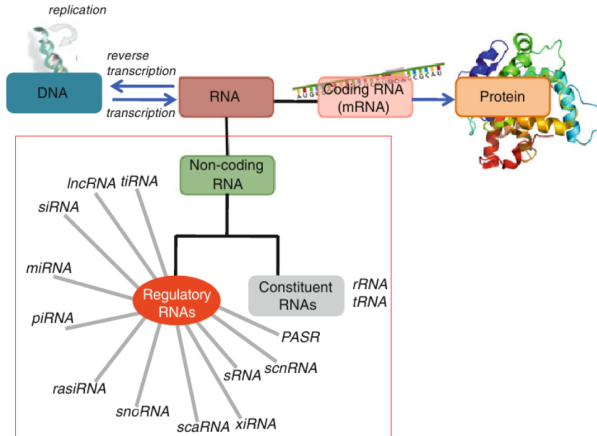
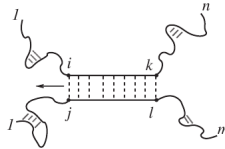




Prédiction des cibles des sRNAs avec rNAV. Un cas d'utilisation de Tulip.



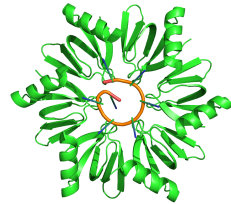
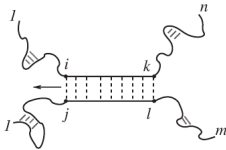
► *Bacterial small RNAs (sRNA)*



► *Bacterial small RNAs (sRNA)*

- Longueur comprise entre ~ 50 à 500 nucléotides (nt)
- Action basée sur un appariement court ($10 \sim 25$ nt) avec d'autres ARNs à l'aide de la protéine Hfq

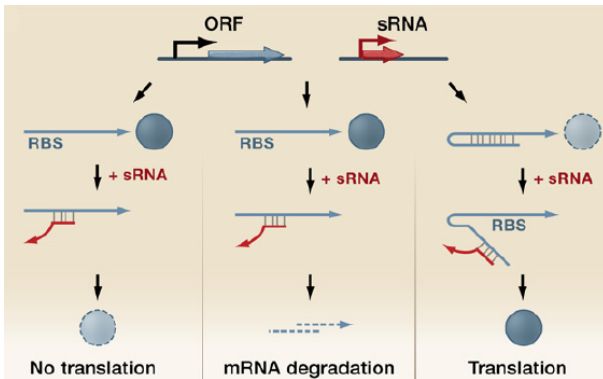
Les sRNAs



Hfq (PDB 1KP2)

▶ *Bacterial small RNAs* (sRNA)

- ▶ Longueur comprise entre ~ 50 à 500 nucléotides (nt)
- ▶ Action basée sur un appariement court ($10 \sim 25$ nt) avec d'autres ARNs à l'aide de la protéine Hfq
- ▶ Régulation de l'expression des gènes lors de changement des conditions environnementales
- ▶ Action sur les voies métaboliques ou voies de réponse au stress (transporteurs...)



[Waters and Storz, 2009]

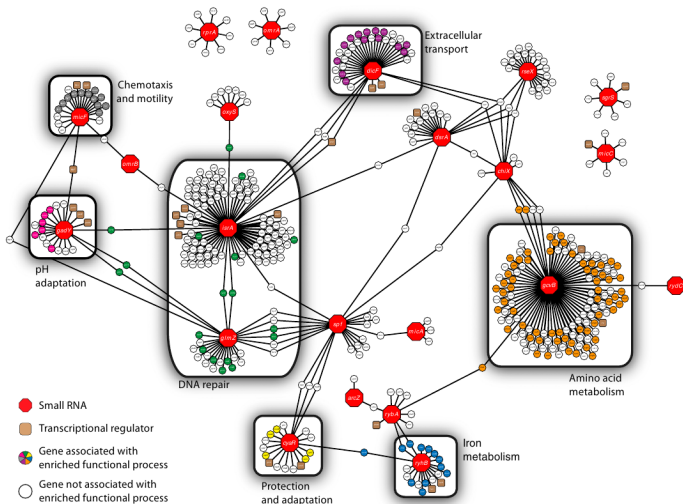
- Majorité des régulations connues sont négatives.

