



HAL
open science

Analyse génomique de l'ADN extracellulaire du Root Extracellular Trap (RET) et caractérisations omiques des "root Associated Cap-Devrived Cells" (AC-DC) chez le soja Glycine max (L.) Merr.1917

Marie Chambard

► **To cite this version:**

Marie Chambard. Analyse génomique de l'ADN extracellulaire du Root Extracellular Trap (RET) et caractérisations omiques des "root Associated Cap-Devrived Cells" (AC-DC) chez le soja Glycine max (L.) Merr.1917. Sciences agricoles. Normandie Université, 2020. Français. NNT : 2020NORMR020 . tel-03787899

HAL Id: tel-03787899

<https://theses.hal.science/tel-03787899v1>

Submitted on 26 Sep 2022

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.



Normandie Université

THESE

Pour obtenir le diplôme de doctorat

Spécialité ASPECTS MOLECULAIRES ET CELLULAIRES DE LA BIOLOGIE - 420006

Préparée au sein de l'Université de Rouen Normandie

**Analyse génomique de l'ADN extracellulaire du
« *Root Extracellular Trap* » (RET) et caractérisations omiques des
« *root Associated Cap-Derived Cells* » (AC-DC)
chez le soja *Glycine max* (L.) Merr., 1917**

Présentée et soutenue par
Marie CHAMBARD

Thèse soutenue publiquement le 24 septembre 2020
devant le jury composé de

Mme Sabine CARPIN	MCF, HDR, Laboratoire de Biologie des Ligneux et des Grandes Cultures (LBLGC - EA 1207), Université d'Orléans	Rapporteur
M Olivier van WUYTSWINKEL	PR, Laboratoire de Biologie des Plantes et Innovation (BIOPI EA 3900) Université de Picardie Jules Verne	Rapporteur
Mme Jacqueline GRIMA-PETTENATI	Directrice de recherche CNRS, Laboratoire de Recherche en Sciences Végétales (LRSV - UMR 5546), Université de Toulouse	Examinatrice
M Azeddine DRIOUICH	PR, Laboratoire de Glycobiologie et Matrice Extracellulaire Végétale (Glyco-MEV - EA 4358), Université de Rouen Normandie	Examinateur
Mme Marie-Laure FOLLET-GUEYE	MCF, HDR, Laboratoire de Glycobiologie et Matrice Extracellulaire Végétale (Glyco-MEV - EA 4358), Université de Rouen Normandie	Directrice de thèse
Mme Isabelle BOULOGNE	MCF, Laboratoire de Glycobiologie et Matrice Extracellulaire Végétale (Glyco-MEV - EA 4358), Université de Rouen Normandie	Co-encadrante de thèse
Mme Isabelle TOURNIER	MCF, Responsable de la plateforme Rouennaise de génomique, Université de Rouen Normandie	Membre invitée
M Eric N'GUEMA ONA	Chef de projet, Centre Mondial de l'Innovation du Groupe Roullier, Saint Malo	Membre invité

Thèse dirigée par Dr. Marie-Laure FOLLET-GUEYE et Dr. Isabelle BOULOGNE, Laboratoire de Glycobiologie et Matrice Extracellulaire Végétale (Glyco-MEV – EA 4358), Université de Rouen Normandie.

Remerciements

Mes remerciements vont tout d'abord à Mme Sabine Carpin, Maître de conférences, HDR au laboratoire de Biologie des Ligneux et des Grandes Cultures (LBLGC – EA 1207) de l'Université d'Orléans et M Olivier van Wuytswinkel, Professeur au laboratoire de Biologie des Plantes et Innovation (BIOPI – EA 3900) de l'Université de Picardie Jules Verne, pour avoir accepté d'évaluer mes travaux de thèse en tant que rapporteurs.

Je remercie également les membres examinateurs et invités de mon jury pour avoir accepté de juger ce travail, Mme Jacqueline Grima-Pettenati, directrice de recherches CNRS au laboratoire de Recherche en Sciences Végétales (LRSV – UMR 5546) de l'Université de Toulouse, Mme Isabelle Tournier, Maître de conférence et responsable de la Plateforme Rouennaise de Génomique de l'Université de Rouen Normandie, M Éric N'Guema Ona, chef de projet au Centre Mondial de l'Innovation du Groupe Roullier, et M Azeddine Driouich, Professeur au laboratoire de Glycobiologie et Matrice Extracellulaire Végétale (GlycoMEV – EA 4358) de l'Université de Rouen Normandie.

Ces travaux de thèse ont été réalisés au sein du laboratoire de Glycobiologie et Matrice Extracellulaire Végétale (GlycoMEV – EA 4358) de l'Université de Rouen Normandie. Je remercie la région Normandie ainsi que la Structure Fédérative de Recherche Normandie Végétale (SFR NorVege) pour le financement de ce projet. Je remercie également l'école doctorale Normande de Biologie Intégrative, Santé Environnement (ED 497 - NBISE) grâce à laquelle j'ai pu suivre des formations à la recherche, et à l'enseignement par le biais de deux contrats de monitorats. Je remercie également l'ED NBISE pour avoir participé au financement de mes stages de formation à la recherche en Amérique du Nord réalisés au cours de ma thèse.

Je remercie également la coopérative agricole La Dauphinoise, pour nous avoir fourni les graines de soja nécessaires à ce projet, ainsi que le service de synthèse de peptides et criblage fonctionnel (SPCF) de la plateforme PRIMACEN de l'Université de Rouen Normandie pour nous avoir fourni le peptide PEP-13 utilisé dans ce projet.

Je remercie également les Prs Jean-Claude Mollet et Muriel Bardor, co-directeurs du laboratoire GlycoMEV, pour m'avoir accueillie au sein de leur laboratoire.

Mes remerciements vont ensuite à mes directrices et co-encadrantes de thèse, les Dr Marie-Laure Follet-Gueye et Dr Isabelle Boulogne, merci pour votre soutien sans failles au cours de ces trois années de thèse, merci de m'avoir fait confiance pour la réalisation de celle-ci, merci pour tout ce que vous m'avez appris lors de cette expérience, et surtout merci pour m'avoir donné l'occasion de me lancer dans ce défi qu'est la réalisation d'une thèse de doctorat.

Je remercie les Dr Alexis Carreras et Dr Marie-Laure Follet-Gueye pour m'avoir acceptée en stage au laboratoire GlycoMEV lorsque je n'étais qu'une jeune L3. Grâce à cette expérience, j'ai pu rencontrer le monde étrange de la recherche, apprenant ainsi l'Art du cassage de graines dans le respect de la statistique Champenoise.

Je remercie ensuite le Pr Azeddine Driouich, qui lorsque j'étais en Master 1, m'a recommandée auprès des Dr Romain Castilleux, Dr Marc Ropitiaux ainsi que le Dr Maïté Vicré qui m'ont ensuite acceptée en stage de Master 1 au sein du laboratoire GlycoMEV.

Je les remercie donc pour ce stage, qui m'a appris l'art de la culture du soja, de la biochimie, et de l'imagerie des zoospores et du RET, et m'a également révélé La Prophétie de l'Elue ! C'est d'ailleurs cette expérience qui m'a confortée dans l'idée de réaliser un jour un doctorat.

Je remercie une nouvelle fois le Pr Azeddine Driouich et le Dr Maïté Vicré, qui lorsque mon stage de Master 2 à leurs côtés à été annulé peu avant son commencement, m'ont vivement recommandée auprès du Dr Éric N'Guema Ona, chef de projet au CMI du Groupe Roullier.

Je remercie donc le Dr Éric N'Guema Ona pour m'avoir acceptée en stage de Master 2 pendant six mois au CMI du Groupe Roullier à Saint Malo, cette expérience fût très enrichissante, merci donc pour tout ce que tu m'as appris. Je remercie également Éric pour m'avoir soutenue et encouragée dans ma démarche de recherche de doctorat et de m'avoir recommandée auprès de mes directrices de thèse.

Parce que ces travaux de thèse n'auraient jamais pu aller aussi loin sans vous, mille mercis à toute mon équipe de fagotteuses, fagotteurs, trampouilleuses et trempouilleurs diplômés ou amateurs : Isabelle Boulogne, Carole Plasson, Marie-Laure Follet-Gueye, Sophie Bernard, Gaëlle Durambur, Aline Planchon, Mathilde Camiade, Angèle Tocqueville, Charlotte Toustou, Marcou Ropitoux, Quentin Arnaudin, Jesmilien Dilubenzi, ainsi qu'à mon armée de stagiaires pour une manip ou pour la vie (Lysa, Iman, Cassy, Baptiste, Rafael, Arthur), avec une mention spéciale à Lysa, ma première stagiaire, qui a accepté de revenir en stage de Master 1 avec moi, j'espère avoir été une bonne encadrante ! ;)

Un grand merci à Mohamed Amine Ben Mlouka et Pascal Cosette de la plateforme PISSARO pour leurs conseils et leur aide précieuse sur la partie protéomique de mon projet.

Merci également à Ludivine Beaussire, Céline Derambure, Sophie Coutant et Isabelle Tournier de la plateforme de Génomique Rouennaise pour m'avoir accueillie lors de mes manips ADNex, mais aussi pour m'avoir conseillé et aidé à avancer sur cette partie. Sans vous nous ne serions pas allées aussi loin dans ce projet.

Merci au Dr Pascale Beauregard de l'Université de Sherbrooke ainsi qu'à toute son équipe pour m'avoir accueillie dans votre laboratoire, pour m'avoir appris un peu de votre savoir-faire en microbiologie et m'avoir fait découvrir une petite partie du Canada.

Merci également au Dr Anjali Iyer-Pascuzzi de l'Université de Purdue ainsi qu'à toute son équipe pour m'avoir accueillie si chaleureusement dans votre laboratoire, grâce à vous j'ai découvert une petite partie de l'Indiana et mon nouveau logiciel préféré : Mapman !

Merci à Isabelle Boulogne pour avoir rendu ce voyage possible, je n'avais jamais voyagé seule aussi loin, grâce à toi j'ai découvert un continent et de nouveaux laboratoires avec un fonctionnement bien différent du nôtre. Ce voyage a été très enrichissant, merci de m'avoir donné cette chance.

Merci enfin à tous les membres du laboratoire GlycoMEV pour votre accueil, votre gentillesse et votre aide. Merci également à tous mes amis-collègues doctorants, post-docs ou techniciens du GlycoMEV : Aline, Marc, Mathilde, Sabine la Nantaise, Romain, Alexis, Coralie, Charlotte, Angèle, Rodolphe, Pierre-Louis, Jérémy, Yohana, Maxime, Jesmilien, Quentin... Votre grande sagesse et vos présences gracieuses et enjouées ont illuminé mon quotidien au cours de ces trois années que j'ai eu la chance inouïe de passer à vos côtés. Une mention

spéciale à mes co-occupants de bureau : Aline, Mathilde, Marc, Yohana et Romain (oui, oui, tu as partagé le bureau de l'Elue pendant UN mois !), merci pour ces moments musicaux, merci pour nos discussions, pour vos conseils emplies de plus ou moins sagesse et pour votre gentillesse.

Merci à l'équipe des sportives, à l'équipe des pauses thé, et à l'ensemble des membres de la Table Américano du Colbert, merci aux amateurs de Montbard et même à ceux qui trouvent ça dégou, merci pour tous ces bons souvenirs.

Ô Aline, voici tout un paragraphe dédié uniquement à ta majestuosité et à ta grandiosité, aucun mot ne pouvant égaler ta grandeur, puisse ce modeste éloge satisfaire tes exigences.

Ô Mathilde, voici tout un paragraphe dédié à ta grandiosité et à ta majestuosité, aucun éloge ne pouvant égaler ta personne, puissent ces modestes lignes combler ta demande. (PS : tu me dois une statue maintenant)

Quentin, tu es resté à mes côtés tout au long de la rédaction de ce manuscrit, tu m'as soutenue et consolée quand le stress et la fatigue s'accumulaient, un immense merci à toi. ♥

Pour finir, merci à ma famille, qui m'a toujours soutenue et a toujours cru en moi, je n'en serais pas là aujourd'hui sans tout cela.

Table des matières

Introduction générale	1
I. Contexte de l'étude	2
II. Les mécanismes de défense des plantes.....	3
II-1 Généralités	3
II-2 La notion d'éliciteurs.....	5
II-2-1 Les éliciteurs microbiens.....	6
II-2-2 Les éliciteurs endogènes ou DAMPs	8
II-2-2-1 Description de la paroi primaire des cellules végétales	9
II-2-2-2 Les DAMPs	14
II-3 Les mécanismes de défense racinaire.....	15
II-3-1 Organisation d'une racine en croissance chez les plantes angiospermes eu- dicotylédones	15
II-3-1-1 Les différents territoires cellulaires.....	15
II-3-1-2 Le méristème subapical de racine ou <i>Root Apical Meristem</i> (RAM)	17
II-3-2 Des initiales de la coiffe aux cellules frontières	18
II-3-2-1 La coiffe racinaire	18
II-3-2-2 Les cellules frontières ou <i>root-Associated, Cap-Derived Cells</i> (AC-DCs).....	19
II-3-3 Le RET et l'immunité racinaire	22
II-3-3-1 Spécificité de l'immunité racinaire	22
II-3-3-2 Le piège extracellulaire de racine ou <i>Root Extracellular Trap</i> (RET).....	23
II-3-3-3 L'implication du RET dans l'immunité racinaire	26
II-3-3-4 L'ADN extracellulaire	28
III. Le soja (<i>Glycine max</i> (L.) Merr.).....	35
III-1 La culture du soja	36
III-2 Les bioagresseurs et éliciteurs du soja	37
III-2-1 Les bioagresseurs du soja	37
III-2-2 La réponse de défense du soja	40

III-2-3 Informations sur le génome du soja et études « omiques »	42
III-2-3-1 Les études transcriptomiques de la racine du soja	42
III-2-3-2 Les études protéomiques de la racine du soja	46
IV. Problématique et objectifs.....	49
Matériel et méthodes.....	51
I. Matériel biologique et éliciteur.....	52
I-1 Stérilisation des graines	52
I-2 Mise en culture des graines	53
I-3 Récolte des AC-DC et de l' ADN extracellulaire (ADNex).....	54
II. Analyse transcriptomique des AC-DC et des racines	56
II-1 Récolte et séquençage de l'ARN des AC-DC et des racines.....	56
II-2 Analyse bio-informatique.....	57
III. Protéomique des AC-DC et racines	60
III-1 Extraction des protéines solubles des AC-DC et des racines	60
III-2 Analyse par spectrométrie de masse.....	62
III-3 Analyse bio-informatique.....	63
IV. Analyse génomique de l'ADN extracellulaire (ADNex).....	65
IV-1 Purification et quantification de l'ADNex	65
IV-2 Analyse qualité, préparation de bibliothèques et séquençage	65
IV-3 Analyse bio-informatique.....	66
Résultats	69
I. Effet d'une élicitation par PEP-13 sur le transcriptome et le protéome des AC-DC et des racines de soja	70
I-1 Analyses transcriptomiques	70
I-1-1 Identification des gènes différentiellement exprimés suite à l'élicitation	70
I-1-2 Analyses des processus biologiques et voies métaboliques impliquées dans la réponse à PEP-13.....	73
I-1-2-1 Analyse des voies métaboliques et des processus biologiques impliquant les gènes communs aux AC-DC et aux racines.....	74
I-1-2-2 Analyse des voies métaboliques et processus biologiques impliquant des gènes spécifiquement régulés chez les AC-DC ou chez la racine	77

I-1-3 Analyse détaillée des gènes différentiellement exprimés en réponse à PEP-13.....	84
I-1-3-1 Chez les AC-DC.....	84
I-1-3-2 Chez les racines	92
I-1-4 Tentative d'identification du récepteur de PEP-13.....	97
I-2 Analyses protéomiques.....	100
I-2-1 Identification des protéines différentiellement exprimées suite à l'élicitation	101
I-2-2 Analyse des processus biologiques impliquées dans la réponse à PEP-13.....	105
I-2-3 Analyse détaillée des protéines différentiellement exprimées en réponse à PEP-13	108
I-3 Comparaison des résultats obtenus entre l'analyse protéomique et transcriptomique..	112
II. Analyse génomique de l'ADNex du RET du soja.....	115
II-1 Analyse de l'alignement des séquences.....	115
II-1-1 Visualisation globale des alignements de séquences	115
II-1-2 Visualisation de l'alignement des séquences d'ADNex avec IGV	119
II-2 Analyse différentielle de l'alignement de séquences d'ADNex suite à l'élicitation.....	124
II-2-1 Analyse différentielle sur la totalité du génome <i>via</i> le logiciel DESq2.....	124
II-2-2 Analyse différentielle sur l'ADN chromosomique.....	131
II-2-3 Détection et identification des zones d'ADN chromosomique présentant un alignement plus important de séquences en conditions élicitée ou non élicitée.....	133
Discussion	137
I. Etude transcriptomique et protéomique des AC-DC et des racines du soja.....	138
I-1 Réponse à PEP-13 des AC-DC	138
I-2 Réponse à PEP-13 des racines.....	144
I-3 Réponses à PEP-13 des AC-DC et des racines	148
I-4 Identification du récepteur de PEP-13	151
II. Analyse génomique de l'ADNex du RET du soja.....	153
II-1 Différences de couverture entre l'ADN chromosomique et l'ADN des organites	153
II-2 Différences dans l'alignement des <i>reads</i> sur l'ADNchr	156
Conclusion et perspectives	159
Références bibliographiques	165
I. Annexe I : Etude des AC-DC et des racines du Lin (<i>Linum usitatissimum</i> L.)	196

I.	Matériels et méthodes	196
	I-1 Stérilisation des graines	196
	I-2 Mise en culture des graines	197
	I-3 Approches microscopiques	198
	I-3-1 Contre coloration à l'encre de chine.....	198
	I-3-2 Immunomarquage et microscopie à épifluorescence.	198
	I-3-3 Composés phénoliques et microscopie à épifluorescence.....	199
	I-4 Analyse transcriptomique	199
	I-4-1 Récolte du RET de lin	199
	I-4-2 Extraction de l'ARN et analyse transcriptomique	199
	I-5 Analyse protéomique	200
	I-5-1 Récolte des AC-DC et des racines	200
	I-5-2 Extraction des protéines solubles des AC-DC et des racines	200
	I-5-3 Analyse par spectrométrie de masse.....	202
	I-5-4 Analyse bio-informatique	202
II.	Résultats	204
	II-1 Caractérisation microscopique du RET de lin en condition de stimulation immunitaire ou non.....	204
	II-1-1 Contre-coloration du mucilage du RET à l'encre de chine.....	204
	II-2 Caractérisation transcriptomique des AC-DC de lin.....	206
	II-3 Caractérisation protéomique du RET de Lin	208
III.	Discussion	209
IV.	Conclusion et perspectives.....	210
II.	Annexe II : Etude du transcriptome des AC-DC de racines d' <i>Arabidopsis thaliana</i> (L.) Heynh	214
	I. Matériels et méthodes	214
	I-1 Stérilisation des graines.....	214
	I-2 Mise en culture des graines.....	215
	I-3 Récolte de l'ARN des AC-DC.....	216
	I-4 Extraction et quantification de l'ARN des AC-DC	216
	I-5 Analyse qualité de l'ARN	216

II.	Séquençage.....	217
III.	Analyses bio-informatiques	218
IV.	Résultats préliminaires	220
	IV-1 Analyse de la variance entre les différents échantillons.....	220
	IV-2 Analyse des gènes spécifiques aux différents types cellulaires	221
	IV-3 Analyse des voies spécifiques aux AC-DC et aux cellules de la columelle.....	224
	IV-4 Analyse Mapman de l'expression des gènes des AC-DC et des cellules de la columelle .	226
V.	Conclusion et perspectives	230
III.	Annexe III : Figures supplémentaires.....	232
I.	Transcriptomique des AC-DC et des racines	232
	I-1 Gènes sous régulés chez les AC-DC.....	232
	I-2 Gènes sous régulés chez les racines	242
I.	Analyse génomique de l'ADNex	244
	II-1 Alignement des séquences	244
	II-2 Observation de la couverture des chromosomes avec IGV	248
	II-3 Graphiques d'alignement des <i>reads</i> sur les chromosomes et les organites (réplicat 7) ...	267
	Résumé	290
	Abstract.....	290

Liste des figures

Figure 1 : Schéma présentant les mécanismes impliqués dans la PTI (<i>Pattern-triggered immunity</i>)	4
Figure 2 : Représentation schématique de la reconnaissance de différents MAMPs, DAMPs et HAMPs associés à leurs récepteurs membranaires	6
Figure 3 : Schéma de la composition polysaccharidique de la paroi végétale	9
Figure 4 : Synthèse et structure schématique de la cellulose	10
Figure 5 : Structure schématique des différentes hémicelluloses et présentation des enzymes impliquées dans leur synthèse	11
Figure 6 : Structure schématique des différents types de pectines	12
Figure 7: Structure schématique des <i>ArabinoGalactan Proteins</i> (AGP)	13
Figure 8 : Structure schématique des extensines	14
Figure 9: Les différents territoires cellulaires d'une racine en croissance d' <i>Arabidopsis thaliana</i>	16
Figure 10 : Schéma montrant l'organisation du RAM fermé chez <i>Arabidopsis thaliana</i>	17
Figure 11 : Structure de l'apex racinaire d' <i>Arabidopsis thaliana</i>	18
Figure 12 : Les différents morphotypes des cellules frontières, leur localisation et leur abondance chez le soja (<i>Glycine max</i> (L.) Merr.).....	21
Figure 13 : Schéma de la réponse de défense de la racine d' <i>Arabidopsis thaliana</i> et de ses variations en fonction des tissus racinaires considérés.	22
Figure 14 : Visualisation microscopique et représentation schématique du RET (<i>Root Extracellular Trap</i>).....	24
Figure 15 : Comparaison du RET (<i>Root Extracellular Trap</i>) et du NET (<i>Neutrophil Extracellular Trap</i>).....	25
Figure 16 : Représentation schématique de la formation lytique ou non-lytique du NET	29
Figure 17 : Tests d'infections de racines de pois par <i>N. haematococca</i> en présence ou non de DNase.....	31
Figure 18 : Observations microscopiques de l'ADNex du pois lors d'une infection par <i>R. solanacearum</i>	32

Figure 19 : Mécanismes pouvant être impliqués dans la reconnaissance de l'ADN extracellulaire comme DAMP chez les cellules végétales	33
Figure 20 : Dessin botanique, inflorescence et graines de soja (<i>Glycine max</i> (L.) Merr.)	35
Figure 21 : Les 7 plantes les plus cultivées dans le monde en 2017 (en tonnes)	36
Figure 22 : Exemples de symptômes de différentes maladies pouvant affecter le soja	39
Figure 23: Systèmes de stérilisation et de mise en germination des graines de soja	53
Figure 24 : Système de récolte du RET de soja en puits	54
Figure 25 : Tests de stérilité des échantillons de RET du soja	55
Figure 26 : Schéma du <i>workflow</i> utilisé pour l'analyse bio-informatique des données de RNAseq de racines et d'AC-DC du soja	57
Figure 27 : Score de qualité des bases séquencées	58
Figure 28 : Pourcentage de bases N par cycle	58
Figure 29 : <i>Workflow</i> utilisé pour l'analyse bio-informatique des gènes différentiellement exprimés chez les AC-DC et les racines en condition élicitée ou témoin	60
Figure 30 : Résultats du dosage protéique (a) et gels SDS-PAGE (b) pour les AC-DC (<i>root Associated Cap-Derived Cells</i>) et pour les racines de soja	61
Figure 31 : <i>Workflow</i> utilisé pour l'analyse bio-informatique des protéines différentiellement exprimées chez les AC-DC et les racines en condition élicitée avec PEP-13 ou non	64
Figure 32 : <i>Workflow</i> réalisé pour l'analyse bio-informatique de l'ADNex du RET du soja	66
Figure 33: Analyse de l'expression différentielle des gènes des AC-DC (<i>root-Associated Cap-Derived Cells</i>) de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE).....	71
Figure 34 : Analyse de l'expression différentielle des gènes des racines de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)	72
Figure 35 : Diagramme de Venn comparant les gènes différentiellement exprimés chez les racines et chez les AC-DC (<i>root-Associated Cap-Derived Cells</i>) de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE).....	73
Figure 36 : Processus biologiques (<i>Gene Ontology</i> , GO) des 64 gènes communément régulés en réponse à PEP-13 chez les AC-DC (<i>root-Associated Cap-Derived Cells</i>) et les racines de soja	75

Figure 37 : Processus biologiques (<i>Gene Ontology</i> , GO) des 20 gènes régulés de façon différente en réponse à PEP-13 chez les AC-DC (<i>root-Associated Cap-Derived Cells</i>) et les racines de soja	76
Figure 38 : Processus biologiques (<i>Gene Ontology</i> , GO) des 2 211 gènes spécifiquement régulés (684 gènes sur-exprimés et 1 527 gènes sous-exprimés) en réponse à PEP-13 chez les AC-DC (<i>root-Associated Cap-Derived Cells</i>) de soja	78
Figure 39 : Processus biologiques (<i>Gene Ontology</i> , GO) des 469 gènes (279 gènes sur-exprimés et 190 gènes sous-exprimés) spécifiquement régulés en réponse à PEP-13 chez les racines de soja	79
Figure 40 : Représentation générale (<i>overview</i>) des 1084 gènes différentiellement exprimés chez les AC-DC (<i>root-Associated Cap-Derived Cells</i>) de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)	86
Figure 41 : Représentation générale (<i>overview</i>) des 261 gènes différentiellement exprimés chez les racines de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)	93
Figure 42 : Diagramme de Venn comparant les gènes sur-exprimés chez les AC-DC (<i>root associated cap-derived cells</i>) et chez les racines de soja en réponse à PEP-13 et étant non annotés ou annotés comme récepteurs.....	97
Figure 43 : Prédiction de la conformation 3D de GLYMA20G100500	99
Figure 44 : Analyses en composante principale des protéomes des AC-DC et des racines de soja	100
Figure 45 : Analyse en composante principale de l'ensemble des protéomes élicités ou témoin des racines et des AC-DC de soja	101
Figure 46 : Analyse de l'expression différentielle des protéines des AC-DC (<i>root-Associated Cap-Derived Cells</i>) de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE).....	102
Figure 47 : Analyse de l'expression différentielle des protéines des racines de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)	103
Figure 48 : Diagramme de Venn comparant les protéines différentiellement exprimées chez les racines (13) et chez les AC-DC (31) de soja en conditions élicitée avec PEP-13 et témoin non élicitée	104
Figure 49 : Processus biologiques (<i>Gene Ontology</i> , GO) des 31 protéines régulées en réponse à PEP-13 chez les AC-DC (<i>root-Associated Cap-Derived Cells</i>) de soja	105
Figure 50 : Processus biologiques (<i>Gene Ontology</i> , GO) des 13 protéines régulées en réponse à PEP-13 chez les racines de soja	106

Figure 51 : Représentation générale (<i>overview</i>) des 16 protéines différentiellement exprimées chez les AC-DC (<i>root-Associated Cap-Derived Cells</i>) de soja entre les conditions élicitée avec PEP-13 (E) ou témoin non élicitée (NE)	109
Figure 52 : Représentation générale (<i>overview</i>) des 9 protéines différentiellement exprimées chez les racines de soja entre les conditions élicitée avec PEP-13 (E) ou témoin non élicitée (NE).....	110
Figure 53 : Diagrammes de Venn comparant les gènes et protéines différentiellement exprimés en réponse à PEP-13 chez les AC-DC et chez les racines	113
Figure 54 : Visualisation IGV de l’alignement des <i>reads</i> des différents échantillons d’ADNex de soja sur le chromosome GM18 en position 7 742 099 – 7 817 639	120
Figure 55 : Visualisation IGV de l’alignement des <i>reads</i> des différents échantillons d’ADNex de soja sur l’ADN de la mitochondrie en position 1 – 66 781	121
Figure 56 : Visualisation IGV de l’alignement des <i>reads</i> des différents échantillon d’ADNex de soja sur l’ADN des plastes en position 1 - 50 503.....	122
Figure 57 : Analyse en composante principale de la variance entre l’alignement des <i>reads</i> de l’ADNex de soja des différents échantillons sur l’ADN mitochondrial	125
Figure 58 : Analyse en composante principale de la variance entre l’alignement des <i>reads</i> de l’ADNex de soja des différents échantillons sur l’ADN plastidial	126
Figure 59 : Analyse en composante principale de la variance entre l’alignement des <i>reads</i> d’ADNex de soja des différents échantillons sur l’ADN chromosomique	127
Figure 60 : Analyse différentielle de l’alignement des <i>reads</i> d’ADNex de soja sur l’ADN mitochondrial	128
Figure 61 : Analyse différentielle de l’alignement des <i>reads</i> d’ADNex de soja sur l’ADN plastidial	129
Figure 62 : Analyse différentielle de l’alignement des <i>reads</i> d’ADNex de soja sur l’ADN chromosomique	130
Figure 63 : Distribution des <i>counts</i> d’ADNex de soja alignés sur le chromosome GM19 en condition non élicitée	133
Figure 64 : Distribution des <i>counts</i> d’ADNex de soja alignés sur le chromosome GM19 en condition élicitée	134
Figure 65 : Synthèse de l’observation de mécanismes activés en réponse à PEP-13 chez les AC-DC de racines de soja	143
Figure 66 : Synthèse de l’observation de mécanismes activés en réponse à PEP-13 chez les racines de soja	148

Figure 67 : Schéma des différentes classes de récepteurs connues ainsi que les domaines qui les composent	152
Figure 68 : Coloration de l'apex racinaire de soja au lugol en conditions élicitée avec PEP-13 et témoin non élicitée.....	156

Liste des tableaux

Tableau 1 : Récapitulatif des différentes études du transcriptome du soja.	45
Tableau 2 : Récapitulatif des différentes études du protéome du soja.	48
Tableau 3 : Analyse de la concentration et de la qualité des 12 échantillons d'ARN totaux des racines et des AC-DC du soja	56
Tableau 4: Dosage puis calcul de la concentration en ADNex des échantillons élicités (E) ou non (NE) du RET du soja	65
Tableau 5 : Rapport qualité du séquençage de l'ADNex du RET du soja généré par BWA.....	67
Tableau 6 : Processus biologiques représentés par les 1 527 gènes régulés négativement chez les AC-DC de soja en réponse à PEP-13.....	80
Tableau 7 : Mise en évidence des trente voies métaboliques les plus représentées chez les AC-DC de soja en condition élicitée (E) ou non (NE) grâce à l'annotation KEGG et l'outil BlastKOALA.	82
Tableau 8 : Mise en évidence des trente voies métaboliques les plus représentées chez les racines de soja en condition élicitée (E) ou non (NE) grâce à l'annotation KEGG et l'outil BlastKOALA.	83
Tableau 9 : Détail des 325 gènes sur-exprimés chez les AC-DC de soja en réponse à PEP-13 classés en fonction de leur catégorie de <i>bins</i>	87
Tableau 10 : Détail des 155 gènes sur-exprimés chez les racines de soja en réponse à PEP-13 classés en fonction de leur catégorie de <i>bins</i>	94
Tableau 11 : Gènes identifiés chez les AC-DC (<i>root associated cap-derived cells</i>) et les racines de soja comme potentiels récepteurs de PEP-13 grâce à DRAGO2.....	98
Tableau 12 : Processus biologiques représentés par les 26 protéines régulées négativement chez les AC-DC de soja en réponse à PEP-13.....	108
Tableau 13 : Détail des 16 protéines différentiellement exprimées chez les AC-DC de soja en réponse à PEP-13, classées en fonction de leur catégorie de <i>bins</i>	111
Tableau 14 : Détail des 9 protéines différentiellement exprimées chez les racines de soja en réponse à PEP-13, classées en fonction de leur catégorie de <i>bins</i>	111
Tableau 15 : Comparaison des processus biologiques pour les analyses transcriptomique et protéomique chez les AC-DC et les racines de soja impliqués dans la réponse à PEP-13 détectés lors de l'analyse WEGO2.0	114

Tableau 16 : Analyse du nombre de séquences de l'ADNex alignées (<i>counts</i>) sur le génome complet (chromosomes, mitochondrie et plaste) du soja pour le réplicat 7 NE (ADNex en condition non élicitée).....	117
Tableau 17 : Analyse du nombre de séquences d'ADNex alignées (<i>counts</i>) sur le génome complet (chromosomes, mitochondrie et plaste) du soja pour le réplicat 7E (ADNex en condition élicitée).....	118
Tableau 18 : Calcul du taux de couverture moyen de l'ADNex extrait en condition élicitée et non élicitée (Nombre de <i>reads</i> x longueur des <i>reads</i> / taille de la séquence) réalisé avec Excel	123
Tableau 19 : Identification de zones d'ADN chromosomique où les alignements de <i>reads</i> d'ADNex de soja présentent les <i>fold change</i> (FC) les plus importants en condition élicitée (E) et en condition non élicitée (NE).	132
Tableau 20 : Identification des zones de l'ADNchr pour lesquelles le nombre d'alignement des <i>reads</i> est supérieur à 100 <i>counts</i> pour tous les échantillons (7, 8 et 9 E et NE).....	135
Tableau 21 : comparaison de la réponse à PEP-13 des AC-DC et de la racine du soja aux réponses obtenues dans la bibliographie à des éliciteurs provenant de <i>Phytophthora sojae</i>	150

Liste des abréviations

AC-DC : root associated cap-derived cells

ADNex : ADN extracellulaire

AGP : arabinogalactane protéine

CC : coiled-coil

CDPK : Ca²⁺-dependant protein kinase

DAMP : damage-associated molecular pattern

DNaseq : DNA sequencing

ETI : effector-triggerred immunity

FAO : food and agriculture organization

FC : fold change

Flg22 : flagelline 22

GNK2 : ginkobilobin 2

GO : gene ontologies

HAMP : herbivore-associated molecular pattern

HRGP : hydroxyproline rich glycoprotein

ITAQ : Isobaric tag for relative and absolute quantitation

LC : liquid chromatography

LRR : Leucine-Rich Repeat

MALDI-TOF : matrix assisted LASER desorption ionization

MAPK : mitogen activated kinase

MAMP : microbe-associated molecular pattern

MER : zone méristématique

MLO : mildew resistance locus

MS : mass spectrometry

NAMP : nematode associated molecular pattern

NBS : nucleotide binding site

NET : neutrophil extracellular trap

OG : oligogalacturonides

PAMP : pathogen-associated molecular pattern

Pb : paires de bases

PEP-13 : peptide 13

PET111 : marqueur des cellules de la columelle

PGPR : plant growth promoting rhizobacteria

PTI : PAMP-triggered immunity

PR : Pathogenesis Related

PRR : pattern recognition receptor

QC : centre quiescent

RAM : root apical meristem

RET : root extracellular trap

RG-I : rhamnogalacturonanes de type I

RG-II : rhamnogalacturonanes de type II

RLK : receptor like kinase

RNAseq : RNA sequencing

ROS : reactive oxygen species

RPW8 : resistance to powdery mildew 8

RLCK : receptor-like cytoplasmic kinase

STK : serine threonine kinase

TF : transcription factor

TIR : toll/interleukine receptor

WER : marqueur des cellules latérales de la
coiffe

WOX5 : wuschel-related homebox 5 /
marqueur des cellules du centre quiescent

WR : whole root

Introduction générale

- I. Contexte de l'étude
- II. Les mécanismes de défense des plantes
 - II-1 Généralités
 - II-2 La notion d'éliciteurs
 - II-3 Les mécanismes de défense racinaire
- III. Le soja (*Glycine max* (L.) Merr.)
 - III-1 La culture du soja
 - III-2 Les bioagresseurs du soja
- IV. Problématique et objectifs

Introduction générale

I. Contexte de l'étude

La culture des végétaux représente aujourd'hui un intérêt économique majeur avec une production mondiale extrêmement importante (9 milliards de tonnes en 2017 selon la FAO STATS). Les végétaux sont en effet cultivés pour l'alimentation humaine et animale, garantissant ainsi la sécurité alimentaire mondiale. Ils sont aussi utilisés pour la fabrication de matériaux (bois, fibres, textiles...), de produits phytopharmaceutiques ou cosmétiques (Wilson, 2008). Au vu de l'importance de la production végétale pour l'économie mondiale, la protection des cultures est donc primordiale afin d'éviter l'insécurité alimentaire et les pertes économiques pouvant découler des baisses de rendement de la production végétale (Bandara *et al.*, 2020).

La protection des cultures est aujourd'hui encore assurée en majorité par l'utilisation de produits phytosanitaires de synthèse (insecticides, herbicides, fongicides...). Il a cependant été démontré que ceux-ci sont rémanents dans l'environnement et présentent un danger pour la santé des êtres vivants (Fenner *et al.*, 2013). Ces substances sont d'ailleurs désormais interdites à la vente pour les particuliers depuis le 1^{er} janvier 2019 (Loi Labbé, modifiée par la loi n° 2015-992 du 17 août 2015 - art. 68 (V)). Les jardiniers amateurs et les agriculteurs ne pouvant se passer de solutions de protection de leurs cultures, une politique de remplacement de ces produits de synthèse chimique par des produits plus respectueux de l'environnement et de l'humain a été engagée (Exemple du glyphosate : Sortir du glyphosate, Ministère de l'agriculture et ministère de la transition écologique et solidaire, 2020).

C'est dans ce contexte que les plans Ecophyto et Ecophyto II+ ont été lancés lors du Grenelle de l'Environnement de 2007, du plan d'actions de 2018 et du comité d'orientation stratégique et de suivi (COS) de 2020 (Ministère de l'agriculture et de l'alimentation, 2020a). Ces plans visent à réduire de 50% l'utilisation des phytosanitaires d'ici 2025, tout en promouvant leur remplacement par des méthodes dites de biocontrôle. Ces méthodes consistent en l'utilisation de mécanismes naturels permettant la protection des plantes et

comprenant l'utilisation d'organismes vivants ou de molécules d'origine végétale, animale ou minérale (Ministère de l'agriculture et de l'alimentation, 2020b).

Le biocontrôle regroupe un ensemble de techniques variées, telle que l'utilisation de macroorganismes dits « auxiliaires » (par exemple *Trichogramma brassicae* Bezdenko pour lutter contre la pyrale du maïs), de microorganismes (par exemple le Serenade ASO®, un biofongicide contenant la souche *Bacillus subtilis* QST 713) ou encore l'utilisation de substances naturelles comme les stimulateurs de défense (éliciteurs et potentialisateurs) qui vont activer les voies de défense des plantes (par exemple le Vacciplant®, un extrait de laminarine) (Ecophytopic, 2020).

Le développement de ces dernières méthodes nécessite cependant une bonne connaissance de l'ensemble des mécanismes impliqués dans la défense des plantes. Si ceux-ci ont été largement étudiés au niveau des parties aériennes (Jones and Dangl, 2006), les mécanismes de défense au niveau racinaire sont encore très peu connus (Chuberre *et al.*, 2018 ; Driouich *et al.*, 2013). Pourtant la racine est un organe stratégique pour le développement, la protection et la survie de la plante. Elle peut être également le siège de stockage par l'accumulation de substances de réserve, le lieu de symbioses et l'objet de l'attaque de nombreux bioagresseurs telluriques tels que des insectes, bactéries, champignons et oomycètes (Bais *et al.*, 2006 ; Egamberdieva *et al.*, 2008 ; Mendes *et al.*, 2011). Ainsi, afin de promouvoir le développement des méthodes de biocontrôle, une meilleure connaissance des mécanismes de défense racinaire est nécessaire, et en particulier ceux mis en place par les plantes d'intérêt agronomique et économique.

II. Les mécanismes de défense des plantes

II-1 Généralités

Une particularité importante de l'immunité végétale réside dans le fait que la majorité des cellules de l'organisme sont potentiellement capables de détecter la présence d'un agent pathogène. Cette détection est tout d'abord réalisée suite à la perception de molécules d'origine microbienne, les *Microbial-Associated Molecular Patterns* (MAMPs) ou d'origine endogène, les *Damage-Associated Molecular Patterns* (DAMPs) (Abdul Malik *et al.*, 2020). Les MAMPs et les DAMPs sont des éliciteurs (peptides, glycopeptides,

oligosaccharides, nucléotides, acides gras, etc.) détectés par les cellules végétales *via* des récepteurs transmembranaires de la membrane plasmique appelés *Pattern Recognition Receptors* (PRRs) (Henry *et al.*, 2012 ; Choi et Klessig, 2016 ; Hou *et al.*, 2019). Ce premier niveau de reconnaissance déclenche la mise en place de l'immunité primaire appelée *Pattern-Triggered Immunity* (PTI) et permet de protéger la plante contre la plupart des invasions pathogènes (Figure 1).

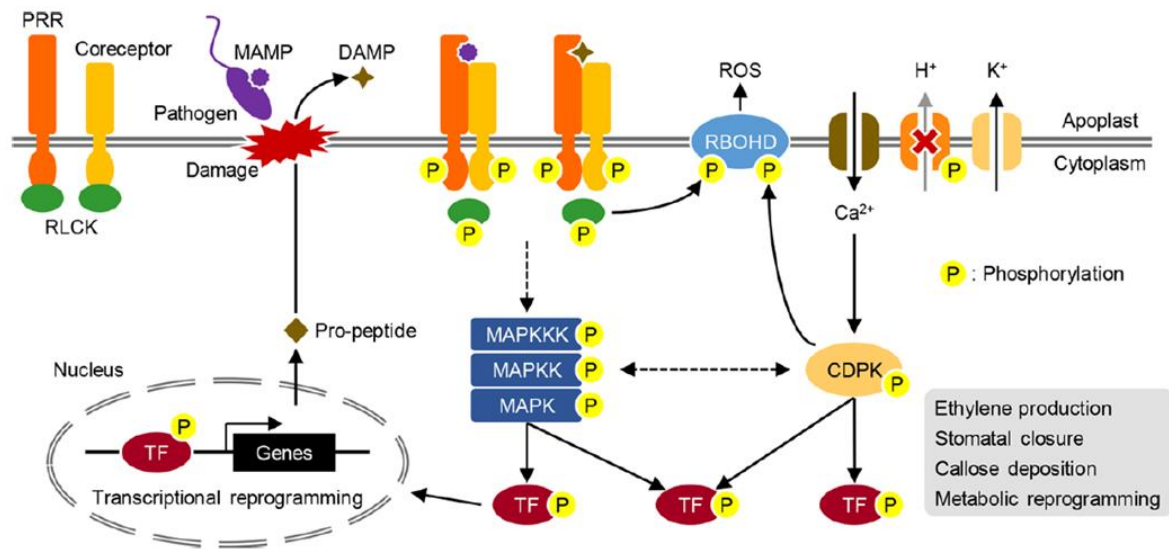


Figure 1 : Schéma présentant les mécanismes impliqués dans la PTI (Pattern-triggered immunity)

La détection des MAMPs (*Microbial-Associated Molecular Patterns*) ou DAMPs (*Damage-Associated Molecular Patterns*) par les récepteurs PRRs (*Pattern Recognition Receptors*) induit la mise en place de différents mécanismes comme l'activation de la voie des MAPKs (*Mitogen Associated Protein Kinases*), la production de ROS (*Reactive Oxygen Species*), et un influx d'ions Ca²⁺. Plus tardivement, on observe la fermeture des stomates, des dépôts de callose, ou encore l'activation de gènes de défense permettant la production d'éthylène. RLCK : *Receptor-Like Cytoplasmic Kinase* ; CDPK : *Ca²⁺-Dependant Protein Kinase* ; RBOHD : *Respiratory Burst Oxidase Homologue Protein D* ; TF : Facteur de transcription. Figure issue de l'étude de Saijo *et al.*, 2018.

La PTI peut plus tardivement se traduire également par la production de phytoalexines. Enfin, parmi les autres réponses cellulaires les plus tardives (de quelques heures à quelques jours) des dépôts de callose et des retards de croissance sont couramment observés (Boller et Felix, 2009 ; Millet *et al.*, 2010). Généralement la fermeture des stomates ainsi qu'une inhibition de la croissance caulinaire est observée après quelques jours d'exposition au pathogène (Yu *et al.*, 2017). Certains phytopathogènes contournent l'immunité primaire en sécrétant des effecteurs (facteurs de virulence) qui en s'opposant à la PTI favorisent l'infection. Ces facteurs de virulence de nature protéique sont parfois considérés comme des éliciteurs spécifiques. Des récepteurs cytoplasmiques, nommés protéines de résistance R ou

NB-LRR (*Nucleotide Binding-Leucine Rich Repeat*), peuvent alors être synthétisés. La reconnaissance des facteurs de virulence par ces nouveaux récepteurs intracellulaires déclenche un second niveau d'immunité ou immunité secondaire plus spécialisée : l'ETI (*Effector-Triggered Immunity*). Les réponses immunitaires primaire et secondaire se caractérisent par des réponses cellulaires semblables avec toutefois des intensités de réponse plus importantes lors de la réponse immunitaire spécialisée de type ETI. L'activation de ces deux réponses immunitaires permet donc, la plupart du temps, de résister à l'attaque du pathogène. Cependant, celles-ci ne sont efficaces que si le pathogène est détecté précocement par la plante, et cela grâce à la reconnaissance de ses éliciteurs associés. Les éliciteurs sont donc un élément majeur de l'immunité végétale.

II-2 La notion d'éliciteurs

Il existe une grande diversité d'éliciteurs dont le nom dépend de l'organisme d'origine (Figure 2). Ils sont appelés PAMPs (*Pathogen Associated Molecular Patterns*) pour les motifs moléculaires associés aux pathogènes, MAMPs (*Microbe Associated Molecular Patterns*) pour les motifs associés aux microorganismes en général qu'ils soient bénéfiques ou pathogènes, NAMPs (*Nematode Associated Molecular Patterns*) pour les molécules issues de nématodes (Choi et Klessig, 2016 ; Yu *et al.*, 2017) ou encore HAMPs (*Herbivore Associated Molecular Patterns*) pour les molécules associées aux herbivores (Bonaventure *et al.*, 2011). De plus, lors de l'attaque d'un pathogène, des molécules endogènes vont être libérées ou dégradées dans le milieu extracellulaire et vont être reconnues comme un signal de danger (Choi et Klessig, 2016 ; Yu *et al.*, 2017). Ces molécules sont appelées DAMPs (*Damage-Associated Molecular Patterns*). Enfin, d'autres molécules non issues de macro ou microorganismes pathogènes ou de la plante elle-même peuvent également présenter des propriétés élicitrices et induire des mécanismes de défense (Henry *et al.*, 2012). Parmi celles-ci, on retrouve notamment les composés analogues aux phytohormones, ou encore des extraits d'autres plantes ou d'organismes comme par exemple la laminarine issue de l'algue brune *Laminaria digitata* ou le chitosan extrait de la chitine contenue dans les carapaces de crustacés (Henry *et al.*, 2012).

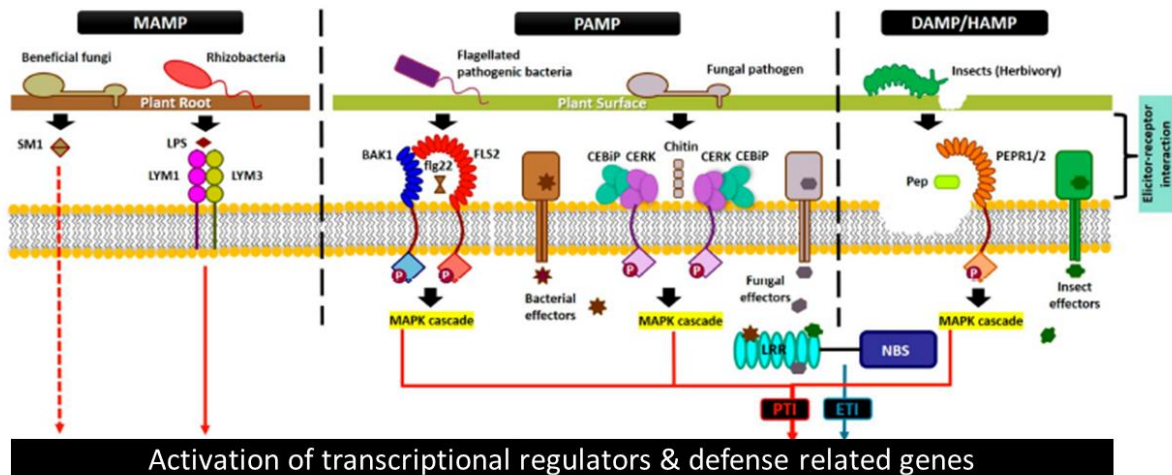


Figure 2 : Représentation schématique de la reconnaissance de différents MAMPs, DAMPs et HAMPs associés à leurs récepteurs membranaires

La reconnaissance des MAMPs (*Microbe Associated Molecular Patterns*), PAMPs (*Pathogen Associated Molecular Patterns*), DAMPs (*Damage-Associated Molecular Patterns*), HAMPs (*Herbivore Associated Molecular Patterns*) par différents récepteurs va activer la transcription de régulateurs et de gènes impliqués dans la réponse de défense de la plante. Cette reconnaissance peut induire l'activation d'une cascade de MAPKs (*Mitogen-Activated Protein Kinases*) et aboutir aux réponses immunitaires primaire PTI (*PAMP-Triggered Immunity*) et secondaire ETI (*Effector-Triggered Immunity*). Figure adaptée de Abdul Malik *et al.*, 2020.

En plus d'être d'origine diverse, les éliciteurs sont aussi de nature diverse. Il peut s'agir de fragments de polysaccharides (oligogalacturonides (OGs), chitine...) (Ridley *et al.*, 2001 ; Boller, 2005), de peptides (flagelline 22 (flg22), peptide 13 (PEP-13), peptidoglycannes, *etc.*) (Albert, 2013), de métabolites (acide fusarique, ...) (Plancot *et al.*, 2013) et même d'acides nucléiques (ADN) (Gómez-Gómez et Boller, 2002 ; Duran-Flores et Heil, 2017).

II-2-1 Les éliciteurs microbiens

Les éliciteurs microbiens peuvent provenir de différents types d'organismes : les bactéries, les virus, les champignons ou encore les oomycètes.

Parmi les éliciteurs bactériens, le plus connu est la flagelline 22 qui est un peptide de 22 acides aminés constitutif du flagelle des bactéries et qui est reconnu chez *Arabidopsis thaliana* par le PRR FLS2 (*Flagellin-Sensitive 2*) qui forme un complexe avec BAK1 (*BRI1-Associated Receptor Kinase 1*) (Chinchilla *et al.*, 2007 ; Boller et Felix, 2009). D'autres éliciteurs peptidiques bactériens peuvent aussi être cités comme csp22 (*cold shock protein 22*) qui est capable d'induire une réponse de défense chez les *Solanaceae* (Boller et Felix, 2009) et dont le récepteur potentiel NbCSPR (*Competence Stimulating Peptide Receptor*) a été décrit chez le tabac (Yu *et al.*, 2017) ; ou encore EF-Tu (*Elongation Factor Thermo*

unstable) ainsi que ses peptides associés elf18 et elf26 qui sont reconnus par le récepteur EFR (EF-Tu Receptor) (Boller et Felix, 2009 ; Yu *et al.*, 2017).

La paroi des cellules bactériennes peut elle aussi être reconnue par la plante. En effet chez *Arabidopsis thaliana* les peptidoglycanes (PNG) sont reconnus par les récepteurs LYM1 et LYM3 (Lysin Motif proteins) et les lipopolysaccharides (LPS) par le récepteur LORE (*Lipo-Oligosaccharide-specific Reduced Elicitation*) (Yu *et al.*, 2017).

La paroi des champignons est quant à elle reconnue par les plantes en raison de la chitine et des β -glucanes qu'elle contient. La chitine est ainsi détectée par les récepteurs LYK5 (*Lysine motif receptor like Kinase*) et CERK1 (*Chitine Elicitor Receptor Kinase 1*) chez *Arabidopsis thaliana* (L.) Heynh et par CEBiP (*Chitine Elicitor Binding Protein*) chez le riz. Le récepteur des β -glucanes n'est pas encore connu à ce jour (Boller et Felix, 2009 ; Yu *et al.*, 2017, Chaliha *et al.*, 2018). Des enzymes produites par les champignons pour la dégradation de la paroi des plantes peuvent aussi être des molécules élicitrices. En effet, par exemple, la xylanase EIX (*Ethylene Inducing Xylanase*) est détectée par les récepteurs LeEIX1 et LeEIX2 chez la tomate (*Solanum lycopersicum* L.) (Yu *et al.*, 2017). Une endopolygalacturonase (PG) provenant de *Botrytis cinerea* Persoon est également capable d'activer la PTI chez la vigne (Poinssot *et al.*, 2003) et son récepteur associé semble être RLP42 (*Receptor-Like Protein 42*) chez *A. thaliana* (Yu *et al.*, 2017).

Les oomycètes sont eux aussi détectés par les plantes. Parmi les éliciteurs issus de ces microorganismes, on retrouve :

- le peptide nlp20 (*necrosis and ethylene-inducing peptide*) qui est reconnu par le récepteur RLP23 chez *A. thaliana* (Yu *et al.*, 2017),
- l'élicitrine INF1 (*Phytophthora INFestans 1*) qui est reconnu par le récepteur ELR (*Elicitin Response*) décrit chez la pomme de terre (Yu *et al.*, 2017),
- les glucanes, l'ergosterol ou la *cellulose-binding lectin* issus de leur paroi (Boller et Felix, 2009) et dont les mécanismes de reconnaissance sont encore inconnus,
- la glycoside-hydrolase XEG1 (*Xyloglucan-specific Endo- β -1,4-Glucanase*) (Yu *et al.*, 2017), dont le PRR associé est toujours inconnu,
- le peptide de 13 acides aminés, PEP-13 provenant d'une transglutaminase des oomycètes du genre *Phytophthora* (Albert, 2013).

Bien que présentant un intérêt majeur pour de nombreuses cultures (dont fait partie le soja), le récepteur de PEP13 est encore inconnu (Brunner *et al.*, 2002 ; Halim *et al.*, 2004 ; Wang *et al.*, 2018). L'effet éliciteur de ce peptide a été pour la première fois mis en évidence sur des cultures cellulaires de persil (*Petroselinum crispum* (Mill.) Fuss) à partir d'extraits de *Phytophthora sojae* Kaufm. & Gerd. (Parker *et al.*, 1991). Ces travaux se sont inspirés de l'étude de Parker *et al.* en 1988 qui avaient mis en évidence l'activité élicitrice de fractions de la paroi de ce même oomycète sur le persil et le soja. PEP-13 a ensuite été étudié à plusieurs reprises sur le persil (Nürnbergger *et al.*, 1994 ; Diekmann *et al.*, 1994 ; Jabs *et al.*, 1997 ; Zimmermann *et al.*, 1997 ; Ligterink *et al.*, 1997 ; Nennstiel *et al.*, 1998 ; Brunner *et al.*, 2002) et sur la pomme de terre (*Solanum tuberosum* (L.)) (Brunner *et al.*, 2002 ; Halim *et al.*, 2004 ; Wang *et al.*, 2019).

II-2-2 Les éliciteurs endogènes ou DAMPs

Une caractéristique majeure des cellules végétales, est la présence d'une paroi cellulaire ou matrice extracellulaire végétale (Figure 3). Cette paroi est une composante essentielle de la défense des plantes puisque c'est la première structure rencontrée par les pathogènes. Elle est connue pour s'adapter aux différents stress biotiques (Plancot *et al.*, 2013) ou abiotiques (Wang *et al.*, 2018) auxquels la plante va être soumise. De plus, les fragments de polysaccharides pariétaux sont une source importante d'éliciteurs endogènes ou DAMPs.

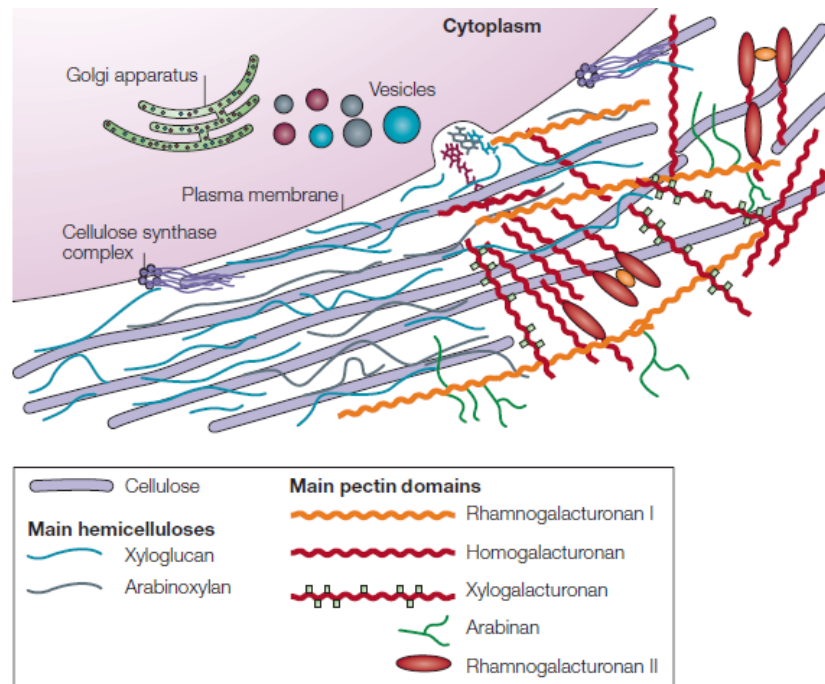


Figure 3 : Schéma de la composition polysaccharidique de la paroi végétale

La paroi est localisée autour de la cellule végétale et est composée de cellulose, d'hémicelluloses, de pectines et de glycoprotéines (ces dernières n'apparaissent pas sur le schéma). La synthèse des composants non celluloses est réalisée dans le système endomembranaire *via* notamment l'appareil de Golgi. Ils sont ensuite acheminés jusqu'à la paroi grâce à des vésicules de sécrétion. Seule la cellulose fait exception à ce mode de synthèse car celle-ci est polymérisée directement au niveau de la membrane plasmique grâce aux complexes de cellulose synthase. Schéma issu de l'étude de Cosgrove *et al.*, 2005.

II-2-2-1 Description de la paroi primaire des cellules végétales

La paroi des cellules végétales est constituée de près de 90% de polysaccharides et de 10% de protéines (glycoprotéines et protéoglycannes). A l'exception de la cellulose, les différents composés de la paroi (hémicelluloses, pectines et protéines) sont synthétisés et assemblés dans le système endomembranaire de sécrétion essentiellement au sein de l'appareil de Golgi. Les composés pariétaux sont ensuite acheminés jusqu'à l'extérieur de la cellule par un transport vésiculaire.

La cellulose ($\beta(1,4)$ -D-glucane) est un polymère homogène de cellobiose (constituée de deux glucopyranoses liés par une liaison osidique $\beta(1\rightarrow4)$) organisé en fibres (Figure 4) (Cosgrove, 2005 ; Lerouxel *et al.*, 2006). Elle est synthétisée au niveau de la membrane de la cellule et libérée directement dans la paroi, grâce à des enzymes nommées CESA (*Cellulose Synthases*). Les complexes enzymatiques CESA forment des rosettes au niveau de la membrane plasmique (Figure 4). La cellulose peut être retrouvée sous deux formes : amorphe et cristalline. La cellulose cristalline est beaucoup plus souple que la forme amorphe. Un relâchement de la cellulose est possible grâce à des enzymes de remodelage de la paroi. De plus, des protéines non enzymatiques de la paroi, les expansines, faciliteraient la coupure des liaisons non covalentes entre les microfibrilles de cellulose et les composés hémicellulosiques (Cosgrove, 2000).

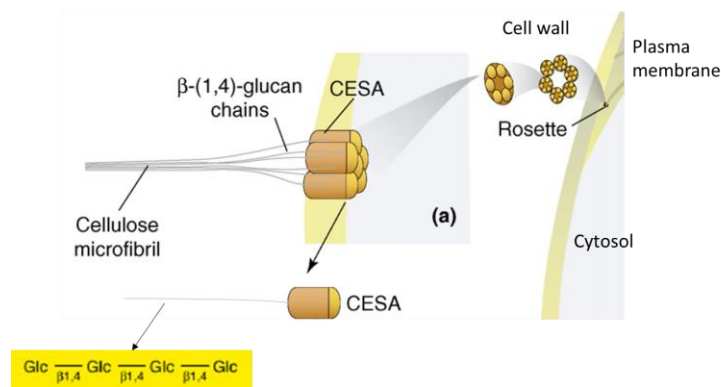


Figure 4 : Synthèse et structure schématique de la cellulose

La cellulose est synthétisée par un complexe protéique d'enzymes nommé CESA (*Cellulose Synthases*) organisées en rosettes dans la membrane plasmique. La fibre de cellulose est constituée de six microfibrilles, chacune constituée de six chaînes de glucanes ($\beta(1,4)$ -D-glucanes). Figure adaptée de l'étude de Lerouxel *et al.*, 2006

Les hémicelluloses sont des polymères ramifiés, avec une chaîne principale composée de glucose, xylose ou mannose liés en $\beta(1-4)$. Les hémicelluloses sont classées en plusieurs catégories en fonction de leur chaîne principale. On parle alors de xyloglucanes, mannanes et xylanes (Figure 5). De plus, cette composition peut varier en fonction de l'espèce végétale ainsi que des tissus (Pauly *et al.*, 2013). Les xyloglucanes sont composés d'une chaîne principale de $\beta(1-4)$ glucose substituée de chaînes latérales de xyloses, galactoses et fucoses (Scheller et Ulskov, 2010). Les mannanes sont composées d'une chaîne principale de mannoses liés en $\beta(1-4)$ ou de glucoses et mannoses liés en $\beta(1-4)$, on parle alors de glucomannanes. Cette chaîne principale peut être ramifiée par des galactoses branchés en $\alpha(1-6)$. Les xylanes sont constitués d'une chaîne principale de $\beta(1-4)$ xylanes, pouvant être substituée d'arabinoses ou d'acides glucuroniques. On les appelle alors des arabinoxylanes ou des glucuronoarabinoxylanes (Scheller et Ulvskov, 2010).

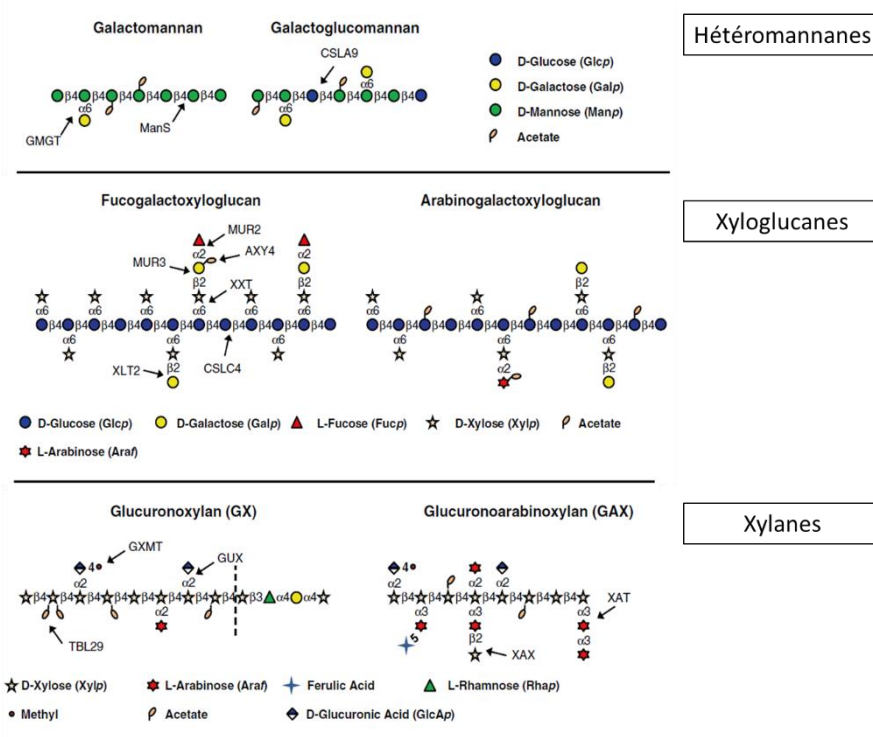


Figure 5 : Structure schématique des différentes hémicelluloses et présentation des enzymes impliquées dans leur synthèse

La structure des hémicelluloses varie en fonction de leur composition (nature de la chaîne principale et des chaînes latérales), de l'espèce végétale et du tissu concerné. Figure adaptée de l'étude de Pauly *et al.*, 2013.

Les pectines sont des polymères chargés riches en acide galacturonique (GalA). La majorité des composés pectiques comportent une chaîne principale de GalA liés en $\alpha(1-4)$

qui peut être méthylestérifiée ou O-acétylée. Lorsqu'il n'y a pas de ramification, il s'agit alors d'homogalacturonanes (Figure 6). Cette chaîne principale peut être substituée de xylooses : on parle alors de xylogalacturonanes. On trouve ensuite les rhamnogalacturonanes de type 1 (RG-I) dont la chaîne principale est constituée d'acides galacturoniques et de rhamnoses (Figure 6). Le RG-I peut posséder des ramifications composées d'arabinanes, galactanes ou d'arabinogalactanes. Le rhamnogalacturonane de type 2 (RG-II) est le polysaccharide le plus complexe des pectines. En effet, sa chaîne principale d'acide galacturonique peut être substituée par quatre chaînes latérales différentes contenant un grand nombre de monosaccharides pouvant être liés entre eux par plusieurs types de liaisons (Figure 6). On retrouve des résidus osidiques rares tels que l'apiose ou le KDO (acide 2-keto-3deoxy-D-manno-2octulosonique). Le RG-II peut se dimériser grâce à des liaisons diester de borate au niveau des résidus apiose, ce qui permet de renforcer la paroi (Dumont *et al.*, 2014).

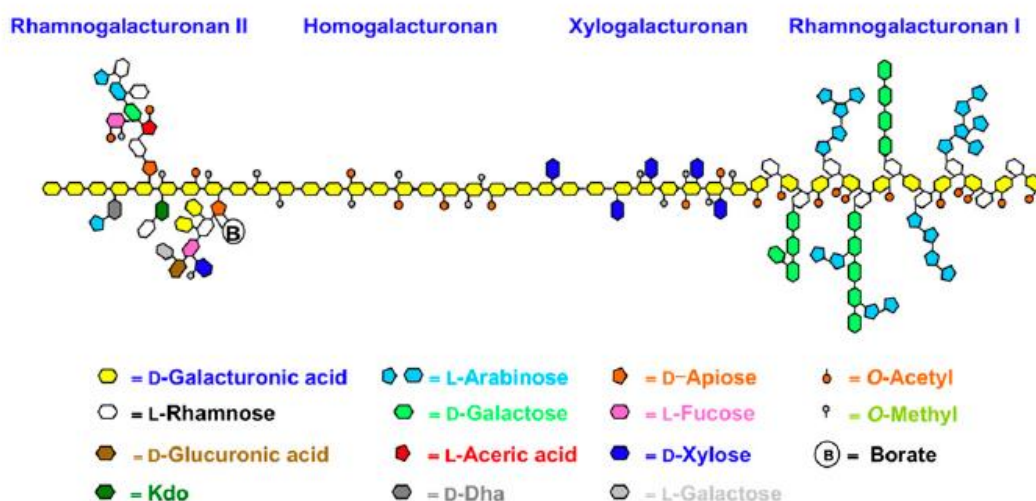


Figure 6 : Structure schématique des différents types de pectines

Il existe quatre types de pectines : les homogalacturonanes, xylogalacturonanes, rhamnogalacturonanes de type 1 (RG-I) et rhamnogalacturonanes de type 2 (RG-II). Kdo = 3-Deoxy-D-manno-2-octulosonic acid; DHA = 3-deoxy-D-lyxo-2-heptulosaric acid. Les homogalacturonanes et les RG-I sont les pectines les plus abondantes de la paroi. Figure issue de l'étude de Harholt *et al.*, 2010.

Les protéines pariétales comprennent les protéines actives de la paroi telles que les expansines et les enzymes de remodelage ainsi que les protéines de structure. Les *Hydroxyproline Rich Glycoproteins* (HRGP) sont une famille de protéines majeure dans la structure pariétale. Ces protéines sont connues pour être impliquées dans la défense de la plante. Les HRGPs comprennent notamment les AGP (*ArabinoGalactan Proteins*) et les extensines. Les AGP sont des protéoglycannes composées d'un corps protéique riche en

hydroxyproline permettant son O-glycosylation, et d'un corps glycanique représentant la majeure partie de cette molécule, et constitué principalement de galactose et d'arabinose. On retrouve aussi dans cette structure de l'acide glucuronique, du fucose et du rhamnose (Nguema-Ona *et al.*, 2013 ; 2014). Les AGP possèdent souvent une ancre GPI (*GlycosylPhosphatidyl Inositol*) qui permet leur fixation à la membrane plasmique (Figure 7). La structure des AGP varie selon les espèces, mais celles-ci sont retrouvées en grand nombre dans tous les organes des plantes étudiées ainsi que dans les exsudats racinaires. Ces protéines jouent un rôle notamment dans l'embryogénèse, l'orientation de la croissance du tube pollinique, la croissance, la mort cellulaire programmée, la communication avec les microorganismes, et dans la protection de la plante contre les stress biotiques (Nguema-Ona *et al.*, 2013 ; 2014).

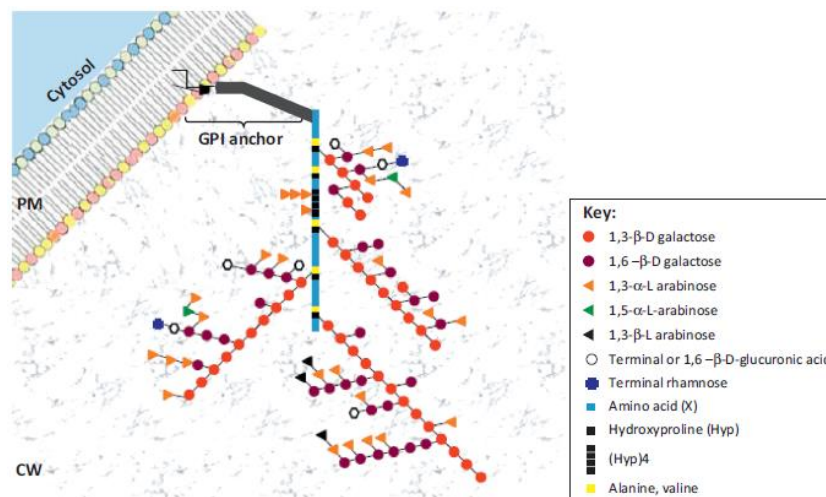


Figure 7: Structure schématique des *ArabinoGalactan Proteins* (AGP)

Les AGP sont des glycoprotéines de la famille des *hydroxyproline rich glycoprotein* (HRGP), possédant un corps protéique riche en hydroxyprolines et un corps glycanique riche en galactose et arabinose. Ces glycoprotéines sont ancrées à la paroi grâce à leur ancre GPI (*GlycosylPhosphatidyl Inositol*). PM = *plasma membrane* ; CW = *cell wall*. Figure issue de l'étude de Nguema-Ona *et al.*, 2013.

Parmi les HRGP, on retrouve aussi les extensines, qui sont des glycoprotéines constituées d'une partie hydrophile qui va être O-glycosylée, et d'une partie hydrophobe qui va permettre un *cross-linking* des extensines entre elles (Figure 8). Les extensines sont impliquées dans le développement embryonnaire, l'élongation cellulaire, la croissance des poils absorbants, la morphologie des cellules bordantes, la structuration de la paroi ainsi que

la défense de la racine contre les bioagresseurs (Nguema-Ona *et al.*, 2014 ; Castilleux *et al.*, 2018).

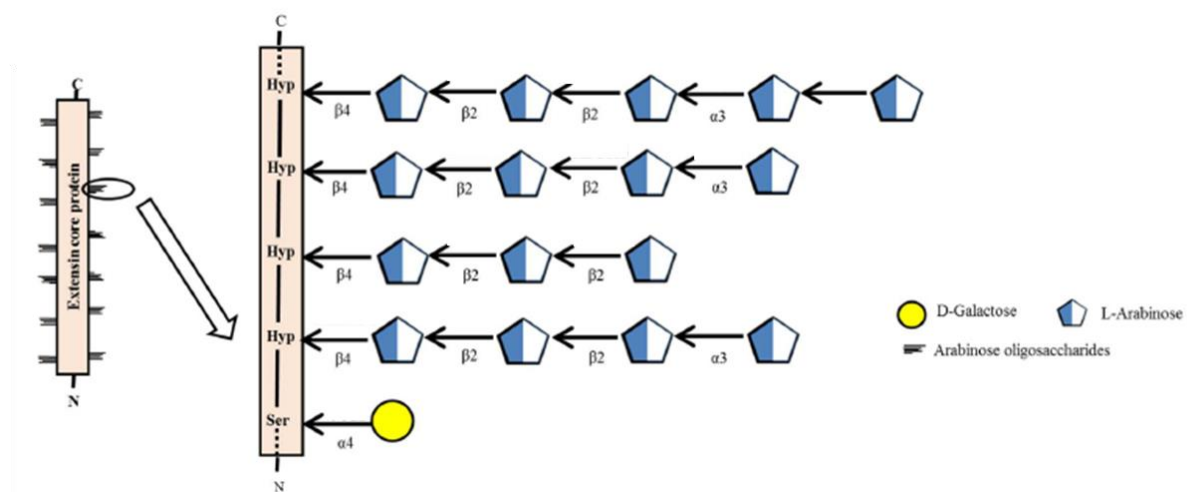


Figure 8 : Structure schématique des extensines

Les extensines sont composées d'un corps protéique sur lequel sont liés des résidus L-arabinose et D-galactose. Figure issue de l'étude de Showalter et Basu, 2016.

II-2-2-2 Les DAMPs

Ces différents composés de la paroi peuvent être modifiés par l'action d'enzymes, permettant l'assouplissement de la paroi ou sa rigidification. Cela permet notamment de mieux résister à l'attaque des pathogènes, rendant plus difficile la dégradation de la paroi par ceux-ci. La paroi végétale constitue donc un premier rempart de défense contre l'attaque des phytopathogènes. De plus, lors de l'attaque d'un bioagresseur, la paroi végétale est dégradée et libère des composants pariétaux (ou des fragments de ceux-ci) reconnus par la plante comme des DAMPs.

Il existe donc différents DAMPs issus de la paroi végétale. Les oligogalacturonides (OG), issus des pectines de type homogalacturonanes (Ferrari *et al.*, 2013 ; Yu *et al.*, 2017), sont les éliciteurs endogènes les mieux caractérisés et les premiers identifiés en 1981 par Hahn *et al.*, chez le soja. Ils sont reconnus par le récepteur WAK1 (*Wall-Associated Kinase 1*) (Azevedo Souza *et al.*, 2017 ; Johnson *et al.*, 2018 ; Hou *et al.*, 2019). Parmi les DAMPs issus de la paroi végétale, on retrouve aussi les oligomères du xyloglucane (Claverie *et al.*, 2018 ; Hou *et al.*, 2019), ou encore la cellobiose issue de la dégradation de la cellulose (Azevedo Souza *et al.*, 2017).

La paroi n'est cependant pas la seule source de DAMPs chez la plante. On retrouve aussi des protéines ou peptides pouvant jouer ce rôle, comme par exemple :

- les systemines, qui sont stockées dans la vacuole et libérées en cas de blessure (Abdul Malik *et al.*, 2020),
- le facteur RALF (*Rapid Alcalinization Factor*) qui est reconnu par le récepteur FER (FERONIA) (Stegmann *et al.*, 2017 ; Hou *et al.*, 2018),
- les protéines de défense PR1 (*Pathogenesis Related 1*) dont la dégradation permet la libération de CAPE1 (*Cationic Antimicrobial Peptide 1*), un peptide de 11 acides aminés également considéré comme un DAMP (Hou *et al.*, 2014 ; Hou *et al.*, 2018),
- GmPep914/890, qui est un peptide libéré chez le soja en cas de blessure (Yamaguchi *et al.*, 2011 ; Hou *et al.*, 2019).

Les DAMPs en particulier et les éliciteurs en général peuvent être détectés par les cellules des tissus aériens, ou par les cellules du tissu racinaire en cas d'attaque par un pathogène du sol (comme par exemple les oomycètes du genre *Phytophthora*). La suite de ce chapitre sera consacrée au compartiment racinaire et à ses mécanismes de défense.

II-3 Les mécanismes de défense racinaire

II-3-1 Organisation d'une racine en croissance chez les plantes angiospermes eudicotylédones

II-3-1-1 Les différents territoires cellulaires

Le système racinaire permet l'ancrage au sol et l'absorption de l'eau et des nutriments nécessaires au bon développement de la plante. Sur une racine en croissance, on distingue trois territoires qui se chevauchent :

- la zone méristématique apicale (Figure 9, A), siège des divisions cellulaires,
- la zone d'élongation (Figure 9, B), responsable en grande partie de la croissance de la racine,
- la zone de différenciation (Figure 9, C) sur laquelle on retrouve des poils absorbants et le développement des racines secondaires.

A ces trois territoires s'ajoute la coiffe racinaire en position apicale qui est une structure protectrice du méristème sub-apical (Petricka *et al.*, 2012).

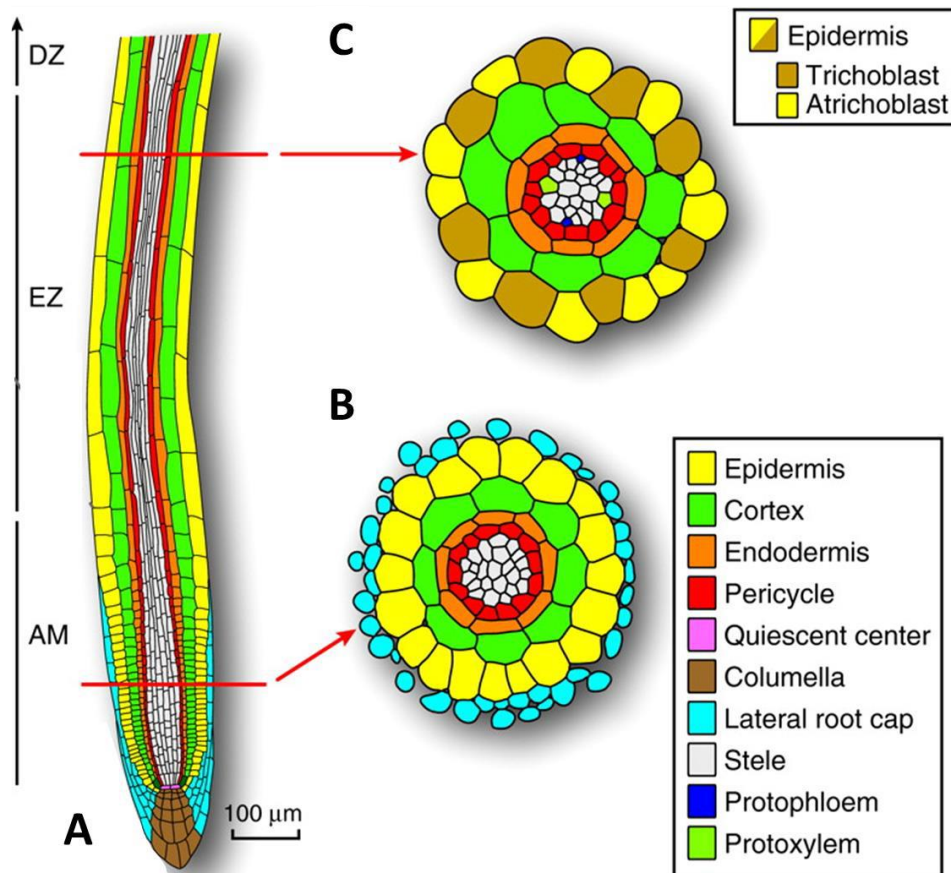


Figure 9: Les différents territoires cellulaires d'une racine en croissance d'*Arabidopsis thaliana*

La racine en croissance est divisée en trois zones : (A) la zone méristématique (AM = *Apical Meristem*) comprenant les initiales méristématiques et leurs cellules dérivées ; (B) la zone d'élongation (EZ = *Elongation Zone*) ; et (C) la zone de différenciation (DZ = *Differentiation Zone*). Figure adaptée de l'étude de Overvoorde *et al.*, 2010.

Chacune de ces zones est composée de différents types cellulaires (Figure 9, Overvoorde *et al.*, 2010). Dans la zone de différenciation, partie la plus mature de la racine, on retrouve des cellules de la stèle qui sont les plus centrales, entourées par les cellules du péricycle puis de l'endoderme et du cortex. Les cellules les plus extérieures sont les cellules du rhizoderme (Overvoorde *et al.*, 2010) où des poils absorbants peuvent se développer. La zone d'élongation, également constituée des cellules de la stèle, du péricycle, de l'endoderme et du cortex, est dédiée à la maturation des cellules racinaires et notamment à l'élongation de la racine. Enfin, le RAM (*Root Apical Meristem*) ou méristème subapical de la racine est le

siège des divisions cellulaires, à l'origine des tissus précédemment décrits et de la coiffe racinaire située à l'extrémité de la racine.

II-3-1-2 Le méristème subapical de racine ou *Root Apical Meristem (RAM)*

Le méristème sub-apical de racine ou *Root Apical Meristem (RAM)* est à l'origine des cellules qui composent tous les tissus racinaires y compris ceux de la coiffe. L'organisation cellulaire du RAM fermé d'*A. thaliana* (Figure 10) est largement décrite dans la littérature (Dolan *et al.*, 1993 ; Van Den Berg *et al.*, 1997 ; Benfey and Scheres, 2000 ; Scheres *et al.*, 2002 ; Rahni *et al.*, 2016).

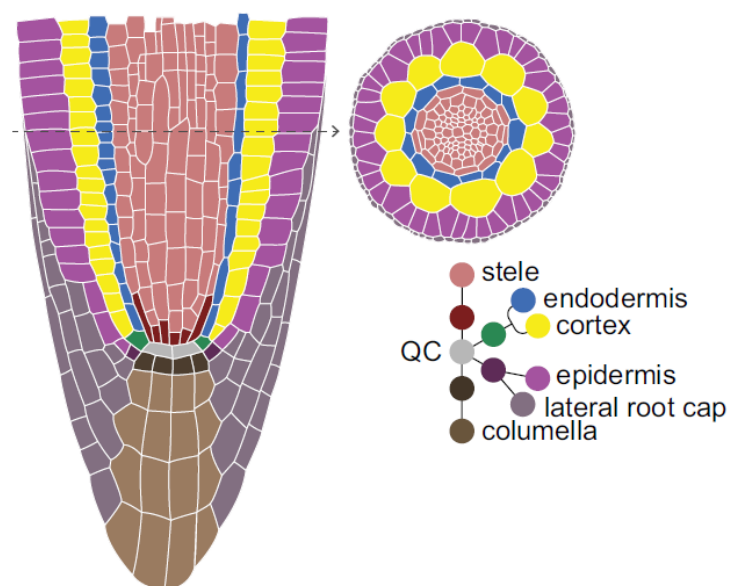


Figure 10 : Schéma montrant l'organisation du RAM fermé chez *Arabidopsis thaliana*

On retrouve quatre types de cellules initiales qui sont à l'origine des différents tissus de la racine : les initiales de la stèle (rouge foncé), les initiales de la coiffe latérale et du rhizoderme (violet), les initiales centrales de la columelle (marron foncé) et les initiales du cortex et de l'endoderme (vert). Figure issue de l'étude de Rahni et Birnbaum, 2019.

