

# Contents

<b>non-protein-coding RNAs as regulators of development in tunicates . . . . .</b>	<b>3</b>
Author A and Author B	
1 Introduction . . . . .	3
2 miRNA families origin and evolutionary perspective . . . . .	4
2.1 Origins and Evolution of MicroRNAs . . . . .	4
2.2 miRNA identification and validation . . . . .	5
2.3 miRNA in clusters . . . . .	9
2.4 miRNA in an perspective evolution . . . . .	11
2.5 To complete the tree of loss and gain of families . . . . .	14
3 miRNAs and its rol in development . . . . .	16
3.1 miRNAs discovery and development . . . . .	16
3.2 miRNAs perspective evolution in development . . . . .	16
3.3 Specific examples . . . . .	18
4 Other ncRNAs associated to development . . . . .	22
4.1 Yellow Crescent RNA . . . . .	22
4.2 MicroRNA-offset RNAs . . . . .	22
4.3 Long Noncoding RNA RMST . . . . .	23
4.4 Splices-leader RNA . . . . .	23
References . . . . .	24



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

Author A and Author B

**Abstract** Each chapter should be preceded by an abstract (10–15 lines long) that summarizes the content. The abstract will appear *online* at [www.SpringerLink.com](http://www.SpringerLink.com) and be available with unrestricted access. This allows unregistered users to read the abstract as a teaser for the complete chapter. As a general rule the abstracts will not appear in the printed version of your book unless it is the style of your particular book or that of the series to which your book belongs.

Please use the 'starred' version of the new Springer `abstract` command for typesetting the text of the online abstracts (cf. source file of this chapter template `abstract`) and include them with the source files of your manuscript. Use the plain `abstract` command if the abstract is also to appear in the printed version of the book.

## 1 Introduction

noncoding RNAs roles in tunicates development date earliest in the 90's from the works of Swalla & Jeffry in which RNAs localized in the yellow crescent or myoplasm, a cytoskeletal domain in oocytes of the ascidian *Styela clava* were discovered [51]. This yellow crescent or YC RNA identified to be present throughout embryonic development was the first example involved in envisioning the future of a growing family of ncRNAs that would play important roles in growth and development in tunicates[51].

---

Author Name  
Name, Address of Institute, e-mail: [writeemail@unal.edu.co](mailto:writeemail@unal.edu.co)

Author Name  
Name, Address of Institute e-mail: [writeemail@com](mailto:writeemail@com)

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* [5]

## 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 [?]. Mechanistically, miRNAs depends on the presence of the evolutionarily even older RNA interference pathways [?, ?] 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 [36]. 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 [37] 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 [4]. 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. [31, 12, 32].

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 simultaneously 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 [?], 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* [41] 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 [20].

Like other gene families, miRNAs form paralogs [?, ?] and hence often appear as families as homologous genes. This forms the basis of the miRBase nomenclature [2]. 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 [47, 18, 17, 58], although the massive restructuring of the miRNA complement of tunicates is an important exception to this rule [?].

The innovation of new miRNA families is an on-going process. Experimental surveys of the miRNA repertoire thus have reported a large number of very young and even species-specific miRNAs [?, ?]. 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 [?]. This is consistent with the fact that stable hairpins are abundant structural elements in random RNAs, which makes it not only possible but actually quite likely that miRNA precursors appear by chance in transcribed genomic regions [?, ?, ?]. 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 [?, ?], 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 [?]. 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 expanding throughout animal evolution in a manner that at least roughly correlates with morphological complexity [20, 47, ?, ?, ?, 18, ?, ?]. Several bursts of miRNA innovation have been observed [20, 18, ?, 23], 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 [?, ?].

## 2.2 miRNA identification and validation

The first microRNA (miRNA) in tunicates was discovered in the year 2000 from the work of *Pasquinelli et al., 2000* (CITE) when they were studying the temporal regulation of *let-7* during development by using samples of small RNAs of a wide range of animal species, in which the ascidian *Ciona intestinalis* was included as well as other vertebrates, hemichordates, mollusc, annelids, arthropod and other bilateral and nonbilateria animals [45]. Later on the year 2003 the same team suggested that *let-7* RNA may control the late temporal transitions during development across animal phylogeny [44] albeit it was not identified on basal metazoans such as cnidarians and poriferans.

Then after the era of genome sequencing became available, it was launched in 2005 the computational screening of whole-genomes of non-model organism as tunicates. Beginning with the *Cionas C. intestinalis* and *C. savigni* a profile-based

strategy was implemented in the ERPIN program [35]. On that work were detected a set of new miRNAs candidates considered as *C. intestinalis* specific such as the members of the family miR-9 and miR-79 and as it was expected, other miRNA families were found homologous between both Cionas like the families miR-124;92;98;325;310-313 and let-7. Coincidentally, by the same year a whole-genomic comparative approach in the urochordate lineage was performed on the species *C. intestinalis*, *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. intestinalis* [40]. After all, the same group in 2007 implemented a structure-based clustering approach in *C. intestinalis* predicted 58 miRNAs, of which only let-7, miR-7, miR-124, and miR-126 coincided with the previously annotated miRNAs [59].

Thus far, the primary focus to identify miRNAs into urochordate lineage has been mainly toward the use of computational approaches but soon came up the use of new hybrid strategies combining computational and experimental studies to validate candidate families previously detected. For instance the first bona fide record for *C. intestinalis* was registered in miRBase only in its Release 11. Those first miRNAs records were derived from the work published in 2007 by Norden-Krichmar et al., [43]. The authors searched for conservation with the seed region of the known mature miRNA sequences from miRBase release 2006 on the whole-genomic sequences of *C. intestinalis* and *C. savignyi*. Those miRNAs were aligned locally using the FASTA/ssearch34 program. Only matches of 90% identity or better were retained. In further steps these authors studied RNA sequences that folds like hairpin structures with the mature miRNA sequence in the stem region including other typical features exhibit in miRNA hairpins. By manual curation of the genomic sequences predicted by the software mfold which folded like hairpin structures, a set of 18 miRNA molecules were detected which appeared conserved in both Cionas. After all, using Northern blot analyses in the adult tissue of *C. intestinalis* the authors confirmed expression of let-7, miR-7, and miR-126, as well as 11 other conserved miRNA families.

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. In this order of ideas, the first repertory of miRNAs based on non-Cionas species was published by Fu et al., in 2008 for the larvacean *O. dioica* [14]. At that time the authors were studying the temporal-spatial expression patterns of conserved miRNAs in different developmental stages of oocytes, 1-cell zygote, 2-8 cell embryos, blastulas, gastrulas, tadpoles (in different stages) and animals from 1 to 6 days from *O. dioica*. In this research, 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 only for 55 miRNAs was detected expression. As a conclusion the authors suggested that those 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. Besides the authors identified a set of sex-specific miRNAs that appeared as male/female gonad differentiation which became apparent and was maintained throughout spermatogenesis [14]. Unexpectedly, the majority of the miRNAs loci in *O. dioica* were located in antisense orientations into the hosted genes in opposite fashion observed in the majority of the known mammalian miRNAs at that 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. intestinalis* 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. intestinalis* 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. dioica* and the colonial tunicate *Botryllus schlosseri* [56]. 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 [26] and in 2017 the prediction of miRNAs families were reported to the species *Halocynthia roretzi*. On the following two sections we will focus on those stages of the fascinated increased screening of the miRNA repertory in tunicates.

