

microARN - BIF7101

Mohamed Amine Remita
Abdoulaye Baniré Diallo *Ph.D.*

25 février 2015

- 1 Interférence par ARN
- 2 C'est quoi un microARN ?
- 3 Biogenèse des microARNs
- 4 Fonctions des microARNs
- 5 microARNs et bioinformatique
 - Prédiction et caractérisation des microARNs
 - Prédiction des gènes cibles
 - microARN et séquençage à haut débit
 - Bases de données, services web et visualisation

Interférence par ARN

- Mécanisme de régulation post-transcriptionnelle de l'expression des gènes
- Implique des petits ARNs simple ou double brin
- Répression ou dégradation de l'ARN messager

Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*

Andrew Fire*, SiQun Xu*, Mary K. Montgomery*, Steven A. Kostas*†, Samuel E. Driver‡ & Craig C. Mello‡

* Carnegie Institution of Washington, Department of Embryology,
115 West University Parkway, Baltimore, Maryland 21210, USA

† Biology Graduate Program, Johns Hopkins University,
3400 North Charles Street, Baltimore, Maryland 21218, USA

‡ Program in Molecular Medicine, Department of Cell Biology,
University of Massachusetts Cancer Center, Two Biotech Suite 213,
373 Plantation Street, Worcester, Massachusetts 01605, USA

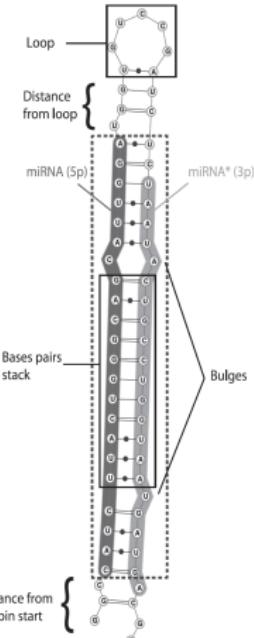
Experimental introduction of RNA into cells can be used in certain biological systems to interfere with the function of an endogenous gene^{1,2}. Such effects have been proposed to result from a simple antisense mechanism that depends on hybridization between the injected RNA and endogenous messenger RNA transcripts. RNA interference has been used in the nematode *Caenorhabditis elegans* to manipulate gene expression^{3,4}. Here we investigate the requirements for structure and delivery of the

microARN - structure primaire

- Découverts chez *C. elegans* en 1993 (*lin-4*) (Lee *et al.*)
 - Nommés « microRNAs » en 2001 (Lagos-Quintana *et al.*)
-
- Petit brin d'ARN
 - Taille de 18 nt à 25 nt
 - Issus d'une plus longue séquence d'ARN appelée pré-miARN
 - Localisés dans plusieurs régions du génome
 - Exprimés par des eucaryotes, procaryotes et virus

