

SMALLA

A toolbox for managing libraries of smallRNA sequences

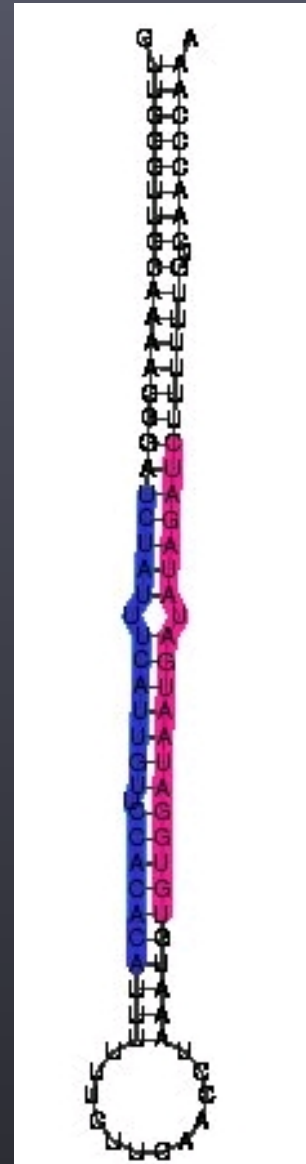
Erika Sallet, Christine Lelandais,
Martin Crespi, Jérôme Gouzy

Contexte biologique

- ▶ **Projet MIRMED**
- ▶ **Objectif** : étudier la diversité des smallRNA/miRNA exprimés chez une plante dans différents tissus et dans différentes conditions, en particulier lors :
 - d'une interaction symbiotique avec une bactérie ou un champignon
 - d'une interaction avec un bactérie pathogène
 - d'un stress
- ▶ **Méthode** : construire et séquencer une vingtaine de banques de smallRNA
 - Technologies 454 et solexa

Les miRNA chez les plantes

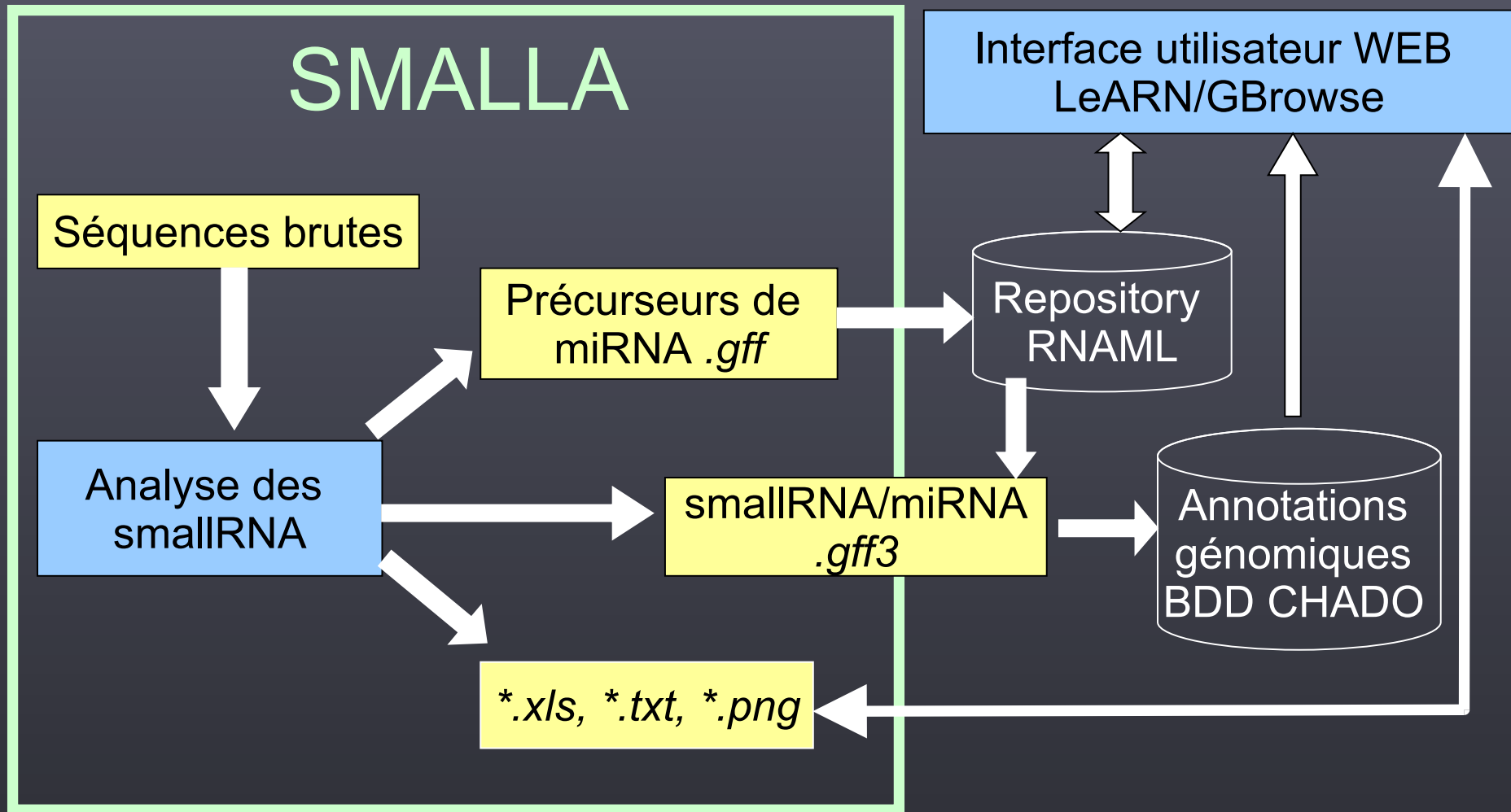
- ▶ smallRNA : séquences de 18 à 24 nt
- ▶ ARN régulateurs impliqués dans de nombreux mécanismes cellulaires
- ▶ 2 familles principales : les siRNA (short interfering RNA) et les miRNA (microRNA)
- ▶ Les miRNA
 - Séquence mature d'environ 21 nt
 - Structure en tige boucle du précurseur
 - Régulateur de l'expression des gènes
 - ▶ Inhibition de la traduction
 - ▶ Dégradation d'ARNm cible