Le RAM est composé en son centre de cellules qui forment le centre quiescent (CQ), qui est lui-même entouré de cellules non différenciées nommées cellules initiales (Rahni et Birnbaum, 2019). Chez *A. thaliana*, il y a 4 types de cellules initiales : les initiales de la stèle formant le cylindre vasculaire, les initiales formant l'endoderme et le cortex, les initiales centrales formant la columelle et les initiales latérales formant les cellules latérales de la coiffe et le rhizoderme (Figure 10). Chaque type d'initiale va subir des divisions et générer sa descendance cellulaire (cellules dérivées) (Rahni et Birnbaum, 2019), formant ainsi des colonnes bien distinctes et organisées dans un RAM fermé (Hamamoto *et al.*, 2006).

Lorsque les initiales et leur descendance ne sont pas organisées, il s'agit alors d'un RAM de type basique-ouvert que l'on retrouve par exemple chez le soja (Hamamoto *et al.*, 2006). Une organisation intermédiaire : le RAM intermédiaire-ouvert, montre un certain niveau d'organisation pour les lignées du rhizoderme et de la coiffe latérale, mais l'identité des initiales n'est pas claire et spécifique. Ce type d'organisation est retrouvé par exemple chez le coton (*Gossypium hirsutum* L.) (Hamamoto *et al.*, 2006).

II-3-2 Des initiales de la coiffe aux cellules frontières

II-3-2-1 La coiffe racinaire

La coiffe racinaire est un ensemble de cellules régulièrement renouvelé. Elle est composée de 2 types cellulaires : les cellules latérales de coiffe et les cellules de la columelle. Comme décrit précédemment, ces 2 types cellulaires sont issus respectivement des cellules initiales latérales et des initiales centrales formant le méristème de coiffe (Arnaud *et al.*, 2010 ; Rahni *et al.*, 2016). La coiffe racinaire a pour fonction première de protéger le RAM lors de sa croissance. Contrairement au reste de la racine, la taille de la coiffe racinaire est peu modifiée par les processus de prolifération et d'élongation cellulaires qui ont lieu dans le RAM et ces cellules dérivées. La coiffe libère dans la rhizosphère ses cellules les plus périphériques appelées cellules frontières ou AC-DC (*root-Associated Cap-Derived Cells* ; Driouich *et al.*, 2013 ; 2019 ; Figure 11).

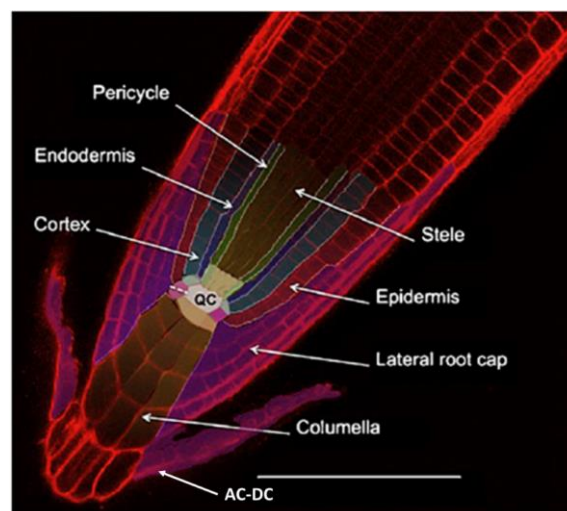


Figure 11 : Structure de l'apex racinaire d'*Arabidopsis thaliana*

Observation réalisée au microscope confocal avec une coloration au iodure de propidium. On observe ici la structure de l'apex avec ses différents types cellulaires mis en évidence (stèle, péricycle, endoderme, épiderme, cortex, centre quiescent (QC), cellules latérales de la coiffe, columelle et AC-DC de racine (*root-Associated, cap-derived cells*)). Barre d'échelle = 100 μ m. Figure adaptée de l'étude de Arnaud *et al.*, 2010.

Les cellules de la columelle ont la particularité d'accumuler des grains d'amidon (statolithes) dans leur cytoplasme (Arnaud *et al.*, 2010). Ces cellules sont appelées statocytes, et sont capables, grâce à ces grains d'amidon, de percevoir la gravité et d'orienter la croissance racinaire. Ces cellules ainsi que les cellules latérales de la coiffe, vont progressivement mûrir vers les extrémités de la coiffe, avant de s'en décrocher grâce à l'action d'enzymes modifiant et dégradant certains polysaccharides de la paroi. Ainsi les pectines de type homogalacturonanes sont dé-estérifiées par des pectines méthylestérases et dégradées par des polygalacturonases (Wen *et al.*, 1999 ; Mravec *et al.*, 2014 ; 2017), ce qui contribue à la libération des AC-DC (Figure 11).

Deux types d'AC-DC ont été décrits dans la littérature : les AC-DC de type cellules bordantes (BC) (Hawes *et al.*, 1998) qui se détachent individuellement de la coiffe racinaire, et les AC-DC de type apparenté aux cellules bordantes (BLC) (Vicré *et al.*, 2005) qui se détachent de la coiffe tout en restant accrochées entre elles.

II-3-2-2 Les cellules frontières ou root-Associated, Cap-Derived Cells (AC-DCs)

A l'origine, les cellules frontières nommées *sloughed peripheral root cap cells* étaient considérées comme des cellules mortes détachées de l'apex racinaire. Il fallut attendre l'étude de Hawes et Pueppke en 1986 pour montrer qu'elles étaient métaboliquement actives. En 1998, Hawes *et al.*, les ont nommés cellules bordantes (BC = border cells) et ont observé une accumulation de ces cellules en réponse à une infection du coton (*Gossypium hirsutum* (L.)) par les zoospores d'un oomycète pathogène (*Pythium dissotocum* Drechsler) ou à une modification de la teneur en CO₂ de leur milieu de culture. Cette étude a ainsi montré que ces cellules sont vivantes, capables de réagir en fonction de leur environnement et semblent jouer un rôle dans la rhizosphère. Jusqu'en 2005, les différentes études réalisées sur l'apex d'*A. thaliana* (Brigham *et al.*, 1995 ; Hawes *et al.*, 2002) n'avaient pas permis l'observation de cellules bordantes sur cette plante. Vicré *et al.*, en 2005 ont mis en évidence des cellules semblables aux BC, mais qui restaient associées à la racine en formant des filaments de cellules attachées entre elles. Celles-ci furent nommées « cellules apparentées aux cellules bordantes » (BLC = *border-like cells*). Jusqu'en 2011, il était admis que les plantes possédaient soit des BC, soit des BLC selon l'espèce. En 2011, une étude de Endo *et al.*, a montré que la plante *Acacia mangium* Willd., possédait des BC mais aussi des BLC, indiquant qu'il pouvait y avoir plusieurs modes de détachement de ces cellules pour

une même plante. Cela a de nouveau été observé chez *Glycine max* (L.) Merr., par Ropitiaux en 2018 et par Carreras *et al.*, en 2020 chez *Acacia tortilis* subs. *raddiana* Savi et *Tamarindus indica* (L.). L'existence de différents morphotypes au sein de ces AC-DC a également été décrit. On distingue ainsi des AC-DC sphériques, intermédiaires ou allongées (Cannesan *et al.*, 2011 ; Carreras *et al.*, 2020 ; Ropitiaux 2018 ; Figure 12). De plus, au sein d'un morphotype on peut retrouver quelques différences entre les formes de cellules : les cellules sphériques pouvant par exemple être soit rondes (Figure 12, b), soit légèrement incurvées (Figure 12, c), les intermédiaires pouvant être droites ou incurvées (Figure 12, d et e) et les cellules allongées pouvant être isolées ou organisées en feuillets (Figure 12, f et g). Enfin, ces morphotypes semblent avoir une localisation particulière au niveau des zones de la racine (Figure 12, h).

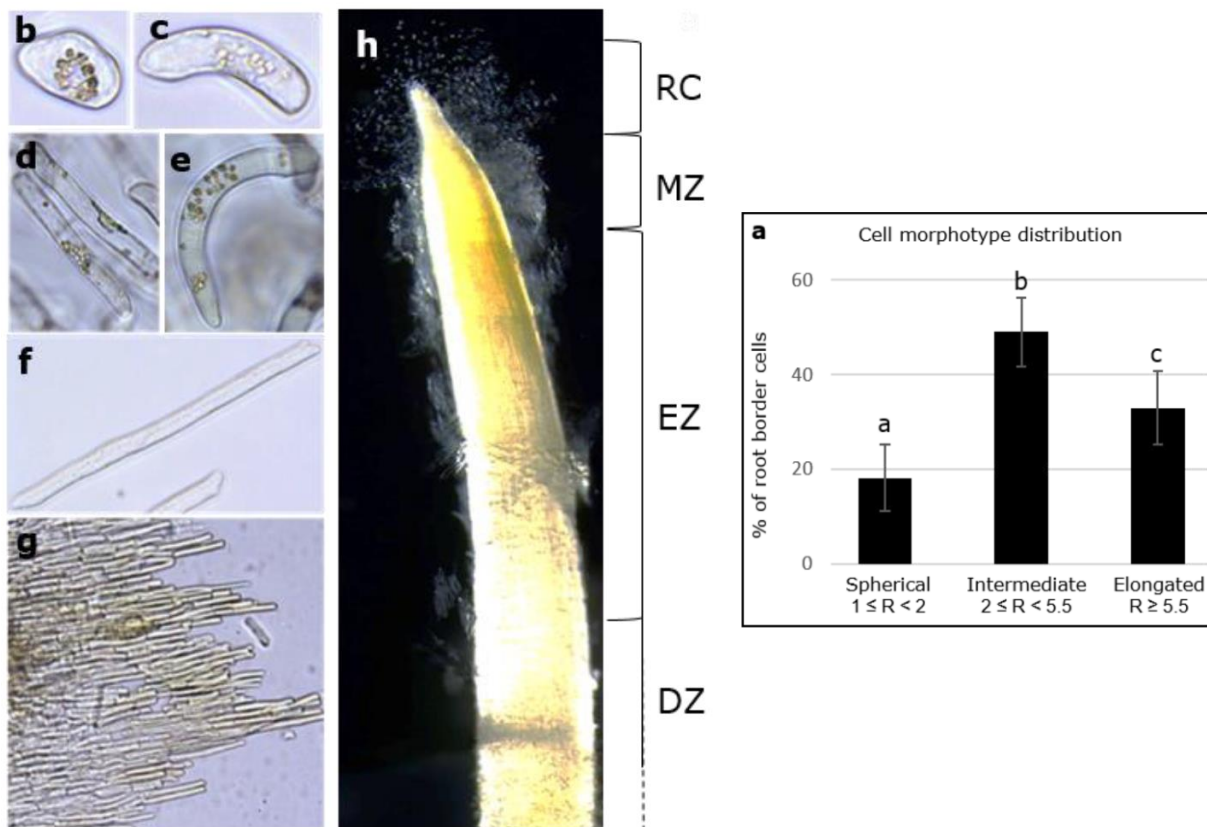


Figure 12 : Les différents morphotypes des cellules frontières, leur localisation et leur abondance chez le soja (*Glycine max* (L.) Merr.)

Histogramme représentant l'abondance de chaque morphotype en pourcentage (a) : on observe que les intermédiaires (d, e) semblent plus abondantes que les allongées (f, g) et sphériques (b, c). Afin de déterminer les morphotypes, un ratio longueur/largeur (R) a été calculé puis attribué à chaque morphotype (sphériques : $1 \leq R < 2$; intermédiaires : $2 \leq R < 5,5$; allongées : $R \geq 5,5$). Pour chaque morphotype, on retrouve des nuances, les sphériques pouvant être rondes (b) ou légèrement incurvées (c) ; les intermédiaires peuvent être droites (d) ou incurvées (e) ; les allongées peuvent être isolées (f) ou en feuillets (g). Chacun de ces morphotypes semble localisé spécifiquement au niveau de certaines zones de la racine (h), les sphériques au niveau de la coiffe racinaire (RC), les intermédiaires au niveau de la zone méristématique (MZ), et les allongées au niveau de la zone d'élongation (EZ) et la zone de différenciation (DZ). Figure issue de l'étude de Ropitiaux, 2018.

Cette hétérogénéité au sein des cellules frontières de la coiffe racinaire a conduit Driouich *et al.*, en 2019, à donner un nom générique à cet ensemble de cellules : les AC-DC de racine (*root-Associated, Cap-Derived Cells*).

Ces cellules, métaboliquement actives et dont la quantité augmente en réponse à une infection (Hawes *et al.*, 1998), vont jouer un rôle dans l'immunité racinaire, en formant une zone bien particulière et spécifique à la racine : le RET (*Root Extracellular Trap*).

II-3-3 Le RET et l'immunité racinaire

II-3-3-1 Spécificité de l'immunité racinaire

La racine, de par sa localisation souterraine, est soumise à des stress biotiques ou abiotiques différents de la partie aérienne de la plante. Ces différences peuvent donc être à l'origine de spécificités dans la réponse de la racine notamment aux stress biotiques. Ces spécificités ont été discutées par Chuberre *et al.*, en 2018. En effet, les mécanismes de défense mis en place au niveau des racines se distinguent de ceux décrits au niveau du système aérien et varient en fonction des tissus racinaires considérés. Les auteurs indiquent que les zones de différenciation, d'élongation et méristématiques ne semblent pas répondre de la même façon à une élicitation (Figure 13).

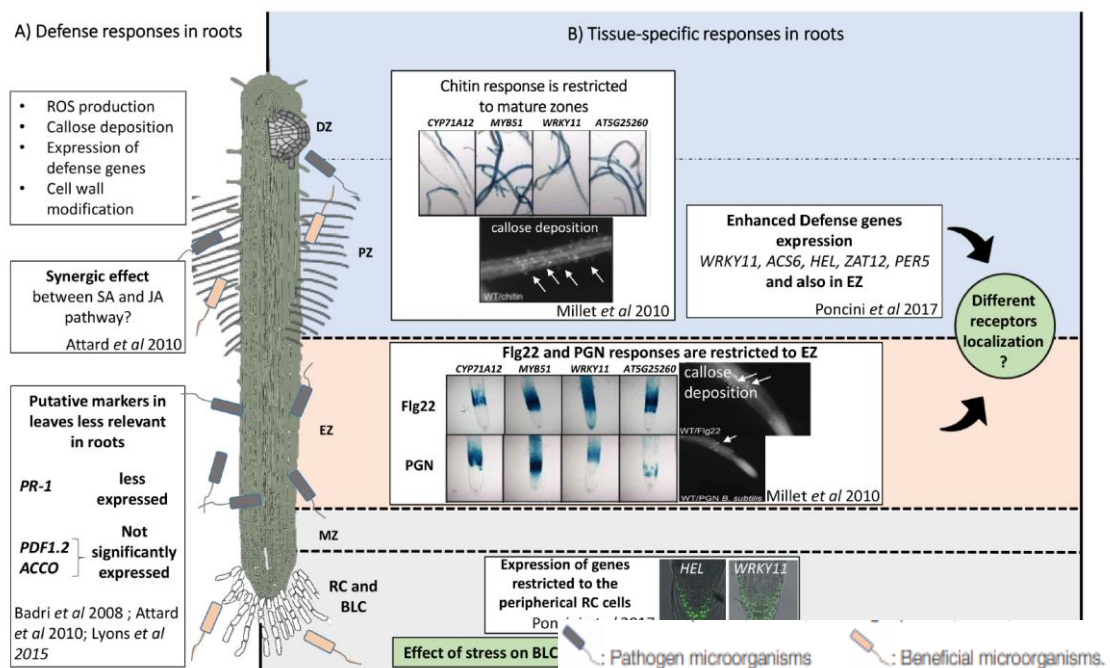


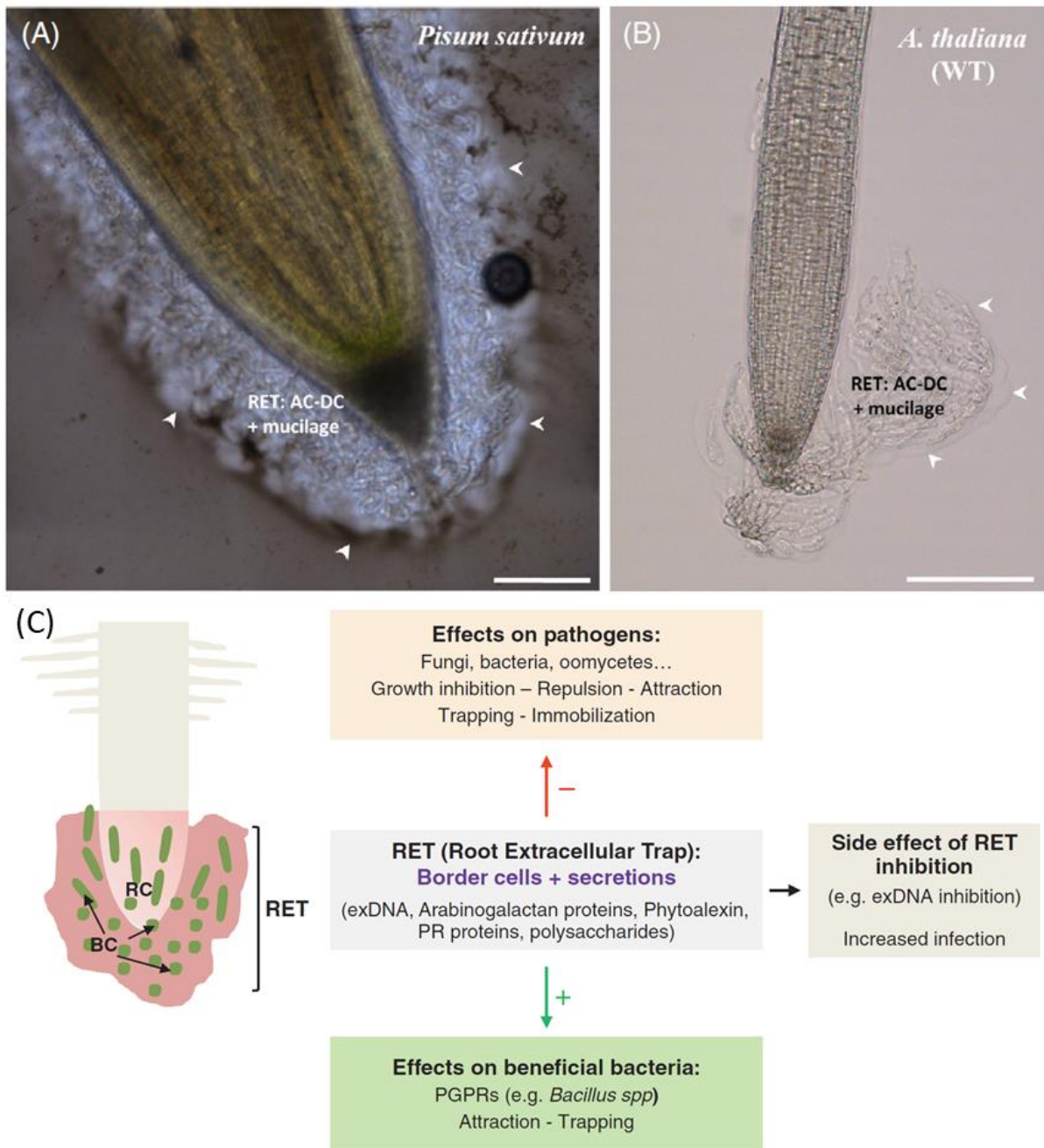
Figure 13 : Schéma de la réponse de défense de la racine d'*Arabidopsis thaliana* et de ses variations en fonction des zones racinaires considérées

La réponse globale de la racine (A) passe par la production de ROS (*Reactive Oxygen Species*), dépôts de callose, l'expression de gènes de défense ainsi que des modifications de la paroi. Contrairement aux parties aériennes, l'acide jasmonique et l'acide salicylique ne semblent pas avoir d'effet antagoniste dans la racine, de plus les marqueurs de défense habituellement utilisés pour la partie aérienne ne semblent pas adaptés à la racine (PR-1 pour l'acide salicylique ; PDF1.2 pour l'acide jasmonique ; ACCO pour l'éthylène). Il existe des réponses spécifiques aux différentes zones de la racine (B) en fonction de l'éliciteur utilisé, indiquant une éventuelle différence de localisation des récepteurs capables de détecter ces éliciteurs. Figure issue de l'étude de Chuberre *et al.*, 2018.

Millet *et al.*, en 2010 ont par exemple observé que la chitine induisait la formation de dépôts de callose et l'expression de certains gènes de défense uniquement au niveau de la zone de différenciation (Figure 13, B). Une réponse similaire est observée au niveau de la zone d'élongation lorsque la racine est élicitée avec la flg22 ou avec des peptidoglycanes (Millet *et al.*, 2010). En 2017, Poncini *et al.*, ont montré que l'expression des gènes du facteur de transcription WRKY11 (nommés ainsi en raison de leurs peptides N-terminaux WRKYGQK) et de HEL (*Hevein-like protein*) semble se restreindre aux cellules périphériques de l'apex en réponse à une élicitation avec la flg22. On observe donc des réponses de défense différentes d'une zone racinaire à une autre. La réponse de défense d'une autre zone spécifique à la racine reste donc à discuter : celle du RET.

II-3-3-2 Le piège extracellulaire de racine ou Root Extracellular Trap (RET)

Les AC-DC, étroitement liées à un mucilage, forment le RET (*Root Extracellular Trap*) ou piège extracellulaire de racine. Ce modèle a été décrit pour la première fois en 2013 (Figure 14) et discuté à nouveau en 2019 par Driouich *et al* (Figure 15), par analogie avec le NET (*Neutrophil Extracellular Trap*) retrouvé chez les mammifères. En effet, ces deux structures réagissent à la présence d'un pathogène, de PAMPs (partie II-2-1) ou de DAMPs (partie II-2-2) (Brinkmann et Zychlinsky, 2012 ; Zawrotniak et Rapala-Kozik, 2013 ; Schauer *et al.*, 2014 ; Boeltz *et al.*, 2019 ; Driouich *et al.*, 2019). Elles sont constituées de molécules similaires : protéines antimicrobiennes, peroxydases, ADN extracellulaire et histones (Figure 15, b ; Urban *et al.*, 2009 ; Brinkmann et Zychlinsky, 2012 ; Wen *et al.*, 2017 ; Driouich *et al.*, 2019). Cependant, les spécificités du RET sont la présence de polysaccharides, de ROS et NO dans les sécrétions (Knee *et al.*, 2001 ; Driouich *et al.*, 2016 ; Galloway *et al.*, 2017 ; Mravec *et al.*, 2017 ; Wang *et al.*, 2017) ; ce type de molécules n'ayant pas été détectées dans le NET (Brinkmann *et al.*, 2004 ; Halverson *et al.*, 2015 ; Selders *et al.*, 2017) (Figure 15, b).



Current Opinion in Plant Biology

Figure 14 : Visualisation microscopique et représentation schématique du RET (Root Extracellular Trap)

Le RET est situé à l'apex racinaire, composé des AC-DC et de leurs sécrétions mucilagineuses. (A) RET du pois en contre-coloration à l'encre de chine en microscopie photonique. (B) RET d'*Arabidopsis thaliana* en microscopie photonique. (A, B) Le mucilage est délimité par les pointes de flèches blanches. (C) Modèle schématique du RET. Celui-ci peut avoir un effet négatif sur les microorganismes pathogènes (immobilisation, inhibition de la croissance, etc.) ou positif sur les microorganismes bénéfiques (attraction, piège). Barres d'échelle (A) et (B) : 50 μ m. Figure adaptée des études de Driouch *et al.*, 2013 et 2019.

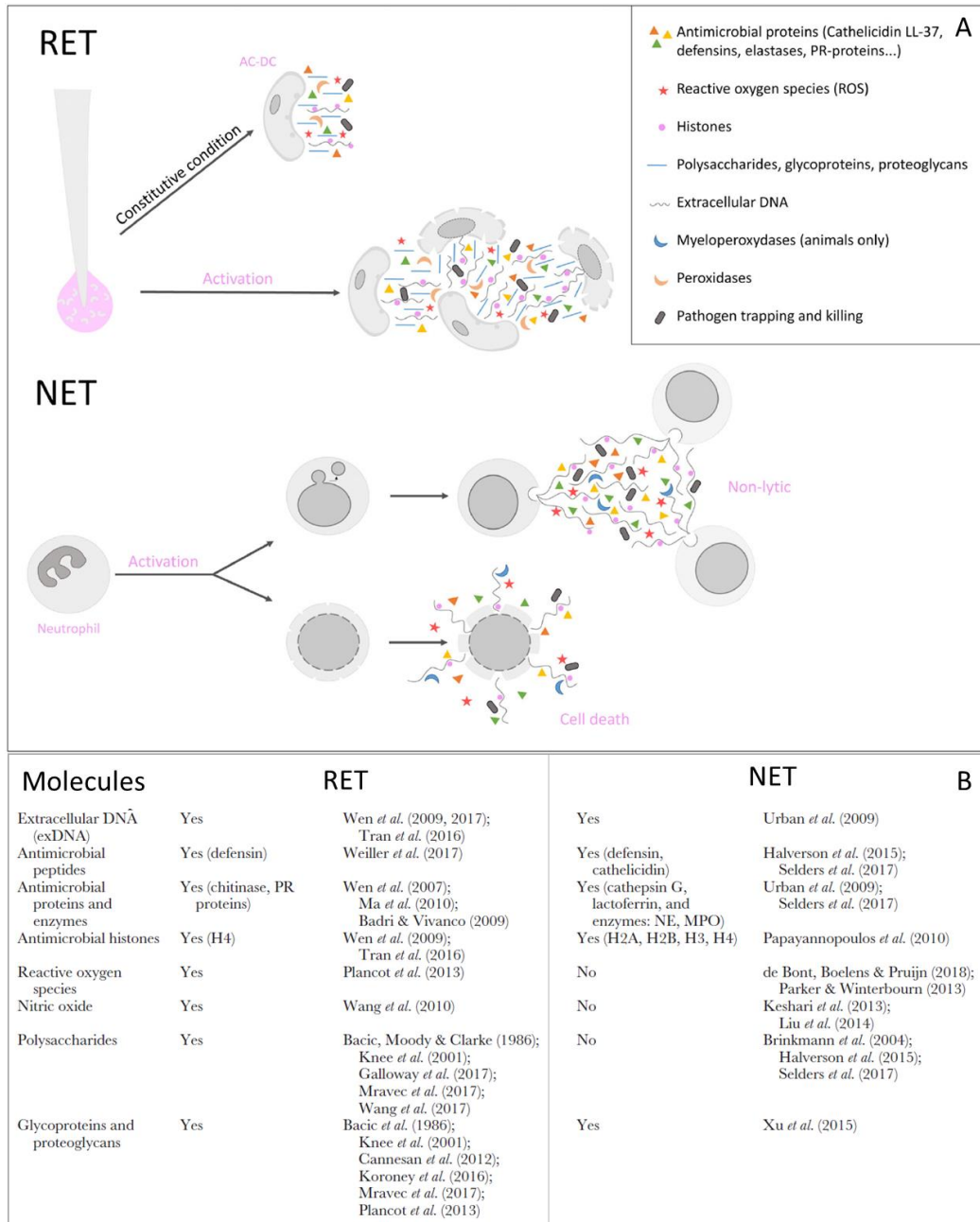


Figure 15 : Comparaison du RET (Root Extracellular Trap) et du NET (Neutrophil Extracellular Trap)

Schéma montrant l'activation du RET et du NET (A). PR = Pathogenesis Related. La formation du RET et du NET peut être stimulée par la reconnaissance d'un PAMP (Pathogen Associated Molecular Pattern) ou DAMP (Damage Associated Molecular Pattern). Le RET est aussi présent de façon constitutive (hors réponse immunitaire), mais en quantité moindre. Deux modes de formation du NET sont possibles : la formation non-lytique (n'entraînant pas la mort cellulaire) et la formation lytique par mort cellulaire. En (B), un tableau montrant certaines différences ou similarités dans la composition du RET et du NET. Figure adaptée de l'étude de Driouich *et al.*, 2019.

Du fait de sa localisation dans le sol, la racine est en permanence confrontée à un grand nombre de microorganismes (Mendes *et al.*, 2013) pouvant avoir une interaction avec celle-ci, de nature bénéfique (exemple des PGPR : *Plant Growth Promoting Rhizobacterias*) ou pathogène pour la plante. Le RET joue un rôle important dans la réponse de la racine à ces interactions (Driouich *et al.*, 2013 ; 2019), et donc dans la défense de la racine.

II-3-3-3 L'implication du RET dans l'immunité racinaire

La position à l'interface sol/apex racinaire du RET lui confère une importance majeure dans l'immunité racinaire. Cependant les fonctions exactes du RET dans la mise en place de mécanismes de défense commencent seulement à être explorées. En effet, les études de Plancot *et al.*, en 2013 et Castilleux *et al.*, en 2018 ont montré une réponse des AC-DC du RET à l'élicitation des racines d'*Arabidopsis thaliana* (L.) Heynh, et de lin (*Linum Usitatissimum* (L.)) liée à la réorganisation des épitopes des extensines pariétales. Dans le cas de ces deux plantes, les extensines semblent donc jouer un rôle dans les mécanismes de défense du RET.

Les AC-DC de racines sont généralement associées à un mucilage composé de polysaccharides, de métabolites secondaires, de protéines antimicrobiennes et d'ADN extracellulaire (Wen *et al.*, 2007 ; Driouich *et al.*, 2013 ; Driouich *et al.*, 2019). Cet ensemble est produit par la racine et par les AC-DC. Les AC-DC seraient cependant les actrices majoritaires de la formation de ce mucilage (Wen *et al.*, 2007). Le mucilage, constitué de composés de haut et de bas poids moléculaires, contient également des molécules exsudées par les racines et qui peuvent y être piégées.

Les composés de bas poids moléculaires sont d'une taille inférieure à 1,5 kDa (Dixon, 2001 ; Marouf et Reynaud, 2007). Il s'agit notamment de monomères ou de molécules faiblement polymérisées, comme par exemple des mono- ou oligosaccharides, des acides aminés ou oligopeptides, des fragments d'acides nucléiques, des acides organiques ou des métabolites spécialisés (Timotiwu et Sakurai, 2002 ; Bais *et al.*, 2006 ; Ashiuchi *et al.*, 2015 ; Barilli *et al.*, 2015). Certains de ces composés sont connus pour avoir un rôle dans la défense. Parmi ceux-ci, on peut citer l'exemple de la pisatine qui est une phytoalexine produite par le pois en réponse à une inoculation par un oomycète *Aphanomyces euteiches* Drechsler (Cannesan *et al.*, 2011). De plus, des composés tels que les acides aminés et

monosaccharides peuvent attirer les microorganismes par chimiotactisme (Bais *et al.*, 2006), permettant d'attirer par exemple des bactéries bénéfiques PGPR, mais aussi les pathogènes (Cannesan *et al.*, 2012).

Les composés de hauts poids moléculaires sont les plus représentés au sein du RET. On y retrouve principalement des polysaccharides, des protéines pouvant être glycosylées et de l'ADN extracellulaire (Bais *et al.*, 2006 ; Wen *et al.*, 2009 ; Driouich *et al.*, 2013).

Les polysaccharides les plus retrouvés dans le mucilage sont notamment des pectines comme les homogalacturonanes ou les xylogalacturonanes chez *Arabidopsis thaliana* (L.) Heynh (Vicré *et al.*, 2005 ; Driouich *et al.*, 2013). En 2019, Ropitiaux *et al.*, ont montré la présence de xyloglucane et de cellulose formant un maillage qui relie les AC-DC de pois (*Pisum sativum* (L.)) entre elles. La présence d'hétéromananes, de xyloglucanes, de cellulose et de xylogalacturonanes a aussi été observée dans le mucilage du RET du Soja (*Glycine max* (L.) Merr.) (Ropitiaux 2018).

En 2007, Wen *et al.*, ont montré qu'il y avait une sécrétion d'environ 1,25 µg de protéines solubles dans l'eau par racine de pois, répartis en plus de 100 protéines différentes. Parmi ces protéines on retrouve notamment des extensines, des endo-xyloglucane transférases, des invertases pariétales, des β-galactosidases, des cellulases, des α-mannosidases, des oligosaccharyl transferases et des histones H4 (Wen *et al.*, 2007). Ces composés sont des acteurs de la synthèse et du métabolisme de la paroi, mais peuvent aussi jouer un rôle dans la défense de la plante. En effet, les histones H4 peuvent modifier les charges à la surface de la paroi des bactéries ; ce qui déstabilise les membranes et provoque leur lyse cellulaire (Brinkmann et Zychlinsky, 2012, Tran *et al.*, 2016). En 2010, Ma *et al.*, ont identifié 2848 protéines dans le RET du maïs. Parmi ces protéines, 24,6% d'entre elles sont reconnues comme étant impliquées dans le métabolisme, 6,7% ont une fonction de liaison, et 4,8% sont impliquées dans la réponse aux maladies ou la défense de la plante (Ma *et al.*, 2010). Il est à noter que 41,2% des protéines de ce protéome de RET de maïs n'étaient pas annotées. Des chitinases, impliquées dans la réponse de défense face aux champignons phytopathogènes ont aussi été retrouvées dans le mucilage du pois, d'*Arabidopsis* et du colza (Ma *et al.*, 2010). En 2016, Weiller *et al.*, ont montré la présence de défensines dans les sécrétions d'*Heliophila coronopifolia* (Burman), des protéines impliquées elles aussi dans la défense. Des AGP (partie II-2-2-1) ont aussi été retrouvées dans les exsudats racinaires

d'*Arabidopsis thaliana* (L.) Heynh, de pois (*Pisum sativum* (L.)), de pomme de terre (*Solanum tuberosum* (L.)) (Cannesan *et al.*, 2012 ; Driouich *et al.*, 2013 ; Nguema-Ona *et al.*, 2013 ; Koroney *et al.*, 2016) et de plantes ligneuses sahéliennes (Careras *et al.* 2020). Un effet chimio-attractant des AGP du pois et du colza sur les zoospores d'*Aphanomyces euteiches* a été démontré par Cannesan *et al.*, en 2012.

Parmi ces composés de haut poids moléculaire, on retrouve aussi de l'ADN extracellulaire (ADNex) (Wen *et al.*, 2009).

II-3-3-4 L'ADN extracellulaire

II-3-3-4-1 Chez les animaux et microorganismes

Chez les mammifères, le rôle des neutrophiles dans le système immunitaire est bien connu. En effet ceux-ci sont capables de migrer du sang vers les tissus infectés et vont pouvoir détecter, puis phagocyter et finalement éliminer les pathogènes (Brinkmann *et al.*, 2004 ; Papayanopoulos, 2018). Afin de détruire ces pathogènes, les neutrophiles vont produire un piège capable de les retenir : le NET (*Neutrophil Extracellular Trap*). En 2004, Brinkmann *et al.*, ont montré la présence d'ADN extracellulaire (ADNex) dans ce NET en grande quantité formant des complexes avec des histones et des élastases. Ils ont mis en évidence que la production du NET n'était pas le résultat d'une simple apoptose, mais que celle-ci pouvait être une étape programmée avant la mort cellulaire des neutrophiles. Suite à cela, il a été montré que la formation du NET pouvait être réalisée de manière létale (Figure 16, A) ou non lytique (Figure 16, B) (Papayannopoulos, 2018 ; Galluzzi *et al.*, 2018). La formation du NET est déclenchée par la détection d'un pathogène, c'est-à-dire l'activation d'un récepteur membranaire qui va reconnaître des PAMPs et déclencher une cascade de mécanismes aboutissant à la formation du NET par NETosis lytique ou non-lytique (Yipp *et al.*, 2012). En 2009, Yousefi *et al.*, ont montré que l'ADNex du NET pouvait être d'origine mitochondrial, ce qui correspondrait également à une formation non-lytique du NET (Yousefi *et al.*, 2009 ; 2019).

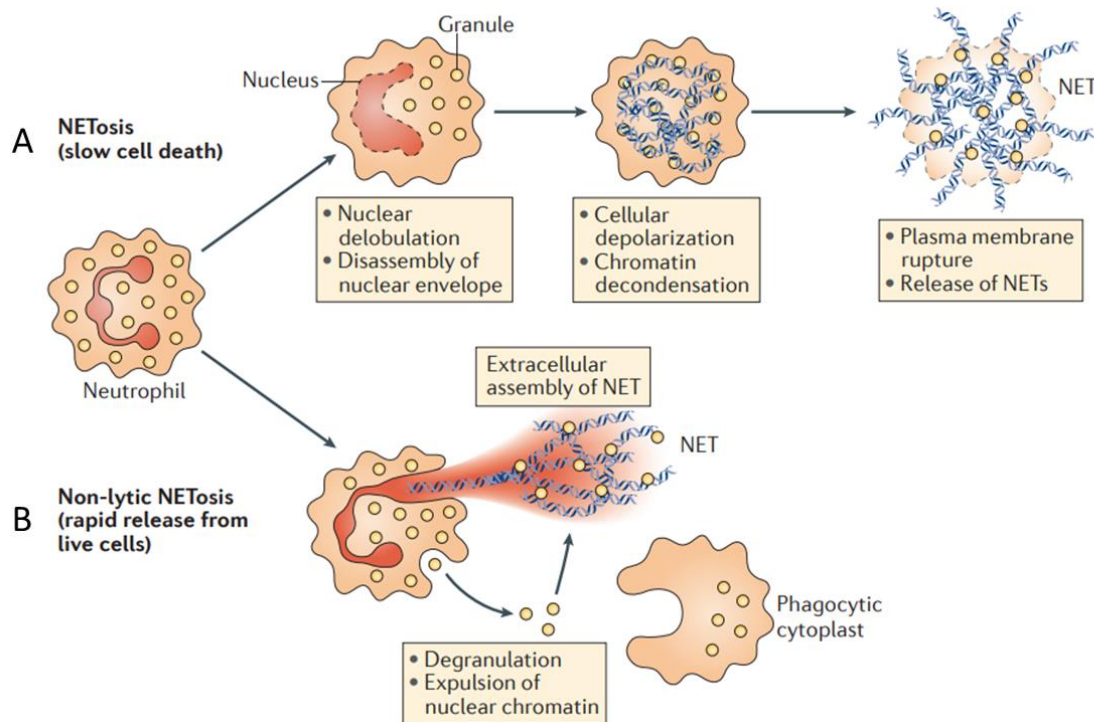


Figure 16 : Représentation schématique de la formation lytique ou non-lytique du NET

Le NET peut être formé de deux manières, la première appelée NETosis (A) commence par la perte de l'intégrité du noyau et de son enveloppe nucléaire, puis se poursuit par la dépolérisation de la cellule et la décondensation de l'ADN et se termine par la rupture de la membrane plasmique permettant de libérer le NET. La seconde voie de formation du NET est la non-lytique NETosis (B) qui consiste en l'expulsion de l'ADN et en une dégranulation où le neutrophile reste vivant et capable de phagocyter les microorganismes pathogènes malgré la perte de son noyau. Figure issue de Papayannopoulos, 2018.

Les neutrophiles ne sont cependant pas les seules cellules du système immunitaire des mammifères à pouvoir former des pièges extracellulaires (ETs : *Extracellular Traps*). Les monocytes (Webster *et al.*, 2010), les mastocytes (von Köckritz-Blickwede *et al.*, 2008), les macrophages (Aulik *et al.*, 2012) et les polynucléaires éosinophiles (Simon *et al.*, 2011 ; Dworski *et al.*, 2011) sont aussi capables de former des ETs. Chez l'Homme, la présence d'ADNex peut résulter d'une inflammation, d'une allergie, d'une infection, de la coagulation ou encore d'une maladie auto-immune (Simon *et al.*, 2013 ; Delgado-Rizo *et al.*, 2017 ; Bont *et al.*, 2018). Cet ADN peut donc être retrouvé dans le sang, la lymphe, la bile, le lait maternel, l'urine, la salive, le liquide amniotique, le liquide spinal, *etc.* (Thierry *et al.*, 2016). La quantification de l'ADNex peut de plus être utilisée dans la détection des cancers. En effet, il a été montré que la quantité d'ADNex augmentait chez les patients atteints d'un cancer (Thierry *et al.*, 2016 ; Elazezy et Jousse, 2018). De plus, l'analyse de séquences cibles de celui-ci peut aussi être utilisée pour le diagnostic de certaines maladies génétiques, comme par exemple le syndrome de Down (Palomaki *et al.*, 2011) ou la trisomie 18 et 13 (Palomaki

et al., 2012) chez le fœtus grâce au plasma maternel. L'ADNex peut aussi être à l'origine de certaines maladies ou complications, par exemple des thromboses, une septicémie ou des métastases (Papayannopoulos, 2018).

Les pièges extracellulaires ne sont pas présents uniquement chez les mammifères (Vargas *et al.*, 2017). Ils sont aussi présents chez d'autres vertébrés tels que le crocodile (Kozłowski *et al.*, 2016), le poisson (Palic *et al.*, 2007), le poulet (Pieper *et al.*, 2017) ou chez des invertébrés comme une amibe (*Dictyostelium discoideum* Rapper) et un ver de terre (*Eisenia andrei* Savigny) (Homa *et al.*, 2015). On les retrouve aussi chez les microorganismes tels que les bactéries et les champignons (Allesen-Holm *et al.*, 2006 ; Baum *et al.*, 2009 ; Martins *et al.*, 2010).

Les bactéries peuvent former un biofilm afin d'adhérer à une surface. Un biofilm est une communauté structurée dans une matrice composée de polysaccharides, de protéines, et d'ADNex (Baum *et al.*, 2009 ; Okshevsky *et al.*, 2015). Cet ADNex semble permettre la structuration du biofilm car un traitement à la DNase conduit à sa déstabilisation (Allesen-Holm *et al.*, 2006). De plus, il permet le transfert horizontal de gènes, la réparation de dommages de l'ADN, la résistance aux pathogènes et aux peptides anti-microbiens, la tolérance aux antibiotiques (Tetz et Tetz 2010) et peut être une source de nutriments pour la communauté bactérienne (Ibáñez de Aldecoa *et al.*, 2017).

L'ADNex est aussi présent chez d'autres microorganismes tels que les archées ou les champignons des genres *Candida* et *Aspergillus* où il semble également impliqué dans l'établissement et la stabilité du biofilm et la tolérance aux antifongiques (Martins *et al.*, 2010 ; Ibáñez de Aldecoa *et al.*, 2017).

L'ADNex est donc retrouvé chez un grand nombre d'organismes vivants, ce qui explique notamment sa présence dans les écosystèmes tels que le sol ou les environnements aquatiques (Ibáñez de Aldecoa *et al.*, 2017 ; Pathan *et al.*, 2020). Dans l'eau, l'ADNex fait partie de la matière organique dissoute (DOM : *Dissolved Organic Matter*) et est utilisé comme source de nutrition par les organismes aquatiques (Lennon, 2007). Dans le sol et les sédiments, cet ADN permet d'apporter 47% des besoins en phosphate, ainsi que 4% du carbone et 7% de l'azote nécessaires à la nutrition des microorganismes (Vuillemin *et al.*,

2017). L'ADNex peut donc provenir de bactéries, d'archées, de virus, de champignons mais aussi d'eucaryotes tels que les plantes.

II-3-3-4-2 Chez les plantes

En 2009, Wen *et al.*, ont mis en évidence la présence d'ADN extracellulaire (ADNex) au sein du mucilage du RET de pois (*Pisum sativum* L.) grâce à des sondes fluorescentes, SYTOX green et DAPI (4',6-diamidino-2-phenylindole). Ils ont ensuite récolté cet ADNex à raison de 10 ng de 5 à 150 kb par racine, et procédé au premier séquençage après amplification de certaines séquences par clonage bactérien. L'identification des séquences obtenues a permis d'observer en majorité un alignement sur le génome du pois (*P. sativum*). De plus, la mise en contact de racines avec un nucléotide contenant du phosphate radioactif (^{32}P) a conduit à la production ADNex marqué, sans augmentation du nombre de cellules mortes. Cette observation a permis à Wen *et al.*, d'émettre l'hypothèse d'une synthèse et excrétion de l'ADNex par des cellules vivantes.

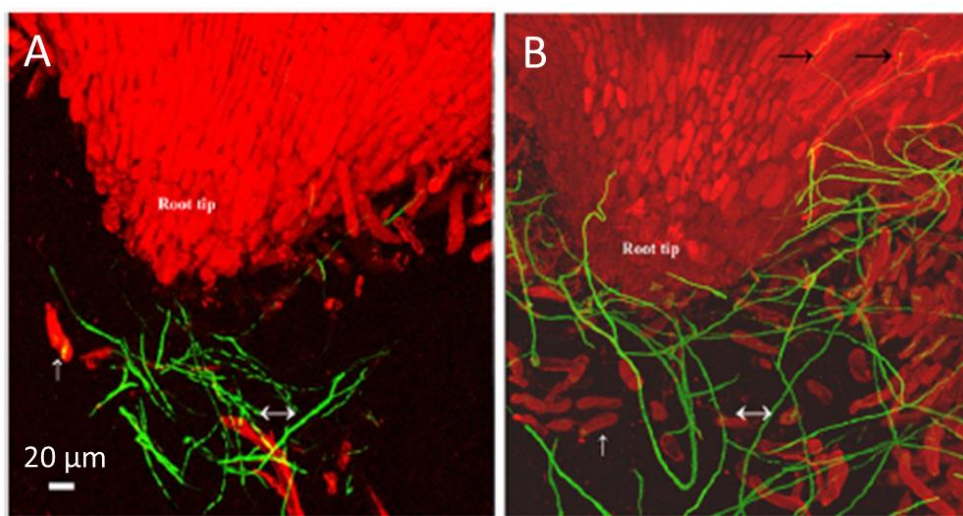


Figure 17 : Tests d'infections de racines de pois par *N. haematococca* en présence ou non de DNase

Images réalisées en microscopie confocale à balayage laser (excitation à 490 et 568 nm). Le champignon *N. haematococca* est marqué à la GFP (green fluorescent protein) et apparaît donc en vert ; la racine apparaît en rouge par autofluorescence. Les observations sont réalisées après 36 heures d'inoculation. (A) infection en condition témoin, le champignon semble retenu à distance de l'apex. (B) Infection en présence d'une unité de DNase I, le champignon atteint l'apex racinaire et peut donc l'infecter plus facilement. Figure issue de Wen *et al.*, 2009.

Le rôle de l'ADNex dans la défense a été mis en évidence grâce à un test d'infection d'une racine de pois par le champignon phytopathogène *Nectria haematococca* Berk. & Broome (Figure 17) en présence ou non d'une endonucléase (DNase I) et d'une exonucléase (BAL31). Ces tests ont abouti à une infection beaucoup plus importante de la racine lors du

traitement avec ces deux nucléases qu'en condition témoin (Figure 17). Ce résultat semble indiquer d'une part la présence d'ADN extracellulaire dans le RET, et d'autre part son implication dans la défense de la racine.

En 2016, Tran *et al.*, ont aussi pu observer l'ADNex du pois lors d'une infection par *Ralstonia solanacearum* (Figure 18, A). Ils ont mis en évidence la production d'ADNex par les AC-DC en réponse à cette infection, ainsi que la production de DNases par le pathogène (Figure 18, B). Ce résultat souligne donc l'importance de l'ADNex dans la défense de la racine et le fait que la production de DNases par les phytopathogènes pourrait être un élément important de leur virulence (Tran *et al.*, 2016 ; Park *et al.*, 2019).

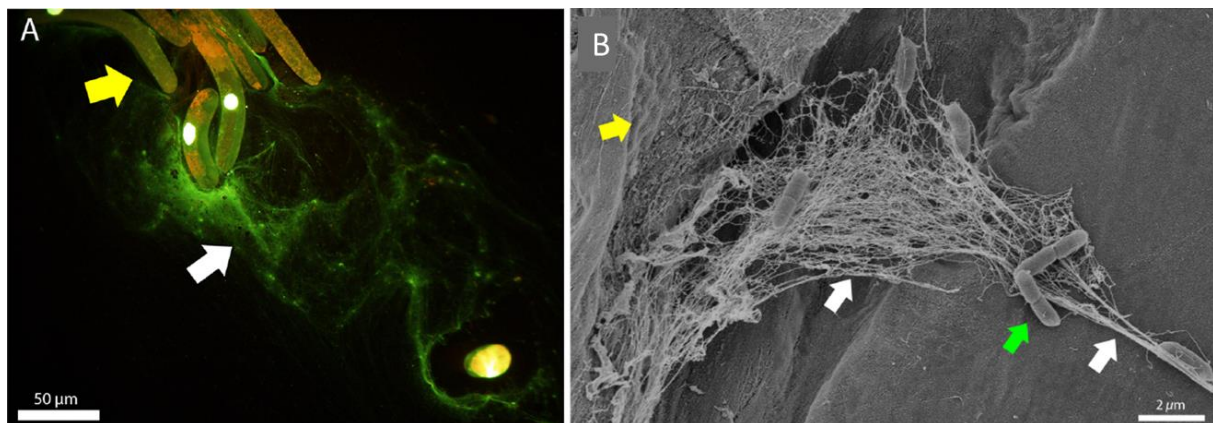


Figure 18 : Observations microscopiques de l'ADNex du pois lors d'une infection par *R. solanacearum*

(A) Image de microscopie à fluorescence d'AC-DC de pois (flèche jaune) et d'ADNex (flèche blanche, coloré en vert au SYTOX green). (B) Image de microscopie électronique à balayage montrant des filaments d'ADNex (flèches blanches) piégeant des bactéries (flèche verte). L'ADNex semble ici provenir d'une AC-DC en cours de lyse cellulaire (flèche jaune). Figure issue de Tran *et al.*, 2016.

Tran *et al.* (2016) ont observé que des bactéries piégées dans le RET étaient davantage lysées que celles placées dans l'eau. De plus, l'action d'une DNase ou d'anticorps anti histones H4 dans le RET permettait de rétablir le taux de survie de ces bactéries. L'ADNex et les histones H4 semblent donc jouer un rôle important dans la défense de la racine du pois contre *R. solanacearum*.

En 2017, Duran-Flores et Heil ont émis l'hypothèse que tout comme celui du NET, l'ADNex d'origine végétale pourrait être reconnu comme un DAMP par les cellules végétales et ainsi activer les défenses de la plante. Afin de vérifier cela, ils ont mis en contact des plantes de haricot (*Phaseolus vulgaris* (L.)) avec des fragments d'ADN génomique de cette même plante et de deux autres Fabacées (*Phaseolus lunatus* (L.) et *Acacia farnesiana* (L.)).

Seuls les fragments d'ADN de haricot ont permis d'activer efficacement i) la production de ROS dans les feuilles, ii) la voie des MAPK dans les cellules en suspension et les feuilles et iii) d'inhiber la croissance des racines. Ces résultats indiquent donc que la plante est capable de reconnaître son propre ADN fragmenté et que celui-ci est bien capable d'activer les défenses immunitaires. Au contraire, l'ADN génomique d'une autre espèce du genre *Phaseolus*, *P. lunatus* (L.), ou celui d'*Acacia farnesiana* (L.) n'ont pas ou très faiblement, induit l'activation de ces mécanismes de défense. La capacité des cellules végétales à reconnaître l'ADN correspondant à leur propre génome impliquerait donc la présence d'un récepteur membranaire capable de détecter la présence de celui-ci. L'activation de ce récepteur (cGAS ou TLR9 chez les mammifères, toujours inconnu chez les plantes) induirait des variations de concentration en Ca^{2+} cytosolique et la production de ROS pouvant activer les défenses immunitaires de la plante (Figure 19, Gallucci et Maffei, 2017).

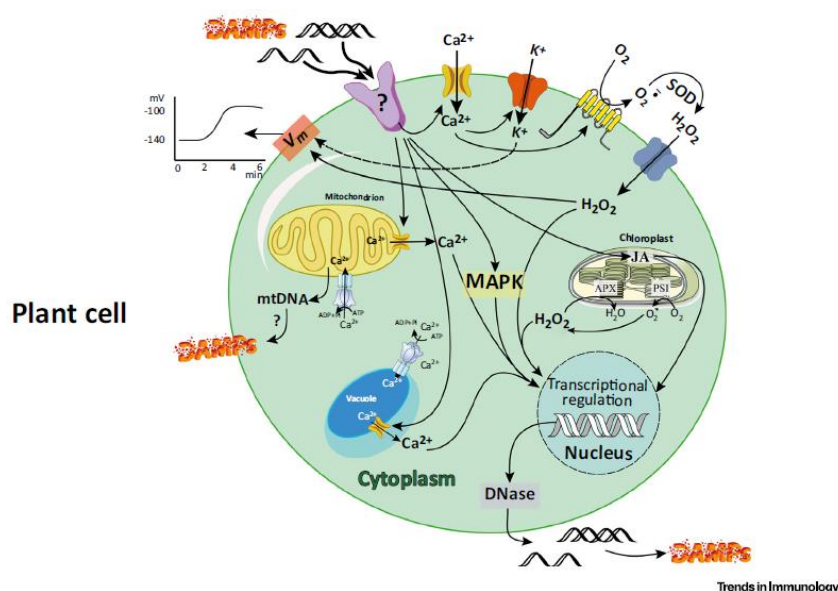


Figure 19 : Mécanismes pouvant être impliqués dans la reconnaissance de l'ADN extracellulaire comme DAMP chez les cellules végétales

La perception de l'ADN_{ex} par un récepteur induit des variations de Ca^{2+} cytosolique conduisant à la production de ROS (anions superoxydes et peroxyde d'hydrogène) et provoquent l'ouverture de canaux K^+ . Cette entrée d'ions K^+ provoque la dépolarisation de la membrane. Ces phénomènes pourraient moduler l'expression de gènes de défense de la plante. Des DNases pourraient être à l'origine de la fragmentation d'ADN qui serait reconnu comme DAMP. L'ADN mitochondrial pourrait lui aussi être impliqué dans ces mécanismes et constituer également l'une des sources de ce DAMP. Figure issue de Gallucci et Maffei, 2017.

Afin de vérifier si l'ADN_{ex} est bien produit par des cellules vivantes du RET et non par la lyse cellulaire, Wen *et al.*, (2017) ont observé des cellules frontières de pois ayant subi un traitement de congélation/décongélation ou de forte centrifugation. Aucune structure semblable à de l'ADN_{ex} n'a pu être détectée suite à ce traitement. Une seconde expérience

a montré un RET nouvellement formé, 30 minutes après retrait du RET par agitation dans l'eau, qui contenait des structures d'ADNex. Ces observations suggèrent donc que l'ADNex pourrait être libéré dans le RET par un mécanisme n'incluant pas uniquement la mort cellulaire.

L'ADNex du RET semble donc nécessaire à la défense de la plante et notamment de sa racine (Wen *et al.*, 2009 ; Tran *et al.*, 2016), tout comme celui du NET chez les mammifères (Brinkmann et Zychlinsky, 2012 ; Halverson *et al.*, 2015). De plus, le RET semble pouvoir libérer de l'ADNex de manière non létale (Wen *et al.*, 2009 ; 2017), tout comme le NET, dont la formation peut être létale ou non (Galluzzi *et al.*, 2018). Si l'origine de l'ADNex du NET est maintenant bien décrite (Yousefi *et al.*, 2009 ; Papayanopoulos, 2017), celle du RET ne l'est pas encore.

Cependant, des études génomiques et microscopiques de l'ADNex ont été réalisées chez deux fabacées : le pois (Wen *et al.*, 2009 ; Tran *et al.*, 2016) et le soja (Ropitiaux, 2018). Ces travaux ont permis de détecter la présence d'ADNex et d'indiquer que cet élément du RET pourrait être un acteur de la défense racinaire chez les Fabacées en général et chez le soja en particulier.

III. Le soja (*Glycine max* (L.) Merr.)

Le soja cultivé (*Glycine max* (L.) Merr. 1917) est une plante de la famille des Fabacées (Figure 20). Elle établit préférentiellement une symbiose avec la bactérie *Bradyrhizobium diazoefficiens* (Delamuta *et al.*, 2013) qui lui permet de fixer l'azote de l'air. Originnaire d'Asie, elle est aujourd'hui cultivée dans le monde entier pour ses graines riches en protéines, glucides, lipides et vitamines (Schmutz *et al.*, 2010 ; Shen *et al.*, 2018 ; CIQUAL ANSES, 2020).

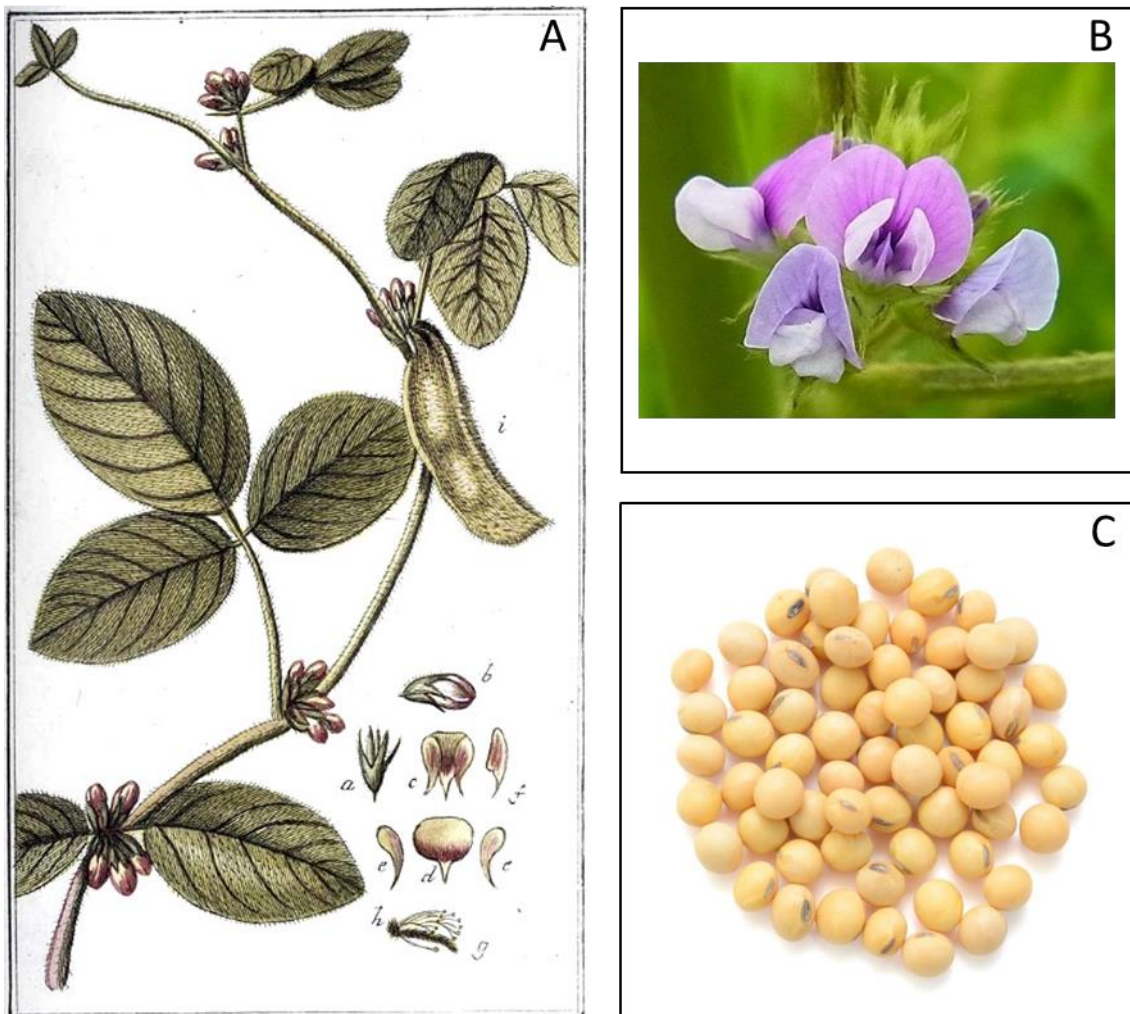
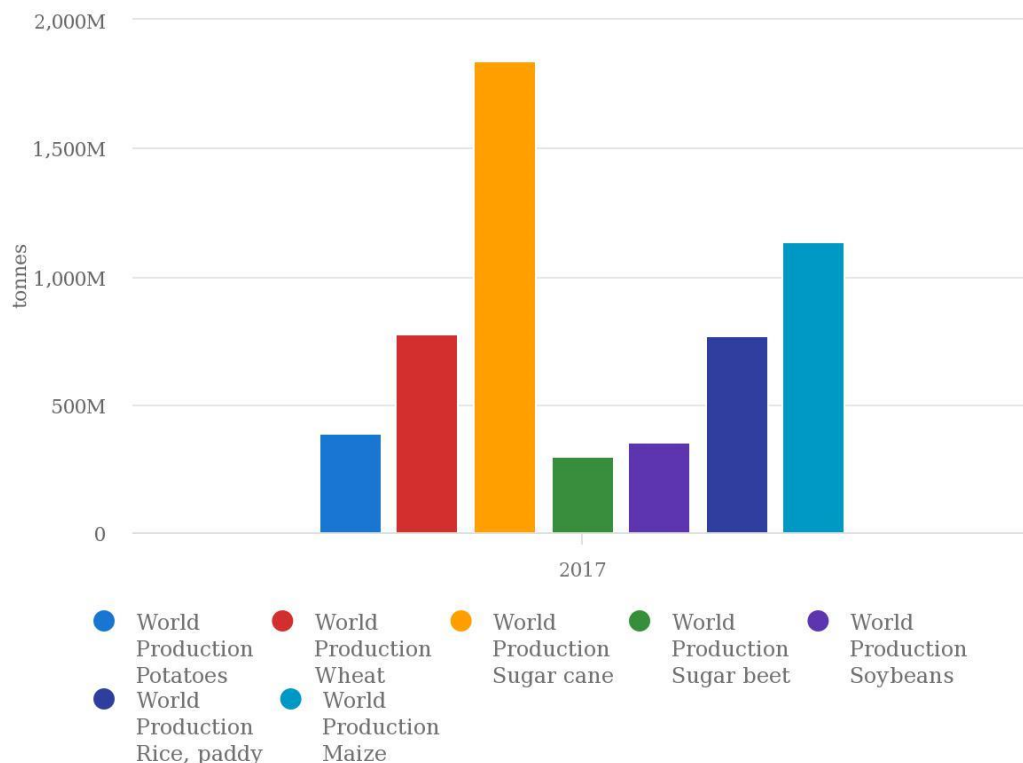


Figure 20 : Dessin botanique, inflorescence et graines de soja (*Glycine max* (L.) Merr.)

Le soja cultivé ou *Glycine max* est une Fabacée dont les parties aériennes (feuilles, tiges et fruits) sont pubescentes. Ses feuilles sont trifoliées (A), ses fleurs sont des racèmes papilionacées de petite taille et localisés à la base des feuilles (B) et son fruit est une gousse pendante contenant des graines globuleuses (C) (Tison et Foucault, 2014). Sources des photos : A : booksofdante.wordpress.com ; B : pinterest.com ; et C : charakusa.com

III-1 La culture du soja

Les principaux pays producteurs de soja sont les Etats-Unis, le Brésil, l'Argentine et la Chine. En 2017, la production mondiale de soja atteignait 336 millions de tonnes, ce qui place le soja à la 6^{ème} place des plantes les plus cultivées dans le monde (FAO STATS, 2020) (Figure 21).



Source: FAOSTAT (Dec 31, 2019)

Figure 21 : Les 7 plantes les plus cultivées dans le monde en 2017 (en tonnes)

La canne à sucre (1,8 milliard de tonnes), le maïs (1,1 milliard de tonnes), le blé (0,77 milliard de tonnes), le riz (0,76 milliard de tonnes), la pomme de terre (0,38 milliard de tonnes), le soja (0,35 milliard de tonnes) et la betterave à sucre (0,30 milliard de tonnes) sont les 7 plantes les plus cultivées dans le monde. Figure issue de l'étude de la FAO STATS, 2019.

Longtemps restée marginale, la culture du soja connaît une forte progression en France avec des surfaces qui ont quadruplé entre 2013 (42 000 ha) et 2018 (154 000 ha) pour une production qui atteint actuellement 400 000 tonnes (Terre-net, 2020). Même si les bassins historiques (Sud-Ouest et Bourgogne) restent prépondérants, l'aire de culture du soja s'étend, grâce à l'inscription récente de variétés précoces et très précoces (groupes de précocité 000, et 0000) qui permettent d'introduire cette culture à haut potentiel dans des régions agricoles du Nord et de l'Ouest de la France. Ainsi en Normandie, quelques essais de culture de soja ont été mis en place au cours des dernières années (Carré 2019, REUSSIR) et

cette culture fait désormais l'objet d'un intérêt régional grandissant *via* le financement d'appels à manifestation d'intérêt "Protéines végétales" par la Région Normandie.

III-2 Les bioagresseurs et éliciteurs du soja

III-2-1 Les bioagresseurs du soja

La culture du soja doit faire face à l'attaque de plus de 200 bioagresseurs dont environ 35 sont connus pour être économiquement importants (Wrather *et al.*, 2010 ; Hartman *et al.*, 2015). Parmi ceux-ci, on retrouve des insectes (par exemple *Spodoptera litura* Fabricius, *Melanagromyza sojae* Zehnter, ou encore *Myloccerus undecimpustulatus* Faust (Gaur et Mogalapu 2018), des bactéries, des virus, des nématodes, des champignons, et des oomycètes (Hartman *et al.*, 2015).

Les maladies bactériennes notables sont dues par exemple à *Pseudomonas syringae* pv. *glycinea* Coerper Young, Dye & Wilkie, responsable de la brûlure bactérienne du soja qui provoque l'apparition de tâches jaunes à brun pâle sur les feuilles (Figure 22, B) (Zou *et al.*, 2005 ; Zabala *et al.*, 2006).

Certains virus peuvent aussi provoquer des pathologies chez le soja, comme par exemple le virus de la mosaïque du soja, qui provoque des tâches vert sombre et vert clair sur les jeunes feuilles, qui aboutissent à terme à une perte des feuilles (Figure 22, E).

Les nématodes sont également de potentiels bioagresseurs de cette plante. On peut citer parmi ceux-ci :

- *Heterodera glycines* Ichinohe, un nématode à kystes, qui provoque la maladie du nanisme jaune (Figure 22, D). Cette maladie est caractérisée par un retard de croissance et un jaunissement de la plante (Alkharouf *et al.*, 2006 ; Kandoth *et al.*, 2011),
- *Meloidogyne incognita* Kofoid & White (Jardim *et al.*, 2018) et *Meloidogyne javanica* Treub (Alekcevetich *et al.*, 2019), sont deux nématodes à galles.

Le soja est aussi sensible aux attaques fongiques. Parmi les champignons phytopathogènes notables, on peut citer :

- *Septoria glycines* Hemmi, qui cause la septoriose du soja et qui se caractérise par la formation de taches brunes sur les feuilles (Batzer *et al.*, 2016),

- *Sclerotinia sclerotiorum* Lib. de Bary, qui provoque la pourriture de la tige du soja (Zhang *et al.*, 2016),
- *Phakopsora pachyrhizi* Syd. & P. Syd., responsable de la rouille du soja qui forme des taches brunes sur les feuilles et provoque leur jaunissement (Figure 22, A) (Panthee *et al.*, 2009),
- *Fusarium oxysporum* f. sp. *Glycines* Armstr. & J.K. Armstr., provoquant la fusariose du soja (Figure 22, A) (Lanubile *et al.*, 2015 ; Chang *et al.*, 2019).

Enfin, les oomycètes sont aussi de potentiels agresseurs du soja. On trouve par exemple :

- *Phytophthora sojae* Kaufm. & Gerd qui est responsable de la pourriture racinaire du soja (Figure 22, C) et provoque à lui seul chaque année des pertes économiques pouvant aller jusqu'à deux milliards de dollars. Les symptômes de cette maladie sont le flétrissement, la verse et les lésions du collet et de la racine (Tyler, 2007),
- Le genre *Pythium spp* qui est associé à la fonte des semis (Navi *et al.*, 2019),
- *Peronospora manshurica* (Naumov) Sydow qui est responsable du mildiou du soja et dont les symptômes sont reconnaissables par l'apparition de petites tâches jaunes sur les feuilles (Figure 22, C) (Dong *et al.*, 2019).

La grande variété des bioagresseurs du soja ainsi que les pertes engendrées par ceux-ci (95,48 milliards de dollars entre 1996 et 2016) (Bandara *et al.*, 2020) soulignent la nécessité de s'intéresser aux interactions du soja avec ceux-ci mais aussi aux mécanismes de défense de cette plante afin de pouvoir développer de nouvelles méthodes de protection.

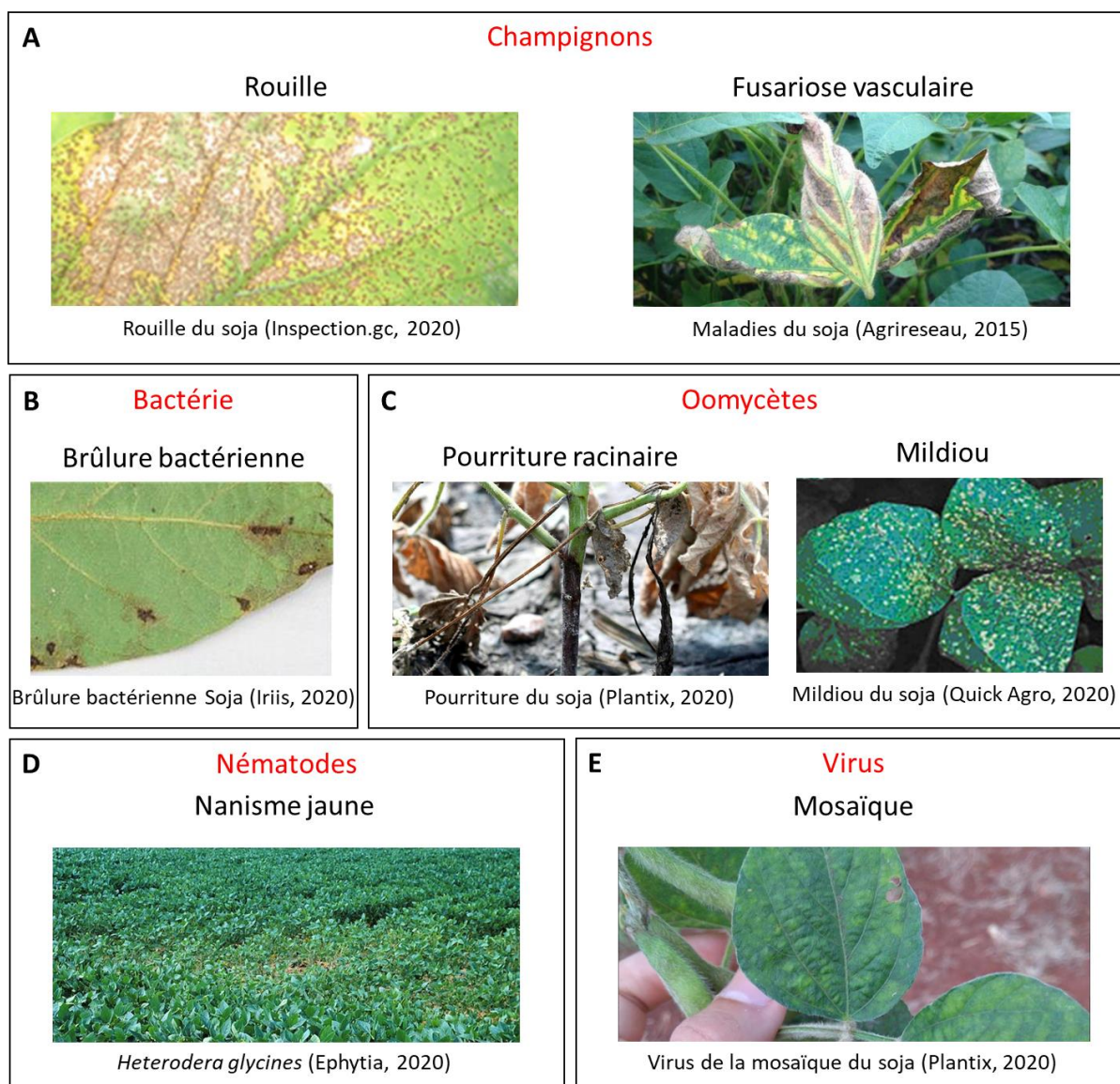


Figure 22 : Exemples de symptômes de différentes maladies pouvant affecter le soja

La rouille et la fusariose (A) sont provoquées par des champignons ; la brûlure bactérienne (B) est provoquée par une bactérie ; la pourriture racinaire et le mildiou (C) sont provoqués par un oomycète ; le nanisme jaune (D) par un nématode ; et la mosaïque (E) est provoquée par un virus. Les maladies sont le plus souvent reconnaissables à leurs symptômes, qui se traduisent par exemple par des taches sur les feuilles (A, B, C : mildiou, E) pouvant être de différentes formes, tailles et couleurs, ou encore par un jaunissement de la plante (C et D), un retard de croissance (D), une pourriture des tissus (A et C), ou la verse des plantes (C).

III-2-2 La réponse de défense du soja

Le soja étant une plante d'intérêt agronomique mondial, de nombreuses études ont été réalisées en présence de pathogènes afin de mieux comprendre les mécanismes de défense de celle-ci (Bonhoff *et al.*, 1987 ; Zou *et al.*, 2005 ; Alkharouf *et al.*, 2006 ; Zabala *et al.*, 2006 ; Klink *et al.*, 2007 ; Panthee *et al.*, 2009 ; Lanubile *et al.*, 2015 ; Dong *et al.*, 2018 ; Bai *et al.*, 2019 ; Ranjan *et al.*, 2019). Ces études consistent à explorer les réponses du soja à l'application de différents stimuli (PAMPs, DAMPs et effecteurs). Ce type d'approche, de par ses réponses multifactorielles, ne permet pas toujours d'identifier précisément les acteurs responsables de l'activation des voies de défense de la plante. Ainsi, en plus de simplifier l'accès aux réponses de la plante et dans une optique de développement de produits de biocontrôle, il est intéressant d'étudier la réponse du soja aux éliciteurs.

Parmi les études ayant porté sur la réponse de cette plante à un éliciteur, on peut citer celle de Parker *et al.*, en 1988. Cette étude a mis en évidence la production de glycéollines en réponse à un extrait glycanique de la paroi de *Phytophthora sojae* Kaufm. & Gerd. La partie protéique de cet extrait n'avait pas induit la production de glycéollines. Après cette étude de 1988, l'effet de cette fraction protéique n'a plus été étudié sur le soja. Cette fraction, correspondant à l'éliciteur PEP-13, a été étudiée sur d'autres plantes telle que le persil, où il induit la production de phytoalexines (Parker *et al.*, 1991).

L'effet des β -glucanes de la paroi de *Phytophthora sojae* a de nouveau été étudié en 2003 par Graham *et al.* Lors de cette étude, l'élicitation a été combinée avec une blessure des cotylédons. Les auteurs ont observé la transcription de gènes de protéines PR en réponse à la blessure et une expression plus importante en cas d'élicitation. Les PR-protéines les plus exprimées dans la totalité des cotylédons lors de cette expérience sont PR-2 (endoglucanase), PR-4 (*WIN-like protein*) et PR-6 (*Kunitz trypsin inhibitors*). En 2007, Daxberger *et al.*, ont de nouveau étudié l'effet des β -glucanes sur le soja, en s'intéressant cette fois à l'activation de la voie des MAP Kinases. L'élicitation permet bien d'activer la voie des MAPK et plus particulièrement des MAPKK-1, MAPK-3 et MAPK-6 du soja.

En 2010, Pearce *et al.*, ont montré que le peptide GmSubPep (issu du soja) permettait l'expression de gènes de défense codants pour des protéines telles que la chitinase, PDR12 (*salicylic acid-inducible ATP-binding cassette transporter*), ou une chalcone synthase. Un

autre peptide issu également du soja, GmPEP914 permet la synthèse de phytoalexines chez cette plante (Yamaguchi *et al.*, en 2011).

En 2016, Pawlowski *et al.*, ont testé l'effet de différents éliciteurs (benzothiadiazole, chitosan, phénylalanine et acide salicylique) sur la résistance du soja à plusieurs pathogènes (*Macrophomina phaseolina* Tassi Goid. ; *Phytophthora sojae* Kaufm. & Gerd. et *Sclerotinia sclerotinium* Lib. de Bary). Lors de cette expérience, il a été observé une meilleure survie des plantes ayant été élicitées et infectées en laboratoire. Les essais en champs n'ont cependant pas montré de différences significatives entre ces conditions testées.

Un dernier exemple d'étude de la réponse du soja à un éliciteur (issus encore du soja lui-même) est celle de Lee *et al.*, en 2018 où l'application des peptides GmPep1, GmPep2 et GmPep3 en condition d'infection par les nématodes *M. incognita* ou *H. glycines* a permis de protéger le soja. La transcription de gènes de réponse aux nématodes a été détectée lors de cette étude. De plus, GmPep1 semble induire la production de son précurseur (GmPROPEP1), une RBOHD (*Respiratory Burst Oxidase protein D*) ainsi qu'une PME1 (*Pectin Methylesterase Inhibitor*) ou encore l'accumulation de ROS (*Reactive Oxygen Species*). Les peptides GmPep2 et GmPep3 semblent aussi permettre l'accumulation de ROS ainsi que des RBOHD.

Parmi ces exemples, différents types d'éliciteurs ont donc été testés sur le soja : des composés glycaniques (Parker *et al.*, 1988 ; Graham *et al.*, 2003), des peptides (Pearce *et al.*, 2010 ; Yamaguchi *et al.*, 2011 ; Lee *et al.*, 2018) et l'application directe de phytohormones (Pawlowski *et al.*, 2016). Cependant, PEP-13, un éliciteur peptidique provenant d'un pathogène du soja (*P. sojae*) et bien étudié chez le persil ou encore la pomme de terre (*cf.* III-2-1), a très peu été étudié sur le soja.

Afin d'étudier au mieux la réponse de défense du soja, l'utilisation de techniques à haut débit comme le séquençage des ARN (RNAseq) ou encore l'étude du protéome (protéomique) en condition de stimulation immunitaire semblent être la plus adaptée. Cependant, afin de pouvoir utiliser ces techniques, un génome de référence et de bonnes connaissances de celui-ci sont nécessaires.

III-2-3 Informations sur le génome du soja et études « omiques »

Les techniques de séquençage à haut débit se sont démocratisées depuis leurs premiers pas en 2000 avec le génome d'*Arabidopsis thaliana* (première plante dont le génome ait été entièrement séquencé). Le génome du soja (variété Williams 82) a fait l'objet d'un premier séquençage et d'une annotation en 2010 par Schmutz *et al.* D'autres séquençages ont depuis été menés en 2015 sur la variété Enrei (Shimomura *et al.*, 2015) puis en 2018 sur les variétés Lee (consortium, pas de publication associée) et Zhonghuang 13 (Shen *et al.*, 2018). Ces différents séquençages ont permis d'estimer la taille du génome du soja à environ 1 Gb pour $2n = 40$ chromosomes (Doyle *et al.*, 2003) (979,046 Mb pour la variété Williams 82 de Schmutz *et al.*, contre 1017,57 Mb pour la variété Zhonghuang de Shen *et al.*) et d'estimer le nombre de gènes codants des protéines à 46 430 (Schmutz *et al.*, 2010), contre 27 029 gènes codants pour *Arabidopsis thaliana* (Swarbreck *et al.*, 2007). Un deuxième séquençage du génome et une nouvelle annotation de la variété Williams 82 ont été conduits en 2016 (JGI Phytozome, 2020), portant le nombre de gènes à 59 847 et 88 647 transcrits dont 71 219 protéines. Ces connaissances sur le génome du soja ainsi que son annotation permettent donc d'entreprendre des études de transcriptomique et de protéomique sur celui-ci.

III-2-3-1 Les études transcriptomiques de la racine du soja

Les nouvelles connaissances sur le génome du soja et ses transcrits ont conduit à un grand nombre d'études à haut débit de son transcriptome (Tableau 1). Dès 2010, plusieurs travaux portant sur les différents organes (Severin *et al.*, 2010), les stress abiotiques (Ge *et al.*, 2010 ; Song *et al.*, 2016), ou biotiques (Libault *et al.*, 2010 ; Lanubile *et al.*, 2015 ; Kang *et al.*, 2018 ; Chang *et al.*, 2019) ont été réalisés.

En 2010, Severin *et al.*, présentent un atlas des gènes transcrits en fonction des différents organes de la plante (racines, nodules, fleurs, feuilles et graines à différents stades). Cette étude RNAseq a mis en évidence des profils transcriptionnels distincts mettant en avant des variations importantes du transcriptome en fonction des organes, des types tissulaires et cellulaires ou du stade de développement. Une étude en microarray de Haerizadeh *et al.*, en 2011 a aussi comparé différents territoires cellulaires, il s'agissait du méristème racinaire, comparé à un mélange de zones non-méristématiques (zones de différenciation de la racine, de la feuille et de la tige). Ils ont identifié 5012 gènes régulés positivement ainsi que 4136

gènes régulés négativement. Parmi les gènes les plus positivement régulés de la zone méristématique, sont retrouvés des gènes impliqués dans la défense et le métabolisme. Ces travaux ont donc montré une spécificité de cette zone racinaire par rapport aux tissus non méristématiques.

Ge *et al.*, (2010) identifient sur la racine du soja sauvage (*Glycine soja* (L.) Siebold & Zucc) des gènes dont l'expression est modulée par un stress alcalin grâce à une puce de microarray spécifique au soja (Affymetrix® Soybean GeneChip®). Ils ont notamment mis en évidence un début de réponse de la racine après 3 heures de traitement, avec un pic de l'expression différentielle des gènes après 6 heures de traitement. Par la suite, l'étude RNAseq de Song *et al.*, en 2016 s'est portée sur la réponse des racines du soja au stress de déficit hydrique. Ils ont observé une expression différentielle de 6609 gènes, dont des gènes impliqués dans la réponse aux hormones végétales (auxine et éthylène), dans le métabolisme des glycomolécules ou molécules associées à la paroi (*Xyloglucan endoTransglucosylase/Hydrolase* (XTH), lignine, flavonoïdes...) et des facteurs de transcription permettant la régulation de la croissance racinaire.

Libault *et al.*, en 2010 examinent la réponse des poils absorbants à une inoculation par *Bradyrhizobium japonicum* Kirchner (bactérie symbiotique formant des nodules) et mettent en évidence grâce à une technique RNAseq Illumina l'expression différentielle de 1973 gènes de réponse à l'infection, dont des gènes homologues à NFR5 (récepteur kinase des facteurs Nod) ou NIN (facteur de transcription bZIP). En 2015, Lanubile *et al.*, se sont intéressés à la réponse des racines du soja lors d'une infection par deux souches de *Fusarium oxysporum* Schldl., l'une pathogène du soja et l'autre inoffensive. Les résultats montrent une réponse très marquée après 72 heures d'infection. De plus, la souche pathogène induit une réponse plus importante de la racine de soja que la souche non pathogène (un plus grand nombre de gènes fortement différentiellement exprimés), soulignant une différence de réponse des racines en fonction de la virulence du microorganisme les infectant.

