

How to improve the detection of miRNA homologs? an outlook from tunicate genomes.

34th TBI Winterseminar in Bled

Cristian A. Velandia Huerto¹

Advisor: Prof. Dr. *Peter F. Stadler*¹

Collaborations: *Ali Yazbeck*¹

¹Inst.f.Informatik, Universität Leipzig, Leipzig, Germany.

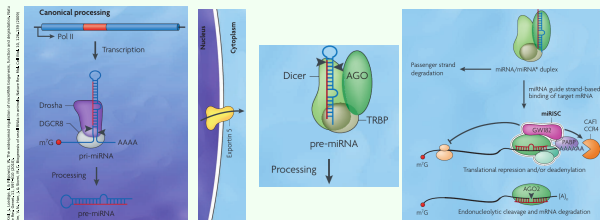
February 15, 2019

Context: Do you know miRNAs?

What is a miRNA?

- a class of small RNA (~ 21-24 bases).
- Endogenous and single strand RNAs.
- Function: *Regulation* of gene expression (via Post-transcriptional gene silencing).
- Produced by microbes, sponges, metazoan, plants and viruses.
- Important role in development and physiology.
- Biogenesis pathway is different between plants and animals [Compartmentalized].

Biogenesis



- Trimming **Pri-miRNA**.
- Transport **Pre-miRNA**.
- Cleavage stem-loop.
- Release miRNA duplex.

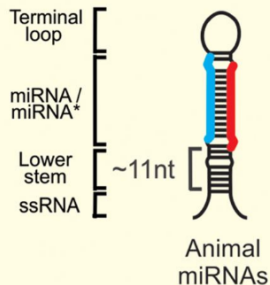
Current approaches of detection

Experimental

- Northern blot.
- Microarrays.
- *In situ* hybridization.
- Amplification techniques.

Computational

- Homology detection.
- *De novo* detection.



Bologna, N. et al. (2012)

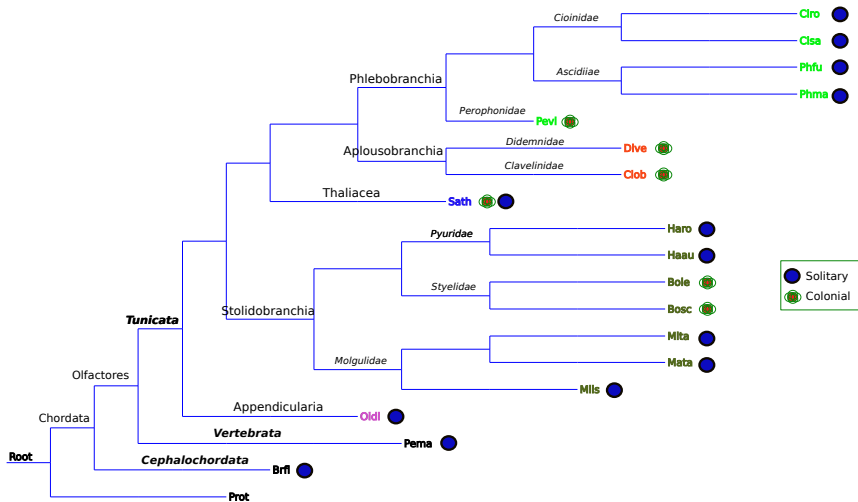
- Due experimental inability to detect all candidates: **Computational approaches.**
- Comprehensive understanding of evolution and functional adaptations of miRNAs requires a *Comprehensive annotation.*

Velandia-Huerto, CA. *et. al.*, Evolution and Phylogeny of MicroRNAs-Protocols, Pitfalls, and Problems. Submitted manuscript (2019).