### 2.2.1 Validation and detection of miRNA families in *Cionas* in this decade

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 after preparing small RNA libraries from various developmental stages including unfertilized eggs, early embryos, late embryos and adults from *C. intestinalis* was performed high-throughput sequencing of cDNA with an Illumina 1G Genome Analyzer experiments. These sequencing led to document 80 miRNAs families for *C. intestinalis*. Unexpectedly, were detected a distinct species of small RNAs processed outside of the miRNA precursors which were termed as moRs or miRNA-offset RNAs [48]. Later on, after extracting non-coding conserved regions of whole genome alignments between *C. intestinalis* 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 [27]. Then in order to validate those candidate, RT-PCR and PAGE were conducted to design a custom microarray. After screened them for miRNA expression were identifying that 244 of the 458 miRNA predictions were represented either in their microarray data or in the Illumina database constructed previously for small RNA derived from *C. intestinalis* by [48]. Although they failed to predict 39 previously characterized miRNAs, it was suggested in this work that *C. intestinalis* genome may encode about 300 miRNA genes. Then to increase the miRNAs collection in *C. intestinalis* a novel computational strategy for the systematic, whole-genome identification of microRNA from high throughput sequencing information was developed in 2010 by [19]. That method, known as miRTRAP, incorporated not only the 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. In short words these strategy relies on the way how the the biochemical machinery processes pre-miRNA hairpins produces short RNA products. This approach is highly depended on the deep of the small RNAs mapped to a given locus and is highlighted by the authors that the approach requires an accurate assignment of small RNA sequences on their relative positions along the hairpin, that is, miR/miR\*, moR/moR\* and loop [19]. Again a new approach took advantage of importance to detect miRNAs from the high-throughput sequencing of small RNAs available from [48]. These approach known as miRDeep2 improved the algorithm of its first version miRDeep [13] and let to identify with an accuracy of 98.6% and 99.9% canonical and non-canonical miRNAs in different species. These approach reported 313 known and 127 novel ones miRNAs in *C. intestinalis*. In the same year the program miRRim2 [53] was applied to the *C. intestinalis* genome, in which some candidates identified from the work of [19] and the several promising candidates were detected. In 2013, [29] was investigated the expression patterns of the cluster miR-1 and miR-133 in *C. intestinalis* 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 The new era to get deep insights into the repertory of miRNA 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. Annotated small ncRNAs sequences from metazoan species and profile markov models from RFAM



were applied with blast and hmmer correspondingly and the best candidates were validated by structural alignments against metazoon-specific covariance models.

Please summary of our Dvexillum paper [56] including the first reported preliminary annotation for colonial tunicate *B. schlosseri* beside the one for Dvexillum. For the preliminary assembled of the genome sequence for the Southern Ocean salp *S. thompsoni* [26] 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 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* [48] (previously known as *C. intestinalis* today reclassified) was used to identify conserved miRNAs in *H. roretzi* [57] .

## 2.3 miRNA in clusters

### 2.3.1 Yanetal2014.pdf

[60] *mir181 is an ancient microRNA (miRNA) gene family that originated in urochordata. Although their functions were subjected to extensive studies in recent years, their evolutionary process remains largely unknown. Here we systematically investigated the homologous genes of the mir-181 family by a sequence similarity search. Representative sequences of the mir-181 gene family were used to reconstruct their evolutionary history. Our results indicated that this family could have derived from multiple duplications, which include two rounds of whole genome duplications and one round of segmental replication. Functional annotation of the target genes of the mir-181 family suggested that this family could participate in some important biological processes including transcriptional and translational regulation, signaling transduction etc. This analysis presented a complex evolutionary dynamics for the origination of a miRNA gene family.*

### 2.3.2 Wangetal2017.pdf

citeWang2017 *Background: miRNAs play essential roles in the modulation of cellular functions via degradation and/or translation attenuation of target mRNAs. They have been surveyed in a single ascidian genus, Ciona. Recently, an annotated draft genome sequence for a distantly related ascidian, Halocynthia roretzi, has become*

available, but miRNAs in *H. roretzi* have not been previously studied. *Results:* We report the prediction of 319 candidate *H. roretzi* miRNAs, obtained through three complementary methods. Experimental validation suggests that more than half of these candidate miRNAs are expressed during embryogenesis. The majority of predicted *H. roretzi* miRNAs appear specific to ascidians or tunicates, and only 32 candidates, belonging to 25 families, are widely conserved across metazoans. *Conclusion:* Our study presents a comprehensive identification of candidate *H. roretzi* miRNAs. This resource will facilitate the study of the mechanisms for miRNA-controlled gene regulatory networks during ascidian development. Further, our analysis suggests that the majority of *Halocynthia* miRNAs are specific to ascidian or tunicates, with only a small number of widely conserved miRNAs. This result is consistent with the general notion that animal miRNAs are less conserved between taxa than plant ones.

### 2.3.3 deSouzaGomezetal2012.pdf

[11] *MicroRNAs (miRNAs) are small noncoding RNA molecules which are processed into ~20-24nt molecules that can regulate the gene expression post-transcriptionally. MiRNA gene clusters have been identified in a range of species, where in miRNAs are often processed from polycistronic transcripts. In this study, a computational approach is used to investigate the extent of evolutionary conservation of the miR-71/2 cluster in animals, and to identify novel miRNAs in the miRNA cluster miR-71/2. The miR-71/2 cluster, consisting of copies of the miR-71 and miR-2 (including miR-13) families, was found to be Protostome-specific. Although, this cluster is highly conserved across the Protostomia, the miR-2 family is completely absent from the Deuterostomia species, while miR-71 is absent from the Vertebrata and Urochordata. The evolutionary conservation and clustering propensity of the miR-71/2 family across the Protostomes could indicate the common functional roles across the member species of the Protostomia.,*

### 2.3.4 Stadler2007.PDF

[59] *The RFAM database defines families of ncRNAs by means of sequence similarities that are sufficient to establish homology. In some cases, such as microRNAs and box H/ACA snoRNAs, functional commonalities define classes of RNAs that are characterized by structural similarities, and typically consist of multiple RNA families. Recent advances in high-throughput transcriptomics and comparative genomics have produced very large sets of putative noncoding RNAs and regulatory RNA signals. For many of them, evidence for stabilizing selection acting on their secondary structures has been derived, and at least approximate models of their structures have been computed. The overwhelming majority of these hypothetical RNAs cannot be assigned to established families or classes. We present here a structure-based clustering approach that is capable of extracting putative RNA classes from genome-wide surveys for structured RNAs. The LocARNA (local align-*

ment of RNA) tool implements a novel variant of the Sankoff algorithm that is sufficiently fast to deal with several thousand candidate sequences. The method is also robust against false positive predictions, i.e., a contamination of the input data with unstructured or nonconserved sequences. We have successfully tested the LocARNA-based clustering approach on the sequences of the RFAM-seed alignments. Furthermore, we have applied it to a previously published set of 3,332 predicted structured elements in the *Ciona intestinalis* genome (Missal K, Rose D, Stadler PF (2005) Noncoding RNAs in *Ciona intestinalis*. *Bioinformatics* 21 (Supplement 2): i77-i78). In addition to recovering, e.g., tRNAs as a structure-based class, the method identifies several RNA families, including microRNA and snoRNA candidates, and suggests several novel classes of ncRNAs for which to date no representative has been experimentally characterized.