En 2018, Kang *et al.*, analysent par RNAseq la réponse du soja à l'infection par le nématode *Heterodera glycines* Ichinohe, en présence ou non de *Bacillus simplex* Meyer & Gottheil, une bactérie PGPR (*Plant Growth Promoting Rhizobacteria*). Au niveau racinaire, 6792 gènes sont communément exprimés en présence ou en absence de la bactérie bénéfique. Ces résultats ont été complétés par une analyse métabolomique et ont montré

que les plantes inoculées avec la bactérie bénéfique produisaient une plus grande quantité de métabolites nématocides. Cette expérience a donc permis d'observer un rôle de biocontrôle de la bactérie *B. simplex* sur le soja grâce au couplage des méthodes de transcriptomique et de métabolomique.

Tableau 1 : Récapitulatif des différentes études du transcriptome du soja

Ref	Variété	Objet d'étude	Contexte	Méthode utilisée	Résultats
Severin et al., 2010	A81-356022	Jeunes feuilles ; fleurs ; gousse ; coquille de gousse ; graines ; racine ; nodule	Réalisation d'un atlas des gènes	RNAseq	(1) Mise en évidence de profils transcriptionnels en fonction de l'origine des tissus (racine, partie aérienne, graine) ainsi qu'au sein d'un même tissu en fonction de son stade de développement. (2) Mise en évidence d'une corrélation entre l'expression et la longueur des gènes. (3) Identification de 177 gènes responsables du remplissage des graines.
Ge et al., 2010	Glycine soja (soja sauvage)	Racine	Stress alcalin (NaHCO ₃)	Microarray	Mise en évidence d'une réponse à partir de 3h de traitement et identification de gènes de réponse à ce stress.
Libault et al., 2010	Non spécifié	Poils absorbants	Inoculation avec <i>B. japonicum kirchner</i>	RNAseq et microarray	Mise en évidence de l'expression différentielle de 1973 gènes de réponse à l'inoculation
Haerizadeh et al., 2011	Bragg	Méristème racinaire et mélange de feuilles, tiges, racine sans méristème	Comparaison du méristème racinaire aux tissus non méristématiques	Microarray	(1) Mise en évidence d'une spécificité du méristème racinaire par rapport au reste de la plante. (2) Identification de 502 gènes régulés positivement et 4136 gènes régulés négativement dans la zone méristématique par rapport au reste de la plante.
Lanubile et al., 2015	Forrest	Racines	Infection avec une souche de <i>Fusarium oxysporum Schiltl.</i> pathogène ou une souche non pathogène	RNAseq	(1) Mise en évidence de changement du transcriptome plus important en réponse à la souche pathogène ; (2) observation d'une réponse très marquée à partir de 72h d'infection.
Song et al., 2016	Williams 82	Racine	Stress de déficit hydrique	RNAseq	Observation de l'expression différentielle de 6609 gènes dont les gènes impliqués dans la réponse à l'auxine et l'éthylène, le métabolisme de glycomolécules et de facteurs de transcription impliqués dans la régulation de la croissance racinaire.
Kang et al., 2018	Liao15	Racines	Infection par le nematode <i>Heterodera glycines ichinoe</i> en condition d'inoculation avec la PGR <i>Bacillus simplex Meyer & Gottlieb</i> Sneb545 ou non	RNAseq et métabolomique	(1) Mise en évidence de 6792 gènes communs entre la condition inoculée avec <i>B. simplex</i> ou non. (2) Complémentation des résultats de transcriptomique par une étude métabolomique montrant une plus grande quantité de métabolites nématocides dans les plantes inoculées avec <i>B. simplex</i> par rapport aux plantes non inoculées avec celle-ci.

III-2-3-2 Les études protéomiques de la racine du soja

La protéomique est également un outil d'intérêt pour analyser les réponses du soja aux stress biotiques et abiotiques (Tableau 2).

Komatsu *et al.*, en 2009 ont réalisé une étude explorant la réponse de plantules de soja à un excès d'eau en couplant analyses transcriptomique et protéomique par Nanospray LTQ XL Orbitrap MS. Un ralentissement important de la croissance racinaire lors du stress, ainsi qu'une augmentation de l'expression de gènes impliqués dans la synthèse de l'éthylène, la fermentation alcoolique, la défense contre les pathogènes et le relâchement de la paroi ont été observés. Une diminution de l'expression de gènes codants pour des inhibiteurs de protéases a également été mise en évidence.

En 2010, Alam *et al* ont suivi la réponse protéomique des racines de soja lors d'un déficit hydrique. Le protéome des racines a été analysé en spectrométrie de masse MALDI-TOF. L'expression différentielle de protéines impliquées dans le métabolisme des glycomolécules et de l'azote, les modifications de la paroi, la transduction des signaux, la défense de la cellule ainsi que la tolérance à la sécheresse ont été mises en évidence. La même année, Sobhanian *et al.*, ont étudié la réponse du soja face à un stress salin et ont observé, par spectrométrie de masse MALDI-TOF, la régulation positive ou négative de quatorze protéines sur les 235 détectées dans les racines. Parmi ces protéines, cinq régulées négativement correspondent à des protéines du métabolisme de la plante, et deux protéines régulées positivement correspondent à des protéines de défense. Les autres étant des protéines non identifiées. Ces résultats suggéraient donc une baisse du métabolisme au profit de la mise en place de défenses pour lutter contre le stress salin.

D'autres études comme celles de Wan *et al.*, en 2005 (utilisant un MALDI-TOF MS/MS) et de Bai *et al.*, en 2019 (utilisant un ITRAQ LC-MS/MS), se sont portées respectivement sur la réponse du soja à une inoculation avec *Bradyrhizobium japonicum* ou *Funneliformis mosseae* Walker & Schüßler. *F. mosseae* est un champignon mycorhizien à arbuscules (AMF : *Arbuscular Mycorrhizal Fungus*) et *B. japonicum*, une bactérie symbiotique fixatrice d'azote. Les travaux de Bai *et al.*, ont mis en évidence 131 protéines différentiellement exprimées dans la racine de soja lorsque la plante est inoculée avec le champignon. Parmi les protéines dont l'expression est régulée positivement, des protéines liées à la défense, à la croissance et au développement de la plante ont été identifiées. Les travaux de Wan *et al.*, ont montré

l'induction de la synthèse de protéines de défense telles que des lipoxygénases, des agglutinines ou des lectines, mais aussi de nouvelles protéines qui n'avaient pas encore été décrites dans la réponse à *la bactérie*, telles qu'une phospholipase D ou une phénylalanine-amonia lysase. De plus, dans cette étude, la comparaison de la réponse de la racine (sans poils absorbants) à celle des poils absorbants a montré des différences entre ces deux échantillons, portant sur la production d'une chitinase et d'une phosphoenolpyruvate carboxylase.

Une autre étude du protéome de soja en MALDI-TOF MS/MS a comparé le protéome de l'apex racinaire à celui de la zone différenciée de la racine (Mathesius *et al.*, 2011). 342 protéines présentes à la fois dans l'apex racinaire et dans la zone différenciée ont été identifiées. Cependant les abondances de ces protéines différaient entre ces deux territoires racinaires. Il a donc été mis en évidence une différence de protéomes au sein des tissus de la racine. Ces résultats sont en accord avec ceux décrits par, Wan *et al.*, en 2005, et Severin *et al.*, en 2010.

Les différents territoires cellulaires de la racine, et notamment ceux du soja, présentent donc des spécificités transcriptomiques et protéomiques se caractérisant par des réponses distinctes face à des stress biotiques et abiotiques (Wan *et al.*, 2005 ; Mathesius *et al.*, 2011 ; Severin *et al.*, 2010 ; Libaut *et al.*, 2010 ; Haerizadeh *et al.*, 2011). Ces différences de réponse ont aussi été récemment mises en évidence dans une revue de Chuberre *et al.*, en 2018 grâce à plusieurs travaux menés sur la plante modèle : *Arabidopsis thaliana*. Parmi toutes ces études du transcriptome et du protéome ayant montré une différence entre les tissus ou types cellulaires, aucune n'a pour l'instant étudié le transcriptome ou le protéome des AC-DC de racine par approches à haut débit.

Tableau 2 : Récapitulatif des différentes études du protéome du soja

Ref	Variété	Objet d'étude	Contexte	Méthode utilisée	Résultats
Wan et al., 2005	Williams 82	Racine et poils absorbants	Inoculation avec <i>B. japonicum kirchner</i>	MALDI-TOF MS/MS	(1) Mise en évidence d'une différence de réponse entre la racine et les poils absorbants. (2) Observation de la synthèse de protéines de défense (lipoxygénases, agglutinines, lectines) ainsi que de nouvelles protéines n'ayant pas été décrites au préalable dans la réponse à <i>B. japonicum</i> (phospholipase D, phénylalanine-amonia lyase)
Komatsu et al., 2009	Enrei	Jeune plante	Stress de surplus hydrique	Nanospray LTQ XL Orbitrap MS et RNAseq	(1) Mise en évidence d'un ralentissement important de la croissance racinaire et augmentation de l'expression de gènes de la synthèse de l'éthylène, de la fermentation hydroalcoolique, de la défense contre les pathogènes et relâchement de la paroi. (2) Identification de 34 protéines différentiellement exprimées après 12h de stress.
Sobhanian et al., 2010	Enrei	Racines; feuilles; hypocotyle	Stress salin	MALDI-TOF MS	Observation de l'expression différentielle de 14 protéines, parmi ces protéines, 5 sont régulées négativement et correspondent à des protéines du métabolisme de la plante et deux protéines sont régulées positivement et sont des protéines de défense.
Alam et al., 2010	Taegwang	Racines	Stress de déficit hydrique	MALDI-TOF MS	(1) Mise en évidence de l'expression différentielle de protéines impliquées dans le métabolisme des glycomolécules et de l'azote ainsi que des protéines impliquées dans la modification de la paroi, la transduction des signaux, et dans la tolérance à la sécheresse. (2) Observation d'une accumulation de la proline, des ROS et d'une peroxydation des lipides.
Mathesius et al., 2011	Bragg	apex racinaire ; zone de différenciation	Comparaison des tissus	MALDI-TOF MS/MS	(1) Identification de 432 protéines présentes dans les deux zones. (2) Observation d'une expression différentielle de ces protéines.
Bai et al., 2019	Heinong 48	Racine	Inoculation avec <i>Funnelliformis mosseae</i> Walker & Schüßler	ITRAQ LC-MS/MS	Mise en évidence de l'expression différentielle de 131 protéines dont des protéines de défense, impliquées dans la croissance ou le développement de la plante qui étaient positivement régulées par l'inoculation.

IV. Problématique et objectifs

Ce projet de thèse a été financé par la région Normandie *via* la Structure Fédérative de Recherche Normandie Végétal FED 4277 grâce à un RIN Doctorant 100% et au RIN Recherche EPURE (*Enhancing Plant nUtRition and hEalth*). Ce projet était porté par les laboratoires AGHYLE (UniLaSalle), UMR EVA (Université de Caen Normandie), LMSM et Glyco-MEV de l'Université de Rouen Normandie. Le projet EPURE comportait trois objectifs : (1) obtenir de meilleures connaissances des mécanismes de nutrition des plantes, (2) mieux comprendre les mécanismes de la défense racinaire et (3) le rôle des microorganismes dans la nutrition et la santé des plantes.

Ce projet de thèse s'inscrivait dans l'objectif n°2 de ce programme : mieux comprendre et caractériser les mécanismes de défense racinaire. Pour cela nous nous sommes intéressés au RET et à deux de ses acteurs majeurs encore peu étudiés : les AC-DC (root-associated, cap-derived cells) et l'ADN extracellulaire. Le modèle végétal principal choisi a été le soja pour son intérêt mondial et bientôt régional et en raison des études antérieures du RET de ce modèle réalisées au sein du laboratoire Glyco-MEV. Dans ce cadre, nous avons donc deux objectifs :

I. **L'étude transcriptomique et protéomique des AC-DC du RET du soja.** En effet, de nombreuses études de transcriptomique et de protéomique ont été réalisées à partir de la racine ou des parties aériennes de cette plante, mais aucune n'a été réalisée sur les AC-DC, qui sont pourtant des cellules vivantes, métaboliquement actives et impliquées dans la défense des plantes. Une étude précédente au laboratoire GlycoMEV, réalisée par des approches d'imagerie a permis d'obtenir des données sur ces AC-DC et notamment sur la composition et l'organisation de leur paroi et du mucilage associé. Cette étude a aussi permis de mettre en évidence des interactions du RET avec des oomycètes du genre *Aphanomyces* et *Phytophthora* (Ropitiaux 2018). D'autres études ont montré la réponse des AC-DC d'une autre Fabacée (le pois) à une élicitation par approche microscopique et l'observation de l'activation de gènes de défense par qRT-PCR (Cannesan *et al.*, 2011 ; 2012 ; Plancot *et al.*, 2013). Afin de poursuivre le travail engagé et de mieux comprendre le rôle des AC-

DC dans la défense de la racine, nous avons choisi des méthodologies à haut débit de transcriptomique RNA-seq et de protéomique LC-Orbitrap.

Pour cela, nous avons dû mettre au point des techniques permettant la récolte des AC-DC puis l'extraction de leurs ARN et protéines en quantité et qualité suffisantes pour ces analyses. Les résultats de ces travaux seront présentés dans le chapitre 1 des résultats.

II. **La caractérisation génomique de l'ADNex du RET du soja.** L'ADNex a déjà été observé à de nombreuses reprises par imagerie chez le soja (Ropitiaux 2018) et chez une autre Fabacée, le pois (Wen *et al.*, 2007 ; Tran *et al.*, 2017). L'ADNex a été séquencé pour la première fois en 2009 chez le pois (Wen *et al.*, 2009). Cependant, les questions relatives à son origine, à la manière dont il est libéré, à l'effet de stimulations immunitaires sur sa production, ainsi qu'à son rôle dans le RET restent à élucider. Afin de répondre autant que possible à ces questions, nous avons dû mettre au point une technique de récolte stérile de l'ADNex, ainsi qu'une technique d'extraction permettant de le récupérer en quantité et qualité suffisantes, pour réaliser un séquençage haut débit DNA-seq de celui-ci. L'ensemble de ce travail sera présenté dans le chapitre 2 des résultats.

Les résultats obtenus sur le modèle soja nous ont permis d'apporter des connaissances supplémentaires sur ces éléments du RET que sont les AC-DC. Ils nous ont permis de transposer les méthodologies développées en transcriptomique et protéomique à un autre modèle agronomique d'intérêt régional (le lin) afin d'évaluer si les résultats obtenus sont spécifiques à l'espèce ou s'ils peuvent être généralisés. Cette étude préliminaire en cours de finalisation est présentée en Annexe 1.

Ces deux espèces (soja et lin) sont séquencées mais les approches "omiques" développées dans notre étude restent limitées lorsque les génomes, transcriptomes et annotations de référence ne sont pas optimums comme tel est le cas pour ces deux modèles agronomiques. Nous avons donc voulu compléter nos connaissances sur les AC-DC en utilisant également une approche transcriptomique sur la plante modèle *Arabidopsis thaliana* en réalisant un gène atlas associé à une analyse méta-transcriptomique. Ces données préliminaires en cours de finalisation sont présentées en Annexe 2.

Matériels et méthodes

I. Matériel biologique et éliciteurs

I-1 Stérilisation des graines

I-2 Mise en culture des graines

I-3 Récolte des AC-DC et de l'ADN extracellulaire (ADNex)

II. Analyse transcriptomique des AC-DC et des racines

II-1 Récolte et séquençage de l'ARN des AC-DC et des racines

II-2 Analyse bio-informatique

III. Protéomique des AC-DC et racines

III-1 Extraction des protéines solubles des AC-DC et des racines

III-2 Analyse par spectrométrie de masse

III-3 Analyse bio-informatique

IV. Analyse génomique de l'ADN extracellulaire (ADNex)

IV-1 Purification et quantification de l'ADNex

IV-2 Analyse qualité, préparation des bibliothèques et séquençage

IV-3 Analyse bio-informatique

Matériels et méthodes

I. Matériel biologique et éliciteur

Dans cette étude, la plante utilisée est le soja (*Glycine Max* (L.) Merr.) de la variété CASTETIS. Cette variété de soja était recommandée aux agriculteurs en 2010 par le réseau commun CTPS/CETIOM pour sa faible sensibilité à la verse et sa richesse en protéines, ce qui lui a valu d'être choisie pour les études précédentes réalisées au laboratoire Glyco-MEV. Dans un souci de continuité des analyses engagées précédemment sur le RET du soja, la variété CASTETIS a donc été conservée dans ce projet. Les graines ont été fournies par la coopérative agricole La Dauphinoise (Vienne, France).

Afin de provoquer une réponse de défense des plantes de soja, celles-ci ont été mises en contact avec l'éliciteur PEP-13, peptide issu à l'origine des oomycètes du genre *Phytophthora*, dont certaines espèces sont connues pour être pathogènes du soja. Ce peptide a été synthétisé par le service de synthèse de peptides et criblage fonctionnel (SPCF) de la plateforme PRIMACEN (Université de Rouen Normandie, France). La séquence de ce peptide est la suivante : VWNQPVRGFKVYE, et sa pureté a été estimée à 99,9% par HPLC.

I-1 Stérilisation des graines

Les graines sont placées dans des pots à prélèvements ouverts disposés sous une cloche à vide elle-même placée sous une sorbonne (Figure 23, A). Un bécher contenant 100 mL de javel (hypochlorite de sodium à 9,6° de chlore actif) est déposé au centre de la cloche. 3 mL d'acide chlorhydrique fumant (ACS, Reag. ISO, Reag. Ph. Eur., fumant, $\geq 37\%$) sont ensuite ajoutés très rapidement dans le bécher de javel et la cloche est aussitôt refermée pour conserver le gaz dichlore immédiatement formé. Les graines de soja sont ainsi incubées toute la nuit (environ 17 heures). Après incubation, la cloche est ouverte sous la sorbonne afin d'évacuer le gaz résiduel. Une seconde incubation d'environ 17 heures dans le dichlore est réalisée à nouveau selon la même procédure décrite précédemment. Une fois ce second temps d'incubation terminé, un maximum de gaz est évacué stérilement à l'aide d'une pompe munie d'un filtre à 0,2 μm sous la sorbonne pendant cinq minutes. La cloche est ouverte et les pots de graines sont refermés rapidement pour maintenir leur stérilité. Les graines sont ensuite conservées à 4°C en attendant leur utilisation.

I-2 Mise en culture des graines

Sous une hotte à flux laminaire, les graines sont rincées avec de l'eau pure stérile et mises à imbiber dans de l'eau pure stérile environ 17 heures à l'obscurité. Après imbibition, dix graines sont déposées stérilement par boîte de Pétri carrée (120 mm), contenant un disque de papier Whatman (diamètre 11 cm, grade 91, GE Healthcare) imbibé avec 4,75 mL de milieu $\frac{1}{2}$ MS (Duchefa Biochemie) liquide stérile à pH 5,7 +/- PEP-13 à 2 $\mu\text{g}/\text{mL}$ (soit 1,233 $\mu\text{mol}/\text{L}$). Les boîtes sont ensuite scellées avec du film alimentaire et placées cinq jours dans une enceinte de culture à 26°C avec une photopériode de 18h jour et 6h nuit, 80-90% RH (Figure 23, B et C).

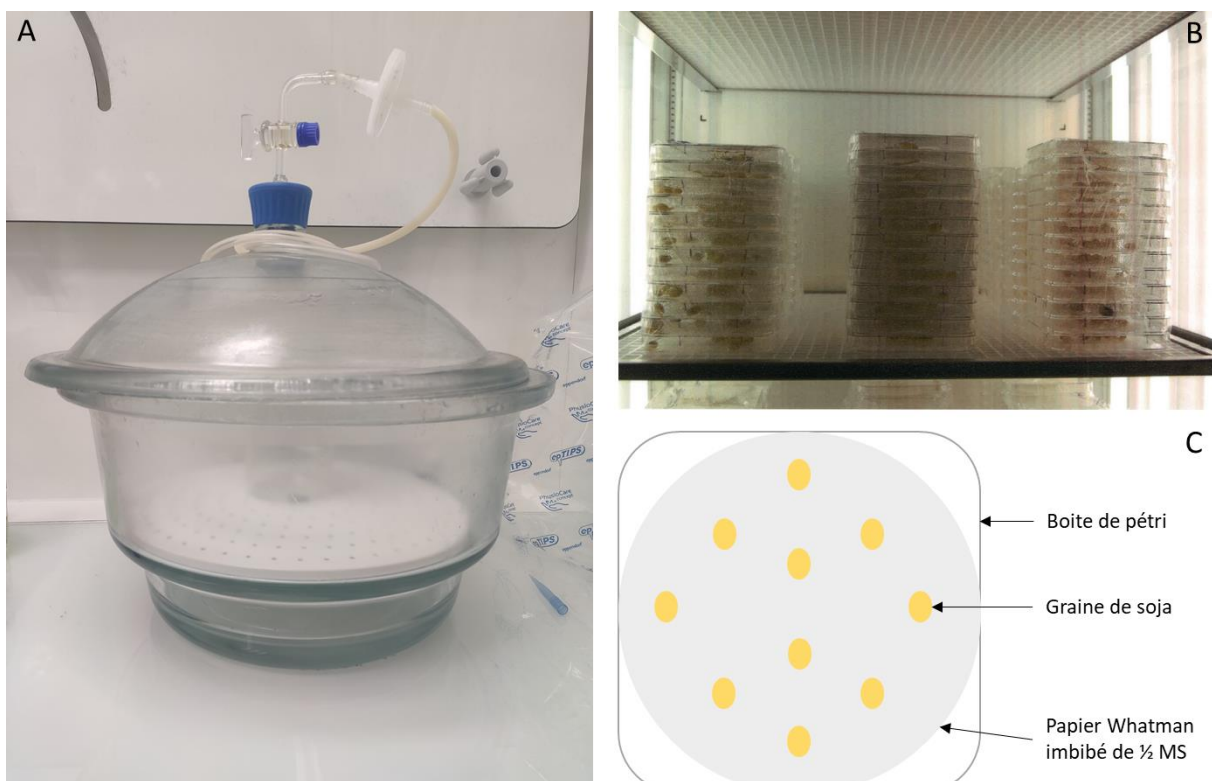


Figure 23: Systèmes de stérilisation et de mise en germination des graines de soja

A : Cloche à vide permettant la stérilisation des graines dans le gaz dichlore ; B : Enceinte de culture contenant des boîtes de graines de soja, les boîtes sont scellées et empilées avec du film alimentaire ; C : Schéma du système de mise en germination des graines de soja en boîte de Pétri.

I-3 Récolte des AC-DC et de l' ADN extracellulaire (ADNex)

Sous une hotte à flux laminaire, dans une plaque 24 puits stérile (Nunclon™ Delta Surface, Thermo Fischer Scientific), 400 µL d'eau exempte de nucléases (*Water, Sterile Nuclease-free*, TFS Fisher BioReagents) sont déposés dans chaque puit. Cinq racines sont prélevées stérilement à l'aide d'une pince et d'un scalpel et déposées dans chaque puit (Figure 24). Entre chaque boîte, les outils de prélèvement (pince et scalpel) sont stérilisés 10 secondes dans un stérilisateur à billes chauffé à 250°C (Steri 250, Keller). Lorsque la plaque est remplie, les racines sont agitées avec une pince afin de décrocher leur RET, et cinq racines sont à nouveau placées puis agitées dans chaque puit. Au total, chaque puit contient donc le RET de dix plantes. Les racines sont gardées dans un pot à prélèvement et conservées dans un congélateur de laboratoire à - 80°C.

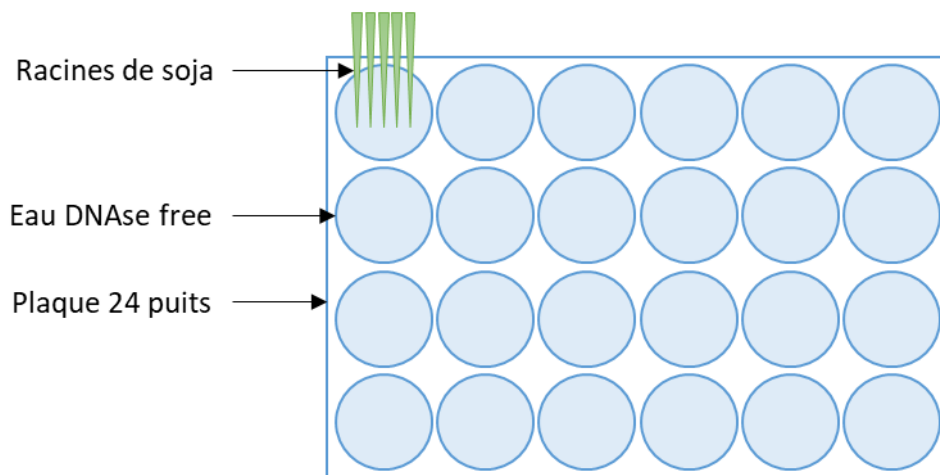


Figure 24 : Système de récolte du RET de soja en puits

Les racines sont déposées dans les puits et agitées afin d'en décrocher le RET. Au total, chaque puit contient donc le RET de dix plantes.

Un aliquot de chaque puit est prélevé à l'aide d'une anse stérile à usage unique et étalé sur du milieu LB solide (1,5% d'agar bactériologique, European bacteriological agar, Conda SA) coulé dans deux autres plaques 24 puits afin de tester la stérilité des RET récoltés. Ces plaques de LB sont ensuite incubées 48 heures, l'une à 37°C et l'autre à 26°C (Figure 25).

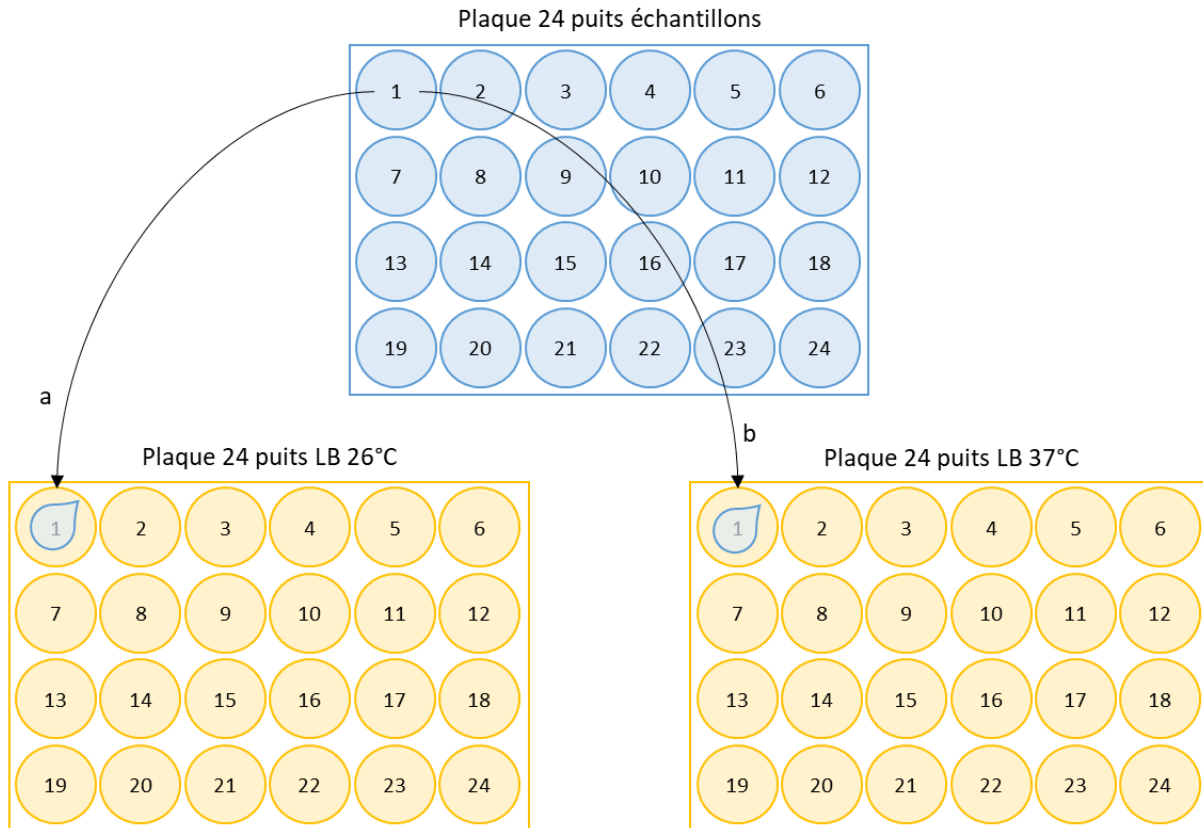


Figure 25 : Tests de stérilité des échantillons de RET du soja

Un aliquot de chaque échantillon (n°1 à 24) est prélevé à l'aide d'une anse stérile, deux dépôts sur milieu LB sont réalisés, le premier (a) sur une plaque qui sera stockée 48h à 26°C, et le second (b) sur une plaque qui sera stockée 48h à 37°C.

Le contenu des puits de RET est ensuite transféré dans un microtube 1,5 mL (FlipTube, GEMÜ GmbH) et centrifugé 10 min à 3000 g afin de culoter les AC-DC. Le surnageant est prélevé et déposé dans un nouveau tube 1,5 mL, celui-ci contient l'ADNex. Les tubes contenant les culots d'AC-DC sont conservés dans un congélateur de laboratoire à -80°C. Les tubes de surnageant sont incubés 10 minutes à 100°C afin de dénaturer les protéines et les enzymes susceptibles de dégrader l'ADN des échantillons. Ceux-ci sont conservés à -80°C. Au bout de 48 heures, les tubes sont triés selon le résultat des tests de stérilité réalisés précédemment. Ceux qui présentent une contamination sont éliminés, les autres sont conservés pour la suite des analyses.

Les surnageants contenant l'ADNex sont traités à la protéinase K (Qiagen). Pour cela 20 µL de solution sont ajoutés dans chaque tube et ceux-ci sont incubés 30 minutes à 60°C. Suite à cela, les tubes correspondant à une même récolte et une même condition sont

regroupés entre eux et concentrés au speedvac afin d'obtenir un volume final minimum dans un seul tube.

II. Analyse transcriptomique des AC-DC et des racines

II-1 Récolte et séquençage de l'ARN des AC-DC et des racines

Les AC-DC et les racines sont récupérés et conservés dans le tampon de lyse du kit RNA/DNA Purification kit (Norgen) et broyées à l'azote liquide. Les ARN totaux des échantillons sont ensuite extraits selon les recommandations du constructeur de ce kit. Les ARN sont ensuite re-purifiés avec un kit à billes RNAClean XP (Agencourt, Beckman). La qualité des ARN est vérifiée par un bioanalyzer 2100 Agilent, puis les ARNs sont quantifiés par fluorescence avec le kit QuantiFluor® RNA system (Promega) (Tableau 3). Douze banques d'ARN messagers sont ensuite réalisées à partir de 5 ng d'ARN total à l'aide du kit NEBNext® Single Cell/Low input RNA Library Prep kit for Illumina (NEB).

Tableau 3 : Analyse de la concentration et de la qualité des 12 échantillons d'ARN totaux des racines et des AC-DC du soja

Six échantillons d'AC-DC et de racines ont été préparés. Pour ces deux zones, on retrouve trois réplicats en condition témoin ou non élicité (NE) et trois réplicats en condition élicité avec PEP-13 (E).

Echantillon	Concentration (ng/μL)	Quantité totale (ng)	RIN (RNA integrity number)
AC-DC NE R1	7,7	231,0	6,9
AC-DC NE R2	4,48	134,4	8,0
AC-DC NE R3	5,81	174,3	9,3
AC-DC E R1	16,1	483,0	5,7
AC-DC E R2	8,4	252,0	9
AC-DC E R3	10,5	315,0	10,0
Racine NE R1	13	390,0	7,2
Racine NE R2	15	450,0	8,0
Racine NE R3	14	420,0	6,7
Racine E R1	13	390,0	6,2
Racines E R2	14	420,0	5,8
Racines E R3	13	390,0	8,7

La qualité des banques a ensuite été vérifiée sur un bioanalyser 2100 Agilent, puis par qPCR. Les douze banques ont été regroupées en un pool qui est déposé sur une *flow cell High Output v2*. Les clusters sont alors amplifiés et séquencés dans un appareil Illumina NextSeq500 en *single reads* de 75 paires de bases. Ces récoltes, extractions et séquençages ont été réalisés sur la plate-forme ProfileXpert de l'Université de Lyon 1.

II-2 Analyse bio-informatique

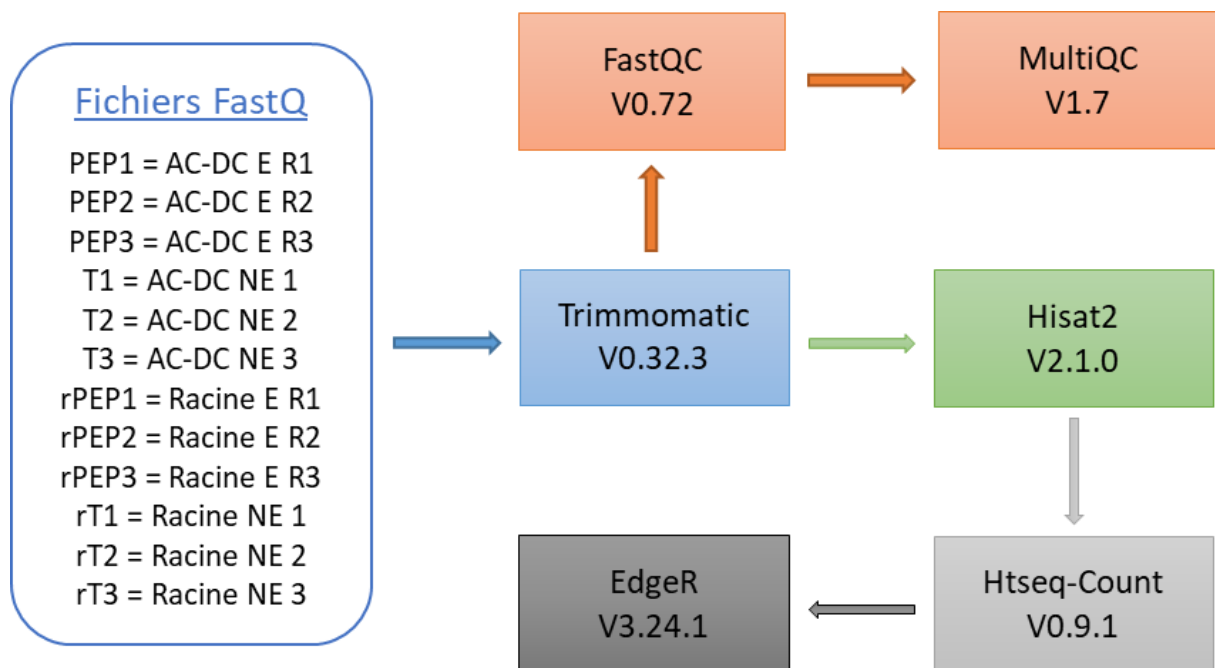


Figure 26 : Schéma du *workflow* utilisé pour l'analyse bio-informatique des données de RNAseq de racines et d'AC-DC du soja

Les différents outils de l'interface d'analyse Galaxy (<https://usegalaxy.org/>) utilisés pour cette analyse sont FastQ V0.72, MultiQC V1.7, Trimmomatic V0.32.3, Hisat2 V2.1.0, HTSeq-Count V0.9.1 et EdgeR V3.24.1. Les étapes représentées en orange correspondent à des étapes de contrôle qualité, en bleu le nettoyage des séquences, en vert l'alignement et en gris clair le comptage, et l'analyse différentielle en gris foncé. E : Elicité ; NE : Non élicité

Les fichiers FastQ fournis pour la plateforme ProfileXpert ont été téléchargés sur l'interface d'analyse Galaxy (Figure 26). Ces fichiers FastQ ont ensuite été nettoyés (suppression des *reads* de faible qualité) à l'aide de Trimmomatic V0.36.6 (Bolger *et al.*, 2014) avec les options *Illuminaclip* afin de couper les séquences d'adaptateurs Illumina et *Slidingwindow* pour enlever les bases dont le séquençage est de mauvaise qualité (Inférieur à un score *Phred* de 20). Une vérification de la qualité du séquençage a ensuite été réalisée avec FastQC (V0.72) puis MultiQC (V1.7) (Ewels *et al.*, 2016) (Figures 27 et 28).

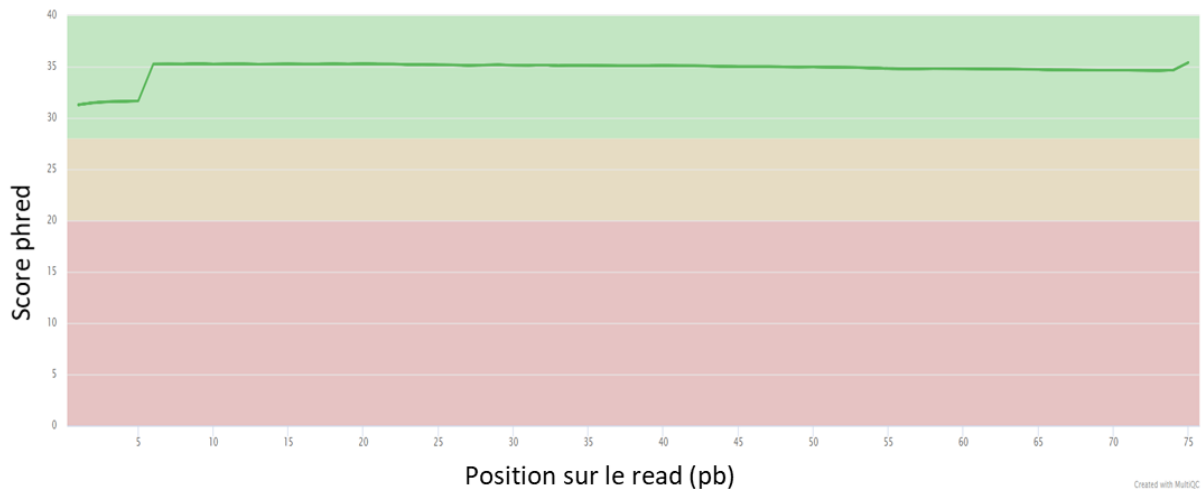


Figure 27 : Score de qualité des bases séquencées

On observe ici le score *phred* moyen des bases séquencées. Un score *phred* de 30 correspond à une précision de 99,9% et est considéré comme un seuil à partir duquel le séquençage des bases est de très bonne qualité. Le score maximal est 40, soit une précision de 99,99%. Ici, on observe donc que le séquençage est de très bonne qualité pour tous les *reads*.



Figure 28 : Pourcentage de bases N par cycle

On observe ici le pourcentage de bases N (non déterminées) en fonction de la position sur le *read*. Pour ces échantillons, une valeur très proche ou égale à 0% a été obtenue pour toutes les positions.

Les *reads* ont ensuite été alignés sur le génome du soja (*Gmax* Wm82 a4 v1, Phytozome V13) grâce à Hisat2 V2.1.0 (Kim *et al.*, 2015). Le comptage des gènes est réalisé avec HTSeq-Count V0.9.1 (Anders *et al.*, 2015) avec le fichier d'annotation (*Gmax* Wm82 a4 v1, Phytozome V13). Les fichiers de comptage ainsi obtenus ont permis de réaliser une analyse de l'expression différentielle des gènes avec EdgeR V3.24.1 (Robinson *et al.*, 2010) entre les conditions élicitée avec PEP-13 (E) vs non élicitée (NE) et AC-DC vs racines.

L'analyse de l'expression différentielle des gènes ayant permis d'obtenir pour chaque gène un FC (*fold change*) et une *p-value* pour chaque comparaison (E vs NE et AC-DC vs Racines), des *volcanoplots* ont été réalisés sur Excel afin de trier les gènes (Figure 29). Seuls ceux possédant un FC supérieur à 1,5 et une *p-value* inférieure à 0,05 ont été conservés. Ces listes de gènes ont ensuite été comparées grâce à Jvenn (Bardou *et al.*, 2014), WEGO 2.0 (wego genomics) et KEGG mapper (kegg.jp) (Figure 29). Suite à cela, les listes ont été analysées par PANTHER (pantherdb.org) avec un test d'*overrepresentation* des GO (*Gene Ontologies*) correspondant à des processus biologiques (test de Fisher et calcul du *false discovery rate*). Enfin, les identifiants des gènes des listes ont été convertis grâce à UNIPROT (uniprot.org) et la fonction *retrieve ID/mapping* et l'option de sortie *cross-ref STRING*. Cette conversion a permis d'analyser les listes grâce à Mapman (Thimm *et al.*, 2004) avec leurs identifiants et leurs FC entre les conditions élicitée ou non élicitée. Pour cela, une trame *overview* a été créée afin d'obtenir des figures personnalisées correspondant aux catégories de *bins* d'intérêt. Le logiciel Mapman utilisant sa propre annotation des gènes, des *bins*, correspondant à des *genes ontologies* (GO), ont été automatiquement attribués à chacun des gènes et ceux-ci ont été classés en fonctions de leurs *bins*.

Enfin, afin d'identifier un récepteur putatif de PEP-13 (Figure 29), les listes de protéines sur-exprimées en présence de PEP-13 dans les AC-DC et dans les racines (FC supérieur ou égal à 1,5 et *p-value* inférieure à 0,05) ont été triées afin de conserver les protéines non annotées et celles annotées comme récepteurs dans la base de données UNIPROT. Ce tri a permis d'obtenir 99 protéines chez les racines et 175 protéines chez les AC-DC. Ces deux listes ont été analysées avec DRAGO2 (PRGdb.org) afin de détecter les éventuels domaines récepteurs de ces protéines. Pour finir, une modélisation de la structure 3D d'un récepteur potentiel a été réalisée grâce à I-TASSER (Yang *et al.*, 2015).

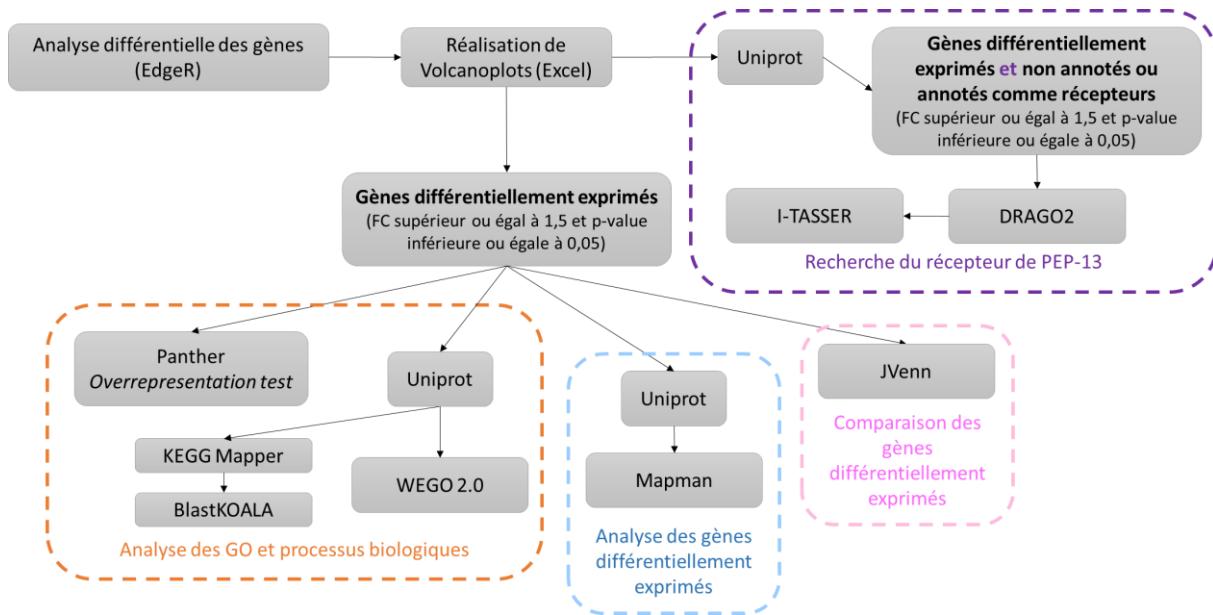


Figure 29 : Schéma du workflow utilisé pour l'analyse bio-informatique des gènes différentiellement exprimés chez les AC-DC et les racines en condition élicitée ou témoin

L'analyse différentielle avec EdgeR a permis d'obtenir un *fold change* (FC) et une *p-value* pour chacun des gènes détectés lors du séquençage ARN. Des *volcanoplots* ont ensuite été réalisés avec Excel afin de trier les gènes en fonction de leur FC (supérieur ou égal à 1,5) et de leur *p-value* (inférieure ou égale à 0,05). Les listes de gènes obtenues pour les AC-DC et pour les racines ont ensuite été comparées avec JVENN (encadré rose) et annotées avec Uniprot afin d'obtenir l'identifiant correspondant à la banque de donnée de Mapman (encadré bleu), la recherche du récepteur (encadré violet) ou encore les GO (*gene ontologies*) nécessaires à l'analyse WEGO 2.0 (encadré orange). Les séquences des transcrits (au format FASTA) ont été utilisées comme entrée dans KEGG Mapper puis BlastKOALA (encadré orange) ainsi que pour DRAGO2 puis I-TASSER (encadré violet).

III. Protéomique des AC-DC et racines

III-1 Extraction des protéines solubles des AC-DC et des racines

Les tubes contenant les AC-DC sont regroupés de façon à obtenir 540 plantes par réplicat, et repris dans du tampon d'extraction (urée 7M, thiourée 2M, 0,5% ampholytes pH 4-6,5, TBP 2 mM, 20mM DTT, 2% Chaps, 0,5% C7Bz0). Les racines sont broyées à l'azote liquide à l'aide d'un pilon et d'un mortier, puis remises en solution dans le tampon d'extraction et de nouveau broyées à l'aide d'un broyeur (Homogenizer 150, Fisherbrand). Les étapes d'extraction des protéines et d'analyse par spectrométrie de masse sont ensuite réalisées sur la plate-forme PISSARO de l'Université de Rouen Normandie.

Les échantillons sont ensuite soniqués pendant 20 minutes (pulses de 3 secondes, intensité 24%, Sonificateur Microprocesseur 750W, Sonics Materials) puis centrifugés à 1000 g pendant 30 minutes.

Afin de déterminer la quantité de protéines extraites (Figure 30, a), un dosage de Bradford est réalisé sur deux dilutions de l'échantillon : l'une à 1/10 et l'autre à 1/50. 20µL de chaque échantillon sont ensuite mélangés avec 1 mL de réactif de Bradford (Protein Assay®, Bio-Rad). Les échantillons sont alors incubés 30 minutes à l'obscurité puis leur absorbance est mesurée à 595 nm à l'aide d'un spectrophotomètre. La concentration en protéines est ensuite calculée en comparant les absorbances obtenues avec celles d'une gamme de concentrations de BSA (*Bovine Serum Albumin*, Sigma-Aldrich). Une vérification de ce dosage est réalisée en faisant migrer les échantillons sur un gel SDS-PAGE à 12 % d'acrylamide, coloré au nitrate d'argent (Figure 30, b).

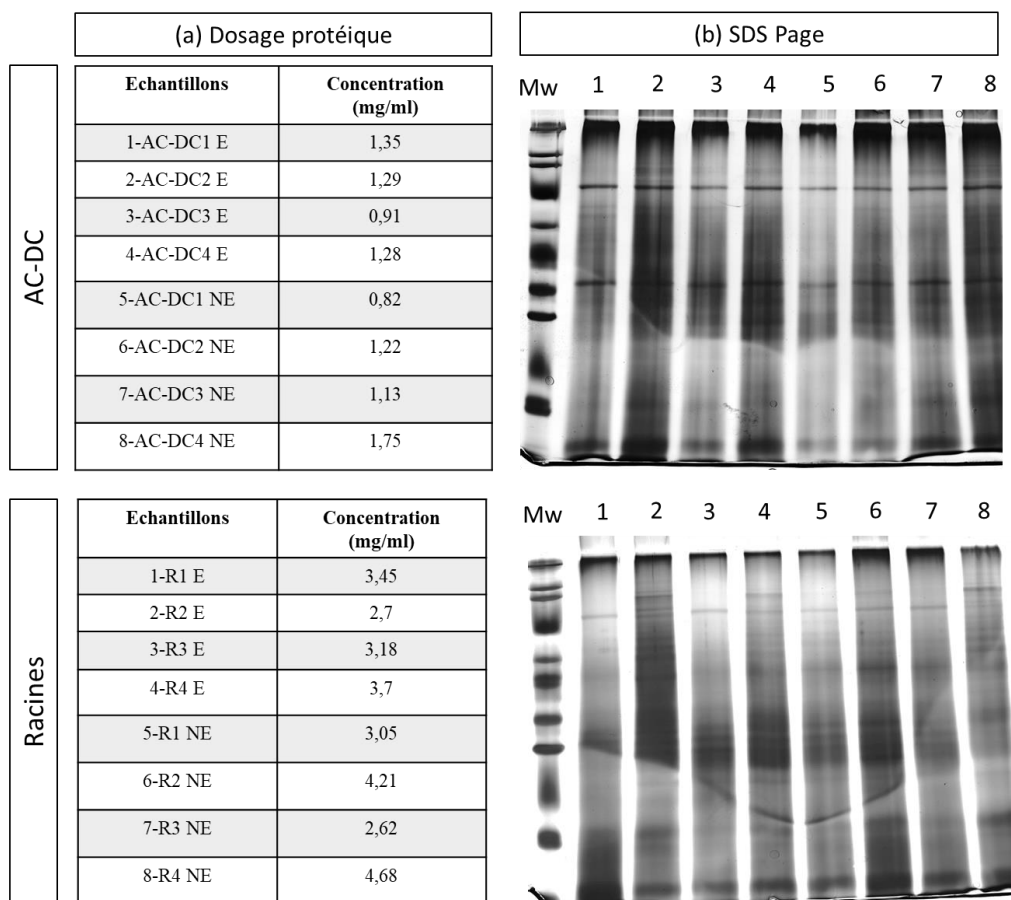


Figure 30 : Résultats du dosage protéique (a) et gels SDS-PAGE (b) pour les AC-DC (*root Associated Cap-Derived Cells*) et pour les racines de soja

Pour chaque échantillon un numéro est attribué dans la colonne « échantillon », celui-ci correspond au numéro de piste sur le gel SDS-PAGE.

30 µg de protéines sont ensuite déposés sur un gel SDS-PAGE à 7 % d'acrylamide (Acrylamide 9,3 mL ; Tris pH 6,8 5 mL ; eau pure 25,05 mL) afin de les concentrer. La migration est réalisée pendant 45 min à 10 mA puis 2 heures à 20 mA. Le gel est ensuite coloré 30 secondes dans une solution de bleu de Coomassie (acide acétique 10 mL ; éthanol

50 mL ; eau pure 40 mL ; bleu de Coomassie 0,5 g) puis décolorés à l'aide d'une solution de décoloration (acide acétique 10 mL ; éthanol 50 mL ; eau pure 40 mL) jusqu'à observation d'une bande contenant les protéines d'intérêt. Les bandes sont excisées à l'aide d'un scalpel, puis lavées trois fois dans 80 µL d'eau pure sous agitation pendant 10 minutes. Elles sont ensuite incubées deux fois dans des bains successifs de DTT (dithiothréitol 7,7 mg ; eau pure 10 mL) puis de iodoacétamide (18,5 mg ; eau pure 10 mL). Les fragments de gels sont conditionnés dans une solution de tampon carbonate/acétonitrile (V/V) suivi d'un bain d'acétonitrile. Les protéines sont digérées dans une solution trypsique à 10 µg/mL dans du tampon carbonate pendant 3 heures sous agitation à 37°C. Les peptides sont ensuite récupérés après différents bains d'eau pure, TFA 1% et d'acétonitrile. Les solutions finales sont enfin séchées au speedvac pendant 4 heures. Les échantillons sont repris dans une solution d'acide formique (Formic acid, FA) à 0,1 %.

Les échantillons ont été analysés grâce à un spectromètre de masse LTQ Orbitrap Elite (Thermo Scientific) équipé d'une source nano-ESI et couplé à une chaîne chromatographique nano-débit (Easy-nLC II, Thermo Scientific). Pour ces analyses, 1 µL de l'échantillon a été chargé sur la colonne d'enrichissement (Cap Trap C₈, 0,5 x 2 mm, Michrom bioressources) à une pression de 200 bars dans 0,1 % FA. La séparation a été effectuée sur une colonne en phase inverse (C₁₈, L 153 mm, ID 5 µm, 100 Å, Nikkyo Technos). Le gradient, d'un débit de 300 nL/min, est assuré comme suit : de 2 à 40% de tampon B (CH₃CN/0,1 % AF) en 105 min, de 40 à 80 % de B en 4 min et 80 % de B pendant 15 min (tampon A : H₂O/0,1 % AF).

III-2 Analyse par spectrométrie de masse

Le spectromètre de masse est utilisé en mode positif avec une tension de capillaire fixée à environ 1,5 kV et une température de la source fixée à 200°C. Les échantillons ont été analysés par fragmentation CID à l'aide d'une méthode référencée « Top20 ». Le balayage MS initial est enregistré dans l'analyseur Orbitrap avec une résolution de 60000 sur une gamme de masse m/z 400-2000. Les ions mono-chargés n'ont pas été sélectionnés et le seuil de signal d'intensité pour la sélection du précurseur est fixé à 500. Lors de ces cycles d'analyses, les 20 ions précurseurs les plus intenses ont été sélectionnés pour l'étape de MS/MS. Les peptides cibles déjà sélectionnés pour la MS/MS ont été exclus de manière dynamique pendant 30 secondes. La fragmentation a été réalisée dans l'analyseur trape ionique linéaire avec une énergie de collision normalisée de 35 %. Dans l'analyseur Orbitrap,

toutes les mesures ont été réalisées après calibration interne à l'aide du pic de m/z à 445,12002 correspondant à un polydiméthylcyclosiloxane (*lock mass*).

III-3 Analyse bio-informatique

Les fichiers de données brutes ont été traités à l'aide du logiciel Proteome Discoverer (Thermo Scientific ; version 1.3) sur la plateforme PISSARO (Université de Rouen Normandie). Les peptides ont été identifiés grâce à l'intégration du moteur de recherche Mascot (Matrix Science ; version 2.2) utilisé contre la base de données protéiques UNIPROT *Glycine max*. Les paramètres suivants ont été utilisés : un clivage manquant et les modifications variables concernant la carbamidométhylation (Cys) et l'oxydation (Met). La tolérance de masse sur l'ion précurseur et les ions fils ont été respectivement de 10 ppm et de 0,5 Da. Les résultats ont été filtrés pour un taux maximal de faux positifs de 5%. Lors de ces analyses, uniquement les protéines identifiées par au moins deux peptides associés à un degré de confiance élevé (*High Confidence*) ont été conservées. La comparaison des différentes conditions d'analyses par analyse en composante principale et la quantification sont réalisées par le logiciel Progenesis LC-MS (nonlinear DYNAMICS).

Les conditions élicitées (E) et non élicitées (NE) des racines et des AC-DC sont ensuite comparées en calculant un *fold change* (FC) des abondances relatives calculées pour chaque condition (Figure 31). Des *volcanoplots* ont été réalisés grâce à Excel afin de trier les protéines et de conserver celles ayant un FC supérieur ou égal à 1,5 et une *p-value* inférieure à 0,05. Ces listes de protéines ont été comparées grâce à Jvenn (Bardou *et al.*, 2014). Suite à cela, les listes ont été analysées par PANTHER (pantherdb.org) avec un test d'*overrepresentation* des GO (*Gene Ontologies*) correspondant à des processus biologiques (test de Fisher et calcul du false discovery rate). Enfin, les GO ont été obtenus grâce à Uniprot et les identifiants des gènes des listes ont été convertis grâce à UNIPROT (uniprot.org) et la fonction *retrieve ID/mapping* et l'option de sortie *cross-ref STRING*. Cette conversion a permis d'analyser les listes grâce à Mapman (Thimm *et al.*, 2004) en utilisant leurs identifiants et leurs FC entre les conditions élicitée (E) et non élicitée (NE). Pour cela, une trame a été créée afin d'obtenir des figures personnalisées correspondant aux catégories de *bins* d'intérêt. Le logiciel Mapman utilisant sa propre annotation des gènes, des *bins*, correspondant à des *genes ontologies* (GO), ont été automatiquement attribués à chacun des gènes et ceux-ci ont été classés en fonctions de leurs *bins* (Figure 31).

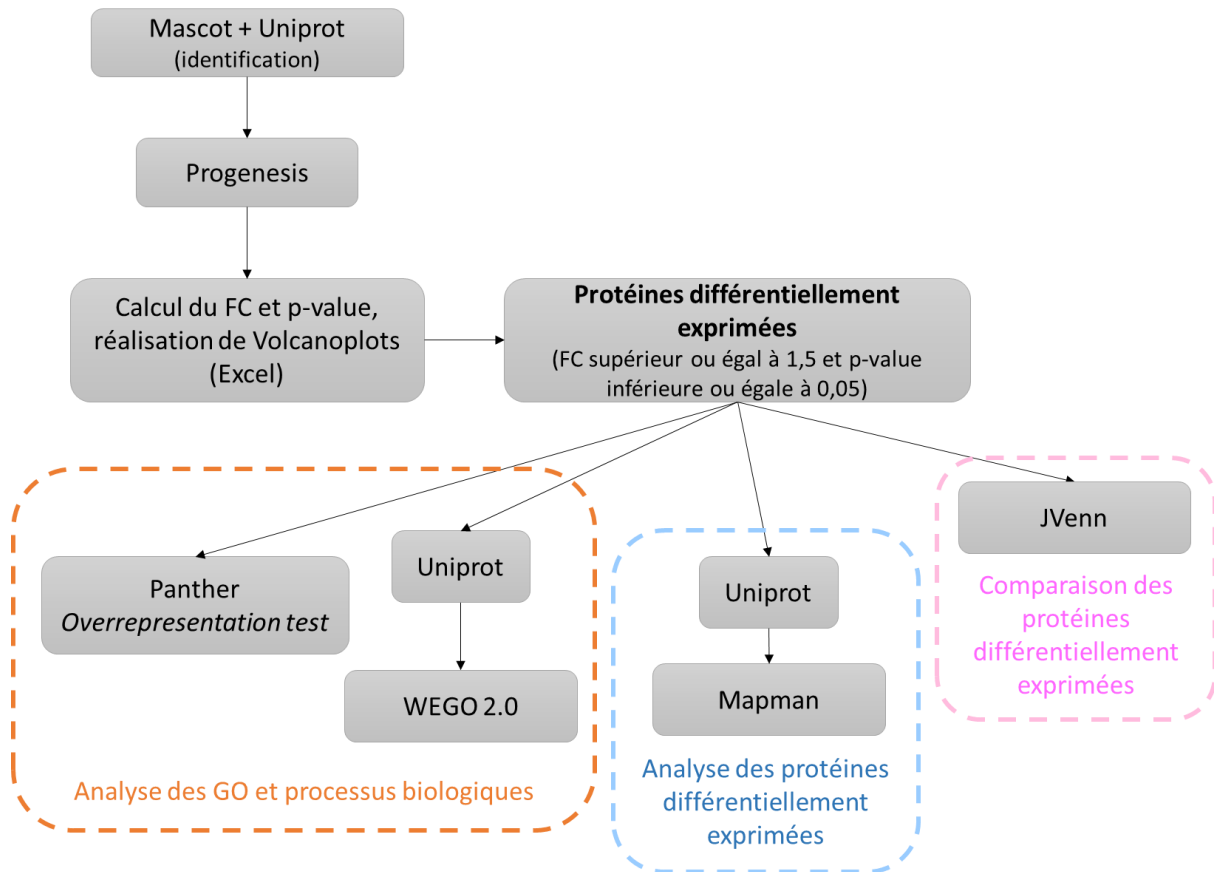


Figure 31 : Schéma du workflow utilisé pour l'analyse bio-informatique des protéines différentiellement exprimées chez les AC-DC et les racines en condition élicitee avec PEP-13 ou non

L'identification des peptides à été réalisée par Mascot avec la base de données Uniprot, une quantification et des analyses en composante principale ont été réalisées par Progenesis. Le calcul du *fold change* (FC) et de la *p-value* pour chacune des protéines détectées lors de l'analyse protéomique a permis de réaliser des *volcanoplots* avec Excel afin de trier les protéines en fonction de leur FC (supérieur ou égal à 1,5) et de leur *p-value* (inférieure ou égale à 0,05). Les listes de protéines obtenues pour les AC-DC et pour les racines ont ensuite été comparées avec JVenn (encadré rose) et annotées avec Uniprot afin d'obtenir l'identifiant correspondant à la banque de donnée de Mapman (encadré bleu), ou encore les GO (*gene ontologies*) nécessaires à l'analyse WEGO 2.0 (encadré orange).

IV. Analyse génomique de l'ADN extracellulaire (ADNex)

IV-1 Purification et quantification de l'ADNex

La purification de l'ADNex est réalisée à partir des tubes concentrés au speedvac (environ 300 plantes par réplicat). 100 μ L par échantillon sont déposés dans un microtube de 2 mL. L'ADN contenu dans ces 100 μ L d'échantillon est purifié grâce au kit DNeasy® PowerClean® Pro Cleanup (Qiagen). Il est ensuite élué dans 100 μ L de solution EB (*Elution Buffer* 10 mM Tris-HCl, pH 8.5, Qiagen). La concentration en ADN est ensuite évaluée (Tableau 4, a) par dosage spectrophotométrique (Nanodrop 2000c spectrophotometer, Thermo scientific).

Tableau 4: Dosage puis calcul de la concentration en ADNex des échantillons élicités (E) ou non (NE) du RET du soja

Echantillon	(a) Concentration en ADN (ng/ μ L)	(b) Concentration finale dans 120 μ L (ng/ μ L)
7 NE	1,7	1,42
7 E	1,3	1,08
8 NE	3,1	2,58
8 E	1,8	1,50
9 NE	2,4	2,00
9 E	1,7	1,42

IV-2 Analyse qualité, préparation de librairies et séquençage

Les échantillons d'ADNex sont complétés à 120 μ L avec de l'eau exempte de nucléases (*Water, Sterile Nuclease-free*, TFS Fisher BioReagents) (Tableau 4, b) puis sont soniqués 360 secondes (sonicateur S220-Series, Covaris) afin d'obtenir des fragments d'environ 200 paires de bases (pb). Les échantillons sont ensuite visualisés sur une TapeStation (2200 TapeStation system, Agilent) grâce au kit High Sensitivity D1000 ScreenTape System (Agilent). Les librairies ont ensuite été préparées grâce au kit NEBNext® Ultra™ II, en suivant les instructions du constructeur pour un échantillon contenant 1 ng d'ADN. Lors de l'étape 1.1.5 le temps d'incubation à 37°C est réduit à 5 min car l'ADN est déjà fragmenté. Le séquençage *paired-end* des librairies est ensuite réalisé sur un séquenceur Illumina MiSeq® avec le kit V2

Micro 300 cycles sur la Plateforme Rouennaise de génomique de l'Université de Rouen Normandie.

IV-3 Analyse bio-informatique

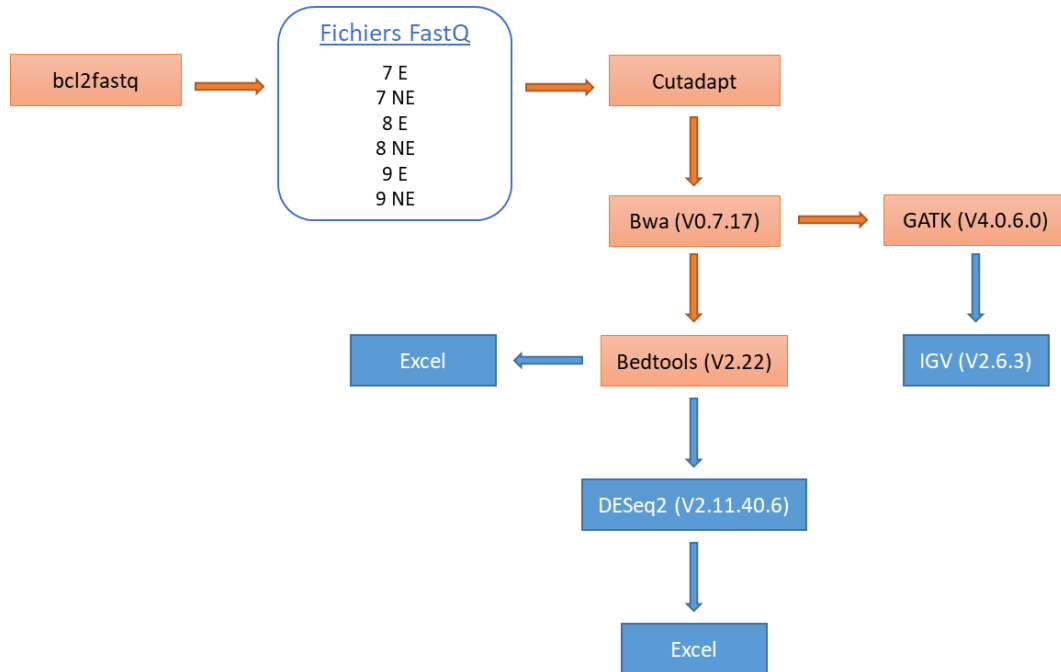


Figure 32 : Workflow réalisé pour l'analyse bio-informatique de l'ADNex du RET du soja

Les différents outils utilisés pour cette analyse sont Cutadapt, Bwa (V0.7.17), GATK (V4.0.6.0), Bedtools (V2.22) ainsi qu'IGV (V2.6.3) et DESeq2 (V2.11.40.6) de l'interface d'analyse Galaxy (<https://usegalaxy.org/>). Les étapes oranges sont celles réalisées par la plateforme Rouennaise de génomique ; les étapes bleues ont été réalisées au sein du laboratoire GlycoMEV. E : élicité ; NE : non élicité.

La première partie de cette analyse bio-informatique a été réalisée par la Plateforme Rouennaise de génomique de l'Université de Rouen Normandie. Elle comprenait (Figure 32) :

- le démultiplexage des fichiers de séquençage avec bcl2fastq (V2.02),
- l'élimination des adaptateurs grâce à Cutadapt,
- l'alignement des séquences sur le génome du soja (*Gmax Wm82 a4 v1*) effectué avec bwa (V0.7.17) (Tableau 5). L'outil GATK (V4.0.6.0) a ensuite été utilisé avec les commandes *SortSam*, *MarkDuplicates* et *Build Bam index* afin d'obtenir les fichiers d'alignement binaires (BAM) ainsi que leur index (BAI),
- le découpage du génome en 1000 pb et comptage grâce à Bedtools (V2.22) en utilisant la commande *coverage*.

Tableau 5 : Rapport qualité du séquençage de l'ADNex du RET du soja généré par BWA

Sample : échantillon ; *Genome version* : version du génome utilisée ; *Number of reads* : nombre de *reads* séquencés ; *Duplicates* : duplicats ; *Total unique reads* : nombre total de *reads* uniques (non dupliqués) ; *Paired in sequencing reads* : nombre de *reads* appairés ; *Total mapped reads* : pourcentage total de *reads* alignés sur le génome ; *Properly paired mapped reads* : pourcentage de paires de *reads* correctement alignées ; *Singleton mapped reads* : pourcentage de *reads* alignés seuls.

Sample	Genome version	Number of reads	Duplicates (%)	Total unique reads	Paired in sequencing reads	Total mapped reads (%)	Properly paired mapped reads (%)	Singleton mapped reads (%)
7 E	W82-A4-V1_mainChr	1 533 684	0,15	1 531 424	1 533 684	91,27	85,44	0,15
7 NE	W82-A4-V1_mainChr	1 569 636	0,14	1 567 485	1 569 636	98,81	95,05	0,13
8 E	W82-A4-V1_mainChr	1 410 722	0,13	1 408 928	1 410 722	97,35	91,68	0,12
8 NE	W82-A4-V1_mainChr	1 368 994	0,14	1 367 118	1 368 994	97,98	93,77	0,11
9 E	W82-A4-V1_mainChr	1 039 830	0,23	1 037 485	1 039 830	98,38	92,43	0,16
9 NE	W82-A4-V1_mainChr	1 584 592	0,21	1 581 343	1 584 592	93,96	89,33	0,16

Les fichiers BAM et BAI ont été visualisés grâce à IGV (*Integrative genomic viewer*, V2.6.3) afin de déterminer la couverture des séquences sur les 20 chromosomes du génome et sur les ADN mitochondrial et plastidial de *G. max*.

La distribution des *counts* obtenus après découpage a été analysée avec Excel en réalisant des graphiques représentant le nombre de *counts* en fonction de la position de ceux-ci sur le génome pour chaque chromosome, pour les ADN mitochondrial et plastidial.

Les comptages par zones de 1000 pb ont ensuite été analysés avec l'outil d'analyse différentielle DESeq2 V2.11.40.6 (Love *et al.*, 2014). Des *volcanoplots* ont été réalisés avec Excel afin de trier les zones de 1000 pb et de conserver celles ayant un *fold change* (FC) supérieur ou égal à 1,5 et une *p-value* inférieure à 0,05.

Résultats

- I. Effet d'une élicitation par PEP-13 sur le transcriptome et le protéome des AC-DC et des racines de soja
 - I-1 Analyses transcriptomiques
 - I-2 Analyses protéomiques
 - I-3 Comparaison des résultats obtenus entre l'analyse protéomique et l'analyse transcriptomique
- II. Analyse génomique de l'ADNex du RET du Soja
 - II-1 Analyse de l'alignement des séquences
 - II-2 Analyse différentielle de l'alignement des séquences d'ADNex suite à l'élicitation

Résultats

I. Effet d'une élicitation par PEP-13 sur le transcriptome et le protéome des AC-DC et des racines de soja

L'objectif de cette étude est d'observer et de comparer les réponses des AC-DC et des racines de soja à une élicitation avec PEP-13 (peptide issu à l'origine des oomycètes du genre *Phytophthora*, dont certaines espèces sont connues pour être pathogènes du soja) dont l'effet a été très peu décrit. Pour cela, un protocole de récolte des AC-DC et racines et notamment de leurs ARNs et protéines en quantité et qualité suffisantes a tout d'abord été mis en place (*cf.* Matériels et Méthodes II et III). Suite à cela, une analyse transcriptomique (I, partie 1) et une analyse protéomique (I, partie 2) ont été réalisées sur ces deux zones en condition élicitée avec PEP-13 (E) et en condition témoin non élicitée (NE). Pour chacune des zones, ces deux conditions sont comparées.

I-1 Analyses transcriptomiques

I-1-1 Identification des gènes différentiellement exprimés suite à l'élicitation

L'analyse bio-informatique réalisée sur les fichiers de séquençage ARN a permis tout d'abord d'obtenir une liste des gènes associés avec leur expression différentielle ou *fold change* (FC) et à une valeur statistique (*p-value*) obtenus grâce à un test exact basé sur une méthode bayésienne empirique avec une correction pour tests multiples. Ces deux valeurs (FC et *p-value*) sont représentées sous forme de graphiques de type *volcanoplots*. Pour ces *volcanoplots*, le seuil statistique choisi est une *p-value* inférieure ou égale à 0,05 et un FC supérieur ou égal à 1,5 dont les valeurs sont représentées respectivement en puissance logarithmique Log10 et Log2 (matérialisées par les lignes verticales et horizontales en pointillé sur les graphiques).

La Figure 33 est le graphique *volcanoplot* représentant les gènes exprimés dans les AC-DC en condition élicitée et témoin. Il met en évidence la présence d'un grand nombre de gènes différentiellement exprimés entre ces deux conditions (au-dessus du seuil statistique minimal choisi). De plus, il semblerait qu'il y ait plus de gènes avec une expression

différentielle plus importante en condition non élicitée qu'en condition élicitée (présence d'un plus grand nombre de points bleus que de points rouges). Enfin, le Log2 du FC maximal atteint pour la condition élicitée est d'environ 6, soit un *fold change* d'environ 64 entre la condition élicitée et non élicitée. Pour la condition non élicitée, le Log2 du FC maximal est d'environ 9,5 soit un *fold change* d'environ 726 entre les deux conditions.

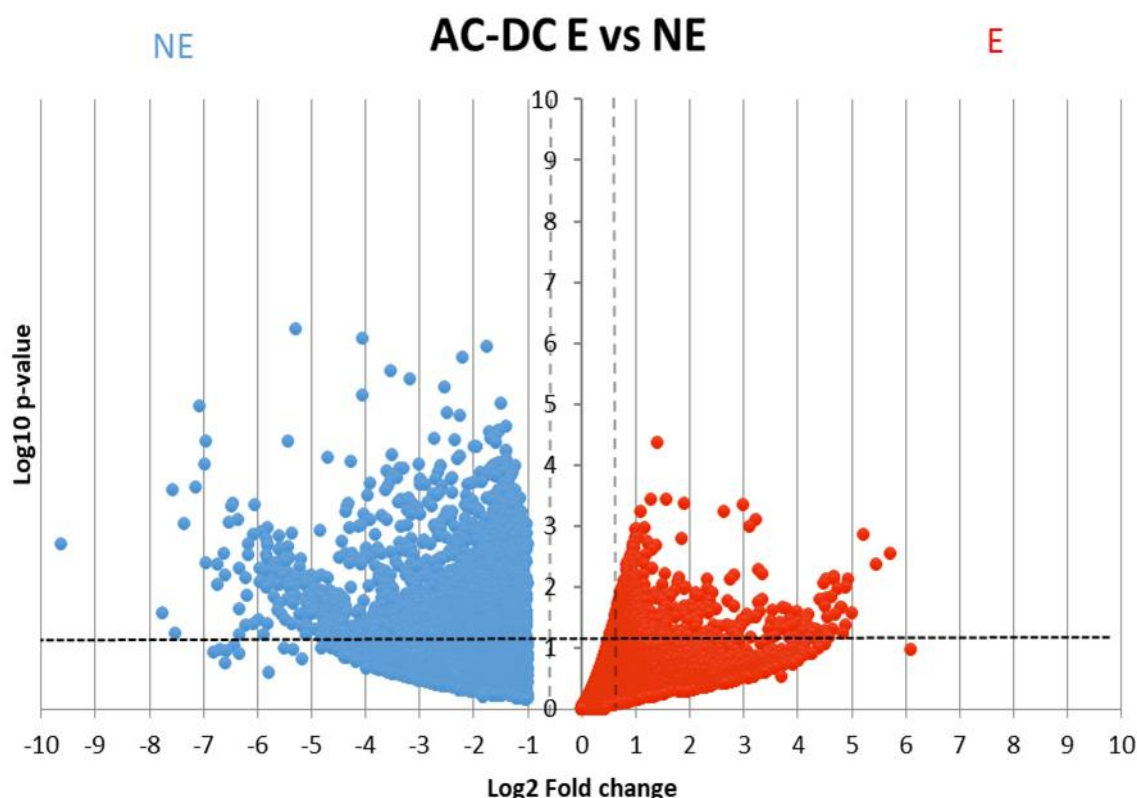


Figure 33: Analyse de l'expression différentielle des gènes des AC-DC (*root-Associated Cap-Derived Cells*) de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)

Sur ce graphique, chaque gène ayant été détecté est représenté par un point. Ces points sont placés grâce à la valeur absolue du Log10 de leur *p-value* et au Log2 de leur *fold change*. La partie gauche de ce graphique (bleue) correspond aux gènes ayant été détectés en plus grande quantité dans la condition non élicitée (NE) et la partie droite de ce graphique (rouge) correspond aux gènes ayant été détectés en plus grande quantité dans la condition élicitée (E). Dans cette étude, les gènes ayant une *p-value* inférieure ou égale à 0,05 (soit $|\text{Log}_{10}| = 1,301029996$, représenté par une ligne horizontale noire en pointillés) et un FC supérieur ou égal à 1,5 (soit $\text{Log}_2 = 0,5849625$, représenté par une ligne verticale grise en pointillés) ont été sélectionnés car considérés comme différentiellement exprimés selon le seuil statistique choisi. Graphique réalisé avec Excel.

La Figure 34 est le graphique *volcanoplot* représentant les gènes exprimés dans les racines en condition élicitée (E) et témoin (NE). Il met en évidence la présence de gènes différentiellement exprimés entre ces deux conditions. Pour ce tissu, il semblerait que la distribution des gènes entre les conditions élicitée (E) et non élicitée (NE) soit plutôt équilibrée. De plus, on remarque que le Log2 du FC maximal atteint en condition non élicité

est d'environ 6 (soit FC = 64) pour la condition non élicitée et d'environ 7 (soit FC = 128) pour la condition élicitée.

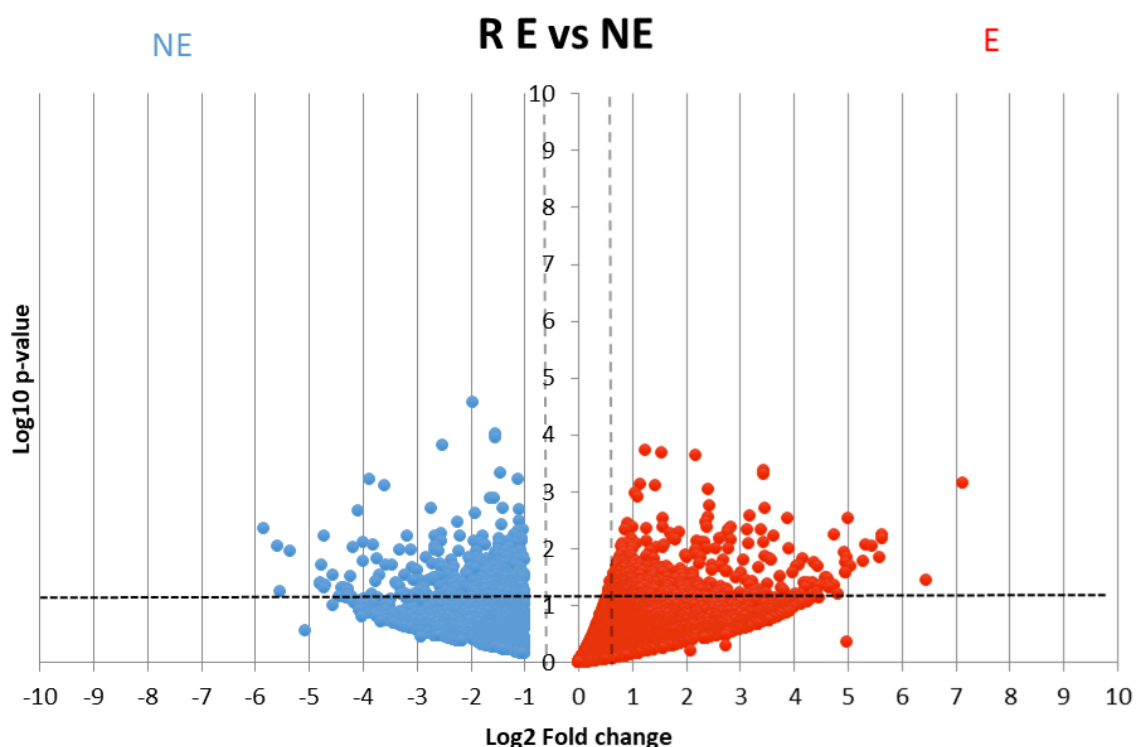


Figure 34 : Analyse de l'expression différentielle des gènes des racines de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)

Sur ce graphique, chaque gène ayant été détecté est représenté par un point. Ces points sont placés grâce à la valeur absolue du Log10 de leur *p-value* et au Log2 de leur *fold change*. La partie gauche de ce graphique (bleue) correspond aux gènes ayant été détectés en plus grande quantité dans la condition non élicitée (NE) et la partie droite de ce graphique (rouge) correspond aux gènes ayant été détectés en plus grande quantité dans la condition élicitée (E). Dans cette étude, les gènes ayant une *p-value* inférieure ou égale à 0,05 (soit $\text{Log}_{10}(= 1,301029996)$, représenté par une ligne horizontale noire en pointillés) et un FC supérieur ou égal à 1,5 (soit $\text{Log}_2 = 0,5849625$, représenté par une ligne verticale grise en pointillés) ont été sélectionnés car considérés comme différentiellement exprimés selon le seuil statistique choisi. Graphique réalisé avec Excel.

Si l'on compare les graphiques obtenus (Figures 33 et 34) pour les racines et pour les AC-DC, on observe tout d'abord que les deux graphiques sont différents, il y aurait donc une différence de réponse entre les racines et les AC-DC. De plus, il semblerait qu'il y ait moins de gènes différentiellement exprimés chez les racines par rapport aux AC-DC supposant donc une réponse à l'élicitation plus importante chez celles-ci comparées aux racines.

Pour la suite des analyses, seuls les gènes dont le FC est supérieur ou égal à 1,5 et la *p-value* inférieure ou égale à 0,05 ont été sélectionnés pour les deux zones. Ainsi, entre les conditions élicitée et non élicitée, une liste de 2295 gènes différentiellement exprimés chez les AC-DC et une liste de 553 gènes différentiellement exprimés chez les racines ont été

obtenues afin d'étudier plus en détail la réponse de ces deux zones à l'élicitation avec PEP-13.

I-1-2 Analyses des processus biologiques et voies métaboliques impliquées dans la réponse à PEP-13

Tout d'abord, les listes de gènes différentiellement exprimés en réponse à PEP-13 chez les AC-DC et chez les racines (FC supérieur ou égal à 1,5 et *p-value* inférieure à 0,05) ont été comparées grâce à un diagramme de Venn afin de savoir si ces gènes sont semblables pour ces deux zones (Figure 35).

