: origin and evolution of micrornas. *Evolution & Development*, 13(1):15–27, 2011.
15. H Cerutti and J A Casas-Mollano. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet*, 50:81–99, 2006.
16. Jerry S. Chen, Abygail M. Gumbayan, Robert W. Zeller, and Joseph M. Mahaffy. An Expanded Notch-Delta Model Exhibiting Long-Range Patterning and Incorporating MicroRNA Regulation. *PLoS Computational Biology*, 10(6), 2014.

17. Jerry S. Chen, Matthew San Pedro, and Robert W. Zeller. mir-124 function during ciona intestinalis neuronal development includes extensive interaction with the notch signaling pathway. *Development*, 138(22):4943–4953, 2011.
18. Jian-Fu Chen, Elizabeth M Mandel, J Michael Thomson, Qiulian Wu, Thomas E Callis, Scott M Hammond, Frank L Conlon, and Da-Zhi Wang. The role of microrna-1 and microrna-133 in skeletal muscle proliferation and differentiation. *Nature Genetics*, 38:228 EP –, 12 2005.
19. Li-Chun Cheng, Erika Pastrana, Masoud Tavazoie, and Fiona Doetsch. mir-124 regulates adult neurogenesis in the subventricular zone stem cell niche. *Nature Neuroscience*, 12:399 EP –, 03 2009.
20. Alejandra M. Clark, Leonard D. Goldstein, Maya Tevlin, Simon Tavaré, Shai Shaham, and Eric A. Miska. The microrna mir-124 controls gene expression in the sensory nervous system of caenorhabditis elegans. *Nucleic Acids Research*, 38(11):3780–3793, 2010.
21. Joanne R. Collier, Nicholas A.M. Monk, Philip K. Maini, and Julian H. Lewis. Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling. *Journal of Theoretical Biology*, 183(4):429 – 446, 1996.
22. Z Dai, Z Chen, H Ye, L Zhou, L Cao, Y Wang, S Peng, and L Chen. Characterization of microRNAs in cephalochordates reveals a correlation between microRNA repertoire homology and morphological similarity in chordate evolution. *Evol Dev.*, 11:41–49, 2009.
23. Frederic Delsuc, Herve Philippe, Georgia Tsagkogeorga, Paul Simion, Marie-Ka Tilak, Xavier Turon, Susanna Lopez-Legentil, Jacques Piette, Patrick Lemaire, and Emmanuel J. P. Douzery. A phylogenomic framework and timescale for comparative genomics and evolutionary developmental biology of tunicates. *bioRxiv*, 2017.
24. France Denoeud, Simon Henriet, Sutada Mungpakdee, Jean-Marc Aury, Corinne Da Silva, Henner Brinkmann, Jana Mikhaleva, Lisbeth Charlotte Olsen, Claire Jubin, Cristian Cañestro, Jean-Marie Bouquet, Gemma Danks, Julie Poulain, Coen Campsteijn, Marcin Adamski, Ismael Cross, Fekadu Yadetie, Matthieu Muffato, Alexandra Louis, Stephen Butcher, Georgia Tsagkogeorga, Anke Konrad, Sarabdeep Singh, Marit Flo Jensen, Evelyne Huynh Cong, Helen Eikeseth-Otteraa, Benjamin Noel, Véronique Anthouard, Betina M. Porcel, Rym Kachouri-Lafond, Atsuo Nishino, Matteo Ugolini, Pascal Chourrout, Hiroki Nishida, Rein Aasland, Snehalata Huzurbazar, Eric Westhof, Frédéric Delsuc, Hans Lehrach, Richard Reinhardt, Jean Weissenbach, Scott W. Roy, François Artiguenave, John H. Postlethwait, J. Robert Manak, Eric M. Thompson, Olivier Jaillon, Louis Du Pasquier, Pierre Boudinot, David A. Liberles, Jean-Nicolas Volff, Hervé Philippe, Boris Lenhard, Hugues Roest Crollius, Patrick Wincker, and Daniel Chourrout. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science*, 330(6009):1381–1385, 2010.
25. Marc R. Friedländer, Sebastian D. Mackowiak, Na Li, Wei Chen, and Nikolaus Rajewsky. miRDeep2 accurately identifies known and hundreds of novel microrna genes in seven animal clades. *Nucleic Acids Res.*, 40:37–52, 2012.
