



HAL
open science

Comparaison de réseaux de gènes pour explorer le rôle des transcrits anti-sens

Marc Legeay, Béatrice Duval

► **To cite this version:**

Marc Legeay, Béatrice Duval. Comparaison de réseaux de gènes pour explorer le rôle des transcrits anti-sens. Grands Graphes et Bioinformatique - Atelier de la 16ème Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC'16), Jan 2016, Reims, France. hal-02516463

HAL Id: hal-02516463

<https://univ-angers.hal.science/hal-02516463>

Submitted on 23 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparaison de réseaux de gènes pour explorer le rôle des transcrits anti-sens

Marc Legeay^{*,**}, Béatrice Duval^{*}

^{*}LERIA - Université d'Angers - UNAM, 2 bd Lavoisier, 49045 Angers FRANCE
{prénom}.{nom}@univ-angers.fr

^{**}Institut de Recherche en Horticulture et Semences (IRHS),
UMR1345 INRA-Université d'Angers-AgroCampus Ouest, Centre Angers-Nantes,
42 rue Georges Morel - BP 60057, 49071 Beaucouzé FRANCE

Résumé. Un des problèmes clés en bioinformatique est de comprendre les mécanismes de régulation au sein d'une cellule. Notre travail concerne l'étude des réseaux de gènes chez le pommier, avec la particularité d'y intégrer les acteurs encore mal connus que sont les ARN anti-sens. Pour explorer l'impact des transcrits anti-sens, nous proposons ici la comparaison des deux réseaux obtenus par une méthode d'inférence très conservative. Nous pouvons ainsi étudier les interactions directes entre les gènes qui sont modifiées si l'on fait intervenir les transcrits anti-sens dans la méthode d'inférence. Un ensemble de motifs caractéristiques autour de ces modifications permet de révéler des ensembles d'acteurs sens et anti-sens intéressants.

1 Motivations et données

Afin de modéliser les mécanismes de régulation au sein d'une cellule, de nombreux travaux se sont intéressés à la construction de réseaux de gènes à partir de la mesure d'expression de leurs transcrits. Les transcrits anti-sens (AS) sont des molécules d'ARN endogènes dont la totalité ou une partie de leur séquence est complémentaire avec d'autres transcrits. Ainsi, l'ARN anti-sens peut s'hybrider avec l'ARNm ce qui entraîne une dégradation de l'ARN et donc l'inhibition du gène. Les différents mécanismes d'action des transcrits anti-sens ne sont pas encore complètement connus (Pelechano et Steinmetz, 2013), mais il apparaît que leur rôle a été largement sous-estimé jusqu'ici. Ainsi une étude récente sur le pommier a détecté une transcription anti-sens pour 65% des gènes exprimés, ce qui laisse envisager un fort potentiel de régulation par les anti-sens, notamment dans les processus de réponse à un stress. Nous nous intéressons donc à la construction de réseaux de gènes faisant intervenir ces nouveaux acteurs que sont les ARN anti-sens.

Les données de transcription dont nous disposons ont été obtenues grâce à une puce qui couvre l'ensemble des gènes codants prédits sur le pommier avec la particularité d'inclure, pour chaque locus identifié, une sonde sens et une sonde anti-sens, soit au total 126 022 sondes. Nous nous intéressons à un contexte biologique qui est celui de la maturation de la pomme pour lequel nous disposons de 22 échantillons mesurés au moment de la récolte, données H (pour Harvest) et 60 jours après la récolte, données 60DAH (60 Days After Harvest). Après

normalisation des données, nous avons identifié 931 sondes Sens (S) et 694 sondes anti-sens (AS) différemment exprimées avec 200 gènes ($S \cap AS$) pour lesquels à la fois le sens et l'anti-sens ont une expression significativement différente. Ces 1 625 gènes forment ce que nous appelons dans la suite *les gènes d'intérêt, Sens et Anti-Sens*.

Une étude fonctionnelle de cet ensemble de gènes d'intérêt a montré l'importance de prendre en compte les transcrits AS (Legeay et al., 2015). En effet, en comparant les termes de GO significativement représentés dans l'ensemble S à ceux significativement représentés dans $S \cup AS$, nous avons mis en évidence *les termes révélés par les anti-sens*, c'est-à-dire les fonctions biologiques qui ne seraient pas apparues dans cette analyse fonctionnelle sans la prise en compte des transcrits anti-sens. Une des fonctions les plus significatives ainsi révélée est *response to cold* (avec une p-valeur de 10^{-5}), or ce processus est clairement en jeu dans cette expérience puisque les pommes sont conservées au froid après la récolte.