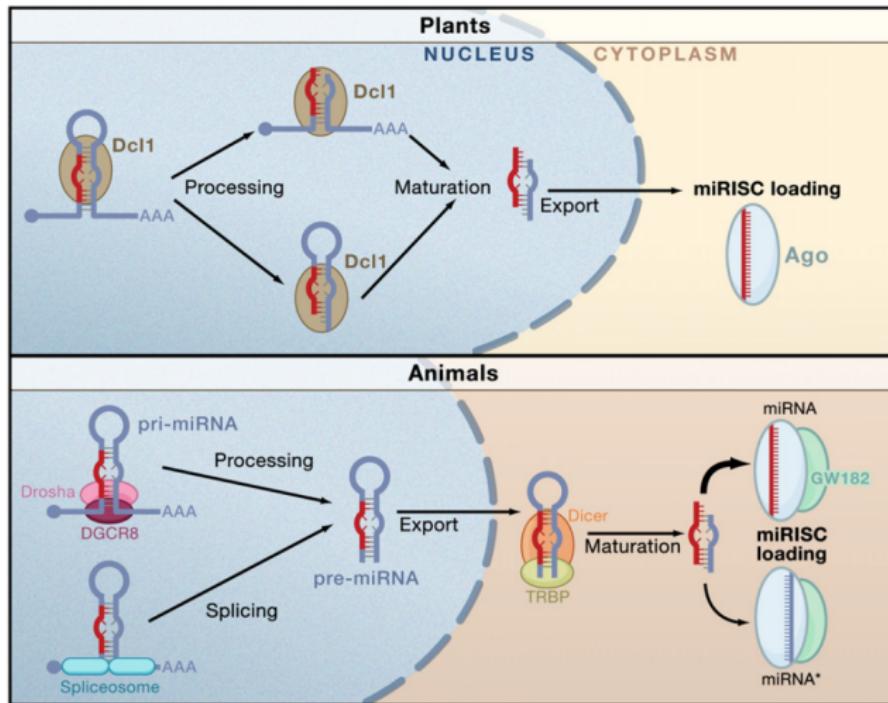
microARN - structure secondaire

- pré-microARN, séquence d'une centaine de nucléotides
- Pliée en épingle cheveux (tige-boucle)
- Complémentarité imparfaite (boucle, hernies)
- Duplex microARN 5p - microARN 3p



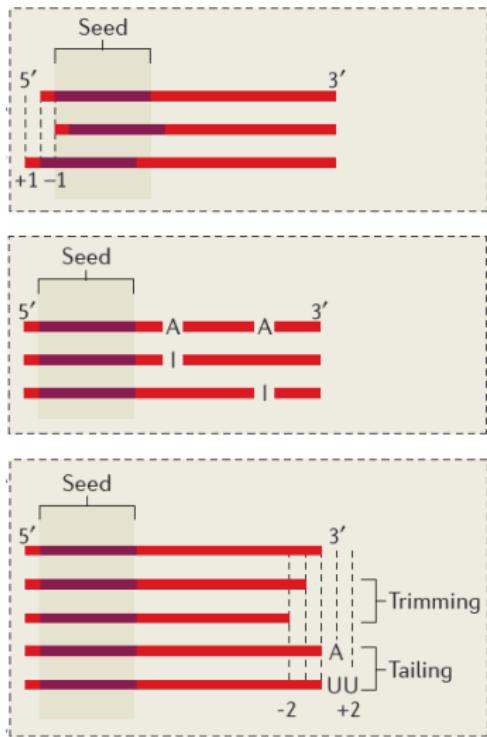
Structure secondaire d'un pré-miARN (Leclercq et al., 2013)

Biogenèse



Biogenèse des microARN chez les plantes et les animaux (Carthew et Sontheimer, 2009)

Isomirs



Différents types d'isomirs (isoformes) (Ameres et Zamore, 2013, modifié)

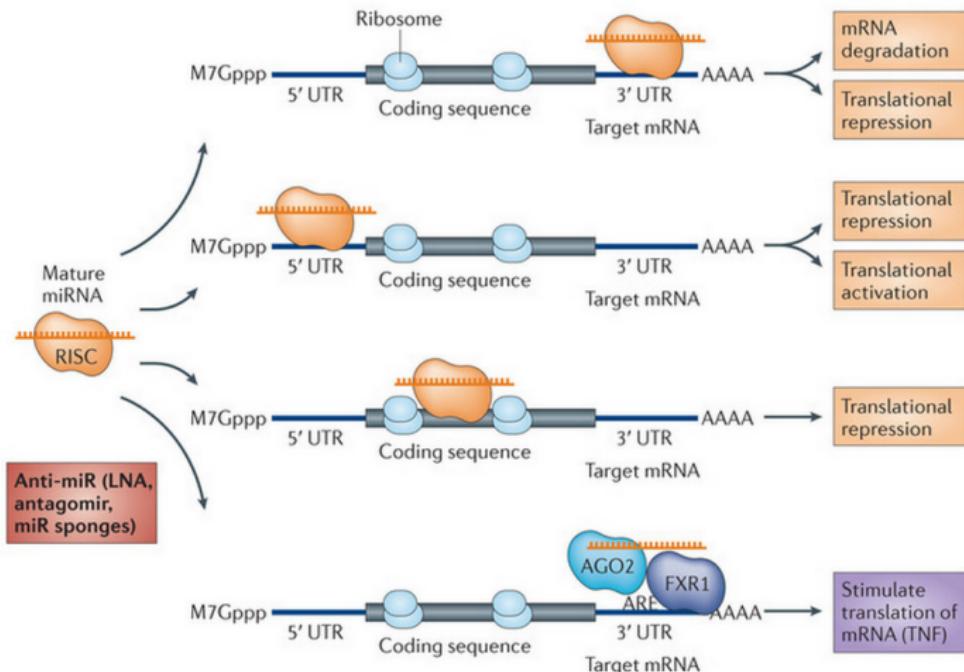
microARNs chez les animaux et les plantes