Besoins

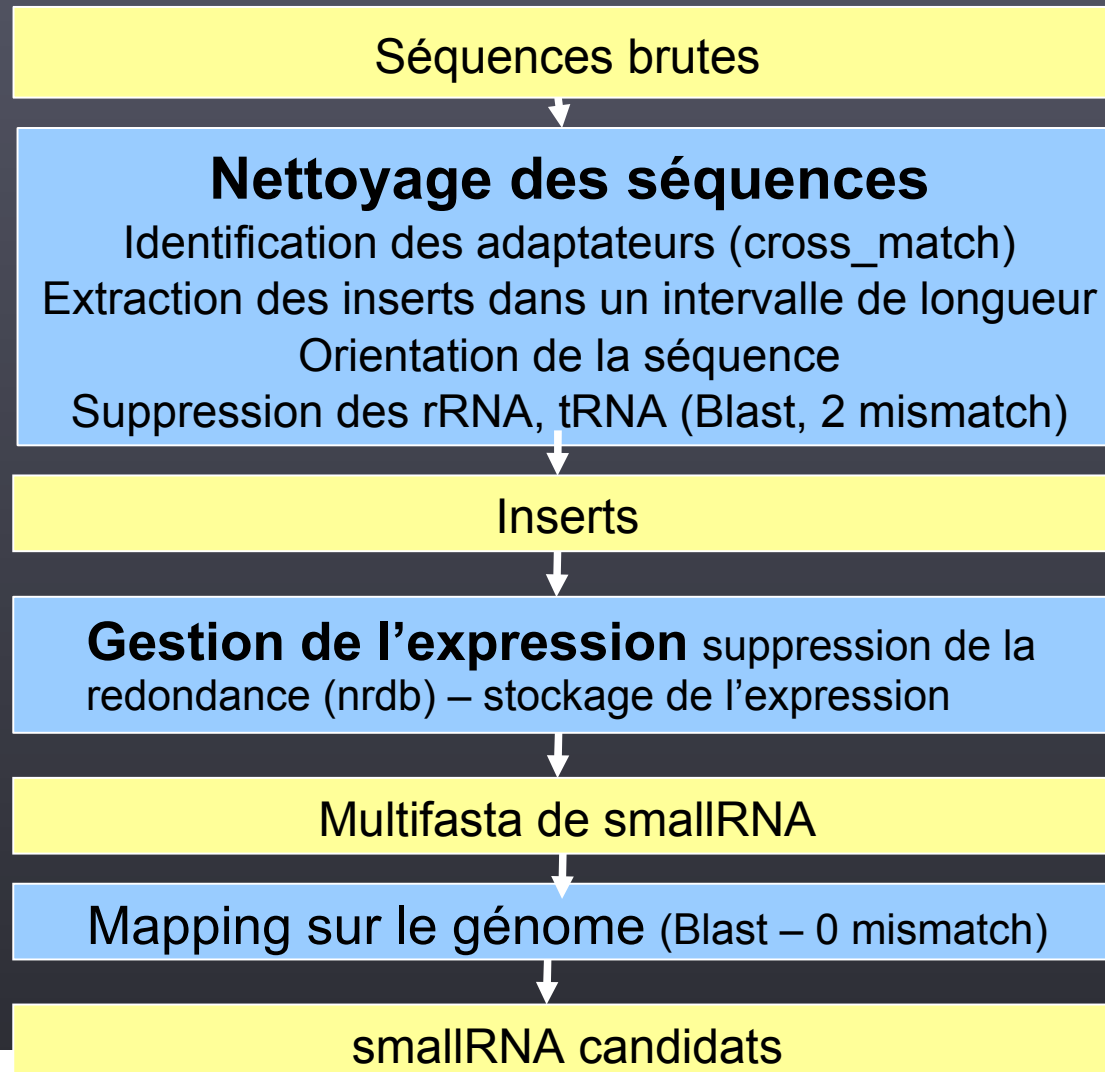
- ▶ Cartographier globalement les smallRNA sur le génome
- ▶ Caractériser les miRNA
 - Détecter les précurseurs de miRNA sur le génome
 - Annoter les précurseurs
 - ▶ miRNA:miRNA*, autres
 - Détecter les ARNm cibles potentiels

