SMALLA
a toolbox for managing libraries of smallRNA sequences

Erika Sallet[1,2] Christine Lelandais[3] Martin Crespi[3] Jérôme Gouzy[1]
[1] Laboratoire des Interactions Plantes Micro-organismes (INRA/CNRS), [2] Plateforme Bioinformatique du Génopole Toulouse Midi-Pyrénées, [3] Institut des Sciences du Végétal, CNRS
Erika.Sallet@toulouse.inra.fr, Jerome.Gouzy@toulouse.inra.fr

SMALLA est disponible sous licence CECILL2 et intégré au package LeARN téléchargeable à partir de http://symbiose.toulouse.inra.fr/LeARN (version >= 1.5).
SMALLA a été réalisé dans le cadre du projet MIRMED http://medicago.toulouse.inra.fr/Mt/RNA/MIRMED/LeARN/cgi-bin/learn.cgi (site accessible après publication)

Contexte

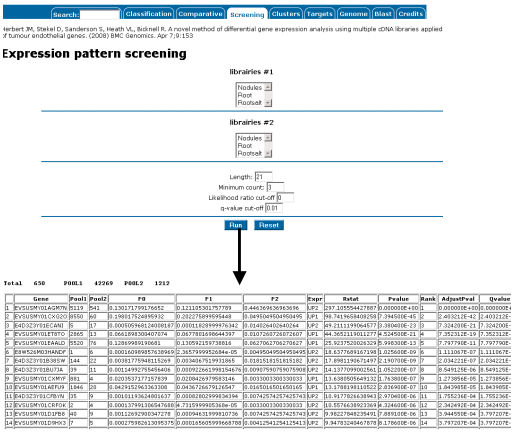
Les technologies de séquençage actuelles rendent possible l'analyse de l'expression des molécules régulatrices que sont les smallRNA (20-25nt). Ces analyses produisent des dizaines de millions de lectures qu'il est nécessaire de nettoyer et de filtrer pour séparer les molécules potentiellement actives des produits de dégradation. Cette étape accomplie, il devient possible de cartographier ces séquences sur le génome d'intérêt pour identifier les gènes ayant potentiellement généré les précurseurs de ces molécules. Analyser la redondance intra banque permet ensuite de comparer les profils d'expression entre différentes conditions.
Parmi l'ensemble des smallRNA, les miRNA jouent un rôle majeur dans la régulation de l'expression ou de la traduction des ARN messagers de gènes codant pour des protéines. Les précurseurs des miRNA ont une structure secondaire particulière qu'il est possible de caractériser (1), de classifier et d'annoter en utilisant les séquences générées (2) pour finalement compléter l'analyse en identifiant les ARN messagers ciblés par les miRNA (3,4).

Outils communs à toutes les classes de smallRNA

SMALLA est une boite à outil bioinformatique qui permet de gérer une grande partie du processus d'analyse des smallRNA et plus particulièrement des miRNA. SMALLA est basé sur l'architecture de LeARN (5) qui est une plateforme destinée à l'annotation des ARN non codant pour des protéines (ncRNA) à l'échelle génomique et qui propose de nombreuses fonctionnalités pour visualiser et éditer l'annotation de ncRNA. Ainsi SMALLA peut d'une part être installé localement et paramétré selon l'organisme d'intérêt et les données disponibles, et d'autre part être utilisé pour visualiser et éditer les précurseurs de miRNA.

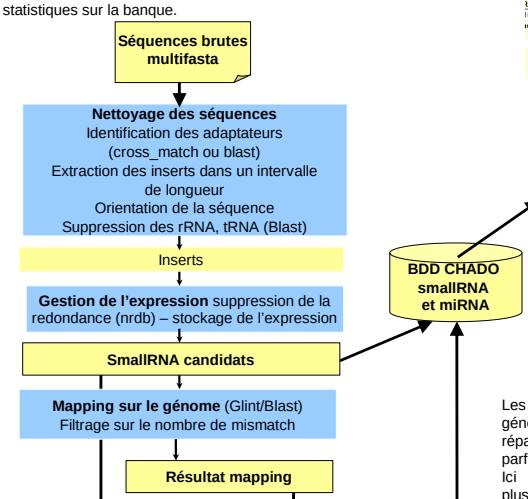
Etude de l'expression des smallRNA dans différentes banques

L'interface utilisateur propose un formulaire web qui permet la recherche de molécules différentiellement exprimées entre différentes conditions



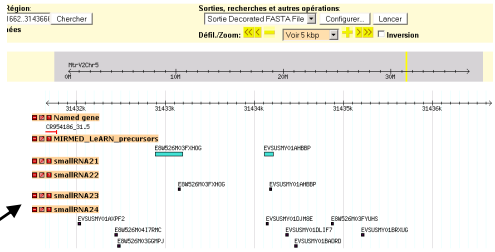
Nettoyage et détection des smallRNA dans les résultats de séquençage

SMALLA propose un ensemble de programmes permettant de nettoyer les séquences produites par les séquenceurs de type 454 ou solexa. Ce processus inclut la génération de statistiques sur la banque.

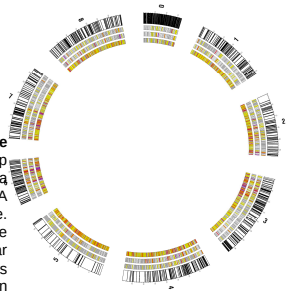


Cartographie des smallRNA sur le génome

Représentation locale. GBrowse permet de visualiser les smallRNA le long de la séquence génomique. Les données sont chargées depuis une base de données CHADO.