Tunicates as genomic model

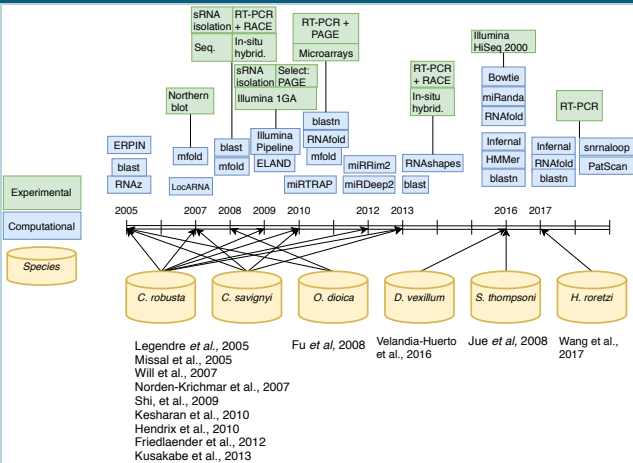


Tunicates as genomic model



Tunicates as genomic model

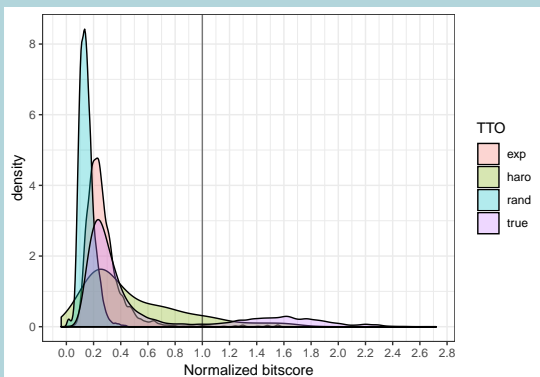
Current studies of miRNAs on tunicates



Velandia-Huerto C.A., *et al.* (2018) Nonprotein-Coding RNAs as Regulators of Development in Tunicates. Results and Problems in Cell Differentiation, vol 65. Springer, Cham

Systematically evaluation: *Halocynthia roretzi* miRNA annotation

Plotting cmsearch results...



- **Exp:** *C. robusta* homology miRNAs sequences by blastn searches (Str 4) in this study.
- **Haro:** Reported candidates in Wang, *et al.* (2017).
- **Rand:** CDS from hsa (v. GRCh38), sizes 80, 150 nt, Rand sampled (with replacement, 95% of confidence, $\alpha = 0.05$, $n=4694$, 10 replicas), shuffled nucleotides with EMBOSS and 100 steps.
- **True:** MirGeneDB.

Re-definition

Pre-defined gathering scores GA were re-defined as nGA as follows:

$$nGA = GA * 0.32 \quad (1)$$

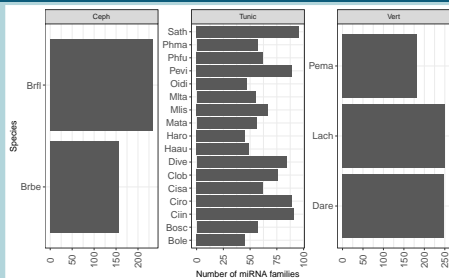
Tunicates miRNAs complement

After applying new *nGA* filters from *cmsearch* results...



- 27/37 candidates were detected applying *nGA*.
- 4 candidates correspond to miRNA families without reported CM. (New candidates?)
- 6 loci were discarded by our strategy.

Absolute frequency miRNA families along chordates



Results of new strategy

- Evident increment of miRNA families/loci.
- Tunicates < families than Cephalochordates and Vertebrates.

↑ Candidates: Improve **sensitivity**.

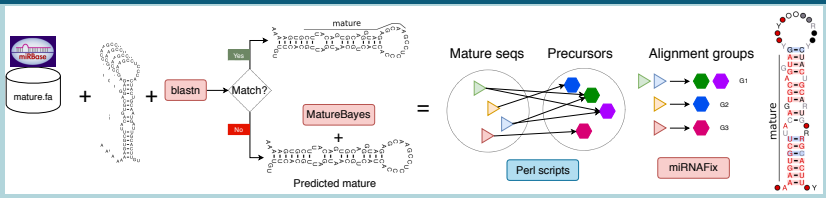
miRBase and RFAM missannotation

Some considerations...

- Definition miRNA families on miRBase and RFAM are different, so equal names **do not** mean equal biological entities!.
- Mature sequences are not reported on for RFAM miRNA families.
- miRBase is the biggest repository of reported mature sequences.
- Not all miRBase sequences have a correspondent CM on RFAM.

MIRFix annotation (Yazbeck A, (2018) personal comm.)

First Approach: Mature sequences are like keys in a %hash

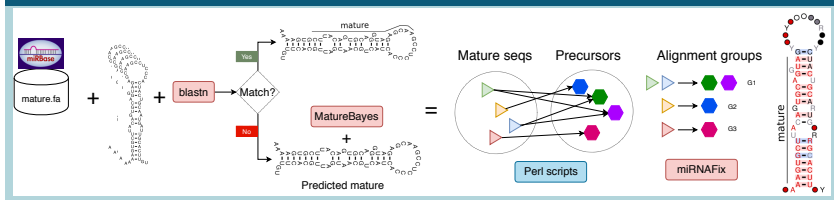


H. roretzi let-7 loci evaluation

miRNA	Mature	Valid?	Alignment Species
S93,68608-68698,-	miRBase	YES	Cionidae, Ascidiidae, Molgulidae, Styelidae, Pyuridae, Phlebobranchia
S93,67725-67805,-	Predicted	YES	Pyuridae
S93,68282-68361,-	miRBase	YES	Cionidae, Ascidiidae, Pyuridae, Styelidae, Phlebobranchia
S93,68444-68525,-	Predicted	YES	Pyuridae, Molgulidae, Styelidae
S648,2294-2373,-	miRBase	YES	Nematoda, Cephalochordata, Styelidae, Pyuridae, Vertebrata

MIRFix annotation (Yazbeck A, (2018) personal comm.)

