



---

## **Construction d'un réseau de co-expression pour des gènes impliqués dans la réponse au nitrate**

---

BOTHOREL Benoît  
GONÇALVES FRASCO Charlotte  
JACQUET Pierre

**Contact :** *[benoit.bothorel@etu.u-bordeaux.fr](mailto:benoit.bothorel@etu.u-bordeaux.fr)*

## Introduction

Les nitrates ( $NO_3^-$ ) sont de puissants eutrophisants qui sont indispensables au développement des plantes. Ils sont par ailleurs considérés comme des polluants lorsqu'ils sont retrouvés en quantité anormale dans les sols. Les engrains utilisés dans l'agriculture contiennent des nitrates, et ont pour but de favoriser la croissance des plantes[1]. Cependant l'utilisation excessive d'engrais peut entraîner une contamination des sols. Si le sol est trop concentré en nitrate les plantes ne sont plus capables d'extraire l'eau du sol, à cause du gradient osmotique, ces dernières finissent donc par faner.[2]

L'étude des gènes impliqués dans la réponse aux nitrates pourrait permettre de déterminer des moyens d'action comme par exemple faire en sorte que les plantes puissent avoir une gestion optimale du nitrate.[3] C'est dans cette optique qu'un réseau de co-expression des gènes est réalisé. Il permet de mettre en évidence les gènes qui sont corrélés ou anti-corrélos au niveau d'un processus biologique donné.

En premier lieu, un degré de co-expression est calculé en utilisant la corrélation de Pearson. Ensuite, un seuil est choisi pour déterminer à partir de quel moment on considère la co-expression comme étant significative. Il existe des outils sous R, comme la fonction *pickSoftThreshold()*, qui permettent de déterminer un seuil optimal.[4] Des filtres sont appliqués pour ne garder que les informations importantes, notamment en utilisant des données comme les interactions protéine-protéine (PPI).

Enfin, après avoir obtenu la liste des gènes les plus significatifs, Cytoscape est utilisé pour créer un graphe de visualisation des interactions entre les différents gènes. Cette représentation est ensuite utilisée pour émettre des hypothèses sur différents mécanismes et régulation au niveau du processus biologique étudié.

# Matériel & Méthodes

## Matériel

### Données

Les données dont nous avons disposé pour cette étude ont été préparés à partir des données aimablement mises à disposition par Tomás C. Moyano, Elena A. Vidal, Orlando Contreras-López et Rodrigo A. Gutiérrez du « Plant Systems Biology Lab » de l'université pontificale catholique du Chili. Les données proviennent du laboratoire et de différentes bases de données publiques (cf. Annexes). Les données sont regroupées dans différents fichiers textes :

- P\_EXPR.txt : Table qui donne 146 mesures pour une gène donné d'*Arabidopsis Thaliana* (AT)
- P\_GOI.txt : Table contenant les noms des locus des gènes d'AT connus pour être impliqués dans la réponse au nitrate
- P\_PPI.txt : Table contenant la liste des paires de gènes codant pour des protéines en interaction
- P\_TF\_LABEL.txt : Table donnant les identifiants des locus de gènes ayant un facteur de transcription et la famille de ce dernier

## R

R est un langage et un environnement pour l'informatique statistique et les graphiques. Celui-ci fournit une grande variété de techniques statistiques (modélisation linéaire et non linéaire, tests statistiques classiques, classification, regroupement, ...) et graphiques, et est extrêmement extensible.

L'un des points forts de R est la facilité avec laquelle des diagrammes bien conçus, de qualité publication, peuvent être générés.

R est disponible en tant que logiciel libre selon les termes de la licence publique générale GNU General de la Free Software Foundation sous forme de code source. [5]

## KEGG

KEGG (*Kyoto Encyclopedia of Genes and Genomes*) est une base de connaissances pour l'analyse systématique des fonctions des gènes, liant les informations génomiques aux informations fonctionnelles d'ordre supérieur.[6]

Les informations génomiques sont stockées dans la base de données **GENES**, qui est une collection de catalogues de gènes pour tous les génomes complètement séquencés et certains génomes partiels avec une annotation à jour des fonctions des gènes.

Les informations fonctionnelles d'ordre supérieur sont stockées dans la base de données **PATHWAY**, qui contient des représentations graphiques de processus cellulaires, tels que le métabolisme, le transport membranaire et autre.

La troisième base de données de KEGG, **LIGAND**, contient des informations sur les composés chimiques, les molécules d'enzyme et les réactions enzymatiques.

## Cytoscape

Cytoscape est un projet logiciel open source pour l'intégration de réseaux d'interaction biomoléculaire avec des données d'expression à haut débit et d'autres états moléculaires. [7]

Cytoscape est utilisé conjointement avec de vastes bases de données d'interactions protéine – protéine, ADN – protéine et génétique qui sont de plus en plus disponibles pour l'homme et les organismes modèles.

Le logiciel Cytoscape est extensible via une architecture utilisant des plugins, permettant le développement rapide d'analyses et de fonctionnalités supplémentaires. Par exemple pour ce projet, les plugins suivants ont été utilisés :

- ClueGo : ClueGO est un plug-in Cytoscape qui améliore l'interprétation biologique de grandes listes de gènes.[8] ClueGO intègre à la fois les termes Ontologie (GO) ainsi que les voies KEGG et crée un réseau de termes GO. Il peut analyser et comparer deux listes de gènes et permet de visualiser les termes regroupés de manière fonctionnelle sous la forme de réseaux et graphiques. D'autres outils tels que BiNGO [9] évaluent les termes GO surreprésentés et reconstruisent l'arbre ontologique hiérarchique, alors que ClueGO utilise des statistiques kappa pour relier les termes dans le réseau. ClueGO génère donc une structure de réseau dynamique en considérant déjà initialement les listes de gènes d'intérêt.
- GLay : GLay fournit une collection variée d'algorithmes de structure de communauté et de fonctions d'agencement de graphe polyvalents pour la mise en cluster de réseaux et la visualisation structurée. [10]
- WordCloud : WordCloud permet de générer des "nuages" de mots organisés en cluster ou non, à partir des colonnes spécifiés provenant de la table de noeuds. Cela permet de visualiser rapidement les termes les plus utilisés dans un réseau.

## Méthodes

### R

Le script R effectué (Cf. Annexes), a permis de réaliser tout d'abord un filtrage des données d'expression correspondant aux gènes d'intérêt. Ensuite, il a été utilisé afin d'obtenir la mesure de la corrélation de tous les groupes de gènes. Un seuil a ensuite été établi à 0.85 afin de récupérer les corrélations linéaires "fortes" entre les variables. Il a été décidé d'affecter, grâce à une boucle dans ce script, un type de corrélation à chaque couple de gènes (corrélé ou anti-corrélé) afin que ce critère soit exploité au cours de notre analyse. Les données du fichier renseignant les interactions Protéine-Protéine (PPI) a ensuite été ajouté dans le but, à nouveau, d'alimenter notre analyse.

## Cytoscape

L'importation des données produites à l'aide du script R sous Cytoscape, a permis d'obtenir un réseau de gènes très dense (*Fig. 1*). Afin d'en dégager des données utilisables pour l'analyse, plusieurs traitements ont dû être élaborés.