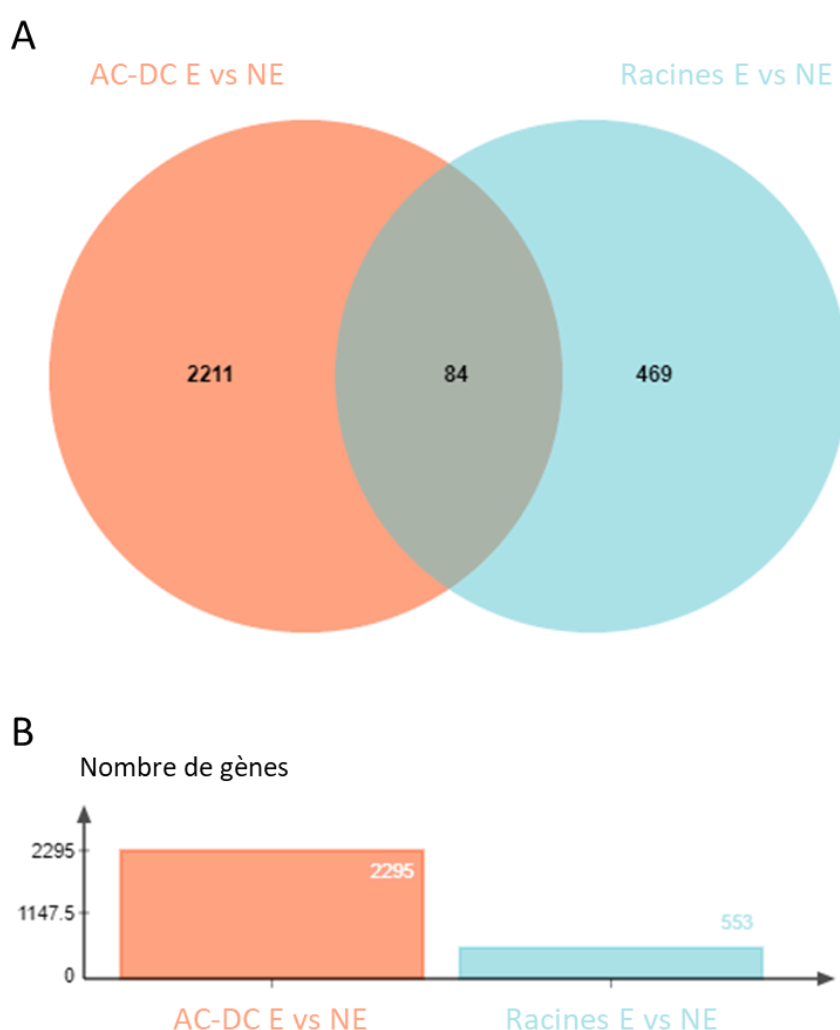


Figure 35 : Diagramme de Venn comparant les gènes différentiellement exprimés chez les racines et chez les AC-DC (*root-Associated Cap-Derived Cells*) de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)

Les listes des gènes différentiellement exprimés (FC supérieur ou égal à 1,5 et *p-value* inférieure à 0,05) sont représentées par un cercle de couleur orange chez les AC-DC et de couleur bleue chez les racines (A). L'intersection entre ces deux cercles correspond aux gènes retrouvés dans les deux listes. Ainsi, 84 gènes sont communs aux AC-DC et aux racines parmi respectivement les 2295 et 553 gènes différentiellement exprimés entre les conditions élicitée ou témoin trouvés dans chaque tissu (B). Graphique réalisé avec Jvenn.

La comparaison de ces deux listes permet donc d'observer, pour un FC supérieur ou égal à 1,5 et une *p-value* inférieure à 0,05, la présence de 84 gènes communs, de 469 gènes spécifiques à la réponse de la racine et 2211 gènes spécifiques à la réponse des AC-DC. Il y a donc bien une différence de réponse à PEP-13 entre ces deux structures racinaires. Afin d'observer plus en détail ces différences, une analyse des voies métaboliques impliquées dans la réponse à PEP-13 pour ces deux zones a ensuite été réalisée.

1-1-2-1 Analyse des voies métaboliques et des processus biologiques impliquant les gènes communs aux AC-DC et aux racines

Si l'on s'intéresse aux 84 gènes communs à la racine et aux AC-DC en réponse à PEP-13, on peut tout d'abord observer que 64 d'entre eux ont un FC similaire, c'est-à-dire positif dans les deux listes (31 gènes) ou négatif dans les deux listes (33 gènes). Afin d'observer plus en détail ces gènes et notamment leur fonction dans les processus biologiques, ces deux listes de gènes ont été analysées avec WEGO 2.0 (Figure 36). Cette analyse permet de mettre en évidence que les racines et les AC-DC vont réguler positivement ou négativement des gènes impliqués dans les processus biologiques représentés en rouge lorsqu'ils sont sur-exprimés dans la condition élicitée ou en bleu lorsqu'ils sont sous-exprimés en réponse à PEP-13.

On observe tout d'abord qu'un grand nombre de catégories de *Gene Ontology* (GO) (réponse aux stimuli, processus métaboliques et cellulaires, organisation et synthèse des composants cellulaires, développement, localisation, signalisation, régulation et croissance) possèdent à la fois des gènes négativement et positivement régulés chez les racines et les AC-DC en réponse à PEP-13. Il semble donc que pour ces voies, les AC-DC et les racines en réponse à PEP-13 régulent négativement certains de leurs acteurs et positivement d'autres acteurs.

Certaines voies sont représentées uniquement par des gènes sur ou sous-exprimés dans les deux zones. Comme par exemple la catégorie correspondant à la "régulation négative de processus biologiques" qui est uniquement représentée par des gènes négativement régulés par PEP-13 dans les deux zones racinaires étudiées, et la catégorie correspondant à la "régulation positive de processus biologiques" qui est représentée uniquement par des gènes positivement régulés chez les AC-DC et les racines en réponse à PEP-13. Il semble donc qu'un certain nombre de processus biologiques soient communément induits en réponse à

PEP-13. D'autres voies, comme la reproduction sont uniquement sous-exprimées en réponse à PEP-13. Les processus spécifiques aux organismes multicellulaires, sont eux uniquement sur-exprimés en réponse à PEP-13 dans les racines et les AC-DC.

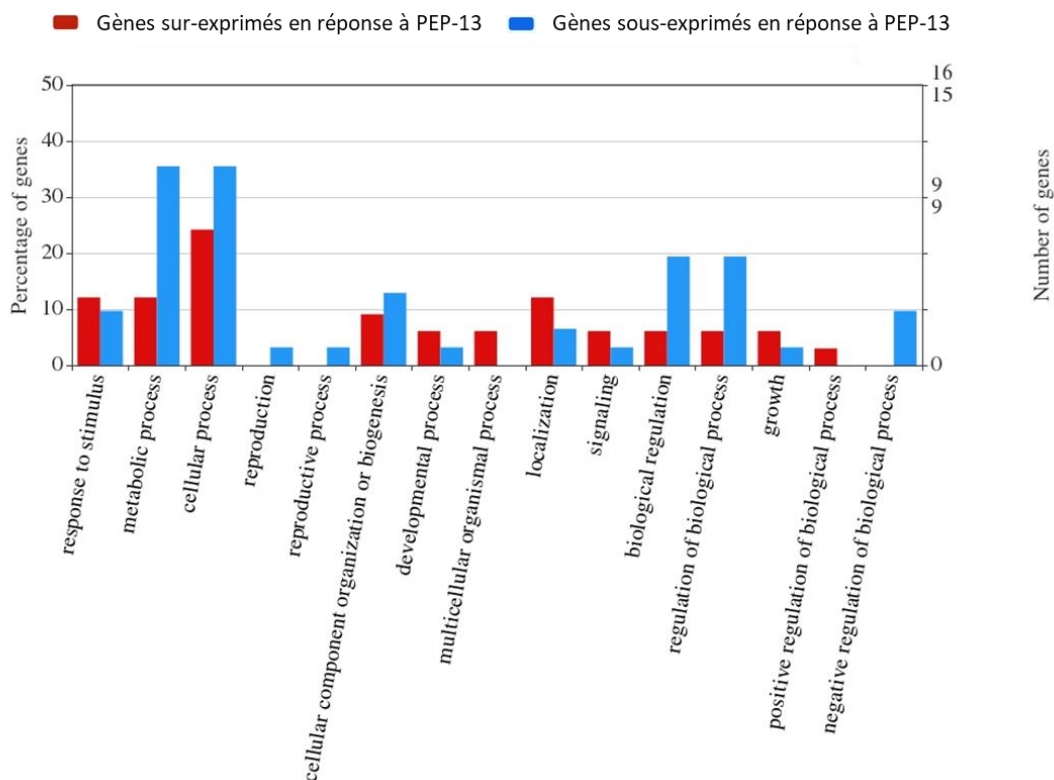


Figure 36 : Processus biologiques (Gene Ontology, GO) des 64 gènes communément régulés en réponse à PEP-13 chez les AC-DC (root-Associated Cap-Derived Cells) et les racines de soja
 Pour chaque gène, un certain nombre de Gene Ontology (GO) ont été assignés en fonction de leur annotation (axe des abscisses). Pour chaque GO, les gènes comptabilisés en nombre et pourcentage de gènes (axes des ordonnées). Enfin, pour un FC supérieur ou égal à 1,5 et une *p-value* inférieure à 0,05, la liste des gènes sous-exprimés en réponse à PEP-13 dans les AC-DC et dans les racines est représentée en bleu, et celle des gènes positivement régulés par PEP-13 dans les AC-DC et dans les racines est représentée en rouge. Graphique réalisé avec WEGO 2.0.

Enfin, parmi les 84 gènes communs aux AC-DC et aux racines, 20 sont régulés de façon différente entre les deux zones. En effet, en réponse à PEP-13, quatre d'entre eux sont régulés positivement chez les AC-DC et négativement chez les racines et seize d'entre eux sont négativement régulés chez les AC-DC et positivement régulés chez les racines. Ces deux listes de gènes ont été étudiées avec WEGO 2.0 afin d'observer les processus biologiques impliqués (Figure 37).

- Gènes négativement régulés dans les AC-DC et positivement régulés dans les racines par PEP-13
- Gènes positivement régulés dans les AC-DC et négativement régulés dans les racines par PEP-13

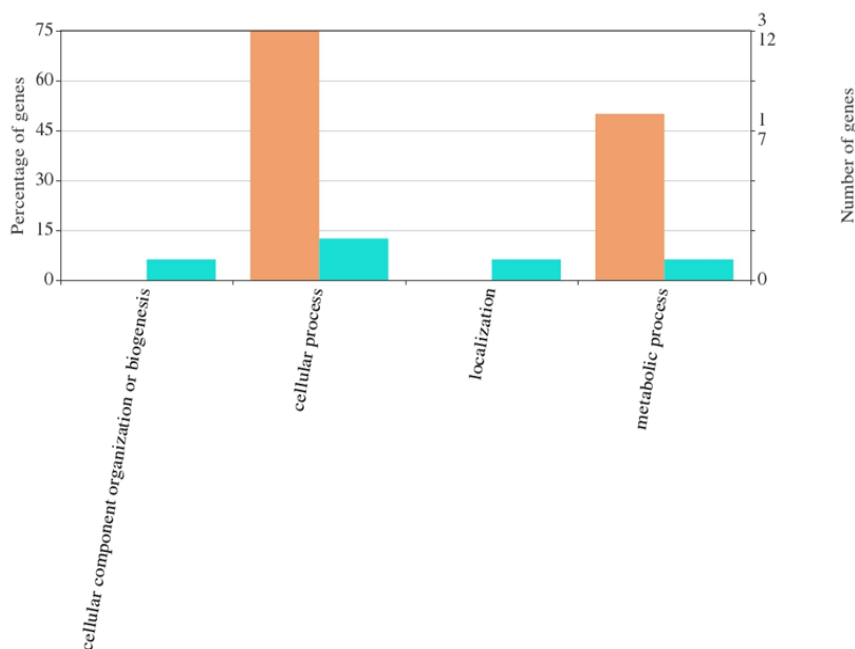


Figure 37 : Processus biologiques (*Gene Ontology, GO*) des 20 gènes régulés de façon différente en réponse à PEP-13 chez les AC-DC (*root-Associated Cap-Derived Cells*) et les racines de soja

Pour chaque gène, un certain nombre de *Gene Ontology* (GO) ont été assignés en fonction de leur annotation (axe des abscisses). Pour chaque GO, les gènes comptabilisés en nombre et pourcentage de gènes (axes des ordonnées). Les gènes positivement régulés chez les AC-DC et négativement régulés chez les racines en réponse à PEP-13 sont représentés en orange. Les gènes négativement régulés chez les AC-DC et positivement régulés chez les racines en réponse à PEP-13 sont représentés en bleu. Graphique réalisé avec WEGO 2.0.

Cette analyse permet tout d’abord de remarquer que les gènes négativement régulés chez les AC-DC et positivement régulés chez les racines sont impliqués dans des processus biologiques d’organisation et de synthèse des composants cellulaires, de processus cellulaires et métaboliques et de localisation. Les gènes positivement régulés par les AC-DC et négativement régulés par les racines sont eux impliqués dans les processus cellulaires et métaboliques. Ces différentes catégories seraient donc en partie régulées de façon inverse entre les deux zones en réponse à PEP-13.

I-1-2-2 Analyse des voies métaboliques et processus biologiques impliquant des gènes spécifiquement régulés chez les AC-DC ou chez la racine

Si l'on s'intéresse maintenant à la réponse individuelle de ces deux zones en réponse à PEP-13, on obtient deux listes de gènes avec un FC supérieur ou égal à 1,5 et une *p-value* inférieure à 0,05. La première liste contient 2 211 gènes différentiellement exprimés spécifiquement chez les AC-DC et la seconde 469 gènes différentiellement exprimés spécifiquement chez les racines en réponse à PEP-13. Parmi ces listes, on peut différencier les gènes régulés positivement ; soit 684 pour les AC-DC et 279 pour les racines, et les gènes négativement régulés ; soit 1 527 pour les AC-DC et 190 pour les racines.

A partir de ces listes, une nouvelle analyse WEGO 2.0 a été réalisée afin d'observer les processus biologiques spécifiques à la réponse à PEP-13 des AC-DC (Figure 38) et des racines (Figure 39).

Pour les AC-DC en réponse à PEP-13 (Figure 38), on observe tout d'abord les 684 gènes sur-exprimés en rouge, et les 1 527 gènes sous-exprimés en réponse en bleu. Parmi les catégories de processus biologiques régulées positivement ou négativement en réponse à PEP-13, il semble que les catégories de GO les plus importantes soient la localisation, les processus métaboliques, les processus cellulaires, la régulation, la réponse aux stimuli, et l'organisation et synthèse des composés cellulaires (protéines, membranes, *etc.*).

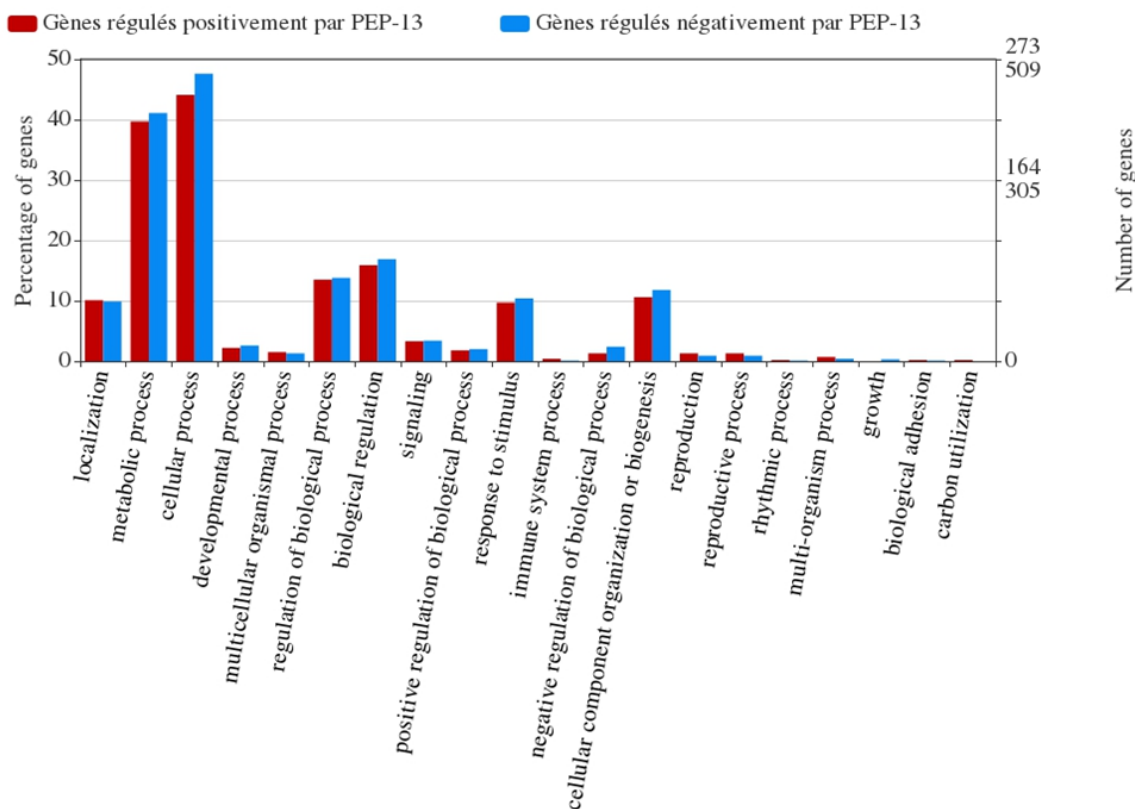


Figure 38 : Processus biologiques (*Gene Ontology*, GO) des 2 211 gènes spécifiquement régulés (684 gènes sur-exprimés et 1 527 gènes sous-exprimés) en réponse à PEP-13 chez les AC-DC (*root-Associated Cap-Derived Cells*) de soja

Pour chaque gène, un certain nombre de *Gene Ontology* (GO) ont été assignés en fonction de leur annotation (axe des abscisses). Pour chaque GO, les gènes comptabilisés en nombre et pourcentage de gènes (axes des ordonnées). Enfin, pour un FC supérieur ou égal à 1,5 et une *p-value* inférieure à 0,05, la liste des gènes sous-exprimés en réponse à PEP-13 dans les AC-DC est représentée en bleu, et celle des gènes positivement régulés par PEP-13 dans les AC-DC est représentée en rouge. Graphique réalisé avec WEGO 2.0.

Pour les racines en réponse à PEP-13 (Figure 39), on observe les 279 gènes sur-exprimés en condition élicitée en rouge, et les 190 gènes sous-exprimés en bleu. Pour ce tissu, les catégories de processus biologiques les plus représentées sont la réponse aux stimuli, le métabolisme et les processus cellulaires, la localisation, l'organisation et la synthèse de composés, et la régulation. De plus, on remarque que les catégories de processus biologiques uniquement régulées positivement sont la reproduction, les régulations négative et positive, le développement, et le processus spécifiques aux organismes multicellulaires.

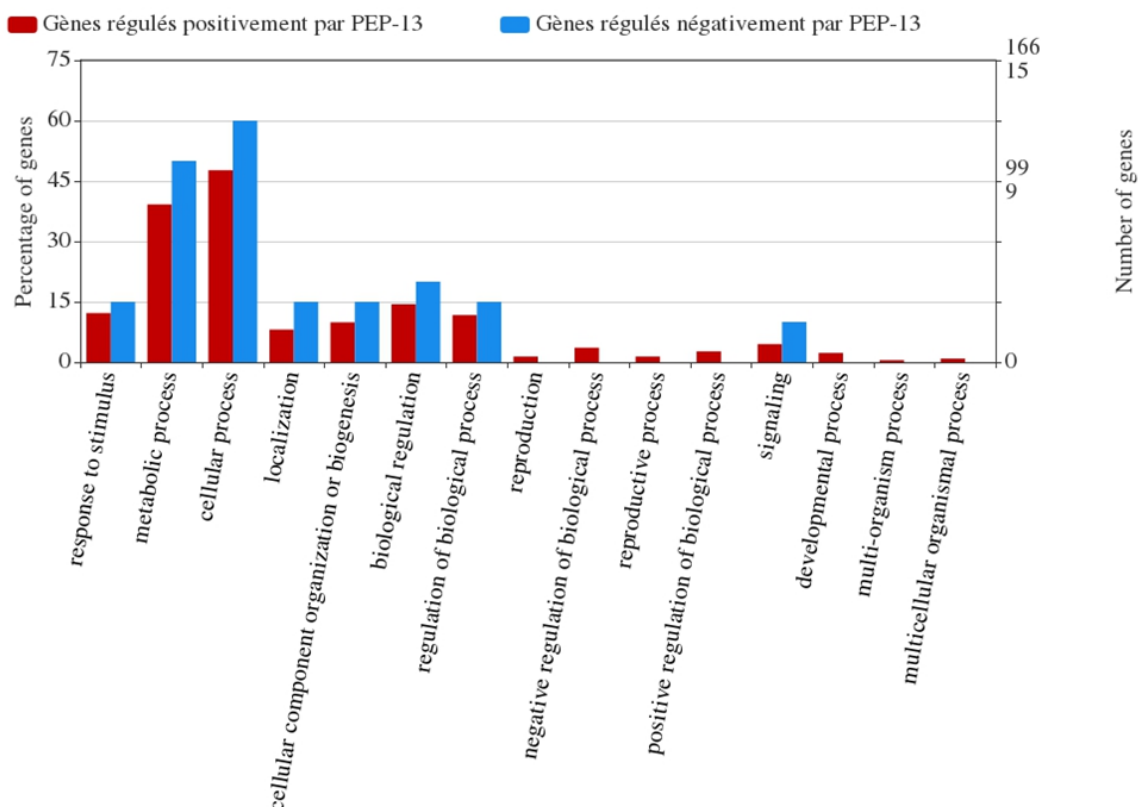


Figure 39 : Processus biologiques (*Gene Ontology*, GO) des 469 gènes (279 gènes sur-exprimés et 190 gènes sous-exprimés) spécifiquement régulés en réponse à PEP-13 chez les racines de soja

Pour chaque gène, un certain nombre de *Gene Ontology* (GO) ont été assignés en fonction de leur annotation (axe des abscisses). Pour chaque GO, les gènes comptabilisés en nombre et pourcentage de gènes (axes des ordonnées). Enfin, pour un FC supérieur ou égal à 1,5 et une *p-value* inférieure à 0,05, la liste des gènes sous-exprimés en réponse à PEP-13 dans les racines est représentée en bleu, et celle des gènes positivement régulés par PEP-13 dans les racines est représentée en rouge. Graphique réalisé avec WEGO 2.0.

Grâce à cette analyse, si l'on compare les processus biologiques principaux mis en jeu en réponse à PEP-13 chez les AC-DC et chez les racines (Figures 38 et 39), on constate qu'ils sont communs aux deux zones, bien que les gènes régulés soient tous différents entre les deux listes.

Afin d'étudier les processus biologiques impliqués dans la réponse de ces deux zones, des tests d'*overrepresentation* ont été réalisés grâce à l'interface PANTHER (*Protein Analysis THrough Evolutionary Relationships Database*). Pour cela, les listes de gènes différentiellement exprimés en réponse à PEP-13 ont été subdivisées en une liste des gènes régulés positivement et l'autre des gènes régulés négativement. Ces listes ont ensuite été traitées avec la liste de référence des gènes de *Glycine max* incluse dans l'interface grâce à un test de Fisher et le calcul de la FDR (*false discovery rate*, *p-value* corrigée). Pour les racines,

aucune voie ne semble statistiquement sur-représentée. La réponse à PEP-13 des racines n'est donc pas détectée par cet outil. Pour les AC-DC, la liste des gènes régulés positivement en réponse à PEP-13 ne permet pas non-plus d'observer une sur-représentation d'une voie particulière. Par contre, la liste des gènes régulés négativement chez les AC-DC en réponse à PEP-13 a permis d'observer la sur-représentation de certains processus biologiques (Tableau 6).

Tableau 6 : Processus biologiques représentés par les 1 527 gènes régulés négativement chez les AC-DC de soja en réponse à PEP-13

GO biological process complete : liste des GO sur-représentés dans la liste ; *Reflist* : nombre total de gènes de *Glycine max* identifiés par ce GO dans leur annotation ; *Input* : nombre de gènes possédant ce GO dans la liste issue de notre étude ; *Expected* : seuil statistique du nombre de gènes à partir duquel une voie est considérée comme statistiquement sur-représentée ; *Fold enrichment* : enrichissement par rapport au seuil statistique ; *FDR* : *False discovery rate*, correspond à la *p-value* corrigée.

GO biological process complete	REFLIST (55865)	Input (1565)	Expected	Fold Enrichment	FDR
ribonucleoprotein complex biogenesis (GO:0022613)	715	42	20,03	2,1	6,93E-03
ribosome biogenesis (GO:0042254)	575	33	16,11	2,05	3,52E-02
amide biosynthetic process (GO:0043604)	1116	59	31,26	1,89	4,21E-03
translation (GO:0006412)	1001	51	28,04	1,82	2,06E-02
peptide biosynthetic process (GO:0043043)	1015	51	28,43	1,79	2,28E-02
cellular amide metabolic process (GO:0043603)	1359	66	38,07	1,73	9,76E-03
organonitrogen compound biosynthetic process (GO:1901566)	2300	99	64,43	1,54	1,20E-02
gene expression (GO:0010467)	2765	119	77,46	1,54	4,62E-03
cellular macromolecule biosynthetic process (GO:0034645)	2376	101	66,56	1,52	1,49E-02
macromolecule biosynthetic process (GO:0009059)	2457	102	68,83	1,48	2,47E-02
cellular nitrogen compound metabolic process (GO:0034641)	5525	221	154,78	1,43	1,83E-04
organic substance biosynthetic process (GO:1901576)	4250	167	119,06	1,4	6,35E-03
cellular biosynthetic process (GO:0044249)	4167	162	116,73	1,39	1,09E-02
organic cyclic compound metabolic process (GO:1901360)	4777	184	133,82	1,37	6,91E-03
nucleobase-containing compound metabolic process (GO:0006139)	4156	158	116,43	1,36	2,87E-02
heterocycle metabolic process (GO:0046483)	4501	171	126,09	1,36	1,63E-02
cellular aromatic compound metabolic process (GO:0006725)	4666	177	130,71	1,35	1,50E-02
biosynthetic process (GO:0009058)	4544	172	127,3	1,35	2,00E-02
primary metabolic process (GO:0044238)	14361	500	402,31	1,24	1,25E-04
nitrogen compound metabolic process (GO:0006807)	12151	423	340,4	1,24	9,49E-04
organic substance metabolic process (GO:0071704)	15239	528	426,9	1,24	1,42E-04
macromolecule metabolic process (GO:0043170)	11159	380	312,61	1,22	1,11E-02
cellular metabolic process (GO:0044237)	14493	485	406,01	1,19	4,39E-03
cellular process (GO:0009987)	18991	634	532,01	1,19	1,87E-04
metabolic process (GO:0008152)	18198	597	509,8	1,17	3,54E-03
biological_process (GO:0008150)	29850	925	836,22	1,11	4,62E-03
Undclassified (UNCLASSIFIED)	26015	640	728,78	0,88	5,28E-03

Parmi ces processus biologiques, on retrouve notamment des voies de synthèse (GO:0043043 ; GO:0043604 ; GO:1901566 ; GO:0034645 ; GO:0009059 ; GO:1901576 ; GO:0044249 ; GO:0009058) et des processus métaboliques (GO:0008152 ; GO:0044237 ; GO:0043170 ; GO:0071704 ; etc.) qui semblent régulées négativement en réponse à PEP-13. On constate également que les voies les plus sur-représentées sont (i) la synthèse des protéines ribosomales ; (ii) la synthèse des ribosomes ; (iii) la synthèse d'amides et (iv) la transcription. Ainsi, une partie du métabolisme et des voies de synthèse, et notamment celle de la synthèse des protéines grâce aux ribosomes, semblent donc être négativement régulées chez les AC-DC en réponse à PEP-13.

Afin de compléter cette analyse, les listes de gènes régulés positivement et négativement en réponse à PEP-13 chez les AC-DC et les racines (FC supérieur ou égal à 1,5 et *p-value* inférieure à 0,05) ont été analysées avec KEGG (*Kyoto Encyclopedia of Genes and Genomes*) grâce à l'outil BlastKOALA (*BLAST-based KO annotation and KEGG mapping*). Pour les différentes listes de gènes intégrées dans cette interface, seulement 36,98 à 45,51% des gènes ont pu être associés à une annotation KEGG. Ce faible taux d'annotation est lié au manque d'annotations au sein des bases de données du génome et des protéines de *Glycine max*. L'analyse a donc été réalisée sur 502 gènes pour les AC-DC élicitées, 968 gènes pour les AC-DC non élicitées, 207 gènes pour les racines élicitées et 125 gènes pour les racines non élicitées. Grâce à l'annotation KEGG des différents gènes utilisés, ceux-ci ont été classés en fonction des voies métaboliques impliquées. Pour chaque tissu un tableau présentant ces trente voies les plus représentées en condition élicitée et en condition non élicitée a été réalisé.

Pour les AC-DC (Tableau 7), on observe un grand nombre de voies représentées aussi bien en condition élicitée et qu'en condition non élicitée (biosynthèse des métabolites secondaires, ribosomes, métabolisme du carbone, interactions plante-pathogène, *etc.*). Ces voies seraient donc régulées à la fois positivement et négativement par PEP-13. De plus, on remarque (en surligné vert) que certaines voies n'apparaissent que pour une seule condition (élicitée ou témoin). Par exemple, la synthèse des ribosomes, des terpénoïdes, des stéroïdes ou encore la senescence sont retrouvées uniquement dans la liste des trente voies les plus représentées chez les AC-DC non élicitées. Ces voies semblent donc régulées négativement par PEP-13. Enfin, les voies de phosphorylation oxydative, de réparation de l'ADN (*nucleotide excision repair* et *base excision repair*), de réplication de l'ADN ou encore la signalisation du phosphatidylinositol sont retrouvées uniquement dans la liste des trente voies les plus représentées chez les AC-DC élicitées. Ces voies semblent donc être régulées positivement par PEP-13.

Tableau 7 : Mise en évidence des trente voies métaboliques les plus représentées chez les AC-DC de soja en condition élicitée (E) ou non (NE) grâce à l'annotation KEGG et l'outil BlastKOALA

Les voies spécifiques à une condition (E ou NE) sont représentées en vert. Ce tableau a été réalisé à partir des 502 gènes sur-exprimés en condition élicitée et des 968 gènes sous-exprimés en réponse à l'élicitation et possédant une annotation KEGG. Tableau réalisé grâce à Excel à partir de l'outil BlastKOALA.

AC-DC NE	Nombre de gènes	AC-DC E	Nombre de gènes
01100 Metabolic pathways	129	01100 Metabolic pathways	76
01110 Biosynthesis of secondary metabolites	76	01110 Biosynthesis of secondary metabolites	35
03010 Ribosome	32	00230 Nucleotide metabolism	9
01200 Carbon metabolism	19	01200 Carbon metabolism	8
03040 Spliceosome	18	00190 Oxidative phosphorylation	7
01230 Biosynthesis of amino acids	16	03013 RNA transport	7
03015 mRNA surveillance pathway	12	00500 Starch and sucrose metabolism	6
04075 Plant hormone signal transduction	12	00520 Amino sugar and nucleotide sugar metabolism	6
03013 RNA transport	11	03015 mRNA surveillance pathway	6
03008 Ribosome biogenesis in eukaryotes	11	04141 Protein processing in endoplasmic reticulum	6
04152 AMPK signaling pathway	11	04144 Endocytosis	6
00230 Nucleotide metabolism	9	01230 Biosynthesis of amino acids	5
00010 Glycolysis / Gluconeogenesis	8	00270 Cysteine and methionine metabolism	5
00500 Starch and sucrose metabolism	8	03040 Spliceosome	5
00940 Phenylpropanoid biosynthesis	8	03420 Nucleotide excision repair	5
04141 Protein processing in endoplasmic reticulum	8	03440 Homologous recombination	5
04016 MAPK signaling pathway - plant	8	04152 AMPK signaling pathway	5
04146 Peroxisome	8	04075 Plant hormone signal transduction	5
00520 Amino sugar and nucleotide sugar metabolism	7	04530 Tight junction	5
00270 Cysteine and methionine metabolism	7	00010 Glycolysis / Gluconeogenesis	4
00900 Terpenoid backbone biosynthesis	7	00562 Inositol phosphate metabolism	4
03018 RNA degradation	7	00330 Arginine and proline metabolism	4
04626 Plant-pathogen interaction	7	00940 Phenylpropanoid biosynthesis	4
04144 Endocytosis	7	03010 Ribosome	4
04530 Tight junction	6	03030 DNA replication	4
00100 Steroid biosynthesis	6	03410 Base excision repair	4
00330 Arginine and proline metabolism	6	04626 Plant-pathogen interaction	4
04218 Cellular senescence	6	04016 MAPK signaling pathway - plant	4
00030 Pentose phosphate pathway	5	04070 Phosphatidylinositol signaling system	4
00040 Pentose and glucuronate interconversions	5	04071 Sphingolipid signaling pathway	4

Pour les racines (Tableau 8), dans la liste des trente voies les plus représentées, on observe des voies retrouvées aussi bien en condition élicitée qu'en condition non élicitée (biosynthèse de métabolites secondaires, métabolisme du carbone, transduction des signaux hormonaux, métabolisme du glutathion, etc.). Ces voies semblent donc être à la fois régulées positivement et négativement en réponse à PEP-13. On observe ensuite des voies spécifiques aux racines non élicitées, comme l'endocytose, le métabolisme de l'azote, la dégradation des acides gras, le métabolisme des xénobiotiques par le cytochrome P450, la signalisation des MAPK, ou encore le métabolisme du pyruvate. Ces voies semblent donc régulées négativement en réponse à PEP-13. Enfin, chez les racines en condition élicitée, on retrouve spécifiquement par exemple la synthèse des protéines dans le réticulum endoplasmique, le métabolisme des acides aminés, la voie des pentoses phosphates, la synthèse de flavonoïdes, la réparation de l'ADN ou encore le métabolisme de l'amidon et du saccharose. Ces voies semblent donc être régulées positivement chez les racines en réponse à PEP-13.

Tableau 8 : Mise en évidence des trente voies métaboliques les plus représentées chez les racines de soja en condition élicitée (E) ou non (NE) grâce à l'annotation KEGG et l'outil BlastKOALA

Les voies spécifiques à une condition (E ou NE) sont représentées en vert. Ce tableau a été réalisé à partir des 207 gènes sur-exprimés en condition élicitée et des 125 gènes sous-exprimés en réponse à l'élicitation et possédant une annotation KEGG. Tableau réalisé grâce à Excel à partir de l'outil BlastKOALA.

Racines NE	Nombre de gènes	Racines E	Nombre de gènes
01100 Metabolic pathways	25	01100 Metabolic pathways	21
01110 Biosynthesis of secondary metabolites	11	01110 Biosynthesis of secondary metabolites	13
01200 Carbon metabolism	4	03010 Ribosome	5
04075 Plant hormone signal transduction	4	01200 Carbon metabolism	4
04144 Endocytosis	3	04141 Protein processing in endoplasmic reticulum	4
01230 Biosynthesis of amino acids	2	00520 Amino sugar and nucleotide sugar metabolism	3
00010 Glycolysis / Gluconeogenesis	2	00270 Cysteine and methionine metabolism	3
00910 Nitrogen metabolism	2	00480 Glutathione metabolism	3
00071 Fatty acid degradation	2	00970 Aminoacyl-tRNA biosynthesis	3
00561 Glycerolipid metabolism	2	04075 Plant hormone signal transduction	3
00564 Glycerophospholipid metabolism	2	04712 Circadian rhythm - plant	3
00250 Alanine, aspartate and glutamate metabolism	2	01230 Biosynthesis of amino acids	2
00480 Glutathione metabolism	2	00030 Pentose phosphate pathway	2
00790 Folate biosynthesis	2	00052 Galactose metabolism	2
00940 Phenylpropanoid biosynthesis	2	00562 Inositol phosphate metabolism	2
00980 Metabolism of xenobiotics by cytochrome P450	2	00260 Glycine, serine and threonine metabolism	2
00982 Drug metabolism - cytochrome P450	2	00940 Phenylpropanoid biosynthesis	2
03040 Spliceosome	2	00945 Stilbenoid, diarylheptanoid and gingerol biosynthesis	2
03015 mRNA surveillance pathway	2	00941 Flavonoid biosynthesis	2
04016 MAPK signaling pathway - plant	2	03040 Spliceosome	2
04146 Peroxisome	2	03013 RNA transport	2
04216 Ferroptosis	2	03018 RNA degradation	2
04217 Necroptosis	2	03420 Nucleotide excision repair	2
01212 Fatty acid metabolism	1	03430 Mismatch repair	2
01220 Degradation of aromatic compounds	1	04070 Phosphatidylinositol signaling system	2
00020 Citrate cycle	1	04150 mTOR signaling pathway	2
00040 Pentose and glucuronate interconversions	1	00500 Starch and sucrose metabolism	1
00052 Galactose metabolism	1	00190 Oxidative phosphorylation	1
00520 Amino sugar and nucleotide sugar metabolism	1	00564 Glycerophospholipid metabolism	1
00620 Pyruvate metabolism	1	00062 Fatty acid elongation	1

Enfin, si l'on compare les résultats obtenus pour les AC-DC et les racines (Tableaux 7 et 8), on observe tout d'abord que certaines voies sont communes aux AC-DC et aux racines, comme par exemple la synthèse de métabolites secondaires, le métabolisme du carbone ou encore la synthèse de phénylpropanoïdes. Les deux zones, racines et AC-DC, semblent donc modifier la régulation de ces voies en réponse à PEP-13. De plus, on observe la présence d'un plus grand nombre de voies spécifiques à une condition (voies métaboliques surlignées en vert) chez les racines (Tableau 8) que chez les AC-DC (Tableau 7) parmi les listes des trente voies les plus représentées.

Enfin, certaines voies semblent spécifiquement apparaître pour un tissu. Par exemple, la voie de signalisation AMPK, le métabolisme des nucléotides, la voie des interactions plante-pathogènes, le métabolisme de l'arginine et de la proline, et la mise en place de jonctions serrées ne sont retrouvées que dans la liste des trente voies les plus représentées chez les AC-DC en condition élicitée ou témoin. Chez les racines en condition élicitée ou témoin, on

retrouve les voies du métabolisme des glycérophospholipides, du glutathion et du galactose qui semblent spécifiques à ce tissu. On observe donc à nouveau une différence de réponse à PEP-13 entre ces deux zones.

Afin d'observer plus en détail la réponse de ces deux zones à l'élicitation avec PEP-13, un dernier niveau d'observation a été réalisé : l'analyse détaillée des gènes régulés par PEP-13.

I-1-3 Analyse détaillée des gènes différentiellement exprimés en réponse à PEP-13

Pour cette analyse, les listes de gènes différentiellement exprimés associés au Log2 de leur *fold change* ont été utilisées afin d'observer plus en détail les gènes impliqués dans la réponse des AC-DC et des racines à l'élicitation avec PEP-13. Pour cela, un schéma de visualisation globale des processus biologiques a tout d'abord été réalisé pour chacune des deux zones grâce au logiciel Mapman. Celui-ci a permis de classer les gènes en fonction des processus biologiques dans lesquels ils interviennent puis d'obtenir un tableau présentant chacun d'eux.

I-1-3-1 Chez les AC-DC

Pour cette analyse, seulement 1112 parmi les 2295 gènes différentiellement exprimés identifiés chez les AC-DC (cf. partie I-1-1) ont pu être annotés par Mapman, soit 48,45% des gènes. Ce faible taux d'annotation est lié au manque d'annotations au sein des bases de données du génome et des protéines de *Glycine max*.

Sur le schéma *overview* (Figure 40) réalisé avec Mapman, 1084 gènes différentiellement exprimés chez les AC-DC en réponse à PEP-13 peuvent être observés. Ces gènes sont représentés sous forme de carrés dont la couleur correspond au Log2 de leur *fold change* (FC) avec une couleur d'autant plus foncée que le FC est grand. Ainsi, les gènes représentés par un carré de couleur dégradée du rouge au rose sont des gènes positivement régulés en réponse à l'élicitation, tandis que les gènes représentés par un carré en dégradé de bleu sont des gènes négativement régulés. Sur cette représentation, on observe majoritairement des gènes négativement régulés par PEP-13 (759 gènes contre 325 positivement régulés). La proportion entre les gènes négativement et positivement régulés semble changer en fonction de la catégorie. En effet, pour la catégorie *Redox regulation*, on observe uniquement des gènes négativement régulés par l'élicitation avec PEP-13. En revanche, pour

la catégorie *signal regulation*, on retrouve plus de gènes positivement régulés que de gènes négativement régulés.

L'élicitation avec PEP-13 semble donc induire la diminution de l'expression d'un grand nombre de gènes, et donc l'inactivation de voies (comme par exemple celle de la régulation redox), au profit de l'augmentation de l'expression d'autres gènes.

Les gènes sont présents dans toutes les catégories de *bins* (cf Matériel et méthodes II-2), et en majorité dans les catégories protéines, ARN et dans la catégorie *unknown/unclassified* qui correspond aux gènes dont la fonction n'est pas encore connue (ou non recensée pour des *gene ontologies*). De plus, certaines catégories semblent peu représentées, comme par exemple *metal handling* qui serait donc peu impactée par l'élicitation avec PEP-13. Enfin, pour la catégorie *stress* on observe une grande majorité de gènes dont l'expression est négativement régulée, et un seul positivement régulé.

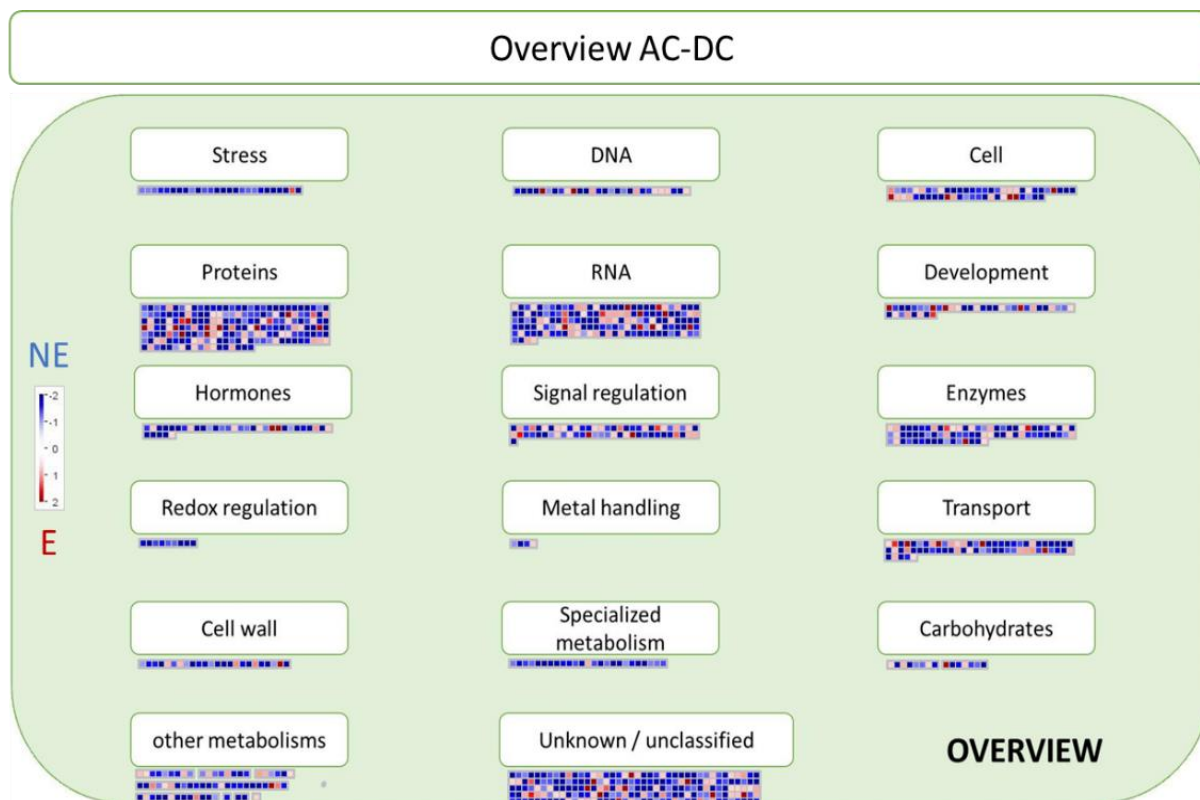


Figure 40 : Représentation générale (*overview*) des 1084 gènes différemment exprimés chez les AC-DC (*root-Associated Cap-Derived Cells*) de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)

Cette représentation permet d'observer les gènes régulés négativement ou positivement chez les AC-DC en réponse à PEP-13 en fonction des processus biologiques dans lesquels ils interviennent. Chaque gène est représenté par un carré dont l'intensité de la couleur dépend du Log_2 de son *fold change* (FC). Ainsi, les gènes sur-exprimés en réponse à PEP-13 sont représentés en dégradé de rouge, et les gènes sous-exprimés en réponse à PEP-13 sont représentés en dégradé de bleu. Graphique réalisé avec Mapman.

Pour la suite de cette étude, seuls les 325 gènes sur-exprimés en réponse à PEP-13 ont été observés (les gènes sous-exprimés seront présentés en annexe III, partie I-1). Pour cela, le détail de ces gènes sur-exprimés est visualisé sous forme de tableau classé par catégories de *bins* (Tableau 9), indiquant l'identifiant du transcrit, le nom de celui-ci ainsi que son FC (exprimé sous forme de logarithme binaire). Les détails concernant les gènes les plus exprimés seront exposés dans la discussion de ce manuscrit.

Tableau 9 : Détail des 325 gènes sur-exprimés chez les AC-DC de soja en réponse à PEP-13 classés en fonction de leur catégorie de bins

Une couleur est attribuée au log2 du FC en fonction de sa valeur, plus celle-ci est élevée plus celui-ci apparaîtra en rouge foncé.

Stress		
ID	Transcript name	Log(FC)
glyma10g28020.1	Germin-like protein	1,3741108
Cell wall		
ID	Transcript name	Log(FC)
glyma08g42270.1	NAD(P)-bd_dom domain-containing protein	0,85022026
glyma03g37550.1	nearly identical (1688)AT1G02730 1,4-beta-D-xylan synthase/ cellulose synthase	0,8394282
glyma18g00630.2	Xyloglucan endotransglucosylase/hydrolase (EC 2.4.1.207)	1,0907273
glyma08g15650.1	Pectinesterase (EC 3.1.1.11)	2,3479564
glyma19g41960.1	Pectinesterase (EC 3.1.1.11)	1,1299162
Carbohydrates		
ID	Transcript name	Log(FC)
glyma14g03300.1	nearly identical (1526)AT1G04920 sucrose-phosphate synthase/ transferase, transferring glycosyl groups	2,8368278
glyma17g04160.1	highly similar to (919)AT3G06500 putative beta-fructofuranosidase / invertase / saccharase/ beta-	0,74499476
glyma12g36280.1	Glyco_transf_20 domain-containing protein	0,657153
glyma18g04360.1	Phos_pyr_kin domain-containing protein	0,8378309
glyma18g44820.1	Alpha-galactosidase (EC 3.2.1.22) (Melibiase)	0,64027613
Hormones		
ID	Transcript name	Log(FC)
glyma08g47240.1	TF-B3 domain-containing protein	0,717696
glyma17g01170.1	Cytochrome b561 and DOMON domain-containing protein	0,785281
glyma18g51090.1	highly similar to (897)AT3G30300 similar to EDA30 (embryo sac development arrest 30)	0,7205277
glyma08g05370.1	nearly identical (1321)AT2G01830 CRE1, WOL1, AHK4, WOL WOL (WOODEN LEG)	0,67555124
glyma13g21120.1	Fe2OG dioxygenase domain-containing protein	3,1195898
glyma13g28970.1	Fe2OG dioxygenase domain-containing protein	2,4914544
glyma15g08370.1	AP2-EREBP transcription factor	0,73741853
glyma20g33800.1	AP2/ERF domain-containing protein	1,0500289
glyma09g38930.1	moderately similar to (246)AT1G19640 Jasmonic acid caroxyl methyltransferase; jasmonate O-	0,6303031
Signal regulation		
ID	Transcript name	Log(FC)
glyma06g47210.1	SWIM-type domain-containing protein	0,8949516
glyma13g01810.2	Photolyase/cryptochrome alpha/beta domain-containing protein	1,243055
glyma14g04820.1	SWIM-type domain-containing protein	0,916933
glyma01g03130.1	moderately similar to (400)AT1G12970 leucine-rich repeat family protein	0,7168035
glyma04g41860.1	Protein kinase domain-containing protein	0,7109476
glyma08g01640.1	Protein kinase domain-containing protein	0,62854755
glyma17g34380.1	Protein kinase domain-containing protein	0,63092816
glyma04g07080.1	Receptor-like serine/threonine-protein kinase (EC 2.7.11.1)	0,69830024
glyma08g46960.1	Receptor-like serine/threonine-protein kinase (EC 2.7.11.1)	0,7974251
glyma09g15200.1	Protein kinase domain-containing protein	1,2045686
glyma02g04150.1	Protein kinase domain-containing protein	0,9745547
glyma09g15200.1	Protein kinase domain-containing protein	1,2045686
glyma02g41140.1	moderately similar to (252)AT5G37660 receptor-like protein kinase-related	0,75815135
glyma04g40460.1	Endo/exonuclease/phosphatase domain-containing protein	1,4334966
glyma06g04950.1	TPR_REGION domain-containing protein	0,927638
glyma06g14300.1	EF-hand domain-containing protein	0,8837467
glyma07g09070.1	highly similar to (632)AT3G61050 NTMC2TYPE4, NTMC2T4 NTMC2T4/NTMC2TYPE4; lipid binding	0,71300894
glyma18g45350.1	weakly similar to (102)AT1G24620 polcalcin, putative / calcium-binding pollen allergen, putative	1,6389558
glyma19g40050.1	HD domain-containing protein	0,7774576
glyma04g01540.1	nearly identical (2833)AT1G49340 At phosphatidylinositol 4-kinase alpha; inositol or phosphatidylinositol	0,7319004
glyma07g07980.1	PRONE domain-containing protein	2,366159
glyma07g09330.1	PRONE domain-containing protein	5,0026164
glyma09g40480.1	SEC7 domain-containing protein	0,71750164
glyma15g01660.1	RanBD1 domain-containing protein	0,720202
glyma18g52450.1	moderately similar to (327)AT5G59840 Ras-related GTP-binding family protein	0,65347654

Proteins		
ID	Transcript name	Log(FC)
glyma06g00420.1	moderately similar to (320)AT2G43030 ribosomal protein L3 family protein	0,9446181
glyma13g39270.2	Ribosomal protein (Fragment)	0,6505946
glyma03g00470.1	Eukaryotic translation initiation factor 3 subunit B (eIF3b) (eIF-3-eta) (eIF3 p110)	0,6288659
glyma15g03020.1	hs to (768)AT1G54270 eukaryotic translation initiation factor 4A-2; helicase ATP-dependent	0,6474985
glyma18g02820.1	S1 motif domain-containing protein	0,89443684
glyma19g00800.1	Elongation factor Ts, mitochondrial (EF-Ts) (EF-TsMt)	0,8694393
glyma10g31310.1	Protein pelota homolog (EC 3.1.-.-)	0,66028476
glyma06g14420.1	SAM domain-containing protein	2,3397486
glyma08g21380.1	hs to (601)AT1G51980 mitochondrial processing peptidase alpha subunit, putative	0,9589164
glyma11g10310.1	moderately similar to (376)AT2G47450 CHAOS; chromatin binding	1,3034582
glyma11g08390.1	Vacuolar protein sorting-associated protein 35	0,9942043
glyma17g37870.1	Vacuolar protein sorting-associated protein 35	0,8191126
glyma08g01610.1	C2 NT-type domain-containing protein	0,6639859
glyma15g06170.1	moderately similar to (375)AT5G51220 ubiquinol-cytochrome C chaperone family protein	1,5054123
glyma02g02990.1	Serine/threonine-protein phosphatase (EC 3.1.3.16)	0,7754455
glyma03g39690.1	highly similar to (711)AT1G09020 Sucrose NonFermenting 4	0,7266649
glyma05g08640.1	Protein kinase domain-containing protein	0,6088411
glyma06g19500.1	Protein kinase domain-containing protein	4,139867
glyma07g36050.1	PPM-type phosphatase domain-containing protein	1,0510561
glyma09g03980.1	highly similar to (761)AT3G06620 protein kinase family protein	0,67676675
glyma09g38510.1	PPM-type phosphatase domain-containing protein	0,8735491
glyma11g15190.1	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase (EC 4.2.1.134)	0,8989453
glyma12g06790.1	PPM-type phosphatase domain-containing protein	0,7198296
glyma12g35510.1	Protein kinase domain-containing protein	1,9116721
glyma13g42800.1	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B	3,3047202
glyma17g05480.1	Protein kinase domain-containing protein	1,4609238
glyma06g01480.1	Protein kinase domain-containing protein	0,87011445
glyma07g05230.1	Protein kinase domain-containing protein	0,5968514
glyma11g04200.1	Protein kinase domain-containing protein	2,1040497
glyma14g06230.1	Protein kinase domain-containing protein	2,1331887
glyma17g36910.1	highly similar to (684)AT2G16250 leucine-rich repeat transmembrane protein kinase, putative	4,197906
glyma19g40820.1	Serine/threonine protein kinase	1,1682988
glyma10g28690.1	AMP_N domain-containing protein	1,9129878
glyma17g36690.1	EGF-like domain-containing protein	0,7469988
glyma18g06030.1	nearly identical (1347)AT1G06900 metalloendopeptidase	0,9781208
glyma04g02440.1	highly similar to (801)AT1G20160 ATSBT5.2 ATSBT5.2; subtilase	3,0115483
glyma11g05410.1	highly similar to (890)AT5G67360 ARA12 ARA12; subtilase	1,2909682
glyma07g00470.1	highly similar to (682)AT1G78880 balbiani ring 1-related / BR1-related	0,611502
glyma07g02360.1	highly similar to (646)AT1G78880 balbiani ring 1-related / BR1-related	0,6912451
glyma13g29660.1	MPN domain-containing protein	0,94100213
glyma03g38180.1	RING-type domain-containing protein	0,6638924
glyma06g15090.1	nearly identical (1752)AT2G22010 zinc finger (C3HC4-type RING finger) family protein	0,9853519
glyma10g30510.1	RING-type domain-containing protein	0,59991413
glyma11g01330.1	RING-type domain-containing protein	0,801512
glyma19g40340.1	RING-CH-type domain-containing protein	0,853035
glyma06g47600.1	moderately similar to (462)AT5G07670 F-box family protein	0,7553056
glyma17g12270.1	F-box domain-containing protein	0,64561826
glyma17g08200.1	Ubiquitin carboxyl-terminal hydrolase (EC 3.4.19.12)	0,9589499
glyma17g29610.1	nearly identical (1160)AT3G49600 UBP26 (ubiquitin-specific protease 26)	0,74318683
glyma19g30650.1	Ubiquitin carboxyl-terminal hydrolase (EC 3.4.19.12)	0,9015175
glyma04g01630.1	hs to (521)AT4G35350 XYLEM CYSTEINE PEPTIDASE 1; cysteine-type peptidase	2,2695782
glyma10g31430.1	Peptidase A1 domain-containing protein	0,9356521
glyma13g02190.1	Peptidase A1 domain-containing protein	0,62394
glyma06g03960.1	SERPIN domain-containing protein	0,85839
glyma06g17940.1	AAA domain-containing protein	0,76991653
glyma12g10230.1	Replication factor C subunit 1	0,76318955
glyma02g44080.1	T-complex protein 1 subunit eta (TCP-1-eta) (CCT-eta)	0,6907675
glyma03g00840.1	GrpE protein homolog	0,69914305
glyma14g04770.1	T-complex protein 1 subunit eta (TCP-1-eta) (CCT-eta)	0,99109113
glyma19g30860.1	weakly similar to (149)AT1G14980 CHAPERONIN 10	0,749138
glyma19g36280.1	Glycosyltransferases (EC 2.4.-.-)	0,94234717

Enzymes		
ID	Transcript name	Log(FC)
glyma04g10200.1	NADPH-cytochrome P450 reductase (CPR) (P450R) (EC 1.6.2.4)	0,8897577
glyma17g02160.1	DLH domain-containing protein	0,7351415
glyma12g09240.1	highly similar to (626)AT2G27690 cytochrome P450, family 94, subfamily C, polypeptide 1; oxygen binding	0,9625162
glyma11g08520.1	Peroxidase (EC 1.11.1.7)	0,64334595
glyma13g16590.1	Peroxidase (EC 1.11.1.7)	0,7608109
glyma14g38170.1	Peroxidase (EC 1.11.1.7)	0,7395987
glyma07g05310.1	Metallophos domain-containing protein	0,9223738
glyma08g22680.1	Phytoeyanin domain-containing protein	0,8654641
glyma03g34420.1	Glycosyltransferase (EC 2.4.1.-)	0,6975346
glyma09g01980.1	Hexosyltransferase (EC 2.4.1.-)	0,5979449
glyma10g07160.1	Glycosyltransferase (EC 2.4.1.-)	1,4263307
glyma20g26180.1	hs to (583)AT1G71070 glycosyltransferase family 14 protein / core-2/l-branching enzyme family protein	0,97855264
glyma18g47960.1	moderately similar to (447)AT3G03330 short-chain dehydrogenase/reductase (SDR) family protein	0,710888
glyma05g02950.1	Lipase_GDSL domain-containing protein	0,59829414
glyma14g02570.1	Lipase_GDSL domain-containing protein	3,3437998
glyma04g38590.1	Beta-galactosidase (EC 3.2.1.23)	0,6564337
glyma17g06280.1	Beta-galactosidase (EC 3.2.1.23)	0,93443227
glyma17g10060.1	alpha-1,2-Mannosidase (EC 3.2.1.-)	0,7169903
glyma03g28870.1	moderately similar to (369)AT3G57270 BETA-1,3-GLUCANASE 1; hydrolase, hydrolyzing O-glycosyl	1,7040588
glyma08g08730.1	Cas1_AcylT domain-containing protein	0,6236937
glyma03g42160.1	FAD-binding PCMH-type domain-containing protein	0,7384058
glyma04g39160.1	moderately similar to (454)AT1G48910 YUC10; monooxygenase	0,71825707
Cell		
ID	Transcript name	Log(FC)
glyma02g05650.1	Kinesin-like protein	0,78266615
glyma05g30630.1	highly similar to (730)AT1G13180 DISTORTED TRICHOMES 1; structural constituent of cytoskeleton	3,2342865
glyma10g02020.1	nearly identical (1057)AT2G47500 microtubule motor	0,60593206
glyma10g29050.1	highly similar to (866)AT5G27000 ARABIDOPSIS THALIANA KINESIN 4; microtubule motor	0,72565675
glyma14g08000.1	F-box domain-containing protein	1,126366
glyma17g05970.1	nearly identical (2425)AT5G20490 Myosin-like protein XIX; motor/ protein binding	0,63426185
glyma17g08230.1	Formin-like protein	0,9156765
glyma18g45370.1	Kinesin-like protein	0,6462352
glyma19g00850.1	highly similar to (712)AT3G12110 ACTIN-11; structural constituent of cytoskeleton	0,77856463
glyma02g43970.1	Structural maintenance of chromosomes protein	0,7355434
glyma04g38670.1	highly similar to (644)AT5G63860 UVB-RESISTANCE 8	0,80059445
glyma06g32850.1	TPR_REGION domain-containing protein	0,9049998
glyma14g03830.1	nearly identical (1583)AT5G19420 Ran GTPase binding / chromatin binding / zinc ion binding	1,8723685
glyma03g03990.1	Cyclin	2,8381448
glyma15g28380.1	Peptidyl-prolyl cis-trans isomerase (PPlase) (EC 5.2.1.8)	0,9900674
glyma01g03890.1	AP-1 complex subunit gamma	0,6402725
glyma06g07760.1	t-SNARE coiled-coil homology domain-containing protein	0,8246412
glyma11g37650.1	PH domain-containing protein	3,675837
glyma18g53730.1	weakly similar to (117)AT1G05785 Got1-like family protein	3,9832928
Development		
ID	Transcript name	Log(FC)
glyma06g04130.1	Patatin (EC 3.1.1.-)	3,7986574
glyma01g40890.1	moderately similar to (362)AT1G32400 TOM2A TOM2A (TOBAMOVIRUS MULTIPLICATION 2A)	0,8455735
glyma02g12220.1	NAC domain-containing protein	0,8366235
glyma02g16570.1	WD_REPEATS_REGION domain-containing protein	0,71733415
glyma04g07460.1	highly similar to (894)AT4G32551 RON2, LUG LUG (LEUNIG)	1,9123042
glyma07g35080.1	NusB domain-containing protein	0,92678785
glyma08g03590.1	NAC domain-containing protein	4,6786127
glyma10g33830.1	highly similar to (795)AT5G07590 WD-40 repeat protein family	1,0228174
glyma11g01570.1	nearly identical (1809)AT3G18110 EMB1270 EMB1270 (EMBRYO DEFECTIVE 1270)	0,6602813
glyma12g36500.1	highly similar to (703)AT2G16405 transducin family protein / WD-40 repeat family protein	0,8545248
glyma13g16580.1	WD_REPEATS_REGION domain-containing protein	0,652069
glyma15g10620.1	SBP-type domain-containing protein	0,65387857
glyma16g26740.1	NAC domain-containing protein	4,8830056
glyma18g44970.1	moderately similar to (345)AT3G08640 alphavirus core protein family	1,4850285
glyma19g41840.1	nearly identical (1678)AT5G27030 TPR3 TPR3 (TOPLESS-RELATED 3)	0,5932826

Transport		
ID	Transcript name	Log(FC)
glyma08g02420.1	moderately similar to (290)AT4G11150 VACUOLAR ATP SYNTHASE SUBUNIT E1	0,58776
glyma16g24220.1	V-type proton ATPase proteolipid subunit	1,4934946
glyma16g02030.1	moderately similar to (253)AT2G46800 ZINC TRANSPORTER; zinc ion transmembrane transporter	0,8532427
glyma08g02290.1	Potassium transporter	0,75871545
glyma03g36310.1	ABC transporter domain-containing protein	0,6628951
glyma08g43810.1	nearly identical (1749)AT3G13080 ATMRP3 multidrug resistance-associated protein 3	1,1196926
glyma17g12130.1	highly similar to (933)AT3G54540 ATGCN4 general control non-repressible 4	0,9904964
glyma02g13730.1	MFS domain-containing protein	2,1020465
glyma16g27330.1	highly similar to (565)AT1G22710 SUCROSE-PROTON SYMPORTER 2; carbohydrate transmembrane	0,9848119
glyma09g29880.1	Cyclic nucleotide-binding domain-containing protein	0,937424
glyma02g42050.1	Aa_trans domain-containing protein	2,974309
glyma06g15180.1	OBG-type G domain-containing protein	0,9885199
glyma08g19240.1	STAS domain-containing protein	0,75574124
glyma13g43670.1	STAS domain-containing protein	0,95080394
glyma04g07190.1	TPT domain-containing protein	2,4209015
glyma01g03090.1	Protein DETOXIFICATION (Multidrug and toxic compound extrusion protein)	1,057428
glyma04g06750.1	highly similar to (907)AT2G34190 xanthine/uracil permease family protein	0,8579135
glyma06g46880.1	DYW_deaminase domain-containing protein	0,99778944
glyma08g40530.1	Phospholipid-transporting ATPase (EC 7.6.2.1)	0,86662424
glyma12g33340.1	Phospholipid-transporting ATPase (EC 7.6.2.1)	0,5987224
glyma20g28380.3	CRAL-TRIO domain-containing protein	2,111705
Specialized metabolism		
ID	Transcript name	Log(FC)
glyma04g10200.1	NADPH--cytochrome P450 reductase (CPR) (P450R) (EC 1.6.2.4)	0,8897577
Other metabolisms		
ID	Transcript name	Log(FC)
glyma02g04790.1	highly similar to (854)AT2G17650 AMP-dependent synthetase and ligase family protein	0,9565585
glyma08g11980.1	AMP-binding domain-containing protein	0,6125697
glyma19g28300.1	nearly identical (1243)AT5G36880 acetyl-CoA synthetase, putative / acetate-CoA ligase, putative	1,1481647
glyma17g07480.1	diacylglycerol kinase (EC 2.7.1.107)	0,680659
glyma15g02710.1	Phospholipase D (EC 3.1.4.4)	4,8134546
glyma15g16270.1	Phospholipase (EC 3.1.4.4)	1,1707071
glyma14g07120.1	Delta-1-pyrroline-5-carboxylate synthase [Includes: Glutamate 5-kinase; Gamma-glutamyl phosphate	0,7867403
glyma15g00780.1	3-isopropylmalate dehydrogenase (EC 1.1.1.85)	0,7444613
glyma18g51400.1	Proline dehydrogenase (EC 1.5.5.2)	1,1664565
glyma13g37570.1	Lipoyl synthase, mitochondrial (EC 2.8.1.8) (Lipoate synthase) (LS) (Lip-syn) (Lipoic acid synthase)	0,6130037
glyma08g20080.1	hs to (682)AT1G22940 THIAMINE REQUIRING 1; catalytic/ thiamin-phosphate diphosphorylase	0,9866866
glyma07g30970.1	Flavokinase domain-containing protein	0,8959013
glyma06g13210.1	PABS domain-containing protein	0,66510385
glyma20g23670.1	Dihydroorotate dehydrogenase (quinone), mitochondrial (DHODEHASE) (EC 1.3.5.2)	0,80203885
glyma15g11310.1	moderately similar to (488)AT4G04880 adenosine/AMP deaminase family protein	0,621286
glyma04g39500.1	Ribonucleoside-diphosphate reductase (EC 1.17.4.1)	0,8512941
glyma04g09800.1	Glucose-6-phosphate isomerase (EC 5.3.1.9)	0,8394278
glyma09g01050.1	Pyrophosphate-fructose 6-phosphate 1-phosphotransferase subunit beta	0,97778517
Metal handling		
ID	Transcript name	Log(FC)
glyma07g39770.1	weakly similar to (159)AT5G50740 ATFP3 farnesylated protein 3, metal ion binding	0,67371655
DNA		
ID	Transcript name	Log(FC)
glyma01g44570.2	IPPC domain-containing protein	2,9886694
glyma10g31180.1	SAP domain-containing protein	0,706197
glyma06g12210.1	AAA domain-containing protein	3,2203126
glyma12g09220.1	DNA polymerase epsilon catalytic subunit (EC 2.7.7.7)	0,97217566
glyma08g09400.1	UmuC domain-containing protein	0,98983616
glyma19g33150.1	moderately similar to (348)AT3G12040 DNA-3-methyladenine glycosylase (MAG)	0,66331375
glyma01g00780.1	Meiotic nuclear division protein 1 homolog	0,68706644
glyma18g00530.1	Zinc-hook domain-containing protein	0,7570579
glyma01g03730.1	moderately similar to (210)AT1G23750 DNA-binding protein-related	0,6591508

RNA		
ID	Transcript name	Log(FC)
glyma02g08480.1	Polyadenylate-binding protein (PABP)	0,67034745
glyma02g08870.1	hs to (716)AT1G43190 polypyrimidine tract-binding protein /heterogeneous nuclear ribonucleoprotein,	0,7194458
glyma08g11160.1	Sm domain-containing protein	0,9960101
glyma07g31350.1	PSP domain-containing protein	0,659185
glyma17g02470.1	RRM domain-containing protein	0,92898613
glyma01g39960.1	RNase III domain-containing protein	2,3607616
glyma08g21290.2	moderately similar to (285)AT3G21350 RNA polymerase transcriptional regulation mediator-related	1,7814004
glyma13g44120.1	highly similar to (801)AT1G52620 pentatricopeptide (PPR) repeat-containing protein	0,80936927
glyma13g42550.1	C2H2-type domain-containing protein	0,89559245
glyma15g13700.1	moderately similar to (477)AT4G21160 ARF-GAP DOMAIN 12; ARF GTPase activator	0,9276568
glyma20g02690.1	highly similar to (754)AT5G37340 zinc finger (ZPR1-type) family protein	0,95993495
glyma20g33480.1	weakly similar to (183)AT5G04390 zinc finger (C2H2 type) family protein	1,4941134
glyma20g39160.1	weakly similar to (156)AT5G06770 KH domain-containing protein / zinc finger (CCCH type) family protein	0,99816495
glyma17g05920.1	weakly similar to (124)AT1G17590 CCAAT-binding transcription factor family protein	4,8049436
glyma07g04040.1	PHD-type domain-containing protein	0,89675814
glyma07g02220.1	hs to (877)AT1G79840 GL2 GL2 (GLABRA 2); DNA binding / transcription factor	2,718024
glyma08g11110.1	Agamous-like 3 protein	0,8153225
glyma01g42650.1	moderately similar to (236)AT5G02320 MYB3R-5 (myb domain protein 3R-5); DNA binding / transcription	1,3972957
glyma06g46880.1	DYW_deaminase domain-containing protein	0,99778944
glyma11g11570.1	moderately similar to (243)AT5G16770 myb domain protein 9; DNA binding	0,87171847
glyma13g41470.1	weakly similar to (194)AT3G08500 MYB83 (myb domain protein 83); DNA binding / transcription factor	0,62296355
glyma02g01420.1	WRKY transcription factor	0,8969536
glyma02g46690.1	moderately similar to (456)AT4G26640 WRKY20 (WRKY DNA-binding protein 20); transcription factor	0,98332965
glyma08g40530.1	Phospholipid-transporting ATPase (EC 7.6.2.1)	0,86662424
glyma12g33340.1	Phospholipid-transporting ATPase (EC 7.6.2.1)	0,5987224
glyma17g13280.1	Phospholipid-transporting ATPase (EC 7.6.2.1)	0,7845025
glyma20g39050.1	moderately similar to (373)AT3G12250 TGA1a-related gene 6	0,74086666
glyma07g15640.1	Auxin response factor	0,6660341
glyma10g15990.1	nearly identical (1755)AT3G57300 transcriptional activator, putative	0,75495446
glyma15g01040.1	MBD domain-containing protein	0,7072738
glyma16g05390.1	ICEa	1,2975736
glyma02g40990.1	RING-type E3 ubiquitin transferase (EC 2.3.2.27)	0,6968355
glyma01g41460.1	VEFS-Box domain-containing protein	4,352041
glyma03g01550.1	very weakly similar to (94.0)AT3G61260 DNA-binding family protein / remorin family protein	0,70838284
glyma08g01770.1	highly similar to (528)AT4G11560 bromo-adjacent homology (BAH) domain-containing protein	0,9280221
glyma12g04650.1	Putative rRNA methyltransferase (EC 2.1.1.-) (2'-O-ribose RNA methyltransferase SPB1 homolog)	1,7857494
glyma12g07590.1	nearly identical (1094)AT5G60410 ATSIZ1/SIZ1	0,77020705
glyma19g43850.1	PPC domain-containing protein	3,2812762
glyma20g24690.1	moderately similar to (471)AT5G04290 KOW domain-containing transcription factor family protein	0,58693844
glyma13g37970.1	B box-type domain-containing protein	0,6067376
glyma18g49390.1	ZF-HD dimerization-type domain-containing protein	4,7047324
glyma02g38800.1	hs to (614)AT4G02990 mitochondrial transcription termination factor family protein / mTERF family	0,8680011
glyma03g01550.1	very weakly similar to (94.0)AT3G61260 DNA-binding family protein / remorin family protein	0,70838284
glyma04g08540.1	moderately similar to (313)AT4G32720 ATLA1 ATLA1 (ARABIDOPSIS THALIANA LA PROTEIN 1); RNA	0,9411619
glyma05g34970.1	UBX domain-containing protein	0,9778135
glyma07g38920.1	DUF1084 domain-containing protein	0,70835334
glyma13g27570.1	moderately similar to (428)AT1G11650 ATRBP45B ATRBP45B; RNA binding	0,91742396
glyma18g06460.1	HIT-type domain-containing protein	0,6403229
glyma20g02690.1	highly similar to (754)AT5G37340 zinc finger (ZPR1-type) family protein	0,95993495
glyma01g39560.1	moderately similar to (257)AT2G42240 similar to RNA-binding protein, putative	0,93125737
glyma09g36880.1	RRM domain-containing protein	0,72713137
glyma13g01740.1	RRM domain-containing protein	0,7574969,

I-1-3-2 Chez les racines

Pour cette analyse, seulement 269 parmi les 553 gènes différentiellement exprimés identifiés chez les racines (cf. partie I-1-1) ont pu être annotés par Mapman, soit 48,64% des gènes. Ce faible taux d'annotation est lié au manque d'annotations au sein des bases de données du génome et des protéines de *Glycine max*.

Sur le schéma *Overview* (Figure 41), 261 gènes différentiellement exprimés chez les racines en réponse à PEP-13 sont visibles. Ces gènes sont présents dans toutes les catégories de *bins*, à l'exception de *metal handling*. Cette catégorie de gènes ne semble donc pas régulée par PEP-13. Parmi tous les gènes visibles sur ce schéma, on observe des gènes dont l'expression est positivement (rouge) et négativement (bleu) régulée en réponse à PEP-13. De plus, il apparaît qu'un plus grand nombre de gènes seraient positivement régulés en réponse à PEP-13 (155 gènes contre 106 négativement régulés). Enfin, en fonction des catégories de *bins*, on observe différentes proportions entre les gènes sous et sur-régulés. En effet, par exemple, les catégories ADN, *developpement* ou encore *regulation redox* sont représentées principalement par des gènes dont l'expression est régulée positivement.

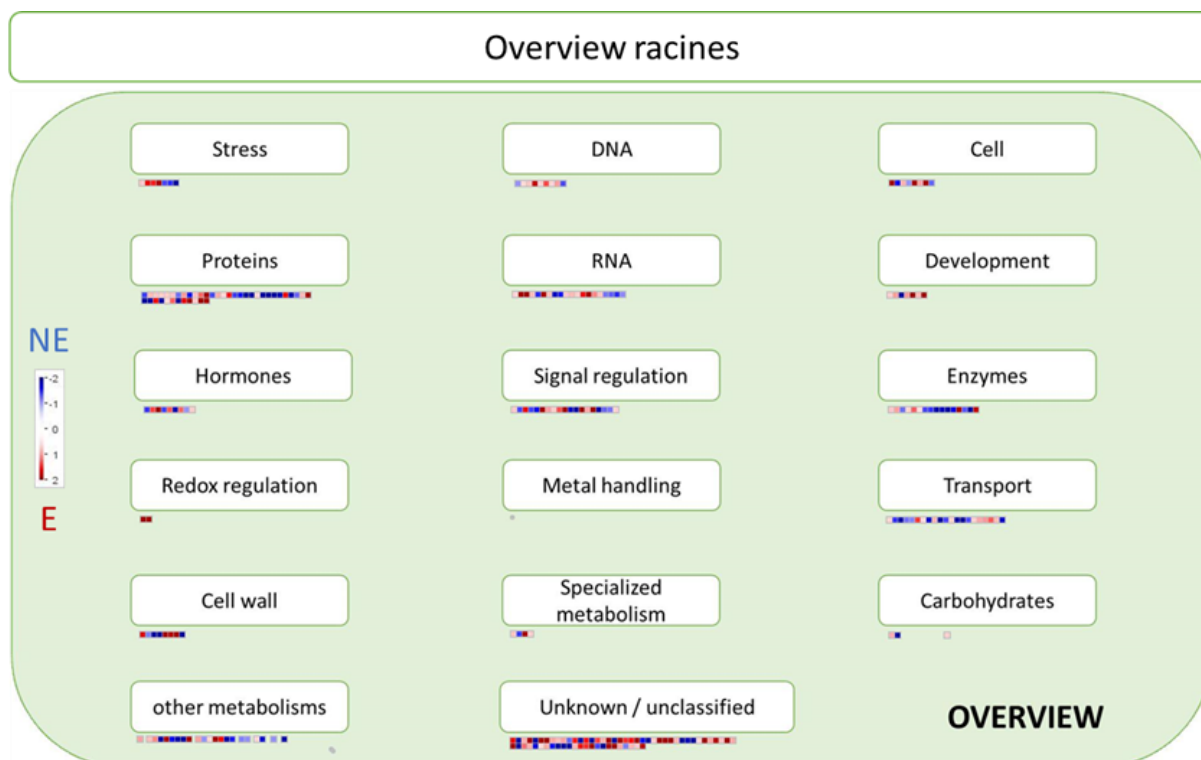


Figure 41 : Représentation générale (*overview*) des 261 gènes différentiellement exprimés chez les racines de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)

Cette représentation permet d’observer les gènes régulés négativement ou positivement chez les racines en réponse à PEP-13 en fonction des processus biologiques dans lesquels ils interviennent. Chaque gène est représenté par un carré dont l’intensité de la couleur dépend du Log2 de son *fold change* (FC). Ainsi, les gènes sur-exprimés en réponse à PEP-13 sont représentés en dégradé de rouge, et les gènes sous-exprimés en réponse à PEP-13 sont représentés en dégradé de bleu. Graphique réalisé avec Mapman.

Pour la suite de cette analyse, seuls les 155 gènes dont l’expression est régulée positivement chez les racines ont été pris en compte (les gènes sous-exprimés seront présentés en annexe III, partie I-2). Pour cela, le détail de ces gènes sur-exprimés est présenté sous forme de tableau classé par catégories de *bins* (Tableau 10), indiquant l’identifiant du transcrit, le nom de celui-ci ainsi que son FC (exprimé sous forme de logarithme binaire). Les détails concernant les gènes les plus exprimés seront exposés dans la discussion de ce manuscrit.

Tableau 10 : Détail des 155 gènes sur-exprimés chez les racines de soja en réponse à PEP-13 classés en fonction de leur catégorie de *bins*

Une couleur est attribuée au log du FC en fonction de sa valeur, plus celle-ci est élevée plus celui-ci apparaîtra en rouge foncé.