2 Comparaison de Réseaux d'interactions

De nombreuses méthodes d'inférence de réseaux de gènes ont été proposées ces dernières années. C3NET (Conservative Causal Core Network) (Altay et Emmert-Streib, 2010) est une méthode qui s'appuie sur l'information mutuelle et une étape de maximisation pour ne retenir pour chaque gène g qu'un seul voisin, celui qui donne la plus forte information mutuelle avec g . Cette démarche permet donc de produire un « cœur de réseau » où seules les interactions directes entre deux gènes g_1 et g_2 apparaissent, les interactions indirectes résultant d'une interaction avec un troisième gène n'étant pas considérées.

Afin d'étudier le rôle des anti-sens dans les réseaux de régulation, nous proposons de comparer deux réseaux obtenus grâce à C3NET, le réseau R_S impliquant uniquement les acteurs sens de S et le réseau R_{SAS} impliquant les acteurs sens et anti-sens de $S \cup AS$. Notre comparaison a pour but d'identifier quelles interactions directes sont modifiées si on prend en compte les transcrits anti-sens¹. Même si l'information mutuelle entre 2 gènes est une mesure symétrique, l'algorithme utilisé par C3NET retourne d'abord un graphe orienté où chaque nœud se connecte à au plus un voisin ; pour cela, on calcule l'information mutuelle entre toutes les paires de gènes et on seuille la matrice obtenue pour ne retenir que les valeurs significatives ; puis on ne garde dans chaque ligne non nulle de la matrice que la valeur maximum, ce qui donne le voisin retenu dans cette ligne.

Lorsque nous intégrons les acteurs anti-sens, nous nous intéressons plus particulièrement aux sens qui se connectent à un anti-sens, car cela correspond à une liaison directe du réseau R_S qui est remplacée par une liaison sens vers anti-sens dans R_{SAS} .

Pour faire apparaître ces modifications, nous construisons le graphe G en ajoutant les arcs de R_S au réseau R_{SAS} . Dans une visualisation réalisée sous Cytoscape (Shannon et al., 2003), nous colorons les arcs de G suivant leur appartenance aux 2 réseaux : un arc est vert s'il est présent uniquement dans R_{SAS} , rouge s'il est présent uniquement dans R_S et gris s'il est présent dans les deux. Avec ce code couleur, une liaison directe du réseau R_S qui est remplacée par une liaison sens vers anti-sens dans R_{SAS} se traduit par un nœud sens dont partent un arc rouge et un arc vert (Figure 1a). Le graphe G obtenu pour les données de l'expérience 60DAH est présenté dans la Figure 1c.

1. La méthode DC3NET (Altay et al., 2011) ne permet pas de comparer des réseaux possédant des acteurs différents.

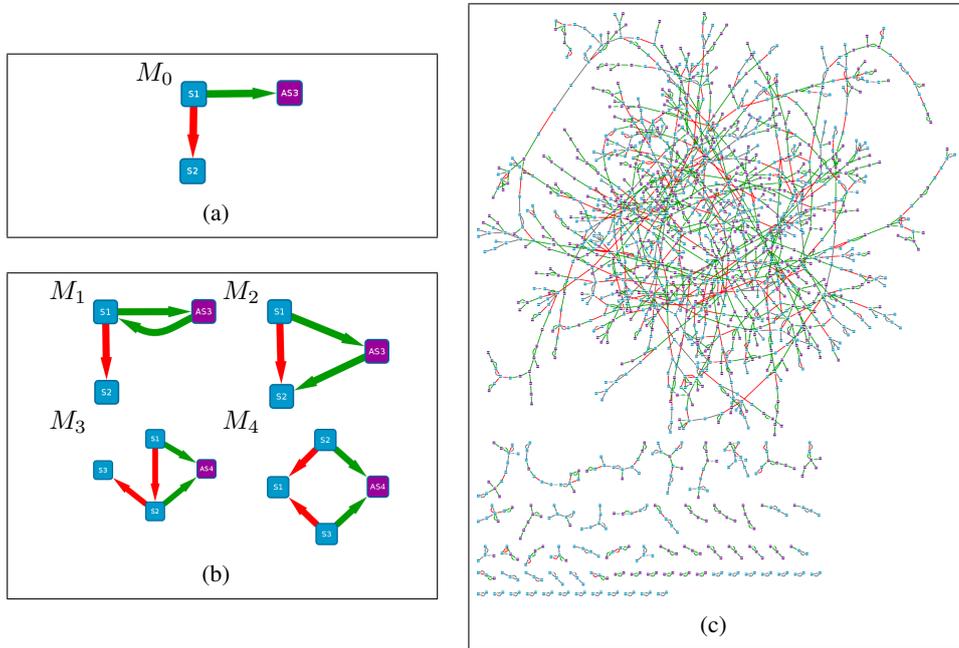


FIG. 1: Motifs et graphe de comparaison d'un réseau Sens avec un réseau Sens et Anti-Sens. Les nœuds bleus représentent les sens et les nœuds violets représentent les anti-sens. **(1a)** Schéma de modification d'une interaction directe. **(1b)** Motifs observés lorsqu'une interaction directe est modifiée. **(1c)** Graphe de comparaison entre les réseaux R_S et R_{SAS} pour l'expérience 60DAH.