- ▶ Des centaines de sRNAs ont été identifiés par *RNA-sequencing*
- ▶ Seulement ~ 70 sRNAs ont des cibles connues [Richter and Backofen, 2011]
- ▶ Quels effets ont ces nouveaux sRNAs ? Quelles sont leurs cibles ?

Prédiction des cibles des sRNAs

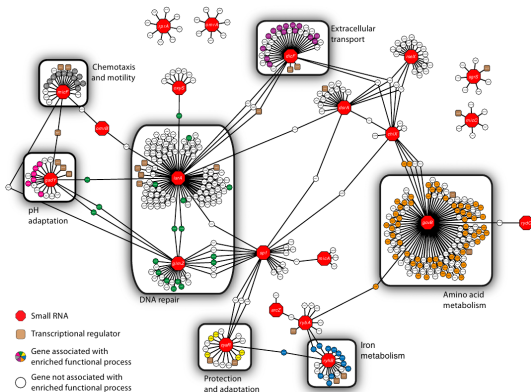
- ▶ Plusieurs programmes bioinformatiques permettent de prédire des cibles candidates pour les sRNAs.
- ▶ Caractérisation de la fonction demeure un défi important (>100 cibles prédites par sRNA).
- ▶ Méthodes non adaptées à un contexte d'utilisation haut débit (un sRNA - genome).

Formulation du problème sous la forme d'un graphe



[Modi et al., 2011]

Formulation du problème sous la forme d'un graphe



[Modi et al., 2011]

- ▶ Données de prédiction
- ▶ Utilisation interactive de la visualisation

i-RNA : Approche haut débit

Objectifs :

1. Intégrer les différents outils de prédiction des interactions sRNA-mRNA
 - ▶ Sélection de la méthode la plus sensible et spécifique
 - ▶ Prédiction à l'échelle du génome

i-RNA : Approche haut débit

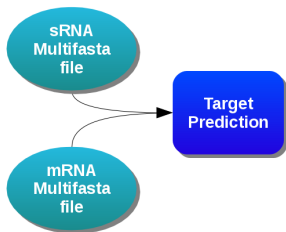
Objectifs :

1. Intégrer les différents outils de prédiction des interactions sRNA-mRNA
 - ▶ Sélection de la méthode la plus sensible et spécifique
 - ▶ Prédiction à l'échelle du génome
2. Améliorer la caractérisation fonctionnelle par l'analyse par enrichissement

i-RNA : Approche haut débit

Objectifs :

1. Intégrer les différents outils de prédiction des interactions sRNA-mRNA
 - ▶ Sélection de la méthode la plus sensible et spécifique
 - ▶ Prédiction à l'échelle du génome
2. Améliorer la caractérisation fonctionnelle par l'analyse par enrichissement
3. Améliorer l'expérience de l'utilisateur en utilisant la visualisation par les graphes (solution "big data")

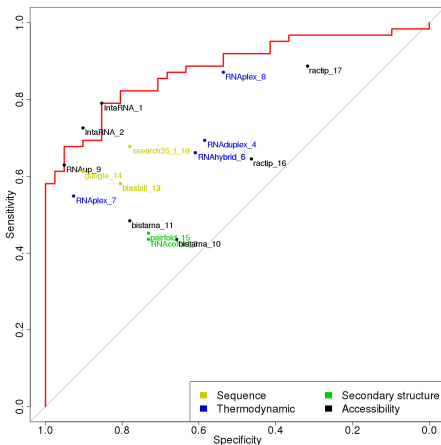


Prédiction des cibles

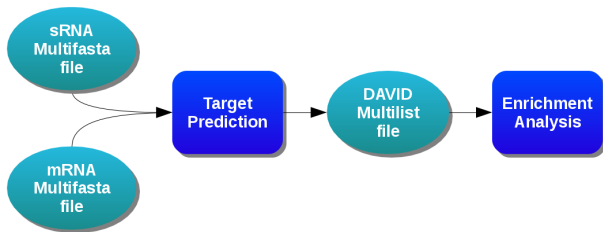
Logiciels	e-valeur	Score	MFE	Énergie
Blastall	X			
Ssearch	X			
Yass	X			
BistaRNA		X		
Guugle		X		
Ractip		X		
Pairfold			X	
RNA duplex			X	
RNA hybrid			X	
RNAplex			X	
IntaRNA				X
RNAcofold				X
RNAup				X