	Plants	Animals
Number of miRNA genes present	100 < number < 200 of genes	100 < number < 500 of genes
Location within genome	Predominantly intergenic regions	Intergenic regions introns
Presence of miRNA clusters	Uncommon	Common
MiRNA biosynthesis	Dicer-like	Drosha, Dicer
Mechanism of repression	mRNA-cleavage (methylation ?)	Translational repression
Location of miRNA binding motifs within target genes	Predominantly the open-reading frame	Predominantly the 3-UTR
Number of miRNAbinding sites within target genes	Generally one	Generally multiple
Function of known target genes	Regulatory genes-crucial for development, enzymes	Regulatory genes-crucial for development, structural proteins, enzymes
(Millar <i>et al.</i> , 2005)		

Implications des microARNs

- Développement et croissance
- Processus biologiques
- Stress biotiques (réponses immunitaires)
- Stress abiotiques (Sécheresse, froid, salinité, etc.)
- Implication dans le cancer
- ... etc.

Mécanismes de fonctions des microARNs



Différentes fonctions des microARNs (Ling et al., 2013, modifiée)

microARNs et bioinformatique

- Prédiction des microARNs
- Prédiction des gènes cibles
- microARN et séquençage à haut débit (NGS)
- Bases de données, services web et visualisation

Prédiction par homologie

- Méthodes comparatives
- Conservation chez des espèces proches ou lointaines
 - Conservation de la structure secondaire
 - Conservation de la séquence
- srnaloop (Grad *et al.*, 2003) (*C. elegans*)
- MiRscan (Lim *et al.*, 2003) (Nématodes)
- MiRseeker (Lai *et al.*, 2003) (Mouches)
- Phylogenetic shadowing (Berezikov *et al.*, 2005)
- Blast

Prédiction *ab initio*

- Approches basées sur des filtres
- Approches basées sur l'apprentissage automatique (machine learning)
- Approches basées sur le duplex microARN :ARNm
- Approches mixtes traitant les données NGS

Prédiction par filtres

- Identification des potentielles séquences en épingle à cheveux qui
 - portent des séquences inversement répétées
 - ou s'alignent avec des petits ARNs séquencés
- Filtres :
 - Taux de GC
 - Longueur du précurseur
 - Minimum Folding Energy (MFE)
 - Nombre des nucléotides non appariés dans le duplex miARN :miARN*
 - Nombre des hernies (bulges) et des boucles internes
 - Différence de taille entre le miARN et miARN*
- MiRscanII (Ohler *et al.*, 2004)
- MIRcheck (Jones-Rhoades et Bartel, 2004)
- MIRFINDER (Bonnet *et al.*, 2004)

Prédiction par apprentissage automatique

- Utilisation d'un ensemble d'attributs/caractéristiques
- Entrainement des données avec deux ensembles
 - Données positives : microARNs et leurs précurseurs validés expérimentalement ou potentiellement vrais
 - Données négatives : D'autres types d'ARN (ARNr, ARNt, etc.) ou séquences génomiques extraites aléatoirement.
- ProMir (Nam *et al.*, 2005)
- Triplet-SVM (Xue *et al.*, 2005)
- miR-abela (Sewer *et al.*, 2005)
- miPred (Jiang *et al.*, 2007)
- HHMMIR (Kadri *et al.*, 2009)
- mirDup (Leclercq *et al.*, 2013)
- ...etc.