Stress		
ID	Transcript names	Log(FC)
glyma08g10960.1	moderately similar to (208)AT1G67360 rubber elongation factor (REF) family protein	0,75935256
glyma02g01020.1	moderately similar to (407)AT1G02360 chitinase, putative	1,7058877
glyma09g02420.1	moderately similar to (360)AT3G14470 disease resistance protein (NBS-LRR class), putative	2,1978064
glyma16g03780.1	TIR domain-containing protein	1,5108746
Cell Wall		
ID	Transcript names	Log(FC)
glyma08g15680.1	nearly identical (1225)AT1G78570 RHAMNOSE BIOSYNTHESIS1; UDP-L-rhamnose synthase/ UDP-glucose 4,6-dehydratase	1,7578726
glyma09g37090.1	Expansin	3,1668162
glyma17g15670.1	moderately similar to (243)AT4G17030 ATEXLB1 (ARABIDOPSIS THALIANA EXPANSIN-LIKE B1)	1,8870727
glyma05g32380.1	PMEI domain-containing protein	2,3724911
Carbohydrates		
ID	Transcript names	Log(FC)
glyma07g01090.1	Beta-fructofuranosidase (EC 3.2.1.26)	0,6900553
glyma14g01430.1	nearly identical (1119)AT1G55740 A. thaliana seed imbibition 1; hydrolase, hydrolyzing O-glycosyl compounds	0,8953937
Hormones		
ID	Transcript names	Log(FC)
glyma03g38030.1	Fe2OG dioxygenase domain-containing protein	1,3861163
glyma09g37890.1	Fe2OG dioxygenase domain-containing protein	1,3051107
glyma13g21120.1	Fe2OG dioxygenase domain-containing protein	2,6218588
glyma20g33800.1	AP2/ERF domain-containing protein	1,2424468
glyma14g35940.1	Usp domain-containing protein	0,7241514
Signal regulation		
ID	Transcript names	Log(FC)
glyma19g31110.1	Phosphatidylinositol 4-phosphate 5-kinase (EC 2.7.1.68)	4,0510473
glyma03g00950.1	moderately similar to (484)AT5G17790 VAR3 (VARIEGATED 3); binding	0,9919649
glyma03g05630.1	PRA1 family protein	2,6830425
glyma12g32020.1	Glutamate receptor	0,72754586
glyma02g17240.1	highly similar to (732)AT1G03010 phototropic-responsive NPH3 family protein	0,6671637
glyma06g46120.1	moderately similar to (315)AT3G21700 GTP binding	0,93731904
glyma07g09330.1	PRONE domain-containing protein	4,9607115
glyma08g11140.1	moderately similar to (207)AT3G07880 Rho GDP-dissociation inhibitor family protein	1,3421267
glyma09g34300.1	nearly identical (1134)AT1G31930 XLG3 (EXTRA-LARGE GTP-BINDING PROTEIN 3); signal transducer	3,514107
glyma11g33100.1	moderately similar to (305)AT5G45130 RHA1	0,7314509
Proteins		
ID	Transcript names	Log(FC)
glyma02g10060.1	highly similar to (674)AT3G58140 phenylalanyl-tRNA synthetase class IIc family protein	0,72453153
glyma02g03750.1	AA_TRNA_LIGASE_II domain-containing protein	0,8411204
glyma01g45530.1	PseudoU_synth_2 domain-containing protein	0,73136175
glyma17g02970.1	nearly identical (1615)AT5G49030 OVA2 (OVULE ABORTION 2); ATP binding / aminoacyl-tRNA ligase	0,76558083
glyma10g36070.1	weakly similar to (152)AT1G35680 50S ribosomal protein L21, chloroplast / CL21 (RPL21)	0,76636434
glyma01g43110.1	S10_pectin domain-containing protein	0,9988927
glyma03g39480.1	weakly similar to (155)AT3G05560 60S ribosomal protein L22-2 (RPL22B)	0,7857041
glyma03g35530.1	weakly similar to (195)AT5G46430 60S ribosomal protein L32 (RPL32B)	1,250446
glyma15g04670.1	Ribosomal_L23eN domain-containing protein	4,7370462
glyma02g08010.1	moderately similar to (347)AT5G18110 NCBP (NOVEL CAP-BINDING PROTEIN); RNA binding / translation initiation factor	0,9090984
glyma13g44070.1	Nuclear pore protein	0,6469292
glyma08g21380.1	highly similar to (601)AT1G51980 mitochondrial processing peptidase alpha subunit, putative	1,5444384
glyma15g10470.1	Protein kinase domain-containing protein	0,6613854
glyma17g15720.1	ATP-dependent Clp protease proteolytic subunit	0,8605938
glyma03g41810.1	Sde2_N_Ubi domain-containing protein	1,6340258
glyma02g05190.1	Proteasome subunit alpha type (EC 3.4.25.1)	3,3842962
glyma07g05190.1	RING-type domain-containing protein	0,6534389
glyma20g38050.1	RING-type domain-containing protein	1,2560668
glyma03g41530.1	F-box domain-containing protein	7,1263742
glyma19g44130.1	F-box domain-containing protein	1,7039046
glyma17g08200.1	Ubiquitin carboxyl-terminal hydrolase (EC 3.4.19.12)	0,74444294
glyma13g21180.1	Peptidase A1 domain-containing protein	2,225299
glyma10g07290.1	Peptidylprolyl isomerase (EC 5.2.1.8)	4,3470616

Enzymes		
ID	Transcript names	Log(FC)
glyma03g03320.1	PMEI domain-containing protein	2,22463
glyma10g25500.1	moderately similar to (439)AT3G24040 glycosyltransferase family 14 protein / core-2/I-branching enzyme family protein	0,7544098
glyma17g36100.1	Hexosyltransferase (EC 2.4.1.-)	0,97302425
glyma18g00620.1	Glycosyltransferase (EC 2.4.1.-)	0,6473919
glyma15g08700.1	SGNH_hydro domain-containing protein	1,8901474
glyma02g05240.1	Alpha-mannosidase (EC 3.2.1.-)	0,5926045
glyma06g02760.1	Endoglucanase (EC 3.2.1.4)	1,27957
Cell		
ID	Transcript names	Log(FC)
glyma13g40580.1	Kinesin motor domain-containing protein	0,8839751
glyma20g29590.1	highly similar to (717)AT5G07270 ankyrin repeat family protein	2,4811997
glyma20g37340.1	Kinesin motor domain-containing protein	4,44838
glyma20g38510.1	ANK_REP_REGION domain-containing protein	0,9793996
glyma13g33450.1	Syntaxin-18_N domain-containing protein	3,547322
Development		
ID	Transcript names	Log(FC)
glyma01g36620.1	Rhodanese domain-containing protein	0,9920425
glyma01g43620.1	GRAS domain-containing protein	2,0111651
glyma03g30210.1	PMR5N domain-containing protein	0,7015297
glyma14g04510.1	nearly identical (1362)AT2G27040 AGO4 (ARGONAUTE 4); nucleic acid binding	0,85857165
glyma15g10620.1	SBP-type domain-containing protein	1,0379298
glyma16g26660.1	FT-like protein (FT2a) (FT3 protein) (Flowering locus T) (Flowering locus T-like protein 3)	2,2067451
Transport		
ID	Transcript names	Log(FC)
glyma06g07990.1	Plasma membrane ATPase (EC 7.1.2.1)	0,662052
glyma06g16670.1	Probable magnesium transporter	1,4127433
glyma11g14370.1	moderately similar to (310)AT2G29410 MTPB1; efflux transmembrane transporter/ zinc ion transmembrane transporter	0,6255647
glyma07g02140.1	highly similar to (675)AT1G52190 proton-dependent oligopeptide transport (POT) family protein	0,87323934
glyma08g02290.1	Potassium transporter	0,7088135
glyma15g00620.1	moderately similar to (311)AT1G80760 NIP6;1 (NOD26-like intrinsic protein 6;1); water channel	0,6949705
glyma01g03090.1	Protein DETOXIFICATION (Multidrug and toxic compound extrusion protein)	1,0198045
glyma04g06750.1	highly similar to (907)AT2G34190 xanthine/uracil permease family protein	0,8230362
glyma12g32010.1	Protein DETOXIFICATION (Multidrug and toxic compound extrusion protein)	0,93658096
glyma20g28380.3	CRAL-TRIO domain-containing protein	1,2849158
Specialized metabolism		
ID	Transcript names	Log(FC)
glyma08g10960.1	moderately similar to (208)AT1G67360 rubber elongation factor (REF) family protein	0,75935256
glyma17g07670.1	highly similar to (872)AT5G57800 CER3/FLP1/WAX2/YRE (ECERIFERUM 3); catalytic	2,3483984
glyma12g32640.1	weakly similar to (194)AT1G03495 transferase	0,71502894
Other metabolisms		
ID	Transcript names	Log(FC)
glyma10g00440.1	3-ketoacyl-CoA synthase (EC 2.3.1.-)	5,588424
glyma02g04790.1	highly similar to (854)AT2G17650 AMP-dependent synthetase and ligase family protein	1,0854774
glyma08g11980.1	AMP-binding domain-containing protein	0,7987894
glyma08g22420.1	Abhydrolase_2 domain-containing protein	2,8160908
glyma08g24510.1	NADH-cytochrome b5 reductase (EC 1.6.2.2)	0,6472452
glyma05g31630.1	highly similar to (770)AT4G39660 AGT2 (ALANINE:GLYOXYLATE AMINOTRANSFERASE 2); alanine-glyoxylate transaminase	0,9662264
glyma10g40750.1	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	0,64364755
glyma09g37490.1	moderately similar to (317)AT5G39890 similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G15120.1)	3,110371
glyma16g03080.1	SATase_N domain-containing protein	1,637793
glyma05g32640.1	moderately similar to (482)AT3G55010 ATPURM/PUR5; phosphoribosylformylglycinamide cyclo-ligase	0,9601251
Redox regulation		
ID	Transcript names	Log(FC)
glyma03g30800.1	Glutathione peroxidase	5,6325154
glyma04g08380.1	moderately similar to (213)AT4G32690 GLB3 (2-on-2 hemoglobin like gene 3)	6,4543123

DNA		
ID	Transcript names	Log(FC)
glyma06g11980.1	AAA domain-containing protein	0,62201333
glyma10g31180.1	SAP domain-containing protein	0,77583313
glyma01g00780.1	Meiotic nuclear division protein 1 homolog	0,711015
glyma10g42180.1	moderately similar to (349)AT3G02540 RAD23-3 (PUTATIVE DNA REPAIR PROTEIN RAD23-3)	1,8783112
glyma17g13050.1	DNA_MISMATCH_REPAIR_2 domain-containing protein	0,6718924
glyma18g53470.1	moderately similar to (330)AT5G38470 DNA repair protein RAD23, putative	1,3641057
glyma06g06460.1	Transmembrane 9 superfamily member	0,99668115
RNA		
ID	Transcript names	Log(FC)
glyma16g32890.1	CPSF_A domain-containing protein	0,65725136
glyma01g39960.1	RNase III domain-containing protein	4,74329
glyma07g37280.1	Bet_v_1 domain-containing protein	5,2798715
glyma07g09230.1	moderately similar to (303)AT3G22320 RPABC24.3 (RNA POLYMERASE I, II AND III 24.3 KDA SUBUNIT)	0,8337755
glyma08g04950.1	nearly identical (1590)AT3G18290 EMB2454 (EMBRYO DEFECTIVE 2454); protein binding / zinc ion binding	0,7078029
glyma03g01000.1	Homeobox domain-containing protein	1,5651357
glyma08g20270.1	weakly similar to (182)AT1G72740 DNA-binding family protein / histone H1/H5 family protein	2,3931844
glyma13g29160.1	weakly similar to (130)AT4G18390 TCP family transcription factor, putative	1,1085379
glyma02g08020.1	AP2/ERF domain-containing protein	5,4364114
glyma07g15640.1	Auxin response factor	0,821488
glyma04g36790.1	Elongator complex protein 3 (EC 2.3.1.-)	0,780216
glyma10g28290.1	BHLH transcription factor	0,7554386
glyma16g02640.1	very weakly similar to (88.2)AT4G00950 MEE47 (maternal effect embryo arrest 47); transcription factor	0,8363928

I-1-4 Tentative d'identification du récepteur de PEP-13

Le récepteur de PEP-13 étant inconnu, une tentative d'identification de celui-ci a été réalisée. Pour cela, les 326 gènes sur-exprimés (FC supérieur ou égal à 1,5 et *p-value* inférieure ou égale à 0,05) dans les racines et 719 gènes dans les AC-DC en réponse à PEP-13 (cf. I-1-1) ont été analysés dans le but de détecter des domaines correspondants à des récepteurs.

Ces gènes ont tout d'abord été triés afin de ne conserver que ceux étant non annotés ou annotés comme récepteurs. Ce tri a permis d'obtenir 77 gènes chez les racines et 174 gènes chez les AC-DC. Ces gènes ont ensuite été comparés avec Jvenn (Figure 42) afin d'identifier les gènes communs.

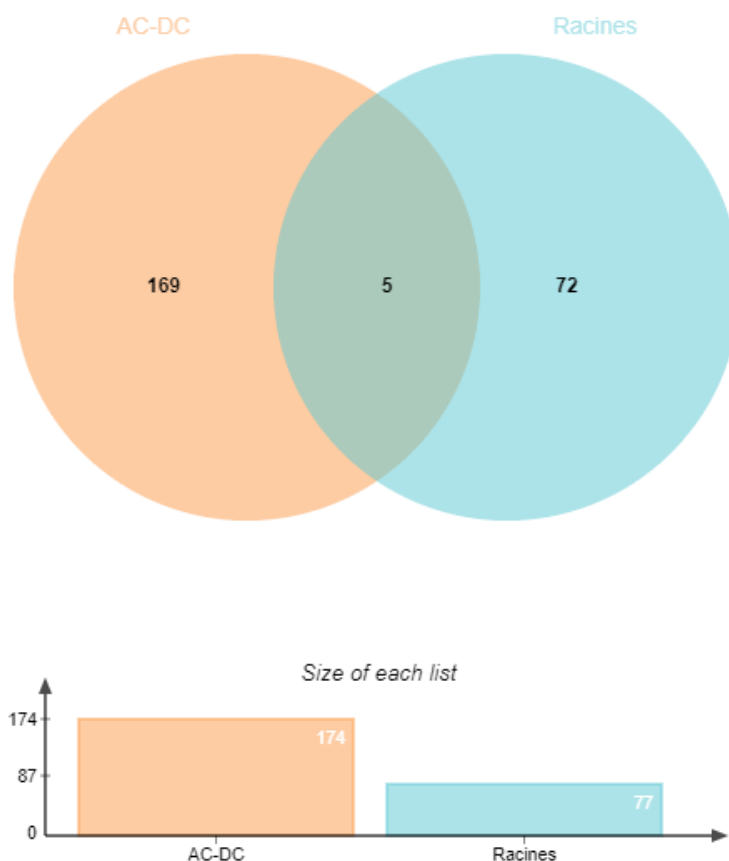


Figure 42 : Diagramme de Venn comparant les gènes sur-exprimés chez les AC-DC (*root associated cap-derived cells*) et chez les racines de soja en réponse à PEP-13 et étant non annotés ou annotés comme récepteurs

Sur ce graphique, les listes des gènes sur-exprimés (FC supérieur ou égal à 1,5 et *p-value* inférieure à 0,05) chez les AC-DC (orange) et chez les racines (turquoise) et étant non annotés ou annotés comme récepteurs sont représentées par un cercle. L'intersection entre ces deux cercles correspond aux gènes retrouvés dans les deux listes. Ainsi, 5 gènes sont communs chez les AC-DC et chez les racines parmi les 174 et 77 gènes sur-exprimés en réponse à PEP-13 et étant non annotés ou annotés comme récepteurs. Graphique réalisé avec Jvenn.

Ces cinq gènes, GLYMA_18G270300, GLYMA_06G296500, GLYMA_20G145500, GLYMA_15G123000 et GLYMA_10G168600, ont ensuite été analysés avec DRAGO2 afin de déterminer si ceux-ci possèdent ou non un domaine récepteur. Aucun domaine récepteur n’a pu être détecté chez les transcrits de ces cinq gènes. Pour un FC supérieur ou égal à 1,5 et une *p-value* inférieure à 0,05, le récepteur de PEP-13 ne semble donc pas être détecté parmi les gènes sur-exprimés communément chez ces deux zones. Celui-ci pourrait cependant être sur-exprimé (FC supérieur à 1,5 et *p-value* inférieure à 0,05) uniquement dans l’une de ces deux zones.

Une nouvelle analyse DRAGO2 a donc été réalisée sur l’ensemble des gènes sur-exprimés en réponse à PEP-13 et non annotés ou étant annotés comme récepteurs chez les AC-DC (174 gènes) et chez les racines (77 gènes). Parmi ces listes, 16 gènes chez les AC-DC et 4 gènes chez les racines dont le transcrit semble correspondre à un récepteur ont été détectés (Tableau 11). Le récepteur de PEP-13 pourrait donc être l’un de ces 20 gènes.

Tableau 11 : Gènes identifiés chez les AC-DC (*root associated cap-derived cells*) et les racines de soja comme potentiels récepteurs de PEP-13 grâce à DRAGO2

On observe pour chaque gène le logarithme binaire de son FC (Log2(FC)), sa classification par type de récepteur ainsi que les différents domaines détectés par DRAGO2. RLP : *Receptor-like protein* ; RLK : *Receptor-like kinase* ; LRR : *leucine rich repeat* ; KIN : *kinase* ; NBS : *Nucleotide Binding Site* ; CC : *Coiled-Coil* ; TIR : *toll/interleukine receptor*

AC-DC				
ID	Log2(FC)	Classe	Domaine	
GLYMA_01G025600	0,7168	RLP	TM, LRR	
GLYMA_01G036200	1,1356	RLP	LRR, TM	
GLYMA_02G036500	0,9746	RLK	TM, kinase, LRR	
GLYMA_03G143300	0,6480	L	LRR	
GLYMA_03G148800	0,7246	KIN	Kinase	
GLYMA_06G036000	1,0310	RLP	TM, LRR	
GLYMA_07G046800	0,5969	RLK	TM, kinase, LRR	
GLYMA_08G250200	0,7703	N	NBS, TM	
GLYMA_11G013700	0,6603	CN	CC, TM	
GLYMA_11G113600	0,7062	N	NBS	
GLYMA_11G166200	0,9634	RLP	TM, LRR	
GLYMA_15G022500	0,6510	L	LRR	
GLYMA_17G226100	0,6309	RLK	LRR, Kinase, TM	
GLYMA_17G250800	4,1979	RLK	LRR, Kinase, TM	
GLYMA_19G220300	1,1683	KIN	TM, Kinase	
GLYMA_20G232100	0,8048	KIN	TM, Kinase	
Racines				
ID	Log2(FC)	Classe	Domaine	
GLYMA_12G103400	4,0958	KIN	TM, Kinase	
GLYMA_15G116400	2,0890	T	TIR	
GLYMA_19G171100	0,9891	RLP	TM, LRR	
GLYMA_20G100500	4,9704	RLP	LRR, TM	

Enfin, une analyse de modélisation 3D a été réalisée avec I-TASSER sur le transcrit correspondant au gène dont le FC est plus important ($\text{Log}_2(\text{FC})=4,9704$ soit un FC de 31,35) : GLYMA_20G100500 identifié dans la racine. Cette analyse a permis d'obtenir une prédiction de la conformation de la protéine correspondante (Figure 43) ainsi qu'une prédiction de ses ligands possibles. La conformation prédite a obtenu un C-score de -1,03 (ce score pouvant varier de -5 à 2 et 2 étant le score de confiance le plus élevé). Le score de confiance obtenu semble donc plutôt moyen. Les deux ligands avec le meilleur score prédits pour cette structure sont des peptides. Le premier correspond à YIYTQ (C-score = 0,08) ; et le second correspond à PIPPSAPSKRHN (C-score = 0,04). Aucun de ces deux ligands ne semble correspondre à PEP-13. Il semblerait donc que malgré son FC important, ce récepteur ne soit pas celui correspondant à PEP-13.

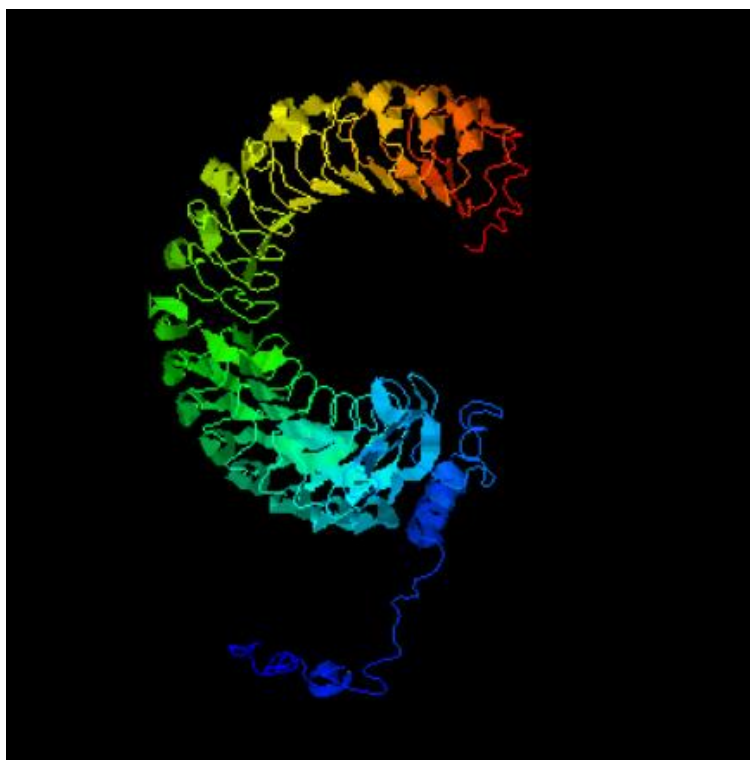


Figure 43 : Prédiction de la conformation 3D de GLYMA20G100500

Cette structure correspond à celle prédite pour GLYMA20G100500, gène codant un récepteur potentiel de type RLP (*receptor like protein*) chez le soja et dont le logarithme binaire du FC atteint 4,9704 en réponse à PEP-13 chez les racines. Ce gène n'a pas été détecté comme sur-exprimé (FC supérieur ou égal à 1,5 et *p-value* inférieure ou égale à 0,05) chez les AC-DC en réponse à PEP-13. Le score de confiance C-score obtenu pour cette structure est de -1,03. Figure obtenue avec I-TASSER.

I-2 Analyses protéomiques

Une analyse protéomique des AC-DC et des racines de soja en condition élicitée avec PEP-13 ou témoin a été réalisée afin de déterminer s'il existe des différences de protéome entre les deux zones étudiées, mais également entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE). Cette analyse a permis d'identifier 1 535 protéines au sein des différents échantillons d'AC-DC et de racines.

Dans un premier temps, une analyse en composante principale des différentes conditions pour chacune des zones a été réalisée grâce au logiciel Progenesis LC-MS (Figure 44). Cette analyse permet d'observer que les deux premières composantes statistiques permettent d'expliquer 84,08 et 84,71% de l'inertie totale des nuages de points. Cela signifie qu'il y a donc une différence statistique entre les populations de peptides identifiés pour chacune des zones, l'une correspondant à la condition élicitée et l'autre à la condition non élicitée. Il y a donc une différence entre les protéomes en condition élicitée avec PEP-13 et en condition témoin non élicitée.

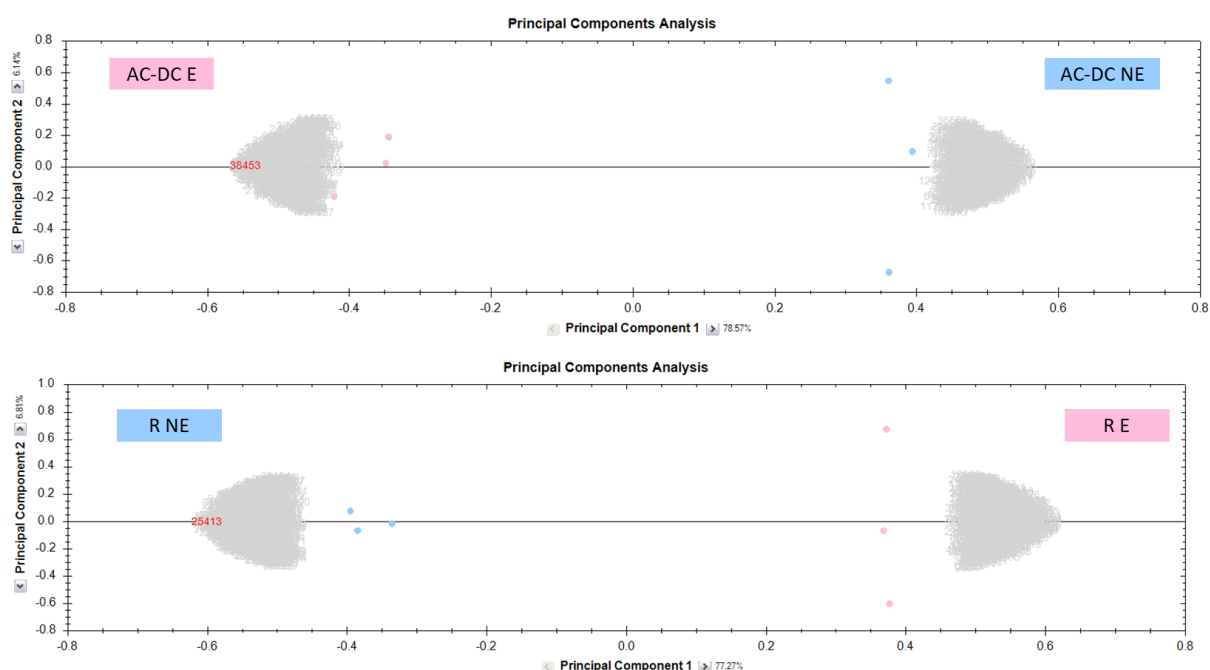


Figure 44 : Analyses en composante principale des protéomes des AC-DC et des racines de soja

Les deux premiers facteurs sont retenus pour l'analyse car ils contribuent à eux seuls à 84,08 et 84,71% de l'inertie totale du nuage de points (F1 : 77,27 et 78,57 % ; F2 : 6,81 et 6,14%). On observe sur ces deux graphes la séparation des populations de peptides identifiés en fonction de la condition élicitée (E) ou témoin (NE).

Afin de comparer les zones entre elles, une seconde analyse en composante principale a été réalisée (Figure 45). Cette analyse permet d'observer que les deux premières composantes statistiques permettent d'expliquer 77,52 % de l'inertie totale du nuage de points. Au sein de ce nuage, on observe la séparation des échantillons en fonction tout d'abord de leur origine (racine ou AC-DC) puis, plus légèrement, en fonction de leur condition élicitée ou témoin. Cela signifie qu'il y a une différence statistique entre le protéome des racines et des AC-DC.

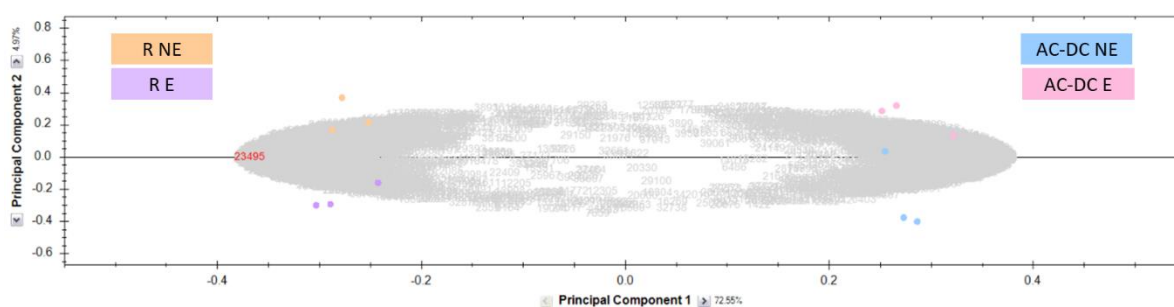


Figure 45 : Analyse en composante principale de l'ensemble des protéomes élicités ou témoin des racines et des AC-DC de soja

Les deux premiers facteurs sont retenus pour l'analyse car ils contribuent à eux seuls à 77,52 % de l'inertie totale du nuage de points (F1 : 72,55 % ; F2 : 4,97 %). On observe sur ce graphe la séparation des populations de peptides identifiés en fonction de la condition élicitée (E) ou témoin (NE) et du tissu dans lequel ils ont été identifiés (racines ou AC-DC).

I-2-1 Identification des protéines différentiellement exprimées suite à l'élicitation

Afin d'étudier plus en détail les protéines impliquées dans la réponse à PEP-13 des AC-DC et des racines, des graphiques *volcanoplots* ont été réalisés pour chacune des zones (Figures 46 et 47).

Pour les AC-DC (Figure 46), le *volcanoplot* met en évidence la présence d'un faible nombre de protéines différentiellement exprimées (FC supérieur ou égal à 1,5 et *p-value* inférieure ou égale à 0,05). En effet, 31 protéines ont été identifiées comme sur-exprimées en condition élicitée et non élicitée. De plus, on observe un plus grand nombre de protéines sur-exprimées dans la condition témoin non élicitée (26 protéines) par rapport à la condition élicitée (5 protéines).

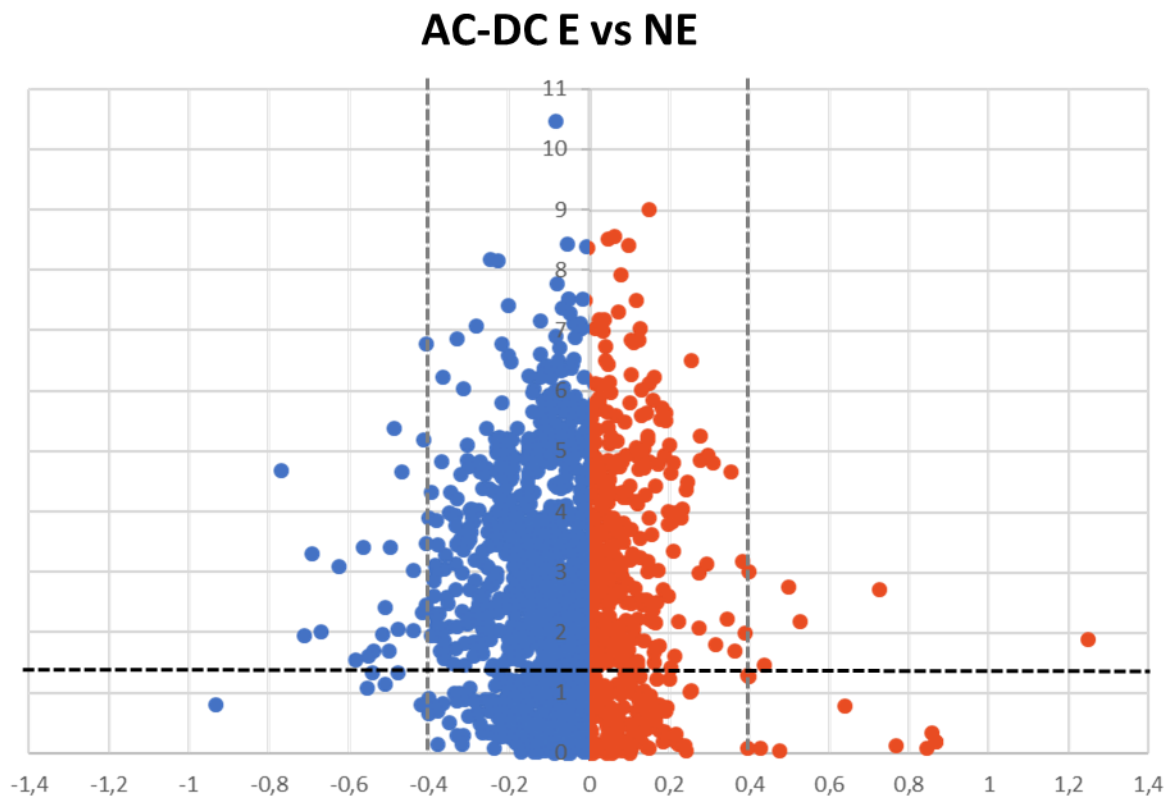


Figure 46 : Analyse de l'expression différentielle des protéines des AC-DC (*root-Associated Cap-Derived Cells*) de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)

Sur ce graphique, chaque protéine ayant été détectée est représentée par un point. Ces points sont placés grâce à la valeur absolue du Log_{10} de leur p -value et au Log_2 de leur *fold change*. La partie gauche de ce graphique (bleue) correspond aux protéines ayant été détectées en plus grande quantité dans la condition non élicitée (NE) et la partie droite de ce graphique (orange) correspond aux protéines ayant été détectées en plus grande quantité dans la condition élicitée (E). Dans cette étude, les protéines ayant une p -value inférieure ou égale à 0,05 (soit $|\text{Log}_{10}| = 1,301029996$, représenté par une ligne horizontale rouge) et un FC supérieur ou égal à 1,5 (soit $\text{Log}_2 = 0,5849625$, représenté par une ligne verticale noire) ont été sélectionnées car considérées comme différentiellement exprimées pour le seuil statistique choisi. Graphique réalisé avec Excel.

Pour les racines (Figure 47), on détecte à nouveau une faible quantité de protéines différentiellement exprimées (FC supérieur ou égal à 1,5 et p -value inférieure ou égale à 0,05). En effet seulement 11 protéines différentiellement exprimées ont été détectées. Parmi ces protéines, 6 sont régulées positivement en réponse à PEP-13 et 7 sont régulées négativement.

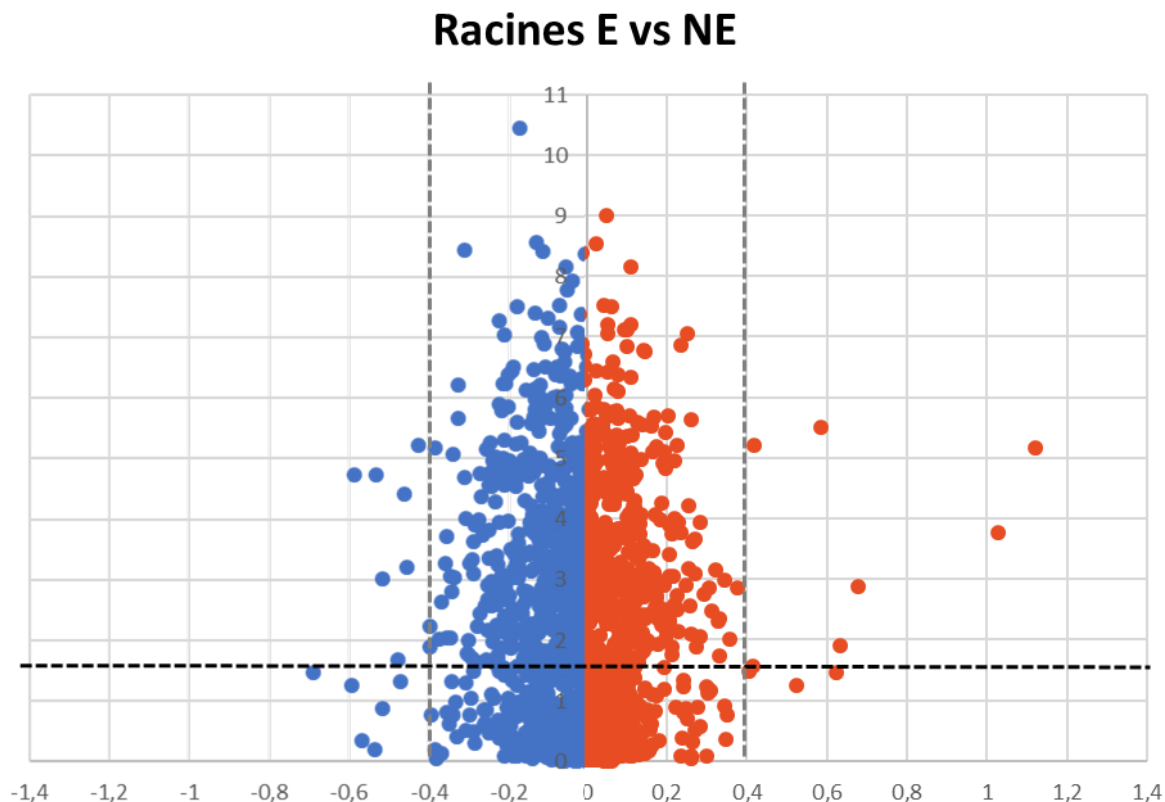


Figure 47 : Analyse de l'expression différentielle des protéines des racines de soja entre les conditions élicitée avec PEP-13 (E) et témoin non élicitée (NE)

Sur ce graphique, chaque protéine ayant été détectée est représentée par un point. Ces points sont placés grâce à la valeur absolue du Log10 de leur *p-value* et au Log2 de leur *fold change*. La partie gauche de ce graphique (bleue) correspond aux protéines ayant été détectées en plus grande quantité dans la condition non élicitée (NE) et la partie droite de ce graphique (orange) correspond aux protéines ayant été détectées en plus grande quantité dans la condition élicitée (E). Dans cette étude, les protéines ayant une *p-value* inférieure ou égale à 0,05 (soit $|\text{Log}_{10}| = 1,301029996$, représenté par une ligne horizontale rouge) et un FC supérieur ou égal à 1,5 (soit $\text{Log}_2 = 0,5849625$, représenté par une ligne verticale noire) ont été sélectionnées car considérées comme différentiellement exprimées pour le seuil statistique choisi. Graphique réalisé avec Excel.

Si l'on compare les résultats obtenus pour les AC-DC et pour les racines, on observe qu'un plus grand nombre de protéines différentiellement exprimées ont été détectées chez les AC-DC (31) par rapport aux racines (13). De plus, environ autant de protéines sur (6) ou sous-exprimées (7) en réponse à PEP-13 ont été détectées chez les racines, en revanche, chez les AC-DC un plus grand nombre de protéines sous-exprimées (26) en réponse à PEP-13 ont été détectées par rapport à la condition élicitée (5).

Afin de vérifier si les protéines détectées comme différentiellement exprimées (FC supérieur ou égal à 1,5 et *p-value* inférieure ou égale à 0,05) chez les AC-DC et chez les racines en réponse à PEP-13 sont similaires ou non, un diagramme de Venn a été réalisé afin de comparer les deux listes (Figure 48). Cette analyse permet d'observer qu'aucune protéine

parmi les 44 détectées et différentiellement exprimées n'est commune aux deux zones. Il y aurait donc bien une différence de réponse protéomique à PEP-13 entre AC-DC et racines.

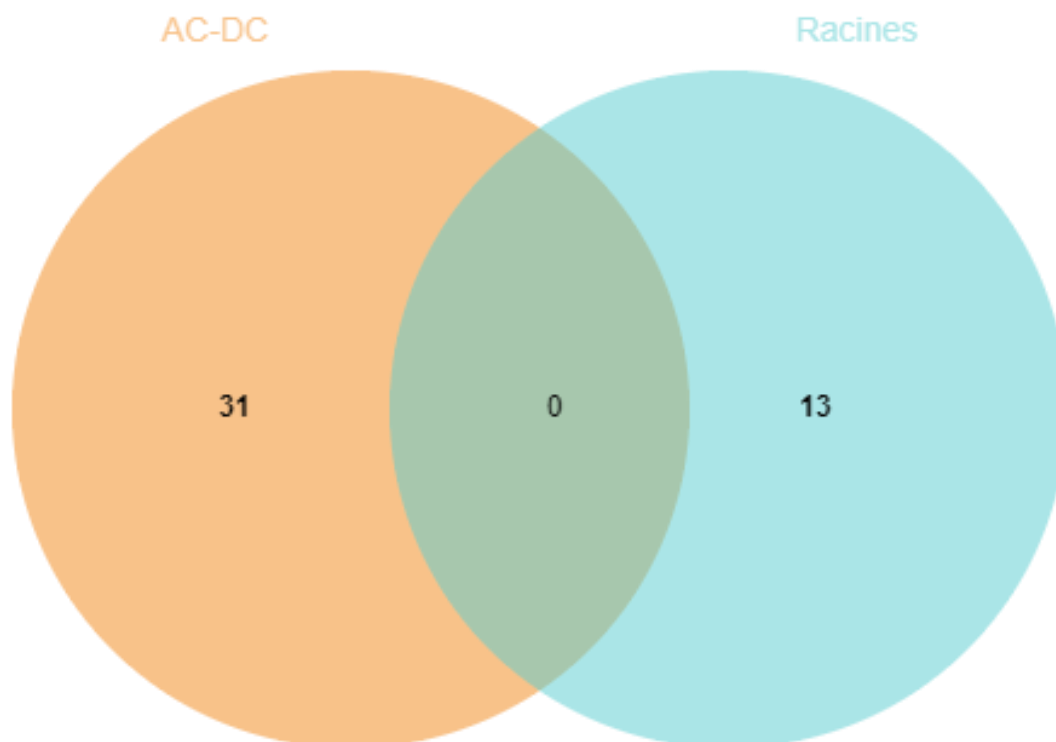


Figure 48 : Diagramme de Venn comparant les protéines différentiellement exprimées chez les racines (13) et chez les AC-DC (31) de soja en conditions élicitée avec PEP-13 et témoin non élicitée

Sur ce graphique les listes des protéines différentiellement exprimées (FC supérieur ou égal à 1,5 et *p-value* inférieure à 0,05) chez les AC-DC (orange) et chez les racines (bleu) sont représentées par un cercle. L'intersection entre ces deux cercles correspond aux protéines retrouvées dans les deux listes. Ainsi, il n'y a aucune protéine commune aux AC-DC et aux racines parmi les 31 et 13 protéines différentiellement exprimés entre les conditions élicitée et témoin. Graphique réalisé avec Jvenn.

I-2-2 Analyse des processus biologiques impliquées dans la réponse à PEP-13

Afin d'étudier plus en détail la réponse des AC-DC et des racines à PEP-13, une analyse des GO a été réalisée grâce à WEGO 2.0. Pour cela, les listes des 26 et 7 protéines négativement régulées en réponse à PEP-13 respectivement chez les AC-DC et chez les racines et des 5 et 6 protéines positivement régulées en réponse à PEP-13 respectivement chez les AC-DC et les racines ont été utilisées.

Pour les AC-DC (Figure 49), on observe que les processus biologiques les plus représentés par les protéines détectées sont les processus cellulaires et métaboliques, ainsi que la régulation biologique, la synthèse et l'organisation des composés cellulaires et la réponse aux stimuli. Ces processus semblent à la fois régulés positivement et négativement par PEP-13. Enfin, on observe des processus qui semblent régulés uniquement positivement (régulation) et uniquement négativement (localisation, développement, processus spécifiques aux organismes multi-cellulaires) par PEP-13.

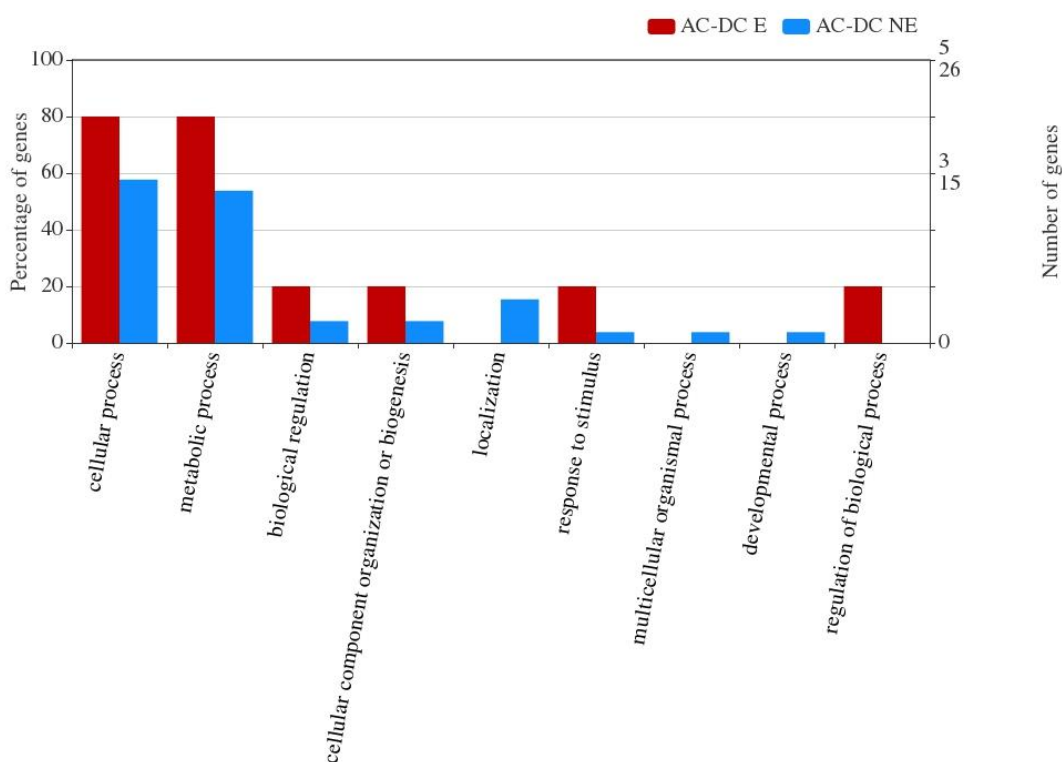


Figure 49 : Processus biologiques (*Gene Ontology*, GO) des 31 protéines régulées en réponse à PEP-13 chez les AC-DC (*root-Associated Cap-Derived Cells*) de soja

Pour chaque protéine, un certain nombre de *Gene Ontology* (GO) ont été assignés en fonction de leur annotation (axe des abscisses). Pour chaque GO, les protéines sont comptabilisées en nombre et pourcentage de gènes (axes des ordonnées). Enfin, pour un FC supérieur ou égal à 1,5 et une *p-value* inférieure à 0,05, la liste des protéines sous-exprimées en réponse à PEP-13 dans les AC-DC est représentée en bleu, et celle des protéines positivement régulées par PEP-13 dans les AC-DC est représentée en rouge. Graphique réalisé avec WEGO 2.0.

Chez les racines (Figure 50), les processus biologiques les plus représentés par les protéines différentiellement exprimées en réponse à PEP-13 sont les processus métaboliques et cellulaires, la synthèse et l'organisation des composants cellulaires, et la régulation. Ces processus sont régulés à la fois positivement et négativement par PEP-13. De plus, on observe des processus biologiques qui semblent uniquement régulés positivement en réponse à PEP-13 (régulation positive et négative, localisation) ou uniquement régulés négativement en réponse à PEP-13 (réponse aux stimuli).

Pour chaque protéine, un certain nombre de *Gene Ontology* (GO) ont été assignés en fonction de leur

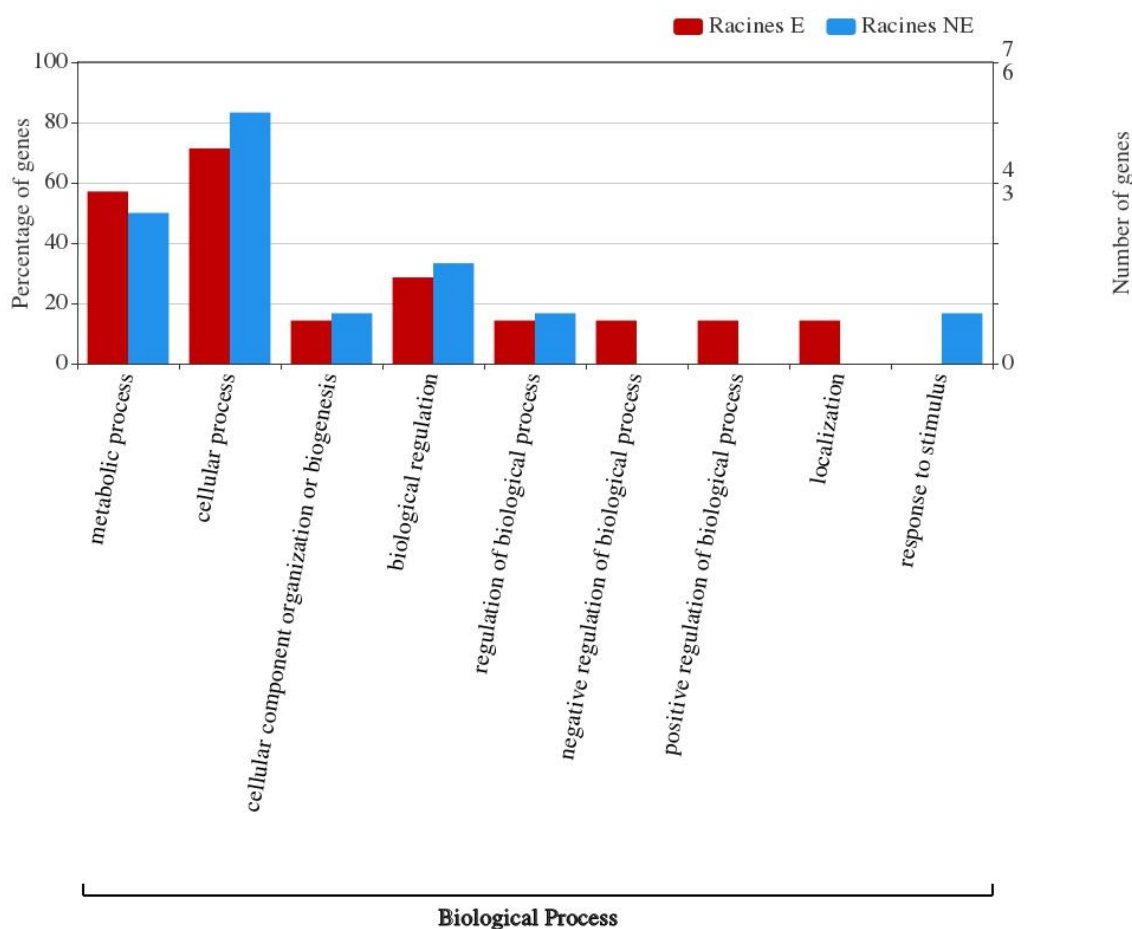


Figure 50 : Processus biologiques (*Gene Ontology*, GO) des 13 protéines régulées en réponse à PEP-13 chez les racines de soja

annotation (axe des abscisses). Pour chaque GO, les protéines sont comptabilisées en nombre et pourcentage de gènes (axes des ordonnées). Enfin, pour un FC supérieur ou égal à 1,5 et une *p-value* inférieure à 0,05, la liste des protéines sous-exprimées en réponse à PEP-13 dans les AC-DC est représentée en bleu, et celle des protéines positivement régulées par PEP-13 dans les AC-DC est représentée en rouge. Graphique réalisé avec WEGO 2.0.

Si l'on compare les processus biologiques impliqués dans la réponse à PEP-13 des racines et des AC-DC, on observe des processus communément régulés (processus métaboliques et cellulaires, la synthèse et l'organisation des composants cellulaires, et la régulation), mais aussi des processus différemment régulés. En effet la localisation est un processus régulé positivement dans la racine en réponse à PEP-13, alors que dans les AC-DC, il est régulé négativement. La réponse aux stimuli est régulée uniquement négativement dans les racines, alors qu'elle est régulée positivement et négativement en réponse à PEP-13 dans les AC-DC. Cette analyse permet donc d'observer que les processus biologiques communs impliqués dans la réponse à PEP-13 chez les AC-DC et chez les racines sont différemment régulés.

Afin de compléter cette analyse des processus biologiques régulés en réponse à l'élicitation avec PEP-13, un second outil a été utilisé : le test d'*overrepresentation* de PANTHER. Pour cette analyse, les quatre listes de protéines régulées positivement (26 et 7) et négativement (5 et 6) respectivement dans les AC-DC et dans les racines en réponse à PEP-13 ont à nouveau été utilisées.

Le test d'*overrepresentation* a permis de mettre en évidence 15 voies statistiquement sur-représentées parmi les 26 protéines régulées négativement en réponse à PEP-13 dans les AC-DC (Tableau 12). Les autres listes n'ont pas permis de détecter de sur-représentation de voies métaboliques à l'aide de ce test. Parmi ces voies, on retrouve notamment le métabolisme de la phosphatidylcholine, la synthèse d'acide dicarboxylique et d'amides, la transcription ou encore la synthèse de peptides. Ces processus biologiques semblent donc régulés négativement en réponse à PEP-13.

Tableau 12 : Processus biologiques représentés par les 26 protéines régulées négativement chez les AC-DC de soja en réponse à PEP-13

GO biological process complete : liste des GO sur-représentés dans la liste ; *Reflist* : nombre total de protéines de *Glycine max* identifiés par ce GO dans leur annotation ; *Input* : nombre de protéines possédant ce GO dans la liste issue de notre étude ; *Expected* : seuil statistique à partir duquel une voie est considérée comme statistiquement sur-représentée ; *Fold enrichment* : enrichissement par rapport au seuil statistique ; FDR : *False discovery rate*, correspond à la *p-value* corrigée.

GO biological process complete	REFLIST (55865)	Input (26)	Expected	Fold Enrichment	FDR
phosphatidylcholine metabolic process (GO:0046470)	34	2	.02	> 100	5.48E-02
dicarboxylic acid biosynthetic process (GO:0043650)	38	2	.02	> 100	5.21E-02
amide biosynthetic process (GO:0043604)	1117	6	.52	11.54	2.24E-02
translation (GO:0006412)	1002	5	.47	10.72	4.76E-02
peptide biosynthetic process (GO:0043043)	1016	5	.47	10.57	4.52E-02
cellular amide metabolic process (GO:0043603)	1360	6	.63	9.48	3.37E-02
peptide metabolic process (GO:0006518)	1159	5	.54	9.27	5.00E-02
organonitrogen compound biosynthetic process (GO:1901566)	2299	8	1.07	7.48	2.82E-02
cellular nitrogen compound biosynthetic process (GO:0044271)	2248	7	1.05	6.69	4.87E-02
cellular biosynthetic process (GO:0044249)	4155	9	1.93	4.65	4.79E-02
organic substance biosynthetic process (GO:1901576)	4239	9	1.97	4.56	4.79E-02
biosynthetic process (GO:0009058)	4534	9	2.11	4.27	4.70E-02
cellular metabolic process (GO:0044237)	14436	16	6.72	2.38	5.08E-02
metabolic process (GO:0008152)	18161	18	8.45	2.13	5.24E-02
cellular process (GO:0009987)	21765	21	10.13	2.07	2.69E-02

I-2-3 Analyse détaillée des protéines différentiellement exprimées en réponse à PEP-13

Dans le but d'étudier plus en détail les protéines impliquées dans la réponse à PEP-13 des AC-DC et des racines, une analyse a été réalisée avec le logiciel Mapman. Pour cela, les listes des protéines différentiellement régulées chez les AC-DC et les racines associées au logarithme binaire de leur FC ont été utilisées. Le logiciel Mapman a ensuite assigné des *bins* correspondant à des GO (*cf* Matériels et méthodes II-2) aux protéines. Celles-ci ont ensuite été classées en fonction de leurs *bins* et assignées à un schéma *overview*.

Chez les AC-DC (Figure 51), 16 protéines (sur les 31 détectées au seuil statistique choisi) ont pu être annotées avec le système de *bins*. Pour la catégorie stress, une protéine régulée positivement a été détectée et semble donc positivement régulée en réponse à PEP-13. Les protéines assignées aux catégories *cell*, *proteins*, *development*, *hormones*, *enzymes*, *transport*, *other metabolisms* et *unknown/unclassified* semblent être régulées négativement en réponse à PEP-13. Enfin, pour la catégorie *specialized metabolism*, deux protéines ont été détectées, l'une régulée positivement, et l'autre négativement.

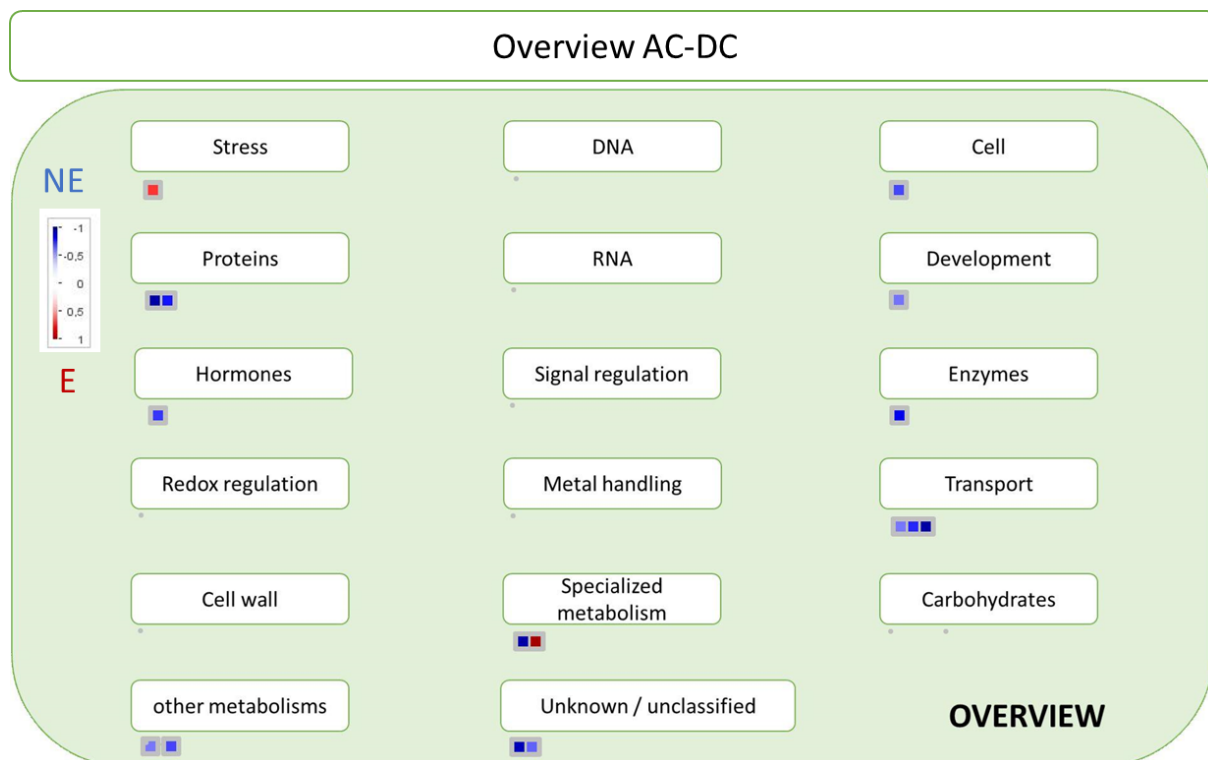


Figure 51 : Représentation générale (*overview*) des 16 protéines différentiellement exprimées chez les AC-DC (*root-Associated Cap-Derived Cells*) de soja entre les conditions élicitée avec PEP-13 (E) ou témoin non élicitée (NE)

Cette représentation permet d'observer les protéines régulées négativement ou positivement chez les AC-DC en réponse à PEP-13 en fonction des processus biologiques dans lesquels elles interviennent. Chaque protéine est représentée par un carré dont l'intensité de la couleur dépend du Log2 de son *fold change* (FC). Ainsi, les protéines sur-exprimées en réponse à PEP-13 sont représentées en dégradé de rouge, et les protéines sous-exprimées en réponse à PEP-13 sont représentées en dégradé de bleu. Graphique réalisé avec Mapman.

Pour les racines (Figure 52), 9 protéines (sur les 13 détectées au seuil statistique choisi) ont été annotées avec des *bins*. Les protéines détectées et assignées aux catégories *development*, *transport* et *unknown/unclassified* sont régulées positivement chez les racines en réponse à PEP-13. Les protéines détectées et assignées aux catégories *redox regulation* et *other metabolism* semblent elles régulées négativement en réponse à PEP-13. Enfin, pour la catégorie *proteins*, deux protéines régulées négativement et une protéine régulée positivement ont été détectées.

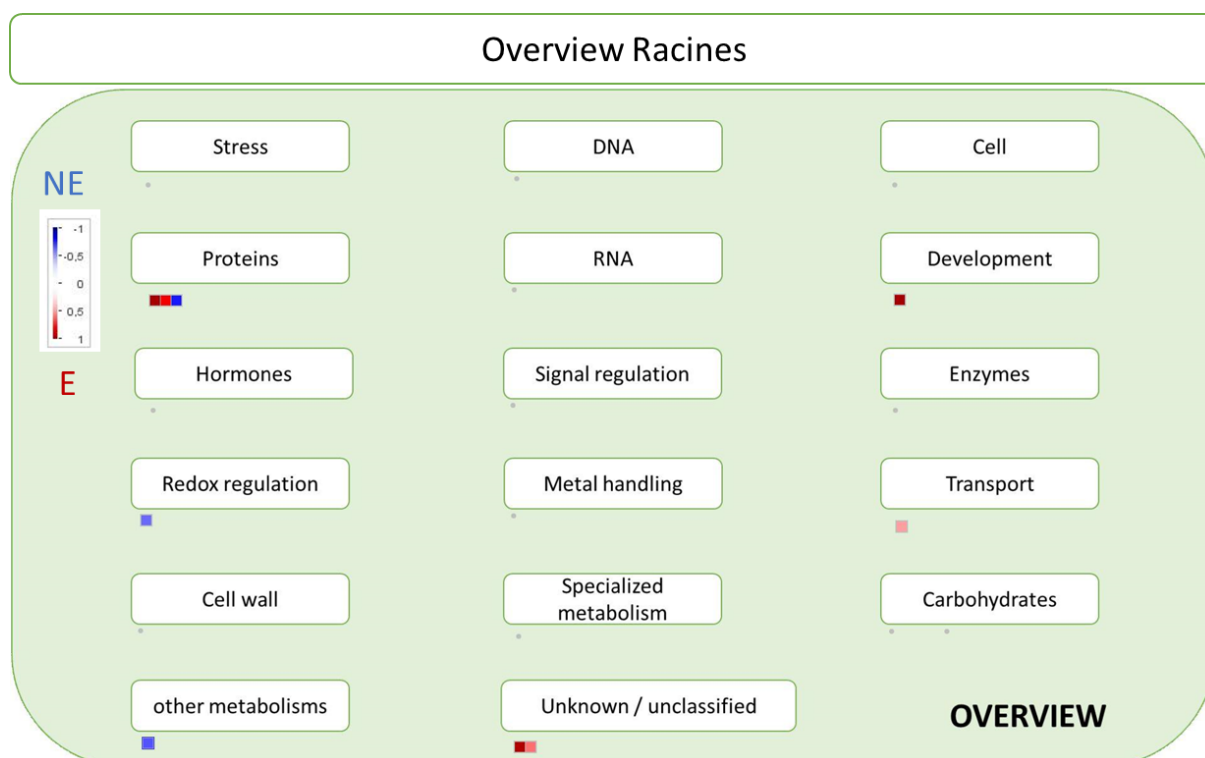


Figure 52 : Représentation générale (*overview*) des 9 protéines différemment exprimées chez les racines de soja entre les conditions élicitée avec PEP-13 (E) ou témoin non élicitée (NE)

Cette représentation permet d'observer les protéines régulées négativement ou positivement chez les racines en réponse à PEP-13 en fonction des processus biologiques dans lesquels elles interviennent. Chaque protéine est représentée par un carré dont l'intensité de la couleur dépend du Log2 de son *fold change* (FC). Ainsi, les protéines sur-exprimées en réponse à PEP-13 sont représentées en dégradé de rouge, et les protéines sous-exprimées en réponse à PEP-13 sont représentées en dégradé de bleu. Graphique réalisé avec Mapman.

Si l'on compare la réponse des AC-DC et des racines à PEP-13, on observe des processus biologiques régulés de façon similaire (*other metabolism*), ou de façon différente entre les deux zones (*development*, *transport*, *proteins* et *unknown/unclassified*). Enfin, certaines catégories semblent régulées spécifiquement dans la racine (*redox regulation*) ou dans les AC-DC (*hormones*, *stress*, *cell*, *enzymes*, *specialized metabolism*).

Pour la suite de cette étude, le détail des protéines est visualisé sous forme de tableau classé par catégories de *bins* (Tableaux 13 et 14), indiquant l'identifiant du transcrit, le nom de celui-ci ainsi que son FC (exprimé sous forme de logarithme binaire). Les détails concernant ces gènes les plus exprimés seront exposés dans la discussion de ce manuscrit.

Tableau 13 : Détail des 16 protéines différentiellement exprimées chez les AC-DC de soja en réponse à PEP-13, classées en fonction de leur catégorie de bins

Une couleur est attribuée au log2 du FC en fonction de sa valeur, plus celle-ci est élevée plus celui-ci apparaîtra en rouge foncé ; et plus elle est basse, plus elle apparaîtra en bleu foncé.

Stress		
ID	Protein name	Log(FC)
glyma05g38110.1	moderately similar to (276)AT4G11650 ATOSM34 (OSMOTIN 34)	0,7186282
Cell		
ID	Protein name	Log(FC)
glyma10g34310.1	Coatomer subunit alpha	-0,6876234
Proteins		
ID	Protein name	Log(FC)
glyma20g25120.2	nearly identical (1165)AT1G25350 OVA9 (OVULE ABORTION 9); glutamine-tRNA ligase	-0,9974677
glyma09g31760.1	nearly identical (1030)AT3G25800 PDF1 protein phosphatase type 2A regulator	-0,77406126
Development		
ID	Protein name	Log(FC)
glyma10g07410.1	weakly similar to (105)AT2G36640 ATECP63 (EMBRYONIC CELL PROTEIN 63)	-0,59820586
Hormones		
ID	Protein name	Log(FC)
glyma09g01110.1	Fe2OG dioxygenase domain-containing protein	-0,71554434
Enzymes		
ID	Protein name	Log(FC)
glyma05g04270.1	highly similar to (880)AT4G12420 SKU5 (skewed 5); copper ion binding	-0,84056777
Transport		
ID	Protein name	Log(FC)
glyma11g12120.1	V-type proton ATPase subunit a	-0,58558744
glyma15g01470.1	nearly identical (2116)AT1G15520 PDR12 (PLEIOTROPIC DRUG RESISTANCE 12)	-0,7439727
glyma13g29180.1	nearly identical (2382)AT2G34660 MRP2 (MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN 2)	-1,0259907
Specialized metabolism		
ID	Protein name	Log(FC)
glyma13g40140.2	nearly identical (1297)AT5G60600 GcpE (CHLOROPLAST BIOGENESIS 4)	-1,1086096
glyma18g38670.1	PKS_ER domain-containing protein	1,0465155
Other metabolisms		
ID	Protein name	Log(FC)
glyma20g38740.1	nearly identical (1107)AT1G50480 THFS (10-FORMYLTETRAHYDROFOLATE SYNTHETASE)	-0,5873188
Unknown / Unclassified		
ID	Protein name	Log(FC)
glyma03g35080.1	Uncharacterized protein	-0,96323246
glyma03g36420.1	Uncharacterized protein	-0,63235146

Tableau 14 : Détail des 9 protéines différentiellement exprimées chez les racines de soja en réponse à PEP-13, classées en fonction de leur catégorie de bins

Une couleur est attribuée au log2 du FC en fonction de sa valeur, plus celle-ci est élevée plus celui-ci apparaîtra en rouge foncé ; et plus elle est basse, plus elle apparaîtra en bleu foncé.

Proteins		
ID	Protein name	Log(FC)
glyma10g02060.1	Ribosomal protein L19	1,6191602
glyma19g39570.1	Ribosomal protein L19	0,8421028
glyma17g09280.2	Ribosomal_L18e/L15P domain-containing protein	-0,76640224
Development		
ID	Protein name	Log(FC)
glyma13g39890.1	moderately similar to (248)KIWEL_ACTCH no original description	1,484601
Hormones		
ID	Protein name	Log(FC)
glyma16g07870.1	Glutaredoxin domain-containing protein	-0,61719877
Transport		
ID	Protein name	Log(FC)
glyma19g35960.1	Cation_ATPase_N domain-containing protein	0,49477175
Other metabolisms		
ID	Protein name	Log(FC)
glyma15g05800.1	FabA domain-containing protein	-0,6574312
Unknown / Unclassified		
ID	Protein name	Log(FC)
glyma13g27810.1	highly similar to (950)AT3G23940 dehydratase family	0,9776553
glyma10g03840.1	H15 domain-containing protein	0,60190666

I-3 Comparaison des résultats obtenus entre l'analyse protéomique et transcriptomique

Deux analyses, l'une protéomique et l'autre transcriptomique, ont été réalisées sur les AC-DC et les racines de soja en conditions élicitée avec PEP-13 ou témoin non élicitée.

Afin d'observer si le transcriptome et le protéome de ces zones sont similaires, une rapide comparaison des identifiants des gènes correspondants aux transcrits (719 E et 1576 NE pour les AC-DC) et aux protéines (5 E et 25 NE pour les AC-DC) détectées comme différentiellement exprimés chez les AC-DC puis chez les racines lors de ces analyses a été réalisée avec Jvenn (Figure 53). Cette analyse a permis d'observer un seul gène commun pour lequel le transcrit et la protéine ont été détectés comme sous-exprimés en réponse à PEP-13 chez les AC-DC : GLYMA_13G326400. Ce gène correspond à une protéine peu annotée qui semble intervenir dans la synthèse des terpénoïdes.

Enfin, si l'on compare les différentes voies identifiées comme étant impliquées dans la réponse à PEP-13 des AC-DC et des racines grâce à la méthode de transcriptomique et de protéomique (Tableau 15 ; cf. I-1-2-2, figures 38 et 39 et I-2-2, figures 49 et 50) on observe des voies régulées de façon similaire :

- chez les AC-DC : processus cellulaires et métaboliques, organisation et synthèse des composés cellulaires, régulation biologique, réponse aux stimuli,
- chez les racines : régulation des processus biologiques, processus cellulaires et métaboliques, organisation et synthèse des composés cellulaires.

On observe également des voies qui sont détectées comme sur et sous-exprimées lors de l'analyse transcriptomique et uniquement détectées dans l'une de ces deux conditions pour l'analyse protéomique (Tableau 15) :

- chez les AC-DC : la régulation des processus biologiques (E), la localisation (NE), les processus multicellulaires (NE), le développement (NE),
- chez les racines : la localisation (E), la réponse aux stimuli (NE).



Figure 53 : Diagrammes de Venn comparant les gènes et protéines différentiellement exprimés en réponse à PEP-13 chez les AC-DC et chez les racines

Sur ce graphique les listes des identifiants des gènes et protéines différentiellement exprimées (FC supérieur ou égal à 1,5 et *p-value* inférieure à 0,05) chez les AC-DC et chez les racines sont représentées par un cercle. L'intersection entre deux cercles correspond aux identifiants retrouvés dans les deux listes. Ainsi, il y a un gène chez les AC-DC pour lequel le transcrit et la protéine correspondante ont été communément détectés comme sous-exprimés avec ces deux méthodes. Pour les racines, aucun identifiant n'est commun aux deux méthodes. La condition élicitee avec PEP-13 est représentée en rose et orange, la condition contrôle est représentée en bleu. p : protéomique ; t : transcriptomique. Graphique réalisé avec Jvenn

Tableau 15 : Comparaison des processus biologiques pour les analyses protéomique et transcriptomique chez les AC-DC et les racines de soja impliqués dans la réponse à PEP-13 détectés lors de l'analyse WEGO 2.0

E : sur-exprimé en condition élicitée ; NE : sous-exprimé en réponse à PEP-13 ; E / NE : voie à la fois sur-exprimée et sous-exprimée en réponse à PEP-13

Processus biologique	AC-DC		Racines	
	Transcriptomique	Protéomique	Transcriptomique	Protéomique
Régulation des processus biologiques	E / NE	E	E / NE	E / NE
Régulation biologique	E / NE	E / NE	E / NE	E / NE
Régulation négative	E / NE		E	NE
Régulation positive	E / NE		E	E
Processus cellulaire	E / NE	E / NE	E / NE	E / NE
Processus métabolique	E / NE	E / NE	E / NE	E / NE
organisation et synthèse des composés cellulaires	E / NE	E / NE	E / NE	E / NE
Localisation	E / NE	NE	E / NE	E
Réponse aux stimuli	E / NE	E / NE	E / NE	NE
Processus multicellulaires	E / NE	NE	E	
Processus multi-organismes	E / NE		E	
Développement	E / NE	NE	E	
Signal	E / NE		E / NE	
Système immunitaire	E / NE			
Croissance	NE			
Reproduction	E / NE		E	
Processus reproductifs	E / NE		E	
Adhésion	E			
Utilisation du carbone	E			

L'analyse protéomique a permis de détecter moins de voies régulées en réponse à PEP-13, ce qui est lié au fait que la quantité de protéines détectées comme différentiellement exprimées en protéomique (31 chez les AC-DC, 13 chez les racines) est nettement inférieure à la quantité de transcrits détectés comme différentiellement exprimés en transcriptomique (2295 chez les AC-DC, 553 chez les racines).

Cependant, les voies détectées grâce aux données de protéomique semblent correspondre à celles détectées en transcriptomique. En effet, à l'exception de la régulation négative chez les racines pour laquelle les données de protéomique ne correspondent pas aux données de transcriptomique (cette voie est sur-exprimée en transcriptomique et sous-exprimée en protéomique), les voies détectées comme sur- et sous-exprimées en transcriptomique ont été détectées comme sur-exprimées et/ou sous-exprimées en protéomique.

II. Analyse génomique de l'ADNex du RET du soja

L'objectif de cette étude est d'apporter une première réponse à la question de l'origine et des séquences concernées de l'ADNex présent dans le RET et de sa possible implication dans la réponse immunitaire. Pour cela, des méthodes de récolte de l'ADNex en milieu stérile et de purification de celui-ci ont été mises au point. Suite à cela, un séquençage de l'ADNex du RET du soja en condition élicitée avec PEP-13 et en condition témoin a été réalisé. Six échantillons, séparés en trois réplicats, nommés réplicats 7, 8 et 9 en condition élicitée avec PEP-13 (E) et témoin non élicitée (NE) ont ainsi été obtenus (*cf* Matériel et méthode, IV). Ces échantillons et conditions ont ensuite été comparés.

II-1 Analyse de l'alignement des séquences

Pour cette étude, le génome complet (chromosomes et organites) du soja a tout d'abord été divisé en zones de 1000 paires de bases sur lesquelles les séquences (*reads*) d'ADNex obtenues ont été alignées. Le positionnement et la couverture des *reads* sur les 20 chromosomes, le plaste et la mitochondrie ont été analysés.

II-1-1 Visualisation globale des alignements de séquences

Les réplicats 7, 8 et 9 ont tout d'abord été analysés pour la condition témoin (NE; non élicité).

Pour le réplicat 7 (Tableau 16), on observe un alignement total de 1 305 138 *reads* sur l'ensemble du génome. Ces *reads* sont répartis entre les 20 chromosomes de l'ADN nucléaire (ADNchr) du soja sur lequel 1 227 257 *reads* se sont alignés, l'ADN plastidial (ADNp) et l'ADN mitochondrial (ADNm) du soja sur lesquels 77 881 *reads* se sont alignés. Ces chiffres permettent donc de remarquer qu'il y a un plus grand nombre de *reads* alignés sur l'ADNchr par rapport à l'ADN provenant des plastides et mitochondries.

Si l'on regarde plus en détail le nombre de *reads* alignés sur ces séquences, on observe une moyenne d'environ 60 000 *reads* qui se sont alignés sur chaque chromosome et 58 871 *reads* qui se sont alignés sur l'ADNm. En revanche, seulement 19 010 *reads* se sont alignés sur l'ADNp. Le nombre de *reads* alignés sur l'ADNm est donc assez proche du nombre moyen de *reads* alignés sur l'ADNchr. Cependant, ces chiffres ne prennent pas en compte la taille de l'ADN chromosomique, de l'ADN d'origine plastidial ou mitochondrial. En effet pour chacun

des chromosomes, on observe une taille de plusieurs dizaines de mégabases (de 34,77 Mb pour le chromosome 11 à 58,02 Mb pour le chromosome 18) avec un total de 949,21 Mb pour l'ensemble des chromosomes. Par contre, l'ADN des organites ne représente que 0,4 Mb pour la mitochondrie et 0,15 Mb pour le plaste. Les tailles de l'ADN chromosomique et des ADN issus des organites étant très éloignées, un ratio du nombre de *counts* alignés divisé par la taille de l'ADN concerné (chromosomique, plastidial ou mitochondrial) en mégabases a été calculé afin de pouvoir comparer les différentes couvertures entre elles. Pour les chromosomes, on obtient donc une somme de 25 986,27 *counts*/Mb pour l'ADNchr total. Ce ratio atteint 147 177,50 *counts*/Mb pour l'ADNm et 126 733,33 *counts*/Mb pour l'ADNp, soit un total de 273 910,83 *counts*/Mb pour l'ensemble de l'ADN des organites. On a donc un ratio moyen environ cent fois plus important pour l'ADN mitochondrial et l'ADN plastidial que pour l'ADN chromosomique. Ces différences de ratio semblent donc mettre en évidence une plus grande couverture de l'ADNex sur l'ADNm et l'ADNp que sur l'ADNchr.

Pour les réplicats 8 et 9 NE, les tableaux sont présentés en annexe III, partie II-1. Les mêmes résultats que pour le réplikat 7 NE ont été observés, c'est-à-dire une différence de couverture entre l'ADNchr et celui de l'ADN des organites.

Tableau 16 : Analyse du nombre de séquences de l'ADNex alignées (*counts*) sur le génome complet (chromosomes, mitochondrie et plaste) du soja pour le réplicat 7 NE (ADNex en condition non élicitée)

Pour les chromosomes, plastes et mitochondrie, un ratio *count*/Mb a été calculé. Ce ratio correspond au nombre de *counts* alignés sur le génome, divisé par sa taille en mégabases (Mb). Enfin, les trois dernières lignes de ce tableau correspondent à la somme totale des counts, de la taille ou des ratios pour l'ensemble des chromosomes (ligne ADNchr), la mitochondrie et le plaste (ligne ADNm et ADNp) et enfin pour le génome complet (ligne ADN total). ADNchr = ADN chromosomique ; ADNm = ADN mitochondrial ; ADNp = ADN plastidial.

		7 NE		
	Chromosome / Organite	Somme des counts	Taille (Mb)	Count/Mb
ADNchr	GM01	72 571	56,83	1 276,98
	GM02	68 079	48,58	1 401,38
	GM03	57 113	45,78	1 247,55
	GM04	64 095	52,39	1 223,42
	GM05	53 068	42,23	1 256,64
	GM06	65 841	51,42	1 280,46
	GM07	56 496	44,63	1 265,87
	GM08	61 344	47,84	1 282,27
	GM09	62 916	50,19	1 253,56
	GM10	69 121	51,59	1 339,81
	GM11	51 386	34,77	1 477,88
	GM12	53 550	40,09	1 335,74
	GM13	59 408	45,87	1 295,14
	GM14	64 346	49,04	1 312,11
	GM15	64 217	51,76	1 240,67
	GM16	48 347	37,89	1 275,98
	GM17	52 893	41,64	1 270,24
	GM18	70 870	58,02	1 221,48
	GM19	66 738	50,75	1 315,03
	GM20	64 858	47,90	1 354,03
Organites	Mitochondrie	58 871	0,40	147 177,50
	Plastes	19 010	0,15	126 733,33
Total	ADNchr	1 227 257	949,21	25 926,27
	Mitochondrie et plastes	77 881	0,55	273 910,83
	ADN total (ADNchr + organites)	1 305 138	949,76	299 837,10

L'alignement des séquences d'ADNex issues des réplicats 7, 8 et 9 a ensuite été analysé pour la condition élicitée (E).

Le réplicat 7E (Tableau 17) a permis d'obtenir 1 209 179 *reads* alignés sur le génome complet. Parmi ces *reads*, 1 166 892 s'alignent sur l'ADNchr et 42 287 s'alignent sur l'ADNm et l'ADNp. Pour chacun des chromosomes, le nombre de *reads* alignés varie entre 68 984 pour le chromosome 1 et 45 569 pour le chromosome 16. Pour les organites, ce nombre est de 29 654 pour l'ADNm et 12 633 pour l'ADNp. Le total des *counts*/Mb de l'ADNchr est de 24 662,33 *counts*/Mb, contre 74 135 *counts*/Mb pour l'ADNm et 84 220 *counts*/Mb pour l'ADNp.

Comme pour les répliquats de la condition non élicitée, les répliquats 8 et 9 E sont présentés en annexe de ce manuscrit (Annexe III, partie II-1). Les mêmes résultats que pour le répliquat 7 E ont été observés, c'est-à-dire une différence de couverture entre l'ADNchr et celui de l'ADN des organites.

Tableau 17 : Analyse du nombre de séquences d'ADNex alignées (counts) sur le génome complet (chromosomes, mitochondrie et plaste) du soja pour le répliquat 7E (ADNex en condition élicitée)

Pour les chromosomes, plastes et mitochondrie, un ratio *count*/Mb a été calculé. Ce ratio correspond au nombre de *counts* alignés sur le génome, divisé par sa taille en mégabases (Mb). Enfin, les trois dernières lignes de ce tableau correspondent à la somme totale des counts, de la taille ou des ratios pour l'ensemble des chromosomes (ligne ADNchr), la mitochondrie et le plaste (ligne ADNm et ADNp) et enfin pour le génome complet (ligne ADN total). ADNchr = ADN chromosomique ; ADNm = ADN mitochondrial ; ADNp = ADN plastidial.

7 E				
Chromosome / Organite	Somme des counts	Taille (Mb)	Count/Mb	
ADNchr	GM01	68 984	56,83	1 213,87
	GM02	64 160	48,58	1 320,71
	GM03	64 160	45,78	1 401,49
	GM04	59 950	52,39	1 144,30
	GM05	51 362	42,23	1 216,24
	GM06	61 937	51,42	1 204,53
	GM07	53 686	44,63	1 202,91
	GM08	59 068	47,84	1 234,70
	GM09	59 683	50,19	1 189,14
	GM10	64 619	51,59	1 252,55
	GM11	48 085	34,77	1 382,95
	GM12	50 117	40,09	1 250,11
	GM13	56 248	45,87	1 226,25
	GM14	60 365	49,04	1 230,93
	GM15	59 668	51,76	1 152,78
	GM16	45 569	37,89	1 202,67
	GM17	49 782	41,64	1 195,53
	GM18	66 287	58,02	1 142,49
	GM19	62 303	50,75	1 227,65
	GM20	60 859	47,90	1 270,54
Organites	Mitochondrie	29 654	0,40	74 135,00
	Plastes	12 633	0,15	84 220,00
Total	ADNchr	1 166 892	949,21	24 662,33
	Mitochondrie et plastes	42 287	0,55	158 355,00
	ADN total (ADNchr + organites)	1 209 179	949,76	183 017,33

Les différents répliquats de séquençage de l'ADNex permettent donc d'observer un ratio des *counts* /Mb plus important pour l'ADN des organites que pour l'ADNchr. Cette différence de ratio semble indiquer une différence de couverture de l'ADNex sur l'ADN chromosomique comparés à l'ADN des organites. Il y aurait donc une meilleure couverture sur la mitochondrie et le plaste en conditions élicitée avec PEP-13 ou témoin non élicitée. Afin de vérifier cette observation, nous avons utilisé le logiciel IGV (*Integrative genomics viewer*)

pour visualiser le positionnement des *reads* séquencés d'ADNex sur les chromosomes, la mitochondrie et le plaste.

II-1-2 Visualisation de l'alignement des séquences d'ADNex avec IGV

Le logiciel IGV permet d'observer, pour une position donnée sur l'ensemble du génome (Figure 54, A), l'alignement des *reads* pour chacun des échantillons (Figure 54, B) ainsi que l'éventuelle présence d'une séquence codante identifiée ayant été annotée à cette position (Figure 54, C). Ainsi, pour les figures 54, 55 et 56, on observe une fenêtre IGV contenant dans une première ligne un repère (Figure 54, A) permettant de visualiser la position observée, six lignes qui correspondent chacune à un échantillon (Figure 54, B, réplicats 7, 8, et 9 en condition élicitée ou non élicitée) et une dernière ligne correspondant à l'annotation de la séquence observée sur laquelle sont inscrits les identifiants des gènes observés (Figure 54, C). Pour chacune des lignes "échantillon" (B), on retrouve une partie supérieure indiquant la quantité de *reads* alignés tout au long de la séquence observée, et une partie inférieure sur laquelle les *reads* alignés à cette position sont représentés par des traits.

Dans un premier temps, la couverture de l'ADNex séquencé sur chacun des vingt chromosomes du soja a été observée et des captures d'écran ont été réalisées pour chacun des chromosomes. La figure 54 est une capture d'écran représentative de ce qui a été observé pour les vingt chromosomes (les captures d'écran correspondant aux 19 autres chromosomes sont présentées en annexe III, partie II-2). Cette capture d'écran sélectionnée est celle de la position 7 742 099 à 7 817 639 du chromosome 18. On remarque que les *reads* d'ADNex sont alignés de façon éparse et peu abondante tout au long de la séquence pour chacun des réplicats. De plus, les *reads* semblent répartis aléatoirement entre les zones codantes ou non. Concernant les zones codantes, on remarque, sur cette portion du chromosome 18, que trois zones codantes apparaissent dont deux correspondant aux gènes GLYMA18G080700 et GLYMA18G081000 et l'une non-annotée.

Enfin, lors de l'observation des différents chromosomes, des zones présentant une augmentation d'alignement de *reads* ont été observées.

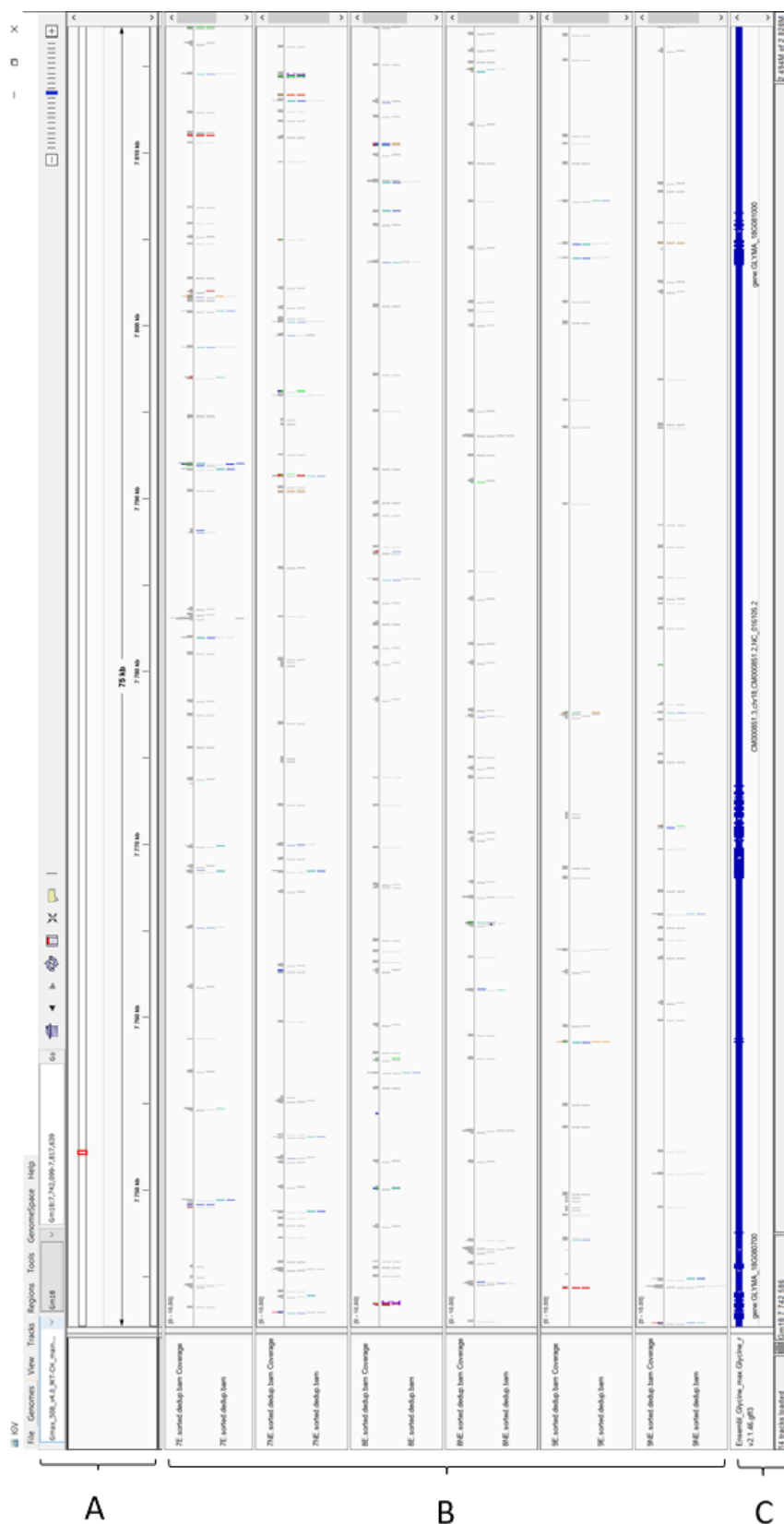


Figure 54 : Visualisation IGV de l’alignement des *reads* des différents échantillons d’ADNex de soja sur le chromosome GM18 en position 7 742 099 – 7 817 639
 A : Séquence observée et repère indiquant la position sur cette séquence ; B : Alignement des *reads* des différents échantillons sur cette séquence ; C : Annotation de cette séquence.