Tout d'abord, l'utilisation de **ClueGO** a permis de dégager les gènes regroupés de manière fonctionnelle à l'aide de la *GO\_BiologicalProcess ontology*. L'étude se basant sur les gènes de la réponse aux nitrates, il a pu être mis en évidence des *GoTerm* intéressants pour l'analyse (*Tab. 1, Fig. 2*). ClueGO a été utilisé à la place de BinGO car il associe directement les GOterms aux gènes dans la table des noeuds. Cela a permis de visualiser plus rapidement les fonctions biologiques et les gènes associés. Les deux outils ont été testé et bien que l'on peut spécifier un "degré" de spécificité des termes dans ClueGO, pour le nitrate, les termes renvoyés par les deux plugins sont les mêmes.

Les termes obtenus sont alors associés aux noeuds afin de créer un graphe reliant les processus biologiques principaux détectés pour les lier aux gènes y répondant le plus (*Fig. 3*). Les noeuds ainsi obtenus présentent donc de nombreuses annotations en fonction de leur rôle biologique. Il a pu être observé la présence de gènes de réponse aux nitrates ("*response to nitrate*"), il s'agit du terme couvrant le plus de gènes. Cependant, les termes "*nitrate metabolic process*" et "*nitrate assimilation*" sont également présent mais moins significatifs. Les gènes impliqués dans les processus métaboliques du nitrate ont donc été mis en évidence (*Fig. 4*).

Afin d'isoler les gènes les plus intéressants pour l'analyse, il a été pertinent de réaliser un sous-graphe reprenant les noeuds associés au nitrate ainsi que les noeuds proches. Il est ainsi possible d'observer que les noeuds impliqués dans la réponse au nitrate le sont également dans d'autres processus comme GLT1 avec le cycle d'assimilation de l'ammoniac (*Fig. 5*).

Ayant associé les GOterms à chaque noeud, ils peuvent alors être visualisés sous formes de nuage de mots avec le plugin **WordCloud**. Ce plugin permet de visualiser plus facilement les termes associés à des groupes de noeuds et permet ainsi d'identifier les clusters les plus intéressants (*Fig. 6*). En cliquant sur un mot, tout les noeuds associés sont automatiquement sélectionnés.

Les principaux clusters ont ensuite été réalisé à l'aide du plugin **GLay**. Le graphe principal a donc été clusterisé en utilisant les sous-graphes les plus connectés (*Fig. 7*). Les gènes associés au nitrate identifiés par ClueGo ont donc permis de mettre en évidence trois clusters principaux.

Le cluster 3 a été isolé et annoté en premier (le moins étendu). La visualisation des nuages de mots associés à chaque groupe permet de déterminer plus facilement leurs fonctions clés (*Fig. 8*). La même chose a été réalisée avec le cluster 2. Cela a permis de mettre en évidence un graphe ultra-connecté. Les gènes intervenant dans les mêmes fonctions au sein du graphe ont pu être isolés. Tous les gènes n'ont pas automatiquement été annotés avec ClueGo à partir de la *GO\_Biological process*. Grâce à cette représentation, il a pu être mis en évidence que les deux gènes AT1G78050 et AT2G156200 sont fortement connectés avec les autres noeuds (*Fig. 9*).

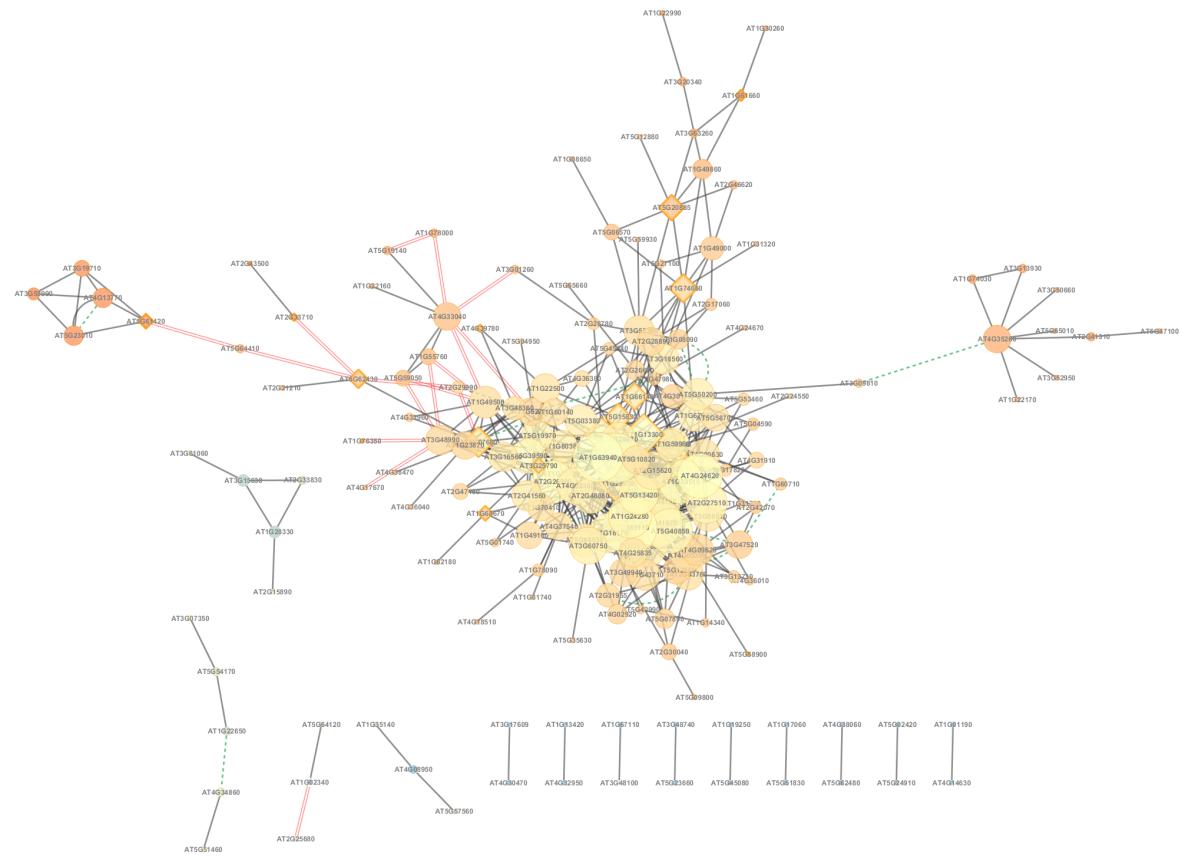
Le cluster 1 n'a pas été annoté intégralement. Celui-ci présentait un seul gène de réponse au nitrate et beaucoup de gènes supplémentaires dans un graphe également très connecté. Le gène en question est AT3G48360 aussi appelé BT2 dans la figure 5 qui n'est connecté avec aucune fonction biologique annoté (autre que celle du nitrate) dans le réseau créé par ClueGO en lien avec nos gènes d'intérêt.

## KEGG

Les gènes présentés précédemment (*Tab. 1*) AT5G53460 (GLT1), AT2G15620 (NIR1) et AT1G08090 (NRT2 :1) sont impliqués dans le métabolisme de l'azote comme on peut le voir, sur la figure 10, en utilisant la base de données **KEGG**.

## Résultats

Le graphe obtenu en *Figure 1* prouve la compléxité du réseau de gène devant être exploité. Grâce aux types de corrélation ajouté pour chaque interaction entre les gènes, il est tout de même possible de visualiser les relations anti-corrélées (double ligne rouge), les interactions Protéines-Protéines (ligne en pointillé vert) ou encore les gènes étant aussi des facteurs de transcription (losanges). Cependant la densité étant telle, d'autres manipulation sont nécessaire pour en dégager des informations pertinentes.