## 2.4 miRNA in an perspective evolution

### 2.4.1 Introduction

### 2.4.2 Pignatelli2016.pdf

[46] Annotation of orthologous and paralogous genes is necessary for many aspects of evolutionary analysis. Methods to infer these homology relationships have traditionally focused on protein-coding genes and evolutionary models used by these methods normally assume the positions in the protein evolve independently. However, as our appreciation for the roles of non-coding RNA genes has increased, consistently annotated sets of orthologous and paralogous ncRNA genes are increasingly needed. At the same time, methods such as PHASE or RAXML have implemented substitution models that consider pairs of sites to enable proper modelling of the loops and other features of RNA secondary structure. Here, we present a comprehensive analysis pipeline for the automatic detection of orthologues and paralogues for ncRNA genes. We focus on gene families represented in Rfam and for which a specific covariance model is provided. For each family ncRNA genes found in all Ensembl species are aligned using Infernal, and several trees are built using different substitution models. In parallel, a genomic alignment that includes the ncRNA genes and their flanking sequence regions is built with PRANK. This alignment is used to create two additional phylogenetic trees using the neighbour-joining (NJ) and maximum-likelihood (ML) methods. The trees arising from both the ncRNA and genomic alignments are merged using TreeBeST, which reconciles them with the species tree in order to identify speciation and duplication events. The final tree is used to infer the orthologues and paralogues following Fitch's definition. We also determine gene gain and loss events for each family using CAFE. All data are accessible through the Ensembl Comparative Genomics ('Compara') API, on our FTP site and are fully integrated in the Ensembl genome browser, where they can be accessed in a user-friendly manner. Database URL: <http://www.ensembl.org>.

### 2.4.3 Dai2009.pdf

[10] *Cephalochordates, urochordates, and vertebrates comprise the three extant groups of chordates. Although higher morphological and developmental similarity exists between cephalochordates and vertebrates, molecular phylogeny studies have instead suggested that the morphologically simplified urochordates are the closest relatives to vertebrates. MicroRNAs (miRNAs) are regarded as the major factors driving the increase of morphological complexity in early vertebrate evolution, and are extensively characterized in vertebrates and in a few species of urochordates. However, the comprehensive set of miRNAs in the basal chordates, namely the cephalochordates, remains undetermined. Through extensive sequencing of a small RNA library and genomic homology searches, we characterized 100 miRNAs from the cephalochordate amphioxus, Branchiostoma japonicum, and B. floridae. Analysis of the evolutionary history of the cephalochordate miRNAs showed that cephalochordates possess 54 miRNA families homologous to those of vertebrates, which is threefold higher than those shared between urochordates and vertebrates. The miRNA contents demonstrated a clear correlation between the extent of miRNA overlapping and morphological similarity among the three chordate groups, providing a strong evidence of miRNAs being the major genetic factors driving morphological complexity in early chordate evolution.*

### 2.4.4 Candini2012.pdf

[8] *MicroRNAs (miRNAs) are small non-coding RNAs that negatively regulate gene expression and thus control diverse biological processes. The high interest in miRNAs as an important mediator of post-transcriptional gene regulation has led to the discovery of miRNAs in several organisms. The present article outlines and discusses the current status of miRNAs information on the basal chordate amphioxus and the evolution of miRNAs in metazoans.*

### 2.4.5 Stadler2006.pdf

[22] *UNLABELLED: Recently, genome-wide surveys for non-coding RNAs have provided evidence for tens of thousands of previously undescribed evolutionary conserved RNAs with distinctive secondary structures. The annotation of these putative ncRNAs, however, remains a difficult problem. Here we describe an SVM-based approach that, in conjunction with a non-stringent filter for consensus secondary structures, is capable of efficiently recognizing microRNA precursors in multiple sequence alignments. The software was applied to recent genome-wide RNAz surveys of mammals, urochordates, and nematodes. AVAILABILITY: The program RNAmicro is available as source code and can be downloaded from <http://www.bioinf.uni-leipzig/Software/RNAmicro>.*

#### 2.4.6 SamGriffit2007.pdf

[16] *MicroRNAs (miRNAs) modulate transcript stability and translation. Functional mature miRNAs are processed from one or both arms of the hairpin precursor. The miR-100/10 family has undergone three independent evolutionary events that have switched the arm from which the functional miRNA is processed. The dominant miR-10 sequences in the insects Drosophila melanogaster and Tribolium castaneum are processed from opposite arms. However, the duplex produced by Dicer cleavage has an identical sequence in fly and beetle. Expression of the Tribolium miR-10 sequence in Drosophila S2 cells recapitulates the native beetle pattern. Thus, arm usage is encoded in the primary miRNA sequence, but outside the mature miRNA duplex. We show that the predicted messenger RNA targets and inferred function of sequences from opposite arms differ significantly. Arm switching is likely to be general, and provides a fundamental mechanism to evolve the function of a miRNA locus and target gene network.*, author = Griffiths-Jones, Sam and Hui, Jerome H L and Marco, Antonio and Ronshaugen, Matthew

#### 2.4.7 Hertel2015.pdf

[21] *MicroRNAs are important regulatory small RNAs in many eukaryotes. Due to their small size and simple structure, they are readily innovated de novo. Throughout the evolution of animals, the emergence of novel microRNA families traces key morphological innovations. Here, we use a computational approach based on homology search and parsimony-based presence/absence analysis to draw a comprehensive picture of microRNA evolution in 159 animal species. We confirm previous observations regarding bursts of innovations accompanying the three rounds of genome duplications in vertebrate evolution and in the early evolution of placental mammals. With a much better resolution for the invertebrate lineage compared to large-scale studies, we observe additional bursts of innovation, e.g., in Rhabditoidea. More importantly, we see clear evidence that loss of microRNA families is not an uncommon phenomenon. The Enoplea may serve as a second dramatic example beyond the tunicates. The large-scale analysis presented here also highlights several generic technical issues in the analysis of very large gene families that will require further research.*

#### 2.4.8 Yanetal2014.pdf

[60] *Mir-181 is an ancient microRNA (miRNA) gene family that originated in urochordata. Although their functions were subjected to extensive studies in recent years, their evolutionary process remains largely unknown. Here we systematically investigated the homologous genes of the mir-181 family by a sequence similarity search. Representative sequences of the mir-181 gene family were used to reconstruct their evolutionary history. Our results indicated that this family could have*

derived from multiple duplications, which include two rounds of whole genome duplications and one round of segmental replication. Functional annotation of the target genes of the mir-181 family suggested that this family could participate in some important biological processes including transcriptional and translational regulation, signaling transduction etc. This analysis presented a complex evolutionary dynamics for the origination of a miRNA gene family. © 2014 Elsevier Ltd.