Figure 55 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex de soja sur l'ADN de la mitochondrie en position 1 – 66 781

A : Séquence observée et repère indiquant la position sur cette séquence ; B : Alignement des *reads* des différents échantillons sur cette séquence ; C : Annotation de cette séquence.

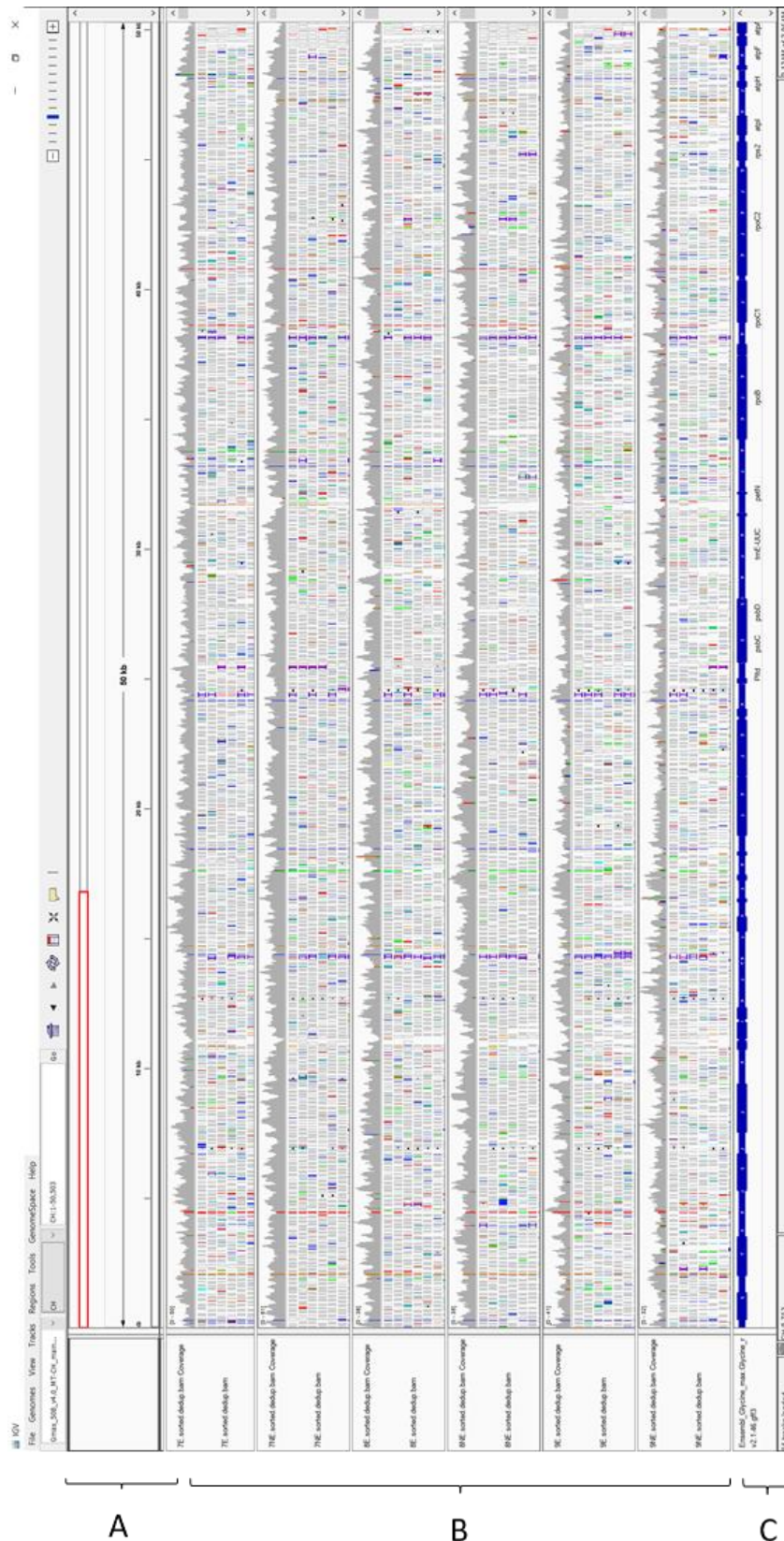


Figure 56 : Visualisation IGV de l’alignement des *reads* des différents échantillon d’ADNex de soja sur l’ADN des plastes en position 1 - 50 503

A : Séquence observée et repère indiquant la position sur cette séquence ; B : Alignement des *reads* des différents échantillons sur cette séquence ; C : Annotation de cette séquence.

La figure 55 représente une capture d'écran réalisée de la couverture de l'ADNex sur l'ADN de la mitochondrie du soja, en position 1 à 66 781, et représentative de l'ensemble de l'ADN de cet organe. On observe un grand nombre de *reads* alignés tout au long de cet ADN, de façon homogène et concernant aussi bien les zones codantes que non codantes. Cette observation confirme qu'il y a donc une bonne couverture de l'ADNex séquencé sur l'ADN mitochondrial.

La figure 56 représente une capture d'écran réalisée de la couverture de l'ADNex sur l'ADN plastidial, en position 1 à 50 503. Un grand nombre de *reads* sont alignés tout au long de la séquence observée, de plus aucune différence n'est observée entre les zones codantes et non codantes. Cette observation confirme qu'il y a donc une bonne couverture de l'ADNex séquencé sur l'ADN plastidial.

Si l'on compare les observations réalisées sur l'ADN chromosomique et celles réalisées sur l'ADN mitochondrial et chloroplastique, il apparaît donc que la couverture de l'ADNchr est beaucoup moins importante que celles de l'ADNm et de l'ADNp. Afin de chiffrer ces observations, un calcul du taux de couverture moyen (nombre de *reads* multipliés par leur longueur et divisé par la taille de la séquence) a été réalisé pour l'ADNchr et l'ADN des organites (ADNm + ADNp) (Tableau 18). Ce calcul permet tout d'abord de confirmer l'observation de la différence de couverture entre l'ADNchr et l'ADNm et ADNp. En effet pour les organites le taux moyen de couverture est de 12,62 X en condition élicitée et de 17,29 X en condition non élicitée. En revanche, pour l'ADN chromosomique, le taux moyen de couverture est de 0,22 X en condition élicitée et 0,25 X en condition non élicitée. Il y a donc bien une différence de couverture en fonction de l'origine (chromosomique ou non) de l'ADN. De plus, ce calcul permet de constater qu'il n'y a pas de différence majeure en termes de taux de couverture en fonction de l'élicitation avec PEP-13 pour l'ADN chromosomique et une couverture légèrement plus importante pour l'ADNex aligné sur les ADN des organites en condition témoin non élicitée.

Tableau 18 : Calcul du taux de couverture moyen de l'ADNex extrait en condition élicitée et non élicitée (Nombre de *reads* x longueur des *reads* / taille de la séquence) réalisé avec Excel

Chromosome / organe	ADNex élicité		ADNex non élicité	
	Moyenne des counts	Taux de couverture	Moyenne des counts	Taux de couverture
ADNchr	1 055 585,33	0,22	1 177 342,00	0,25
Mitochondrie et plastes	34 702,67	12,62	47 543,00	17,29
ADN total (ADNchr + organites)	1 090 288,00	0,23	1 224 885,00	0,26

Afin d'étudier plus en détail les différences entre l'ADNex issu de plantules élicitées et non élicitées, une analyse différentielle des échantillons a été réalisée à partir des *counts* obtenus par zones du génome de 1 000 pb.

II-2 Analyse différentielle de l'alignement de séquences d'ADNex suite à l'élicitation

II-2-1 Analyse différentielle sur la totalité du génome *via* le logiciel DESeq2

Une analyse en composante principale de la variance entre les échantillons a été réalisée grâce au logiciel DESeq2. Pour cette analyse, les données d'alignement des *counts* sur les zones de 1 000 pb du génome ont été utilisées. De plus, ces données ont été divisées manuellement sur Excel en trois fichiers en fonction de l'origine de l'ADN : un fichier pour l'ADNchr, un second pour l'ADNp et un dernier pour l'ADNm. Ces trois fichiers ont donc pu être analysés séparément. Pour l'ADN mitochondrial (Figure 57), on observe donc que les deux composantes principales expliquent 53% de la variance entre les échantillons. De plus, les réplicats 7 et 8 conditions élicitée et non élicitée semblent très proches. Le réplicat 9 semble légèrement à l'écart des autres. Ce graphique montre donc qu'il ne semble pas y avoir de différences majeures entre les conditions élicitée et non élicitée pour l'ADNex provenant de l'ADN mitochondrial.

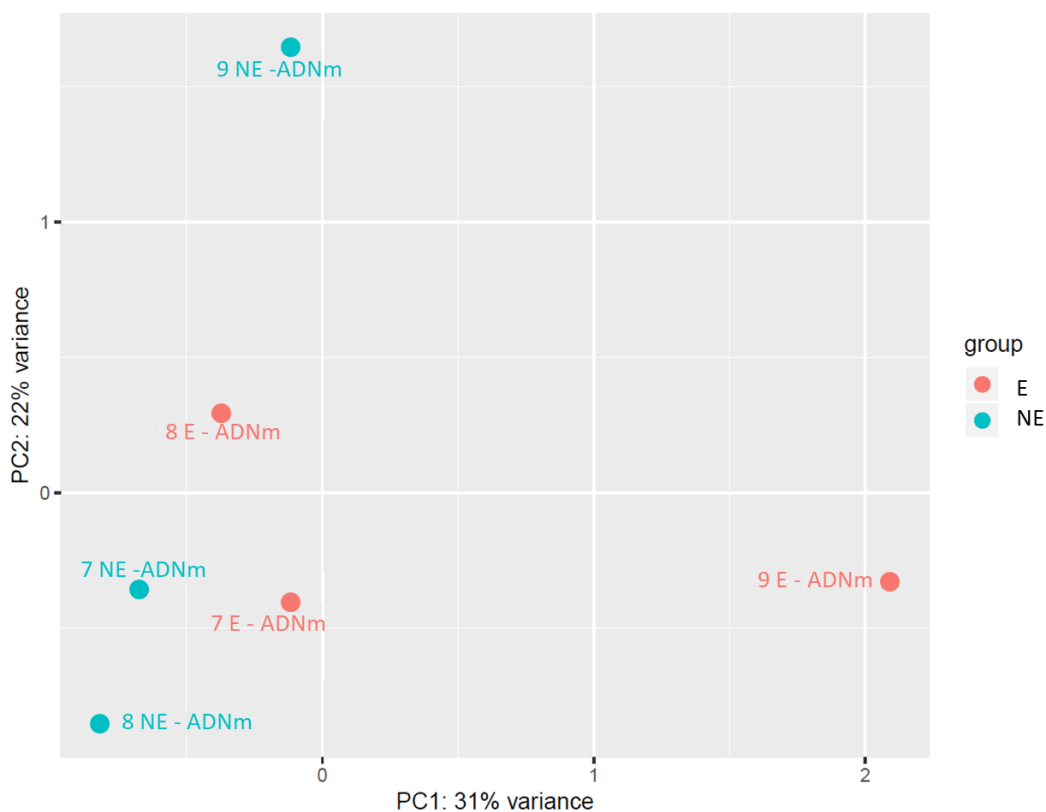


Figure 57 : Analyse en composante principale de la variance entre l'alignement des *reads* de l'ADNex de soja des différents échantillons sur l'ADN mitochondrial

Sur cette figure on observe deux composantes principales (PC1 et PC2) expliquant respectivement 31% et 22% de la variance entre les échantillons, soit au total 53% de cette variance. Chaque échantillon est représenté par un point bleu ou rouge en fonction de sa condition élicitée (E, rouge) ou non élicitée (NE, bleu).

Pour l'ADN plastidial (Figure 58), les composantes principales permettent d'expliquer 56% de la variance. Le réplicat 9 (E et NE) semble (comme pour la mitochondrie) légèrement à l'écart des deux autres qui sont très proches entre eux. Cette analyse montre donc qu'il ne semble pas y avoir de différences entre l'ADNex provenant des plastes en condition élicitée et en condition non élicitée.

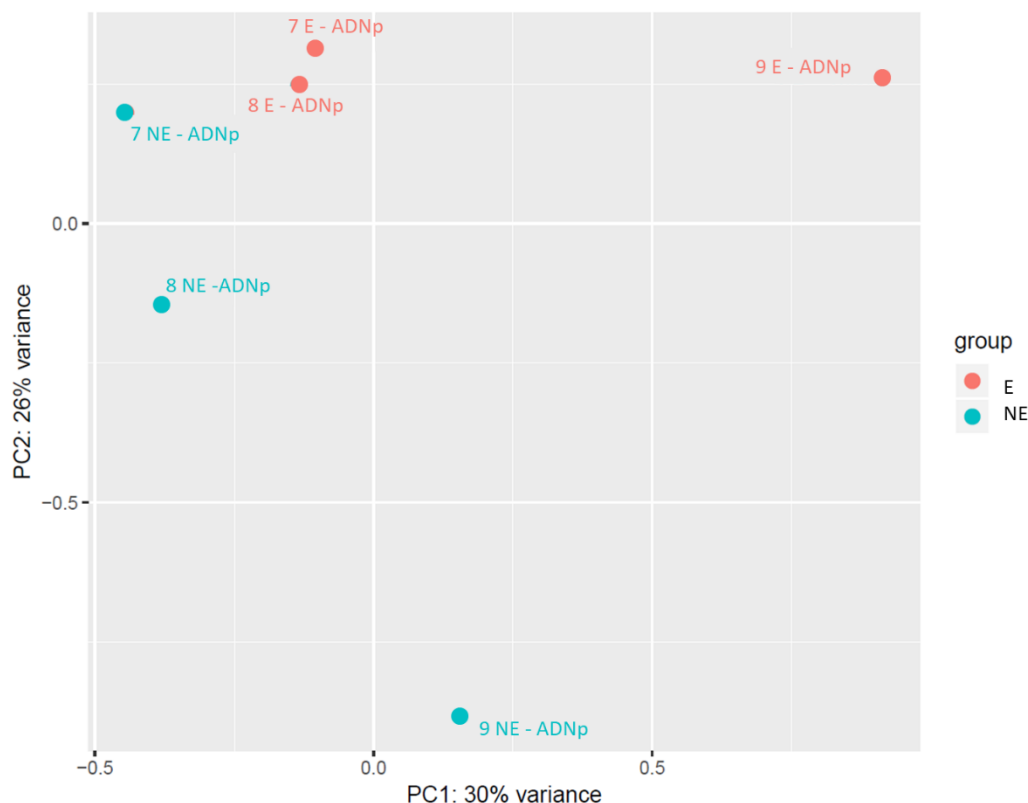


Figure 58 : Analyse en composante principale de la variance entre l'alignement des *reads* de l'ADNex de soja des différents échantillons sur l'ADN plastidial

Sur cette figure on observe deux composantes principales (PC1 et PC2) expliquant respectivement 30% et 26% de la variance entre les échantillons, soit au total 56% de cette variance. Chaque échantillon est représenté par un point bleu ou rouge en fonction de sa condition élicitée (E, rouge) ou non élicitée (NE, bleu).

Enfin, pour l'ADN chromosomique (Figure 59), l'analyse en composante principale permet d'expliquer 53% de la variance des échantillons. Les trois réplicats en condition non élicitée (7NE, 8NE et 9NE) ainsi que le réplicat 9 (condition élicitée) semblent être très proches les uns des autres. La seconde composante (PC2) semble permettre d'éloigner de façon assez importante les réplicats 8 E et 7 E des autres réplicats. Si l'on exclue le réplicat 9, il semble donc qu'il y ait une différence plus marquée entre les conditions élicitée et non élicitée pour l'ADNchr par rapport à l'ADNm et l'ADNp.

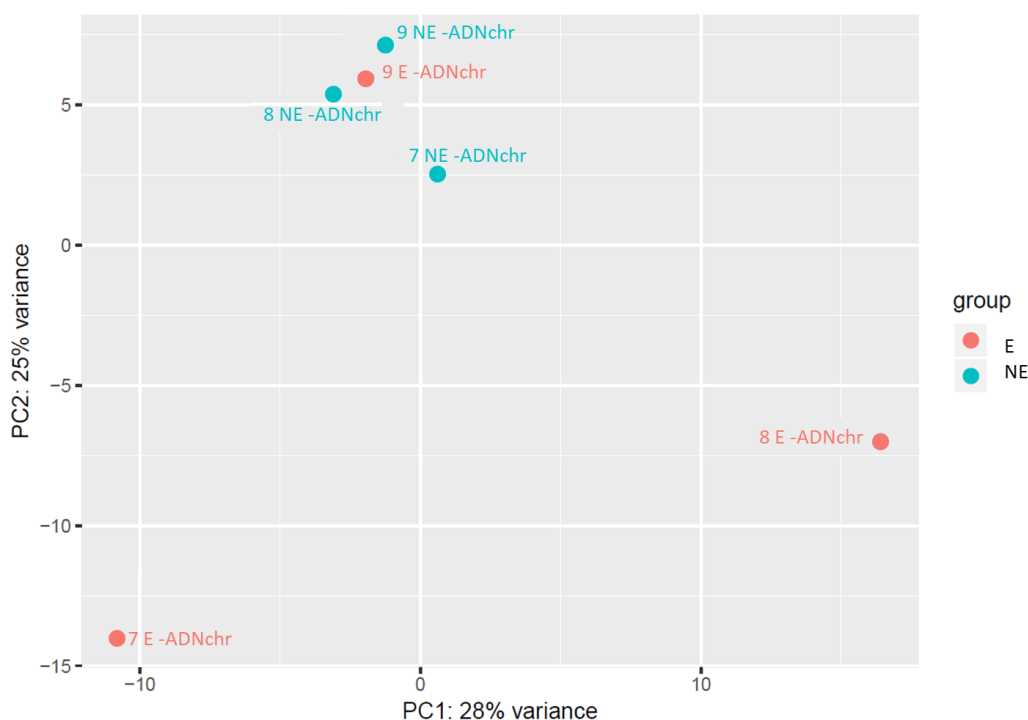


Figure 59 : Analyse en composante principale de la variance entre l'alignement des *reads* d'ADNex de soja des différents échantillons sur l'ADN chromosomique

Sur cette figure on observe deux composantes principales (PC1 et PC2) expliquant respectivement 28% et 25% de la variance entre les échantillons, soit au total 53% de cette variance. Chaque échantillon est représenté par un point bleu ou rouge en fonction de sa condition élicitée (E, rouge) ou non élicitée (NE, bleu).

Afin d'observer plus en détail les éventuelles différences entre les conditions élicitée et non élicitée en fonction de l'origine de l'ADNex, des *volcanoplots* ont été réalisés avec Excel à partir des données d'analyse différentielle de l'alignement des *reads* sur les zones de 1 000 pb obtenues précédemment avec DESeq2.

Pour l'ADN mitochondrial (Figure 60), on observe tout d'abord que les points correspondants chacun à une zone de l'ADNm de 1 000 pb semblent équitablement répartis entre ceux correspondant aux *reads* alignés en condition élicitée et ceux correspondant aux *reads* alignés en condition non élicitée. De plus, peu de points semblent dépasser le seuil statistique correspondant à une *p-value* de 0,05 et parmi ceux-ci, aucun ne dépasse un Log2 de FC de 2. Il n'y a donc pas de différence statistique entre les *reads* en conditions élicitée et non élicitée pour l'ADNex aligné sur l'ADN mitochondrial.

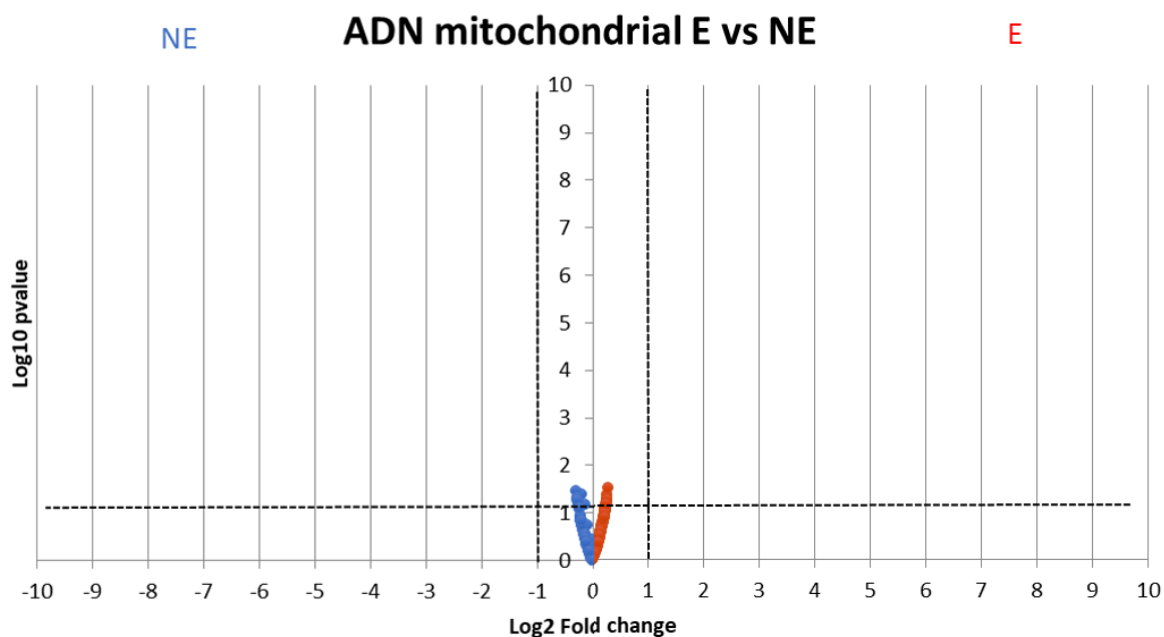


Figure 60 : Analyse différentielle de l'alignement des *reads* d'ADNex de soja sur l'ADN mitochondrial

Sur ce graphique, chaque point représente une zone de 1 000 pb découpée dans la séquence de l'ADN mitochondrial du soja. Le nombre de *reads* alignés sur chaque position en condition élicitée et non élicitée pour les trois réplicats a été utilisé ici afin d'obtenir une valeur de *fold change* (FC) ainsi qu'une *p-value* associée. On observe donc ici le Log2 du FC de chacun des points en fonction du Log10 de leur *p-value*. Les zones présentant un plus fort alignement de *reads* en condition élicitée sont représentées en rouge, et les zones dont l'alignement des *reads* est plus important en condition non élicitée sont représentées en bleu. Le seuil statistique correspondant à une *p-value* de 0,05 est représenté par une ligne horizontale en pointillé et le seuil de FC minimal de 2 est représenté par les lignes verticales en pointillé.

Pour l'ADN plastidial (Figure 61), les mêmes observations que pour l'ADN mitochondrial sont réalisées. On retrouve à nouveau des points qui semblent équitablement répartis entre les conditions élicitée et non élicitée. Peu d'entre eux ont une *p-value* significative (inférieure à 0,05), et aucun d'entre eux ne dépasse un FC de 2. Il n'y a donc pas de différence significative entre l'alignement des *reads* d'ADNex en condition élicitée et non élicitée sur l'ADN plastidial.

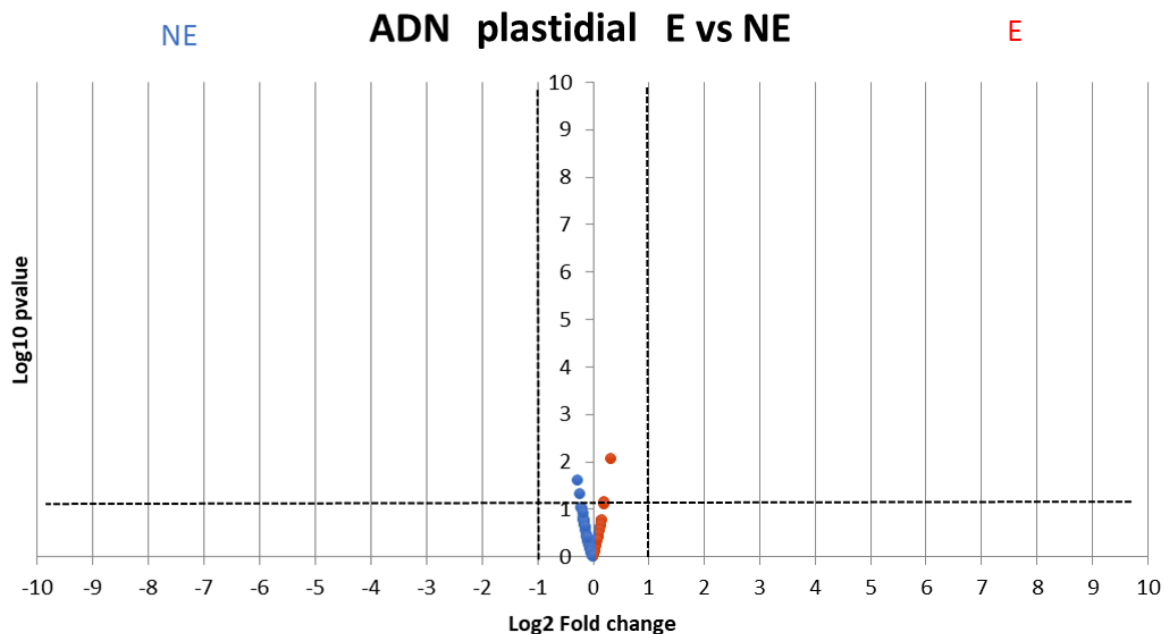


Figure 61 : Analyse différentielle de l'alignement des *reads* d'ADNex de soja sur l'ADN plastidial

Sur ce graphique, chaque point représente une zone de 1 000 pb découpée dans la séquence de l'ADN plastidial du soja. Le nombre de *reads* alignés sur chaque position en condition élicité et non élicité pour les trois réplicats a été utilisé ici afin d'obtenir une valeur de *fold change* (FC) ainsi qu'une *p-value* associée. On observe donc ici le Log2 du FC de chacun des points en fonction du Log10 de leur *p-value*. Les zones présentant un plus fort alignement de *reads* en condition élicitée sont représentées en rouge, et les zones dont l'alignement des *reads* est plus important en condition non élicitée sont représentées en bleu. Le seuil statistique correspondant à une *p-value* de 0,05 est représenté par une ligne horizontale en pointillé et le seuil de FC minimal de 2 est représenté par les lignes verticales en pointillé.

Enfin, pour l'ADN chromosomique du soja (Figure 62), on observe tout d'abord que les points correspondants chacun à une zone de l'ADNchr de 1 000 pb ne semblent pas équitablement répartis. En effet il semble qu'il y ait plus de points bleus, c'est-à-dire un plus grand nombre de *reads* alignés sur des zones de 1 000 pb en condition non élicitée. Un nombre important de points semble dépasser le seuil statistique de la *p-value* inférieure à 0,05. De plus, la majorité de ces points semble dépasser un FC de 2. Cette analyse indique donc qu'il y aurait des différences dans l'alignement des *reads* d'ADNex sur l'ADN chromosomique en fonction de la condition élicitée comparée à la condition non élicitée.

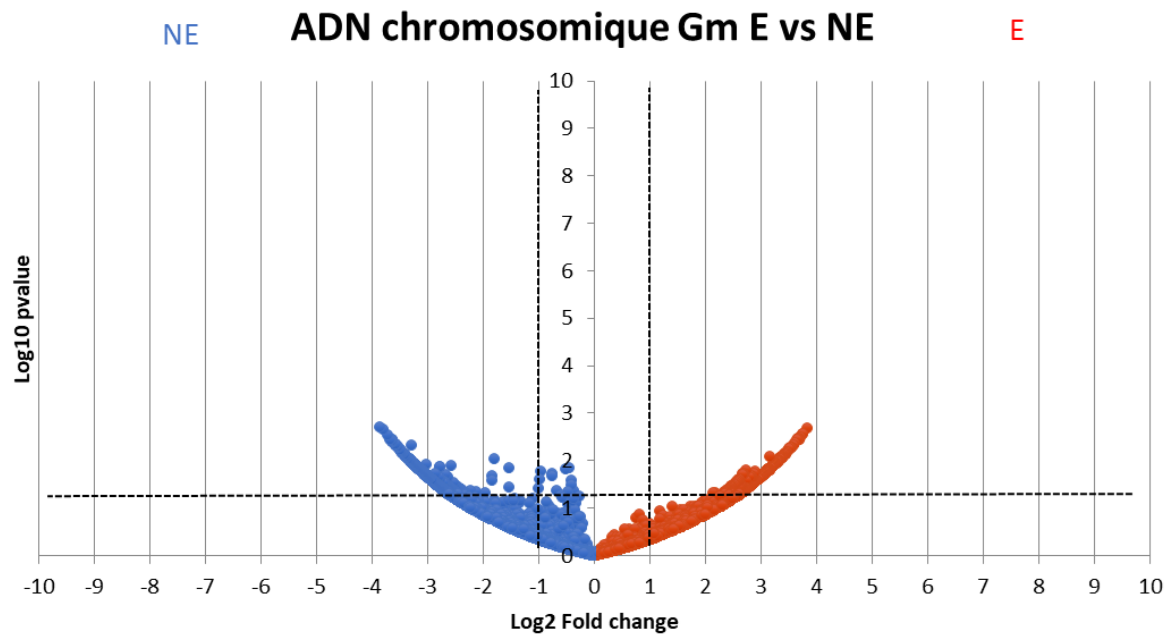


Figure 62 : Analyse différentielle de l'alignement des *reads* d'ADNex de soja sur l'ADN chromosomique

Sur ce graphique, chaque point représente une zone de 1 000 pb découpée dans la séquence de l'ADN chromosomique du soja. Le nombre de *reads* alignés sur chaque position en condition élicitée et non élicitée pour les trois réplicats a été utilisé ici afin d'obtenir une valeur de *fold change* (FC) ainsi qu'une *p-value* associée. On observe donc ici le Log₂ du FC de chacun des points en fonction du Log₁₀ de leur *p-value*. Les zones présentant un plus fort alignement de *reads* en condition élicitée sont représentées en rouge, et les zones dont l'alignement des *reads* est plus important en condition non élicitée sont représentées en bleu. Le seuil statistique correspondant à une *p-value* de 0,05 est représenté par une ligne horizontale en pointillé et le seuil de FC minimal de 2 est représenté par les lignes verticales en pointillé.

La dernière partie de ces résultats consistera donc à observer et identifier pour l'ADN chromosomique des zones présentant (i) une différence de nombre d'alignement des *reads* en conditions élicitée et non élicitée et (ii) un fort nombre d'alignements de *reads* en condition élicitée et non élicitée.

II-2-2 Analyse différentielle sur l'ADN chromosomique

Afin d'étudier plus en détail les zones pour lesquelles l'alignement des *reads* semble être différent entre la condition élicitée et non élicitée, dix des positions avec les FC les plus importants ont été sélectionnées et représentées sous forme de tableau (Tableau 19). Parmi celles-ci, cinq ont un FC positif (plus de *counts* en condition élicitée) et cinq d'entre eux ont un FC négatif (plus de *counts* en condition non élicitée).

Comme observé précédemment sur le *volcanoplot* présentant l'alignement des *reads* d'ADNex sur l'ADN chromosomique (Figure 62), les log2 des FC les plus importants atteignent -3,8522 et 3,8293. Les *counts* moyens (avant leur normalisation) atteignent eux un maximum de 9 *reads* et un minimum de 0. De plus, on observe la présence d'un *count* moyen égal à 0 pour chacune des zones, soit en condition non élicitée lorsque le FC est positif, soit en condition élicitée lorsque le FC est négatif. Le nombre de *counts* étant très faible pour toutes ces zones, les différences observées entre les conditions élicitées et non élicitées semblent donc minimales, malgré leur FC important et leur *p-value* inférieure à 0,05.

Enfin, sept de ces zones ne correspondent à aucune zone ayant été annotée comme codante, dans le génome de *Glycine max* ainsi que dans les génomes des autres plantes (blast avec un seuil minimal de 70% d'identité). Parmi les trois zones correspondant à une annotation, la première (Tableau 19 ; Gm07 : 27 655 001 – 27 656 000) contient l'annotation d'une phosphatase putative DSP8 (*dual specificity protein phosphatase*). Les deux autres correspondent, après un blast, à un rétrotransposon avec un pourcentage d'identité de 85,88% (Tableau 19 ; Gm09 : 27 858 001 – 27 859 000) ainsi qu'à un transporteur avec un pourcentage d'identité de 87,13% (Tableau 19 ; Gm16 : 14 656 001 – 14 657 000).

Tableau 19 : Identification de zones d'ADN chromosomique où les alignements de *reads* d'ADNex de soja présentent les *fold change* (FC) les plus importants en condition élicitée (E) et en condition non élicitée (NE)

Position : position de la zone de 1000 pb sur le chromosome ; count moyen : moyenne des counts obtenue pour les trois réplicats ; Identification : annotation trouvée pour cette position sur le génome du soja ou *blast* de la séquence sur l'ensemble des génomes des plantes ; N/A : zone non codante

Condition	Chromosome	Position (pb)		Counts moyens	log2(FC)	P-value	Identification chez le soja ou blast (%identité)
		START	STOP				
E NE	Gm07	27 655 001	27 656 000	8 0	3,8293	0,0021	PREDICTED: Glycine max putative dual specificity protein phosphatase DSP8 (LOC100787468), transcript variant X1
E NE	Gm15	15 541 001	15 542 000	8 0	3,7595	0,0028	N/A
E NE	Gm07	7 109 001	7 110 000	7 0	3,7505	0,0028	N/A
E NE	Gm09	27 858 001	27 859 000	7 0	3,7060	0,0033	85,88% Glycine max retrotransposon gmw1-45m6-re-2
E NE	Gm16	14 656 001	14 657 000	7 0	3,6978	0,0034	87,13% PREDICTED: Glycine soja organic cation/carnitine transporter 7-like (LOC114422556)
E NE	Gm20	45 220 001	45 221 000	0 9	-3,8522	0,0020	N/A
E NE	Gm07	22 405 001	22 406 000	0 8	-3,8070	0,0022	N/A
E NE	Gm13	15 882 001	15 883 000	0 8	-3,7247	0,0029	N/A
E NE	Gm14	20 131 001	20 132 000	0 8	-3,6824	0,0036	N/A
E NE	Gm01	29 642 001	29 643 000	0 7	-3,6575	0,0036	N/A

II-2-3 Détection et identification des zones d'ADN chromosomique présentant un alignement plus important de séquences en conditions élicitée ou non élicitée

Afin d'identifier les zones d'ADNchr présentant le plus grand nombre d'alignement de *reads*, des graphiques représentant la distribution des *counts* pour chaque zone de 1 000 pb ont tout d'abord été réalisés sur Excel. Ceux-ci ont été faits pour chaque chromosome ainsi que pour les séquences d'ADN mitochondrial et plastidial de l'ensemble des échantillons, et ont permis d'identifier des zones d'intérêt. Afin de limiter le nombre de figures dans cette partie du manuscrit, seule la distribution des *counts* du réplicat 7 (E et NE) sur le chromosome 19 sera présentée (Figures 63 et 64). Les autres graphiques sont présentés en annexe III, partie II-2.

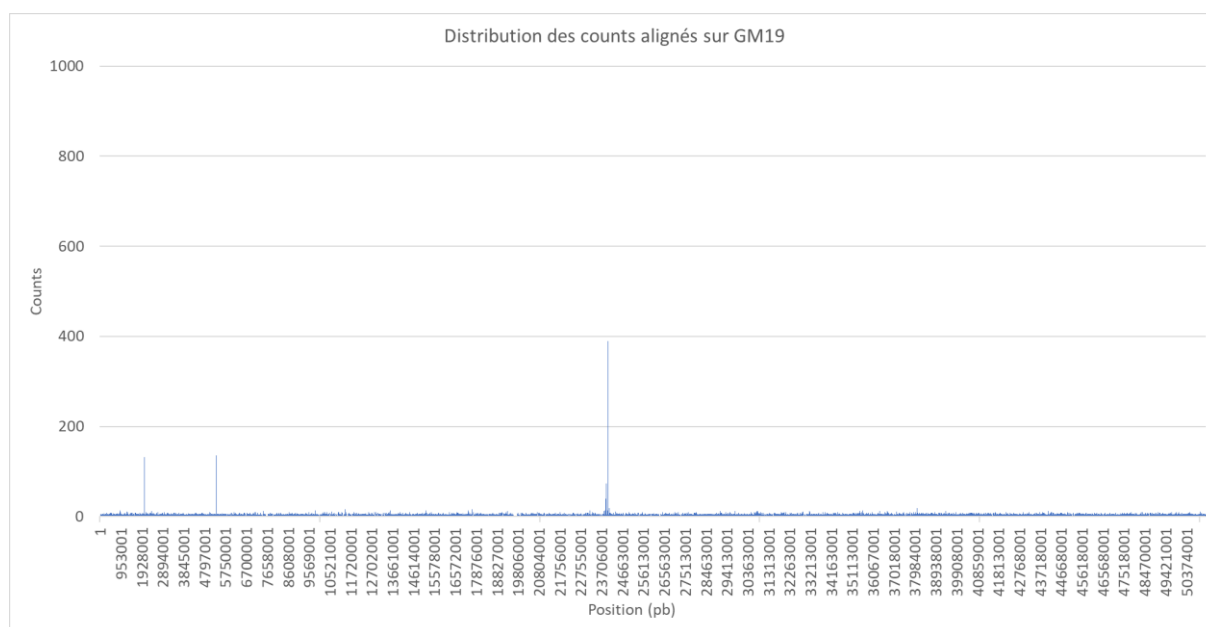


Figure 63 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM19 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 19, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb) et peut varier sur ce chromosome de 1 à 50,75 Mb.

Sur la figure 63, on observe donc le nombre de *counts* alignés sur le chromosome 19 en fonction de leur position sur celui-ci en condition non élicitée. Sur la majeure partie du graphique, une ligne de base est observée aux alentours de 0 *counts* par zone de 1 000 pb. Trois pics sont observés aux positions 2 059 001 pb ; 5 346 001 pb et 24 013 001 pb. Les deux premiers pics atteignent respectivement 132 et 136 *counts*, et le dernier pic atteint 389 *counts*.

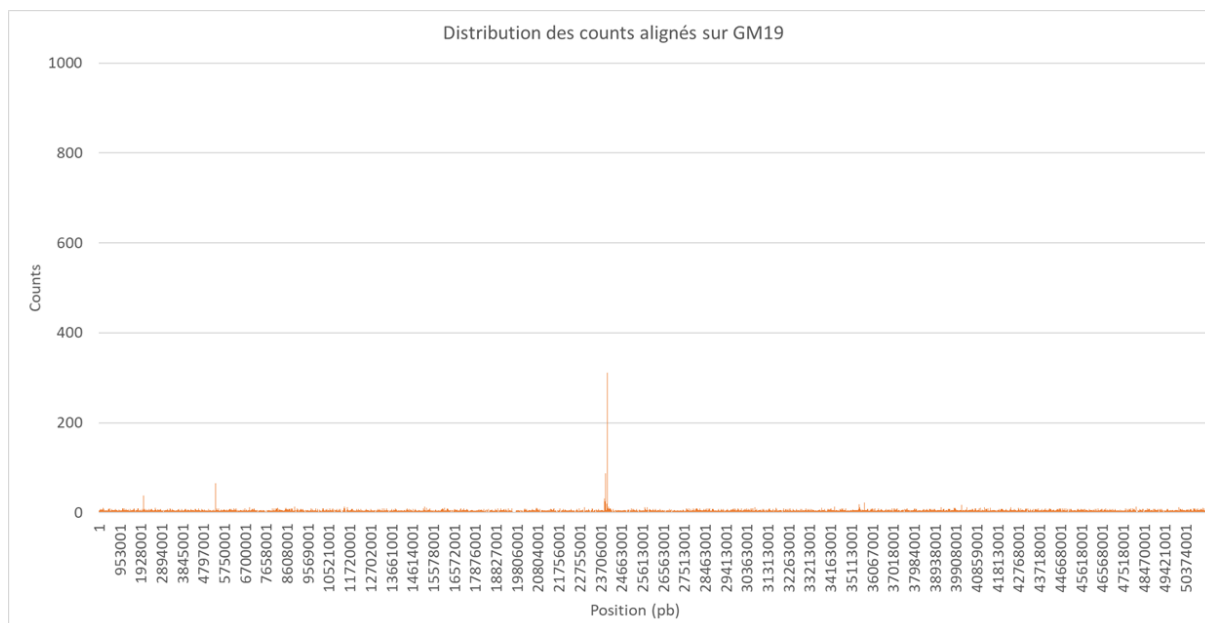


Figure 64 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM19 en condition élicitée

Sur cette figure on observe la distribution des counts sur le chromosome 19, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb) et peut varier sur ce chromosome de 1 à 50,75 Mb.

Sur la figure 64, on observe le nombre de *counts* alignés sur le chromosome 19 en condition élicitée en fonction de leur position sur celui-ci. La grande majorité du chromosome présente un nombre de *counts* proche de 0. Trois pics sont observés aux positions 2 059 001 pb ; 5 346 001 pb et 24 013 001 pb. Ceux-ci atteignent des *counts* de 66, 87 et 312.

Si l'on compare l'alignement des *reads* sur le chromosome 19 en condition élicitée et non élicitée, on observe donc que les trois pics repérés sont présents dans les deux conditions aux mêmes positions, cependant leur nombre de *count* est plus élevé en condition non élicitée par rapport à la condition élicitée. Des pics ont aussi été observés sur les graphiques montrant l'alignement des *reads* sur les autres chromosomes du soja. Afin d'identifier une partie de ces pics, seuls ceux présents pour les 6 échantillons (trois réplicats et deux conditions) et dont le nombre de *reads* alignés est toujours supérieur à 100 ont été sélectionnés.

Tableau 20 : Identification des zones de l'ADNchr pour lesquelles le nombre d'alignement des reads est supérieur à 100 counts pour tous les échantillons (7, 8 et 9 E et NE)

Position : position de la zone de 1000 pb sur le chromosome ; Nombre de reads : nombre de reads séquencés pour chaque réplicat ; Identification : annotation de cette zone sur le génome du soja ou blast de la séquence sur l'ensemble des génomes des plantes. E : élicité ; NE : non élicité ; N/A : zone non codante

Condition	Chromosome	Position (pb)		Nombre de counts			Log2(FC)	p-value	Identification chez le soja ou blast (% identité)	Commentaire		
		START	STOP	R7	R8	R9						
NE	Gm01	15 120 001	15 121 000	463	377	482	-0,2001	0,2168	N/A	Séquence répétée, identifiée spécifiquement chez <i>Glycine max</i>		
E				381	429	288						
NE	Gm02	21 076 001	21 077 000	191	157	178	-0,1269	0,5728	LOC100785390 protein kinase and PP2C-like domain-containing protein [<i>Glycine max</i> (soybean)]	Kinases impliquées dans la réponse de défense des plantes (Saijo <i>et al.</i> , 2018)		
E				147	189	124						
NE		21 073 001	21 074 000	178	143	157	-0,4408	0,1000				
E				130	124	102						
NE	GM04	26 982 001	26 983 000	312	273	398	-0,4529	0,0145	N/A	Séquence répétée, retrouvée chez <i>Glycine max</i> et <i>Glycine soja</i>		
E				219	264	200						
NE		26 981 001	26 982 000	264	261	297	-0,3252	0,0787				
E				225	230	172						
NE	Gm09	25 428 001	25 429 000	356	314	368	-0,3650	0,0423	79,66% <i>Pisum sativum</i> clone Ps-phage20 Ogre retrotransposon, partial sequence	Identifié spécifiquement chez les fabacées		
E				279	297	194						
NE		23 598 001	23 599 000	243	246	244	-0,3565	0,1129			96,05% <i>Glycine max</i> NB-LRR type disease resistance protein Rps1-k-1 (Rps1-k-1) and NB-LRR type disease resistance protein Rps1-k-2 (Rps1-k-2) genes, complete cds.	Séquence répétée, impliquée dans la résistance à <i>Phytophthora sojae</i> . Serait proche ou ferait partie d'une zone d'hétérochromatine (Gao et Bhattacharya 2008). Séquence retrouvée spécifiquement chez <i>Glycine max</i>
E				228	200	121						
NE	21 046 001	21 047 000	939	755	1022	-0,2724	0,0572	93,01% PREDICTED: <i>Glycine soja</i> probable WRKY transcription factor 11 (LOC114391634), mRNA	Facteurs de transcription WRKY impliqués dans la réponse de défense du soja (Song <i>et al.</i> , 2018). Séquence spécifiquement retrouvée chez <i>Glycine max</i> .			
E			821	789	535							
NE	Gm17	25 874 001	25 875 000	382	292	433	-0,4082			0,0332	97,96% PREDICTED: <i>Glycine soja</i> L-ascorbate oxidase homolog (LOC114411184), mRNA	Identifié chez les fabacées et malvacées (<i>Gossypium</i>)
E				281	317	196						
NE		25 875 001	25 876 000	266	233	363	-0,2705	0,1887				
E				244	235	192						
NE	Gm18	25 876 001	25 877 000	244	204	353	-0,3831	0,0719	72,09% PREDICTED: <i>Glycine soja</i> uncharacterized LOC114387464 (LOC114387464), ncRNA	Séquence répétée, retrouvée spécifiquement chez <i>G. max</i>		
E				228	204	158						
NE		32 363 001	32 364 000	202	176	223	-0,2318	0,2649				
E				224	176	137						
NE	Gm19	32 392 001	32 393 000	185	149	201	-0,0097	0,9693	N/A	Séquence répétée		
E				185	166	108						
NE	Gm20	24 013 001	24 014 000	389	318	455	-0,3740	0,0627	Triacyl-glycerol lipase 2	Identifié spécifiquement chez les fabacées		
E				312	352	191						
NE		20 696 001	20 697 000	253	232	355	-0,2738	0,1550			N/A	Séquence répétée, retrouvée spécifiquement chez <i>G. max</i>
E				228	215	217						
NE	20 695 001	20 696 000	167	133	200	-0,1727	0,4652					
E			170	135	117							

Le tableau 20 présente les seize pics observés dans les graphiques de distribution des counts pour lesquels le nombre de reads alignés dépasse 100 counts pour tous les échantillons. Pour l'identification, les seize pics ont été regroupés en onze zones. On observe tout d'abord que pour ces zones, le log2 du fold change est très proche de 0 et la p-value est rarement au-dessous du seuil statistique (0,05), il y aurait donc très peu de changement entre la condition élicitée et la condition non élicitée pour ces zones. Sur ces onze zones, seulement deux sont annotées chez le soja (pics Gm02 et Gm19) et cinq d'entre elles correspondent avec un pourcentage d'identité supérieur à 70% à un gène identifié chez un autre organisme (*Glycine soja* ou *Pisum sativum*). Enfin, parmi toutes ces zones, six

correspondent à des séquences répétées chez le soja (parmi lesquelles 4 sont des zones non codantes).

Discussion

- I. Etude transcriptomique et protéomique des AC-DC et des racines de soja
 - I-1 Réponse à PEP-13 des AC-DC
 - I-2 Réponse à PEP-13 des racines
 - I-3 Réponses à PEP-13 des AC-DC et des racines
 - I-4 Tentative d'identification du récepteur de PEP-13

- II. Analyse génomique de l'ADNex du RET du soja
 - II-1 Différences de couverture entre l'ADN chromosomique et l'ADN des organites
 - II-2 Différences dans l'alignement des *reads* sur l'ADNchr

Discussion

Dans le but de continuer de caractériser plus précisément le RET du soja (et notamment deux de ses acteurs) en condition élicitée par PEP-13 et témoin, une étude du transcriptome et du protéome des AC-DC et des racines par RNAseq et LC-Orbitrap ainsi que le séquençage de l'ADNex ont été réalisés et présentés dans ce manuscrit. Les résultats obtenus pour la partie transcriptomique et protéomique des AC-DC et des racines seront discutés dans la première partie de ce chapitre, et les résultats obtenus pour la partie d'analyse génomique de l'ADNex seront discutés en seconde partie de ce chapitre.

I. Etude transcriptomique et protéomique des AC-DC et des racines du soja

Dans un premier temps, l'analyse transcriptomique et protéomique des AC-DC et des racines en condition de stimulation immunitaire avec PEP-13 ou témoin a permis d'observer la présence de gènes différentiellement exprimés chez les AC-DC et chez les racines en réponse à PEP-13. Dans un second temps, l'étude plus détaillée des résultats obtenus a montré que les réponses des AC-DC et des racines à PEP-13 sont différentes. Dans les deux territoires racinaires, les quantités de gènes et de protéines exprimés sont plus importantes en condition témoin. Par ailleurs, un plus grand nombre de transcrits et de protéines différentiellement exprimées sont détectées chez les AC-DC par rapport aux racines de soja. Enfin, les AC-DC et les racines de soja répondent qualitativement de façon distincte à l'élicitation par PEP13 au niveau transcriptomique et protéomique.

I-1 Réponse à PEP-13 des AC-DC

Ainsi, si l'on observe à nouveau le tableau obtenu avec Mapman pour les AC-DC (Résultats, I-1-3-1 ; Tableau 9) pour la catégorie *stress*, on peut observer qu'un gène codant pour une *germin-like* est positivement régulé. Ces protéines sont des glycoprotéines retrouvées dans la paroi des cellules végétales impliquées dans un grand nombre de processus biologiques, dont la réponse aux stress biotiques (Bernier et Berna, 2001). Selon

Dunwell *et al.*, en 2008, ces protéines possèdent une activité superoxyde dismutase, permettant la synthèse d' H_2O_2 à partir de formes actives de l'oxygène. Elles pourraient donc permettre d'initier l'activation des défenses des cellules voisines à partir des formes actives de l'oxygène générées par des peroxydases. En effet, trois gènes codants des peroxydases sont retrouvés dans la catégorie *enzymes*. De plus, selon Christensen *et al.* (2004), les superoxyde dismutases pourraient jouer un rôle dans le renforcement de la paroi, qui est la première barrière de défense des cellules végétales.

Si l'on observe maintenant les gènes sur-exprimés dans la catégorie *cell wall* (Tableau 9), on note la présence de deux gènes codants pour des pectines méthylestérases (PME). Ces enzymes catalysent la dé-méthylesterification des pectines et sont impliquées dans des mécanismes comme l'adhésion cellulaire ou l'élongation des tiges. Les PME sont aussi responsables du détachement des AC-DC de l'apex racinaire (Mravec *et al.*, 2017) grâce à leur action de déméthylestérification des pectines, qui va promouvoir l'action des endopolygalacturonases dégradant les pectines. A un pH basique et en présence d'ions Ca^{2+} , les PME permettent aussi le renforcement de la paroi grâce à la formation d'un gel de pectines (Micheli, 2001). On observe également des gènes de synthèse des polymères de la paroi (*cellulose synthase*) et de modification de ses polymères comme par exemple une XTH (*Xyloglucan endoTransglucosylase/Hydrolase*). Cette enzyme permet la modification des polymères de xyloglucanes, et notamment la création d'un réseau de xyloglucanes et de cellulose (Song *et al.*, 2018). Ces polymères seraient, selon Ropitiaux *et al.* (2019), essentiels à la cohésion du mucilage et à la liaison des AC-DC entre elles grâce à la formation de ce réseau de xyloglucane et de cellulose.

Dans la catégorie *carbohydrates*, on observe la surexpression de gènes codants pour des enzymes du métabolisme du saccharose (glyma14g03300.1 et glyma17g04160.1), une α -galactosidase qui permet l'hydrolyse de résidus α -D-galactoses ou encore un gène codant pour une enzyme de synthèse du tréhalose (glyma12g36280.1). Ce disaccharide joue notamment un rôle dans le contrôle de la croissance et du développement des plantes, mais aussi dans la résistance aux stress abiotiques et biotiques (Phan et Van Dijck, 2019). En effet, le tréhalose pourrait être considéré comme un éliciteur par la plante, sa détection ayant induit la transcription de gènes de défense tels que le facteur de transcription WRKY6 et une PR-protéine (la β -1,3-glucanase) chez *Arabidopsis*. Parmi les gènes sur-exprimés chez les AC-

DC lors de l'élicitation avec PEP-13, on retrouve un gène codant pour une β -1,3-glucanase (Tableau 9, catégorie *enzymes* ; glyma03g28870.1) et deux gènes codants pour des facteurs de transcription WRKY et WRKY20 (catégorie *transcription factor* ; glyma02g01420.1 et glyma02g46690.1). La surexpression de ces gènes pourrait donc être due à la présence de tréhalose.

D'autres facteurs de transcription impliqués dans la défense sont sur-exprimés en réponse à PEP-13, comme les facteurs de transcription MYB (*myeloblastosis*) et bZIP (*basic leucine zipper*) qui sont connus pour être exprimés en réponse à un stress biotique et notamment chez le soja en réponse au champignon *Phakopsora pachyrhizi* Syd & P. Syd (Aoyagi *et al.*, 2014 ; Alves *et al.*, 2015). Ces facteurs de transcription, MYB et bZIP, sont activés par phosphorylation. Les protéines kinases impliquées n'étant pas identifiées chez le soja, on peut tout de même noter que parmi les gènes sur-exprimés en réponse à PEP-13, 23 gènes codent pour des protéines ayant une activité kinase (Tableau 9, *proteins, signal regulation*). L'activation de ces facteurs de transcription liés à la réponse de défense pourrait donc avoir lieu. Enfin, on retrouve aussi des régulateurs négatifs de la voie des protéines kinases, notamment trois gènes codants des PP2C (protéine phosphatase de type 2c) sont sur-exprimés (Tableau 9, *proteins*).

Les protéines kinases ne sont pas les seules protéines de régulation des signaux retrouvées dans la catégorie *signal regulation*. En effet, on retrouve aussi cinq gènes codants pour des acteurs de la signalisation par les protéines G (glyma07g07980.1 ; glyma07g09330.1 ; glyma09g40480.1 ; glyma15g01660.1 ; glyma18g52450.1), impliquées dans la transduction de divers signaux dont les stress biotiques (Pandey, 2019) ; ainsi que six gènes codants pour des acteurs de la signalisation calcique (glyma04g40460.1 ; glyma06g04950.1 ; glyma06g14300.1 ; glyma07g09070.1 ; glyma18g45350.1 ; glyma19g40050.1). La signalisation calcique, et notamment l'influx de calcium qui la déclenche, est bien connue pour intervenir lors de la détection d'un éliciteur par son PRR associé et pour être un élément important de la PTI (Tuteja et Mahajan, 2007 ; Saijo *et al.*, 2018). L'activation de cette voie semble donc confirmer que PEP-13 aurait bien été détecté par les AC-DC du soja, grâce à un récepteur PRR et aurait activé la PTI.

L'activation de la PTI des cellules végétales implique divers acteurs, parmi lesquels on retrouve notamment la production de la phytohormone éthylène (Saijo *et al.*, 2018). Pour la

catégorie hormones, on observe principalement la surexpression de gènes impliqués dans la voie de l'éthylène (glyma13g21120.1, glyma13g28970.1, glyma15g08370.1, glyma20g33800.1). Cette hormone est aussi connue chez les plantes pour son action lors de la maturation des fruits, de l'abscission, de la sénescence des fleurs, lors de la germination des graines ou encore en réponse à un stress abiotique ou biotique (Mattoo, 2018 ; Saijo *et al.*, 2018). Chez le soja, cette hormone est induite dans les racines en réponse à *Fusarium oxysporum* et jouerait donc un rôle dans la défense de celles-ci (Lanubile *et al.*, 2015). De plus, il a été montré par Yang *et al.*, en 2019 que le phytopathogène *Phytophthora sojae* Kaufm. & Gerd., était capable d'inhiber la synthèse de l'éthylène chez le soja grâce à son effecteur PsAvh238 (*avirulence homologue*) qui interagit avec l'enzyme 1-aminocyclopropane-1-carboxylate synthase (ACS). Enfin, dans la catégorie enzymes du tableau 9, deux lipases avec un motif GDSL sont retrouvées. Ces lipases sont connues pour leur action de modulation du signal de l'éthylène chez *Arabidopsis* ou encore chez le riz (Kim *et al.*, 2014 ; Zhao *et al.*, 2020). L'éthylène pourrait donc être un acteur important de la défense du soja face à *P. sojae*, organisme à partir duquel PEP-13 a été isolé.

Un dernier exemple de gène sur-exprimé en réponse à PEP-13 est celui codant pour la patatine (Tableau 9, *development*). Cette glycoprotéine est une protéine de réserve (Souza Cândido *et al.*, 2011) présentant une activité phospholipase. Les phospholipases sont des enzymes dégradant la membrane plasmique, et relâchant donc divers lipides (diacylglycerol, phosphatidic acid, lysophospholipides...) qui vont être reconnus par la cellule végétale comme des DAMPs, induisant l'activation de cascades de MAP Kinases, des changements de pH cytoplasmique, ou encore la régulation de gènes PR (*pathogenesis-related*) (D'Ambrosio *et al.*, 2017 ; Li *et Wang*, 2019). Deux autres gènes codant pour des phospholipases ont de plus été détectés dans la catégorie *enzymes* (glyma15g02710.1 ; glyma15g16270.1). Il semble donc que les phospholipases, dont la patatine, jouent un rôle dans la réponse des AC-DC à l'éliciteur PEP-13.

Enfin, l'analyse Mapman du protéome des AC-DC a permis d'identifier deux protéines sur-exprimées en réponse à PEP-13 (*cf* Résultats, I-2-3 ; Tableau 12). L'une, classée dans la catégorie *stress*, correspond à une osmotine (glyma05g38110.1). Les osmotines correspondent aux protéines de défense PR-5 et sont connues pour intervenir dans la défense des plantes face aux stress biotiques induits par les champignons, virus ou bactéries

(Hakim *et al.*, 2018), mais aussi dans la réponse de défense à l'oomycète *Phytophthora capsici* Leonian, chez le poivre (*Piper nigrum* L.) (Vandana et Suseela Bhai, 2018). La seconde protéine détectée, classée dans la catégorie *specialized metabolism*, est une protéine impliquée dans la synthèse de la lignine (glyma18g38670.1). Les lignines participent au renforcement de la paroi et notamment dans la réponse du soja à une infection par *Phytophthora sojae* (Yan *et al.*, 2019 ; Wang *et al.*, 2020). Ces deux protéines semblent donc être impliquées dans la réponse des AC-DC du soja à PEP-13.

L'éliciteur PEP-13 semble donc induire, chez les AC-DC (Figure 65), l'activation de la PTI (production de ROS grâce aux peroxydases, activation de la signalisation calcique et facteurs de transcription, production d'éthylène, et d'enzymes kinases). En complément de l'activation de la PTI, le renforcement de la paroi (production de lignines et PME) ainsi que le renforcement du mucilage semblent être activés grâce à la production de PME, de cellulose et la création de réseaux de xyloglucanes et de cellulose grâce à la production de XTH (*Xyloglucan endoTransglycosylase/Hydrolase*).

De plus, des mécanismes de transmission du signal de défense semblent être mis en place. Notamment grâce à l'accumulation de tréhalose, de patatine et de phospholipases qui vont permettre la formation de DAMPs. Ces DAMPs vont être détectés par la cellule elle-même ainsi que celles environnantes, créant ainsi une boucle d'activation de la PTI chez les AC-DC.

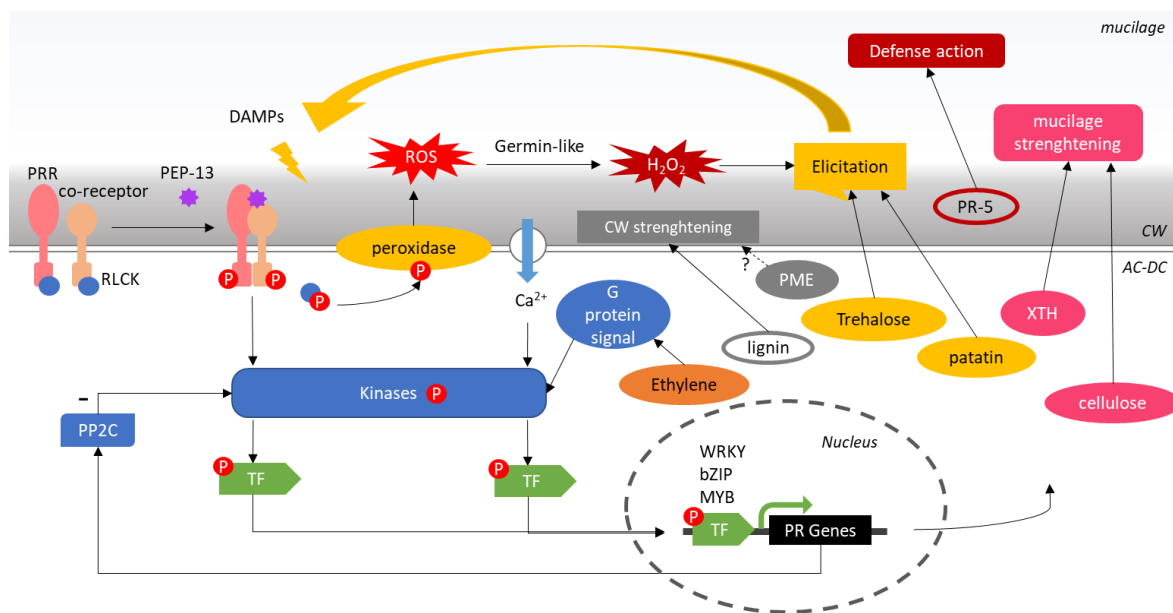


Figure 65 : Synthèse de l'observation de mécanismes activés en réponse à PEP-13 chez les AC-DC de racines de soja

Au bout de 5 jours, pour les AC-DC, la détection de PEP-13 par son PRR associé entraîne la phosphorylation de kinases, aboutissant à l'activation de facteurs de transcription (TF). Les RLCK (*receptor-like cytoplasmic kinase*) vont activer des peroxydases et permettre la synthèse de ROS (*reactive oxygen species*). Ces ROS peuvent être converties en H₂O₂ par les protéines *germin-like*. La réponse à PEP-13 inclut également la synthèse de PME (pectine méthylestérases), de tréhalose, de patatines, d'osmotines ou PR-5 (*pathogenesis related*), d'éthylène, de cellulose ou encore de XTH (*Xyloglucan endoTransglucosylase/Hydrolase*). Ces acteurs vont permettre le renforcement de la paroi (CW : cell wall) et le remodelage du mucilage, ainsi que la production de DAMPs et l'élicitation des cellules environnantes. Enfin, la synthèse de PP2C (protéine phosphatase de type 2C) permet l'inhibition des kinases. Ce schéma est une synthèse des observations réalisées précédemment sur une sélection de transcrits et protéines issus de notre étude, et n'est pas une représentation de la totalité de la réponse des AC-DC à PEP-13. En effet d'autres transcrits et protéines sur-exprimées en condition élicitée n'ont pas été pris en compte pour la réalisation de ce schéma. Figures pleines : données transcriptomiques ; figures sans remplissage : données protéomiques (ici PR-5 et *lignin*). Figure inspirée de celle de Saijo *et al.*, 2018.

I-2 Réponse à PEP-13 des racines

Pour les racines, on observe quatre gènes classés dans la catégorie *stress* et dont l'expression est positivement régulée (Tableau 10). Parmi ceux-ci, on retrouve notamment deux gènes codant pour des protéines permettant la reconnaissance des pathogènes : une protéine possédant un domaine récepteur TIR-NB-LRR (*Toll/interleukin 1 receptor nucleotide-binding leucine-rich repeat* ; glyma16g03780.1), et une protéine possédant un domaine récepteur NBS-LRR (*nucleotide binding – leucine rich repeat* ; glyma09g02420.1). Ces domaines sont connus pour leur capacité à se lier à un effecteur provenant d'un pathogène grâce au domaine LRR, cette liaison induisant l'activation d'une fonction de clivage du NADP⁺ lorsqu'il y a présence d'un domaine TIR, ainsi que l'activation du processus de mort cellulaire et la transcription de gènes de défense grâce au domaine NBS (Chang *et al.*, 2019). Ces deux protéines, dont les transcrits d'ARN sont détectés dans la racine en réponse à PEP-13, pourraient donc jouer un rôle dans la reconnaissance des effecteurs des oomycètes du genre *Phytophthora*.

De plus, une chitinase potentielle a été détectée parmi les gènes positivement régulés par PEP-13. Les chitinases sont des enzymes qui permettent de dégrader la chitine de la paroi des pathogènes fongiques (Malik, 2019). Cependant, une étude de Rocha *et al.*, en 2015 a montré la présence de chitinases dans les exsudats du soja en réponse au nématode *Meloidogyne incognita* Kofoid & White, celle-ci pouvant être expliquée par la présence de chitine dans les œufs produits par ce nématode (Rocha *et al.*, 2015). En revanche, les chitinases ne semblent pas spécifiques de la réponse du soja aux oomycètes (Zhao *et al.*, 2013 ; Dong *et al.*, 2019). Un blast des peptides de la séquence de PEP-13 sur la base de données NCBI des protéines des champignons permet d'observer, pour une couverture de 100% et un pourcentage d'identité de 84,62%, une correspondance avec une protéine de *Catenaria anguillulae* Sorokin (N° d'accès ORZ39265.1) ainsi qu'à d'autres protéines appartenant à des champignons du genre *Xylaria*, *Cladophialophora*, *Rhizophagus* ou encore *Glomus*. Il se pourrait donc que PEP-13 soit reconnu comme un motif appartenant aux oomycètes mais aussi aux champignons par la racine de soja.

La reconnaissance de PEP-13, par son récepteur PRR associé devrait engendrer la transduction d'un signal, le plus souvent sous forme d'une cascade de phosphorylation de protéines kinases. Parmi les gènes de la catégorie *signal regulation* régulés positivement en

réponse à PEP-13, on retrouve notamment une seule kinase (glyma19g31110.1), ainsi qu'un récepteur du glutamate (glyma12g32020.1), une protéine intervenant dans la réponse à la lumière (glyma02g17240.1) et enfin sept gènes codants pour des acteurs de la signalisation par les protéines G. Les protéines G semblent donc être impliquées dans la réponse des racines à PEP-13. Pour rappel, ces protéines sont impliquées dans un grand nombre de processus biologiques, dont la réponse aux stress biotiques (Pandey, 2019). Leur action aboutissant à l'activation de facteurs de transcription.

Parmi les facteurs de transcription régulés positivement en réponse à PEP-13, on retrouve un transcrit qui semble appartenir à la famille des MYB (*myeloblastosis* ; glyma08g20270.1), connu pour être exprimé en réponse à un stress biotique et notamment chez le soja en réponse au champignon *Phakopsora pachyrhizi* Syd & P. Syd (Aoyagi *et al.*, 2014). D'autres facteurs de transcriptions sont aussi retrouvés comme par exemple le facteur de transcription ARF (auxin response factor) ou encore un facteur de transcription de la réponse à l'éthylène ERF (ethylene-response factor ; glyma02g08020.1).

Dans la catégorie *hormones*, cinq gènes positivement régulés apparaissent. Ceux-ci sont tous les cinq annotés comme faisant partie des voies de synthèse, de transduction du signal ou de réponse à l'éthylène. La voie de l'éthylène, indiquant une réponse de défense (Van der Ent et Pieterse, 2018) et notamment lors de la PTI (Saijo *et al.*, 2018), semble donc être activée dans les racines en réponse à PEP-13.

L'activation de la défense des cellules végétales, et notamment lors de la PTI, passe aussi par le renforcement de la paroi. Dans la catégorie *cell wall*, un gène codant pour une pectine méthylestérase (PME) et son inhibiteur a été détecté. Les PME sont impliquées dans le renforcement ou le relâchement de la paroi en fonction du pH et de la présence d'ions Ca^{2+} (Micheli, 2001). Ces enzymes sont aussi impliquées dans le détachement des AC-DC (Mravec *et al.*, 2017). La présence de ce transcrit en réponse à PEP-13 pourrait donc être à l'origine d'un renforcement de la paroi de la racine, ou d'une augmentation de la production d'AC-DC par la racine en réponse à l'éliciteur PEP-13.

Trois enzymes de synthèse de lipides sont aussi retrouvées dans la catégorie *other metabolisms*, dont la 3-ketoacyl-CoA synthase, qui est fortement sur-exprimée en condition élicitee (Log₂ (FC) = 5,58). Il semblerait donc qu'en réponse à PEP-13, les racines augmentent

fortement leur synthèse des lipides. Ces lipides, pouvant former des membranes, pourraient donc être utilisés pour le transport vésiculaire, ou pour l'élongation des cellules racinaires. En effet, la formation des cellules bordantes passerait par une étape d'élongation cellulaire (Mravec *et al.*, 2017). Dans la catégorie *cell wall*, correspondant aux gènes de synthèse ou de modification de la paroi, quatre gènes sont retrouvés. Parmi ceux-ci, deux correspondent à des expansines (glyma09g37090.1 et glyma17g15670.1) qui permettent le relâchement de la paroi (Cosgrove, 2015), nécessaire à la croissance cellulaire. Les racines semblent donc, en réponse à PEP-13, augmenter leur production d'AC-DC. En effet il a été observé une augmentation significative de la quantité d'AC-DC dans le RET du pois lors d'une infection de celui-ci par l'oomycète *Aphanocyces euteiches* Drechsler par Cannesan *et al.*, en 2011. Une réponse similaire pourrait donc être enclenchée en réponse à l'éliciteur PEP-13, provenant des oomycètes du genre *Phytophthora*.

La dernière enzyme apparaissant dans la catégorie *cell wall* est une enzyme de synthèse de rhamnose (glyma08g15680.1). Le rhamnose est notamment retrouvé dans le RGI et dans d'autres polymères pariétaux. Celui-ci pourrait donc, par exemple, être synthétisé dans le but de renforcer la paroi grâce au RGI (Harholt *et al.*, 2010).

Divers gènes codant pour des acteurs de la synthèse et de la dégradation des protéines sont retrouvés dans la catégorie *proteins*. Notamment une sous-unité de type alpha du protéasome (glyma02g05190.1), deux protéines contenant un domaine F-box (glyma03g41530.1 ; glyma19g44130.1), ou encore des protéines ribosomales (glyma10g36070.1 ; glyma03g39480.1 ; glyma03g35530.1 ; glyma15g04670.1), permettant donc la synthèse de protéines. Trois glycosyltransférases sont retrouvées dans la catégorie *enzymes*, ainsi qu'une α -manosidase. Il semble donc qu'en réponse à PEP-13, la production de glycoprotéines (comme des HRGPs) ou de polysaccharides soit positivement régulée. Parmi les glycoprotéines, certaines sont connues pour jouer un rôle dans les mécanismes de défense racinaire, comme par exemple les AGPs qui ont été retrouvées dans les exsudats racinaires de différentes plantes, et qui, chez le pois, possèdent une action chimio-attractante sur les zoospores d'*Aphanomyces euteiches* (Cannesan *et al.*, 2012). De plus, les AGPs ont aussi un rôle dans la croissance cellulaire (Willats et Knox, 1996 ; Nibbering *et al.*, 2020), et pourraient donc jouer un rôle dans la maturation des AC-DC qui vont croître et parfois s'incurver tout en se détachant de l'apex racinaire (Mravec *et al.*, 2017).

Dans la catégorie transport, on retrouve notamment deux transporteurs impliqués dans la détoxification de la cellule (glyma01g03090.1 ; glyma12g32010.1), un transporteur de potassium (glyma08g02290.1) ou encore de magnésium (glyma06g16670.1).

Enfin, dans la catégorie *DNA*, on retrouve principalement des gènes de réparation de l'ADN, il semble donc que la détection de PEP-13 induise la mise en place de complexes de réparation de l'ADN, en prévision d'une éventuelle dégradation de celui-ci.

Pour finir, l'analyse Mapman du protéome des racines de soja en réponse à PEP-13 a permis de détecter cinq protéines sur-exprimées en réponse à l'élicitation. Parmi ces protéines, on retrouve deux protéines ribosomales (glyma10g02060.1 et glyma19g39570.1), impliquées dans la synthèse des protéines ; une protéine permettant le transport transmembranaire du calcium (glyma19g35960.1), nécessaire à l'activation de la PTI (Abdul Malik *et al.*, 2020) ; une histone (glyma10g03840.1) et une protéine chloroplastique (glyma13g27810.1) classées dans la catégorie *unknown/unclassified* et une kiwelline (glyma13g39890.1) dans la catégorie *development*. Les kiwellines sont des protéines connues pour leur implication dans la défense des plantes face aux pathogènes (Bange et Altegoer, 2019) et notamment chez le maïs pour leur interaction avec l'effecteur Cmu1 (*chorismate mutase*) ; qui perturbe la synthèse d'acide salicylique, provenant du champignon *Ustilago maydis* (DC.) Corda (Wildermuth, 2019). Des kiwellines ont également été détectées dans la réponse de la pomme de terre à l'oomycète *Phytophthora infestans* (Ali *et al.*, 2014).

L'éliciteur PEP-13 semble donc induire, chez les racines de soja et au bout de 5 jours d'élicitation (Figure 66), des mécanismes de défense comme la détoxification, la synthèse de chitinases et de kiwellines, ou encore de protéines de détection des effecteurs, la synthèse de lipides, de PME et d'expansines pouvant permettre au niveau de l'apex racinaire la libération d'AC-DC. La transmission du signal de défense semble pouvoir être assurée par les protéines G dont un grand nombre de transcrits permettant leur régulation ont été positivement régulés par PEP-13.

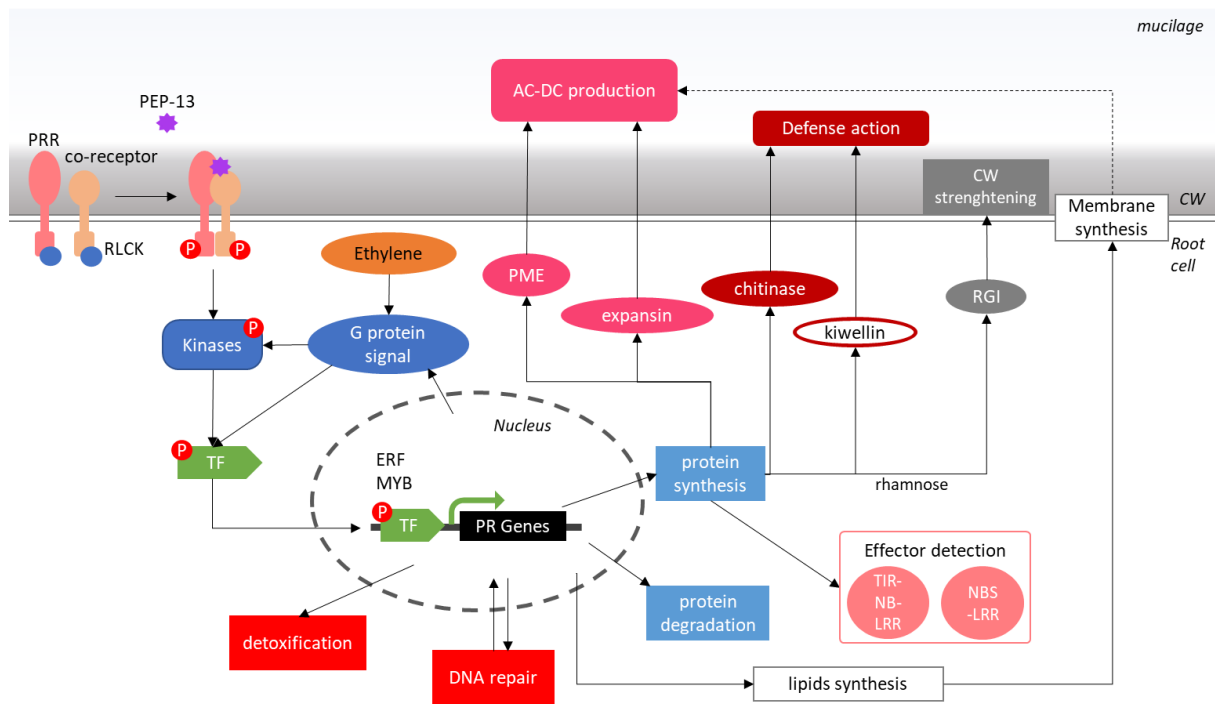


Figure 66 : Synthèse de l'observation de mécanismes activés en réponse à PEP-13 chez les racines de soja

Au bout de cinq jours, pour la racine, la reconnaissance de PEP-13 par son récepteur semble activer des mécanismes de défense comme la détoxification, la réparation de l'ADN, ou la synthèse de chitinases et de kiwellines. De plus, il semble que la production d'AC-DC soit positivement régulée. En effet, certaines protéines permettant leur différenciation (pectine méthylestérase et expansines) ainsi que la synthèse de lipides pourraient être impliquées dans ce mécanisme. Ce schéma est une synthèse des observations réalisées précédemment sur une sélection de transcrits et protéines détectées lors de notre étude, et n'est pas une représentation de la totalité de la réponse des racines à PEP-13. En effet d'autres transcrits et protéines sur-exprimées en condition élicitée n'ont pas été pris en compte pour la réalisation de ce schéma. Figures pleines : données transcriptomiques ; figures sans remplissage : données protéomiques (ici *kiwelling*) ; TF : facteur de transcription ; TIR-NB-LRR : *Toll/interleukin 1 receptor nucleotide-binding leucine-rich repeat* ; NBS-LRR : *nucleotide binding – leucine rich repeat* ; RGI : rhamnogalacturonane de type I ; AGP : *arabinogalactan protein* ; CW : *cell wall*. Figure inspirée de celle de Saijo *et al.*, 2018.

I-3 Réponses à PEP-13 des AC-DC et des racines

L'observation des réponses des AC-DC et des racines à PEP-13 permet donc d'observer tout d'abord une réponse de défense, et notamment l'activation de la voie de l'éthylène dans ces deux zones. Dans un second temps, on observe que la réponse de ces deux éléments semble très différente. En effet, les racines vont mettre en place des mécanismes de défense comme la détoxification ou la réparation de l'ADN, ainsi que la synthèse de différentes molécules permettant d'aboutir à la différenciation des AC-DC et donc à l'augmentation de leur population. Les AC-DC quant à elles semblent réaliser une activation de la PTI, ainsi que la synthèse de molécules permettant la formation de DAMPs (tréhalose, patatines, *germins-like*). Les AC-DC vont de plus activer le renforcement de leur paroi ainsi que le remodelage du mucilage.

Ces réponses distinctes entre les AC-DC et la racine semblent donc confirmer les études ayant précédemment montré une différence des réponses à l'élicitation en fonction des zones racinaires (Millet *et al.*, 2010 ; Haerizadeh *et al.*, 2011 ; Chuberre *et al.*, 2018).

Enfin, si l'on compare les réponses des AC-DC et des racines de notre étude à celles obtenues chez le soja en réponse à des éliciteurs provenant de *P. sojae* (Tableau 21 ; Parker *et al.*, 1988 ; Graham *et al.*, 2003 ; Daxberger *et al.*, 2007), on retrouve des résultats concordants.

On observe tout d'abord que les réponses observées semblent correspondre aux résultats obtenus par Parker *et al.*, en 1988. En effet, la production de glycéollines par les cotylédons et les cellules en culture de soja n'avait pas été observée lors d'une élicitation avec la partie protéique d'un extrait de paroi de *P. sojae* (contenant donc PEP-13). A l'inverse, l'élicitation des cotylédons et des cellules en culture de soja par la partie glycanique de l'extrait de paroi de *P. sojae* avait permis l'accumulation de glycéollines. Parmi les transcrits sur-exprimés en réponse à PEP-13 (avec un FC supérieur ou égal à 1,5 une *p-value* inférieure à 0,05 et possédant une annotation), le *glycinol 4-dimethylallyltransferase*, enzyme responsable de la synthèse de glycéollines et le facteur de transcription MYB29A2 permettant leur synthèse (Jahan *et al.*, 2020), n'ont pas été détectés. Il ne semble donc pas qu'il y ait une production de glycéollines dans les racines ou dans les AC-DC en réponse à PEP-13.

Ensuite, les études de Graham *et al.*, en 2003 et Daxberger *et al.*, en 2007 ont montré chez les cotylédons et les cellules en culture de soja, une production des protéines PR-2, PR-4 et PR-6 (Graham *et al.*, 2003) ainsi que des MAPK1, MAPK3 et MAPK6 (Daxberger *et al.*, 2007) en réponse aux β -glucanes de la paroi de *P. sojae*. Chez les AC-DC, aucune de ces molécules n'a été retrouvée parmi les transcrits sur-exprimés en réponse à PEP-13. Chez les racines, seule l'une de ces molécules : une endoglucanase a pu être retrouvée parmi les transcrits sur-exprimés en réponse à PEP-13 (\log_2 du FC = 1,27 soit environ FC = 2,4 et *p-value* = 0,021).

On observe donc une différence de réponse des racines et des AC-DC par rapport aux cotylédons et aux cellules en culture du soja, face à des éliciteurs provenant de

Phytophthora sojae. La réponse des différents tissus et zones semble dépendre aussi de la nature de l'éliciteur utilisé (ici peptidique ou glycanique).

Tableau 21 : comparaison de la réponse à PEP-13 des AC-DC et de la racine du soja aux réponses obtenues dans la bibliographie à des éliciteurs provenant de *Phytophthora sojae*

Molécules	Eliciteur	Tissu observé	Référence	Présence dans le transcriptome des AC-DC ?	Présence dans le transcriptome des racines ?
Absence de Glycéollines	Partie protéique d'un extrait de paroi de <i>P. sojae</i>	cotylédons et culture cellulaire	Parker <i>et al.</i> , 1988	Non	Non
Glycéollines	Partie glycanique d'un extrait de paroi de <i>P. sojae</i>		Parker <i>et al.</i> , 1988	Non	Non
PR-2 (endoglucanase)	β -glucanes de la paroi de <i>P. sojae</i>	cotylédons	Graham <i>et al.</i> , 2003	Non	Oui, FC = 1,28
PR-4 (Win-like protein)				Non	Non
PR-6 (Kunitz tripsin inhibitor)				Non	Non
MAPK1	β -glucanes de la paroi de <i>P. sojae</i>	culture cellulaire	Daxberger <i>et al.</i> , 2007	Non	Non
MAPK3				Non	Non
MAPK6				Non	Non

Pour finir, la comparaison des résultats obtenus grâce aux analyses transcriptomique et protéomique des réponses des AC-DC et des racines à PEP-13 a montré des résultats différents, notamment dans le nombre de processus biologiques impliqués. En effet, le nombre de protéines détectées comme différentiellement exprimées en protéomique (31 pour les AC-DC et 13 pour les racines) est largement inférieur au nombre de gènes détectés comme différentiellement exprimés en transcriptomique (2295 pour les AC-DC et 553 pour les racines), ce qui peut expliquer en partie cette différence.