Architecture bioinformatique



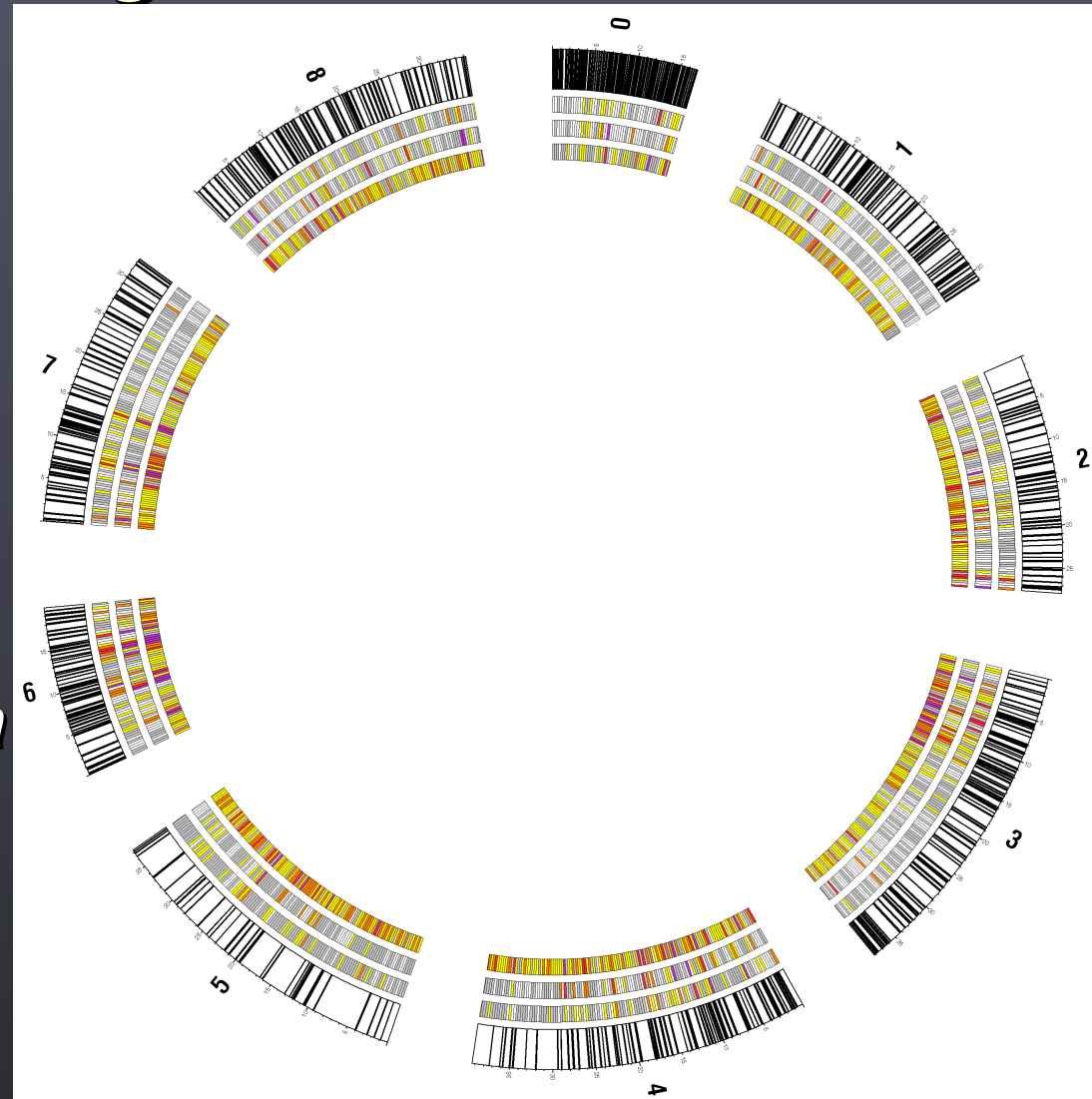
LeARN: a platform for detecting, clustering and annotating non-coding RNAs Céline Noirot, Christine Gaspin, Thomas Schiex, and Jérôme Gouzy BMC Bioinformatics. 2008; 9: 21