## 2.5 To complete the tree of loss and gain of families

Our miRNA families updated with the new two annotated miRNAs Salpa and Halocynthia

[56] A comparative analysis of miRNAs families in tunicates is summarized in Fig. 5 c and d. We find several patterns of recognizable trends as we analyze the conservation, loss and gain of miRNA families in tunicates compared to selected bilaterians. For instance, based in our analysis of conserved miRNA families across the bilaterians, we estimate that ancestor of the Bilateria likely contained approximately 33 miRNA families. In contrast, the ancestor of the Chordata presumably contained members of 37 miRNA families. In the cephalochordate *B. floridae*, we find the occurrence of a unique repertoire of 82 miRNA families that evolved as a consequence of many gains (+50), but also some losses (5) [8]. In the ancestor of *Olfactores* we presume the presence of 72 miRNA families, of which tunicates did not show losses, with the notable exception of *O. dioica* that has undergone substantial losses (60). Within the ascidians, only *B. schlosseri* and *C. savignyi* show dramatic losses of miRNA families, 32 and 17 respectively. *D. vexillum* has lost 8 miRNA families. *D. vexillum* shows 16 families that are absent in other tunicates i.e. mir-430, mir-9, mir-130, mir-190, mir-139, mir-460, mir-315, mir-305, mir-458, mir-185, mir-233, mir-569, mir-944, mir-567, mir-2985 and mir-4068 (Fig. 5d). Comparisons of the miRNAs repertoires between the colonial tunicates (*D. vexillum* and *B. schlosseri*) on the one hand, and solitary tunicates (*O. dioica*, *C. intestinalis* and *C. savignyi*) on the other hand, show 10 microRNA families, i.e. mir-133, mir-186, mir-6, mir-279, mir-340, mir-11, mir-60, mir-592, mir-883 and mir-549 that are specific to colonial tunicates, while only 2 families, i.e. mir-31 and mir-1473, are specific to solitary tunicates. It is worthwhile exploring how colonial specific miRNAs may have been co-opted in ascidians to function in somatic stem cell function, regeneration, budding, or other asexual developmental processes, as miRNAs are known to be important players in stem cell function, and developmental processes of differentiation in vertebrates

### 2.5.1 Jueetal2016.pdf

[26] A preliminary genome sequence has been assembled for the Southern Ocean salp, *Salpa thompsoni* (Urochordata, Thaliacea). Despite the ecological importance

of this species in Antarctic pelagic food webs and its potential role as an indicator of changing Southern Ocean ecosystems in response to climate change, no genomic resources are available for *S. thompsoni* or any closely-related urochordate species. Using a multiple-platform, multiple-individual approach, we have produced a 318,767,936 bp genome sequence, covering more than 50% of the estimated 602 MB (173 MB) genome size for *S. thompsoni*. Using a non-redundant set of predicted proteins, more than 50% (16,823) of sequences showed significant homology to known proteins and ~38% (12,151) of the total protein predictions were associated with Gene Ontology functional information. We have generated 109,958 SNP variant and 9,782 indel predictions for this species, serving as a resource for future phylogenomic and population genetic studies. Comparing the salp genome to available assemblies for four other urochordates, *Botryllus schlosseri*, *Ciona intestinalis*, *Ciona savignyi* and *Oikopleura dioica*, we found that *S. thompsoni* shares the previously-estimated rapid rates of evolution for these species. High mutation rates are thus independent of genome size, suggesting that rates of evolution >1.5 times that observed for vertebrates are a broad taxonomic characteristic of urochordates. Tests for positive selection implemented in PAML revealed a small number of genes with sites undergoing rapid evolution, including genes involved in ribosome biogenesis and metabolic and immune process that may be reflective of both adaptation to polar, planktonic environments as well as the complex life history of the salps. Finally, we performed an initial survey of small RNAs, revealing the presence of known, conserved miRNAs, as well as novel miRNA genes; unique piRNAs; and mature miRNA signatures for varying developmental stages. Collectively, these resources provide a genomic foundation supporting *S. thompsoni* as a model species for further examination of the exceptional rates and patterns of genomic evolution shown by urochordates. Additionally, genomic data will allow for the development of molecular indicators of key life history events and processes and afford new understandings and predictions of impacts of climate change on this key species of Antarctic pelagic ecosystems.

## 2.5.2 Wangetal2017.pdf

[57] *Background: miRNAs play essential roles in the modulation of cellular functions via degradation and/or translation attenuation of target mRNAs. They have been surveyed in a single ascidian genus, Ciona. Recently, an annotated draft genome sequence for a distantly related ascidian, Halocynthia roretzi, has become available, but miRNAs in H. roretzi have not been previously studied. Results: We report the prediction of 319 candidate H. roretzi miRNAs, obtained through three complementary methods. Experimental validation suggests that more than half of these candidate miRNAs are expressed during embryogenesis. The majority of predicted H. roretzi miRNAs appear specific to ascidians or tunicates, and only 32 candidates, belonging to 25 families, are widely conserved across metazoans. Conclusion: Our study presents a comprehensive identification of candidate H. roretzi miRNAs. This resource will facilitate the study of the mechanisms for miRNA-controlled gene reg-*

ulatory networks during ascidian development. Further, our analysis suggests that the majority of *Halocynthia* miRNAs are specific to ascidian or tunicates, with only a small number of widely conserved miRNAs. This result is consistent with the general notion that animal miRNAs are less conserved between taxa than plant ones.

### 3 miRNAs and its rol in development

#### 3.1 miRNAs discovery and development

##### 3.1.1 Bartel2004.pdf

[3] *MicroRNAs (miRNAs) are endogenous 22 nt RNAs that can play important regulatory roles in animals and plants by targeting mRNAs for cleavage or translational repression. Although they escaped notice until relatively recently, miRNAs comprise one of the more abundant classes of gene regulatory molecules in multicellular organisms and likely influence the output of many protein-coding genes.*

##### 3.1.2 Lee1993.pdf

[34] *diverse postembryonic developmental events in C. elegans. /in-4 acts by negatively regulating the level of LIN-14 protein, creating a temporal decrease in LIN-14 protein starting in the first larval stage (L1). We have cloned the C. elegans lin-4 locus by chromosomal walking and transformation rescue. We used the C. elegans clone to isolate the gene from three other Caenorhabditis species; all four Caenorhabditis clones functionally rescue the h-4 null allele of C. elegans. Comparison of the /in-4 genomic sequence from these four species and site-directed mutagenesis of potential open reading frames indicated that /in-d does not encode a protein. Two small /in-4 transcripts of approximately 22 and 61 nt were identified in C. elegans and found to contain sequences complementary to a repeated sequence element in the 3'untranslated region (UTR) of lin-74 mRNA, suggesting that /in-4 regulates h-74 tr*

#### 3.2 miRNAs perspective evolution in development

##### 3.2.1 holland2014.pdf

[24] *Morphological comparisons among extant animals have long been used to infer their long-extinct ancestors for which the fossil record is poor or non-existent. For evolution of the vertebrates, the comparison has typically involved amphioxus*



*and vertebrates. Both groups are evolving relatively slowly, and their genomes share a high level of synteny. Both vertebrates and amphioxus have regulative development in which cell fates become fixed only gradually during embryogenesis. Thus, their development fits a modified hourglass model in which constraints are greatest at the phylotypic stage (i.e., the late neurula/early larva), but are somewhat greater on earlier development than on later development. In contrast, the third group of chordates, the tunicates, which are sister group to vertebrates, are evolving rapidly. Constraints on evolution of tunicate genomes are relaxed, and they have discarded key developmental genes and organized much of their coding sequences into operons, which are transcribed as a single mRNA that undergoes trans-splicing. This contrasts with vertebrates and amphioxus, whose genomes are not organized into operons. Concomitantly, tunicates have switched to determinant development with very early fixation of cell fates. Thus, tunicate development more closely fits a progressive divergence model (shaped more like a wine glass than an hourglass) in which the constraints on the zygote and very early development are greatest. This model can help explain why tunicate body plans are so very diverse. The relaxed constraints on development after early cleavage stages are correlated with relaxed constraints on genome evolution. The question remains: which came first?*

### 3.2.2 Sun2013.pdf

*[50] MicroRNAs (miRNAs) are ~22 nt RNAs that coordinate vast regulatory networks in animals and thereby influence myriad processes. This Review examines evidence that miRNAs have continuous roles in adults in ways that are separable from developmental control. Adult-specific activities for miRNAs have been described in various stem cell populations, in the context of neural function and cardiovascular biology, in metabolism and ageing, and during cancer. In addition to reviewing recent results, we also discuss methods for studying miRNA activities specifically in adults and evaluate their relative strengths and weaknesses. A fuller understanding of continuous functions of miRNAs in adults has bearing on efforts and opportunities to manipulate miRNAs for therapeutic purposes.*