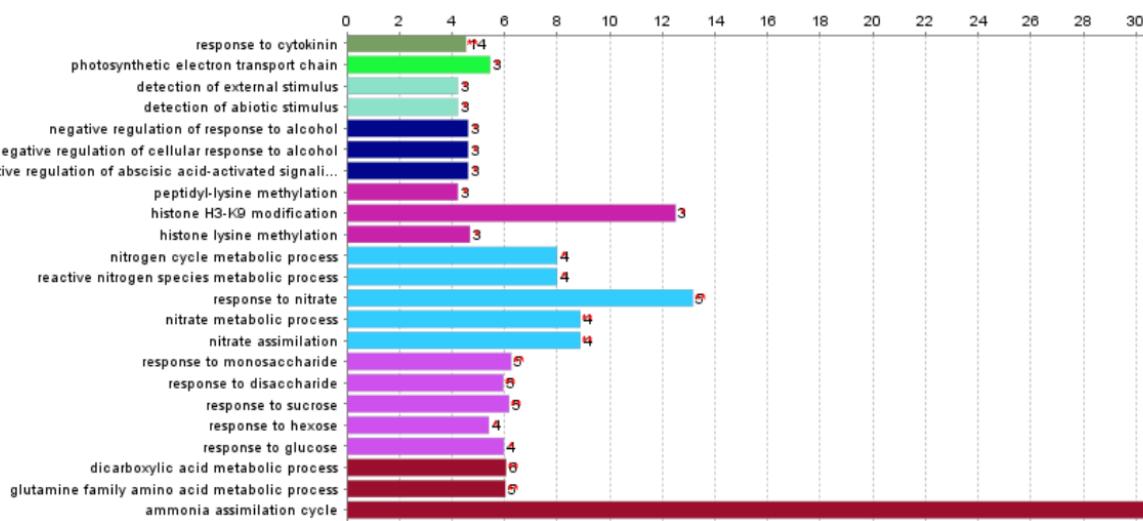


**Figure 1** – Réseau de gènes depuis les données produites par le script R.

*Le graphe ici représenté est très dense. Il est donc difficile d'en dégager des résultats permettant d'effectuer une analyse.*

Les termes GO permettent de décrire les propriétés des produits géniques étudiés et ainsi connaître les fonctions moléculaires réalisées lors des processus biologique par chaque gène. Dans la *Figure 2*, les termes sont regroupés dans un tableau qui permet d'observer les pourcentages de gènes par terme sur tous les gènes reconnus dans la base GO biological process provenant du graphes général. Cette annotation est à titre indicatif mais nous permet tout de même de constater que les termes les plus représentés sont ceux en relation avec le **cycle d'assimilation de l'ammoniac** et la **reponse au nitrate**.

La *Table 1* permet d'observer plus en détails les valeurs retrouvées pour la réponse au nitrate. Ainsi, nous pouvons avoir une information sur la *p-value* mais aussi sur les gènes associés à cette fonction biologique.



**Figure 2** – Termes GO présent dans notre jeu de données

Response to nitrate :

P–Value:  $6.803 \times 10^{-6}$

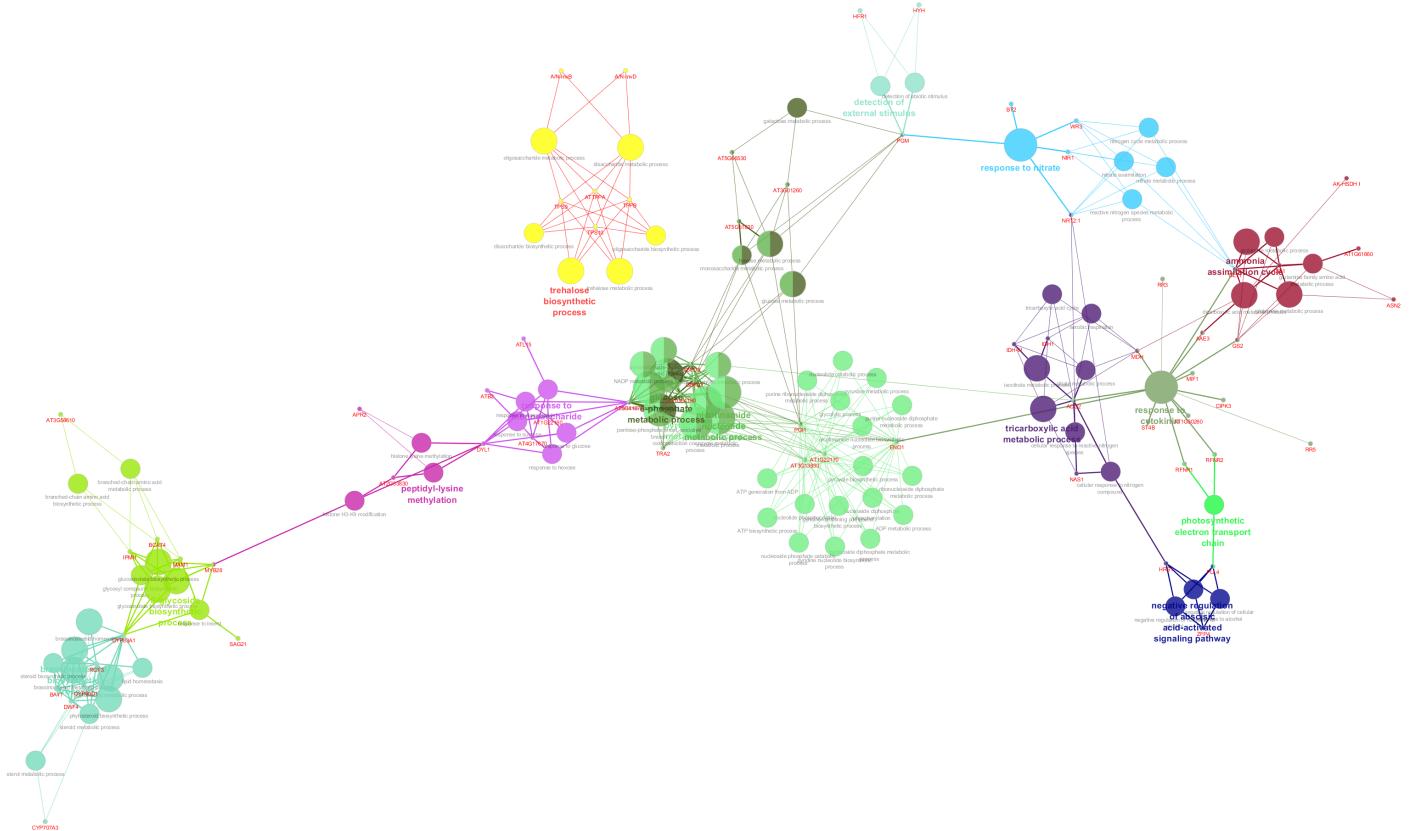
P–Value corrected with Bonferonni step:  $3.878 \times 10^{-4}$

Associated genes found :