Détection des smallRNA dans les résultats de séquençage 454



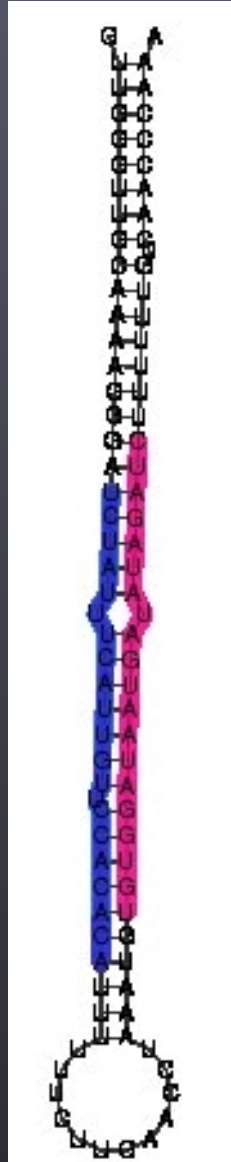
Cartographie globale des smallRNA sur le génome

- ▶ Logiciel Circos
<http://mkweb.bcgsc.ca/circos/>
- ▶ Exemple : mapping des smallRNA de longueur 21, 22 et 24 nt sur les chromosomes de *Medicago truncatula*



Recherche de précurseurs de miRNA

- ▶ Conservation des 20-22 mers mappés sur le génome
- ▶ Recherche de structure tige-boucle avec MiRfold (B. Billoud et al. Biochimie 2005)
- ▶ Filtre sur l'énergie libre et sur la structure du duplex (Jones-Rhoades et al. 2006, Meyers et al. 2008)
- ▶ Annotation des précurseurs avec les autres smallRNA
- ▶ Classification des précurseurs dépendante de l'annotation



Interface LeARN spécifique aux smallRNA

Search:
 Classification Comparative Screening Clusters Targets Genome Blast Credits

Class I Ir Class II Class IIIa Class IIIb Class IV Class V

Class I: mir and mir* only
 (reads corresponding to the seed are more abundant than other reads)

[\[download file\]](#)