### 3.2.3 Kloosterman2006.pdf

*[28] MicroRNAs (miRNAs) control gene expression by translational inhibition and destabilization of mRNAs. While hundreds of miRNAs have been found, only a few have been studied in detail. miRNAs have been implicated in tissue morphogenesis, cellular processes like apoptosis, and major signaling pathways. Emerging evidence suggests a direct link between miRNAs and disease, and miRNA expression signatures are associated with various types of cancer. In addition, the gain and loss of miRNA target sites appears to be causal to some genetic disorders. Here, we discuss the current literature on the role of miRNAs in animal development and disease. ?? 2006 Elsevier Inc. All rights reserved.*

### 3.2.4 Iyengar2014.pdf

[25] *The human brain is one of the most complex biological systems, and the cognitive abilities have greatly expanded compared to invertebrates without much expansion in the number of protein coding genes. This suggests that gene regulation plays a very important role in the development and function of nervous system, by acting at multiple levels such as transcription and translation. In this article we discuss the regulatory roles of three classes of non-protein coding RNAs (ncRNAs)-microRNAs (miRNAs), piwi-interacting RNA (piRNAs) and long-non-coding RNA (lncRNA), in the process of neurogenesis and nervous function including control of synaptic plasticity and potential roles in neurodegenerative diseases. miRNAs are involved in diverse processes including neurogenesis where they channelize the cellular physiology toward neuronal differentiation. miRNAs can also indirectly influence neurogenesis by regulating the proliferation and self renewal of neural stem cells and are dysregulated in several neurodegenerative diseases. miRNAs are also known to regulate synaptic plasticity and are usually found to be co-expressed with their targets. The dynamics of gene regulation is thus dependent on the local architecture of the gene regulatory network (GRN) around the miRNA and its targets. piRNAs had been classically known to regulate transposons in the germ cells. However, piRNAs have been, recently, found to be expressed in the brain and possibly function by imparting epigenetic changes by DNA methylation. piRNAs are known to be maternally inherited and we assume that they may play a role in early development. We also explore the possible function of piRNAs in regulating the expansion of transposons in the brain. Brain is known to express several lncRNA but functional roles in brain development are attributed to a few lncRNA while functions of most of them remain unknown. We review the roles of some known lncRNA and explore the other possible functions of lncRNAs including their interaction with miRNAs.*

## 3.3 Specific examples

### 3.3.1 Chenetal2014.PDF

[9] *Notch-Delta signaling is a fundamental cell-cell communication mechanism that governs the differentiation of many cell types. Most existing mathematical models of Notch-Delta signaling are based on a feedback loop between Notch and Delta leading to lateral inhibition of neighboring cells. These models result in a checkerboard spatial pattern whereby adjacent cells express opposing levels of Notch and Delta, leading to alternate cell fates. However, a growing body of biological evidence suggests that Notch-Delta signaling produces other patterns that are not checkerboard, and therefore a new model is needed. Here, we present an expanded Notch-Delta model that builds upon previous models, adding a local Notch activity gradient, which affects long-range patterning, and the activity of a regulatory microRNA. This model is motivated by our experiments in the ascidian *Ciona intestinalis* showing*

*that the peripheral sensory neurons, whose specification is in part regulated by the coordinate activity of Notch-Delta signaling and the microRNA miR-124, exhibit a sparse spatial pattern whereby consecutive neurons may be spaced over a dozen cells apart. We perform rigorous stability and bifurcation analyses, and demonstrate that our model is able to accurately explain and reproduce the neuronal pattern in Ciona. Using Monte Carlo simulations of our model along with miR-124 transgene over-expression assays, we demonstrate that the activity of miR-124 can be incorporated into the Notch decay rate parameter of our model. Finally, we motivate the general applicability of our model to Notch-Delta signaling in other animals by providing evidence that microRNAs regulate Notch-Delta signaling in analogous cell types in other organisms, and by discussing evidence in other organisms of sparse spatial patterns in tissues where Notch-Delta signaling is active.*

### 3.3.2 Fuetal2008.pdf

*[14] Recent studies reveal correlation between microRNA (miRNA) innovation and increased developmental complexity. This is exemplified by dramatic expansion of the miRNA inventory in vertebrates, a lineage where genome duplication has played a significant evolutionary role. Urochordates, the closest extant group to the vertebrates, exhibit an opposite trend to genome and morphological simplification. We show that the urochordate, larvacean, Oikopleura dioica, possesses the requisite miRNA biogenic machinery. The miRNAs isolated by small RNA cloning were expressed throughout the short life cycle, a number of which were stocked as maternal determinants prior to rapid embryonic development. We identify sex-specific miRNAs that appeared as male/female gonad differentiation became apparent and were maintained throughout spermatogenesis. Whereas 80% of mammalian miRNAs are hosted in introns of protein-coding genes, the majority of O. dioica miRNA loci were located in antisense orientations to such genes. Including sister group ascidians in analysis of the urochordate miRNA repertoire, we find that 11 highly conserved bilaterian miRNA families have been lost or derived to the point they are not recognizable in urochordates and a further 4 of these families are absent in larvaceans. Subsequent to this loss/derivation, at least 29 novel miRNA families have been acquired in larvaceans. This suggests a profound reorganization of the miRNA repertoire integral to evolution in the urochordate lineage.*

### 3.3.3 Hendrixetal2010.pdf

*[19] MicroRNAs (miRs) have been broadly implicated in animal development and disease. We developed a novel computational strategy for the systematic, whole-genome identification of miRs from high throughput sequencing information. This method, miRTRAP, incorporates the mechanisms of miR biogenesis and includes additional criteria regarding the prevalence and quality of small RNAs arising from*

*the antisense strand and neighboring loci. This program was applied to the simple chordate Ciona intestinalis and identified nearly 400 putative miR loci.*

### 3.3.4 Kusakabeetal2013.pdf

[29] *Muscle-specific miR-1/206 and miR-133 families have been suggested to play fundamental roles in skeletal and cardiac myogenesis in vertebrates. To gain insights into the relationships between the divergence of these miRs and muscular tissue types, we investigated the expression patterns of miR-1 and miR-133 in two ascidian Ciona species and compared their genomic structures with those of other chordates. We found that Ciona intestinalis and Ciona savignyi each possess a single copy of the miR-1/miR-133 cluster, which is only 350 nucleotide long. During embryogenesis, Ciona miR-1 and miR-133 are generated as a single continuous primary transcript accumulated in the nuclei of the tail muscle cells, starting at the gastrula stage. In adults, mature miR-133 and miR-1 are differentially expressed in the heart and body wall muscle. Expression of the reporter gene linked to the 850-bp upstream region of the predicted transcription start site confirmed that this region drives the muscle-specific expression of the primary transcript of miR-1/miR-133. In many deuterostome lineages, including that of Ciona, the miR-1/133 cluster is located in the same intron of the mind bomb (mib) gene in reverse orientation. Our results suggest that the origin of genomic organization and muscle-specific regulation of miR-1/133 can be traced back to the ancestor of chordates. Duplication of this miR cluster might have led to the remarkable elaboration in the morphology and function of skeletal muscles in the vertebrate lineage. © 2012 Elsevier B.V. All rights reserved.*