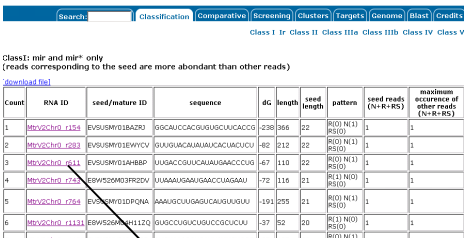
Représentation globale Les graphiques de type heatmap générés par circos (6) représentent la répartition et la densité des smallRNA parfaitement mappés sur le génome. Ici on visualise des smallRNA de plusieurs longueurs (une piste par longueur) sur plusieurs chromosomes (une tranche par chromosome) d'un organisme.

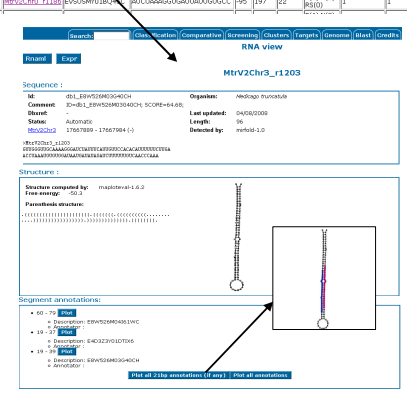


Outils spécifiques aux miRNA

Consultation des précurseurs de miRNA

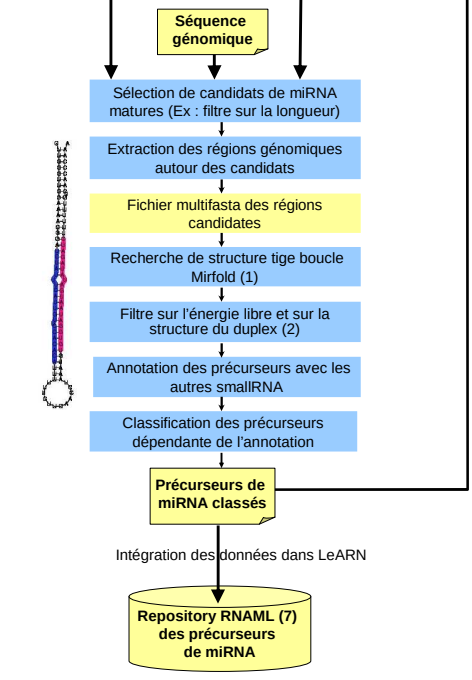
L'interface permet de consulter la classification des précurseurs de miRNA et d'accéder aux informations détaillées sur chaque précurseur (séquence, structure, annotation avec d'autres smallRNA). Les données sont extraites du repository RNAML.





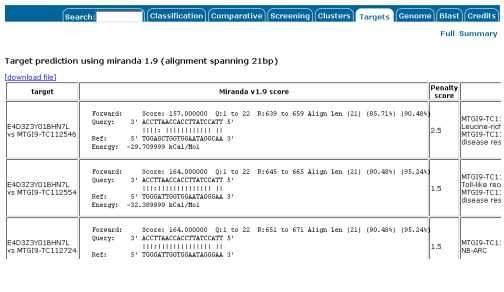
Pipeline de détection de précurseurs de miRNA

Le package contient des programmes qui permettent, sur la base de la cartographie sur le génome de référence, de déterminer les structures secondaires correspondant à des précurseurs de miRNA.



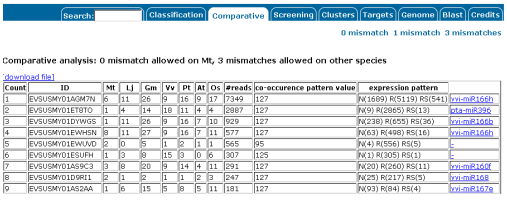
Recherche des cibles des miRNA

SMALLA contient un programme pour détecter les ARN messagers qui pourraient être considérés comme des cibles potentielles de miRNA. Ce programme filtre les résultats produits par MiRanda (3) selon les critères définis par Jones-Rhoades *et al.* (4) basés sur la longueur de l'alignement et sur les appariements.



Analyse comparative des miRNA matures

Nous proposons un programme qui permet une première analyse phylogénomique en comparant les séquences de miRNA matures à la banque de référence miRBase (8). L'interface propose également de visualiser les résultats de mapping des miRNA sur les génomes d'autres espèces afin de déterminer les miRNA spécifiques à l'espèce d'intérêt ou à la famille de l'espèce.



Références

1. Billoud, B., De Paepe, R., Baulcombe, D. and Boccara, M. (2005) *Biochimie*, **87,** 905-910.
2. Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J. *et al.* (2008) *Plant Cell.*
3. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) *PLoS Biol*, **2,** e363.
4. Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. (2006) *Annu Rev Plant Biol*, **57,** 19-53.
5. Céline Noirot, Christine Gaspin, Thomas Schiex, and Jérôme Gouzy *BMC Bioinformatics*. 2008; 9: 21
6. http://mkweb.bcgsc.ca/circos/
7. A. Waugh et al. *RNA*. 2002 Jun;8(6):707-17.
8. Griffiths-Jones, S., Saini, H.K., van Dongen, S. and Enright, A.J. (2008) *Nucleic Acids Res*, **36,** D154-158