- ▶ Données d'entraînement :
 - ▶ 43 sRNAs - 52 mRNAs
 - ▶ 62 interactions / 41 non-interactions identifiées expérimentalement (sRNATarBase)

Prédiction des cibles



- ▶ IntaRNA [Busch et al., 2008] (AUC = 0.878)
- ▶ Complexité en temps : $O(n^2 m^2)$ (n et m : longueur du sRNA et du mRNA considéré)



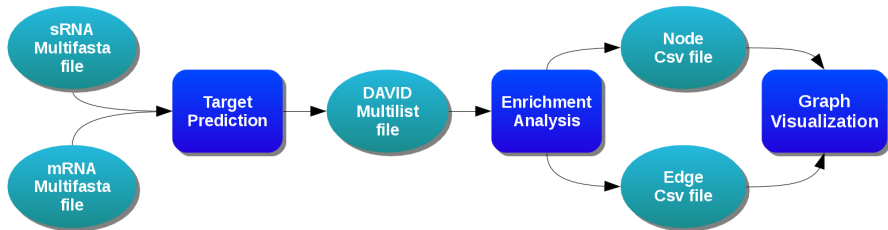
Analyse par enrichissement

- ▶ 68 outils bioinformatiques pour l'analyse par enrichissement [Huang et al., 2009]

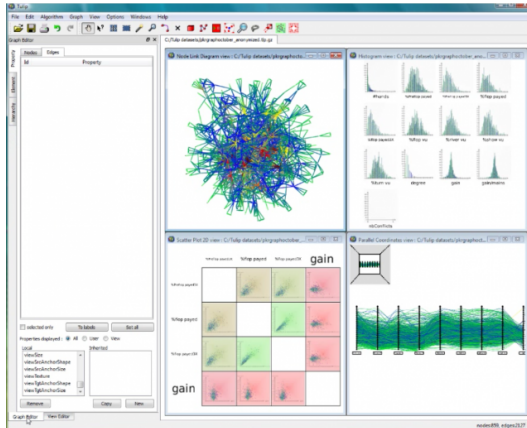
Analyse par enrichissement

- ▶ 68 outils bioinformatiques pour l'analyse par enrichissement [Huang et al., 2009]
- ▶ David knowledgebase [Huang et al., 2007]
 - ▶ Plusieurs organismes disponibles
 - ▶ Plusieurs sources données (GO, KEGG, InterPro, etc...)
 - ▶ Accès disponible par web-service

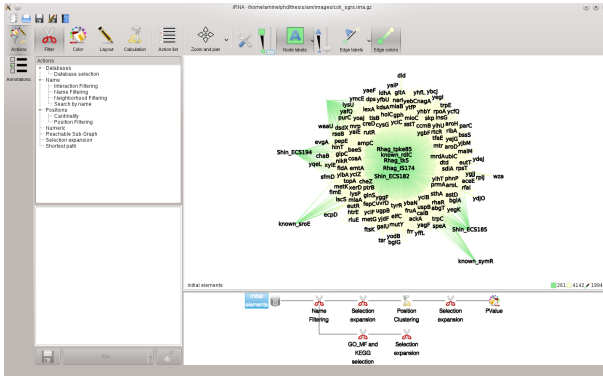
Pipeline d'analyse i-RNA



Visualisation de graphe : Tulip



- ▶ Tulip [Auber, 2003] :
 - ▶ Logiciel dédié à la visualisation de très grand graphes
 - ▶ Solution multi-plateforme développée en C++
 - ▶ Adapté au "Big data"



<http://youtu.be/YKGPY-tHcW4>

rNAV [Dubois*, Ghozlane* et al., 2013]

Prédiction des cibles des sRNAs à l'échelle du génome

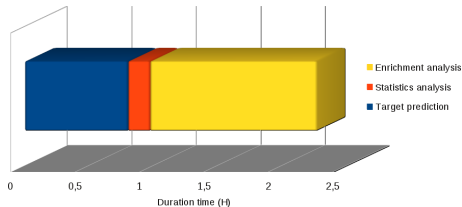
Application à *Escherichia coli* K12 mg1655 :