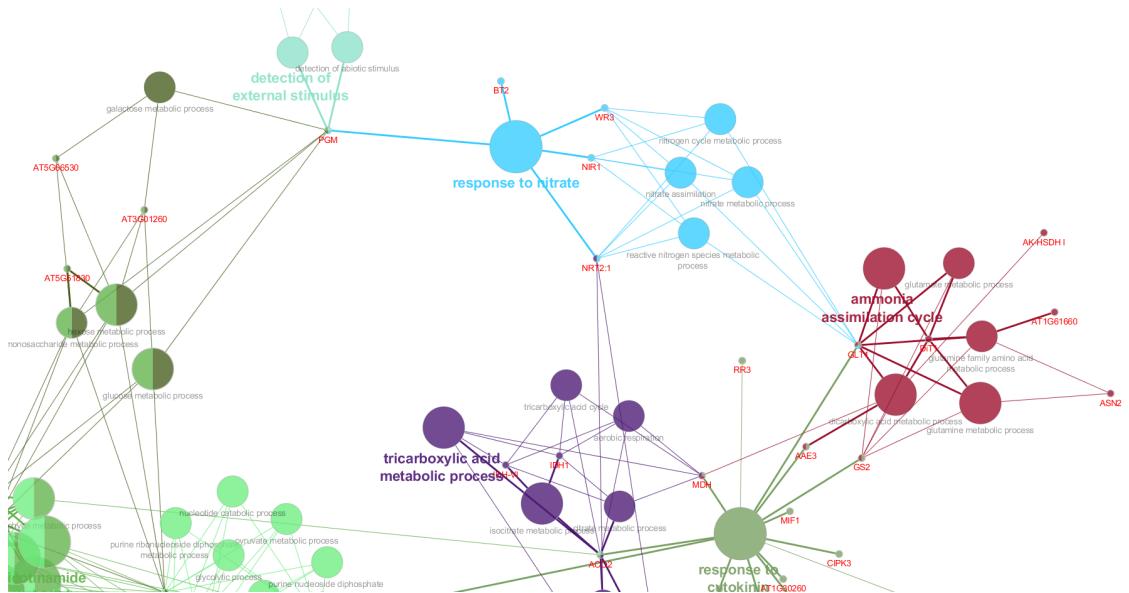
- AT3G48360 (BT2)
- AT5G53460 (GLT1)
- AT2G15620 (NIR1)
- AT1G08090 (NRT2:1)
- AT1G78050 (PGM)
- AT5G50200 (WR3)

**Table 1** – Liste des gènes associés au terme GO "Response to nitrate"

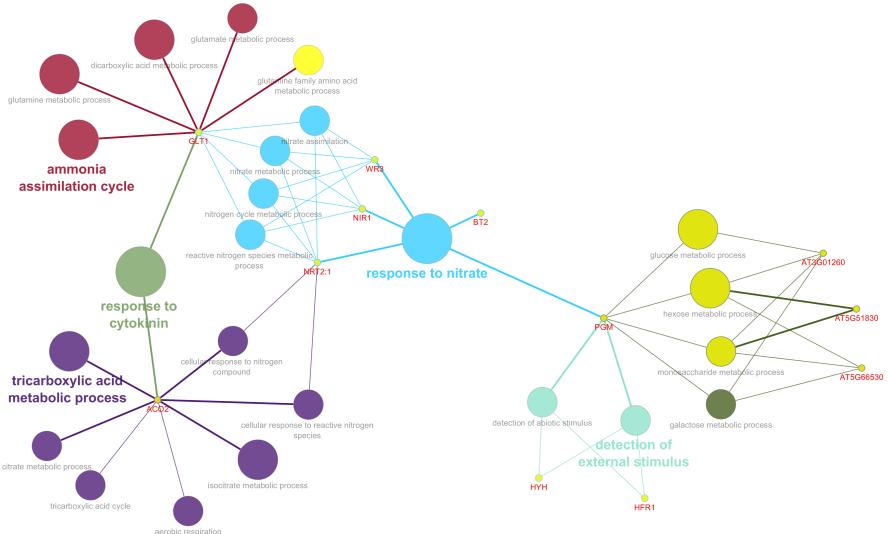
Les termes GO associés au graphe *Figure 3* permettent de constater qu'il y a de nombreuses fonctions autre que celles liées au nitrate et qu'elles forment des groupes facilement identifiables. Grâce au focus réalisé dans la *Figure 4* et au sous-graphe de la *Figure 5*, nous pouvons plus facilement identifier les groupes de gènes directement liés à ceux impliqués dans la réponse au nitrate : *detection of external stimulus*, *tricarboxylic acid metabolic process* et *ammonia assimilation cycle*.



**Figure 3** – Résultats obtenus à l'aide de ClueGO et aux termes retrouvés dans les données.  
Les gènes en rouge sont les gènes impliqués dans les processus biologiques liés au nitrate