### 3.3.5 Pasquinellietal2000.pdf

[45] *Two small RNAs regulate the timing of Caenorhabditis elegans development. Transition from the first to the second larval stage fates requires the 22-nucleotide lin-4 RNA, and transition from late larval to adult cell fates requires the 21-nucleotide let-7 RNA. The lin-4 and let-7 RNA genes are not homologous to each other, but are each complementary to sequences in the 3' untranslated regions of a set of protein-coding target genes that are normally negatively regulated by the RNAs. Here we have detected let-7 RNAs of approximately 21 nucleotides in samples from a wide range of animal species, including vertebrate, ascidian, hemichordate, mollusc, annelid and arthropod, but not in RNAs from several cnidarian and poriferan species, Saccharomyces cerevisiae, Escherichia coli or Arabidopsis. We did not detect lin-4 RNA in these species. We found that let-7 temporal regulation is also conserved: let-7 RNA expression is first detected at late larval stages in C. elegans and Drosophila, at 48 hours after fertilization in zebrafish, and in adult stages of annelids and molluscs. The let-7 regulatory RNA may control late temporal transitions during development across animal phylogeny.*

### 3.3.6 Tangetal2013.pdf

*JoyceTang2013 The formation of the sensory organs and cells that make up the peripheral nervous system (PNS) relies on the activity of transcription factors encoded by proneural genes (PNGs). Although PNGs have been identified in the nervous systems of both vertebrates and invertebrates, the complexity of their interactions has complicated efforts to understand their function in the context of their underlying regulatory networks. To gain insight into the regulatory network of PNG activity in chordates, we investigated the roles played by PNG homologs in regulating PNS development of the invertebrate chordate Ciona intestinalis. We discovered that in Ciona, MyT1, Pou4, Atonal, and NeuroD-like are expressed in a sequential regulatory cascade in the developing epidermal sensory neurons (ESNs) of the PNS and act downstream of Notch signaling, which negatively regulates these genes and the number of ESNs along the tail midlines. Transgenic embryos mis-expressing any of these proneural genes in the epidermis produced ectopic midline ESNs. In transgenic embryos mis-expressing Pou4, and MyT1 to a lesser extent, numerous ESNs were produced outside of the embryonic midlines. In addition we found that the microRNA miR-124, which inhibits Notch signaling in ESNs, is activated downstream of all the proneural factors we tested, suggesting that these genes operate collectively in a regulatory network. Interestingly, these factors are encoded by the same genes that have recently been demonstrated to convert fibroblasts into neurons. Our findings suggest the ascidian PNS can serve as an in vivo model to study the underlying regulatory mechanisms that enable the conversion of cells into sensory neurons. ?? 2013 Elsevier Inc.*

### 3.3.7 Spina2017.pdf

[49] *Here we present a parallel study of mRNA and microRNA expression during oral siphon (OS) regeneration in Ciona robusta, and the derived network of their interactions. In the process of identifying 248 mRNAs and 15 microRNAs as differentially expressed (DE), we also identified 57 novel microRNAs, several of which are among the most highly DE. Analysis of functional categories identified enriched transcripts related to stress responses and apoptosis at the wound healing stage, signaling pathways including Wnt and TGF- during early regrowth, and negative regulation of extracellular proteases in late stage regeneration. Consistent with the expression results we found that inhibition of TGF- signaling blocked OS regeneration. A correlation network was subsequently inferred for all predicted microRNA-mRNA target pairs expressed during regeneration. Network based clustering associated transcripts into 22 non-overlapping groups, functional analysis of which showed enrichment of stress response, signaling pathway and extracellular protease categories could be related to specific microRNAs. Finally, predicted targets of the miR-9 cluster suggest a role in regulating differentiation and proliferative state of neural progenitors through regulation of the cytoskeleton and cell cycle.*

## 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 [51]. 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 [51]. Subsequently most YC transcripts enter the primary muscle cell lineage after cleavage and are also present in the secondary muscle cell lineage [51]. YC RNA was first discovered in the club tunicate *Styela clava* [51]. 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 [51]. 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 [51].

### 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 [7]. During a study focused on identifying miRNAs in the simple chordate *C. intestinalis* moRNAs were first discovered [48]. Unexpectedly, half of the *C. intestinalis* 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 [48]. 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 [7]. 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 [30]. 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 [48]. The expression level of certain moRNAs can even be greater than for their corresponding miRNA [54]. Finally, it can be argued [7] 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 [30, 54]. What functions moRNAs may have, varies. For example, moRNA expression was recorded in solid tumours, together with other small RNAs [39]. In addition the fact that an 18-fold enrichment of moRNAs was observed in the nucleus [52] indicates that at least some moRNAs may have functions related to nuclear processes [7]. 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 [42]. 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 [42]. 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 [42]. 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 [42]. 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 [56]. 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. intestinalis*, *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 [56].

### 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 [55]. 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 [15]. SL RNA trans splicing has not only been described for euglenoids, kinetoplastids, cnidarians, nematodes, and Platyhelminthes [15], but also for deuterostomes like the simple chordate *C. intestinalis* [55] and the appendicularian *O. dioica* [15]. Hereby *O. dioica* was shown to not only trans-splice SL RNAs to mRNAs, as does *C. intestinalis*, but also to use trans splicing in resolving polycistronic transcripts [15]. 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. intestinalis* to 41 nt in trypanosomatids [15]. 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 [38] and the resolution of polycistronic pre-mRNAs [1, 6], to the production of functional mRNAs from RNA polymerase I transcripts [33].

## References

1. Nina Agabian. Trans splicing of nuclear pre-mRNAs. *Cell*, 61(7):1157 – 1160, 1990.
2. 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.
3. David P. Bartel. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, 116(2):281–297, 2004.
4. David P. Bartel. MicroRNA target recognition and regulatory functions. *Cell*, 136:215–233, 2009.
5. Arash Bashirullah, Ramona L Cooperstock, and Howard D Lipshitz. RNA localization. *Annu. Rev. Biochem.*, 67:335–394, 1998.
6. Thomas Blumenthal. Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends in Genetics*, 11(4):132 – 136, 1995.
7. 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.
8. Simona Candiani. Focus on miRNAs evolution: A perspective from amphioxus. *Briefings in Functional Genomics*, 11(2):107–117, 2012.
9. 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.
10. Zhonghua Dai, Zuo Zhou Chen, Hua Ye, Longhai Zhou, Lixue Cao, Yiquan Wang, Sihua Peng, and Liangbiao Chen. Characterization of microRNAs in cephalochordates reveals a correlation between microRNA repertoire homology and morphological similarity in chordate evolution. *Evolution and Development*, 11(1):41–49, 2009.
11. Matheus De Souza Gomes, Mark T A Donoghue, Mohankumar Muniyappa, Roberta Vericiano Pereira, Renata Guerra-S??, and Charles Spillane. Computational identification and evolutionary relationships of the MicroRNA gene cluster miR-71/2 in protostomes. *Journal of Molecular Evolution*, 76(6):353–358, 2013.
12. 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.