- ▶ Données : 261 sRNAs \times 4142 mRNAs = 1081062 comparaisons

Prédiction des cibles des sRNAs à l'échelle du génome

Application à *Escherichia coli* K12 mg1655 :

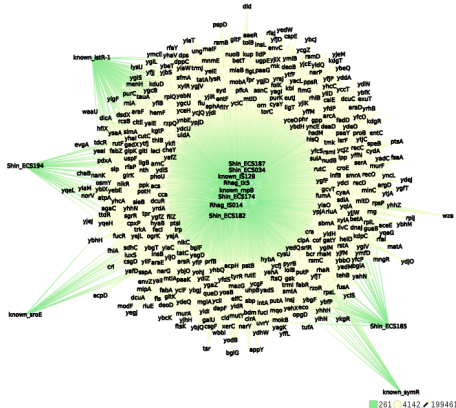
- ▶ Données : 261 sRNAs × 4142 mRNAs = 1081062 comparaisons
- ▶ Temps d'exécution :
 - ▶ 0,8H (prédictions réalisés sur 288 cpus) + 0.17H (analyse statistique réalisée sur 12 cpus) + 1.3H (analyse par enrichissement)



- ▶ Prédictions réalisées en 2.5H pour *Escherichia coli*

Prédiction des cibles des sRNAs à l'échelle du génome

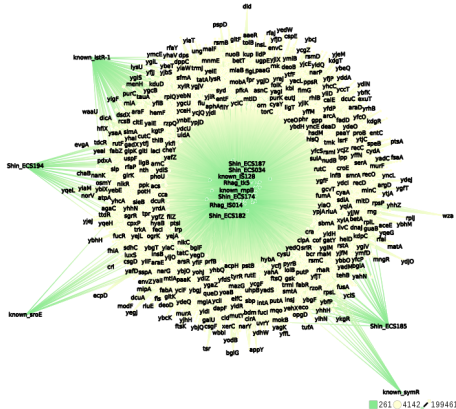
Application à *Escherichia coli* K12 mg1655 :



- ▶ 199461 interactions (1081062 potentielles)

Prédiction des cibles des sRNAs à l'échelle du génome

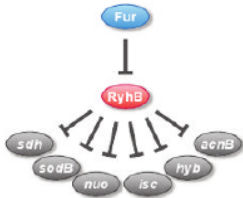
Application à *Escherichia coli* K12 mg1655 :



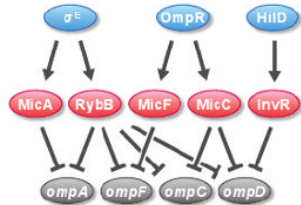
- ▶ 199461 interactions (1081062 potentielles)

Comment analyser ces prédictions ?

- ▶ Différents types de régulations peuvent être identifiées [Beisel and Storz, 2010] :

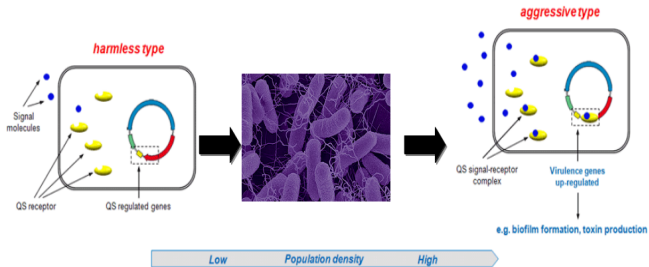


Single-input module (SIM)



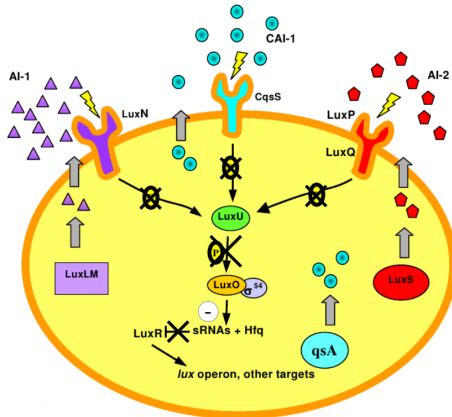
Dense overlapping regulon (DOR)