Count	RNA ID	seed/mature ID	sequence	dG	length	seed length	pattern	seed reads (N+R+RS)	maximum occurrence of other reads (N+R+RS)
1	MtrV2Chr0_r154	EVSUSMY01BAZRJ	GGCAUCCACGUGUGCUUCACCG	-238	366	22	R(0) N(1) RS(0)	1	1
2	MtrV2Chr0_r283	EVSUSMY01EWYCV	GUUGUACAUAUAUCACUACUCU	-82	212	22	R(0) N(1) RS(0)	1	1
3	MtrV2Chr0_r611	EVSUSMY01AHBBP	UUGACCGUUCAUAUGAACCCUG	-67	110	22	R(0) N(1) RS(0)	1	1
4	MtrV2Chr0_r743	E8W526M03FR2DV	UUAAAUGAAUGAACCUAGAAU	-72	116	21	R(1) N(0) RS(0)	1	1
5	MtrV2Chr0_r764	EVSUSMY01DPQNA	AAAUGCUUGAGUCAUGUUGUU	-191	255	21	R(0) N(1) RS(0)	1	1
6	MtrV2Chr0_r1131	E8W526M04H11ZQ	GUGCCUGUCUGUCCGCUCUU	-37	52	20	R(1) N(0) RS(0)	1	1
7	MtrV2Chr0_r1186	EVSUSMY01BQ45C	AUCUAAAGGUGAUUAUUGUGCC	-95	197	22	R(0) N(1) RS(0)	1	1
8	MtrV2Chr0_r1365	E8W526M03HGO6D	CUAGUACAUGUUUAAGUAAU	-30	86	20	R(1) N(0) RS(0)	1	1

Fiche descriptive d'un précurseur

Structure :

Structure computed by: maploteval-1.6.2

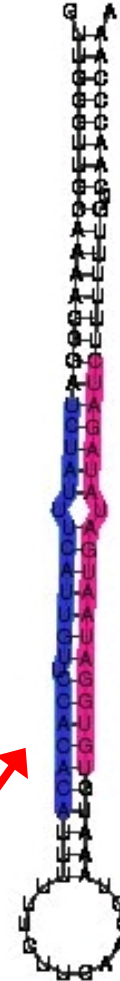
Free-energy: -50.3

Parenthesis structure:

```
.....  
.....
```

Segment annotations:

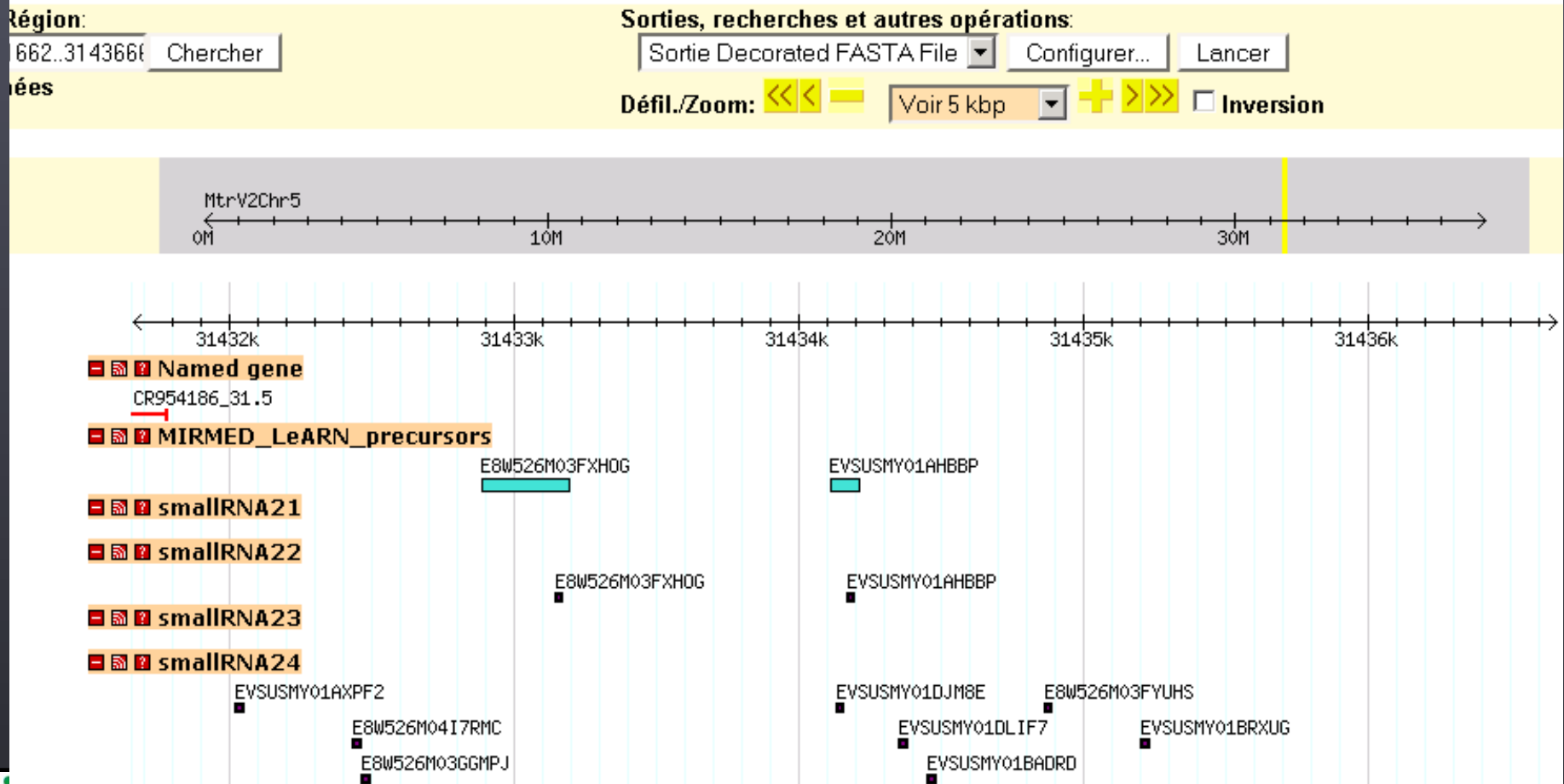
- 60 - 79 **Plot**
 - Description: E8W526M04I61WC
 - Annotator :
- 19 - 37 **Plot**
 - Description: E4D3Z3Y01DTIX6
 - Annotator :
- 19 - 39 **Plot**
 - Description: E8W526M03G40CH
 - Annotator :