26. Xianghui Fu, Marcin Adamski, and Eric M. Thompson. Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*. *Mol. Biol. Evol.*, 25:1067–1080, 2008.
27. Umberto Galderisi, Francesco Paolo Jori, and Antonio Giordano. Cell cycle regulation and neural differentiation. *Oncogene*, 22:5208 EP –, 08 2003.
28. Philippe Ganot, Torben Kallesøe, Richard Reinhardt, Daniel Chourrout, and Eric M. Thompson. Spliced-leader rna trans splicing in a chordate, oikopleura dioica, with a compact genome. *Molecular and Cellular Biology*, 24(17):7795–7805, 2004.
29. Sam Griffiths-Jones, Jerome H L Hui, Antonio Marco, and Matthew Ronshaugen. MicroRNA evolution by arm switching. *EMBO reports*, 12(2):172–177, 2011.
30. A M Heimberg, R Cowper-Sal·lari, M Sémon, P C Donoghue, and Kevin J Peterson. MicroRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. *Proc. Natl. Acad. Sci. USA*, 107:19379–19383, 2010.
31. A. M. Heimberg, L. F. Sempere, V. N. Moy, P. C. J. Donoghue, and K.J. Peterson. MicroRNAs and the advent of vertebrate morphological complexity. *Proc. Natl. Acad. Sci. USA*, 105:2946–2950, 2007.

32. D Hendrix, M Levine, and W Shi. miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. *Genome Biol*, 11(4):R39, 2010.
33. Jana Hertel, Sebastian Bartschat, Axel Wintsche, Christian Otto, The Students of the Bioinformatics Computer Lab 2011, and Peter F. Stadler. Evolution of the let-7 microRNA family. *RNA Biology*, 9:231–241, 2012.
34. Jana Hertel, Manuela Lindemeyer, Kristin Missal, Claudia Fried, Andrea Tanzer, Christoph Flamm, Ivo L. Hofacker, Peter F. Stadler, and The Students of Bioinformatics Computer Labs 2004 and 2005. The expansion of the metazoan microRNA repertoire. *BMC Genomics*, 7:15 [epub], 2006.
35. Jana Hertel and Peter F. Stadler. The expansion of animal microRNA families revisited. *Life*, 5:905–920, 2015.
36. Linda Z. Holland. Genomics, evolution and development of amphioxus and tunicates: The Goldilocks principle, 2015.
37. Tetsuro Ikuta and Hidetoshi Saiga. Organization of hox genes in ascidians: Present, past, and future. *Developmental Dynamics*, 233(2):382–389, 2005.
38. W. Joyce Tang, Jerry S. Chen, and Robert W. Zeller. Transcriptional regulation of the peripheral nervous system in Ciona intestinalis. *Developmental Biology*, 378(2):183–193, 2013.
39. Nathaniel K. Jue, Paola G. Batta-Lona, Sarah Trusiak, Craig Obergfell, Ann Bucklin, Michael J. O'neill, and Rachel J. O'neill. Rapid evolutionary rates and unique genomic signatures discovered in the first reference genome for the southern ocean salp, salpa thompsoni (Urochordata, Thaliacea). *Genome Biology and Evolution*, 8(10):3171–3186, 2016.
40. Raja Keshavan, Michael Virata, Anisha Keshavan, and Robert W. Zeller. Computational identification of ciona intestinalis micrornas. *Zoological Science*, 27(2):162–170, 2010.
41. Dunja Knapp, Herbert Schulz, Cynthia Alexander Rascon, Michael Volkmer, Juliane Scholz, Eugen Nacu, Mu Le, Sergey Novozhilov, Akira Tazaki, Stephanie Protze, Tina Jacob, Norbert Hubner, Bianca Habermann, and Elly M. Tanaka. Comparative transcriptional profiling of the axolotl limb identifies a tripartite regeneration-specific gene program. *PLOS ONE*, 8(5):1–20, 05 2013.
42. Kevin M. Kocot, Michael G. Tassia, Kenneth M. Halanych, and Billie J. Swalla. Phylogenomics offers resolution of major tunicate relationships. *Molecular Phylogenetics and Evolution*, 121:166 – 173, 2018.