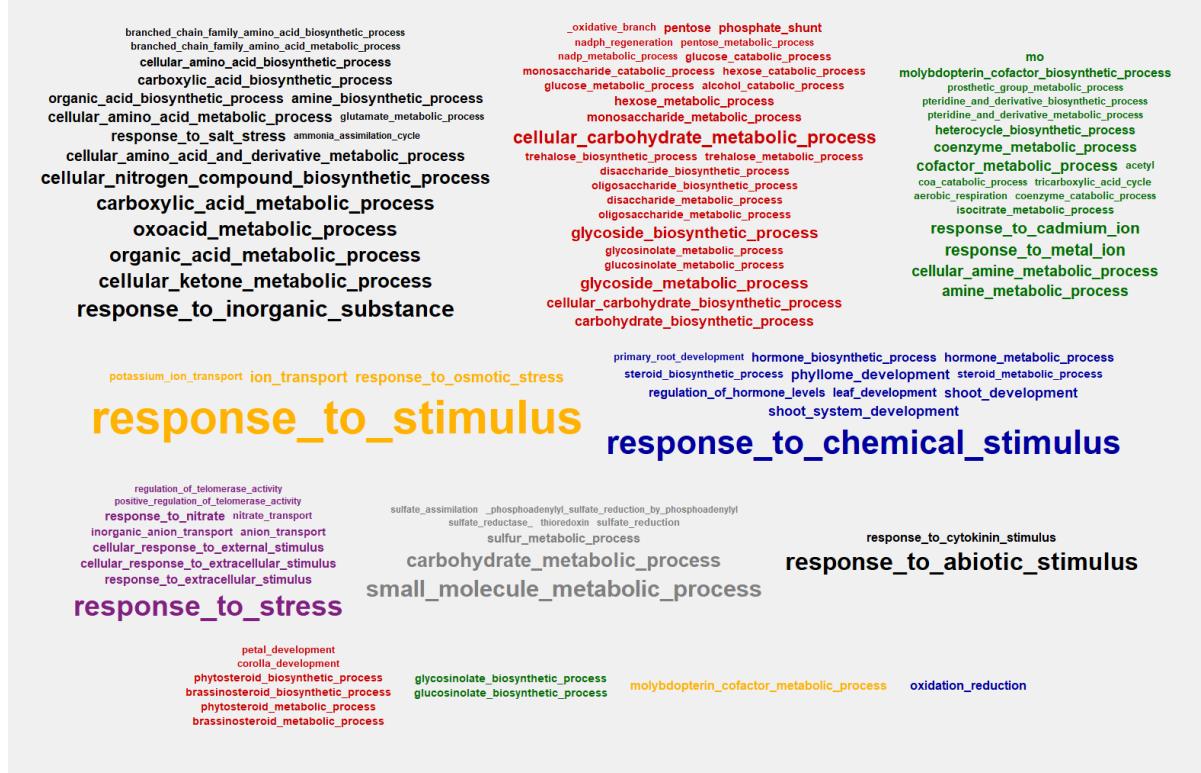


**Figure 4** – Focus de la *Figure 3* au niveau des gènes d'intérêts ayant un lien avec le nitrate



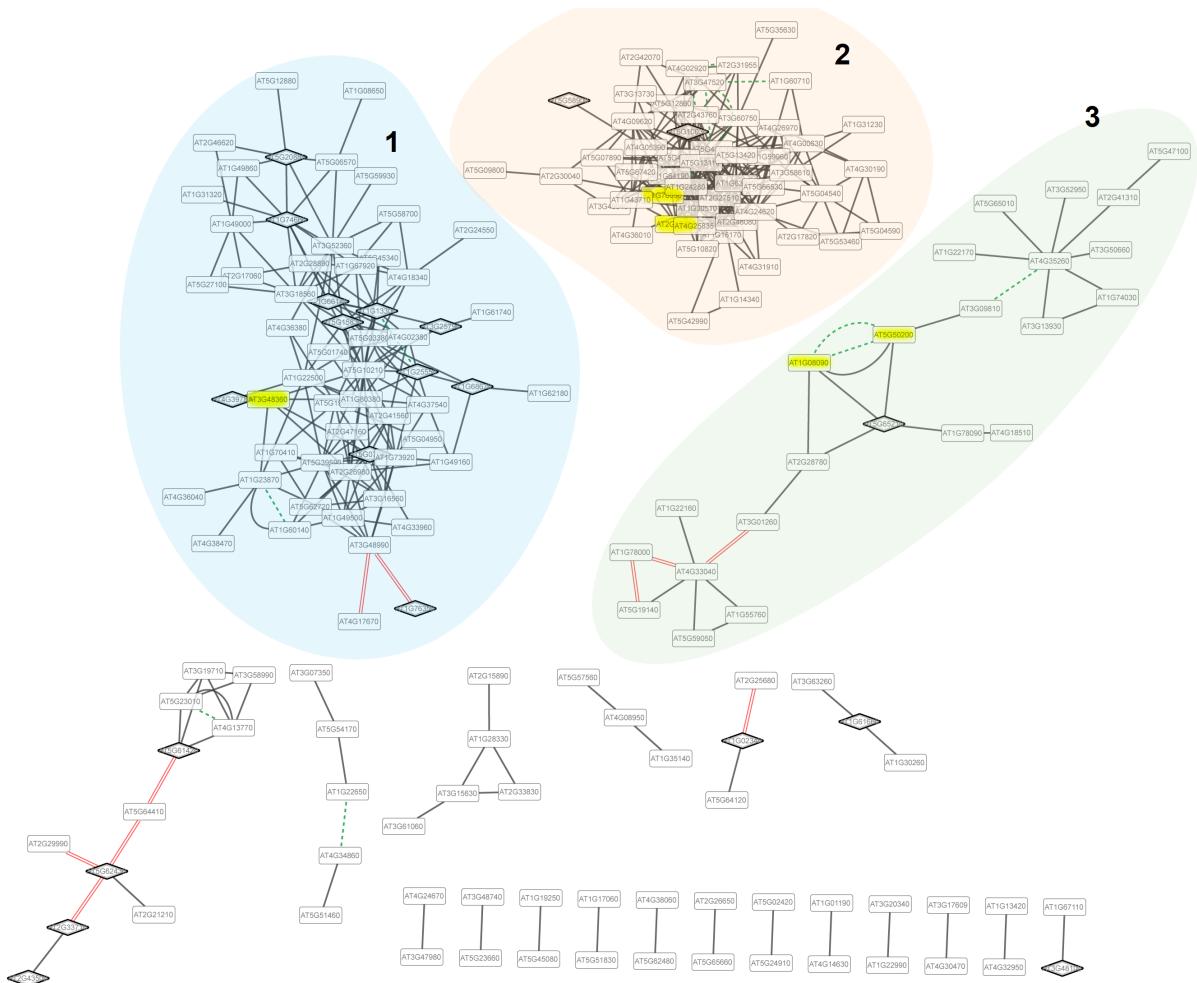
**Figure 5** – Sous-graphe reprenant les noeuds associés au nitrate ainsi que les noeuds proches.

Le nuage de mots *Figure 6* nous permet de constater à nouveau que les termes les plus représentés sont ici "response to stimulus" ou "response to chemical stimulus" ou encore "response to stress", ce qui est logique puisque le nitrate est impliqué dans ces processus biologiques.



**Figure 6** – Nuage de mots permettant de visualiser les termes les plus représentés

Après avoir clusterisé le graphe principal (*Figure 1*), les sous-graphes obtenus *Figure 7* mettent en évidence trois clusters respectivement annotés "1", "2" et "3". Ces clusters contiennent tous des gènes liés au nitrate (en jaune), des facteurs de transcription mais aussi quelques relations de types anti-corréllées ou d'interaction protéines-protéines.

