

## Proposition de sujet de stage

### Développement d'une stratégie pour l'identification de petits ARNs à partir de données de sRNAseq

Nom de l'entreprise : Institut National de la Recherche Agronomique  
Adresse : INRA - Unité Mathématiques et Informatique Appliquées-Toulouse (MIA-T).  
Chemin de Borde Rouge. BP 52627 - 31326 Castanet-Tolosan cedex  
Personnes à contacter : Christine Gaspin ou Matthias Zytnicki  
Téléphone : 05 61 28 52 82 / 05 61 28 50 71

Contact email : [christine.gaspin@toulouse.inra.fr](mailto:christine.gaspin@toulouse.inra.fr), [matthias.zytnicki@toulouse.inra.fr](mailto:matthias.zytnicki@toulouse.inra.fr)

#### Description du sujet

Les petits ARN non codants, sont des acteurs majeurs de la régulation de l'expression des gènes. Ces ARNnc sont notamment impliqués dans la transmission de signaux environnementaux, la maturation de la cellule ou encore certaines pathologies.

Le séquençage haut-débit, et plus particulièrement le sRNA-Seq, est une technique aujourd'hui largement utilisée pour trouver ces petits ARNs. Ce séquençage peut produire des centaines de millions de séquences à un prix abordable.

Un des objectifs principaux du sRNA-Seq est d'identifier certains gènes non-codants, et notamment les micro-ARN (miARN). S'il existe aujourd'hui beaucoup d'outils de recherche de miARN, leur efficacité est souvent limitée.

Par ailleurs, il existe aujourd'hui beaucoup d'espèces dont les miARN ont été relativement caractérisés, et notamment chez les espèces modèles (humain, souris, *D. melanogaster*, *A. thaliana*, etc.). Ces informations, obtenues sur une espèce connue, peuvent aider à rechercher miARN d'une autre espèce, pour peu que les alignements génomiques entre les deux espèces existent.

Dans le cadre de ce stage de M2R, nous exploiterons la conservation de régions non codantes dans les génomes eucaryotes afin d'assigner correctement les loci de production des lectures de sRNA-seq et de proposer des annotations structurales et fonctionnelles valides.

Les développements à réaliser portent sur la conception de modules permettant 1) l'identification des loci ciblés par les lectures courtes, 2) la conservation des loci ciblés au sein d'un ensemble donné de génomes et 3) la caractérisation/classification des profils d'expression au sein de loci conservés.