

Non-protein-coding RNAs as a regulators of development in tunicates

Cristian Velandia

January 19, 2018

miRNA families origin and evolutionary perspective

miRNAs in clusters

One of the most interesting aspects about the patterns of genomic locations of miRNAs is to know 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. 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 [?]. 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 [?] 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. intestinalis* located in an intergenic region [?], [?] and one in *C. savigny* overlapped in an intron [?]. By testing real time PCR co-expression of some miRNAs, their host and adjacent genes in *O. dioca* by [?] 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.

In *C. intestinalis* 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 [?] in contrast to the location of the loci in antisense orientations of protein-coding gene as is 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 [?]. In Cionas have been also detected miRNAs organized in clusters, for example in *C. intestinalis* a putative cluster was detected by [?] 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 [?]. 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.

My main concerns here that the cluster cluster let-7/miR-125/miR-100 observed in *Drosophila* and is included time ago for [?]. In his plot you see a cluster on the right side and another on the left side, I am not sure why the mir-125 is not including if right now is validated in miRBase shaping a cluster 2. Why the mir125 is not here? If it is reported time ago and is validated in miBase and by [?]? And why mir233 is on it and not for Csa and Cin?. Please could you check this let7 cluster in the current plot?. Here I follow with Cristian description of this cluster. Please check about the mir125 and do any reconciliation if is posible with the plot of the paper of [?] in 1.

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 2 mir-10, 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: 1let-7 + 2 mir-10 and 2 let-7 + other families. In this way, tunicates reported the latter group, not including mir-10 on the cluster but including mir-233 or even mir-1473.

Figure 1: evolutionlet7

Figure 2: let7-mir125-mir1473. Current distribution in miRBase

A second miRNA cluster consisting of the miR-182 and miR-183 was also detected in *C. intestinalis* in 2010 by [?] 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 14d. Here the authors also found five additional paralogs of let-7 within a 1-kb stretch previously know as Scaffold_138 which is right now described on the chromosome 4q. The current distribution for both Cionas is shown in 3. Unfortunately is not seen in the Cristian detection method?, could we plotted or is not making sense? Since let7 is a very important regulator of development

Figure 3: let7paralogs

The cluster miR-1/miR-133, expressed specifically in Cionas muscle tissues was also reported by [?]. The authors reported that one copy of this cluster is presented in both Cionas. As is shown in the plot 14b a copy is also presented in *L. chalumnae*. In 2012 a new cluster was proposed in *C. intestinalis* by [?] located on the cromosoma 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 paralogous is located on the middle of those loci. This current distribution in shown of 4 whose loci were validated by [?], [?], [?], and [?]. As was mentioned by [?] 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. intestinalis* 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 [?] today is known to be built by the families miR4000, miR-4001, miR4002, and miR4006 located on chromosome 7q Figures 5 and 6. Another cluster is also located on the same chromosome composed by the families miR-4003, miR4005 and miR4077 in Figure 7. Some other cluster are also found on the cromosoma 1a, 10q and 3p. See this structure on Figures 10, 9 and 8, most of them validated by [?]. Please check the Table 1 which are and which are not to update

Figure 4: mir4091

Figure 5: 7qf

Figure 6: 7qs

Figure 7: 7qother

Figure 8: 3p

Figure 9: 3p-7q

Figure 10: 1q-10q

Some other clusters shared between both Cionas are the cluster 92, 124 and 200 validated by [?], [?], [?] which the structure is seen on Figures 11, 12 and 13