Plot all 21bp annotations (if any) **Plot all annotations**

Visualisation de l'annotation du génome

► Gbrowse



Analyse comparative des miRNA

► Rechercher

- des séquences de miRNA matures sur le génome d'autres espèces
- des similarités avec miRBase V12

Search: Classification Comparative Screening Clusters Targets Genome Blast Credits

0 mismatch 1 mismatch 3 mismatches

Comparative analysis: 0 mismatch allowed on Mt, 3 mismatches allowed on other species

[download file](#)

Count	ID	Mt	Lj	Gm	Vv	Pt	At	Os	#reads	co-occurrence pattern value	expression pattern	
1	EVSUSMY01AGM7N	6	11	26	9	16	9	17	7349	127	N(1689) R(5119) RS(541)	vvi-miR166h
2	EVSUSMY01ET8TO	1	4	14	18	11	4	4	2887	127	N(9) R(2865) RS(13)	pta-miR396
3	EVSUSMY01DYWGS	1	11	26	9	16	7	10	929	127	N(238) R(655) RS(36)	vvi-miR166b
4	EVSUSMY01EWHSN	8	11	27	9	16	7	11	577	127	N(63) R(498) RS(16)	vvi-miR166h
5	EVSUSMY01EWUVD	2	0	5	1	2	1	1	565	95	N(4) R(556) RS(5)	-
6	EVSUSMY01ESUFH	1	3	8	15	3	0	6	307	125	N(1) R(305) RS(1)	-
7	EVSUSMY01AS9C3	3	8	20	9	14	4	11	291	127	N(20) R(260) RS(11)	vvi-miR160f
8	EVSUSMY01D9R11	2	1	2	1	1	2	3	247	127	N(25) R(217) RS(5)	vvi-miR168
9	EVSUSMY01AS2AA	1	6	15	5	8	5	11	181	127	N(93) R(84) RS(4)	vvi-miR167e
104	E8W526M03FVMG6	141	0	0	0	0	0	0	2	64	N(0) R(2) RS(0)	-
105	EVSUSMY01DLS17	1	0	1	0	1	0	0	2	84	N(2) R(0) RS(0)	-
106	EVSUSMY01DUXIB	15	0	31	0	0	0	0	2	80	N(2) R(0) RS(0)	-