13. 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 Research*, 40(1):37–52, 2012.
14. Xianghui Fu, Marcin Adamski, and Eric M. Thompson. Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*. *Molecular Biology and Evolution*, 25(6):1067–1080, 2008.
15. 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.
16. Sam Griffiths-Jones, Jerome H L Hui, Antonio Marco, and Matthew Ronshaugen. MicroRNA evolution by arm switching. *EMBO reports*, 12(2):172–177, 2011.
17. 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.
18. 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.
19. 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.
20. 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.
21. Jana Hertel and Peter Stadler. The Expansion of Animal MicroRNA Families Revisited. *Life*, 5(1):905–920, 2015.
22. Jana Hertel and Peter F. Stadler. Hairpins in a Haystack: Recognizing microRNA precursors in comparative genomics data. *Bioinformatics*, 22(14):197–202, 2006.
23. Jana Hertel and Peter F. Stadler. The expansion of animal microRNA families revisited. *Life*, 5:905–920, 2015.
24. Linda Z. Holland. Genomics, evolution and development of amphioxus and tunicates: The Goldilocks principle, 2015.
25. Bharat R. Iyengar, Ashwani Choudhary, Mayuresh A. Sarangdhar, K. V. Venkatesh, Chetan J. Gadgil, and Beena Pillai. Non-coding RNA interact to regulate neuronal development and function. *Frontiers in Cellular Neuroscience*, 8(February):1–10, 2014.
26. 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.
27. Raja Keshavan, Michael Virata, Anisha Keshavan, and Robert W. Zeller. Computational Identification of *Ciona intestinalis* MicroRNAs. *Zoological Science*, 27(2):162–170, 2010.
28. Wigard P. Kloosterman and R. H A Plasterk. The Diverse Functions of MicroRNAs in Animal Development and Disease. *Developmental Cell*, 11(4):441–450, 2006.
29. 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.
30. 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.
31. 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.

32. 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.
33. 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.
34. R C Lee, R L Feinbaum, and V Ambros. the *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 75: 843–85:843–854, 1993.
35. Matthieu Legendre, André Lambert, and Daniel Gautheret. Profile-based detection of microRNA precursors in animal genomes. *Bioinformatics*, 21(7):841–845, 2005.
36. Shi-Lung Lin, Joseph D. Miller, and Shao-Yao Ying Ying. Intronic MicroRNA (mirna). *J Biomed Biotechnol.*, 2006:26818, 2006.
37. E Lund, S Güttinger, A Calado, J E Dahlberg, and U Kutay. Nuclear export of microRNA precursors. *Science*, 303:95–98, 2004.
38. 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.
39. 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 Sittenbon. Discovery of microRNAs and other small RNAs in solid tumors. *Nucleic Acids Research*, 38(18):6234–6246, 2010.
40. Kristin Missal, Dominic Rose, and Peter F. Stadler. Non-coding RNAs in *Ciona intestinalis*. *Bioinformatics*, 21(SUPPL. 2):77–78, 2005.
41. Eric P Nawrocki and Sean R Eddy. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29:2933–2935, 2013.
42. 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.
43. 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.
44. 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.
45. 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.
46. Miguel Pignatelli, Albert J. Vilella, Matthieu Muffato, Leo Gordon, Simon White, Paul Flicek, and Javier Herrero. ncRNA orthologies in the vertebrate lineage. *Database*, 2016(September 2017):1–20, 2016.
47. L F Sempere, C N Cole, M A McPeck, and K J Peterson. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J Exp Zool B Mol Dev Evol.*, 306B:575–588, 2006.
48. 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.
49. E. J. Spina, E. Guzman, H. Zhou, K. S. Kosik, and W. C. Smith. A microRNA-mRNA expression network during oral siphon regeneration in *Ciona*. *Development*, 144(10):1787–1797, May 2017.
50. Kailiang Sun and Eric C. Lai. Adult-specific functions of animal microRNAs. *Nature Reviews Genetics*, 14(8):535–548, 2013.
51. 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.
52. Ryan J Taft, Ken C Pang, Timothy R Mercer, Marcel Dingler, and John S Mattick. Non-coding RNAs: regulators of disease. *The Journal of Pathology*, 220(2):126–139, 2010.

53. 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.
54. 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.
55. 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.
56. Cristian A. Velandia-Huerto, Adriaan A. Gittenberger, Federico D. Brown, Peter F. Stadler, and Clara I. Bermúdez-Santana. Automated detection of ncRNAs in the draft genome sequence of a colonial tunicate: the carpet sea squirt *Didemnum vexillum*. *BMC Genomics*, 17(1):691, Aug 2016.
57. 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.
58. 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.
59. 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.
60. Zhen Yang, Xueshuai Wan, Zhuoya Gu, Haohai Zhang, Xiaobo Yang, Lian He, Ruoyu Miao, Yang Zhong, and Haitao Zhao. Evolution of the mir-181 microRNA family. *Computers in Biology and Medicine*, 52:82–87, 2014.

