

Compte rendu TP2

Recherche de pré-microARN et Hybridation ARNmessenger / MicroARN

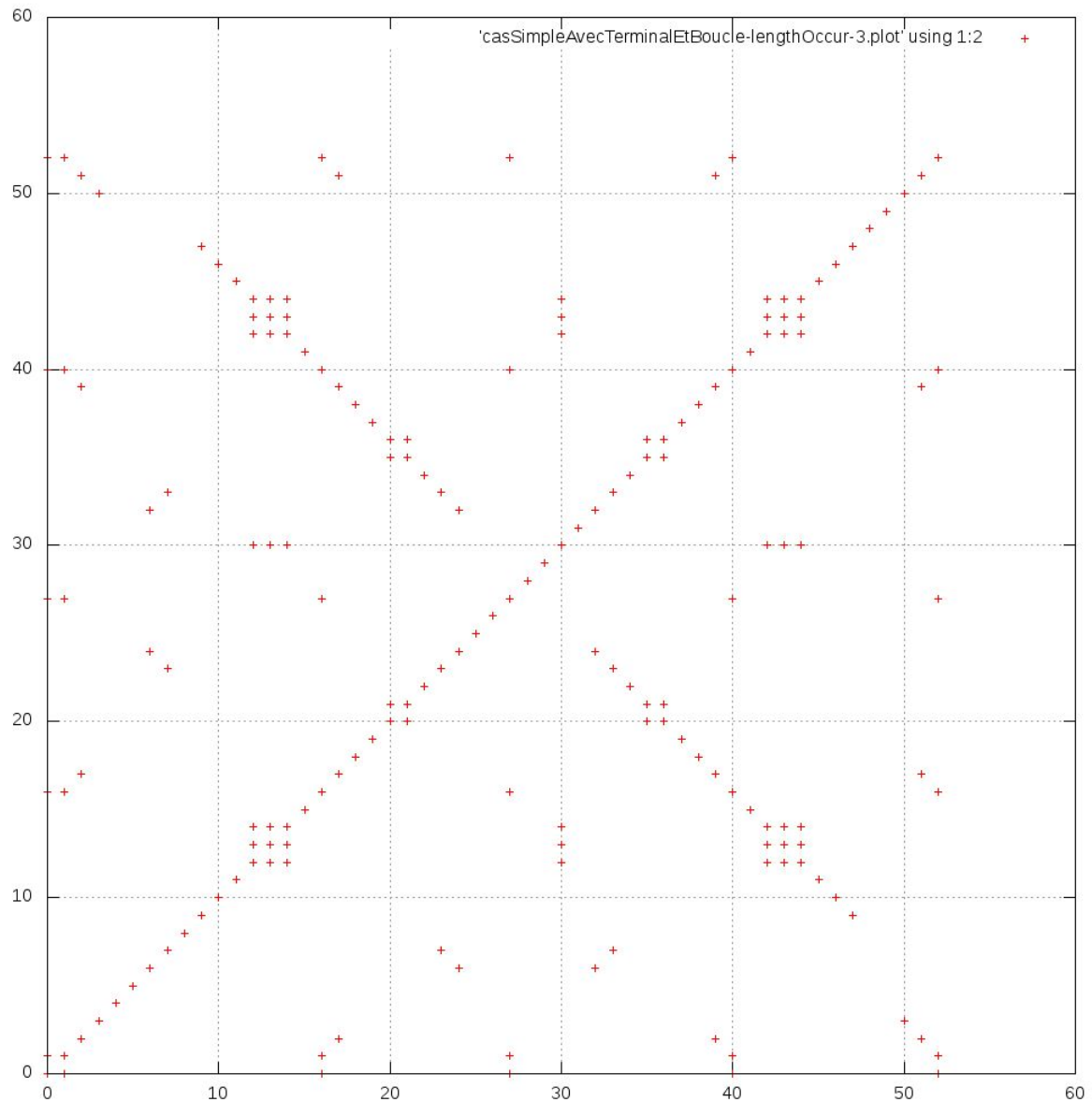
Partie 1

Afin de pouvoir réaliser des appariement de micro ARN à partir de pré-microARN nous avons réalisé quatres exemples (Ces exemples ainsi que les DotPlot associé se trouvent dans le dossier fasta):