Prédiction par apprentissage automatique

Quelques caractéristiques

Attribut	Description
Triplets	Pourcentage de chaque triplet (Défini dans Xue et al. 2005)
Dinucléotides	Pourcentage de AA,AU,AG,AC,UA,UU,UG,UC,GA,GU,GG,GC,CA,CU,CG
Hemies	Positions, longueur et nombre de hernies dans le microARN et environs
Pairage	Nombre de paires de bases dans une fenêtre glissante de longueur 3, 5 ou 7. Présence d'une position avec un 5, 10 ou 20 mer complètement apparié.
Compositions nucléotides	Pourcentage de A, C, G et U
Types de paires de base	Pourcentage de paires de bases dans le duplexe : G-C, G-U et A-U
Boucle	miRNA dans la bouche de l'épingle
Distance du miRNA à la boucle ou au départ et fin de l'épingle	Situation du miRNA entre la boucle et le début et fin de l'épingle
G+C	Contenu GC du miRNA
MFE duplex	Energie minimale du repliement du duplexe miRNA-miRNA*
Longueur	Longueur du miRNA

Prédiction par apprentissage automatique

Machine à vecteurs de support (SVM)

↓ ↓ ↓
 U U C G U
 .() (() () . (() .()
 Triplet element: continuous three structures with middle nucleotide. Taking “C” as the same as “U”

32 triplet element features --- 32-dimension vector:

($\mathbf{u}_{\cdot \cdot \cdot \cdot \cdot}$, $\mathbf{G}_{\cdot \cdot \cdot \cdot \cdot}$, $\mathbf{G}_{\cdot \cdot \cdot \cdot \cdot}$, \dots)

↓

Counting the appearances of the triplet elements:

$$(12, 4, 3, 1, 2, 0, 0, 0, 10, 1, \dots)$$

↓

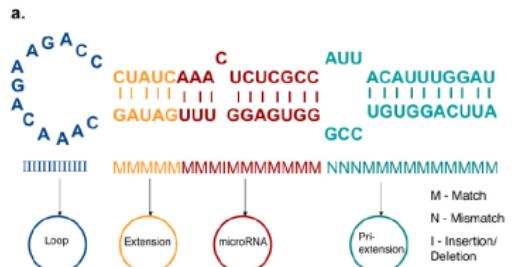
Normalizing the triplet element count vector:

$$(0.1846, 0.0615, 0.0462, 0.0154, 0.0308, 0, 0, 0, 0.1538, 0.0154, \dots)$$

triplet-SVM (Xue et al., 2005)

Prédiction par apprentissage automatique

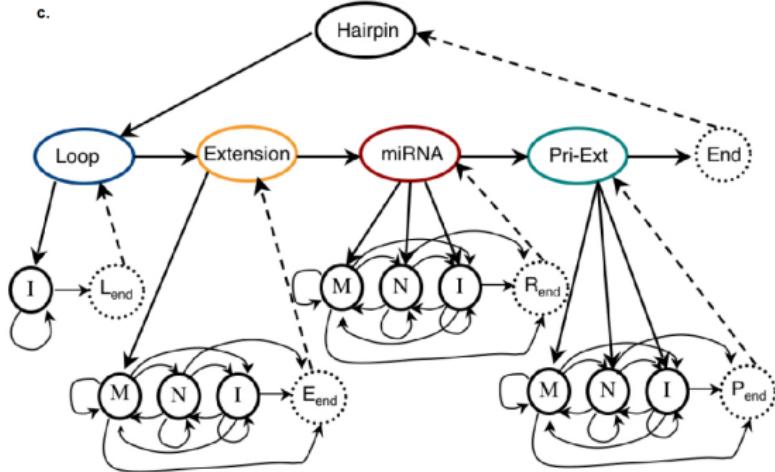
Modèle de Markov caché (HMM)



b.