Cas d'étude : Analyse de la sensibilité au quorum



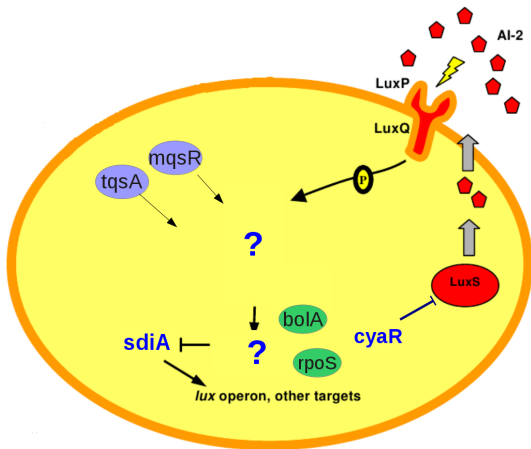
- ▶ Système de communication entre les bactéries
- ▶ Permet de synchroniser l'expression des gènes selon la densité de la population
- ▶ Responsable de la formation de biofilm et de la production de toxine

Cas d'étude : Analyse de la sensibilité au quorum



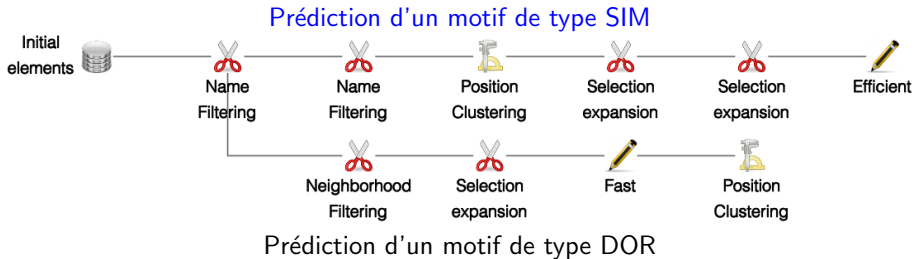
Forte densité de population
Sensibilité au quorum de *Vibrio harvei* [Wang, 2004]

Cas d'étude : Analyse de la sensibilité au quorum



Sensibilité au quorum *Escherichia coli*

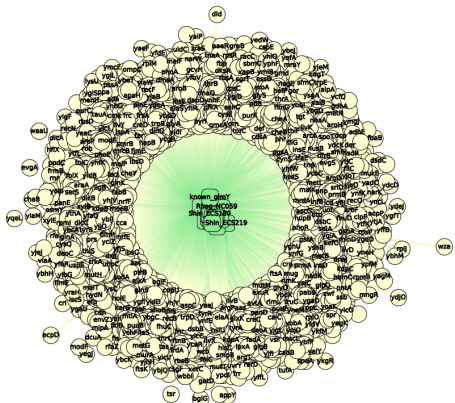
Cas d'étude : Analyse de la sensibilité au quorum



Cas d'étude : Analyse de la sensibilité au quorum

Prédictions d'un motif de type SIM :

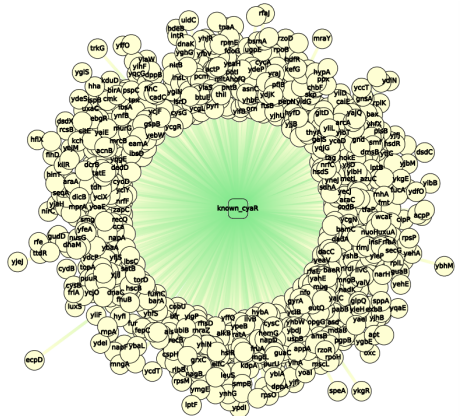
1. Recherche des sRNAs qui interagissent avec LuxS et leurs cibles
33 sRNAs - 4138 mRNAs



Cas d'étude : Analyse de la sensibilité au quorum

Prédictions d'un motif de type SIM :

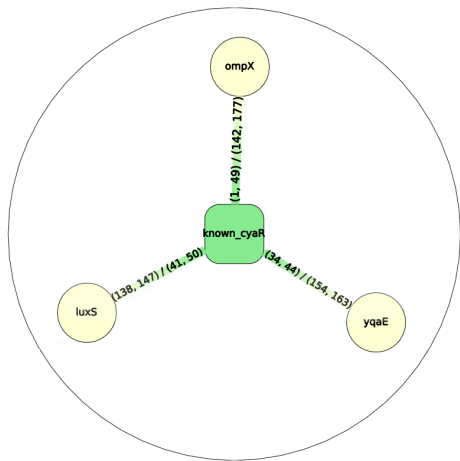
1. Recherche des sRNAs qui interagissent avec LuxS et leurs cibles
33 sRNAs - 4138 mRNAs
2. Interactions de CyaR : 986 cibles