43. Rie Kusakabe, Saori Tani, Koki Nishitsuji, Miyuki Shindo, Kohji Okamura, Yuki Miyamoto, Kenta Nakai, Yutaka Suzuki, Takehiro G. Kusakabe, and Kunio Inoue. Characterization of the compact bicistronic microRNA precursor, miR-1/miR-133, expressed specifically in Ciona muscle tissues. *Gene Expression Patterns*, 13(1-2):43–50, 2013.
44. David Langenberger, Clara Bermudez-Santana, Jana Hertel, Steve Hoffmann, Philipp Khaitovich, and Peter F. Stadler. Evidence for human microrna-offset rnas in small rna sequencing data. *Bioinformatics*, 25(18):2298–2301, 2009.
45. David Langenberger, Clara Bermudez-Santana, Peter F. Stadler, and Steve Hoffmann. Identification and classification of small RNAs in transcriptome sequence data. *Pac. Symp. Biocomput.*, 15:80–87, 2010.
46. David Langenberger, M. Volkan Çakir, Steve Hoffmann, and Peter F. Stadler. Dicer-processed small RNAs: Rules and exceptions. *J. Exp. Zool: Mol. Dev. Evol.*, 320:35–46, 2012.
47. C T Lee, T Risom, and W M Strauss. Evolutionary conservation of microRNA regulatory circuits: an examination of microrna gene complexity and conserved microRNA-target interactions through metazoan phylogeny. *DNA Cell Biol.*, 26:209–218, 2007.
48. Mary Gwo-Shu Lee and Lex H. T. Van der Ploeg. Transcription of protein-coding genes in trypanosomes by rna polymerase i. *Annual Review of Microbiology*, 51(1):463–489, 1997. PMID: 9343357.
49. Matthieu Legendre, André Lambert, and Daniel Gautheret. Profile-based detection of microRNA precursors in animal genomes. *Bioinformatics*, 21(7):841–845, 2005.
50. H Liang and W-H. Li. Lowly expressed human microrna genes evolve rapidly. *Mol Biol Evol*, 26:1195–1198, 2009.

51. Shi-Lung Lin, Joseph D. Miller, and Shao-Yao Ying Ying. Intronic MicroRNA (mirna). *J Biomed Biotechnol.*, 2006:26818, 2006.
52. J Lu, Y Shen, Q Wu, S Kumar, B He, S Shi, R W Carthew, S M Wang, and C I Wu. The birth and death of microRNA genes in *Drosophila*. *Nat Genet.*, 40:351–355, 2008.
53. E Lund, S Güttinger, A Calado, J E Dahlberg, and U Kutay. Nuclear export of microRNA precursors. *Science*, 303:95–98, 2004.
54. A. Marco, M. Ninova, M. Ronshaugen, and S. Griffiths-Jones. Clusters of microRNAs emerge by new hairpins in existing transcripts. *Nucleic Acids Res*, 41:7745–7752, 2013.
55. P A Maroney, J A Denker, E Darzynkiewicz, R Laneve, and T W Nilsen. Most mrnas in the nematode ascaris lumbricoides are trans-spliced: a role for spliced leader addition in translational efficiency. *RNA*, 1(7):714–723, 1995.
56. Rowena McBeath, Dana M Pirone, Celeste M Nelson, Kiran Bhadriraju, and Christopher S Chen. Cell shape, cytoskeletal tension, and rhoa regulate stem cell lineage commitment. *Developmental Cell*, 6(4):483 – 495, 2004.
57. Eti Meiri, Asaf Levy, Hila Benjamin, Miriam Ben-David, Lahav Cohen, Avital Dov, Nir Dromi, Eran Elyakim, Noga Yerushalmi, Orit Zion, Gila Lithwick-Yanai, and Einat Sitbon. Discovery of micrornas and other small rnas in solid tumors. *Nucleic Acids Research*, 38(18):6234–6246, 2010.
58. J Meunier, F Lemoine, M Soumillon, A Liechti, M Weier, K Guschanski, H Hu, P Khaitovich, and H Kaessmann. Birth and expression evolution of mammalian microrna genes. *Genome Res*, 23:34–45, 2012.