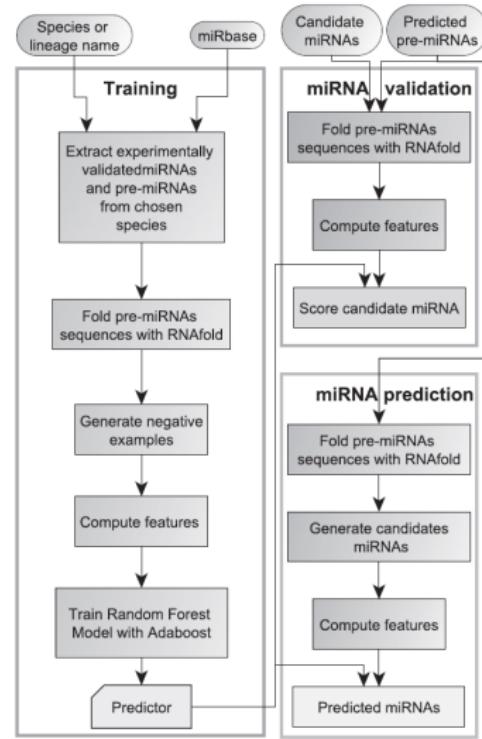
```

CCAGAAGACAAACCUAUCAAACUCUCGCCAUUACAUUUGGAU
-----GAUAGUUU-GGAGUGGGCCUGUGGACUUA
LLLLLLLLLLLLEEEEERRRRRRRRRRRPPPPPPPPPPPP
  
```



HHMMIR (Kadri et al., 2009)

Prédiction par apprentissage automatique

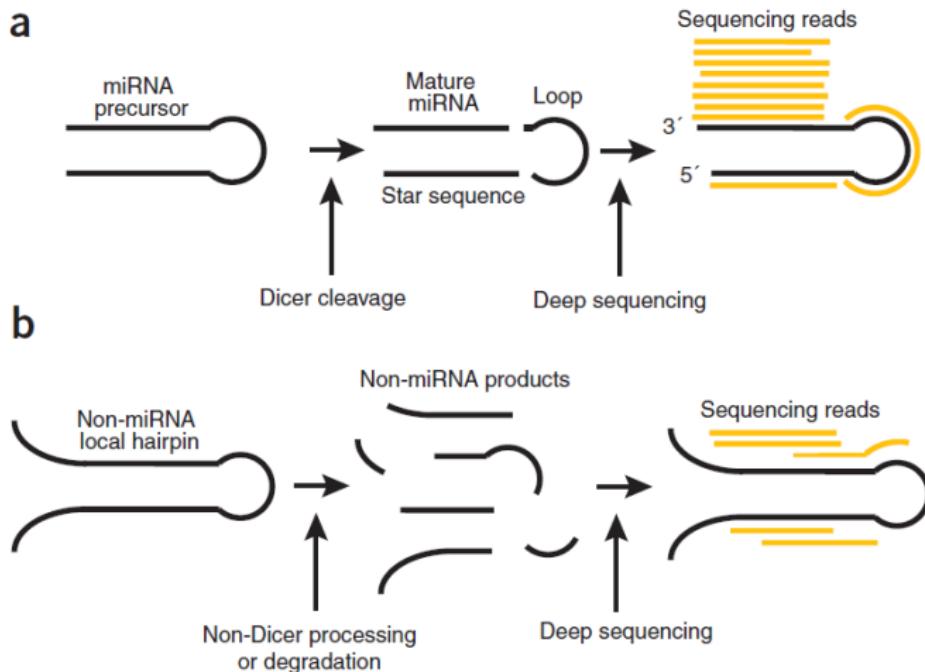


Random Forest avec Adaboost

Prédiction basée sur le duplex microARN : ARNm

- Alignement des régions 3'-UTR de plusieurs génomes conservés
- Identification des motifs courts et conservés
- Extraction des séquences en épingle à cheveux contenant le complémentaire de ces motifs
- (Xie *et al.*, 2005)

Approches mixtes



mirDeep (Friedlander et al., 2008)