Le fait qu'un plus grand nombre de transcrits d'ARN soit détecté par rapport aux protéines ne semble pas très étonnant si l'on observe les résultats obtenus dans la bibliographie pour les différentes analyses protéomiques (ex : Bai *et al.*, 2019 qui ont détectés 131 protéines différentiellement exprimées en réponse à une inoculation avec *Funneliformis mosseae* Walker & Schüßler) et transcriptomiques (ex : Lanubile *et al.*, 2015 qui ont détecté 1659 gènes différentiellement exprimés en réponse à une inoculation avec une souche pathogène de *Fusarium oxysporum* Schltdl.) réalisées sur la racine du soja. De plus, lors d'un stress abiotique comme un stress salin (Sobhanian *et al.*, 2010) ou un stress hydrique d'excès d'eau (Komatsu *et al.*, 2009) le nombre de protéines différentiellement

exprimées est beaucoup plus faible (respectivement 14 et 34 protéines) et se rapproche donc des 44 protéines (31 chez les AC-DC + 13 chez les racines) différentiellement exprimées détectées dans notre analyse.

Cependant, un total de 44 protéines différentiellement exprimées chez les racines et les AC-DC en réponse à PEP-13 semble assez faible par rapport aux 131 obtenues dans l'étude de Lanubile *et al.*, en 2019. Il est donc possible qu'un certain nombre de protéines n'aient pas été détectées avec la méthode utilisée dans notre analyse, mais une différence importante entre ces deux analyses : la nature du traitement appliqué, pourrait expliquer en partie cette différence. En effet, dans notre étude la racine et les AC-DC de soja sont confrontées à une élicitation par un peptide PEP-13 avec un FC minimal de 1,5, alors que dans l'étude de Bai *et al.*, en 2019, la racine de soja est inoculée avec un champignon mycorhizien et le FC minimal retenu est de 1,2. Une inoculation avec un organisme entier pourrait donc induire une réponse plus complexe que la réponse à un éliciteur seul, expliquant donc, avec la différence de FC minimal, un nombre de protéines différentiellement exprimées plus faible dans notre étude. La base de données utilisée pour l'identification des peptides étant la même dans notre étude et celle de Bai *et al.*, (Uniprot) la différence de nombre de protéines détectées ne semble donc pas provenir de ce paramètre. Cependant, un test d'identification des protéines à partir du transcriptome réalisé dans notre étude et donc sur la même variété de soja (CASTETIS) serait susceptible de nous permettre d'identifier un plus grand nombre de protéines.

I-4 Identification du récepteur de PEP-13

Dans le but d'identifier le récepteur de PEP-13, une analyse bio-informatique a été réalisée sur les gènes sur-exprimés en réponse à PEP-13 et étant annotés comme récepteurs ou non annotés sur la base de données Uniprot. Grâce à cette analyse, 20 gènes ont été identifiés comme potentiels récepteurs chez les AC-DC et chez les racines. Cependant, aucun gène commun à ces deux zones n'a été détecté parmi les gènes dont le FC est supérieur ou égal à 1,5 et la *p-value* inférieure à 0,05. Il est donc possible que le récepteur de PEP-13 ne soit pas sur-exprimé (ou avec un FC inférieur à 1,5) en réponse à celui-ci après cinq jours d'élicitation dans l'une ou dans les deux zones. De plus, le récepteur dont le transcrit atteint

le FC le plus important parmi les 20 détectés a été analysé avec I-TASSER afin de prédire un modèle de structure 3D pour celui-ci. La prédiction obtenue semble indiquer que celui-ci ne correspond pas à un récepteur de PEP-13.

Parmi les récepteurs potentiels détectés par DRAGO2, on retrouve des récepteurs de classe CN (*coiled-coil* (CC) *nucleotide binding site* (NBS)), N (NBS), RLK (*receptor like kinase*), RLP (*receptor like protein*), T (*toll/interleukine receptor*), et L (LRR : *leucine rich repeat*) ainsi que des kinases transmembranaires ou non. Selon Santana Silva et Micheli en 2020, 13 classes différentes de récepteurs ont été identifiées (Figure 67). Ces classes sont composées d'un ou plusieurs domaines parmi les dix identifiés sur la figure 67. Cependant, l'outil utilisé pour notre analyse ne permet pas de détecter les domaines RPW8 (*resistance to powdery mildew 8*), GNK2 (*ginkobilobin 2*) et MLO (*mildew resistance locus*). Les potentiels récepteurs de type RPW8NL (RPW8, NBS, LRR), MLO et RLK-GNK2 n'ont donc pas pu être identifiés parmi la liste des gènes analysés.

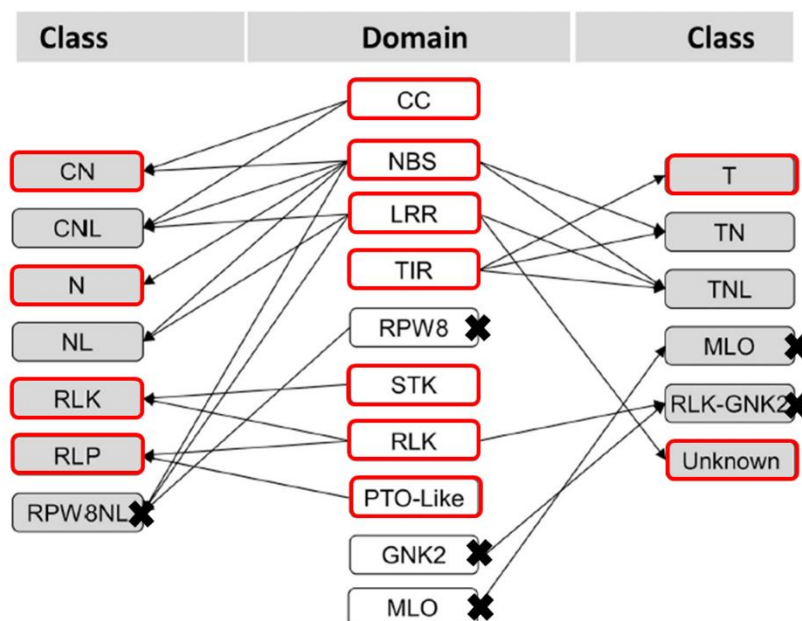


Figure 67 : Schéma des différentes classes de récepteurs connues ainsi que les domaines qui les composent

Treize classes de récepteurs ont été décrites, celles-ci sont composées d'un ou plusieurs domaines parmi les dix identifiés pour le moment. Les domaines et classes de récepteurs encadrés en rouge correspondent à ceux détectés par DRAGO2 dans la liste des vingt gènes détectés comme potentiels récepteurs de PEP13. Les domaines et classes de récepteurs représentés avec une croix noire correspondent à ceux qui ne peuvent pas être détectés par DRAGO2. CC : *coiled-coil* ; NBS : *nucleotide binding site* ; LRR : *leucine rich repeat* ; TIR : *toll/interleukine receptor* ; RPW8 : *resistance to powdery mildew 8* ; STK : *serine threonine kinase* ; RLK : *receptor like kinase* ; PTO : *receptor-like protein* ; GNK2 : *ginkobilobin 2* ; MLO : *mildew resistance locus*. Figure adaptée de Santana Silva et Micheli, 2020.

II. Analyse génomique de l'ADNex du RET du soja

L'analyse génomique réalisée sur l'ADNex du RET de soja en condition élicitee ou témoin a permis d'obtenir des *reads* s'alignant sur l'ensemble de l'ADN (chromosomique, plastidial et mitochondrial) du soja. Des différences d'alignement ont été observées en fonction (i) de l'ADN sur lequel a été aligné l'ADNex, et (ii) de la position des *reads* sur les chromosomes. Par contre, de très faibles différences ont été observées entre les conditions élicitee et non élicitee.

II-1 Différences de couverture entre l'ADN chromosomique et l'ADN des organites

L'alignement des *reads* d'ADNex sur l'ensemble du génome du soja a montré une différence majeure : la différence de couverture entre l'ADNchr et les ADNm et ADNp. En effet, même si un plus grand nombre de *reads* s'alignent sur l'ADN chromosomique par rapport à l'ADN des organites, un meilleur taux de couverture est observé pour les ADN mitochondrial et plastidial (12,62 X en condition élicitee et 17,29 X en condition non élicitee) que pour l'ADNchr (0,22 X en condition élicitee et 0,25 X en condition non élicitee) (Tableau 18). Cette contradiction s'explique par la différence de taille entre l'ADNchr (942,21 Mb) et l'ADN des organites (0,55 Mb).

Il semble donc que l'ADNex soit un mélange des ADNchr, ADNm et ADNp. Cependant, l'ADNchr semble moins bien représenté que les ADNp et ADNm au sein du RET. Ceci pourrait être lié à une meilleure conservation des ADNp et ADNm comparé à l'ADNchr dans le milieu extracellulaire. Selon Sirois et Buckley en 2019, la persistance de l'ADN dans le sol dépend de différentes propriétés extrinsèques (milieu extracellulaire) ou intrinsèques (propriétés de l'ADN). Dans le cas du RET, les ADNchr, ADNp, et ADNm sont libérés dans le même milieu extracellulaire et se retrouvent dans le mucilage du RET. La persistance différentielle de ces séquences d'ADN ne peut donc pas être expliquée par un critère extrinsèque. Parmi les critères intrinsèques pouvant jouer sur la persistance de l'ADN extracellulaire, on retrouve par exemple le pourcentage en bases G et C, ou encore sa masse moléculaire (Nagler *et al.*, 2018 ; Sirois et Buckley 2019). En effet, la stabilité d'une séquence d'ADN est en partie régulée par son ratio de bases G et C par rapport aux bases A et T. Les bases A et T étant moins stables que les bases G et C (Yakovchuk *et al.*, 2006 ; Zhang *et al.*, 2014). Or, si l'on

compare les pourcentages de GC des différentes séquences d'ADN chromosomique (35% GC, Schmutz *et al.*, 2010), mitochondrial (45% GC, Chang *et al.*, 2013) et plastidial (34% GC, Sasaki *et al.*, 2005) du soja, on observe un pourcentage en GC plus élevé chez l'ADN mitochondrial, et un pourcentage similaire de GC entre l'ADN chromosomique et l'ADN plastidial. Cette variable permet d'expliquer une possible meilleure conservation de l'ADNm par rapport à l'ADNchr mais ne justifie pas la persistance supérieure de l'ADNp.

La masse moléculaire de l'ADN, et notamment la taille des séquences semble elle aussi conférer une meilleure stabilité à l'ADN extracellulaire. En effet, selon une étude d'Ogram *et al.*, en 1988, la masse moléculaire des séquences d'ADN serait directement corrélée à leur adsorption par le sol. Au sein du RET et dans nos conditions d'études, l'ADNex n'est pas en contact avec le sol mais avec les autres molécules du RET. Or, selon Douarch 2007, l'ADN peut être adsorbé par une surface chargée positivement. Il est donc envisageable que l'ADNex puisse se lier à des composants du RET chargés positivement (comme les histones, Dangl *et al.*, 2001), améliorant ainsi sa conservation. Cependant, il n'existe encore aucune donnée concernant la taille des séquences d'ADNex en fonction de leur origine (chromosomique ou organelles) qui pourrait permettre d'expliquer une meilleure conservation de l'ADN des organelles par rapport à l'ADN chromosomique.

Il est aussi envisageable que l'ADNex ait été fragmenté suite à une lyse cellulaire des AC-DC ou d'autres cellules racinaires. Dans ce cas, la taille et la conservation des séquences d'ADNex dépendraient donc de l'action de DNAses sur celui-ci. Or, certaines DNAses seraient impactées négativement par la méthylation de l'ADN (Böckelmann *et al.*, 2006). Une différence de méthylation pourrait donc expliquer une différence de conservation de ces ADN.

Si l'on observe les données de la littérature concernant la méthylation de l'ADN des plantes, on peut constater que la méthylation semble être espèce-spécifique (Fulnecek *et al.*, 2002) mais aussi organelle-spécifique. En effet la méthylation de l'ADNchr et de l'ADN des organelles n'est pas la même (Vanyushin, 2006). Deux types principaux de méthylation sont retrouvés chez les plantes : la méthylation m^5C (5-méthylcytosine) et m^6A (N^6 -methyladenine). Or, il a été montré que seule la méthylation de type m^6A est retrouvée dans l'ADN mitochondrial des plantes. Les deux types de méthylation (m^6A et m^5C) ont cependant été retrouvés dans l'ADN des plastes. Il est donc envisageable qu'une différence de

méthylation ait entraîné une moins bonne conservation de l'ADN chromosomique du soja comparé à son ADN plastidial et mitochondrial. Ainsi il est alors envisageable que la conservation de l'ADNex varie en fonction de l'espèce végétale considérée.

Une autre hypothèse quant à la meilleure couverture de l'ADN mitochondrial et plastidial par rapport à l'ADN chromosomique serait celle d'une origine mitochondriale et plastidial de l'ADNex. En effet, il a été montré dans la bibliographie que le NET (*Neutrophil Extracellular Trap*), structure retrouvée chez les mammifères et proche du RET des plantes (Driouich et al., 2019), pouvait être formé à partir d'ADN d'origine mitochondriale (Yousefi et al., 2009 ; Lood et al., 2016). L'élicitation de neutrophiles avec certaines molécules (ex : *granulocyte/macrophage colony-stimulating factor GM-CSF* ; *toll-like receptor 4 TLR4*) permet la formation d'un piège extracellulaire contenant de l'ADN mitochondrial (et ne contenant pas d'ADN chromosomique). La formation de ce NET ne nécessitant pas la mort cellulaire des neutrophiles (vital NETosis ; Yipp et Kubes 2013). Il est donc possible d'envisager une éventuelle libération programmée de l'ADN des organites à l'origine de l'ADNex du RET. Cette hypothèse impliquerait donc une libération d'ADNm et d'ADNp de façon régulière et n'ayant pas été lysé suite à la mort cellulaire, ce qui expliquerait donc la bonne conservation et couverture de l'ADNm et de l'ADNp. La présence d'ADN chromosomique de moins bonne qualité et avec une moins bonne couverture, s'expliquerait par la lyse des cellules de la racine et des AC-DC.

De plus, cette hypothèse permettrait d'expliquer les résultats obtenus par Wen et al., en 2009. En effet, ceux-ci avaient observé que l'ADNex libéré dans le RET du pois semblait nouvellement synthétisé (utilisation de dCTP marqué au ³²P). Or, la division des plastes n'est pas dépendante du cycle cellulaire (Solymosi et al., 2018), et celle-ci semble pouvoir être réalisée de façon régulière (de 1 à 4h chez le blé selon Boffey et Lloyd, 1988), grâce à des mécanismes de régulation spécifique (Okazaki et al., 2009 ; Chang et al., 2017).

Des données préliminaires pouvant appuyer cette hypothèse ont été obtenues au laboratoire (Durambur 2020, données non publiées). En effet une coloration au lugol (colorant de l'amidon) a été réalisée sur le RET de soja en condition élicitée avec PEP-13 ou non (Figure 68). Suite à cette expérience, il semblerait que la quantité d'amyloplastes au sein des cellules de l'apex racinaire augmente en réponse à PEP-13 (expérience réalisée sur une plante de chaque condition). Une augmentation de la quantité d'amyloplastes dans l'apex

racinaire lors de l'élicitation pourrait donc permettre une libération plus importante d'ADN plastidial (Solymosi *et al.*, 2018) provenant des amyloplastes au sein du RET et donc potentiellement aussi expliquer sa présence plus importante comparé à l'ADN chromosomique.

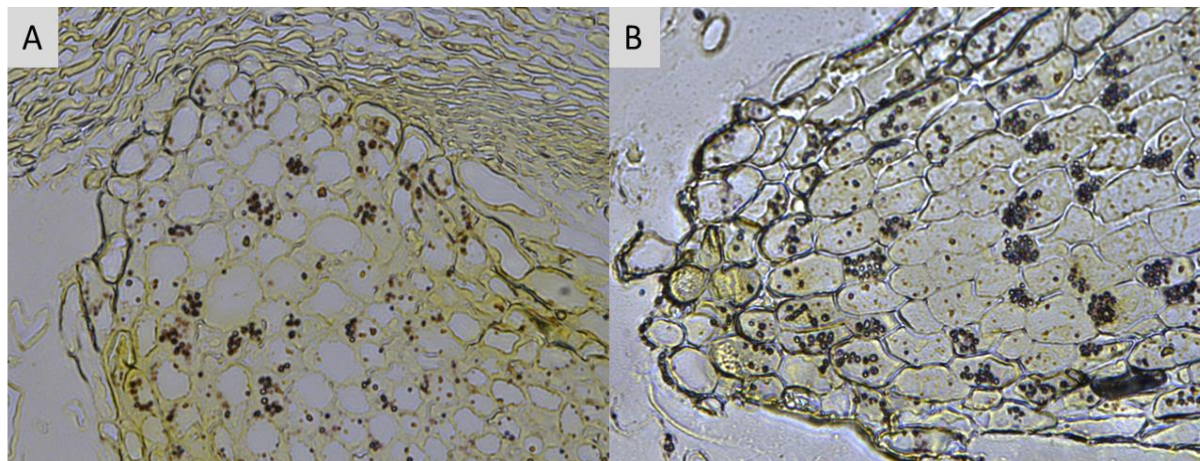


Figure 68 : Coloration de l'apex racinaire de soja au lugol en conditions élicitée avec PEP-13 et témoin non élicitée

A : Apex racinaire en condition témoin non élicitée, on observe des grains d'amidons contenus dans les amyloplastes (grains violacés). B : Apex racinaire en condition élicitée avec PEP-13, un plus grand nombre d'amyloplastes semblent être détectés dans les cellules de l'apex. Coloration réalisée sur une plante pour chaque condition. Durambur 2020, données non publiées.

Une seconde différence a été observée lors de l'alignement des *reads* d'ADNex sur le génome du soja : la formation de pics dans l'alignement des *reads* sur les séquences ADN des chromosomes.

II-2 Différences dans l'alignement des *reads* sur l'ADNchr

L'alignement des *reads* d'ADNex sur l'ADN chromosomique a montré une très faible couverture de chacun des chromosomes. Cependant certaines zones de l'ADNchr présentent des pics du nombre d'alignement de *reads*, ces pics pouvant atteindre plusieurs centaines de *counts*. Il semble donc que certaines zones de l'ADNchr soient plus conservées au sein de l'ADNex. Parmi les zones pour lesquelles les pics sont les plus importants (présents dans les six échantillons et toujours supérieurs à 100 *counts*), on retrouve des séquences spécifiques aux Fabacées, trois séquences correspondant à des séquences impliquées dans la réponse de défense de la plante, ainsi que de nombreuses séquences répétées.

Une première explication de la présence de ces pics est que l'ensemble de ces séquences pourraient être répétées dans le génome de la plante. En effet la variété utilisée dans cette

étude (CASTETIS) est différente de la variété utilisée comme génome de référence du soja (Williams 82, Schmutz *et al.*, 2010). Il est donc possible que la variété CASTETIS possède dans son génome davantage de séquences répétées par rapport à ce génome de référence. Une méthode reconnue dans la détection de variants ayant subi une duplication est la détection de la variation de la profondeur (c'est-à-dire du nombre de *reads*) du séquençage d'une partie du génome (Wala *et al.*, 2018). La présence de pics, traduisant une augmentation locale de la profondeur du séquençage, pourrait donc être le résultat de la présence de répétitions dans le génome du soja de la variété CASTETIS. De plus, la présence de gènes de défense parmi ces séquences répétées pourrait être expliquée par la sélection variétale à l'origine de l'obtention de la variété CASTETIS. En effet, lors de la sélection variétale, les critères comme la résistance aux pathogènes, la sensibilité à la verse, la précocité, ou la teneur en protéines permettent de déterminer si une variété est intéressante ou non. La duplication d'une séquence codant pour un acteur de la défense de la plante face à un pathogène pourrait permettre à celle-ci d'acquérir une meilleure résistance aux stress biotiques grâce notamment au phénomène de sub-fonctionnalisation (Hurles *et al.*, 2004 ; Lefevre *et al.*, 2016), qui permet de diminuer la pléiotropie d'un gène, et serait donc un atout pour la plante.

Une seconde explication à la présence de ces pics serait que ces séquences en particulier pourraient être mieux protégées que le reste de l'ADNchr, par exemple grâce à une méthylation de ces zones (Böckelmann *et al.*, 2006). En effet, l'ADN génomique de la plante est reconnu comme un DAMPs, et celle-ci est capable de différencier son propre ADN de celui d'une autre espèce (Duran-Flores et Heil, 2017). Or, parmi les séquences identifiées, certaines semblent spécifiques aux Fabacées ou même spécifiques à *Glycine max* (Tableau 19). Ces séquences pourraient donc être reconnues comme des DAMPs par les AC-DC au sein du RET, et donc initier une réponse de défense des AC-DC (et de la racine), leur conférant ainsi une meilleure défense en cas d'attaque par un pathogène. L'ADNex pourrait donc avoir une fonction d'éliciteur des AC-DC au sein du RET (et des autres cellules de la racine), grâce à ces séquences spécifiques.

Conclusion et perspectives

Conclusion et perspectives

L'objectif de cette étude était de mieux comprendre le RET et sa réponse de défense. Pour cela, nous avons réalisé dans un premier temps une étude du transcriptome et du protéome des AC-DC et des racines en condition élicitée avec PEP-13 ou non. Cette étude a permis de mettre en évidence une réponse à PEP-13 des AC-DC et des racines. De plus, une réponse différente a été observée entre ces deux zones. En effet, la réponse des racines pourrait conduire à une augmentation de la différenciation des AC-DC, et les AC-DC semblent mettre en place le renforcement de leur paroi et le remodelage de leur mucilage ainsi que la formation de DAMPs et l'activation de la PTI.

Lors de cette étude, les AC-DC du RET de soja ont été prélevées afin d'en réaliser le transcriptome. Cependant, il a été montré qu'il existait trois différents morphotypes d'AC-DC au sein du RET de soja (Ropitiaux, 2018). Ces différents morphotypes pourraient donc avoir des rôles distincts et donc des réponses propres. Afin d'affiner les résultats obtenus, il serait donc intéressant de réaliser le transcriptome de chacun de ces morphotypes, grâce à une technique de *single-cell* RNAseq. Des tests de tris des AC-DC ainsi que d'extraction de leur ARN ont déjà été initiés par le laboratoire Glyco-MEV en collaboration avec la plateforme ProfileXpert et l'entreprise Cellenion dans le cadre du projet collaboratif d'innovation DRAGoNE. Grâce à cette étude, le rôle de chacun des morphotypes d'AC-DC au sein du RET de soja pourra être étudié en condition contrôle et en condition d'élicitation par PEP-13.

L'étude protéomique réalisée sur les AC-DC et les racines de soja nous a permis d'identifier un total de 1 535 protéines, ce qui semble assez peu. En comparaison, une étude récente du protéome de racine de soja réalisée par Wang *et al.*, en 2020 a permis l'identification d'un total de 6 640 protéines grâce à une méthode de labelling TMT (*tandem mass tag*) réalisée à partir de 100 µg de protéines, suivie d'une analyse HPLC LC-MS/MS. Une première différence ayant pu influencer sur la quantité de protéines identifiées est la quantité de protéines utilisées. Wang *et al.*, ont en effet utilisé 100 µg de protéines ; contre, dans notre étude, 30 µg de protéines. De plus, Wang *et al.*, ont ensuite fractionné leurs échantillons par HPLC avant l'analyse LC-MS/MS. Afin d'améliorer les prochaines études protéomiques du soja, une augmentation de la quantité de protéines totales utilisées, ainsi

que l'ajout d'une étape de fractionnement par HPLC pourraient éventuellement être envisagées.

Une tentative d'identification du récepteur de PEP-13 a été réalisée dans cette étude. Pour cela, des domaines récepteurs ont été identifiés grâce à DRAGO2 parmi les gènes sur-exprimés en réponse à PEP-13 (FC supérieur à 1,5 et *p-value* inférieure ou égale à 0,05) et étant non annotés ou annotés comme récepteurs. Cette analyse a permis de détecter 20 gènes dont les transcrits pourraient correspondre à un récepteur de PEP-13. Afin de déterminer si l'un de ces gènes correspond à un récepteur de PEP-13, une modélisation de la structure des transcrits grâce à l'outil bio-informatique I-TASSER (Zhang lab) a été initiée. Une première analyse a été réalisée sur le récepteur potentiel possédant le FC le plus important (GLYMA20G100500). L'analyse réalisée semble montrer que celui-ci ne correspond pas au récepteur de PEP-13. Il s'agit donc de continuer cette analyse sur les 19 autres gènes restants, d'élargir l'étude aux domaines non reconnus par DRAGO2 voire de faire du docking moléculaire à partir de PEP-13.

Enfin, si cette analyse permet d'observer que l'un de ces récepteurs semble pouvoir se lier à PEP-13, alors la réalisation d'un mutant ainsi qu'une analyse de celui-ci ; comme réalisé précédemment pour l'identification de récepteurs tels que FLS2 (*flagelin-sensing 2*) ou EFR (*elongation factor TU receptor*) (Gomez-Gomez et Boller, 2000 ; Zipfel *et al.*, 2006), pourrait être envisagée afin de confirmer si celui-ci permet bien à la plante de détecter PEP-13.

Dans un second temps, l'objectif de cette étude était de mieux comprendre le rôle de l'ADNex du RET, ainsi que d'étudier son origine. Pour cela, l'ADNex a été extrait et séquencé en condition élicitee avec PEP-13 ou témoin. Suite à cette étude, nous avons pu observer qu'il y avait très peu de différences entre la condition élicitee et non élicitee, ainsi que de fortes différences d'alignement des *reads* d'ADNex entre les séquences d'ADN chromosomique et les séquences d'ADN plastidial et mitochondrial. Deux hypothèses quant à l'origine de l'ADNex ont donc été proposées : (i) l'ADNex pourrait provenir de la lyse cellulaire des AC-DC et des autres cellules de la racine, cependant l'ADN des organites serait mieux conservé au sein du RET que l'ADN chromosomique et (ii) l'ADNex pourrait être libéré

à partir des organites des AC-DC et/ou des cellules racinaires, et l'ADN chromosomique extracellulaire proviendrait de la lyse des cellules du RET les plus anciennes.

Afin d'explorer ces hypothèses, une première méthode envisageable serait la mise au point d'une technique microscopique d'imagerie cellulaire en direct (*live cell imaging*), qui permettrait de suivre la libération de l'ADNex du RET, comme cela a déjà été fait pour le NET (Van der Linden *et al.*, 2017). Cette technique permettrait de visualiser directement la libération d'ADNex, soit grâce à un mécanisme de libération de l'ADN des organites, soit par lyse cellulaire.

Une seconde méthode envisageable serait de provoquer la mort cellulaire d'un échantillon d'AC-DC et de cellules de la coiffe (au préalable débarrassées de l'ADNex du RET), et de comparer suite à cela l'ADN extracellulaire obtenu à l'ADNex du RET séquencé dans cette étude. En effet, si l'ADNex provient uniquement de la lyse cellulaire des AC-DC et des cellules de l'apex racinaire, alors le même profil d'alignement devrait être observé.

De plus, afin de vérifier si l'ADN des organites est mieux conservé grâce à une différence de méthylation entre l'ADN des organites et l'ADN chromosomique, une étude de la méthylation de l'ADN des AC-DC et de l'ADNex pourrait être envisagée. En effet il a été montré chez *Arabidopsis* par Kawakatsu *et al.*, en 2016 que l'ADN des cellules de la coiffe racinaire et notamment de la columelle, est l'ADN présentant le plus fort taux de méthylation observé chez *Arabidopsis*, tous tissus étudiés confondus. Les AC-DC provenant de la coiffe pourraient donc présenter une méthylation très importante, pouvant être différente entre les organites et l'ADN chromosomique.

Enfin, il serait intéressant de réaliser des tests d'élicitation du soja avec de l'ADNex, et notamment avec les séquences d'ADN chromosomiques correspondant aux pics détectés dans l'alignement des *reads* d'ADNex sur la séquence de l'ADN chromosomique. En effet, si celles-ci sont préférentiellement conservées dans le RET, alors elles pourraient avoir un rôle spécifique, comme celui d'être détectées comme des DAMPs (Duran-Flores et Heil, 2017), et participerait donc à l'élicitation de la plante. De plus, il serait intéressant de tenter d'identifier le récepteur de l'ADNex, cette identification pourrait être réalisée dans un premier temps à partir du transcriptome réalisé dans cette étude, puis à partir du transcriptome des AC-DC et des racines en condition d'élicitation avec l'ADNex.

Références bibliographiques

A

Abdul Malik, N.A., Kumar, I.S., Nadarajah, K., 2020. Elicitor and Receptor Molecules: Orchestrators of Plant Defense and Immunity. *IJMS* 21, 963.

Alam, I., Sharmin, S.A., Kim, K.-H., Yang, J.K., Choi, M.S., Lee, B.-H., 2010. Proteome analysis of soybean roots subjected to short-term drought stress. *Plant Soil* 333, 491–505.

Albert, M., 2013. Peptides as triggers of plant defence. *Journal of Experimental Botany* 64, 5269–5279.

Alekcevetch, J.C., Passianotto, A.L. de L., Ferreira, E.G.C., dos Santos, A.B., da Silva, D.C.G., Dias, W.P., Belzile, F., Abdelnoor, R.V., Marcelino-Guimarães, F.C., 2019. Genome-Wide Association Study for Resistance to the Meloidogyne javanica Causing Root-Knot Nematode in Soybean. Preprints 2019, 2019120093

Ali, A., Alexandersson, E., Sandin, M., Resjö, S., Lenman, M., Hedley, P., Levander, F., Andreasson, E., 2014. Quantitative proteomics and transcriptomics of potato in response to *Phytophthora infestans* in compatible and incompatible interactions. *BMC Genomics* 15, 497.

Alkharouf, N.W., Klink, V.P., Chouikha, I., Beard, H., MacDonald, M.H., Meyer, S., Knap, H., Khan, R., Matthews, B.F., 2006. Timecourse microarray analyses reveal global changes in gene expression of susceptible *Glycine max* (soybean) roots during infection by *Heterodera glycines* (soybean cyst nematode). *Planta* 224, 838–852.

Allesen-Holm, M., Barken, K.B., Yang, L., Klausen, M., Webb, J.S., Kjelleberg, S., Molin, S., Givskov, M., Tolker-Nielsen, T., 2006. A characterization of DNA release in *Pseudomonas aeruginosa* cultures and biofilms. *Molecular Microbiology* 59, 1114–1128.

Alves, M.S., Soares, Z.G., Vidigal, P.M.P., Barros, E.G., Poddanosqui, A.M.P., Aoyagi, L.N., Abdelnoor, R.V., Marcelino-Guimarães, F.C., Fietto, L.G., 2015. Differential expression of four

soybean bZIP genes during *Phakopsora pachyrhizi* infection. *Funct Integr Genomics* 15, 685–696.

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.

Aoyagi, L.N., Lopes-Caitar, V.S., de Carvalho, M.C.C.G., Darben, L.M., Polizel-Podanosqui, A., Kuwahara, M.K., Nepomuceno, A.L., Abdelnoor, R.V., Marcelino-Guimarães, F.C., 2014. Genomic and transcriptomic characterization of the transcription factor family R2R3-MYB in soybean and its involvement in the resistance responses to *Phakopsora pachyrhizi*. *Plant Science* 229, 32–42.

Arnaud, C., Bonnot, C., Desnos, T., Nussaume, L., 2010. The root cap at the forefront. *Comptes Rendus Biologies* 333, 335–343.

Ashiuchi, M., Oike, S., Hakuba, H., Shibatani, S., Oka, N., Wakamatsu, T., 2015. Rapid purification and plasticization of d-glutamate-containing poly- γ -glutamate from Japanese fermented soybean food natto. *Journal of Pharmaceutical and Biomedical Analysis* 116, 90–93.

Aulik, N.A., Hellenbrand, K.M., Czuprynski, C.J., 2012. *Mannheimia haemolytica* and Its Leukotoxin Cause Macrophage Extracellular Trap Formation by Bovine Macrophages. *Infect. Immun.* 80, 1923–1933.

Azevedo Souza, C., Li, S., Lin, A.Z., Boutrot, F., Grossmann, G., Zipfel, C., Somerville, S.C., 2017. Cellulose-Derived Oligomers Act as Damage-Associated Molecular Patterns and Trigger Defense-Like Responses. *Plant Physiol.* 173, 2383–2398.

B

Bai, L., Sun, H.-B., Liang, R.-T., Cai, B.-Y., 2019. iTRAQ Proteomic Analysis of Continuously Cropped Soybean Root Inoculated With *Funneliformis mosseae*. *Front. Microbiol.* 10, 61.

Bais, H.P., Weir, T.L., Perry, L.G., Gilroy, S., Vivanco, J.M., 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu. Rev. Plant Biol.* 57, 233–266.

Bandara, A.Y., Weerasooriya, D.K., Bradley, C.A., Allen, T.W., Esker, P.D., 2020. Dissecting the economic impact of soybean diseases in the United States over two decades. *PLoS ONE* 15, e0231141.

Bange, G., Altegoer, F., 2019. Plants strike back: Kiwellin proteins as a modular toolbox for plant defense mechanisms. *Communicative & Integrative Biology* 12, 31–33.

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., Klopp, C., 2014. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* 15, 293.

Barilli, E., Rubiales, D., Amalfitano, C., Evidente, A., Prats, E., 2015. BTH and BABA induce resistance in pea against rust (*Uromyces pisi*) involving differential phytoalexin accumulation. *Planta* 242, 1095–1106.

Batzer, J.C., Kandel, Y.R., Bradley, C.A., Chilvers, M.I., Tenuta, A.U., Wise, K.A., Hernández, E., Mueller, D.S., 2016. Effect of Seed Treatment on Early Season Brown Spot Caused by *Septoria glycines* of Soybean. *Plant Health Progress* 17, 223–228.

Baum, M.M., Kainović, A., O’Keeffe, T., Pandita, R., McDonald, K., Wu, S., Webster, P., 2009. Characterization of structures in biofilms formed by a *Pseudomonas fluorescens* isolated from soil. *BMC Microbiol* 9, 103.

Benfey, P.N., Scheres, B., 2000. Root development. *Current biology* 10, 22.

Bernier, F., Berna, A., 2001. Germins and germin-like proteins: Plant do-all proteins. But what do they do exactly? *Plant Physiology and Biochemistry* 39, 545–554.

Böckelmann, U., Janke, A., Kuhn, R., Neu, T.R., Wecke, J., Lawrence, J.R., Szewzyk, U., 2006. Bacterial extracellular DNA forming a defined network-like structure. *FEMS Microbiology Letters* 262, 31–38.

Boeltz, S., Amini, P., Anders, H.-J., Andrade, F., Bilyy, R., Chatfield, S., Cichon, I., Clancy, D.M., Desai, J., Dumych, T., Dwivedi, N., Gordon, R.A., Hahn, J., Hidalgo, A., Hoffmann, M.H., Kaplan, M.J., Knight, J.S., Kolaczowska, E., Kubes, P., Leppkes, M., Manfredi, A.A., Martin, S.J., Maueröder, C., Maugeri, N., Mitroulis, I., Munoz, L.E., Nakazawa, D., Neeli, I., Nizet, V., Pieterse, E., Radic, M.Z., Reinwald, C., Ritis, K., Rovere-Querini, P., Santocki, M., Schauer, C., Schett, G., Shlomchik, M.J., Simon, H.-U., Skendros, P., Stojkov, D., Vandenabeele, P., Berghe,

T.V., Vlag, J. van der, Vitkov, L., Köckritz-Blickwede, M. von, Yousefi, S., Zarbock, A., Herrmann, M., 2019. To NET or not to NET: current opinions and state of the science regarding the formation of neutrophil extracellular traps. *Cell Death & Differentiation* 26, 395–408.

Boffey, S.A., Lloyd, D., 1988. *Division and Segregation of Organelles*. Cambridge University Press.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

Boller, T., 2005. Peptide signalling in plant development and self/non-self perception. *Current Opinion in Cell Biology* 17, 116–122.

Boller, T., Felix, G., 2009. A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. *Annual Review of Plant Biology* 60, 379–406.

Bonaventure, G., VanDoorn, A., Baldwin, I.T., 2011. Herbivore-associated elicitors: FAC signaling and metabolism. *Trends in Plant Science* 16, 294–299.

Bonhoff, A., Rieth, B., Golecki, J., Grisebach, H., 1987. Race cultivar-specific differences in callose deposition in soybean roots following infection with *Phytophthora megasperma* f. sp. *glycinea*. *Planta* 172, 101–105.

Brigham, L.A., Woo, H.H., Hawes, M.C., 1995. Root Border Cells as Tools in Plant Cell Studies. *Methods in Cell Biology* 49, 377–387.

Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D.S., Weinrauch, Y., Zychlinsky, A., 2004. Neutrophil Extracellular Traps Kill Bacteria. *Science* 303, 1532–1535.

Brinkmann, V., Zychlinsky, A., 2012. Neutrophil extracellular traps: Is immunity the second function of chromatin? *The Journal of Cell Biology* 198, 773–783.

Brunner, F., 2002. Pep-13, a plant defense-inducing pathogen-associated pattern from *Phytophthora* transglutaminases. *The EMBO Journal* 21, 6681–6688.

C

Cannesan, M.A., Durand, C., Burel, C., Gangneux, C., Lerouge, P., Ishii, T., Laval, K., Follet-Gueye, M.-L., Driouich, A., Vicré-Gibouin, M., 2012. Effect of Arabinogalactan Proteins from the Root Caps of Pea and Brassica napus on Aphanomyces euteiches Zoospore Chemotaxis and Germination. *PLANT PHYSIOLOGY* 159, 1658–1670.

Cannesan, M.A., Gangneux, C., Lanoue, A., Giron, D., Laval, K., Hawes, M., Driouich, A., Vicré-Gibouin, M., 2011. Association between border cell responses and localized root infection by pathogenic Aphanomyces euteiches. *Annals of Botany* 108, 459–469.

Castilleux, R., Plancot, B., Ropitiaux, M., Carreras, A., Leprince, J., Boulogne, I., Follet-Gueye, M.-L., Popper, Z.A., Driouich, A., Vicré, M., 2018. Cell wall extensins in root–microbe interactions and root secretions. *Journal of Experimental Botany* 69, 4235–4247.

Chaliha, C., Rugen, M.D., Field, R.A., Kalita, E., 2018. Glycans as Modulators of Plant Defense Against Filamentous Pathogens. *Front. Plant Sci.* 9, 928.

Chang, C., Xu, S., Tian, L., Shi, S., Nasir, F., Chen, D., Li, X., Tian, C., 2019. Connection the Rhizomicrobiome and Plant MAPK Gene 12. *Plant Pathol.*, 35(6), 623-634.

Chang, M., Clinton, M., Liu, F., Fu, Z.Q., 2019. NAD⁺ Cleavage: TIR Domain-Containing Resistance Proteins in Action. *Trends in Plant Science* 24, 1069–1072.

Chang, N., Sun, Q., Li, Y., Mu, Y., Hu, J., Feng, Y., Liu, X., Gao, H., 2017. PDV2 has a dosage effect on chloroplast division in Arabidopsis. *Plant Cell Rep* 36, 471–480.

Chang, S., Wang, Y., Lu, J., Gai, J., Li, J., Chu, P., Guan, R., Zhao, T., 2013. The Mitochondrial Genome of Soybean Reveals Complex Genome Structures and Gene Evolution at Intercellular and Phylogenetic Levels. *PLoS ONE* 8, e56502.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J.D.G., Felix, G., Boller, T., 2007. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* 448, 497–500.

Choi, H.W., Klessig, D.F., 2016. DAMPs, MAMPs, and NAMPs in plant innate immunity. *BMC Plant Biol* 16, 232.

Christensen, A.B., Thordal-Christensen, H., Zimmermann, G., Gjetting, T., Lyngkjær, M.F., Dudler, R., Schweizer, P., 2004. The Germinlike Protein GLP4 Exhibits Superoxide Dismutase Activity and Is an Important Component of Quantitative Resistance in Wheat and Barley. *MPMI* 17, 109–117.

Chuberre, C., Plancot, B., Driouich, A., Moore, J.P., Bardor, M., Gügi, B., Vitré, M., 2018. Plant Immunity Is Compartmentalized and Specialized in Roots. *Frontiers in Plant Science* 9, 1692.

Claverie, J., Balacey, S., Lemaître-Guillier, C., Brulé, D., Chiltz, A., Granet, L., Noirot, E., Daire, X., Darblade, B., Héloir, M.-C., Poinssot, B., 2018. The Cell Wall-Derived Xyloglucan Is a New DAMP Triggering Plant Immunity in *Vitis vinifera* and *Arabidopsis thaliana*. *Front. Plant Sci.* 9, 1725.

Cosgrove, D.J., 2015. Plant expansins: diversity and interactions with plant cell walls. *Current Opinion in Plant Biology* 25, 162–172.

Cosgrove, D.J., 2005. Growth of the plant cell wall. *Nat Rev Mol Cell Biol* 6, 850–861.

Cosgrove, D.J., 2000. Loosening of plant cell walls by expansins. *Nature* 407, 321–326.

D

D'Ambrosio, J.M., Couto, D., Fabro, G., Scuffi, D., Lamattina, L., Munnik, T., Andersson, M.X., Álvarez, M.E., Zipfel, C., Laxalt, A.M., 2017. Phospholipase C2 Affects MAMP-Triggered Immunity by Modulating ROS Production. *Plant Physiol.* 175, 970–981.

Daxberger, A., Nema, A., Mithöfer, A., Fliegmann, J., Ligterink, W., Hirt, H., Ebel, J., 2007. Activation of members of a MAPK module in β -glucan elicitor-mediated non-host resistance of soybean. *Planta* 225, 1559–1571.

de Bont, C.M., Koopman, W.J.H., Boelens, W.C., Pruijn, G.J.M., 2018. Stimulus-dependent chromatin dynamics, citrullination, calcium signalling and ROS production during NET formation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1865(11) 1621-1629.

Delamuta, J.R.M., Ribeiro, R.A., Ormeno-Orrillo, E., Melo, I.S., Martinez-Romero, E., Hungria, M., 2013. Polyphasic evidence supporting the reclassification of *Bradyrhizobium*

japonicum group Ia strains as *Bradyrhizobium diazoefficiens* sp. nov. INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY 63, 3342–3351.

Delgado-Rizo, V., Martínez-Guzmán, M.A., Iñiguez-Gutierrez, L., García-Orozco, A., Alvarado-Navarro, A., Fafutis-Morris, M., 2017. Neutrophil Extracellular Traps and Its Implications in Inflammation: An Overview. *Frontiers in Immunology* 8, 81.

Diekmann, W., Herkt, B., Low, Philip S., Nürnberger, T., Scheel, D., Terschüren, C., Robinson, David G., 1994. Visualization of elicitor-binding loci at the plant cell surface. *Planta* 195, 126-137.

Dixon, R.A., 2001. Natural products and plant disease resistance. *Nature* 411, 843–847.

Dolan, L., Janmaat, K., Willemsen, V., Linstead, P., Poethig, S., Roberts, K., Scheres, B., 1993. Cellular organisation of the *Arabidopsis thaliana* root. *Development* 119, 71-84.

Dong, H., Tan, J., Li, M., Yu, Y., Jia, S., Zhang, C., Wu, Y., Liu, Y., 2019. Transcriptome analysis of soybean WRKY TFs in response to *Peronospora manshurica* infection. *Genomics* 111, 1412–1422.

Dong, S., Wang, W., Jiang, Y., Ma, Z., Yan, C., Liu, L., Cui, G., 2019. Antioxidant and Proteomic Analysis of Soybean Response to Drought during Soybean Flowering. *Ekoloji* 28, 2041-2052.

Douarche, C., 2007. Étude de l'adsorption de l'ADN simple brin et double brin aux interfaces. Université de Lille 1.

Doyle, J.J., Doyle, J.L., Rauscher, J.T., Brown, A.H.D., 2003. Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus *Glycine*): Research review. *New Phytologist* 161, 121–132.

Driouich, A., Follet-Gueye, M.-L., Vitré-Gibouin, M., Hawes, M., 2013. Root border cells and secretions as critical elements in plant host defense. *Current Opinion in Plant Biology* 16, 489–495.

Driouich, A., Smith, C., Ropitiaux, M., Chambard, M., Boulogne, I., Bernard, S., Follet-Gueye, M., Vitré, M., Moore, J., 2019. Root extracellular traps versus neutrophil extracellular traps in host defence, a case of functional convergence? *Biological Reviews* 94, 1685-1700.

Dumont, M., Lehner, A., Bouton, S., Kiefer-Meyer, M.C., Voxeur, A., Pelloux, J., Lerouge, P., Mollet, J.-C., 2014. The cell wall pectic polymer rhamnogalacturonan-II is required for proper pollen tube elongation: implications of a putative sialyltransferase-like protein. *Annals of Botany* 114, 1177–1188.

Duran-Flores, D., Heil, M., 2017. Extracellular self-DNA as a damage-associated molecular pattern (DAMP) that triggers self-specific immunity induction in plants. *Brain, Behavior, and Immunity* 72, 78-88.

Dworski, R., Simon, H.-U., Hoskins, A., Yousefi, S., 2011. Eosinophil and neutrophil extracellular DNA traps in human allergic asthmatic airways. *Journal of Allergy and Clinical Immunology* 127, 1260–1266.

E

Egamberdieva, D., Kamilova, F., Validov, S., Gafurova, L., Kucharova, Z., Lugtenberg, B., 2008. High incidence of plant growth-stimulating bacteria associated with the rhizosphere of wheat grown on salinated soil in Uzbekistan. *Environ Microbiol* 10, 1-9.

Elazezy, M., Joosse, S.A., 2018. Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management. *Computational and Structural Biotechnology Journal* 16, 370–378.

Endo, I., Tange, T., Osawa, H., 2011. A cell-type-specific defect in border cell formation in the *Acacia mangium* root cap developing an extraordinary sheath of sloughed-off cells. *Annals of Botany* 108, 279–290.

Ewels, P., Magnusson, M., Lundin, S., Källner, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048.

F

Fenner, K., Canonica, S., Wackett, L.P., Elsner, M., 2013. Evaluating Pesticide Degradation in the Environment: Blind Spots and Emerging Opportunities. *Science* 341, 752–758.

Ferrari, S., 2013. Oligogalacturonides: plant damage-associated molecular patterns and regulators of growth and development. *Frontiers in Plant Science* 4, 49.

Fulneček, J., Matyasek, R., Kovarik, A., 2002. Distribution of 5-methylcytosine residues in 5S rRNA genes in *Arabidopsis thaliana* and *Secale cereale*. *Mol Genet Genomics* 268, 510–517.

G

Galloway, A.F., Pedersen, M.J., Merry, B., Marcus, S.E., Blacker, J., Benning, L.G., Field, K.J., Knox, J.P., 2017. Xyloglucan is released by plants and promotes soil particle aggregation. *New Phytologist* 217, 1128-1136.

Gallucci, S., Maffei, M.E., 2017. DNA Sensing across the Tree of Life. *Trends in Immunology* 1417, 1-14.

Galluzzi, L., Vitale, I., Aaronson, S.A., Abrams, J.M., Adam, D., Agostinis, P., Alnemri, E.S., Altucci, L., Amelio, I., Andrews, D.W., Annicchiarico-Petruzzelli, M., Antonov, A.V., Arama, E., Baehrecke, E.H., Barlev, N.A., Bazan, N.G., Bernassola, F., Bertrand, M.J.M., Bianchi, K., Blagosklonny, M.V., Blomgren, K., Borner, C., Boya, P., Brenner, C., Campanella, M., Candi, E., Carmona-Gutierrez, D., Cecconi, F., Chan, F.K.-M., Chandel, N.S., Cheng, E.H., Chipuk, J.E., Cidlowski, J.A., Ciechanover, A., Cohen, G.M., Conrad, M., Cubillos-Ruiz, J.R., Czabotar, P.E., D'Angiolella, V., Dawson, T.M., Dawson, V.L., De Laurenzi, V., De Maria, R., Debatin, K.-M., DeBerardinis, R.J., Deshmukh, M., Di Daniele, N., Di Virgilio, F., Dixit, V.M., Dixon, S.J., Duckett, C.S., Dynlacht, B.D., El-Deiry, W.S., Elrod, J.W., Fimia, G.M., Fulda, S., García-Sáez, A.J., Garg, A.D., Garrido, C., Gavathiotis, E., Golstein, P., Gottlieb, E., Green, D.R., Greene, L.A., Gronemeyer, H., Gross, A., Hajnoczky, G., Hardwick, J.M., Harris, I.S., Hengartner, M.O., Hetz, C., Ichijo, H., Jäättelä, M., Joseph, B., Jost, P.J., Juin, P.P., Kaiser, W.J., Karin, M., Kaufmann, T., Kepp, O., Kimchi, A., Kitsis, R.N., Klionsky, D.J., Knight, R.A., Kumar, S., Lee, S.W., Lemasters, J.J., Levine, B., Linkermann, A., Lipton, S.A., Lockshin, R.A., López-Otín, C., Lowe, S.W., Luedde, T., Lugli, E., MacFarlane, M., Madeo, F., Malewicz, M., Malorni, W., Manic, G., Marine, J.-C., Martin, S.J., Martinou, J.-C., Medema, J.P., Mehlen, P., Meier, P., Melino, S., Miao, E.A., Molkentin, J.D., Moll, U.M., Muñoz-Pinedo, C., Nagata, S., Nuñez, G., Oberst, A., Oren, M., Overholtzer, M., Pagano, M., Panaretakis, T., Pasparakis, M., Penninger, J.M., Pereira, D.M., Pervaiz, S., Peter, M.E., Piacentini, M., Pinton, P., Prehn, J.H.M., Puthalakath, H., Rabinovich, G.A., Rehm, M., Rizzuto, R., Rodrigues, C.M.P., Rubinsztein, D.C., Rudel, T., Ryan, K.M., Sayan, E., Scorrano, L., Shao, F., Shi, Y., Silke, J., Simon, H.-U., Sistigu,

A., Stockwell, B.R., Strasser, A., Szabadkai, G., Tait, S.W.G., Tang, D., Tavernarakis, N., Thorburn, A., Tsujimoto, Y., Turk, B., Vanden Berghe, T., Vandenabeele, P., Vander Heiden, M.G., Villunger, A., Virgin, H.W., Vousden, K.H., Vucic, D., Wagner, E.F., Walczak, H., Wallach, D., Wang, Y., Wells, J.A., Wood, W., Yuan, J., Zakeri, Z., Zhivotovsky, B., Zitvogel, L., Melino, G., Kroemer, G., 2018. Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ.* 25, 486–541.

Gaur, N., Mogalapu, S., 2018. Pests of Soybean, in: Omkar (Ed.), *Pests and Their Management*. Springer Singapore, Singapore, 137–162.

Ge, Y., Li, Y., Zhu, Y.-M., Bai, X., Lv, D.-K., Guo, D., Ji, W., Cai, H., 2010. Global transcriptome profiling of wild soybean (*Glycine soja*) roots under NaHCO₃ treatment. *BMC Plant Biol* 10, 153.

Gomez-Gomez, L., Boller, T., 2000. FLS2: An LRR Receptor-like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in Arabidopsis. *Molecular Cell* 5, 1003-1011.

Graham, M.Y., Weidner, J., Wheeler, K., Pelow, M.J., Graham, T.L., 2003. Induced expression of pathogenesis-related protein genes in soybean by wounding and the *Phytophthora sojae* cell wall glucan elicitor. *Physiological and Molecular Plant Pathology* 63, 141–149.

H

Haerizadeh, F., Singh, M.B., Bhalla, P.L., 2011. Transcriptome profiling of soybean root tips. *Functional Plant Biol.* 38, 451.

Hakim, Ullah, A., Hussain, A., Shaban, M., Khan, A.H., Alariqi, M., Gul, S., Jun, Z., Lin, S., Li, J., Jin, S., Munis, M.F.H., 2018. Osmotin: A plant defense tool against biotic and abiotic stresses. *Plant Physiology and Biochemistry* 123, 149–159.

Halim, V.A., Hunger, A., Macioszek, V., Landgraf, P., Nürnberger, T., Scheel, D., Rosahl, S., 2004. The oligopeptide elicitor Pep-13 induces salicylic acid-dependent and -independent defense reactions in potato. *Physiological and Molecular Plant Pathology* 64, 311–318.

Halverson, T.W.R., Wilton, M., Poon, K.K.H., Petri, B., Lewenza, S., 2015. DNA Is an Antimicrobial Component of Neutrophil Extracellular Traps. *PLOS Pathogens* 11, e1004593.

Hamamoto, L., Hawes, M.C., Rost, T.L., 2006. The Production and Release of Living Root Cap Border Cells is a Function of Root Apical Meristem Type in Dicotyledonous Angiosperm Plants. *Annals of Botany* 97, 917–923.

Harholt, J., Suttangkakul, A., Vibe Scheller, H., 2010. Biosynthesis of Pectin. *Plant Physiol.* 153, 384–395.

Hartman, G.L., Rupe, J.C., Sikora, E.J., Domier, L.L., Davis, J.A., Steffey, K.L., 2015. *Compendium of Soybean Diseases and Pests*, Fifth edition, The American Phytopathological Society. ed.

Hawes, M.C., Bengough, G., Cassab, G., Ponce, G., 2002. Root Caps and Rhizosphere. *Journal of Plant Growth Regulation* 21, 352–367.

Hawes, M.C., Brigham, L.A., Wen, F., Woo, H.H., Zhu, Y., 1998. Function of root border cells in plant health: Pioneers in the rhizosphere. *Annual review of phytopathology* 36, 311–327.

Hawes, M.C., Pueppke, S.G., 1986. Sloughed Peripheral Root Cap Cells: Yield from Different Species and Callus Formation from Single Cells. *American Journal of Botany* 73, 1466–1473.

Henry, G., Thonart, P., Ongena, M., 2012. PAMPs, MAMPs, DAMPs and others: an update on the diversity of plant immunity elicitors. *Biotechnol. Agron. Soc. Environ.* 16, 257-268.

Homa, J., Rorat, A., Kruk, J., Cocquerelle, C., Plytycz, B., Vandebulcke, F., 2015. Dermal exposure of *Eisenia andrei* earthworms: Effects of heavy metals on metallothionein and phytochelatin synthase gene expressions in coelomocytes: Influence of heavy metals on gene expression in *E. andrei*. *Environmental Toxicology and Chemistry* 34, 1397–1404.

Hou, S., Jamieson, P., He, P., 2018. The cloak, dagger, and shield: proteases in plant–pathogen interactions. *Biochemical Journal* 475, 2491–2509.

Hou, S., Liu, Z., Shen, H., Wu, D., 2019. Damage-Associated Molecular Pattern-Triggered Immunity in Plants. *Front. Plant Sci.* 10, 646.

Hou, S., Wang, X., Chen, D., Yang, X., Wang, M., Turrà, D., Di Pietro, A., Zhang, W., 2014. The Secreted Peptide PIP1 Amplifies Immunity through Receptor-Like Kinase 7. *PLoS Pathog* 10, e1004331.

Hurles, M., 2004. Gene Duplication: The Genomic Trade in Spare Parts. *PLoS Biol* 2, e206.

I

Ibáñez de Aldecoa, A.L., Zafra, O., González-Pastor, J.E., 2017. Mechanisms and Regulation of Extracellular DNA Release and Its Biological Roles in Microbial Communities. *Frontiers in Microbiology* 8, 1390.

J

Jabs, T., Tschöpe, M., Colling, C., Hahlbrock, K., Scheel, D., 1997. Elicitor-stimulated ion fluxes and O₂- from the oxidative burst are essential components in triggering defense gene activation and phytoalexin synthesis in parsley. *Proceedings of the National Academy of Sciences* 94, 4800–4805.

Jahan, M.A., Harris, B., Lowery, M., Infante, A.M., Percifield, R.J., Kovicich, N., 2020. Glyceollin Transcription Factor GmMYB29A2 Regulates Soybean Resistance to *Phytophthora sojae*. *Plant Physiol.* 183, 530-546.

Johnson, J.M., Thürich, J., Petutschnig, E.K., Altschmied, L., Meichsner, D., Sherameti, I., Dindas, J., Mrozinska, A., Paetz, C., Scholz, S.S., Furch, A.C.U., Lipka, V., Hedrich, R., Schneider, B., Svatoš, A., Oelmüller, R., 2018. A Poly(A) Ribonuclease Controls the Cellotriose-Based Interaction between *Piriformospora indica* and Its Host *Arabidopsis*. *Plant Physiol.* 176, 2496–2514.

Jones, J.D.G., Dangl, J.L., 2006. The plant immune system. *Nature* 444, 323–329.

K

Kandath, P.K., Ithal, N., Recknor, J., Maier, T., Nettleton, D., Baum, T.J., Mitchum, M.G., 2011. The Soybean *Rhg1* Locus for Resistance to the Soybean Cyst Nematode *Heterodera glycines* Regulates the Expression of a Large Number of Stress- and Defense-Related Genes in Degenerating Feeding Cells. *PLANT PHYSIOLOGY* 155, 1960–1975.

Kang, W., Zhu, X., Wang, Y., Chen, L., Duan, Y., 2018. Transcriptomic and metabolomic analyses reveal that bacteria promote plant defense during infection of soybean cyst nematode in soybean. *BMC Plant Biology* 18, 86-100.

Kawakatsu, T., Stuart, T., Valdes, M., Breakfield, N., Schmitz, R.J., Nery, J.R., Urich, M.A., Han, X., Lister, R., Benfey, P.N., Ecker, J.R., 2016. Unique cell-type-specific patterns of DNA methylation in the root meristem. *Nature Plants* 2, 16058.

Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12, 357–360.

Kim, H.G., Kwon, S.J., Jang, Y.J., Chung, J.H., Nam, M.H., Park, O.K., 2014. GDSL lipase 1 regulates ethylene signaling and ethylene-associated systemic immunity in Arabidopsis. *FEBS Letters* 588, 1652–1658.

Klink, V.P., Overall, C.C., Alkharouf, N.W., MacDonald, M.H., Matthews, B.F., 2007. A time-course comparative microarray analysis of an incompatible and compatible response by Glycine max (soybean) to *Heterodera glycines* (soybean cyst nematode) infection. *Planta* 226, 1423–1447.

Knee, E.M., Gong, F.-C., Gao, M., Teplitski, M., Jones, A.R., Foxworthy, A., Mort, A.J., Bauer, W.D., 2001. Root Mucilage from Pea and Its Utilization by Rhizosphere Bacteria as a Sole Carbon Source. *MPMI* 14, 775–784.

Komatsu, S., Yamamoto, R., Nanjo, Y., Mikami, Y., Yunokawa, H., Sakata, K., 2009. A Comprehensive Analysis of the Soybean Genes and Proteins Expressed under Flooding Stress using Transcriptome and Proteome Techniques. *J. Proteome Res.* 8, 4766–4778.

Koroney, A.S., Plasson, C., Pawlak, B., Sidikou, R., Driouich, A., Menu-Bouaouiche, L., Vicré-Gibouin, M., 2016. Root exudate of *Solanum tuberosum* is enriched in galactose-containing molecules and impacts the growth of *Pectobacterium atrosepticum*. *Annals of Botany* 118, 797–808.

Kozlowski, H.N., Lai, E.T.L., Havugimana, P.C., White, C., Emili, A., Sakac, D., Binnington, B., Neschadim, A., McCarthy, S.D.S., Branch, D.R., 2016. Extracellular histones identified in crocodile blood inhibit in-vitro HIV-1 infection. *AIDS* 30(13), 2043–2052.

L

Lanubile, A., Muppirala, U.K., Severin, A.J., Marocco, A., Munkvold, G.P., 2015. Transcriptome profiling of soybean (*Glycine max*) roots challenged with pathogenic and non-pathogenic isolates of *Fusarium oxysporum*. *BMC Genomics* 16, 1089.

Lee, M.W., Huffaker, A., Crippen, D., Robbins, R.T., Goggin, F.L., 2018. Plant elicitor peptides promote plant defences against nematodes in soybean: Plant elicitor peptide-induced nematode resistance. *Molecular Plant Pathology* 19, 858–869.

Lefevre, T., Raymond, M., Thomas, F., 2016. *Biologie évolutive, Biologie. De Boeck supérieur.*

Lennon, J.T., 2007. Diversity and Metabolism of Marine Bacteria Cultivated on Dissolved DNA. *AEM* 73, 2799–2805.

Lerouxel, O., Cavalier, D.M., Liepman, A.H., Keegstra, K., 2006. Biosynthesis of plant cell wall polysaccharides — a complex process. *Current Opinion in Plant Biology* 9, 621–630.

Li, S., Yamada, M., Han, X., Ohler, U., Benfey, P.N., 2016. High-Resolution Expression Map of the *Arabidopsis* Root Reveals Alternative Splicing and lincRNA Regulation. *Developmental Cell* 39, 508–522

Li, J., Wang, X., 2019. Phospholipase D and phosphatidic acid in plant immunity. *Plant Science* 279, 45–50.

Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R.J., Bilgin, D.D., Radwan, O., Neece, D.J., Clough, S.J., May, G.D., Stacey, G., 2010. Complete Transcriptome of the Soybean Root Hair Cell, a Single-Cell Model, and Its Alteration in Response to *Bradyrhizobium japonicum* Infection. *Plant Physiol.* 152, 541–552.

Ligterink, W., 1997. Receptor-Mediated Activation of a MAP Kinase in Pathogen Defense of Plants. *Science* 276, 2054–2057.

Lood, C., Blanco, L.P., Purmalek, M.M., Carmona-Rivera, C., De Ravin, S.S., Smith, C.K., Malech, H.L., Ledbetter, J.A., Elkon, K.B., Kaplan, M.J., 2016. Neutrophil extracellular traps enriched in oxidized mitochondrial DNA are interferogenic and contribute to lupus-like disease. *Nat Med* 22, 146–153.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 550.

M

Ma, W., Muthreich, N., Liao, C., Franz-Wachtel, M., Schütz, W., Zhang, F., Hochholdinger, F., Li, C., 2010. The mucilage proteome of maize (*Zea mays* L.) primary roots. *J. Proteome Res.* 9, 2968–2976.

Malik, A., 2019. Purification and properties of plant chitinases: A review. *J Food Biochem* 43, e12762.

Marouf, A., Reynaud, J., 2007. *La botanique de A à Z*, Dunod. ed.

Martins, M., Uppuluri, P., Thomas, D.P., Cleary, I.A., Henriques, M., Lopez-Ribot, J.L., Oliveira, R., 2010. Presence of Extracellular DNA in the *Candida albicans* Biofilm Matrix and its Contribution to Biofilms. *Mycopathologia* 169, 323–331.

Mathesius, U., Djordjevic, M.A., Oakes, M., Goffard, N., Haerizadeh, F., Weiller, G.F., Singh, M.B., Bhalla, P.L., 2011. Comparative proteomic profiles of the soybean (*Glycine max*) root apex and differentiated root zone. *PROTEOMICS* 11, 1707–1719.

Mattoo, A.K., 2018. *The Plant Hormone Ethylene*. CRC Press. ed.

Mendes, R., Garbeva, P., Raaijmakers, J.M., 2013. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol Rev* 37, 634–663.

Micheli, F., 2001. Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends in Plant Science* 6, 414–419.

Millet, Y.A., Danna, C.H., Clay, N.K., Songnuan, W., Simon, M.D., Werck-Reichhart, D., Ausubel, F.M., 2010. Innate Immune Responses Activated in *Arabidopsis* Roots by Microbe-Associated Molecular Patterns. *Plant Cell* 22, 973–990.

Mravec, J., Guo, X., Hansen, A.R., Schückel, J., Kračun, S.K., Mikkelsen, M.D., Mouille, G., Johansen, I.E., Ulvskov, P., Domozych, D.S., Willats, W.G.T., 2017. Pea Border Cell Maturation and Release Involve Complex Cell Wall Structural Dynamics. *Plant Physiol.* 174, 1051–1066.

N

Nagler, M., Podmirseg, S.M., Griffith, G.W., Insam, H., Ascher-Jenull, J., 2018. The use of extracellular DNA as a proxy for specific microbial activity. *Appl Microbiol Biotechnol* 102, 2885–2898.

Navi, S.S., Huynh, T., Mayers, C.G., Yang, X.-B., 2019. Diversity of *Pythium* spp. associated with soybean damping-off, and management implications by using foliar fungicides as seed treatments. *Phytopathol Res* 1, 8.

Nennstiel, D., Scheel, D., Nürnberger, T., 1998. Characterization and partial purification of an oligopeptide elicitor receptor from parsley (*Petroselinum crispum*). *FEBS Letters* 431, 405–410.

Nguema Ona, E., Vire-Gibouin, M., Cannesan, M.A., Driouich, A., 2013. Arabinogalactan proteins in root–microbe interactions. *Trends in Plant Science* 18, 440-449.

Nguema-Ona, E., Vire-Gibouin, M., Gotté, M., Plancot, B., Lerouge, P., Bardor, M., Driouich, A., 2014. Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis and function. *Front. Plant Sci.* 5, 499.

Nibbering, P., Petersen, B.L., Motawia, M.S., Jørgensen, B., Ulvskov, P., Niittylä, T., 2020. Golgi-localized exo- β 1,3-galactosidases involved in AGP modification and root cell expansion in *Arabidopsis*. *BioRxiv*, 947820.

Nürnberger, T., Nennstiel, D., Jabs, T., Sacks, W.R., Hahlbrock, K., Scheel, D., 1994. High affinity binding of a fungal oligopeptide elicitor to parsley plasma membranes triggers multiple defense responses. *Cell* 78, 449–460.

O

Ogram, Andrew., Sayler, G.S., Gustin, Denise., Lewis, R.J., 1988. DNA adsorption to soils and sediments. *Environ. Sci. Technol.* 22, 982–984.

Okazaki, K., Kabeya, Y., Suzuki, K., Mori, T., Ichikawa, T., Matsui, M., Nakanishi, H., Miyagishima, S., 2009. The PLASTID DIVISION1 and 2 Components of the Chloroplast Division Machinery Determine the Rate of Chloroplast Division in Land Plant Cell Differentiation. *Plant Cell* 21, 1769–1780.

Okshevsky, M., Regina, V.R., Meyer, R.L., 2015. Extracellular DNA as a target for biofilm control. *Current Opinion in Biotechnology* 33, 73–80.

Overvoorde, P., Fukaki, H., Beeckman, T., 2010. Auxin Control of Root Development. *Cold Spring Harbor Perspectives in Biology* 2, a001537.

P

Palić, D., Ostojić, J., Andreasen, C.B., Roth, J.A., 2007. Fish cast NETs: Neutrophil extracellular traps are released from fish neutrophils. *Developmental & Comparative Immunology* 31, 805–816.

Palomaki, G.E., Deciu, C., Kloza, E.M., Lambert-Messerlian, G.M., Haddow, J.E., Neveux, L.M., Ehrich, M., van den Boom, D., Bombard, A.T., Grody, W.W., Nelson, S.F., Canick, J.A., 2012. DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. *Genet Med* 14, 296–305.

Palomaki, G.E., Kloza, E.M., Lambert-Messerlian, G.M., Haddow, J.E., Neveux, L.M., Ehrich, M., van den Boom, D., Bombard, A.T., Deciu, C., Grody, W.W., Nelson, S.F., Canick, J.A., 2011. DNA sequencing of maternal plasma to detect Down syndrome: An international clinical validation study: *Genetics in Medicine* 13, 913–920.

Pandey, S., 2019. Heterotrimeric G-Protein Signaling in Plants: Conserved and Novel Mechanisms. *Annu. Rev. Plant Biol.* 70, 213–238.

Panthee, D.R., Marois, J.J., Wright, D.L., Narváez, D., Yuan, J.S., Stewart, C.N., 2009. Differential expression of genes in soybean in response to the causal agent of Asian soybean rust (*Phakopsora pachyrhizi* Sydow) is soybean growth stage-specific. *Theoretical and Applied Genetics* 118, 359–370.

Papayannopoulos, V., 2018. Neutrophil extracellular traps in immunity and disease. *Nature Reviews Immunology* 18, 134-147.

Park, H.-J., Wang, W., Curlango-Rivera, G., Xiong, Z., Lin, Z., Huskey, D.A., Hawes, M.C., VanEtten, H.D., Turgeon, B.G., 2019. A DNase from a Fungal Phytopathogen Is a Virulence Factor Likely Deployed as Counter Defense against Host-Secreted Extracellular DNA. *mBio* 10, e02805-18.

Parker, J.E., Hahlbrock, K., Scheel, D., 1988. Different cell-wall components from *Phytophthora megasperma* f. sp. *glycinea* elicit phytoalexin production in soybean and parsley. *Planta* 176, 75-82.

Parker, J.E., Schulte, W., Hahlbrock, K., Scheel, D., 1991. An Extracellular Glycoprotein from *Phytophthora megasperma* f. sp. *glycinea* Elicits Phytoalexin Synthesis in Cultured Parsley Cells and Protoplasts. *Molecular Plant-microbe interactions* 4, 19–27.

Pathan, S.I., Arfaioli, P., Ceccherini, M.T., Ascher-Jenull, J., Pietramellara, G., 2020. Preliminary evidences of the presence of extracellular DNA single stranded forms in soil. *PLoS ONE* 15, e0227296.

Pauly, M., Gille, S., Liu, L., Mansoori, N., de Souza, A., Schultink, A., Xiong, G., 2013. Hemicellulose biosynthesis. *Planta* 238, 627–642.

Pawlowski, M.L., Bowen, C.R., Hill, C.B., Hartman, G.L., 2016. Responses of soybean genotypes to pathogen infection after the application of elicitors. *Crop Protection* 87, 78–84.

Pearce, G., Yamaguchi, Y., Barona, G., Ryan, C.A., 2010. A subtilisin-like protein from soybean contains an embedded, cryptic signal that activates defense-related genes. *Proceedings of the National Academy of Sciences* 107, 14921–14925.

Petricka, J.J., Winter, C.M., Benfey, P.N., 2012. Control of Arabidopsis Root Development. *Annu. Rev. Plant Biol.* 63, 563–590.

Phan, L.C.H.B.T., Van Dijck, P., 2019. Biosynthesis and Degradation of Trehalose and Its Potential to Control Plant Growth, Development, and (A)biotic Stress Tolerance, in: Hossain, M.A., Kumar, V., Burritt, D.J., Fujita, M., Mäkelä, P.S.A. (Eds.), *Osmoprotectant-Mediated Abiotic Stress Tolerance in Plants*. Springer International Publishing, Cham, pp. 175–199.

Pieper, J., Locke, M., Ruzaike, G., Voigt, S., Methner, U., Berndt, A., 2017. In vitro and in vivo generation of heterophil extracellular traps after Salmonella exposure. *Veterinary Immunology and Immunopathology* 188, 1–11.

Plancot, B., Santaella, C., Jaber, R., Kiefer-Meyer, M.C., Follet-Gueye, M.-L., Leprince, J., Gattin, I., Souc, C., Driouich, A., Vicre-Gibouin, M., 2013. Deciphering the Responses of Root Border-Like Cells of Arabidopsis and Flax to Pathogen-Derived Elicitors. *PLANT PHYSIOLOGY* 163, 1584–1597.

Poinssot, B., Vandelle, E., Bentéjac, M., Adrian, M., Levis, C., Brygoo, Y., Garin, J., Sicilia, F., Coutos-Thévenot, P., Pugin, A., 2003. The Endopolygalacturonase 1 from Botrytis cinerea Activates Grapevine Defense Reactions Unrelated to Its Enzymatic Activity. *MPMI* 16, 553–564.

Poncini, L., Wyrsh, I., Déneraud Tendon, V., Vorley, T., Boller, T., Geldner, N., Métraux, J.-P., Lehmann, S., 2017. In roots of Arabidopsis thaliana, the damage-associated molecular pattern AtPep1 is a stronger elicitor of immune signalling than flg22 or the chitin heptamer. *PLoS ONE* 12, e0185808.

R

Rahni, R., Birnbaum, K.D., 2019. Week-long imaging of cell divisions in the Arabidopsis root meristem. *Plant Methods* 15, 30-44.

Rahni, R., Efroni, I., Birnbaum, K.D., 2016. A Case for Distributed Control of Local Stem Cell Behavior in Plants. *Developmental Cell* 38, 635–642.

Ranjan, A., Westrick, N.M., Jain, S., Piotrowski, J.S., Ranjan, M., Kessens, R., Stiegman, L., Grau, C.R., Conley, S.P., Smith, D.L., Kabbage, M., 2019. Resistance against Sclerotinia sclerotiorum in soybean involves a reprogramming of the phenylpropanoid pathway and up-regulation of antifungal activity targeting ergosterol biosynthesis. *Plant Biotechnol J* 17, 1567–1581.

Ridley, B.L., O'Neill, M.A., Mohnen, D., 2001. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* 57, 929–967.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.

Rocha, R.O., Morais, J.K.S., Oliveira, J.T.A., Oliveira, H.D., Sousa, D.O.B., Souza, C.E.A., Moreno, F.B., Monteiro-Moreira, A.C.O., Antonino de Souza Júnior, J.D., Grossi de Sá, M.F., Vasconcelos, I.M., 2015. Proteome of Soybean Seed Exudates Contains Plant Defense-Related Proteins Active against the Root-Knot Nematode *Meloidogyne incognita*. *J. Agric. Food Chem.* 63, 5335–5343.

Ropitiaux, M., 2018. Le Root Extracellular Trap (RET), un réseau au coeur de la défense racinaire : Caractérisation moléculaire et fonctionnelle chez deux légumineuses, *Glycine max* (Merr.) L. et *Pisum sativum* (L.). Université de Rouen Normandie.

Ropitiaux, M., Bernard, S., Follet-Gueye, M.-L., Vicré, M., Boulogne, I., Driouich, A., 2019. Xyloglucan and cellulose form molecular cross-bridges connecting root border cells in pea (*Pisum sativum*). *Plant Physiology and Biochemistry* 139, 191–196.

S

Saijo, Y., Loo, E.P., Yasuda, S., 2018. Pattern recognition receptors and signaling in plant-microbe interactions. *Plant J* 93, 592–613.

Santana Silva, R.J., Micheli, F., 2020. RRGPredictor, a set-theory-based tool for predicting pathogen-associated molecular pattern receptors (PRRs) and resistance (R) proteins from plants. *Genomics* 112, 2666-2676.

Saski, C., Lee, S.-B., Daniell, H., Wood, T.C., Tomkins, J., Kim, H.-G., Jansen, R.K., 2005. Complete Chloroplast Genome Sequence of *Glycine max* and Comparative Analyses with other Legume Genomes. *Plant Molecular Biology* 59, 309–322.

Schauer, C., Janko, C., Munoz, L.E., Zhao, Y., Kienhöfer, D., Frey, B., Lell, M., Manger, B., Rech, J., Naschberger, E., Holmdahl, R., Krenn, V., Harrer, T., Jeremic, I., Bilyy, R., Schett, G., Hoffmann, M., Herrmann, M., 2014. Aggregated neutrophil extracellular traps limit inflammation by degrading cytokines and chemokines. *Nature Medicine* 20, 511–517.

Scheller, H.V., Ulvskov, P., 2010. Hemicelluloses. *Annu. Rev. Plant Biol.* 61, 263–289.

Scheres, B., Benfey, P., Dolan, L., 2002. Root Development. *The Arabidopsis Book* 1, e0101.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.-C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C., Jackson, S.A., 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183.

Selders, G.S., Fetz, A.E., Radic, M.Z., Bowlin, G.L., 2017. An overview of the role of neutrophils in innate immunity, inflammation and host-biomaterial integration. *Regenerative Biomaterials* 4, 55–68.

Severin, A.J., Woody, J.L., Bolon, Y.-T., Joseph, B., Diers, B.W., Farmer, A.D., Muehlbauer, G.J., Nelson, R.T., Grant, D., Specht, J.E., Graham, M.A., Cannon, S.B., May, G.D., Vance, C.P., Shoemaker, R.C., 2010. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. *BMC Plant Biology* 10, 160.

Shen, Y., Liu, J., Geng, H., Zhang, J., Liu, Y., Zhang, H., Xing, S., Du, J., Ma, S., Tian, Z., 2018. De novo assembly of a Chinese soybean genome. *Science China Life Sciences* 61, 871–884.

Showalter, A.M., Basu, D., 2016. Extensin and Arabinogalactan-Protein Biosynthesis: Glycosyltransferases, Research Challenges, and Biosensors. *Front. Plant Sci.* 7, 814.

Simon, D., Hoesli, S., Roth, N., Staedler, S., Yousefi, S., Simon, H.-U., 2011. Eosinophil extracellular DNA traps in skin diseases. *Journal of Allergy and Clinical Immunology* 127, 194–199.

Simon, D., Simon, H.-U., Yousefi, S., 2013. Extracellular DNA traps in allergic, infectious, and autoimmune diseases. *Allergy* 68, 409–416.

Sirois, S.H., Buckley, D.H., 2019. Factors governing extracellular DNA degradation dynamics in soil. *Environmental Microbiology Reports* 11, 173–184.

Sobhanian, H., Razavizadeh, R., Nanjo, Y., Ehsanpour, A., Jazii, F., Motamed, N., Komatsu, S., 2010. Proteome analysis of soybean leaves, hypocotyls and roots under salt stress. *Proteome Sci* 8, 19.

Solymosi, K., Lethin, J., Aronsson, H., 2018. Diversity and Plasticity of Plastids in Land Plants, in: Maréchal, E. (Ed.), *Plastids, Methods in Molecular Biology*. Springer US, New York, NY, pp. 55–72.

Song, L., Prince, S., Valliyodan, B., Joshi, T., Maldonado dos Santos, J.V., Wang, J., Lin, L., Wan, J., Wang, Y., Xu, D., Nguyen, H.T., 2016. Genome-wide transcriptome analysis of soybean primary root under varying water-deficit conditions. *BMC Genomics* 17, 57.

Song, L., Valliyodan, B., Prince, S., Wan, J., Nguyen, H., 2018. Characterization of the XTH Gene Family: New Insight to the Roles in Soybean Flooding Tolerance. *IJMS* 19, 2705.

Souza Cândido, E., Pinto, M.F.S., Pelegrini, P.B., Lima, T.B., Silva, O.N., Pogue, R., Grossi-de-Sá, M.F., Franco, O.L., 2011. Plant storage proteins with antimicrobial activity: novel insights into plant defense mechanisms. *FASEB j.* 25, 3290–3305.

Stegmann, M., Monaghan, J., Smakowska-Luzan, E., Rovenich, H., Lehner, A., Holton, N., Belkhadir, Y., Zipfel, C., 2017. The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. *Science* 355, 287–289.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P., Huala, E., 2007. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research* 36, D1009–D1014.

T

Tetz, V.V., Tetz, G.V., 2010. Effect of Extracellular DNA Destruction by DNase I on Characteristics of Forming Biofilms. *DNA and Cell Biology* 29, 399–405.

Thierry, A.R., El Messaoudi, S., Gahan, P.B., Anker, P., Stroun, M., 2016. Origins, structures, and functions of circulating DNA in oncology. *Cancer and Metastasis Reviews* 35, 347–376.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A., Rhee, S.Y., Stitt, M., 2004. mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal* 37, 914–939.

Timotiwu, P.B., Sakurai, N., 2002. Identification of mono-, oligo-, and polysaccharides secreted from soybean roots. *Journal of Plant Research* 115, 77–85.

Tran, T.M., MacIntyre, A., Hawes, M., Allen, C., 2016. Escaping Underground Nets: Extracellular DNases Degrade Plant Extracellular Traps and Contribute to Virulence of the Plant Pathogenic Bacterium *Ralstonia solanacearum*. *PLOS Pathogens* 12, e1005686.

Tuteja, N., Mahajan, S., 2007. Calcium Signaling Network in Plants: An Overview. *Plant Signaling & Behavior* 2, 79–85.

Tyler, B.M., 2007. *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. *Molecular Plant Pathology* 8, 1–8.

U

Urban, C.F., Ermert, D., Schmid, M., Abu-Abed, U., Goosmann, C., Nacken, W., Brinkmann, V., Jungblut, P.R., Zychlinsky, A., 2009. Neutrophil Extracellular Traps Contain Calprotectin, a Cytosolic Protein Complex Involved in Host Defense against *Candida albicans*. *PLoS Pathogens* 5, e1000639.

V

van den Berg, C., Willemsen, V., Hendriks, G., Weisbeek, P., Scheres, B., 1997. Short-range control of cell differentiation in the *Arabidopsis* root meristem. *Nature* 390, 287–289.

Van der Ent, S., Pieterse, C.M.J., 2018. Ethylene: Multi-Tasker in Plant-Attacker Interactions, in: Roberts, J.A. (Ed.), *Annual Plant Reviews Online*. John Wiley & Sons, Ltd, Chichester, UK, pp. 343–377.

van der Linden, M., Westerlaken, G.H.A., van der Vlist, M., van Montfrans, J., Meyaard, L., 2017. Differential Signalling and Kinetics of Neutrophil Extracellular Trap Release Revealed by Quantitative Live Imaging. *Sci Rep* 7, 6529.

Vandana, V.V., Suseela Bhai, R., 2018. Differential expression of PR genes in response to *Phytophthora capsici* inoculation in resistant and susceptible black pepper (*Piper nigrum* L.) lines. *Eur J Plant Pathol* 150, 713–724.

Vanyushin, B.F., 2006. DNA Methylation in Plants 56.

Vargas, A., Boivin, R., Cano, P., Murcia, Y., Bazin, I., Lavoie, J.-P., 2017. Neutrophil extracellular traps are downregulated by glucocorticosteroids in lungs in an equine model of asthma. *Respiratory Research* 18, 207.

Vicré, M., Santaella, C., Blanchet, S., Gateau, A., Driouich, A., 2005. Root Border-Like Cells of *Arabidopsis*. Microscopical Characterization and Role in the Interaction with Rhizobacteria. *PLANT PHYSIOLOGY* 138, 998–1008.

von Köckritz-Blickwede, M., Goldmann, O., Thulin, P., Heinemann, K., Norrby-Teglund, A., Rohde, M., Medina, E., 2008. Phagocytosis-independent antimicrobial activity of mast cells by means of extracellular trap formation. *Blood* 111, 3070–3080.

Vuillemin, A., Horn, F., Alawi, M., Henny, C., Wagner, D., Crowe, S.A., Kallmeyer, J., 2017. Preservation and Significance of Extracellular DNA in Ferruginous Sediments from Lake Towuti, Indonesia. *Front. Microbiol.* 8.

W

Wala, J.A., Bandopadhyay, P., Greenwald, N.F., O'Rourke, R., Sharpe, T., Stewart, C., Schumacher, S., Li, Y., Weischenfeldt, J., Yao, X., Nusbaum, C., Campbell, P., Getz, G., Meyerson, M., Zhang, C.-Z., Imielinski, M., Beroukhim, R., 2018. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* 28, 581–591.

Wan, J., Torres, M., Ganapathy, A., Thelen, J., DaGue, B.B., Mooney, B., Xu, D., Stacey, G., 2005. Proteomic Analysis of Soybean Root Hairs After Infection by *Bradyrhizobium japonicum*. *MPMI* 18, 458–467.

Wang, H., He, H., Qi, Y., McLellan, H., Tian, Zhejuan, Birch, P.R.J., Tian, Zhendong, 2019. The oomycete microbe-associated molecular pattern Pep-13 triggers SERK3/BAK1-independent plant immunity. *Plant Cell Reports* 38, 173-182.