59. Kristin Missal, Dominic Rose, and Peter F. Stadler. Non-coding RNAs in Ciona intestinalis. *Bioinformatics*, 21(SUPPL. 2):77–78, 2005.
60. Eric P Nawrocki and Sean R Eddy. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29:2933–2935, 2013.
61. Eric P. Nawrocki and Sean R. Eddy. Infernal 1.1: 100-fold faster rna homology searches. *Bioinformatics*, 29(22):2933–2935, 2013.
62. Shi-Yan Ng, Gireesh K. Bogu, Boon Seng Soh, and Lawrence W. Stanton. The long noncoding rna rmst interacts with sox2 to regulate neurogenesis. *Molecular Cell*, 51(3):349–359, 2013.
63. R Niwa and F J. Slack. The evolution of animal microRNA function. *Curr Opin Genet Dev.*, 17:145–150, 2007.
64. Trina M Norden-Krichmar, Janette Holtz, Amy E Pasquinelli, and Terry Gaasterland. Computational prediction and experimental validation of Ciona intestinalis microRNA genes. *BMC genomics*, 8(1):445, 2007.
65. A. E. Pasquinelli, A. McCoy, E. Jimenez, E. Salo, G. Ruvkun, M. Q. Martindale, and J. Baguna. Expression of the 22 nucleotide let-7 heterochronic RNA throughout the Metazoa: a role in life history evolution? *Evol. Dev.*, 5(4):372–378, 2003.
66. a E Pasquinelli, B J Reinhart, F Slack, M Q Martindale, M I Kuroda, B Maller, D C Hayward, E E Ball, B Degnan, P Müller, J Spring, a Srinivasan, M Fishman, J Finnerty, J Corbo, M Levine, P Leahy, E Davidson, and G Ruvkun. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808):86–89, 2000.
67. Kevin J. Peterson, Michael R. Dietrich, and Mark A. McPeek. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *Bioessays*, 31:736–747, 2009.
68. N Price, R A Cartwright, N Sabath, D Graur, and R B Azevedo. Neutral evolution of robustness in drosophila microRNA precursors. *Mol Biol Evol*, 28:2115–2123, 2011.
69. S E Prochnik, D S Rokhsar, and A A Aboobaker. Evidence for a microRNA expansion in the bilaterian ancestor. *Dev Genes Evol.*, 217:73–77, 2007.
70. Brenda J. Reinhart, Frank J. Slack, Michael Basson, Amy E. Pasquinelli, Jill C. Bettinger, Ann E. Rougvie, H. Robert Horvitz, and Gary Ruvkun. The 21-nucleotide let-7 rna regulates developmental timing in caenorhabditis elegans. *Nature*, 403:901 EP –, 02 2000.
71. Yutaka Satou, Katsuhiko Mineta, Michio Ogasawara, Yasunori Sasakura, Eiichi Shoguchi, Keisuke Ueno, Lixy Yamada, Jun Matsumoto, Jessica Wasserscheid, Ken Dewar, Graham B. Wiley, Simone L. Macmil, Bruce A. Roe, Robert W. Zeller, Kenneth EM Hastings, Patrick

Lemaire, Erika Lindquist, Toshinori Endo, Kohji Hotta, and Kazuo Inaba. Improved genome assembly and evidence-based global gene model set for the chordate ciona intestinalis: new insight into intron and operon populations. *Genome Biology*, 9(10):R152, Oct 2008.

72. L F Sempere, C N Cole, M A McPeek, and K J Peterson. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J Exp Zoolog B Mol Dev Evol.*, 306B:575–588, 2006.
73. Lorenzo F. Sempere, Sarah Freemantle, Ian Pitha-Rowe, Eric Moss, Ethan Dmitrovsky, and Victor Ambros. Expression profiling of mammalian micrornas uncovers a subset of brain-expressed micrornas with possible roles in murine and human neuronal differentiation. *Genome Biology*, 5(3):R13, Feb 2004.
74. S A Shabalina and E V Koonin. Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol.*, 23:578–587, 2008.
75. Weiyang Shi, David Hendrix, Mike Levine, and Benjamin Haley. A distinct class of small RNAs arises from pre-miRNA–proximal regions in a simple chordate. *Nature Structural & Molecular Biology*, 16(2):183–189, 2009.