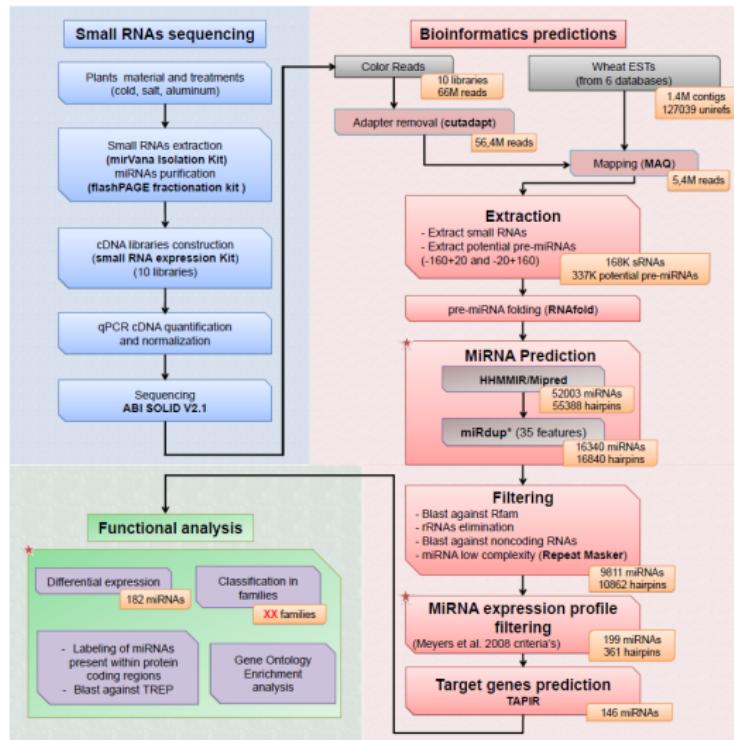
Prédiction des gènes cibles

- Un microARN n'est fonctionnel que s'il cible un gène
- Animaux
 - Cible la région 3'-UTR de l'ARNm
 - Complémentarité microARN :ARNm faible (mismatches, gaps, couples G :U)
- Plantes
 - Cible n'importe quelle région de l'ARNm
 - Complémentarité microARN :ARNm presque parfaite

Prédiction des gènes cibles

- Approches basées sur la conservation du seed 5' (nd 2 à nd 8) et la stabilité thermodynamique du microARN :ARNm
 - TargetScan (Lewis *et al.*, 2005)
 - picTar (Krek *et al.*, 2005)
 - DIANA-microT (Kiriakidou *et al.*, 2004)
- Approches basées sur l'apprentissage automatique
 - Programmation génétique : TargetBoost (Saetrom *et al.*, 2005)
 - SVM : MiTarget (Kim *et al.*, 2006)
- Services web
 - TAPIR (Bonnet *et al.*, 2010)
 - psRNATarget (Dai *et al.*, 2011)

Pipeline d'identification à partir de données NGS



(Agharbaoui et al., 2015)

Rfam

[HOME](#) | [SEARCH](#) | [BROWSE](#) | [FTP](#) | [BLOG](#) | [HELP](#)

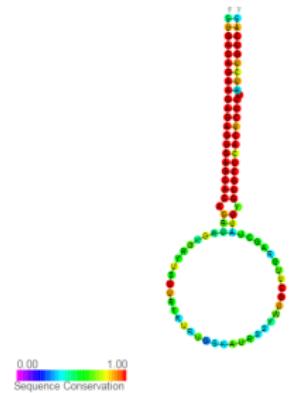
Family: **mir-156 (RF00073)**

Description: *mir-156 microRNA precursor*[Summary](#)
[Sequences](#)
[Alignment](#)[Secondary structure](#)
[Species](#)
[Trees](#)
[Structures](#)
[Motif matches](#)
[Database references](#)
[Curation](#)[Jump to...](#)enter ID/acc [Go](#)

Secondary structure

This section shows a variety of different secondary structure representations for this family. [More...](#)

You can view the secondary structure of the family using the [VARNA](#) applet. You can see more information about VARNA itself [here](#).

[seqcons](#) [bpcons](#) [cor](#) [ent](#) [maxcm](#) [norm](#) [rchie](#)ACAO & ECA
634 sequences

41 species

8 structures

miRBase

miRBase

MANCHESTER IOM

Home | Search | Browse | Help | Download | Blog | Submit | Search

Search miRBase

By miRNA identifier or keyword
 Enter a miRNA accession, name or keyword:

By genomic location
 Select organism, chromosome and start and end coordinates. Leave the start/end boxes blank to retrieve all miRNAs on the selected chromosome.
 Choose species: Chr: Start: End:

For clusters
 Select organism and the desired inter-miRNA distance.
 Choose species: Inter-miRNA distance:

By tissue expression
 Select organism and tissue.
 Choose species: Select tissue: Get experiments

By sequence
Single sequence searches:
 Paste a sequence here to search for similarity with miRBase miRNA sequences (**max size 1000 nts**). You can choose to search against hairpin precursor sequences or mature miRNAs. This search may take a few minutes. Please note: this facility is designed to search for homologs of microRNA sequences, **not to predict their target sites**. For target site prediction, please use [the available bespoke tools](#).
 Search sequences: Search method: Mature miRNAs BLASTN
 Choose BLASTN to search for a miRNA homolog in a longer sequence. SSEARCH is useful for finding a short sequence within the library of miRNAs (for instance, find a short motif in a miRNA or precursor stem-loop, or find mature sequences that are related to your query).
 E-value cutoff:
 Maximum no. of hits:
 Show results only from specific organisms:
 human mouse worm fly Arabidopsis
 or choose a taxonomic classification: No species filter

Or: Select the sequence file you wish to use
 Choisissez un fichier | Aucun fichier choisi
 Search mRNAs | Réinitialiser | Example

Wheat microRNA Portal

Wheat MicroRNA Portal

| Home/Search | Advanced Search | Data | Tools | Citations

Find Search for wheat microRNAs, target genes, ests or microRNA precursors using sequences or keywords

Welcome to the UQAM Wheat microRNA Portal !
 You will find inside a total of **10 small RNA libraries** under different abiotic stress.
 You will also find inside a compilation of published wheat microRNAs from different studies (**5468 wheat microRNAs** - September 2014).
 Finally, you can verify your putative wheat, cereal and plant microRNAs using the web server version of **Mirdeep v1.2** and **MirCheck v1.0**.

Libraries
 View the microRNAs in the investigated wheat libraries
 Aerial tissues from plants in vegetative phase under normal conditions → Go →

Experimental conditions
 View the microRNAs by experimental conditions
 Vernalisation response in winter wheat → Go →

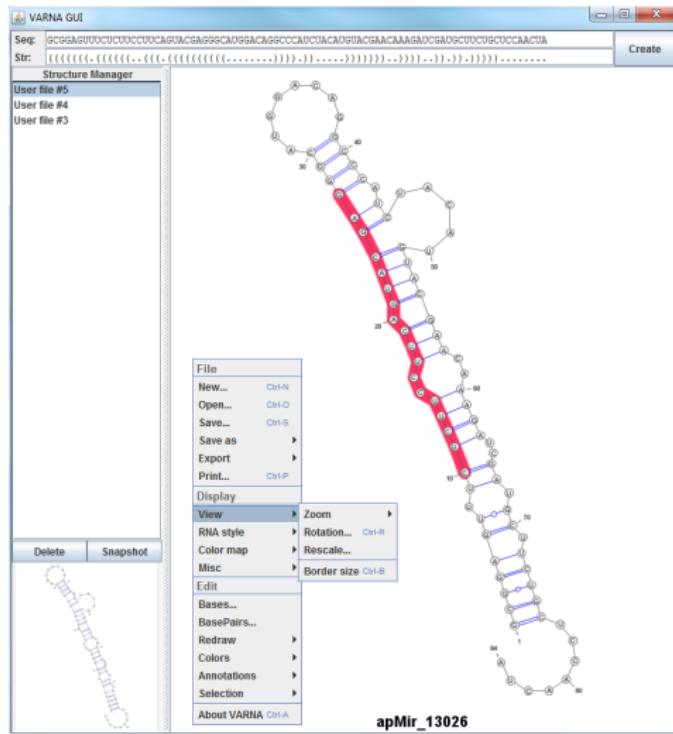
Differential microRNA expression
 Searching for differential miRNA expression between different experimental conditions
 e.g. MicroRNAs upregulated in vernalisation response but downregulated in aluminium response

Searching for sequence identifiers
 You can search for microRNAs by ID or Target ID
 e.g. apMir_20502, Glutathione peroxidase, or gene ontology GO:0008152