Figure 11: mir92

Figure 12: mir124

Figure 13: mir200

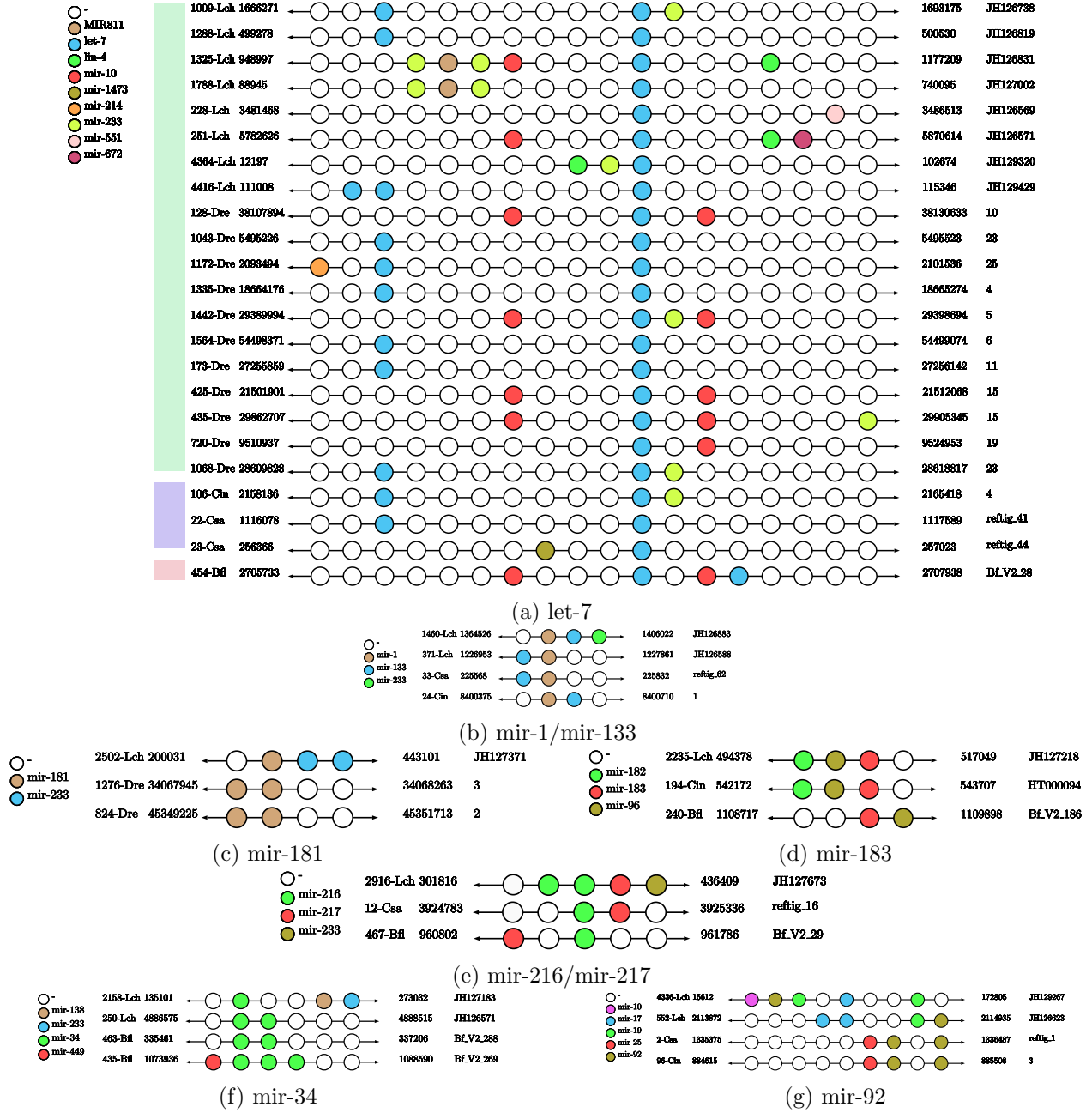


Figure 14: Multiple alignments of miRNA's clusters. **Prot**: Protostomata, **Brfl**: *B. floridae*, **Oidi**: *O. dioica*, **Dvex**: *D. vexillum*, **Ciin**: *C. intestinalis*, **Cisa**: *C. savignyi*, **Ciro**: *C. robusta*, **Sath**: *S. thompsoni*, **Mata**: *M. oculata*, **Mlta**: *M. occulta*, **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*.

Clade	Specie	Chr	Start	End	Size(Mb)	No. miRNAs	Elements
C	<i>B. floridae</i>	Bf.V2.118	216744	220351	3607	5	bfl-mir-4869, bfl-mir-4857, bfl-mir-4862, bfl-mir-4856b, bfl-mir-4856a

T	<i>O. dioica</i>	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
T	<i>B. schlosseri</i>	chrUn	40003	41320	1317	2	mir-233, mir-10
T	<i>C. intestinalis</i>	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
T	<i>C. savignyi</i>	reftig_16	3924783	3925336	553	3	csa-mir-216b, csa-mir-216a, csa-mir-217
T	<i>C. savignyi</i>	reftig_1	1335375	1336487	1112	3	csa-mir-92b, csa- mir-92c, csa-mir- 92a

V	<i>D. rerio</i>	4	28738556	28754891	16335	60	<p>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</p>
---	-----------------	---	----------	----------	-------	----	---

V	<i>L. chalumnae</i>	JH126646.1	1529355	1882777	353422	7	mir-233, mir-233, mir-233, mir-598, mir-672, MIR535, mir-233
---	---------------------	------------	---------	---------	--------	---	---

Table 1: Details of biggest miRNA cluster for chordate species

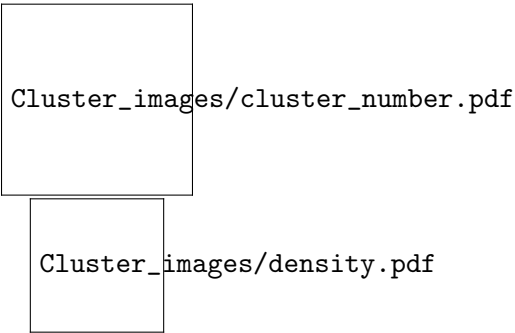


Figure 15: Analysis of the distribution, size and number’s of cluster along chordate species.

Results

Applying the last strategy to detect miRNA's clusters granted the option to study the conserved elements along chordate's genomes. As shown in Figure 15, directly with the location form those miRNA elements have been possible to identify the number and the length of those identified regions along all the studied genomes. In this case, the cluster that contains the greatest number of miRNAs elements (60) is located on *D. rerio* genome: **Chromosome 4:28738556-28754891**, for tunicates on *C. intestinalis*: **Chromosome 7:4153284-4156782** with 23 elements, and in *B. floridae*: **Bf_V2.118: 216744-220351** only 5 miRNAs have been detected (for further details Table 1 describe the miRNAs families located inside the largest clusters for each specie). Additionally, is more frequent to found clusters that contain 2 miRNAs families, but is also important to know that the relation between the number of miRNAs inside a cluster and the cluster's length is higher on tunicates and cephalochordates in comparison to vertebrates, it means that inside the identified clusters along vertebrates the miRNAs elements are more distant between them.

Comparison between miRNA clusters have been calculated in order to access to the most conserved set of miRNA families inside chordate's clusters. As a result, the following miRNAs families were identified being part of specific clusters regions: let-7, mir-1, MIR1122, mir-130, mir-132, mir-133, mir-135, mir-146, mir-15, mir-17, mir-181, mir-183, MIR1846, mir-186, mir-19, mir-193, mir-216, mir-219, mir-23, mir-24, mir-242, mir-25, mir-27, mir-286, mir-29, mir-2985-2, mir-30, mir-34, mir-395, mir-454, mir-489, MIR535, mir-8, mir-9 (Figure 14).