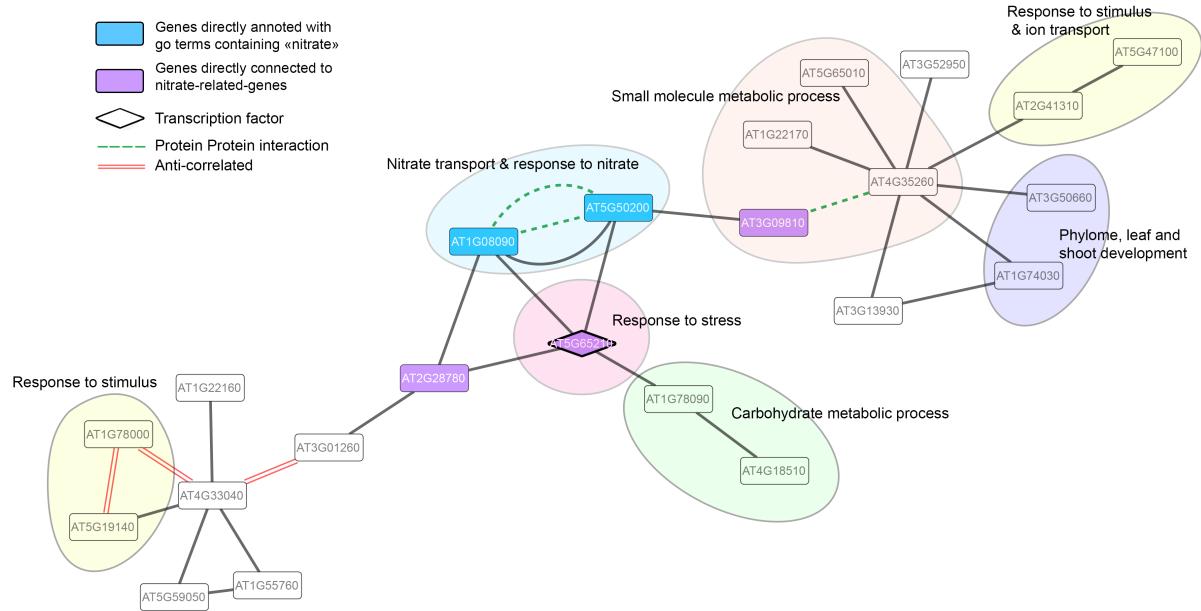


**Figure 7** – Visualisation des sous-graphes contenant les gènes associés au nitrate.

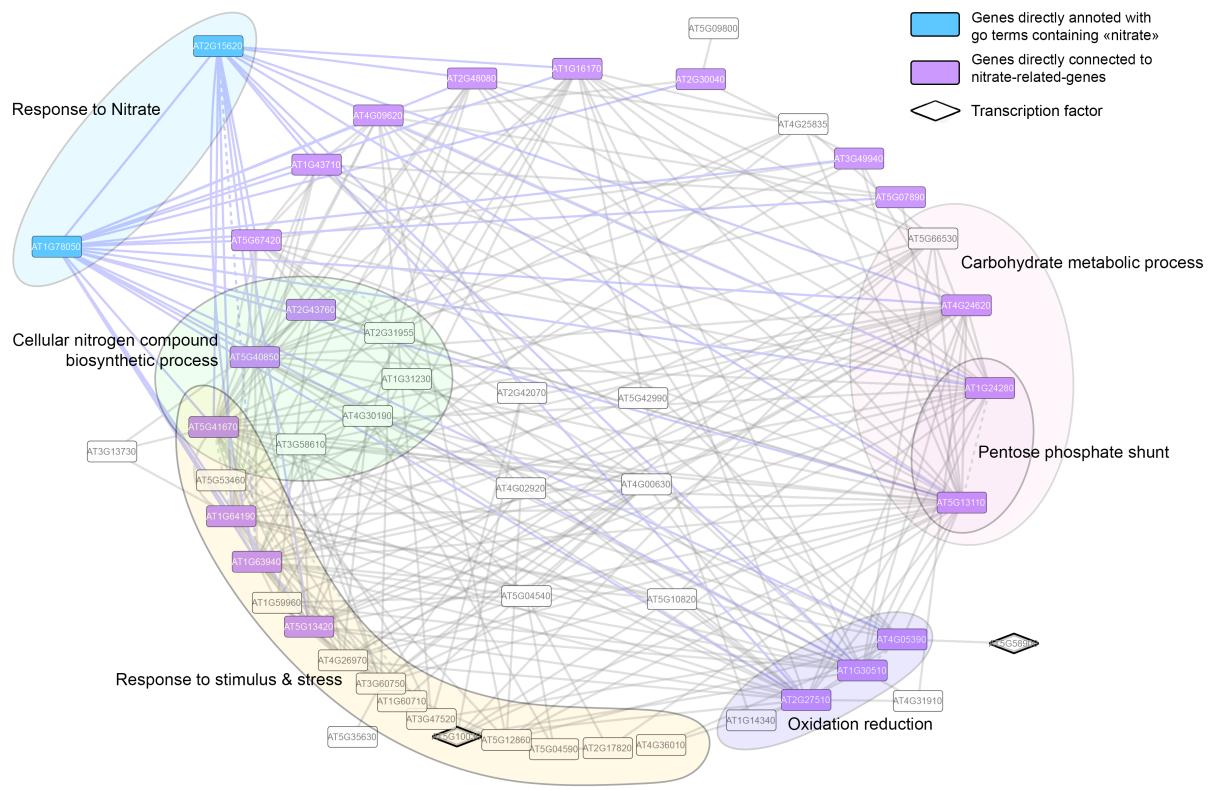
Les gènes en jaune sont les gènes précédemment cités comme étant impliqués dans les processus biologiques liés au nitrate.

La *Figure 8* permet d'observer que les 2 gènes impliqués dans le transport et la réponse au nitrate sont **AT1G08090 (NRT2:1)** et **AT5G50200 (WR3)**. On observe entre eux une interaction de type protéine-proteine. De plus, ces deux gènes sont liés à un facteur de transcription **AT5G65210 (TGA1)** qui lui est impliqué dans la réponse au stress et plus particulièrement à la résistance des maladies. Le gène **WR3** est quant à lui lié au gène **AT3G09810 (IDH-VI)** appartenant aux processus métaboliques de petites molécules.

La *Figure 9* met en évidence 2 gènes principaux de la réponse au nitrate : **AT1G78050 (PGM)** et **AT2G15620 (NIR1)**. Ces derniers sont directement associés à des gènes ayant différentes fonctions biologiques tel que les processus de biosynthèse de composés d'azote cellulaire, les réponses au stimulus et au stress de nouveau, dans les processus d'oxydo-reduction ainsi que dans le processus métabolique des glucides et en particulier la voie des pentoses phosphate.



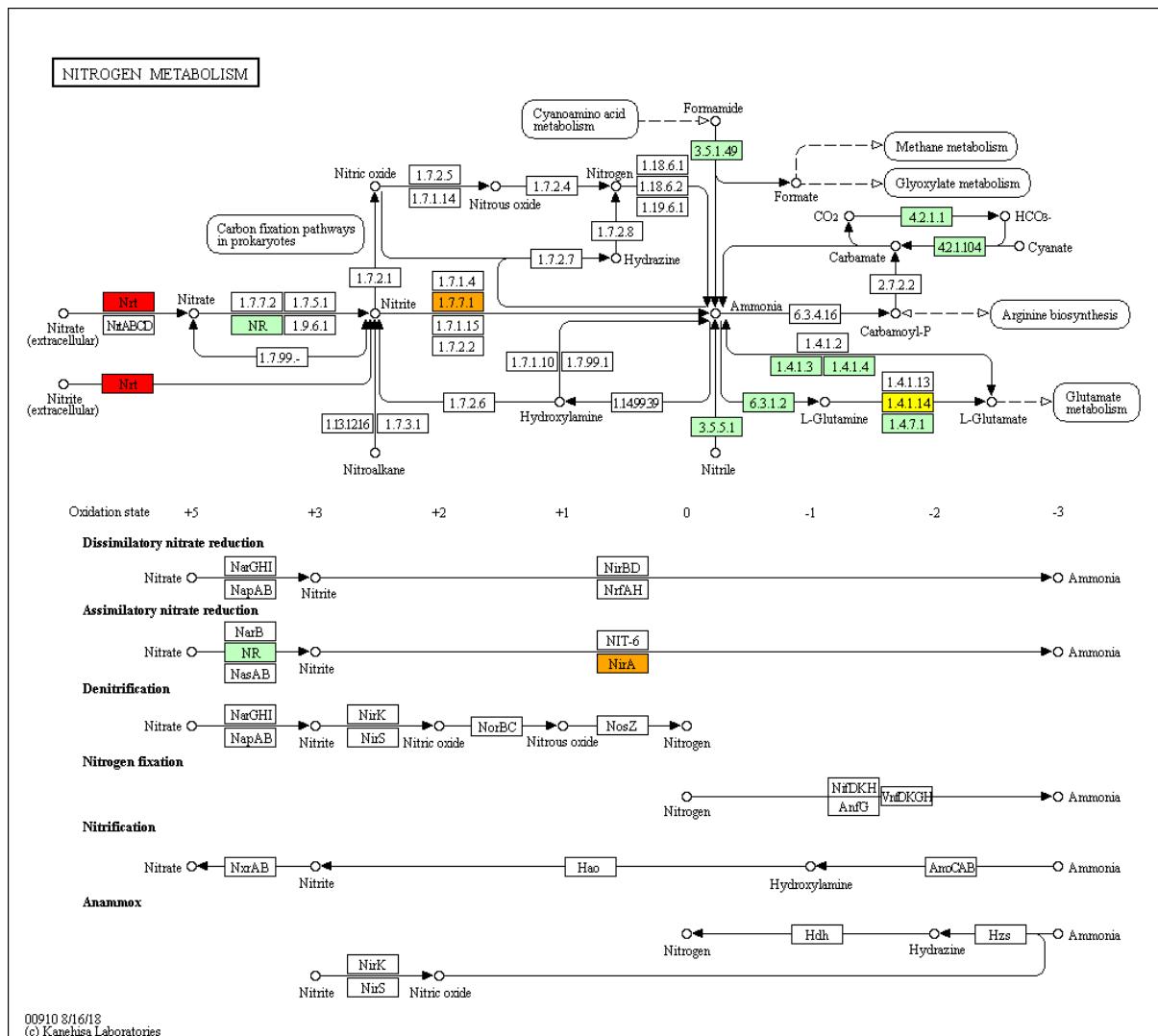
**Figure 8 – Cluster 3 annoté**



**Figure 9 – Cluster 2 annoté**

La *Figure 10* permet d'observer que les trois gènes cités précédemment font partit du métabolisme du nitrogène (azote). Tout d'abord NRT2:1 joue un rôle en tant que transporteur du nitrate et permet donc l'absorption du nitrate extra-cellulaire par la plante.