Quick link
 All the predicted wheat microRNAs and associated hairpins **Total 2239**
 Putative miRNAs found in this database **Total 1369**
 All the conserved wheat microRNAs found in the database **Total 267**
 Clusters of microRNAs found in this database **Total 345**
 All published wheat microRNAs included in this database **Total 5468**

Statistics (version 1.0)
 Latest update: September 2014
 Number of validated microRNAs: **1369**
 Number of associated targets: **6 841**
 Number of associated ests: **4 446**
 Number of small RNAs associated with ests: **168 834**

Visualisation



<http://varna.lri.fr/>



Visualisation

Représentation textuelle

```

2522 #####
2523 #####
2524 # CUAUCCCCUCCAUUUCAAAAUAUGAUGACUCAACUUUGUACUAAUUUAGUACAAGGUUGGCCAACUCAUUUJGGGAGCGGGAGGUACUUAAGA
2525 & (((((((((+ ((((((((((((. (. ((((((((((((....)))))))))))))))))))))))))))).... apPre_20260 TC421300;gi|20100526|gb|BJ276947
2526 > CAUCUCAUUUJGGGAGCGGGG 100,155,21,40,57,22,84,15,15,33 [542] apMir_20989
2527 % UCCCUCCAUCCCCAAAUAAGAU 1,6,2,0,1,1,3,0,0,0 [14] apMir_39667
2528 % CCCUCCAUCCCCAAAUAAGUG 13,5,1,1,0,1,6,1,3,1 [32] apMir_20989
2529 % CCUCCAUCCCCAAAUAAGAUGA 3,4,1,1,0,5,2,0,0,0 [16] apMir_21975
2530 % UUCCAAAUAUGAUGACUCAACUU 1,0,0,0,0,1,1,2,0,0 [5] apMir_79710
2531 % UUCCAAAUAUGAUGACUCAAC 9,15,4,8,1,8,7,4,3,3 [62] apMir_19545
2532 % UCCAAAUAUGAUGACUCAACU 6,7,0,4,0,17,13,12,11,8 [78] apMir_20102
2533 % CCAAAUAUGAUGACUCAACUU 1,2,0,3,0,3,0,4,1,1 [15] apMir_22687
2534 % AAAUAGAUGACUCAACUUUGU 1,4,0,0,0,0,1,0,0,0 [6] apMir_21090
2535 % AAUAGAUGACUCAACUUUGUACUA 1,1,3,1,0,1,1,1,1,0 [10] apMir_42708 apMir_20989
2536 % AUAGAUGACUCAACUUUGUACUAA 5,1,0,1,0,4,3,2,0,2 [18] apMir_74988
2537 % AUAGAUGACUCAACUUUGUACUA 0,2,0,0,0,0,3,0,0,0,2 [7] apMir_41758
2538 % GGCACAUCAUUUJGGGAGCGGA 1,0,1,0,2,0,2,1,0,0 [5] apMir_20990 apMir_20989
2539 % GCCAUCAUUUJGGGAGCGGA 6,1,0,4,2,1,2,1,0,4 [21] apMir_21814 apMir_20989
2540 % CAUCAUUUUUJGGGAGCGGG 17,15,3,2,4,4,7,1,2,3 [58] apMir_24350 apMir_20989
2541 % CAUCAUUUUUJGGGAGCGGA 1,3,0,2,0,1,1,2,3,0 [13] apMir_34735
2542 % CAUCAUUUUUJGGGAGCGGG 5,1,0,1,2,2,0,0,0,0 [11] apMir_58314
2543 % CAUCAUUUUUJGGGAGCGGG 8,10,2,0,4,0,11,1,3,2 [41] apMir_18337
2544 % CAUCAUUUUUJGGGAGCGGA 1,0,0,0,2,2,4,1,0,1 [9] apMir_13964
2545 % AUCUAUUUUJGGGAGCGGG 2,5,0,0,0,1,0,0,1,0 [9] apMir_23921
2546 % AUCUAUUUUJGGGAGCGGG 2,7,0,0,1,0,3,0,1,0 [14] apMir_38531
2547 % AUCUAUUUUJGGGAGCGGG 6,24,4,2,6,5,1,2,3,0 [53] apMir_21967
2548 #####
2549 #####

```

(Agharbaoui et al., 2015)