Recherche des cibles des miRNA

- ▶ Méthode basée sur MiRanda (Enright A et al, Genome Biology, 2003)
- ▶ Filtrage des résultats en fonction de la longueur de l'alignement et des appariements (Jones-Rhoades and Bartel, 2004)

Search: <input type="text"/> Classification Comparative Screening Clusters Targets Genome Blast Credits			
			Full Summary
Target prediction using miranda 1.9 (alignment spanning 21bp)			
download file			
target	Miranda v1.9 score		Penalty score
E4D3Z3Y01BHN7L vs MTGI9-TC112546	Forward: Score: 157.000000 Q:1 to 22 R:639 to 659 Align Len (21) (85.71%) (90.48%) Query: 3' ACCTTAACCACCTTATCCATT 5' : Ref: 5' TGGAGCTGGTGGGAATAGGCAA 3' Energy: -29.709999 kCal/Mol	2.5	MTGI9-TC112546 Leucine-rich repeat MTGI9-TC112546 disease res
E4D3Z3Y01BHN7L vs MTGI9-TC112554	Forward: Score: 164.000000 Q:1 to 22 R:645 to 665 Align Len (21) (90.48%) (95.24%) Query: 3' ACCTTAACCACCTTATCCATT 5' : Ref: 5' TGGGATTGGTGGGAATAGGGAA 3' Energy: -32.389999 kCal/Mol	1.5	MTGI9-TC112554 Toll-like receptor MTGI9-TC112554 disease res
E4D3Z3Y01BHN7L vs MTGI9-TC112724	Forward: Score: 164.000000 Q:1 to 22 R:651 to 671 Align Len (21) (90.48%) (95.24%) Query: 3' ACCTTAACCACCTTATCCATT 5' : Ref: 5' TGGGATTGGTGGGAATAGGGAA 3'	1.5	MTGI9-TC112724 NB-ARC

Exploration

- ▶ Recherche par mot-clé (ID, séquence, ...)
- ▶ Recherche par Blast

Search: NBS-LRR

Classification

Comparative

Searching NBS-LRR in ...

Target prediction using miranda 1.9 (alignment spanning 21

	#MIRMED_ID	target	alignment_score	free_energy	pena
1	E4D3Z3Y01BHN7L	MTGI9-TC112546	157.00	-29.71	2.5

NCBI BLAST BLAST Entrez ?

Choose program to use and database to search:

Program Database

Enter sequence below in [FASTA](#) format

Or load it from disk

Set subsequence: From To

The query sequence is [filtered](#) for low complexity regions by default.
[Filter](#) Low complexity Mask for lookup table only

[Expect](#) [Matrix](#) Perform ungapped alignment

[Query Genetic Codes \(blastx only\)](#)

[Database Genetic Codes \(tblast\[nx\] only\)](#)

AT5G17680.1

Les difficultés

- ▶ Nettoyage des données 454
 - Parfois absence d'un adaptateur 5' ou 3'
 - Gestion des séquences reverse complémentées
 - Bruit dans les séquences (produits de dégradation)
- ▶ Processus long
 - Beaucoup de données
 - De nombreuses étapes relativement lentes (mirfold, miranda, intégration des données dans LeARN)
 - Solutions :
 - ▶ Filtrage précoce
 - ▶ Parallélisation

Conclusion

- ▶ SMALLA, boîte à outils pour :
 - Nettoyer les banques de séquences de smallRNA
 - Détecter des précurseurs de miRNA et les annoter
 - Rechercher les ARNm cibles de miRNA
 - Comparer les molécules avec d'autres miRNA ou génomes
- ▶ En cours
 - Analyse de l'expression des smallRNA
- ▶ Release de LeARN intégrant SMALLA disponible début juin sur le site <http://symbiose.toulouse.inra.fr/LeARN>

Remerciements

- ▶ Plate-forme bioinformatique de la Génopole Toulouse Midi-Pyrénées
 - 9 mois de CDD
 - Cluster de calcul
- ▶ Emeline Deleury INRA Sophia
 - CHADO
- ▶ Les autres participants au projet MIRMED
 - Andréas Niebel, Christophe Roux, Cécile Ben, Karine Gallardo, etc...