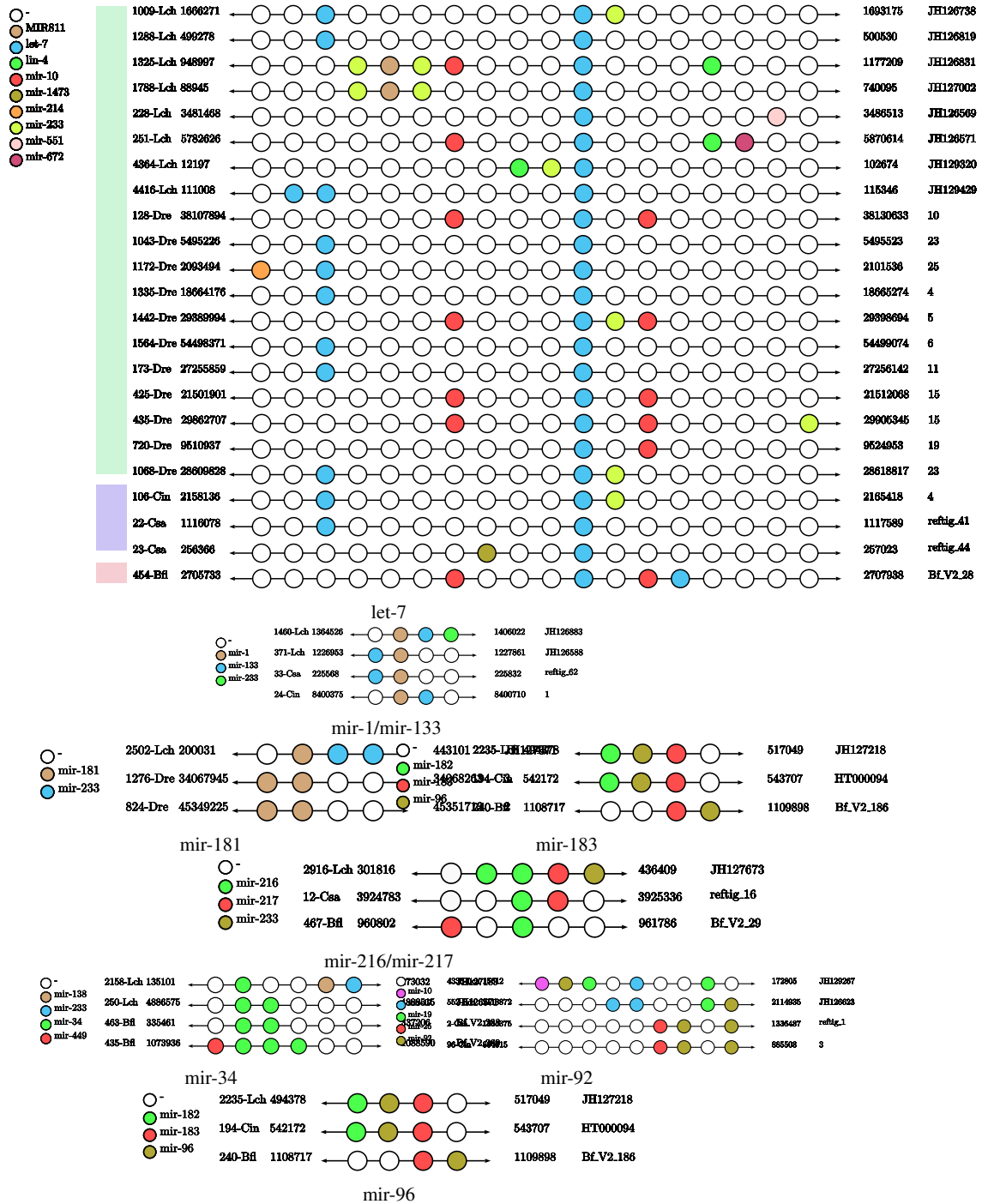


Fig. 2: Multiple alignments of miRNA's clusters. **Prot:** Protostomata, **Brfl:** *B. floridae*, **Oidi:** *O. dioica*, **Dvex:** *D. vexillum*, **Ciin:** *C. intestinalis*, **Cisa:** *C. savignyi*, **Ciro:** *C. robusta*, **Sath:** *S. thompsoni*, **Mata:** *M. oculata*, **Mlta:** *M. occulta*, **Mlis:** *M. occidentalis*, **Bosc:** *B. schlosseri*, **Haro:** *H. roretzi*, **Pema:** *P. marinus*, **Dare:** *D. rerio*, **Lach:** *L. chalumnae*, **Xetr:** *X. tropicalis* and **Anca:** *A. carolinensis*.

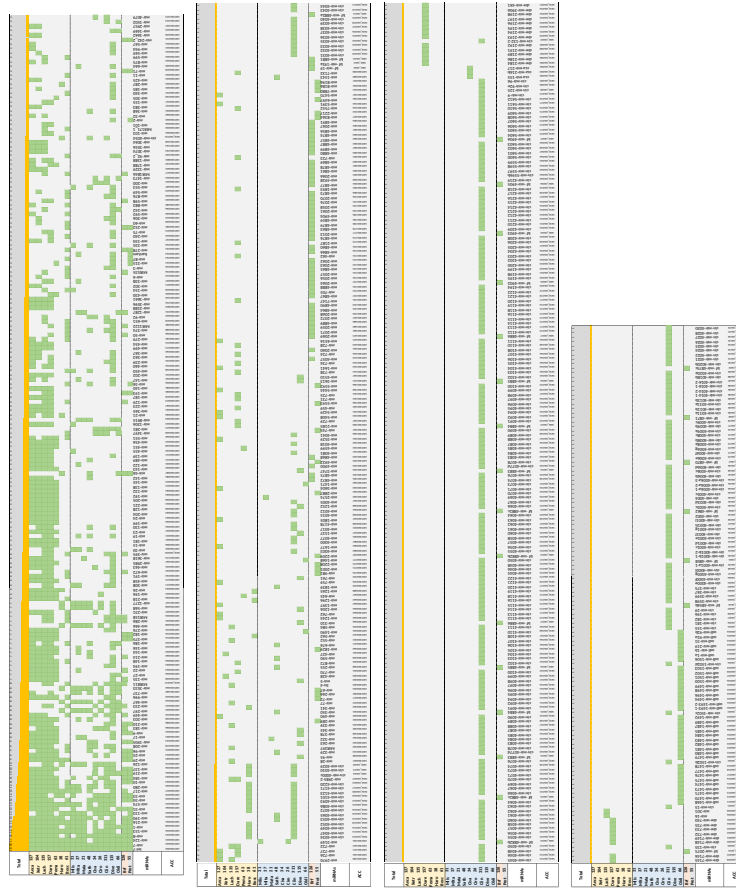
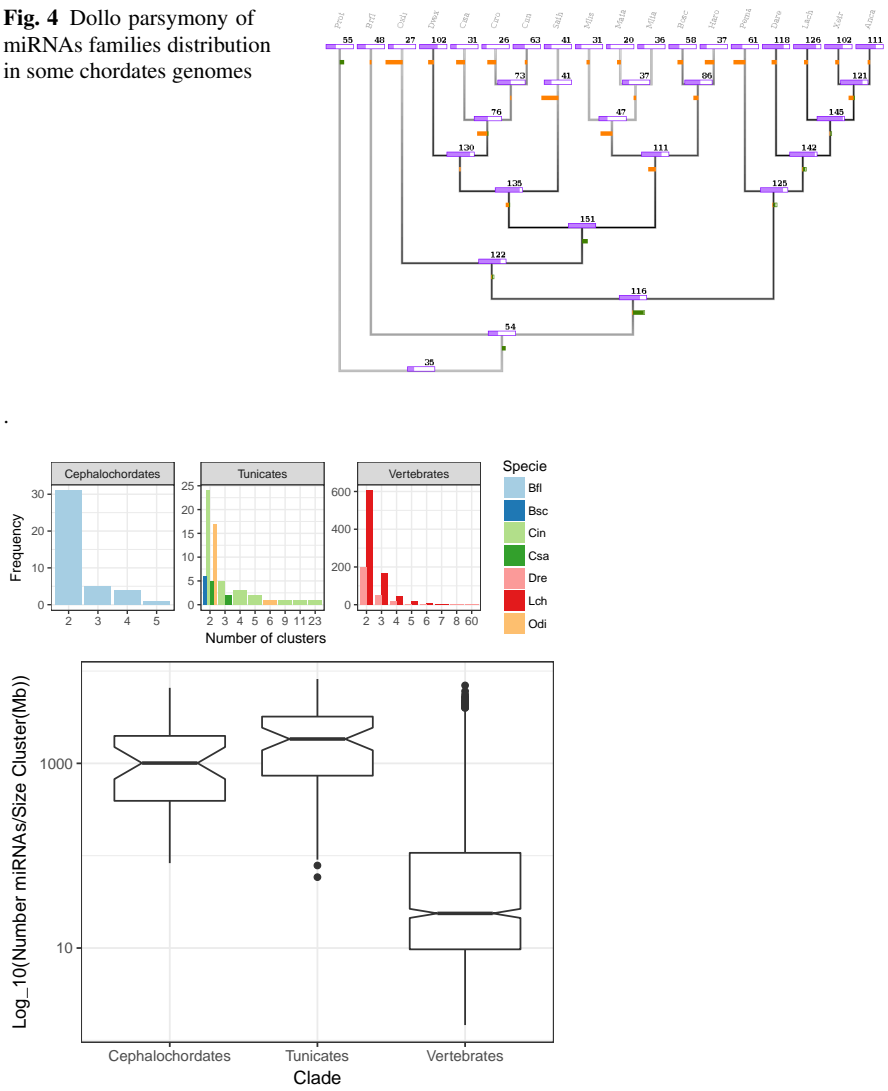


Fig. 3: Absence/Presence Matrix of miRNAs families along Bilaterian species. **Prot**: Protostomata, **Brfl**: *B. floridae*, **Oidi**: *O. dioica*, **Dvex**: *D. vexillum*, **Ciin**: *C. intestinalis*, **Cisa**: *C. savignyi*, **Ciro**: *C. robusta*, **Sath**: *S. thompsoni*, **Mata**: *M. oculata*, **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*.

**Fig. 4** Dollo parsimony of miRNAs families distribution in some chordates genomes



**Fig. 5:** Analysis of the distribution, size and number's of cluster along chordate species.

**Fig. 6** Comparison between miRNAs families along Bilateralian species. Same labels from Figure 3 were used. In this case *vertebrates* group the following species: **Pema**: *P. marinus*, **Dare**: *D. rerio*, **Lach**: *L. chalumnae*, **Xetr**: *X. tropicalis* and **Anca**: *A. carolinensis*