Cas d'étude : Analyse de la sensibilité au quorum

Prédictions d'un motif de type SIM :

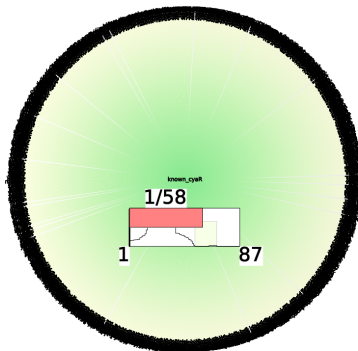
1. Recherche des sRNAs qui interagissent avec LuxS et leurs cibles
33 sRNAs - 4138 mRNAs
2. Interactions de CyaR : 986 cibles



Cas d'étude : Analyse de la sensibilité au quorum

Prédictions d'un motif de type SIM :

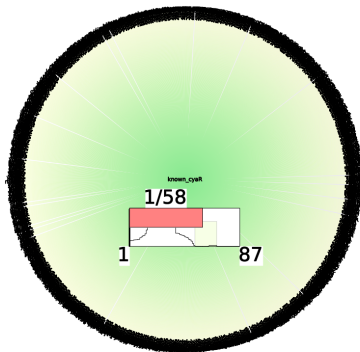
1. Recherche des sRNAs qui interagissent avec LuxS et leurs cibles
33 sRNAs - 4138 mRNAs
2. Interactions de CyaR : 986 cibles
3. Sélection des interactions de CyaR dans la région 1..58 (966 cibles)



Cas d'étude : Analyse de la sensibilité au quorum

Prédictions d'un motif de type SIM :

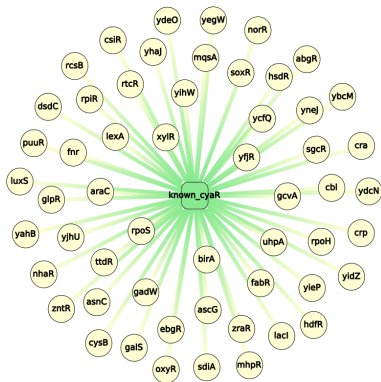
1. Recherche des sRNAs qui interagissent avec LuxS et leurs cibles
33 sRNAs - 4138 mRNAs
2. Interactions de CyaR : 986 cibles
3. Sélection des interactions de CyaR dans la région 1..58 (966 cibles)
4. Sensibilité au Quorum :
 - ▶ Uniprot : *DNA binding :H-T-H* (luxR)



Cas d'étude : Analyse de la sensibilité au quorum

Prédictions d'un motif de type SIM :

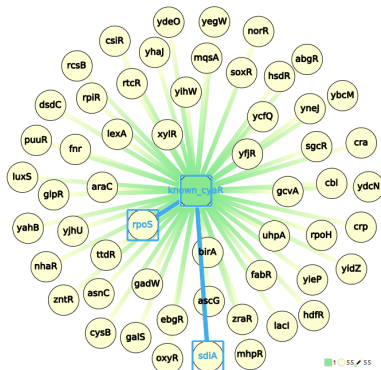
1. Recherche des sRNAs qui interagissent avec LuxS et leurs cibles
33 sRNAs - 4138 mRNAs
2. Interactions de CyaR : 986 cibles
3. Sélection des interactions de CyaR dans la région 1..58 (966 cibles)
4. Sensibilité au Quorum :
 - ▶ Uniprot : *DNA binding :H-T-H* (luxR)
5. 55 cibles identifiées