Autour de ce schéma de modification d'interaction, nommé M_0 , nous observons dans le graphe des motifs plus riches représentés dans la Figure 1b. Le motif M_1 dénote un lien fort entre un sens et un anti-sens. Le motif M_2 révèle une liaison entre $S1$ et $S2$ qui se trouve être indirecte car elle fait intervenir l'acteur intermédiaire $AS3$. Le motif M_3 représente une « attraction » de l'anti-sens : deux sens qui étaient connectés se lient maintenant au même anti-sens. On peut observer une variante de ce motif où $S3$ est confondu avec $S1$. Le motif M_4 décrit les changements de connexions : deux sens sont toujours reliés à un même acteur, mais cet acteur qui était un sens dans R_S est maintenant un anti-sens dans R_{SAS} . On peut observer une variante de ce motif où $S3$ se connecte à $AS4$. Ce motif permet d'identifier les anti-sens qui ont une information mutuelle importante avec des sens.

Nous avons construit les graphes de comparaison pour les expériences H et 60DAH. La Table 1 dénombre les motifs présents dans chacun de ces graphes. Parmi les 931 nœuds de R_S , environ 40% subissent une modification d'interaction, et ce dans les deux expériences.

D'autre part, nous avons croisé l'ensemble de gènes ainsi isolé avec l'étude fonctionnelle menée précédemment (cf Section 1). La fonction `response to cold` est représentée par 26 sens et 37 anti-sens. La Table 1 dénombre également les motifs contenant au moins un de ces gènes, ainsi que les gènes présents dans les motifs. En comparant l'expérience H avec 60DAH, on observe que près de la moitié des gènes `response to cold` interviennent dans un motif

Comparaison de réseaux de gènes pour explorer le rôle des transcrits anti-sens

Expérience		M_0	M_1	M_2	M_3	M_4
H	# motifs	378	117	26	31	8
	# motifs response to cold	31	0	6	1	3
	# gènes response to cold	28	0	7	1	3
60DAH	# motifs	373	102	21	16	5
	# motifs response to cold	29	1	2	1	7
	# gènes response to cold	24	1	2	1	8

TAB. 1: Nombre de motifs, nombre de motifs contenant au moins un gène response to cold et nombre de ces gènes apparaissant dans les motifs, sachant qu'il existe 63 gènes response to cold et 931 nœuds dans R_S .

M_0 . Parmi eux, on observe 9 anti-sens qui sont présents dans M_0 à la fois dans H et 60DAH. De plus ces anti-sens ont la particularité d'avoir une expression qui diminue entre les deux expériences. L'idée ici est de combiner la comparaison des graphes avec l'étude fonctionnelle afin de soumettre à l'interprétation biologique des gènes ayant un comportement remarquable et associés à une fonction spécifique.

Références

- Altay, G., M. Asim, F. Markowetz, et D. E. Neal (2011). Differential C3net reveals disease networks of direct physical interactions. *BMC Bioinformatics* 12(1), 296.
- Altay, G. et F. Emmert-Streib (2010). Inferring the conservative causal core of gene regulatory networks. *BMC Systems Biology* 4(1), 132.
- Legeay, M., B. Duval, J.-P. Renou, et J. Bourbeillon (2015). Construction et Analyse de Réseaux de Gènes Contextuels dans le Domaine Végétal. Poster JOBIM, 6–9 Juillet 2015. www.inra.fr/jobim2015/Media/Fichier/Posters-PDF/Post-036.
- Pelechano, V. et L. M. Steinmetz (2013). Gene regulation by antisense transcription. *Nature Reviews Genetics* 14(12), 880–893.
- Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, et T. Ideker (2003). Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* 13(11), 2498–2504.

Summary

Understanding the regulation mechanisms in a cell is a key issue in bioinformatics. We study gene networks from apple data while integrating anti-sense transcripts. In order to explore the role of anti-sense transcripts in gene networks, we propose to compare two networks computed by a conservative reverse engineering method. Thus we can explore which direct interactions are modified when anti-sense transcripts are considered by the inference method. Specific graph patterns involving those modifications reveal interesting sense and anti-sense actors.