76. Elijah J. Spina, Elmer Guzman, Hongjun Zhou, Kenneth S. Kosik, and William C. Smith. A microrna-mrna expression network during oral siphon regeneration in ciona. *Development*, 144(10):1787–1797, 2017.
77. B J Swalla and W R Jeffery. A maternal RNA localized in the yellow crescent is segregated to the larval muscle cells during ascidian development., 1995.
78. Ryan J Taft, Ken C Pang, Timothy R Mercer, Marcel Dinger, and John S Mattick. Non-coding rnas: regulators of disease. *The Journal of Pathology*, 220(2):126–139, 2010.
79. Andrea Tanzer, Markus Riester, Jana Hertel, Clara Isabel Bermudez-Santana, Jan Gorodkin, Ivo L. Hofacker, and Peter F. Stadler. Evolutionary genomics of microRNAs and their relatives. In Gustavo Caetano-Anolles, editor, *Evolutionary Genomics and Systems Biology*, pages 295–327. Wiley-Blackwell, Hoboken, NJ, 2010.
80. Andrea Tanzer and Peter F. Stadler. Molecular evolution of a microRNA cluster. *J. Mol. Biol.*, 339:327–335, 2004.
81. Goro Terai, Hiroaki Okida, Kiyoshi Asai, and Toutai Mituyama. Prediction of Conserved Precursors of miRNAs and Their Mature Forms by Integrating Position-Specific Structural Features. *PLoS ONE*, 7(9):1–11, 2012.
82. Jennifer L. Umbach and Bryan R. Cullen. In-depth analysis of kaposi's sarcoma-associated herpesvirus microrna expression provides insights into the mammalian microrna-processing machinery. *Journal of Virology*, 84(2):695–703, 2010.
83. Jennifer L. Umbach, Lisa I. Strelow, Scott W. Wong, and Bryan R. Cullen. Analysis of rhesus rhadinovirus micrornas expressed in virus-induced tumors from infected rhesus macaques. *Virology*, 405(2):592 – 599, 2010.
84. Amanda E. Vandenberghe, Thomas H. Meedel, and Kenneth E.M. Hastings. mrna 5-leader trans-splicing in the chordates. *Genes & Development*, 15(3):294–303, 2001.
85. C. A. Velandia-Huerto, A. A. Gittenberger, F. D. Brown, P. F. Stadler, and C. I. Bermudez-Santana. Automated detection of ncrnas in the draft genome sequence of a colonial tunicate: the carpet sea squirt didemnum vexillum. *BMC Genomics*, 17:691, Aug 2016.
86. Kai Wang, Christelle Dantec, Patrick Lemaire, Takeshi A. Onuma, and Hiroki Nishida. Genome-wide survey of miRNAs and their evolutionary history in the ascidian, Halocynthia roretzi. *BMC Genomics*, 18(1):314, 2017.
87. B M Wheeler, A M Heimberg, V N Moy, E A Sperling, T W Holstein, S Heber, and K J Peterson. The deep evolution of metazoan microRNAs. *Evol. Dev.*, 11:50–68, 2009.
88. Sebastian Will, Kristin Reiche, Ivo L. Hofacker, Peter F. Stadler, and Rolf Backofen. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Computational Biology*, 3(4):680–691, 2007.
89. Kasumi Yagi, Yutaka Satou, Françoise Mazet, Sebastian M. Shimeld, Bernard Degnan, Daniel Rokhsar, Michael Levine, Yuji Kohara, and Nori Satoh. A genomewide survey of developmentally relevant genes in ciona intestinalis. *Development Genes and Evolution*, 213(5):235–244, Jun 2003.

90. Jin Zhao, Gavin R. Schnitzler, Lakshmanan K. Iyer, Mark J. Aronovitz, Wendy E. Baur, and Richard H. Karas. Microrna-offset rna alters gene expression and cell proliferation. *PLOS ONE*, 11(6):1–16, 06 2016.
91. Yuhai Zhao, Lin Cong, and Walter J. Lukiw. Plant and animal micrornas (mirnas) and their potential for inter-kingdom communication. *Cellular and Molecular Neurobiology*, 38(1):133–140, Jan 2018.