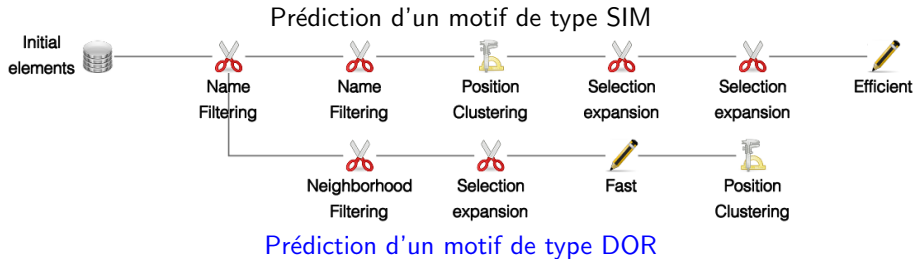
Cas d'étude : Analyse de la sensibilité au quorum

Prédictions d'un motif de type SIM :

1. Recherche des sRNAs qui interagissent avec LuxS et leurs cibles
33 sRNAs - 4138 mRNAs
2. Interactions de CyaR : 986 cibles
3. Sélection des interactions de CyaR dans la région 1..58 (966 cibles)
4. Sensibilité au Quorum :
 - ▶ Uniprot : *DNA binding :H-T-H* (luxR)
5. 55 cibles identifiées
6. *sdiA* (homologue de *luxR*) [Dyszal et al., 2010] et *rpoS* (réponse au stress) [Adnan et al., 2010]



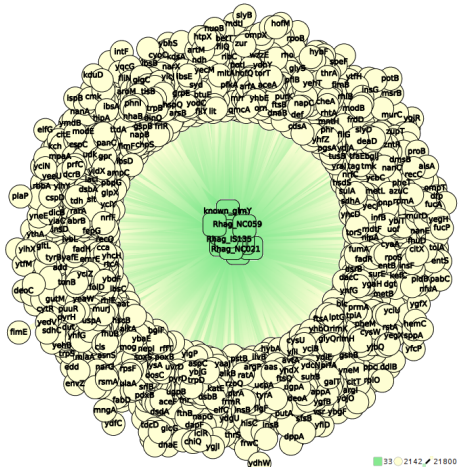
Cas d'étude : Analyse de la sensibilité au quorum



Cas d'étude : Analyse de la sensibilité au quorum

Prédiction d'un motif de type DOR :

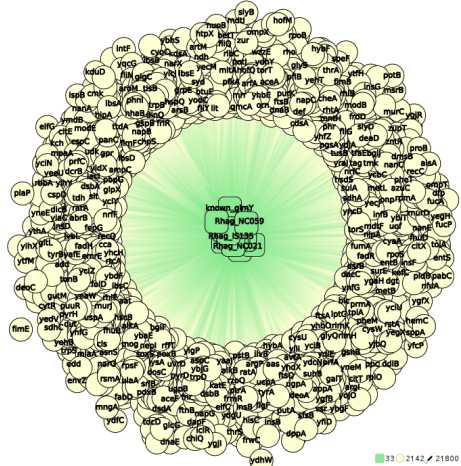
1. mRNA interagissant avec au moins 8 sRNAs en commun
33 sRNAs - 2142 mRNAs



Cas d'étude : Analyse de la sensibilité au quorum

Prédiction d'un motif de type DOR :

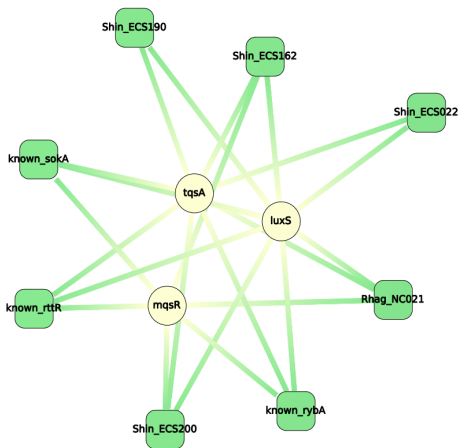
1. mRNA interagissant avec au moins 8 sRNAs en commun
33 sRNAs - 2142 mRNAs
2. Sélection des éléments annotés pour la sensibilité au quorum :
 - ▶ Swissprot et Gene ontology biological pathways (GO_BP)