Dans un second temps, NIR1 permet d'assimiler le nitrate en le synthétisant en ammoniac par un phénomène de réduction. Enfin, GLT1 permet la transformation de la L-Glutamine, issue de l'ammoniac, en L-Glutamate.



**Figure 10 – Métabolisme de l'azote : GLT1(jaune), NIR1(orange) et NRT2 :1(rouge)**

# Discussion & Conclusion

## Discussion

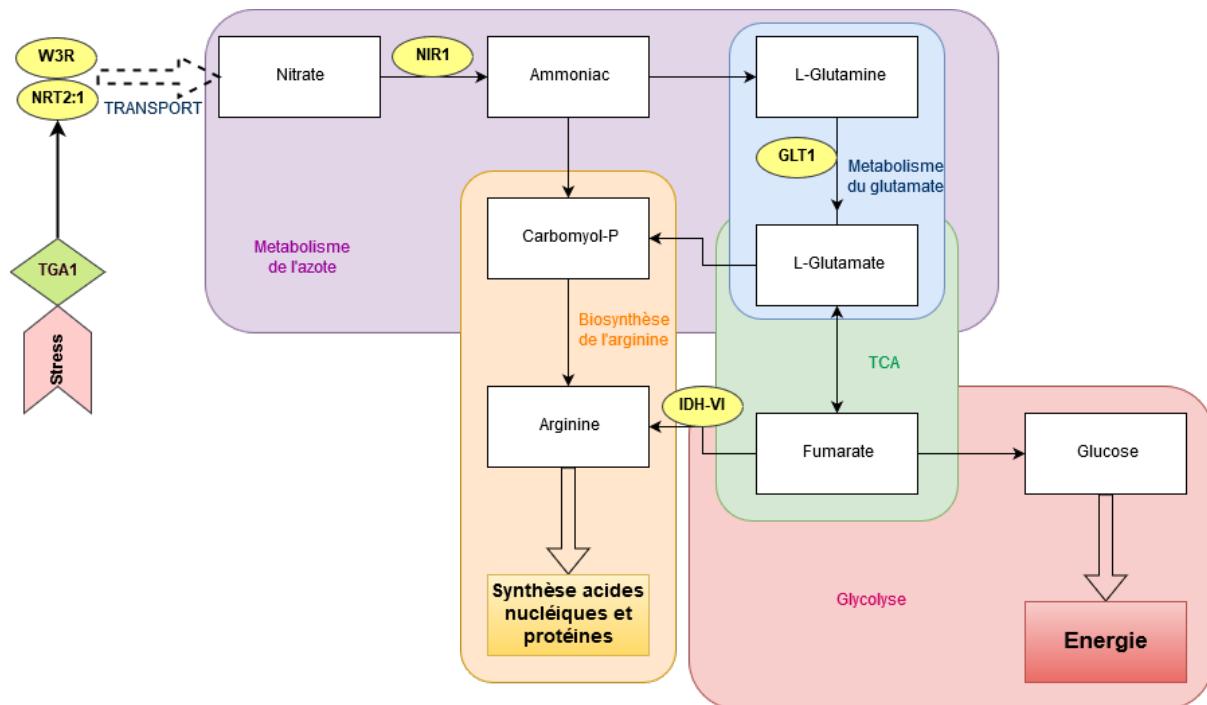
Les résultats obtenus ont permis de dégager plusieurs choses. Tout d'abord, les gènes NRT2:1 et W3R sont des gènes ayant une interaction protéine-protéine et permettant le transport du nitrate du compartiment extracellulaire vers le compartiment intracellulaire. Si le nitrate vient à manquer et donc provoque un stress pour la plante, le facteur de transcription TGA1 régule les quantités du transporteur NRT2:1 afin de palier à ce manque [11].

Le nitrate ainsi transporté dans la cellule va pouvoir, grâce au gène NIR1 synthétiser de l'ammoniac. Ce dernier peut, à son tour synthétiser deux molécules : le L-glutamine ou le carbomyol-P.

Le L-glutamine va pouvoir à son tour synthétiser du L-glutamate grâce au gène GLT1. Ce dernier peut donc entrer dans le cycle du citrate (TCA) afin d'obtenir du fumarate qui pourra soit, grâce au gène IDH-VI, produire de l'arginine, soit entrer dans le cycle de la glycolyse afin de produire du glucose et donc de l'énergie.

Le carbomyol cité précédemment peut, lui aussi permettre la biosynthèse de l'arginine. L'arginine est une molécule indispensable pour la plante puisqu'elle permet la synthèse d'acides nucléiques et donc de protéines.

Le diagramme de la *Figure 11* permet d'avoir une vue d'ensemble de ces résultats.



**Figure 11** – Diagramme réalisé à partir des résultats obtenus

Le losange vert représente un facteur de transcription, les cercles jaunes sont les gènes, les carrés blancs sont les molécules impliquées et les carrés arrondis sont les différents pathways où ces molécules se retrouvent.

## Conclusion

La construction d'un réseau de co-expression, à partir des données de transcriptomiques (*cf. Annexes*), et sa représentation à l'aide du logiciel Cytoscape ont permis de mettre en évidence l'implication de plusieurs gènes dans les mécanismes qui permettent la réponse au nitrate chez la plante.

Des gènes sont impliqués dans le transport du nitrate (**NRT2:1 et W3R**), et d'autres, comme **NIR1, GLT1 et IDH-VI**, dans la transformation du nitrate en molécules utiles pour d'autres mécanismes tels que la glycolyse ou la synthèse de protéines.

Les résultats restent hypothétiques, des expériences en biologie moléculaire pourraient venir corroborer ces résultats. Il est possible de vérifier l'effet d'un ou plusieurs gènes en inactivant les gènes trouvés comme étant impliqués dans la réponse au nitrate à l'aide de la technique du *Knock-out* (KO).