- Un cas simple ne contenant pas de boucle terminale et pas de boucle intermédiaires:

casSimple.fasta

ACGUAGAAACCCCGUAAUAUGUGCACAUAUUACGGGGGUUUCUACGU
((



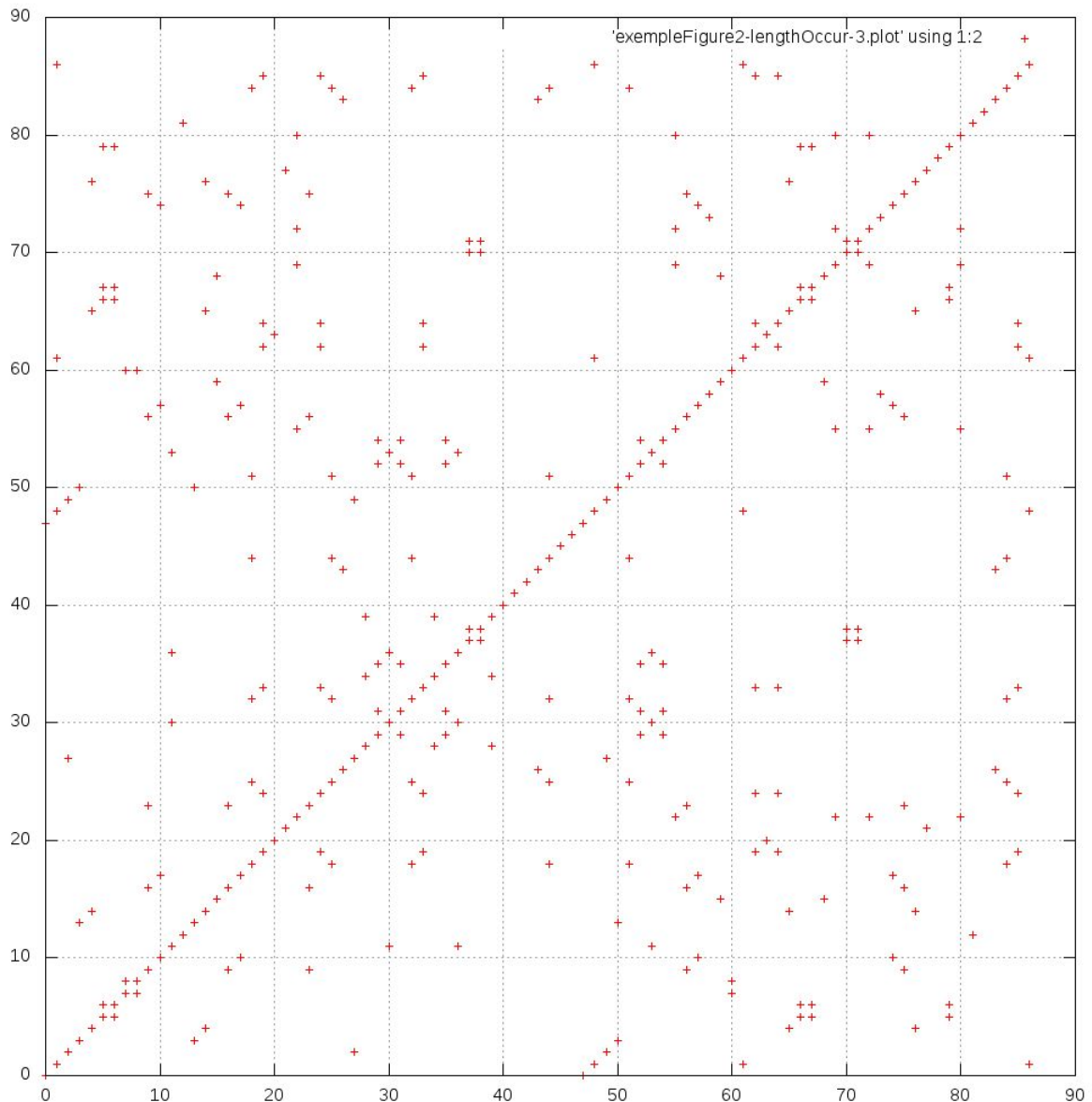
- Un cas complexe avec terminale et boucles diverses:

exempleFigure2.fasta

```

UGCUUCCGGGCCUGUUUCCCUGAGACCUCAAGUGUGAGUGUACUAUUGAUGCUUCACA
CCUGGGCUCUCCGGGUACCAGGACGGUUUGAGCA
(((((((.((((.((((.((((.((((.((((.(.....))))))))).))))).)))
))...))))).)))...))))))

```



Il est possible d'exécuter manuellement la réalisation des dotplot avec la suite de commande suivante:

java -jar {chemin vers l'archive TP1.jar} {chemin vers le fichier fasta} 3 D RC

enfin la suite de commande à exécuter dans gnuplot (exemple pour le fichier casSimple.fasta):

set terminal 'png' size 1000,1000

set grid

set output 'casSimple.png'

plot 'casSimple-lengthOccur-3.plot' using 1:2

exempleFigure2.fasta

.....(((((((((((.((((.(((((((.....(((()))..)))))).))....))))).)

```
./EvaluerAppariement {dossier contenant les fichier micro ARN}
```

Partie 3

Afin de pouvoir évaluer le taux d'hybridation entre un ARN messenger et un micro ARN nous avons décidé de réaliser encore une fois l'algorithme de Needleman et Wunsch en utilisant cette fois comme référentiel d'égalité un caractère unique selon le principe de complémentarité.