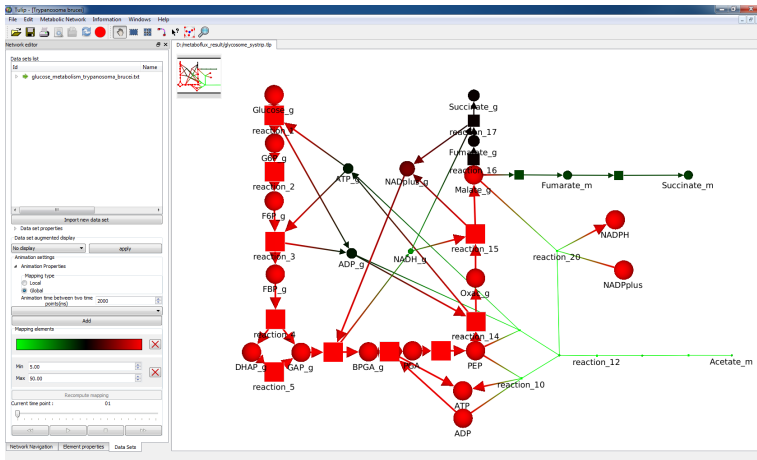
Cas d'étude : Analyse de la sensibilité au quorum

Prédiction d'un motif de type DOR :

1. mRNA interagissant avec au moins 8 sRNAs en commun
33 sRNAs - 2142 mRNAs
2. Sélection des éléments annotés pour la sensibilité au quorum :
 - ▶ Swissprot et Gene ontology biological pathways (GO_BP)
3. 8 sRNAs - 3 mRNAs
4. *tqsA* and *mqsR* sont impliqués dans le contrôle de la formation de biofilm [Rettner and Saier, 2010]







- ▶ Le pipeline iRNA :
 - ▶ Prédiction des cibles pour l'ensemble des sRNAs d'un organisme
 - ▶ Utilisation de la **visualisation** pour l'identification d'un nombre restreint de cibles
- ▶ Apport de Tulip :
 - ▶ Manipulation et retour visuel immédiat
 - ▶ "Big data" sur portable






<http://youtu.be/2rbvVIMOmV>

Systrip [Dubois, Cottret, Ghozlane et al., 2012]





Bibliography I

-  Adnan, M., Morton, G., Singh, J., and Hadi, S. (2010).
Contribution of rpos and bola genes in biofilm formation in escherichia coli k-12 mg1655.
Mol Cell Biochem, 342(1-2) :207–213.
-  Auber, D. (2003).
Tulip-a huge graph visualization framework.
Graph Drawing Software, 2265 :105–126.
-  Beisel, C. L. and Storz, G. (2010).
Base pairing small rnas and their roles in global regulatory networks.
FEMS Microbiol Rev, 34(5) :866–882.
-  Busch, A., Richter, A. S., and Backofen, R. (2008).
Intarna : efficient prediction of bacterial srna targets incorporating target site accessibility and seed regions.
Bioinformatics, 24(24) :2849–2856.

Bibliography II

-  Dyszel, J. L., Soares, J. A., Swearingen, M. C., Lindsay, A., Smith, J. N., and Ahmer, B. M. M. (2010).
E. coli k-12 and ehec genes regulated by sdia.
PLoS One, 5(1) :e8946.
-  Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009).
Bioinformatics enrichment tools : paths toward the comprehensive functional analysis of large gene lists.
Nucleic Acids Res, 37(1) :1–13.
-  Huang, D. W., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M. W., Lane, H. C., and Lempicki, R. A. (2007).
David bioinformatics resources : expanded annotation database and novel algorithms to better extract biology from large gene lists.
Nucleic Acids Res, 35(Web Server issue) :W169–W175.

Bibliography III

-  Modi, S. R., Camacho, D. M., Kohanski, M. A., Walker, G. C., and Collins, J. J. (2011).
Functional characterization of bacterial srnas using a network biology approach.
Proc Natl Acad Sci U S A, 108(37) :15522–15527.
-  Rettner, R. E. and Saier, Jr, M. H. (2010).
The autoinducer-2 exporter superfamily.
J Mol Microbiol Biotechnol, 18(4) :195–205.
-  Richter, A. and Backofen, R. (2011).
Accessibility and conservation in bacterial small rna-mrna interactions and implications for genome-wide target predictions.
-  Wang, L. (2004).
Autoinducer-2 (AI-2) mediated quorum sensing in Escherichia coli.
PhD thesis, University of Maryland, College Park.

Bibliography IV



Waters, L. S. and Storz, G. (2009).
Regulatory rnas in bacteria.
Cell, 136(4) :615–628.