First Approach: Mature sequences are like keys in a %hash

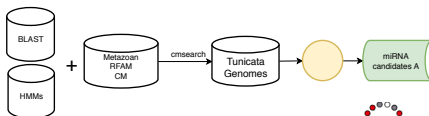


Good try...but not enough.

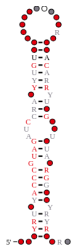
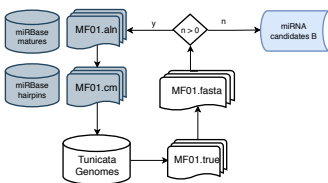
- Grouping rule for miRNAs families is too *strict* and *artificial*.
- Due grouping strategy, some already homologous families have been splitted.

Second Approach: Mixed approach on RFAM and miRBase.

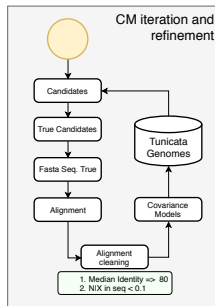
Velandia-Huerto, *et. al*
(2016)



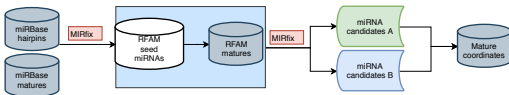
Yazbeck, et al.
Personal comm.



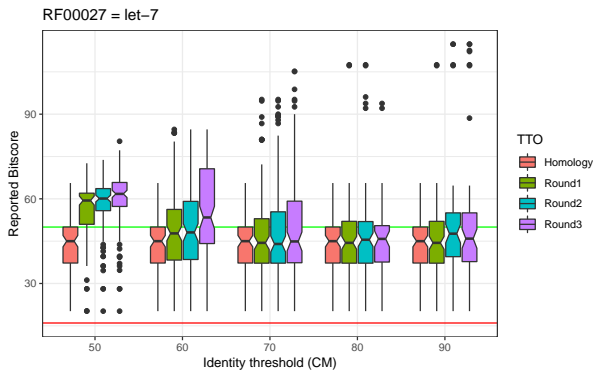
Precursors



Mature (miR) and miR* annotation



CM modifications for miRNAs sequences



Conclusions

- Most conserved miRNAs candidates were successfully detected applying default `cmsearch` threshold parameters, but *derived* miRNA loci have to be detected with a new methods to avoid misannotation.
- Additional criteria could be included effectively taking advantage of important features from miRNAs structure (precursor folding, mature position and inspection by multiple alignments) like the evaluated ones by `miRNAFix`.

Thanks!, Vielen Dank!, Obrigado!, ¡Gracias!



Peter F. Stadler



Ali Yazbeck



Clara I. Bermúdez



Adriaan Gittenberger



Federico D. Brown



UNIVERSITÄT
LEIPZIG



Universiteit
Leiden



UNIVERSIDAD
NACIONAL
DE COLOMBIA



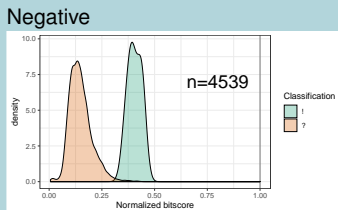
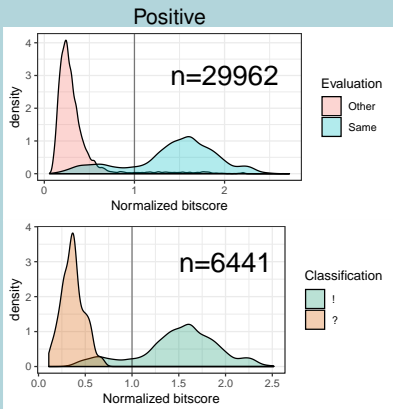
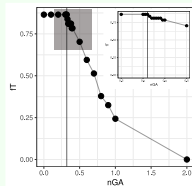
USP Universidade de São Paulo
Instituto
de Biociências

DAAD

Deutscher Akademischer Austausch Dienst
Servicio Alemán de Intercambio Académico



CM re-definition scores

Plotting `cmsearch` results...Haro data evaluated and nGA  $nGA = 0.32$

Assumption: Haro candidates are true candidates.