La récurrence est réalisé sur toutes les sous séquences de l'ARN messenger commençant par une succession de 7 nucléotides complémentaires avec les 7 premiers nucléotides de la séquence de micro ARN. La taille de la sous séquence d'ARN Messenger doit être au maximum supérieure à 5 nucléotides par rapport à la séquence de micro ARN (on considérera que si sous séquence contient plus de 5 gap elle n'est pas hybridable).

Enfin, pour chaque sous séquence calculée un score est calculé selon la formule suivante:

$$2 * \text{nombre de match} + \text{nombre de substitutions} - \text{nombre de gap}$$

Ainsi si deux sous séquences peuvent être trouvée pour un ARN messenger l'alignement obtenant le score le plus élevé sera gardé.

Les résultats obtenus sont tous présents dans le fichier Résultat_Hybridation.pdf

Les résultats présentent le taux d'hybridation sous deux formes:

- forme naïve : **nombre de match / taille du micro ARN * 100**
- forme complexe : **score calculé / (taille du micro ARN * 2) * 100**

Pour chaque ARN messenger les meilleurs cas d'hybridation sont les suivants:

ARN messagé : ARNmessenger-1.fasta
Micro ARN : hsa-miR-19a_MI_ARN.fasta
Résultat:
UCAAGAUCCAUCGACCUUAAACAGG
XX X X XX---X
AGUUUUGCAUAGUUGCACUAC---A
Pourcentage d'hybridation méthode naïf : 68.10%
Pourcentage d'hybridation méthode calculée : 72.7%

ARN messagé : ARNmessenger-2.fasta
Micro ARN : hsa-miR-18a_MI_ARN.fasta
Résultat:

GUUCCGCCUGGGC-UGUGGGCAUGUC
X X-- XX-X-
UAAGGUGCAUCUAG-UGCAG-A-UAG
Pourcentage d'hybridation méthode naïf : 73.9%
Pourcentage d'hybridation méthode calculée : 69.5%

ARN messagé : ARNmessenger-3.fasta
Micro ARN : hsa-miR-20a_MI_ARN.fasta
Résultat:
AUUUUUAUGACUAUUUAUGC-ACAUU
X X- X-X-
UAAAGUGCUUAUAGU-GCAGG-UAG
Pourcentage d'hybridation méthode naïf : 78.2%
Pourcentage d'hybridation méthode calculée : 71.7%

ARN messagé : ARNmessenger-4.fasta	ARN messagé : ARNmessenger-4.fasta
Micro ARN : hsa-miR-17-5p_MI_ARN.fasta	Micro ARN : hsa-miR-20a_MI_ARN.fasta
Résultat:	Résultat:
GUUUUGUGUG-UGUGUGUGUGU-GUG	GUUUUGUGUG-UGUGUGUGUGU-GUG
X-- X- X-- X	X-- X- X-- X
CAAAGUGCU-UACAG-UGCAG-GUAG	UAAAGUGCU-UAUAG-UGCAG-GUAG
Pourcentage d'hybridation méthode naïf : 73.9%	Pourcentage d'hybridation méthode naïf : 73.9%
Pourcentage d'hybridation méthode calculée : 63.0%	Pourcentage d'hybridation méthode calculée : 63.0%

Comme l'énonce le paradox de simpson il est difficile dire si deux séquences sont hybridable car il faut définir plusieurs variables qui peuvent ne pas s'associer. Ainsi Nous avons décidé de définir deux séquences hybridables lorsque leur pourcentage naïf et calculé dépasse les 60%.

Il est possible d'exécuter manuellement le programme de calcul d'hybridation en utilisant la commande suivante:

java -jar {chemin ver Hybridation.jar} {chemin vers le fichier ARN messenger} {chemin vers le fichier micro ARN}

Enfin il est possible de réaliser une suite de test en mettant dans un dossier les fichiers ARN messenger et un autre les micro ARN et lancer la commande suivante:

./EvaluerHybridation.sh {chemin vers le dossier ARN messenger} {chemin vers le dossier micro ARN}

Structure de l'archive

Compte_rendu_TP2.pdf

Resultat_Hybridation.pdf

src (dossier de sources)

TP1.jar (*permet de lister tous les mots de n caractère dans une séquence ADN: `java -jar TP1 {fichier fasta} 3 D RC`*)

script-gnuplot.plot (*permet de générer les dotplot des exemples présenté partie 1*)

fasta (*dossier contenant les fichier exemples partie 1-2*)

arnMessenger (*contient les ARN messenger partie 3*)

microArn (*contient les micro ARN partie 3*)

EvaluerAppariement.sh (*réalise l'appariement sur tous les fichiers d'un dossier précisé*)

EvaluerHybridation.sh (*realiser l'hybridation de tous les fichiers arn messenger et micro arn*)

Appariement.jar

Hybridation.jar