## Références

- [1] B.-M. Chen, Z.-H. Wang, S.-X. Li, G.-X. Wang, H.-X. Song, and X.-N. Wang, “Effects of nitrate supply on plant growth, nitrate accumulation, metabolic nitrate concentration and nitrate reductase activity in three leafy vegetables,” *Plant Science*, vol. 167, pp. 635–643, Sep 2004.
- [2] J. Hewitson, “The effects of nitrate on plant growth,” Oct 2018. <http://www.saps.org.uk>.
- [3] J.-F. Morot-Gaudry, M. Orsel, C. Diaz, F. Daniel-Vedele, and C. Masclaux-Daubresse, “Absorption et assimilation du nitrate et recyclage de l’azote organique chez les plantes : intérêt pour le colza,” *Ol. Corps Gras, Lipides*, vol. 13, pp. 393–402, Nov 2006.
- [4] J. Canales, T. C. Moyano, E. Villarroel, and R. A. Gutiérrez, “Systems analysis of transcriptome data provides new hypotheses about *Arabidopsis* root response to nitrate treatments,” *Front. Plant Sci.*, vol. 5, Feb 2014.
- [5] R Core Team, *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [6] M. Kanehisa and S. Goto, “Kegg : Kyoto encyclopedia of genes and genomes,” *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [7] P. Shannon, A. Markiel, O. Ozier, N. Baliga, J. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, “Cytoscape : A software environment for integrated models of biomolecular interaction networks,” *Genome Res.*, 2013.
- [8] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, and J. Galon, “ClueGO : a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks,” *Bioinformatics*, vol. 25, pp. 1091–1093, Apr 2009.
- [9] S. Maere, K. Heymans, and M. Kuiper, “BiNGO : a Cytoscape plugin to assess over-representation of gene ontology categories in biological networks,” *Bioinformatics*, vol. 21, pp. 3448–3449, Aug 2005.
- [10] J. H. M. D. J. S. F. M. Gang Su, Allan Kuchinsky, “Glay : community structure analysis of biological networks,” *Bioinformatics*, vol. Volume 26, no. Issue 24, p. 3135–3137, 5 December 2010.
- [11] L. Wang and P. R. Fobert, “Arabidopsis clade I TGA factors regulate apoplastic defences against the bacterial pathogen *Pseudomonas syringae* through endoplasmic reticulum-based processes,” *PLoS One*, vol. 8, p. e77378, Sep 2013.

## Annexes

### Origine des données de transcriptome

Publication	Titre de l'expérience	Identifiant (base de données)
Gutierrez (2007)	Transcription profiling by array of <i>Arabidopsis</i> grown in nutrient solutions with various concentrations of nitrate and sucrose.	E-MEXP-828 (ArrayExpress)
Krouk (2010)	High resolution $\text{NO}_3^-$ response of <i>Arabidopsis</i> roots	GSE20044 (GEO)
Ju (2009)	Expression data of 10-day-old wild-type and <i>chl1-5</i> plants exposed to 25 mM nitrate for 0 or 0.5 h	GSE9148 (GEO)
Ruffel (2011)	A systemic view of coordinated root responses to $\text{NO}_3^-$ heterogeneous environment in <i>Arabidopsis</i>	GSE22966 (GEO)
Patterson (2010)	Comparison of root transcriptomes in <i>Arabidopsis thaliana</i> plants supplied with different forms of inorganic nitrogen	GSE29589 (GEO)
Vidal (2013)	Root nitrate response of Ws plants and <i>afb3-1</i> mutant plants.	GSE35544 (GEO)
Alvarez (2014)	Root nitrate response of Col-0 plants and <i>tga1/tga4</i> mutant plants	GSE43011 (GEO)
Wang (2003)	Treatment of <i>Arabidopsis</i> with low concentration of nitrate.	Exp479 (iPlant Collaborative)
Wang (2004)	WT vs. <i>Nitrate Reductase</i> null mutant high nitrate concentration treatment.	Exp480 (iPlant Collaborative)
Wang (2007)	<i>Arabidopsis</i> treated with nitrite and nitrate	Exp481 (iPlant Collaborative)

# Script R

```
1 # Lecture de la table qui donne 146 mesures pour une gène donné
2 d'Arabidopsis Thaliana (AT)
3 data=read.table("P_EXPR.txt", header=T)
4
5 # Lecture de la liste des gènes d'intérêt et chargement dans un vecteur
6 focus<-read.table("P_GOI.txt", header=TRUE)
7 list_focus <-as.vector(focus[, "id"])
8 list_focus
9
10 # Filtrage des données d'expression correspondant aux gènes d'intérêt
11 filter_list <- data[,"id"] %in% list_focus
12 data_focus<-data[filter_list,]
13 data_focus
14
15 # Nécessaire de décaler la première colonne pour récupérer les labels des
16 lignes
17 rownames(data_focus)
18 # Les labels que l'on souhaite sont dans la première colonne
19 data_focus[,1]
20 rownames(data_focus)=data_focus[,1]
21 data_focus
22 data_focus=data_focus[,-1]
23 data_focus
24 data_focus=t(data_focus)
25 data_focus
26
27 # calculer la corrélation entre toutes les colonnes d'une matrice
28 #(nécessaires de donner en paramètres une datafram)
29 result_correlation=cor(data_focus)
30 result_correlation
31
32 #Utiliser la fonction melt pour convertir une matrice en liste d'interaction
33 library("reshape2")
34 list_correlation=melt(result_correlation)
35 list_correlation
36
37 #seuil à 0.85 afin de récupérer des corrélations linéaires forte entre les
38 variables
39 list_correlation=list_correlation[abs(list_correlation["value"])>0.85,]
40
41 #Il est nécessaire de supprimer les paires redondantes (on est parti d'une
42 matrice symétrique)
43 #Création d'une nouvelle colonne Alphabétique pour stipuler si le nom
44 du gène A est ordonné selon l'alphabet par rapport au nom du gène B
45 list_correlation["Alphabétique"]<-
46 as.character(list_correlation[,"Var1"])<as.character(list_correlation[,"Var2"]
47 "])
48 list_correlation
49
50 #Suppression des lignes où l'ordre de classement entre les deux noms de
51 gènes ne suit pas l'ordre alphabétique (permet de supprimer les
52 doublons)
53 list_correlation=list_correlation[list_correlation[,4]==TRUE,]
54 list_correlation
55
56 #Suppression de la colonne transitoire alphabétique - n'est plus
57 nécessaire-
58 list_correlation=list_correlation[,-4]
59 #Supression de la colonne Value
60
61 #Ajout de la colonne Interaction pour donner le type, ici ce sont les paires
62 dont l'expression est corrélée.
```

```

63 #Boucle for pour affecter un type de corrélation à chaque interaction
64 for(i in 1:nrow(list_correlation)){
65   if(list_correlation$value[i] < 0){
66     list_correlation$type[i] = "anti-correlated"
67   }else{
68     list_correlation$type[i] = "correlated"
69   }
70 }
71 list_correlation
72
73 #Chargement des données PPI
74 data_PPI=read.table("P_PPI.txt", sep="\t", header=T)
75 data_PPI
76
77 #Changement du nom des deux premières colonnes pour qu'elles
78 correspondent dans les deux listes
79 colnames(list_correlation)[1]="id1"
80 colnames(list_correlation)[2]="id2"
81 list_correlation
82
83 #echanger l'ordre des colonnes pour garder les values
84 (list_correlation<-list_correlation[,c(1,2,4,3)])
85 colnames(data_PPI)[3]="type"
86 data_PPI$value="unknown"
87 data_PPI
88
89 #Extraction des noms de gènes de la liste de co-expression pour les
90 utiliser comme filtre sur les PPI (on ne s'intéresse pas aux PPI des gènes
91 pour lesquelles pas de données d'expression)
92 id1<-as.vector(list_correlation[,"id1"])
93 id2<-as.vector(list_correlation[,"id2"])
94 list_net_genes <- c(id1,id2)
95 list_net_genes <- unique(list_net_genes)
96 list_net_genes
97 PPI_filtered<- data_PPI[,"id1"] %in% list_net_genes & data_PPI[,"id2"]
98 %in% list_net_genes
99 PPI_filtered<-data_PPI[PPI_filtered,]
100
101 #Fusion des deux listes pour créer un nouveau data frame et le
102 sauvegarder dans un fichier
103 LISTE=rbind(list_correlation,PPI_filtered)
104 LISTE
105 write.table (LISTE,"LISTE085.txt", sep="\t",row.names=FALSE,quote=F)

```