Wang, P., Chen, X., Goldbeck, C., Chung, E., Kang, B.-H., 2017. A distinct class of vesicles derived from the trans -Golgi mediates secretion of xylogalacturonan in the root border cell. *The Plant Journal* 92, 596–610.

Wang, P., Yang, B., Wan, H., Fang, X., Yang, C., 2018. The differences of cell wall in roots between two contrasting soybean cultivars exposed to cadmium at young seedlings. *Environ Sci Pollut Res* 25, 29705–29714.

Wang, X., Zhao, Z., Guo, N., Wang, H., Zhao, J., Xing, H., 2020. Comparative Proteomics Analysis Reveals That Lignin Biosynthesis Contributes to Brassinosteroid-Mediated Response to *Phytophthora sojae* in Soybeans. *J. Agric. Food Chem.* 68, 5496–5506.

Webster, S.J., Daigneault, M., Bewley, M.A., Preston, J.A., Marriott, Helen.M., Walmsley, S.R., Read, R.C., Whyte, M.K.B., Dockrell, D.H., 2010. Distinct Cell Death Programs in Monocytes Regulate Innate Responses Following Challenge with Common Causes of Invasive Bacterial Disease. *J.I.* 185, 2968–2979.

Wen, F., Curlango-Rivera, G., Huskey, D.A., Xiong, Z., Hawes, M.C., 2017. Visualization of extracellular DNA released during border cell separation from the root cap. *American Journal of Botany* 104, 970–978.

Wen, F., VanEtten, H.D., Tsaprailis, G., Hawes, M.C., 2007. Extracellular Proteins in Pea Root Tip and Border Cell Exudates. *PLANT PHYSIOLOGY* 143, 773–783.

Wen, F., White, G.J., VanEtten, H.D., Xiong, Z., Hawes, M.C., 2009. Extracellular DNA Is Required for Root Tip Resistance to Fungal Infection. *PLANT PHYSIOLOGY* 151, 820–829.

Wen, F., Zhu, Y., Hawes, M.C., 1999. Effect of Pectin Methyltransferase Gene Expression on Pea Root Development 11, 1129-1140.

Wildermuth, M.C., 2019. Plants fight fungi using kiwellin proteins. *Nature* 565, 575–577.

Willats, W.G.T., Knox, J.P., 1996. A role for arabinogalactan-proteins in plant cell expansion: evidence from studies on the interaction of b-glucosyl Yariv reagent with seedlings of *Arabidopsis thaliana*. *The plant journal* 9, 919–925.

Wilson, R.F., 2008. Soybean: Market Driven Research Needs, in: Stacey, G. (Ed.), *Genetics and Genomics of Soybean*. Springer New York, New York, NY, pp. 3–15.

Wrather, A., Shannon, G., Balardin, R., Carregal, L., Escobar, R., Gupta, G.K., Ma, Z., Morel, W., Ploper, D., Tenuta, A., 2010. Effect of Diseases on Soybean Yield in the Top Eight Producing Countries in 2006. *Plant Health Progress* 11, 29.

Y

Yakovchuk, P., 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research* 34, 564–574.

Yamaguchi, Y., Barona, G., Ryan, C.A., Pearce, G., 2011. GmPep914, an Eight-Amino Acid Peptide Isolated from Soybean Leaves, Activates Defense-Related Genes. *Plant Physiol.* 156, 932–942.

Yan, Q., Si, J., Cui, X., Peng, H., Chen, X., Xing, H., Dou, D., 2019. The soybean cinnamate 4-hydroxylase gene GmC4H1 contributes positively to plant defense via increasing lignin content. *Plant Growth Regul* 88, 139–149.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 12, 7–8

Yang, B., Wang, Yuyin, Guo, B., Jing, M., Zhou, H., Li, Y., Wang, H., Huang, J., Wang, Yan, Ye, W., Dong, S., Wang, Yuanchao, 2019. The *Phytophthora sojae* RXLR effector Avh238 destabilizes soybean Type2 Gm ACS s to suppress ethylene biosynthesis and promote infection. *New Phytol* 222, 425–437.

Yipp, B.G., Kubes, P., 2013. NETosis: how vital is it? *Blood* 122, 2784–2794.

Yipp, B.G., Petri, B., Salina, D., Jenne, C.N., Scott, B.N.V., Zbytniuk, L.D., Pittman, K., Asaduzzaman, M., Wu, K., Meijndert, H.C., Malawista, S.E., de Boisleury Chevance, A., Zhang, K., Conly, J., Kubes, P., 2012. Infection-induced NETosis is a dynamic process involving neutrophil multitasking in vivo. *Nature Medicine* 18, 1386–1393.

Yousefi, S., Mihalache, C., Kozlowski, E., Schmid, I., Simon, H.U., 2009a. Viable neutrophils release mitochondrial DNA to form neutrophil extracellular traps. *Cell Death Differ* 16, 1438–1444.

Yousefi, S., Mihalache, C., Kozlowski, E., Schmid, I., Simon, H.U., 2009b. Viable neutrophils release mitochondrial DNA to form neutrophil extracellular traps. *Cell Death Differ* 16, 1438–1444.

Yousefi, S., Stojkov, D., Germic, N., Simon, D., Wang, X., Benarafa, C., Simon, H., 2019. Untangling “NETosis” from NETs. *Eur. J. Immunol.* 49, 221–227.

Yu, X., Feng, B., He, P., Shan, L., 2017. From Chaos to Harmony: Responses and Signaling Upon Microbial Pattern Recognition. *Annual review of phytopathology. Annu. Rev. Phytopathol.* 55, 5.1-5.29.

Z

Zabala, G., Zou, J., Tuteja, J., Gonzalez, D.O., Clough, S.J., Vodkin, L.O., 2006. Transcriptome changes in the phenylpropanoid pathway of Glycine max in response to *Pseudomonas syringae* infection. *BMC Plant Biol* 6, 26.

Zawrotniak, M., Rapala-Kozik, M., 2013. Neutrophil extracellular traps (NETs) — formation and implications 60, 277–284.

Zhang, Fuli, Ge, H., Zhang, Fan, Guo, N., Wang, Y., Chen, L., Ji, X., Li, C., 2016. Biocontrol potential of *Trichoderma harzianum* isolate T-aloe against *Sclerotinia sclerotiorum* in soybean. *Plant Physiology and Biochemistry* 100, 64–74.

Zhang, Z., Li, J., Pan, L., Ye, Y., Zeng, X., Song, T., Zhang, X., Wang, E.K., 2014. A Novel Visualization of DNA Sequences, Reflecting GC-Content. *MATCH Commun. Math. Comput. Chem.* 72, 533-550.

Zhao, H., Ma, B., Duan, K.-X., Li, X.-K., Lu, X., Yin, C.-C., Tao, J.-J., Wei, W., Zhang, W.-K., Xin, P.-Y., Man Lam, S., Chu, J.-F., Shui, G.-H., Chen, S.-Y., Zhang, J.-S., 2020. The GDSL Lipase MHZ11 Modulates Ethylene Signaling in Rice Roots. *Plant Cell* 32, 1626–1643.

Zhao, J., Zhang, Y., Bian, X., Lei, J., Sun, J., Guo, N., Gai, J., Xing, H., 2013. A comparative proteomics analysis of soybean leaves under biotic and abiotic treatments. *Molecular Biology Reports* 40, 1553–1562.

Zimmermann, S., Nurnberger, T., Frachisse, J.-M., Wirtz, W., Guern, J., Hedrich, R., Scheel, D., 1997. Receptor-mediated activation of a plant Ca²⁺-permeable ion channel involved in pathogen defense. *Proceedings of the National Academy of Sciences* 94, 2751–2755.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D.G., Boller, T., Felix, G., 2006. Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated Transformation. *Cell* 125, 749–760.

Zou, J., Rodriguez-Zas, S., Aldea, M., Li, M., Zhu, J., Gonzalez, D.O., Vodkin, L.O., DeLucia, E., Clough, S.J., 2005. Expression profiling soybean response to *Pseudomonas syringae* reveals new defense-related genes and rapid HR-specific downregulation of photosynthesis. *Molecular Plant-Microbe Interactions* 18, 1161–1174.

Sites web consultés

Agence Canadienne de l'inspection des aliments, 2020.

<https://www.inspection.gc.ca/protection-des-vegetaux/phytoravageurs-especes-envahissantes/maladies/rouille-du-soja/fra/1327501711301/1327501824978>

Agrireseau, 2020. https://www.agrيرهseau.net/documents/Document_90449.pdf

Carré M-A, Réussir 2019. <https://www.reussir.fr/le-soja-la-conquete-de-la-normandie>

CIQUAL ANSES 2020. <https://ciqual.anses.fr/#/aliments/20901/soja-graine-entiere>

Ecophytopic, 2020. <http://www.ecophytopic.fr/tr/methodes-de-lutte/biocontrôle/auxiliaires-lâchers-inondatifs>)

Goemar, 2020. <https://www.goemar.com/fr/technologies/natural-protect-technologie/homologation>

INRAE, ePhytia, 2020. <http://ephytia.inra.fr/fr/C/22671/Vigi-Semences-Heterodera-glycines-Nematode-a-kystes-du-soja>

INRAE, ePhytia, 2020. <http://ephytia.inra.fr/fr/C/25358/Vigi-Semences-Phytophthora-sojae>

Iriis CRAAQ, 2020. (CRAAQ = Centre de référence en agriculture et agroalimentaire du Québec), <https://www.iriisphytoprotection.qc.ca/Fiche/Bacterie?imageld=1960>

Ministère de l'agriculture, 2020a. <https://agriculture.gouv.fr/le-plan-ecophyto-quest-ce-que-cest>

Ministère de l'agriculture, 2020b. <https://agriculture.gouv.fr/quest-ce-que-le-biocontrole>

Phytozomes, 2020. https://phytozome.igi.doe.gov/pz/portal.html#!info?alias=Org_Gmax

Plantix, 2020. <https://plantix.net/fr/library/plant-diseases/200013/soybean-mosaic-virus>

Quick Agro, 2020. <https://www.quick-agro.fr/soja/maladies/65-mildiou-du-soja.html>

Sortir du glyphosate, Ministère de l'agriculture et ministère de la transition écologique et solidaire, 2020. <https://glyphosate.gouv.fr/solutions-et-alternatives>

Terre-net média, 2020. <https://www.terre-net.fr/observatoire-technique-culturelle/strategie-technique-culturelle/article/une-annee-record-pour-le-soja-francais-217-133686.html>

Annexe I : Etude des AC-DC et des racines du Lin (*Linum usitatissimum* L.)

- I. Matériel et méthodes
- II. Résultats
- III. Discussion
- IV. Conclusion et perspectives

Annexe I : Etude des AC-DC et des racines du Lin (*Linum usitatissimum* L.)

L'objectif de cette étude est de mieux comprendre le rôle du RET du Lin (*Linum usitatissimum* L.) et son action de défense de la racine. Pour cela, le protéome des racines et des AC-DC de lin, en condition contrôle ou élicitée par l'acide fusarique ; issu des champignons phytopathogènes du genre *Fusarium* sp, a été réalisé, ainsi qu'une caractérisation microscopique du RET et une extraction de l'ARN des AC-DC dans l'optique de la réalisation de leur transcriptome. Les parties de caractérisation microscopique et transcriptomique du RET de Lin sont issues du stage de master 1 de L. N'Guessan.

I. Matériels et méthodes

Pour cette étude, des graines de lin de variété Aramis nous ont été fournies par la coopérative Terre de Lin dans le cadre des travaux de thèse de Planchon, 2018. En effet cette variété de lin textile de printemps présente une bonne tolérance à *Fusarium oxysporum* Schltdl. Cette variété a d'ailleurs été recommandée pour les parcelles à faible risque de verse en 2019 par Arvalis (Arvalis, 2020), notamment pour sa forte tolérance à la fusariose (*Fusarium oxysporum*) et à la brûlure (*Chalara elegans* Nag Raj & W.B. Kendr., et *Pythium* sp).

I-1 Stérilisation des graines

Les graines sont placées dans un pot à prélèvements ouvert disposé sous une cloche à vide elle-même placée sous une sorbonne. Un bécher contenant 100 mL de javel (hypochlorite de sodium à 9,6% de chlore actif) est déposé au centre de la cloche. 3 mL d'acide chlorhydrique fumant (ACS, Reag. ISO, Reag. Ph. Eur., fumant, ≥ 37 %) sont ensuite ajoutés très rapidement dans le bécher de javel et la cloche est aussitôt refermée pour conserver le gaz dichlore immédiatement formé. Les graines sont incubées dans ce dispositif pendant deux heures. Suite à cela, un maximum de gaz est évacué stérilement à l'aide d'une

pompe sous la sorbonne pendant cinq minutes. La cloche est ouverte et les tubes de graines sont refermés rapidement pour maintenir leur stérilité. Les graines sont ensuite conservées à 4°C en attendant leur utilisation.

I-2 Mise en culture des graines

Les graines de lin sont mises à imbiber dans l'eau pure stérile sur la nuit et à l'obscurité. Le lendemain, elles sont déposées stérilement dans une boîte de pétri carrée (120 mm) contenant du milieu MS (Duchefa Biochemie) complété avec 1% d'agar bactériologique. Les boîtes sont ensuite scellées avec du film alimentaire et mises en culture verticalement cinq jours dans une armoire à 22°C avec une photopériode de 16h de jour et 8h de nuit. L'élicitation est réalisée après 96h de culture avec de l'acide fusarique à 1µM (Bhouziganne *et al*, 2006). Pour cela les plantules sont recouvertes avec environ 200 µL de cette solution. Pour les plantes témoins, celles-ci sont recouvertes de la même manière avec de l'eau pure stérile. Les boîtes sont refermées et laissées sécher à l'horizontale dans l'armoire de culture avant de les replacer verticalement 3 heures plus tard (Figure S1-1).



Figure S1-1 : Culture de lin (5jours) en milieu MS solide, orientation verticale. A droite, élicitation à l'acide fusarique et à gauche sans élicitation

I-3 Approches microscopiques

I-3-1 Contre coloration à l'encre de chine

Pour une visualisation globale du mucilage présent dans le RET, une contre coloration à l'encre de Chine, diluée au 1/500ème dans de l'eau pure, est effectuée. La solution est déposée au niveau de l'apex racinaire d'une plantule posée sur une lame de microscopie (Driouich *et al*, 2013) et observée à la loupe binoculaire (Leica M125, caméra DFC295).

I-3-2 Immunomarquage et microscopie à épifluorescence.

Le protocole utilisé a été créé au sein du laboratoire (par Gaëlle Durambur) (Castilleux *et al*, 2018 ; Ropitiaux 2018). Des lames à 10 puits sont nettoyées avec de l'éthanol 96°, puis de la Poly-lysine (diluée au 1/10ème) est déposée et séchée une nuit (améliore l'adhérence des racines et du RET).

La pointe racinaire est déposée au centre du puit, puis est fixée avec 40 µL d'une solution de PIPES 1X contenant un produit fixateur, le paraformaldéhyde (PAF à 4%) pendant 40 minutes. Dès l'ajout de PAF, les manipulations se font exclusivement sous sorbonne. La solution de fixation est retirée, et une série de 4 lavages est réalisée avec 40 µL de PBS1X + BSA 1% avec des temps d'incubation de 10 minutes pour chaque lavage. La pointe racinaire est ensuite incubée dans 29 µL d'anticorps primaire LM1 (Smallwood *et al*, 1995) dilué au 1/5e par une solution de PBS1X + BSA1%, en chambre humide, toute la nuit à 4°C. Le lendemain, la solution d'anticorps primaire est retirée, puis une série de 4 lavages avec 40 µL de PBS1X + BSA 1% est effectuée avec 10 minutes de temps d'incubation pour chaque lavage. La racine est ensuite incubée avec 29 µL d'anticorps secondaire (AlexaFluor TM 488 goat anti-rat IgG (H+L); Thermofisher) dilué au 1/200e dans du PBS1X + BSA 1%, dans une chambre humide, pendant 2 heures à l'étuve à 25°C. La solution contenant l'Anticorps secondaire est retirée, et une série de 4 lavages de 10 minutes est réalisée afin d'éliminer le surplus d'anticorps. Un dernier lavage se fait avec 40 µL de PBS1X pendant 10 minutes et la solution est retirée. Pour finir, un « touch » est réalisé, c'est-à-dire que la racine est frottée légèrement dans le puit afin de détacher le RET puis retirée. Enfin 5 µL de PBS1X sont ajoutés. La lame à puit est recouverte d'une lamelle et scellée avec du scotch Anapore.

Les observations de l'immunomarquage de surface se font au microscope à épifluorescence inversé (Leica DMI6000B). Le cube à filtre L5 (λexcitation 360± 40 ;

λémission 470 ± 40) ; est utilisé et l'intensité est réglée à 1, le diaphragme d'ouverture à 11 et de champ à 25.

I-3-3 Composés phénoliques et microscopie à épifluorescence.

Les racines de plantules sont découpées afin de récupérer l'apex racinaire et la zone pilifère. Ces 2 parties d'intérêts sont posées sur une lame en verre recouverte de poly-lysine, préalablement nettoyée à l'éthanol 96°. L'équivalent de 200 μ L d'eau stérile sont ajoutés afin de recouvrir chaque partie, enfin la lamelle est scellée avec du scotch Anapore.

L'observation des composés phénoliques est faite *via* un microscope à épifluorescence inversé (Leica DMI6000B) avec un cube à filtre bleu A4 L5 (λ excitation 480 ± 40 ; λ émission 527 ± 30) correspondant aux ultra-violet (UV). L'intensité est réglée à 5, le diaphragme d'ouverture à 11 et de champ à 30. Les temps sont relevés à une saturation précise située entre le minimum de surexposition le maximum de sous exposition (Logiciel Av57). Un test non paramétrique de Wilcoxon a été réalisé sur excel.

I-4 Analyse transcriptomique

I-4-1 Récolte du RET de lin

Les pointes racinaires sont coupées et trempées dans des cupules, préalablement rincées à l'éthanol et lavées à l'eau RNase-Dnase free, dans 200 μ l de tampon de lyse ARN. Les racines sont agitées afin de décrocher les cellules frontières ainsi que le mucilage dans la cupule. Lorsque la cupule devient trop concentrée en cellules (le liquide devient trouble), il convient de remplacer le liquide. Les échantillons contenant le RET sont conservés dans la glace tout le long de la manipulation. Les racines sont conservées dans des pots à prélèvements, pour la comparaison du transcriptome de racine à celui des cellules frontières. Les échantillons de RET et de racines sont conservés au congélateur à -80°C .

I-4-2 Extraction de l'ARN et analyse transcriptomique

Deux kits d'extraction Macherey-Nagel (MN) permettant d'isoler l'ARN ont été testés : NucleoSpin RNAplus et NucleoSpin RNAplusXS. L'extraction se fait directement sur les échantillons de cellules frontières récoltés. Ces derniers sont homogénéisés et lysés dans un volume de tampon spécifique adaptés à chaque kit (LBP pour RNAplus ou LB1 pour RNAplusXS). Les échantillons obtenus sont filtrés dans une colonne retenant l'ADN génomique et centrifugés 30 secondes à 11 000 g. Une solution fixatrice d'ARN (BS ou BSXS)

est ajoutée aux lysats obtenus et ce mélange est placé sur une colonne retenant les ARN puis centrifugé 15 secondes à 11 000 g. Les ARN, maintenant retenus sur la membrane de silice de la colonne, subissent 3 lavages successifs. Le 1er avec 200 µL de WB1 ou 100 µL de MDB est suivi d'une centrifugation de 15 secondes à 11 000 g. Le 2e avec 600 µL ou 500 µL de WB2 est suivi d'une centrifugation de 15 secondes à 11 000 g. Enfin, le 3e avec 250 µL ou 200 µL de WB2 est suivi d'une centrifugation de 2min à 11 000 g. L'ARN est enfin élué avec de l'eau RNase-free (2 x 30 µL ou 10-20 µL) grâce à une centrifugation d'1 min à 11 000g. Les ARN isolés (4 répliques de 1 µg) sont conservés au congélateur à - 80°C jusqu'à leur envoi à la plateforme de génomique, plateforme MGX (Montpellier), pour séquençage haut débit RNAseq.

I-5 Analyse protéomique

I-5-1 Récolte des AC-DC et des racines

Les plantes d'une boîte sont prélevées à l'aide d'une pince et sont disposées en "fagots" de façon à ce que les apex soient alignés. Les racines sont ensuite sectionnées avec un scalpel, prélevées à l'aide de la pince et agitées pendant 15 secondes dans 200 µL de tampon d'extraction (urée 7M, thiourée 2M, 0,5% ampholytes pH 4-6,5, TBP 2 mM, 20mM DTT, 2% Chaps, 0,5% C7Bz0) préalablement disposés dans une cupule (Centrifichem Cups 0,25 mL ; Kartell) rincées à l'éthanol 70% et à l'eau pure (Water, Sterile Nuclease-free, TFS Fisher BioReagents). Trois réplicats sont réalisés pour la condition élicitée avec l'acide fusarique et pour la condition contrôle non élicitée. Chaque échantillon est constitué des AC-DC d'environ 300 plantes pour les échantillons d'AC-DC, et de 30 racines pour les échantillons de racines.

I-5-2 Extraction des protéines solubles des AC-DC et des racines

Les tubes contenant les AC-DC sont regroupés de façon à obtenir environ 300 plantes par réplicat. Les racines (30 par échantillon) sont broyées à l'azote liquide à l'aide d'un pilon et d'un mortier, puis remises en solution dans le tampon d'extraction (urée 7M, thiourée 2M, 0,5% ampholytes pH 4-6,5, TBP 2 mM, 20mM DTT, 2% Chaps, 0,5% C7Bz0) et de nouveau broyées à l'aide d'un broyeur (Homogenizer 150, Fisherbrand). Les étapes d'extraction des protéines et d'analyse par spectrométrie de masse sont ensuite réalisées sur la plate-forme PISSARO de l'Université de Rouen Normandie.

Les échantillons sont ensuite soniqués pendant 20 minutes (pulses de 3 secondes, intensité 24%, Sonificateur Microprocesseur 750W, Sonics Materials) puis centrifugés à 1000 g pendant 30 minutes.

Afin de déterminer la quantité de protéines extraites (Figure S1-2, A et C), un dosage de Bradford est réalisé sur deux dilutions de l'échantillon : l'une à 1/10 et l'autre à 1/50. 20µL d'échantillon sont ensuite mélangés avec 1mL de réactif de Bradford (Protein Assay®, Bio-Rad). Les échantillons sont alors incubés 30 minutes à l'obscurité puis leur absorbance à 595 nm est mesurée à l'aide d'un spectrophotomètre. La concentration en protéines est ensuite calculée en comparant les absorbances obtenues avec celles d'une gamme de concentrations de BSA (*bovine serum albumin*, Sigma-Aldrich). Une vérification de ce dosage est réalisée en faisant migrer les échantillons sur un gel SDS-PAGE à 12% d'acrylamide, coloré au nitrate d'argent (Figure S1-2, B).

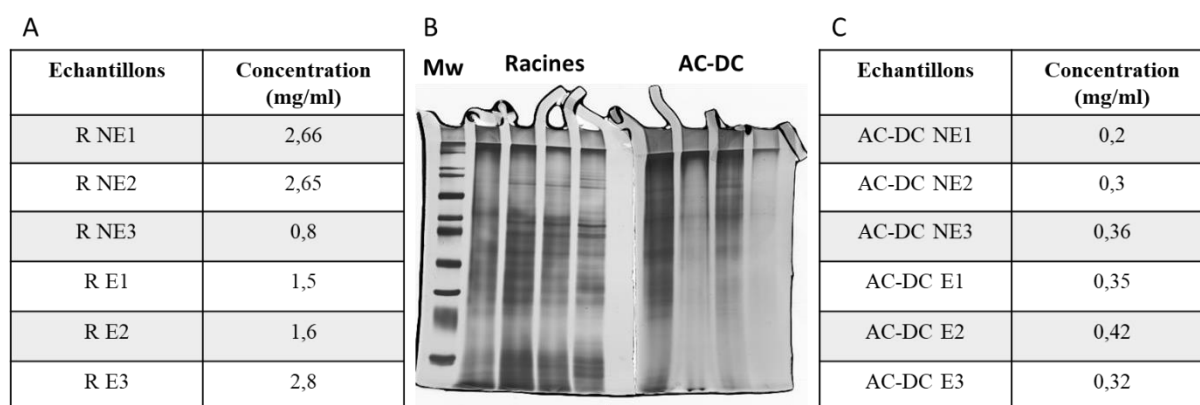


Figure S1-2 : Résultats du dosage protéique (A et C) et gels SDS-PAGE (B) pour les AC-DC et les racines de Lin

Un dosage de Bradford a été réalisé afin d'estimer la concentration en protéines de chacun des échantillons de racines (A) et d'AC-DC (C). Ce dosage a ensuite été confirmé par un gel SDS-PAGE coloré au nitrate d'argent (B).

30 µg de protéines sont ensuite déposés sur un gel SDS-PAGE à 7% d'acrylamide (Acrylamide 9,3 mL ; Tris pH 6,8 5 mL ; eau pure 25,05 mL) afin de les concentrer. La migration est réalisée pendant 45min à 10 mA puis 2 heures à 20 mA. Le gel est ensuite coloré 30 secondes dans une solution de bleu de Coomassie (acide acétique 10 mL ; éthanol 50 mL ; eau pure 40 mL ; bleu de Coomassie 0,5g) puis décolorés à l'aide d'une solution de décoloration (acide acétique 10 mL ; éthanol 50 mL ; eau pure 40 mL) jusqu'à observation d'une bande contenant les protéines d'intérêt. Les bandes sont excisées à l'aide d'un scalpel, puis lavées trois fois dans 80 µL d'eau pure sous agitation pendant 10 minutes. Elles sont ensuite incubées deux fois dans des bains successifs de DTT (dithiothréitol 7,7 mg ; eau pure

10 mL) puis de iodoacétamide (18,5 mg ; eau pure 10 mL). Les fragments de gels sont conditionnés dans une solution de tampon carbonate/acétonitrile (V/V) suivi d'un bain d'acétonitrile. Les protéines sont digérées dans une solution trypsique à 10 µg/mL dans du tampon carbonate pendant 3 heures sous agitation à 37°C. Les peptides sont ensuite récupérés après différents bains d'eau milliQ, TFA 1% et d'acétonitrile. Les solutions finales sont enfin séchées au speedvac pendant 4 heures. Les échantillons sont repris dans une solution d'acide formique (Formic acid, FA) à 0,1%.

Les échantillons ont été analysés sur un spectromètre de masse LTQ Orbitrap Elite (Thermo Scientific) équipé d'une source nano-ESI et couplé à une chaîne chromatographique nano-débit (Easy-nLC II, Thermo Scientific). Pour ces analyses, 1 µL de l'échantillon a été chargé sur la colonne d'enrichissement (Cap Trap C₈, 0,5 x 2 mm, Michrom bioressources) à une pression de 200 bars dans 0,1 % FA. La séparation a été effectuée sur une colonne en phase inverse (C₁₈, L 153 mm, ID 5 µm, 100 Å, Nikkyo Technos). Le gradient, d'un débit de 300 nL/min, est assuré comme suit : de 2 à 40% de tampon B (CH₃CN/0,1% AF) en 105 min, de 40 à 80 % de B en 4 min et 80% de B pendant 15 min (tampon A : H₂O/0,1% AF).

I-5-3 Analyse par spectrométrie de masse

Le spectromètre de masse est utilisé en mode positif avec une tension de capillaire fixée à environ 1,5 kV et une température de la source fixée à 200°C. Les échantillons ont été analysés par fragmentation CID à l'aide d'une méthode référencée « Top20 ». Le balayage MS initial est enregistré dans l'analyseur Orbitrap avec une résolution de 60000 sur une gamme de masse m/z 400-2000. Les ions mono-chargés n'ont pas été sélectionnés et le seuil de signal d'intensité pour la sélection du précurseur est fixé à 500. Lors de ces cycles d'analyses, les 20 ions précurseurs les plus intenses ont été sélectionnés pour l'étape de MS/MS. Les peptides cibles déjà sélectionnés pour la MS/MS ont été exclus de manière dynamique pendant 30 secondes. La fragmentation a été réalisée dans l'analyseur trappe ionique linéaire avec une énergie de collision normalisée de 35 %. Dans l'analyseur Orbitrap, toutes les mesures ont été réalisées après calibration interne à l'aide du pic de m/z à 445,12002 correspondant à un polydiméthylcyclosiloxane (lock mass).

I-5-4 Analyse bio-informatique

Les fichiers de données brutes ont été traités à l'aide du logiciel Proteome Discoverer (Thermo Scientific ; version 1.3) sur la plateforme PISSARO (Université de Rouen

Normandie). Les peptides ont été identifiés grâce à l'intégration du moteur de recherche Mascot (Matrix Science ; version 2.2) utilisé contre la base de données protéiques Phytozome *Linum usitatissimum*. Les paramètres suivants ont été utilisés : un clivage manquant et les modifications variables concernant la carbamidométhylation (Cys) et l'oxydation (Met). La tolérance de masse sur l'ion précurseur et les ions fils ont été respectivement de 10 ppm et de 0,5 Da. Les résultats ont été filtrés pour un taux maximal de faux positifs de 5%. Lors de ces analyses, uniquement les protéines identifiées par au moins deux peptides associés à un degré de confiance élevé (« High Confidence ») ont été conservées. La comparaison des différentes conditions d'analyses par analyse en composante principale ainsi que la quantification sont réalisées par le logiciel Progenesis LC-MS (nonlinear DYNAMICS).

II. Résultats

II-1 Caractérisation microscopique du RET de lin en condition de stimulation immunitaire ou non.

II-1-1 Contre-coloration du mucilage du RET à l'encre de chine

La méthode de contre-coloration à l'encre de chine permet de visualiser le mucilage dense du RET. En effet, l'encre de chine ne colore pas le mucilage ce qui permet une coloration négative pour visualiser celui-ci. Sur la figure SI-3 le mucilage (zone délimité par les pointillés) qui englobe l'apex racinaire et les AC-DC du lin (A,B) est observé. En conditions non élicitées (B,D), le mucilage du RET est présent chez les deux plantes de façon constitutionnelle. Après élicitation avec l'AF (photo A), on observe une augmentation importante de la quantité de mucilage au sein du RET.

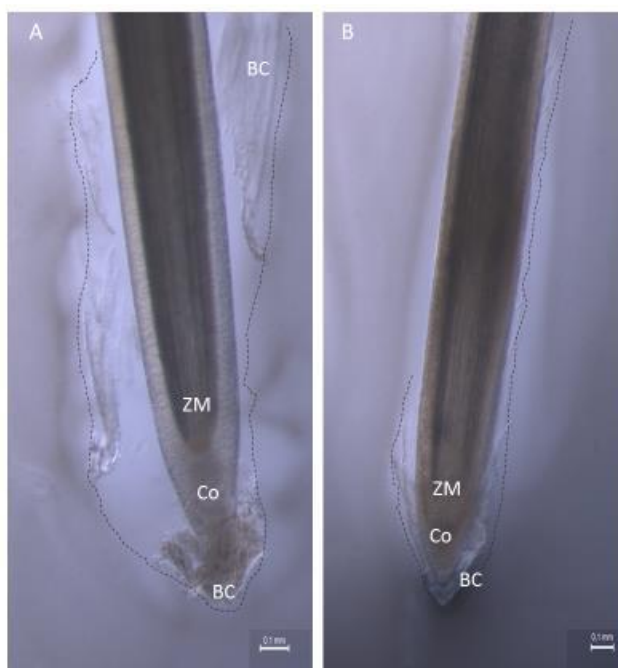


Figure S1-3 : Contre-coloration du RET de lin à l'encre de chine en condition d'élicitation avec l'acide fusarique (A) ou non (B)

Le RET est délimité par la ligne en pointillés. ZM : zone méristématique ; Co : coiffe ; BC : cellules frontières ou AC-DC. Barres d'échelle : 0,1 mm.

L'immunomarquage des extensines avec l'anticorps monoclonal LM1 (dont l'épitope inconnu) du RET de lin (figure SI-4) permet d'observer la présence d'extensines au niveau de la paroi des AC-DC ainsi que dans le mucilage en conditions élicitées ou non élicitées. Chez le lin, il ne semble pas y avoir de différences d'intensité de fluorescence, la quantité

d'extensines ne semble donc pas varier significativement à la suite de l'élicitation par l'acide fusarique.

De plus, des différences dans la quantité et la localisation de la fluorescence sont observées, en fonction des différents morphotypes d'AC-DC et de leur conformation. En effet, pour les intermédiaires et les allongées on détecte un signal intense lorsqu'elles sont incluses dans des amas (a,b,d,e) et un signal très faible lorsqu'elles sont isolées (c,f).

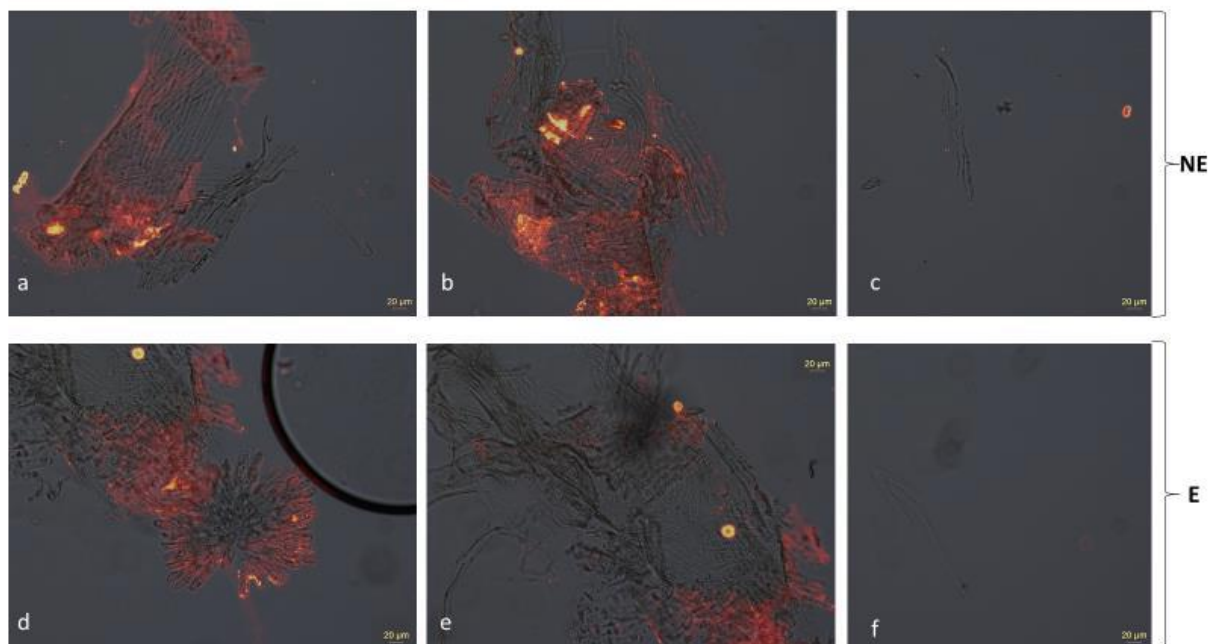


Figure S1-4 : Immunomarcage des extensines via l'anticorps monoclonal LM1 chez le lin en condition de stimulation immunitaire par l'acide fusarique (d,e,f) ou non élicité (a,b, Barre d'échelle 20 µm)

Le suivi de l'auto-fluorescence des racines de lin est possible grâce aux composés phénoliques produits au niveau de la racine. Il est possible d'observer une augmentation de l'auto-fluorescence grâce au relevé du temps d'exposition aux UV. Les photos obtenues figure SI-5 ont été réalisées en mode transmission par lumière blanche et en mode d'épi-illumination UV (A4 : dans le bleu), en se positionnant au niveau du début de la saturation (surexposition). Sachant que le bleu correspond à une sous exposition et le rouge à une surexposition, l'image est acquise lorsque le minimum de pointillés rouges est visible sur l'écran.

L'étude n'a pas relevé de différence significative entre les conditions élicitées ou non comme en témoignent les données des temps d'exposition relevés.

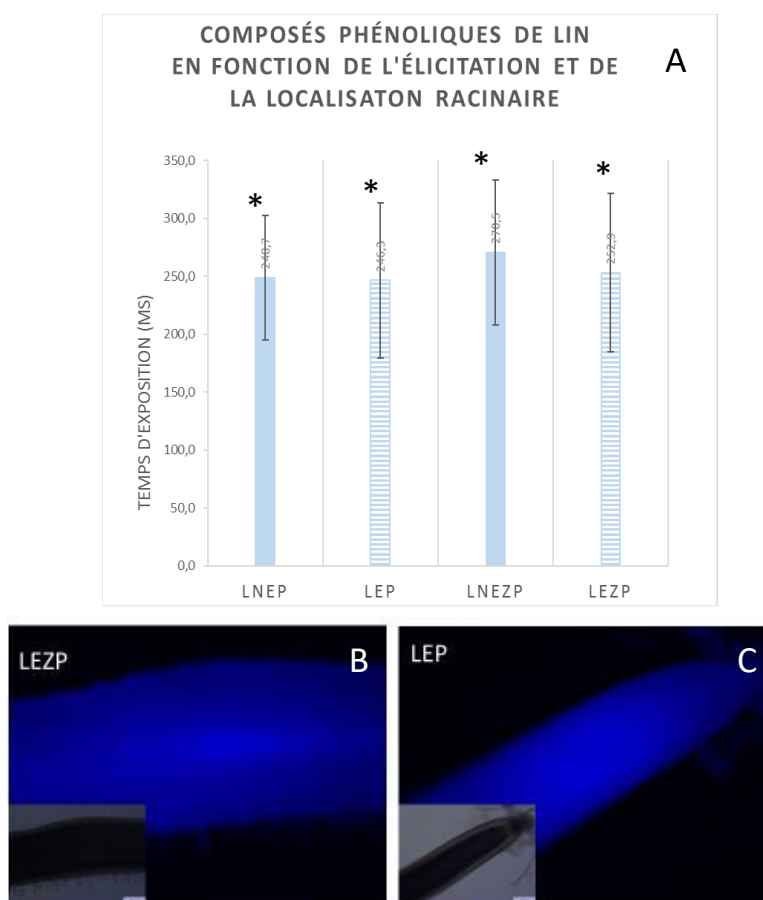


Figure S1-5 : Analyse des composés phénoliques chez le lin en condition de stimulation immunitaire respectivement à la suite d'une élévation par l'AF ou non

A : Histogrammes du temps d'expositions correspondant à la quantité de composés phénoliques en fonction des conditions de stimulation (NE/E) et de la zone racinaire (P/ZP) chez le lin. B, C : Photos d'épifluorescences de racines du lin en conditions d'élévation pour les 2 zones d'intérêts. Des signes (*) différents indiquent des différences significatives entre les valeurs, à l'inverse les signes semblables n'indiquent pas de différences. $P < 0.05$ test (non paramétrique de Wilcoxon).

II-2 Caractérisation transcriptomique des AC-DC de lin

Il a été nécessaire de mettre en place un protocole d'extraction qui permet d'obtenir des échantillons d'ARN avec une qualité et une quantité optimale. Dans ce but, 2 protocoles d'extraction d'ARN sont testés : le kit RNA plus et le kit RNA plus XS, le kit XS permettant l'extraction d'ARN à partir de plus faibles quantités de matériel. Pour chaque kit, une même quantité de matériel est utilisée (environ 100 plantes). Les critères qualitatifs et quantitatifs sont déterminés par des analyses au Nanodrop et au Bioanalyser.

Les résultats du tableau 1 montrent que le kit RNA plus XS a un rendement 7 fois plus important que le RNA plus. Au niveau de la pureté de nos échantillons (Tableau I-1), le ratio A260/ A280 témoigne des contaminations des acides nucléiques par des protéines pour les

ARN issus du kit RNA plus. En revanche, ceux issus du kit RNA plus XS sont de bonne qualité et peu contaminés par l'ADN génomique ou les protéines.

Le second ratio A260/A230, extrêmement faible pour les deux kits, témoigne de la contamination d'un agent absorbant à 230 nm comme par exemple des polysaccharides.

Tableau 1-1 : Tableau de comparaison de la qualité et de la quantité de nos ARN en fonction du kit utilisé (RNAplus ou RNAplusXS)

Kit d'extraction	Nanodrop					Bio-analyser	
	Concentration en ARN (ng/μL)	VE (μL)	Quantité d'ARN totale (μg)	A20/A280	A260/A230	RIN1	RIN2
RNA plus	1,1	60	0,066	0,52	-0,01	8,3	9,3
RNA plus XS	11,5	40	0,46	2,43	0,01	ND	ND

Enfin, le numéro d'intégrité de l'ARN (RIN) est très élevé pour le kit RNA plus (9,3 et 8,3) et les graphiques (Figure SI-6) montrent une faible ligne de base, 2 pics d'ARN ribosomiques à 18S et 25S et un pic des ARN de petites tailles. L'ARN est de bonne qualité pour le RNA plus, la même chose est observée pour le second kit mais, aucun RIN n'est déterminé, cette non-détection s'expliquerait par l'épaule visible sur le graphique au niveau du pic des ARN de petites tailles. Les résultats obtenus pour ce dernier kit témoignent d'une bonne qualité de nos ARN malgré l'absence de RIN.

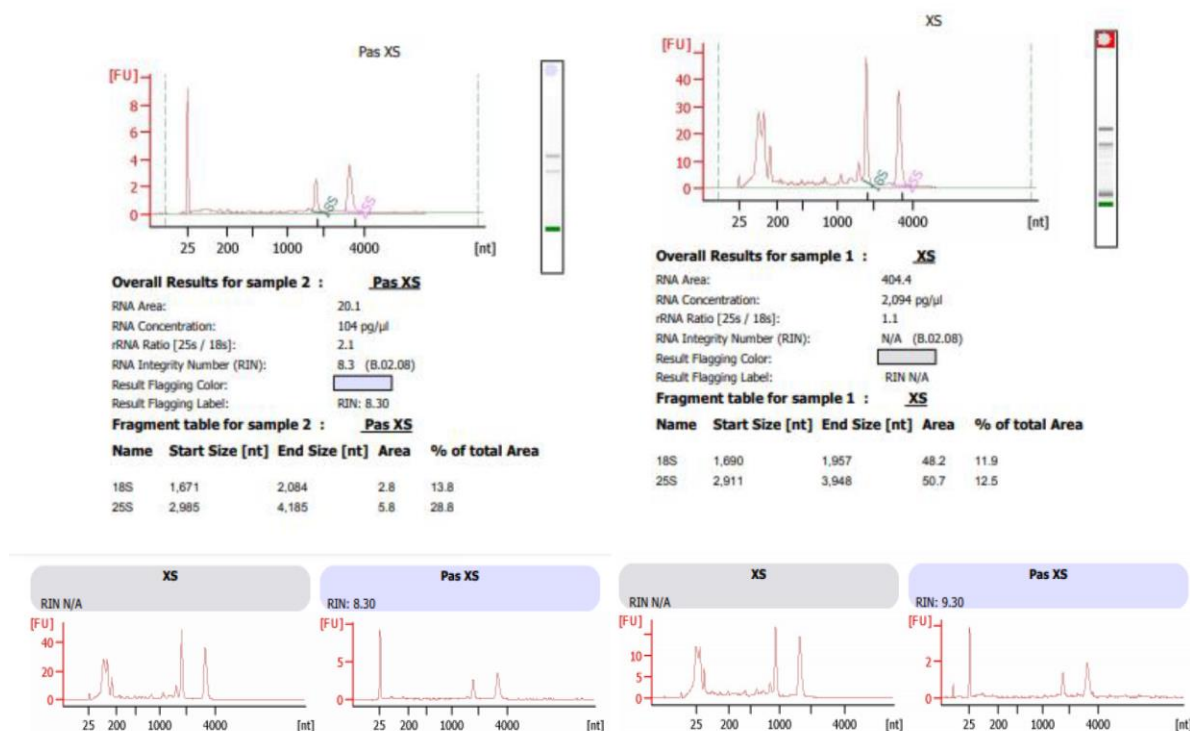


Figure S1-6 : Electrophorégrammes obtenus sur les échantillons d'ARN obtenus par extractions avec 2 kits différents (Pas XS correspond au kit RNAlplus et XS au kit RNAlplus XS)

II-3 Caractérisation protéomique du RET de Lin

Cette première analyse préliminaire a permis d'identifier un total de 478 protéines. Pour chaque condition, le nombre de protéines identifiées pour le meilleur des trois réplicats est présenté dans le tableau SI-2.

Tableau S1-2 : Nombre de protéines identifiées pour le meilleur réplicat de chaque condition

Echantillons	Nombre de protéines
Lin-Racine NE	312
Lin-Racine E	244
Lin-AC-DC NE	168
Lin-AC-DC E	188

Une seconde analyse a été réalisée sur un seul échantillon de racines non élicitées, réalisé à partir d'un plus grand nombre de racines, cet échantillon a permis l'identification de 683 protéines.

III. Discussion

Le but de ce travail était de caractériser les effets d'une molécule élicitrice issue d'un pathogène du Lin sur le RET de celui-ci. Différentes approches microscopiques et moléculaires ont été utilisées, afin de mieux connaître les différents mécanismes de défense mis en place lors d'une stimulation immunitaire.

Le modèle étudié, semble réagir à l'élicitation. En effet la quantité de mucilage au sein du RET déterminée par la contre coloration à l'encre de chine montre une augmentation en condition d'élicitation (Figure SI-3). Il pourrait sembler que ce mucilage résulte soit d'une activité de sécrétion accrue et/ou d'une production d'AC-DC en plus grand nombre par la coiffe. Ces résultats concordent avec ceux présentés par Driouich *et al*, en 2013 sur le pois en réponse à une infection par *Pseudomonas aeruginosa* ou *Fusarium solani* f. sp. Pisi ou par Cannesan *et al*, en 2011 qui montraient une augmentation de la quantité d'AC-DC de pois à la suite d'une infection par *Aphanomyces euteiches*. Bien sûr, cette technique de détection ne permet pas la quantification des AC-DC mais une méthode de quantification pourrait être une perspective de travail.

L'immunomarquage a permis de montrer une importante variabilité au sein de la même espèce selon le type d'AC-DCs (Figure SI-4). Les études de Plancot *et al*, en 2013 et de Castilleux *et al*, en 2018, ont montré que les extensines peuvent être un bon marqueur de défense, et que celles-ci étaient modulées en réponse à une élicitation de la plantule. Les résultats obtenus ne sont pas en accord avec les résultats de ces études. Cependant, il convient d'être prudents car dans ces études précédentes, les modèles utilisés (*Arabidopsis*), les éliciteurs (peptidoglycan et flagelin 22) et la méthode d'imagerie (confocal) n'étaient pas les mêmes.

Pour améliorer le protocole de détection du possible remodelage des extensines chez le lin, il est envisageable d'utiliser un autre anticorps comme JIM12 ou le microscope confocal à la place du microscope à épifluorescence.

D'après les résultats de cette étude, une différence significative de la quantité des composés phénoliques est notable entre l'apex racinaire et la zone pilifère (Figure SI-5, a). Il y aurait plus de composés phénoliques dans l'apex racinaire que dans la zone pilifère. Cette méthode de test aux UV n'a pas permis de montrer une différence significative entre

plantules de lin élicitées et les plantules témoins. Cette méthode mise au point au laboratoire Glyco-MEV, avait cependant montré son efficacité sur les racines de plantules de tomate variété Saint-Pierre élicitée à la flagelline 22 pendant 48h (Durambur et Gotté, données non publiées). Il semblerait donc que cette méthode donne des résultats différents soulignant une variabilité de la distribution des composés phénoliques en fonction du type de plante et d'éliciteurs.

IV. Conclusion et perspectives

Le but de cette étude était de mieux comprendre le rôle du RET de lin et notamment son rôle dans la défense de la racine. Pour cela, une caractérisation microscopique, un protocole de récolte des AC-DC et des racines, puis d'extraction de l'ARN et des protéines solubles et d'analyse protéomique ont été mis au point.

Les résultats du premier axe d'étude ont permis de montrer que les composés phénoliques ne semblent pas être de bons marqueurs de stimulation immunitaire pour nos modèles étudiés.

Concernant l'immunomarquage des extensines, il s'agit d'un premier screening qu'il conviendra d'améliorer, les différences fines observées au microscope à épifluorescence notamment pour le soja semblent suffire. Tout de même, il pourrait être intéressant, par la suite d'envisager d'observer de manière plus précise la réorganisation des extensines notamment via un microscope confocal à balayage laser, utilisé dans les travaux de Castilleux *et al*, en 2018 ou par microscopie électronique à transmission.

Le second axe de ce travail, l'axe de caractérisation « omique » du RET de lin, est toujours en cours de réalisation. Les échantillons de RET et de racines ont été récoltés et les tests réalisés ont permis de choisir le protocole d'extraction des ARN avec le kit RNA Plus XS.

De nouvelles récoltes d'ARN seront réalisées et envoyées à la plateforme MGX. L'objectif est d'effectuer la caractérisation transcriptomique qui permettra d'obtenir des informations supplémentaires sur les fonctions du RET et notamment sur les fonctions défensives mises en place lors d'une élicitation.

Pour la partie protéomique, les premiers tests ont permis d'identifier jusqu'à 683 protéines pour un échantillon, ce qui est un résultat encourageant. Pour la suite de cette étude, de nouvelles récoltes sont envisagées afin d'obtenir des échantillons avec un nombre de plantes plus important.

Annexe II : Etude du transcriptome des AC-DC de racines d'*Arabidopsis thaliana* (L.) Heynh

- I. Matériel et méthodes
- II. Séquençage
- III. Analyses bio-informatiques
- IV. Résultats préliminaires
- V. Conclusion et perspectives

Annexe II : Etude du transcriptome des AC-DC de racines d'*Arabidopsis thaliana* (L.) Heynh

Pour cette étude, l'objectif était d'obtenir un atlas des gènes exprimés dans les AC-DC de racines (*root-Associated Cap-Derived Cells*) d'*Arabidopsis thaliana*, plante modèle, et de les comparer aux données bibliographiques d'autres types cellulaires de la racine de cette plante grâce à une étude métatranscriptomique. Pour cela nous avons tout d'abord dû mettre en place un protocole de récolte des AC-DC et d'extraction de leur ARN en qualité et quantité suffisantes pour permettre leur séquençage, puis l'analyse bio-informatique des données obtenues.

I. Matériels et méthodes

Pour ces analyses, la plante modèle *A. thaliana* écotype Col-0 a été utilisée. En effet cette plante est un modèle physiologique car son génome est le plus étudié et donc le mieux annoté, ce qui nous a permis d'obtenir un maximum d'informations sur les gènes exprimés dans les AC-DC d'*Arabidopsis*.

I-1 Stérilisation des graines

Les graines sont placées dans un microtube de 1,5 mL (FlipTube, GEMÜ GmbH) ouvert puis dans un pot à prélèvements ouvert disposé sous une cloche à vide elle-même placée sous une sorbonne. Un béccher contenant 100 mL de javel (hypochlorite de sodium à 9,6% de chlore actif) est déposé au centre de la cloche. 3 mL d'acide chlorhydrique fumant (ACS, Reag. ISO, Reag. Ph. Eur., fumant, $\geq 37\%$) sont ensuite ajoutés très rapidement dans le béccher de javel et la cloche est aussitôt refermée pour conserver le gaz dichlore immédiatement formé. Les graines sont incubées dans ce dispositif pendant deux heures. Suite à cela, un maximum de gaz est évacué stérilement à l'aide d'une pompe sous la

sorbonne pendant cinq minutes. La cloche est ouverte et les tubes de graines sont refermés rapidement pour maintenir leur stérilité. Les graines sont ensuite conservées à 4°C en attendant leur utilisation.

I-2 Mise en culture des graines

Les graines d'*Arabidopsis* stériles sont rincées trois fois avec 1 mL d'eau pure stérile, puis déposées (à l'aide d'une micropipette 20-200 µL muni d'un cône coupé stérilement) sur du milieu *Arabidopsis* (Duchefa Biochemie), 1% d'agar bactériologique supplémenté avec du $\text{Ca}(\text{NO}_3)_2$ à 2mM dans une boîte de Pétri carrée (120 mm). Environ 100 à 150 graines sont déposées en rang dans chaque boîte (Figure SII-1, A). Celles-ci sont ensuite scellées avec du film alimentaire et mises en culture verticalement onze jours dans une enceinte de culture à 19-21°C avec une photopériode de 16h de jour et 8h de nuit, 80-90% RH.

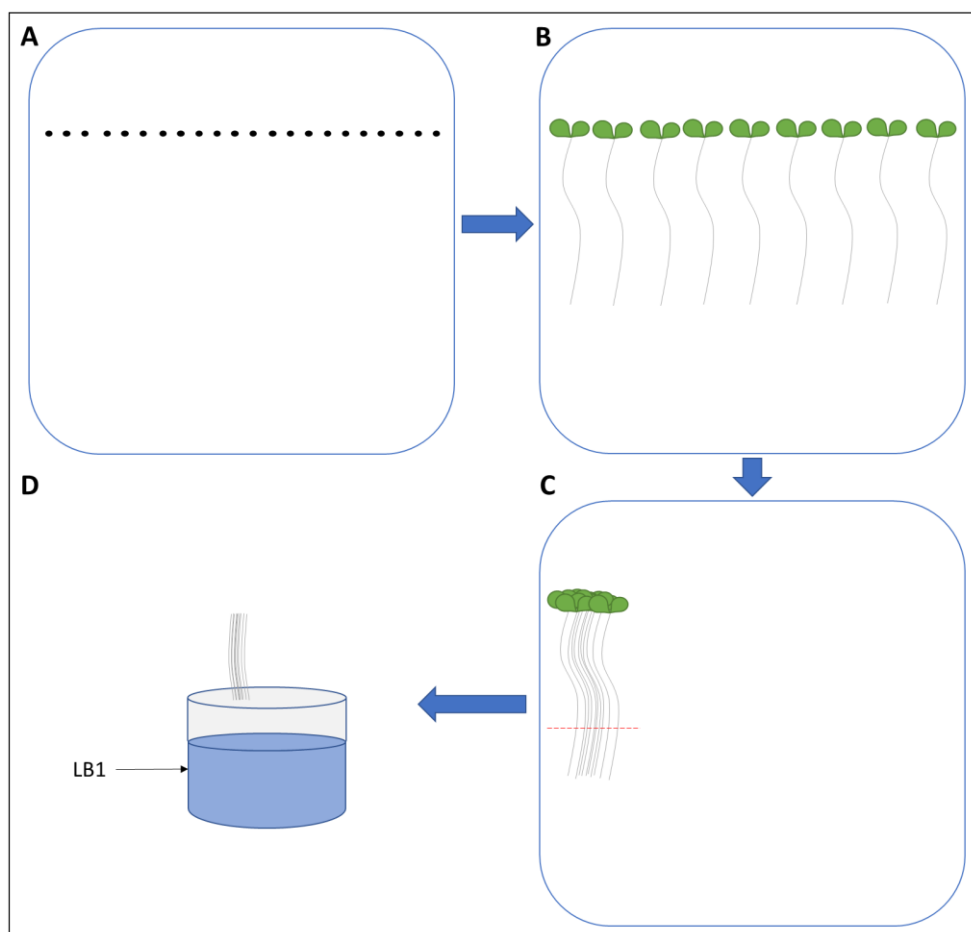


Figure SII-1 : Représentation schématique de la mise en culture des graines d'*Arabidopsis* et de la récolte des AC-DC

(A) Les graines sont mises à germer en ligne dans une boîte de Pétri ; après 11 jours de culture, (B) les plantules sont regroupées de façon à aligner leurs apex racinaires (C). Ceux-ci sont ensuite prélevés et agités dans le tampon de lyse LB1 du kit NucleoSpin® RNA Plus XS (Macherey nagel) (D).

I-3 Récolte de l'ARN des AC-DC

Les plantes d'une boîte (environ 100) sont prélevées à l'aide d'une pince et sont disposées en "fagots" de façon à ce que les apex soient alignés (Figure SII-1, C). Les racines sont ensuite sectionnées avec un scalpel, prélevées à l'aide de la pince et agitées (Figure SII-1, D) pendant 15 secondes dans 200 µL de tampon de lyse LB1 du kit NucleoSpin® RNA Plus XS (Macherey nagel) préalablement disposés dans une cupule (CENTRIFICHEM CUPS 0,25 mL ; KARTELL) rincée à l'éthanol 70% et à l'eau exempte de nucléases (Water, Sterile Nuclease-free, TFS Fisher BioReagents). Le tampon de lyse est régulièrement renouvelé au cours de la récolte. Les AC-DC contenues dans le tampon sont récoltées au fur et à mesure, placées dans un tube Eppendorf 2 mL et conservées dans la glace le temps de la récolte. Une fois celle-ci terminée, les tubes de tampon de lyse contenant les AC-DC sont conservés dans un congélateur de laboratoire à -80° C. Une récolte d'environ 20 000 plantes correspond à un échantillon.

I-4 Extraction et quantification de l'ARN des AC-DC

L'extraction de l'ARN des AC-DC d'Arabidopsis est réalisée grâce au kit NucleoSpin® RNA Plus XS selon les instructions du constructeur. Le volume de tampon de lyse utilisé pour un échantillon étant trop grand pour être filtré par une seule colonne d'élimination d'ADN génomique, plusieurs colonnes « gDNA » sont utilisées. Puis, les aliquots d'échantillon sont regroupés et filtrés dans une seule colonne « RNA binding » afin de récupérer un maximum d'ARN. La concentration en ARN est ensuite évaluée par dosage fluorométrique (Quantus™ fluorometer, Promega). Un aliquot de 2 µL est prélevé pour l'analyse qualité et l'ensemble (échantillons et aliquots) est conservé à -80°C.

I-5 Analyse qualité de l'ARN

La qualité des ARN des AC-DC est évaluée grâce au bioanalyseur (2100 Bioanalyzer system, Agilent), en utilisant le kit RNA 6000 Pico Kit Quick Start (Agilent) (Figure SII-2).

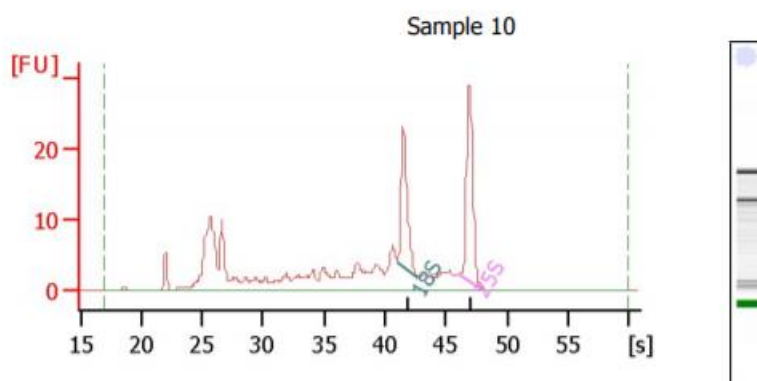


Figure SII-2 : Résultat obtenu au bioanalyseur pour l'échantillon d'AC-DC d'Arabidopsis C1

On observe sur ce graphique deux pics correspondant aux ARN 18S (42 secondes) et 25S (47 secondes), ainsi qu'une ligne de base faible et un pic de dégradation (25 secondes) peu élevé. Le RIN (*RNA integrity number*) de cet échantillon est estimé par l'appareil à 7,2. Ces critères indiquent une qualité correcte de l'échantillon ARN.

Trois échantillons (C1, C2 et C3) dont la qualité et la quantité d'ARN est suffisante ont été envoyés sur la plateforme MGX pour la création de bibliothèques et le séquençage.

II. Séquençage

Une re-vérification de la qualité des ARN envoyés (Figure SII-3), la préparation des banques et le séquençage ont été réalisés par la plateforme MGX (Montpellier GenomiX). Des banques d'ARNm ont été préparées grâce au kit *TruSeq Stranded mRNA Sample Preparation* (Illumina), incluant une étape de sélection des ARN polyadénylés. La validation des banques est réalisée grâce à une quantification de l'ADN sur Fragment Analyzer (kit *Standard Sensitivity NGS*, Illumina) ainsi que par qPCR (Light Cycler 480, Roche).

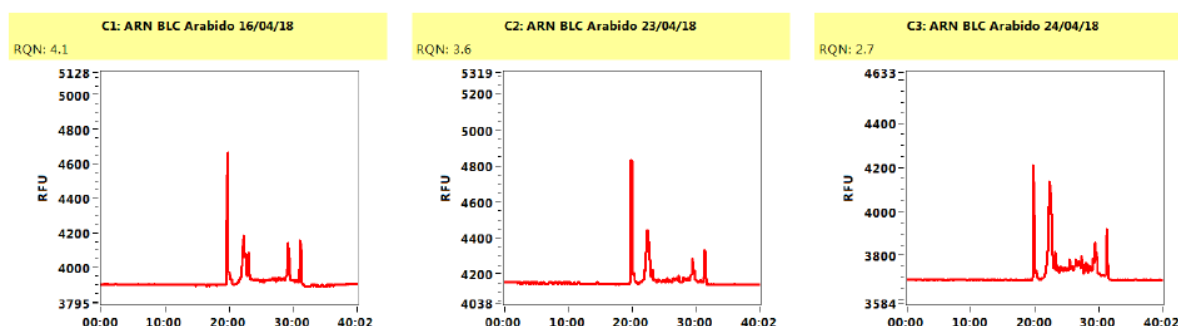


Figure SII-3 : Seconde analyse qualité des trois échantillons d'ARN d'AC-DC envoyés (C1, C2 et C3)

On observe sur ces graphiques deux pics correspondant aux ARN 18S et 25S, ainsi qu'une ligne de base faible et un pic de dégradation peu élevé pour C1 et C2. Le RQN (*RNA quality number*) de ces échantillons est estimé par l'appareil à 4,1 pour C1 et 3,6 pour C2. Ces critères indiquent une qualité correcte de ces échantillons ARN. L'échantillon C3 présente une ligne de base élevée, ainsi qu'un pic de dégradation important et un RQN estimé à 2,7. Ces critères indiquent que cet échantillon a subi une dégradation.

Suite à ces étapes, une seule banque de qualité suffisante pour le séquençage a pu être réalisée à partir de l'échantillon C1. En effet, l'échantillon C3 s'est dégradé lors du transport et a donc été mis de côté, et la réalisation d'une banque d'ARN à partir de l'échantillon C2 a échoué. Le séquençage *paired-end* de cette unique librairie a ensuite été réalisé sur un HiSeq 2500 d'Illumina à l'aide du kit *cluster generation* (Illumina).

III. Analyses bio-informatiques

La première partie de cette analyse bio-informatique a été réalisée par la plateforme MGX (Montpellier GenomiX). Elle comprenait :

- Le démultiplexage des fichiers de séquençage, réalisé avec *bcl2fastq*
- Une analyse qualité des *reads* (Figures SII-4 et SII-5), réalisée avec *Sequencing Analysis Viewer* et *FastQC*, qui a permis d'observer la bonne qualité des *reads* et du séquençage.

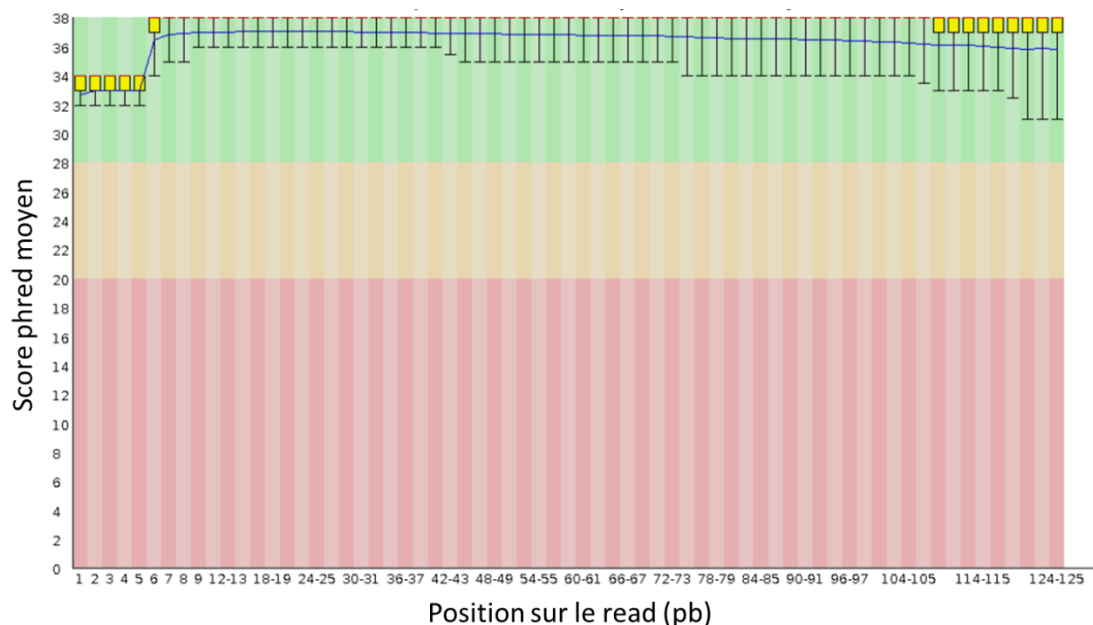


Figure SII-4 : Scores de qualité des bases séquencées (exemple des reads 1)

On observe ici le score *phred* moyen des bases séquencées (ligne bleue). Un score phred de 30 correspond à une précision de 99,9% et est considéré comme un seuil à partir duquel le séquençage des bases est de très bonne qualité. Le score maximal est 40, soit une précision de 99,99%. Des résultats similaires ont été observés pour les *reads* 2.



Figure SII-5 : Pourcentage de bases N par cycle (exemple des *reads* 1)

On observe ici le pourcentage de bases N (non déterminées) en fonction de la position sur le *read*. Ce pourcentage atteint 1% pour la première base des *reads* 1, et est à 0% pour toutes les autres positions, montrant donc une très bonne qualité du séquençage. Des résultats similaires ont été observés pour les *reads* 2.

- Un screening des contaminants, réalisé avec *FastQ Screen*. Ce screening a permis de mettre en évidence que la grande majorité des *reads* s'alignait sur le génome d'*Arabidopsis*, montrant qu'il ne semblait donc pas y avoir de contamination de l'échantillon avec un autre organisme.
- Le trimming des adaptateurs, réalisé avec *Trim Galore!*. Celui-ci permet de retirer les séquences d'adaptateurs dans les *reads*, de supprimer les bases de mauvaise qualité (score phred inférieur à 20) et d'éliminer les *reads* de moins de 20 bases. A la suite de cette étape, il restait environ 279 millions de *reads*.
- Les séquences ont ensuite été alignées avec TopHat2 et le comptage des gènes a été réalisé avec *FeatureCounts* (V1.6.2).

À la suite de cela, le nombre de copies des gènes exprimés (*genecount*) ainsi obtenu a été comparé à d'autres profils transcriptomiques retrouvés dans la littérature, notamment ceux de Li *et al.*, 2016, qui ont réalisé une analyse du transcriptome de différents types cellulaires de la racine d'*Arabidopsis*. Pour cela, les fichiers FastQ correspondants aux transcriptomes de la columelle (PET), des cellules latérales de la coiffe (WER), des cellules du centre quiescent (WOX) ainsi que de l'ensemble des cellules de la zone méristématique (MER) et de la racine (WR) ont été téléchargés sur Galaxy avec l'outil *Download and extract*

reads in FASTA/Q (V2.9.1.3). L'ensemble des fichiers FastQ, ainsi que celui correspondant au transcriptome des AC-DC, ont ensuite été "nettoyés" avec *Trimmomatic* (V0.36.5) avec les options *Illuminaclip* et *Sliding windows*. Cette étape a permis de sélectionner les *reads* de bonne qualité, de supprimer les séquences d'adaptateurs ainsi que les bases de mauvaise qualité. Suite à cela, les *reads* ont été alignés sur le génome d'Arabidopsis (Version TAIR10) avec *Hisat2* (V2.1.0), le comptage des gènes a ensuite été réalisé avec *htseq-count* (V0.9.1) et le fichier d'annotation *Araport11* (*Athaliana_447_Araport11*).

Les *genecounts* ainsi obtenus ont été analysés à l'aide de différents outils afin de faire :

- des analyses en composante principale avec *DESeq2* (V2.11.40.6) pour observer la variance entre les différents types cellulaires,
- des comparaisons des listes des gènes ayant été détectés au moins une fois (*genecount*>0) pour chaque type cellulaire et réplicat avec *JVenn* (jvenn.toulouse.inra.fr). Cette étape permet de mettre en évidence des listes de gènes détectés spécifiques d'un type cellulaire comparé à un autre, ou encore la liste de leurs gènes communs.
- des analyses des listes des gènes spécifiques détectés dans un type cellulaire ou dans plusieurs types cellulaires avec l'outil *statistical overrepresentation test* de Panther (pantherdb.org). Ce test permet de mettre en évidence la sur-représentation statistique d'une voie métabolique, grâce à la comparaison du nombre de gènes appartenant à une voie dans la liste donnée et le nombre de gènes appartenant à cette même voie dans le génome.
- des analyses des *genecounts* grâce à *Mapman*, afin de comparer l'expression des gènes pour différentes voies. Cette approche permet d'ajouter une notion quantitative à la détection des gènes, puisqu'elle prend en compte les *counts* pour chaque gène.

IV. Résultats préliminaires

IV-1 Analyse de la variance entre les différents échantillons

Dans un premier temps, la distance entre les échantillons a été observée afin de déterminer si l'échantillon d'AC-DC est proche ou non des autres types cellulaires

racinaires (Figure SII-6). Cette analyse a permis de remarquer que l'échantillon d'AC-DC apparaît très éloigné de tous les types cellulaires et tissus racinaires étudiés par Li *et al.* La columelle (PET), les cellules latérales de la coiffe (WER), le centre quiescent (WOX) et la zone méristématique (MER) sont très proches entre eux mais un peu plus éloignés de la racine entière (WR).

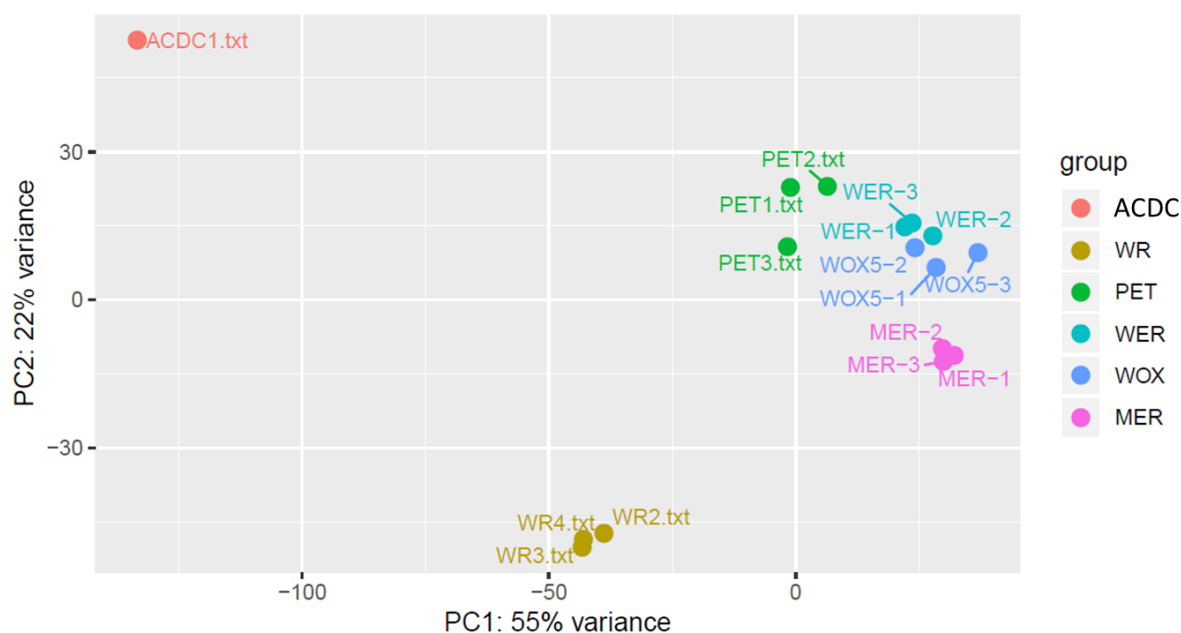


Figure SII-6 : Analyse en composante principale (ACP) de la variance entre les AC-DC et les différents échantillons de Li *et al.*, 2016

AC-DC = cellules frontières (*root Associated Cap-Derived cells*) ; WR = racine entière (*whole root*) ; PET = marqueur des cellules de la columelle ; WER = marqueur des cellules latérales de la coiffe ; WOX = marqueur du centre quiescent ; MER = zone méristématique. L'ACP est expliquée à 77% par les axes 1 et 2. Graphique réalisé avec DeSeq2.

IV-2 Analyse des gènes spécifiques aux différents types cellulaires

L'étape suivante a été d'observer la répartition des gènes détectés au moins une fois sur un diagramme de Venn, afin de comparer l'échantillon d'AC-DC de racine ainsi que les différents types cellulaires étudiés par Li *et al.* Pour chacun de ces types cellulaires, les trois réplicats ont été analysés. Lorsque l'on compare les gènes détectés dans les trois réplicats de cellules de la columelle (PET) avec ceux des AC-DC (Figure SII-7), on observe 4447 gènes qui semblent être spécifiques aux AC-DC. Certains gènes sont retrouvés dans le transcriptome des AC-DC ainsi que dans au moins l'un des réplicats de cellules de la columelle ; par exemple, 209 gènes détectés dans le transcriptome des AC-DC le sont aussi dans celui du réplicat 1 du transcriptome de cellules de la columelle, et 9769 gènes sont détectés dans

chacun des transcriptomes de la columelle et des AC-DC. Parmi tous les gènes détectés dans au moins l'un des réplicats de la columelle, 4028 n'ont pas été détectés dans les AC-DC. De plus, 12011 gènes sont détectés au moins une fois dans les trois réplicats de la columelle, et parmi ceux-ci 2242 sont spécifiques à celle-ci.

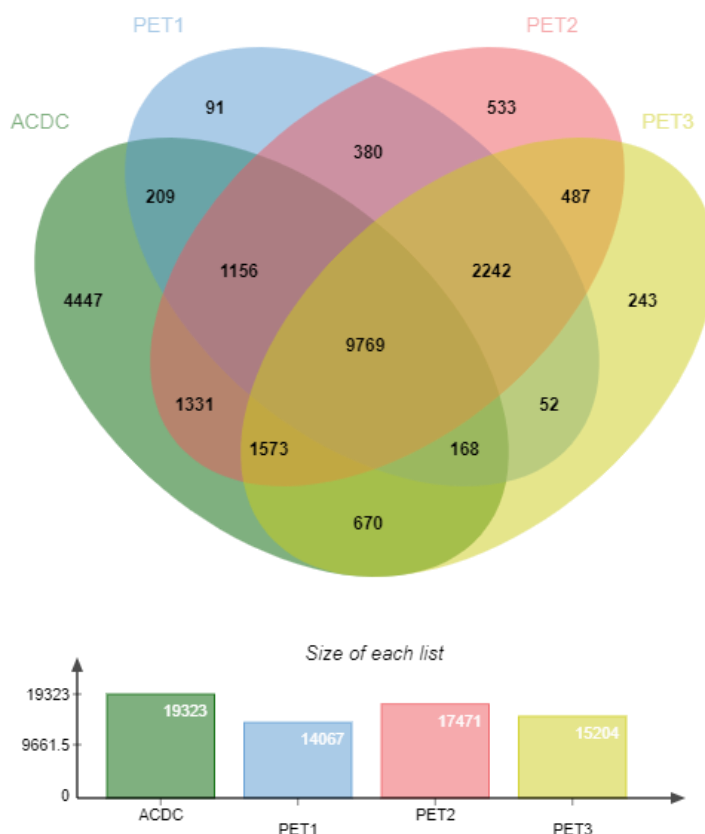


Figure SII-7 : Diagramme de Venn des listes de gènes détectés au moins une fois dans les trois réplicats de transcriptomes de cellules de la columelle (Li *et al.*, 2016) et dans le transcriptome des AC-DC

Pour chaque échantillon (AC-DC, PET1, PET2, PET3) la liste de leurs gènes détectés est représentée par une ellipse de couleur. Pour les AC-DC, 19323 gènes ont été détectés, contre 14067 pour PET1, 17471 pour PET2, et 15204 pour PET3. Pour chaque liste, on observe le nombre de gènes qui leur sont spécifiques, il y a donc 4447 gènes spécifiquement détectés dans le transcriptome des AC-DC, 91 dans le transcriptome de PET1, 533 dans celui de PET2, et 243 dans le transcriptome de PET3. Les zones de chevauchement entre les ellipses correspondent aux gènes en communs entre ces listes. Il y a donc 209 gènes ayant été détectés dans le transcriptome des AC-DC et celui de PET1, 1331 gènes détectés dans les transcriptomes des AC-DC et PET2, et 243 gènes détectés dans ceux des AC-DC et PET3. 9769 gènes ont été détectés dans les transcriptomes des AC-DC et des trois réplicats de la columelle, et 2242 gènes ont été détectés spécifiquement dans les trois réplicats de la columelle. Graphique réalisé avec JVenn à partir des données de cette analyse et de celle de Li *et al.*, 2016.

Lorsque l'on compare les gènes détectés dans le transcriptome des AC-DC à celui des trois réplicats du transcriptome des cellules latérales de la coiffe (WER) (Figure SII-8), on observe que 1578 gènes sont détectés uniquement dans le transcriptome des AC-DC, et 4325 gènes sont détectés spécifiquement dans les trois réplicats des cellules latérales de la

coiffe. De plus, 15461 gènes ont été détectés dans l'ensemble des trois réplicats de cellules latérales de la coiffe et dans les AC-DC.

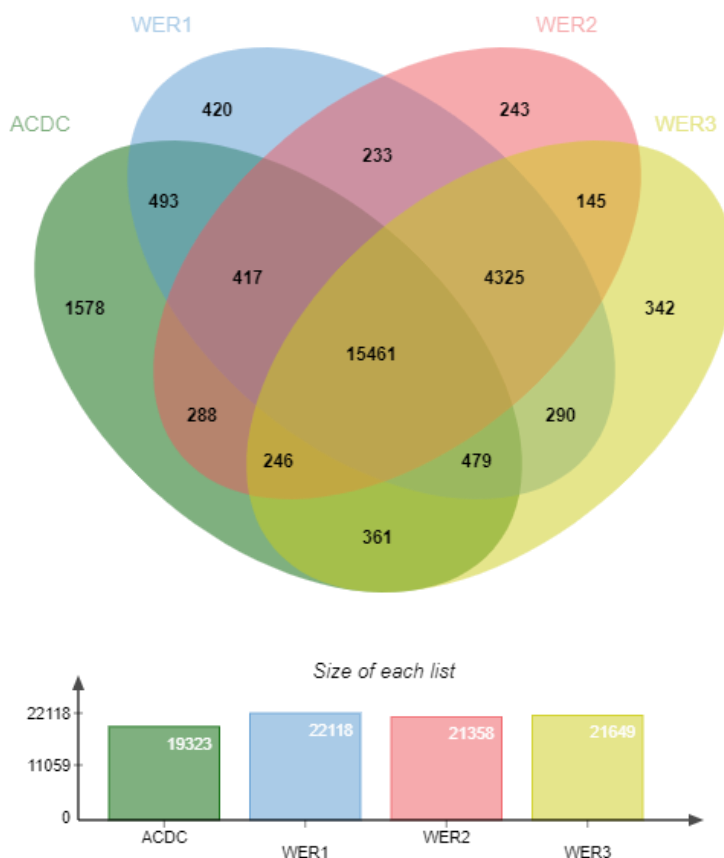


Figure SII-8 : Diagramme de Venn des listes de gènes détectés au moins une fois dans les trois réplicats de transcriptomes de cellules latérales de la coiffe (Li *et al.*, 2016) et dans le transcriptome des AC-DC

Pour chaque échantillon (AC-DC, WER1, WER2, WER3) la liste de leurs gènes détectés est représentée par une ellipse de couleur. Pour les AC-DC, 19323 gènes ont été détectés, contre 22118 pour WER1, 21358 pour WER2, et 21649 pour WER3. Pour chaque liste, on observe le nombre de gènes qui leur sont spécifiques, il y a donc 1578 gènes spécifiquement détectés dans le transcriptome des AC-DC, 420 dans le transcriptome de WER1, 243 dans celui de WER2, et 342 dans le transcriptome de WER3. 493 gènes ont été détectés dans le transcriptome des AC-DC ainsi que celui de WER1, 288 gènes ont été détectés dans les transcriptomes des AC-DC et WER2, et 361 gènes ont été détectés dans ceux des AC-DC et WER3. 15461 gènes ont été détectés dans les transcriptomes des AC-DC et des trois réplicats de cellules de la columelle, et 4325 gènes ont été détectés spécifiquement dans les trois réplicats de cellules latérales de la coiffe. Graphique réalisé avec JVenn.

Il convient maintenant de réaliser le même type d'analyse pour les cellules du centre quiescent (WOX) et de la zone méristématique (MER).

IV-3 Analyse des voies spécifiques aux AC-DC et aux cellules de la columelle

Les listes de gènes ayant été spécifiquement détectés dans un tissu ont été analysées par un test d'*overrepresentation* de Panther afin d'observer les voies auxquelles appartiennent ces gènes. Les gènes correspondants aux 4447 gènes spécifiquement détectés dans les AC-DC en comparaison avec le transcriptome des cellules de la columelle (Figure SII-7) ont permis de mettre en évidence trois voies spécifiques aux AC-DC par rapport à la columelle (Tableau SII-1).

Tableau SII-1 : Voies représentées par les gènes détectés uniquement dans les AC-DC en comparaison avec la columelle. Seules les voies avec un enrichissement minimum de deux fois sont affichées. FDR = *False discovery rate*. Tableau généré par le test d'*overrepresentation* de Panther

Only in AC-DC vs columella					
GO biological process complete	REFLIST (27416)	Input (4267)	Expected	Enrichment	FDR
positive regulation of proteasomal ubiquitin-dependent protein catabolic process (GO:0032436)	44	18	6,85	2,63	2,93E-02
positive regulation of proteasomal protein catabolic process (GO:1901800)	45	18	7	2,57	3,29E-02
positive regulation of ubiquitin-dependent protein catabolic process (GO:2000060)	45	18	7	2,57	3,28E-02

Sur les 4447 gènes entrés dans Panther, 4267 ont pu être assignés à leur annotation (Tableau SII-1, colonne *Input*), permettant d'obtenir les *gene ontologies* (GO) associés à chacun des gènes entrés. Les gènes ayant des GO appartenant à la catégorie *biological process* ont ensuite été sélectionnés et comptés par catégories de GO. Ainsi, 18 gènes possèdent une annotation GO correspondant à la régulation positive du catabolisme des protéines par le protéasome (GO : 1901800) dans la liste des gènes spécifiques aux AC-DC en comparaison avec la columelle. Ces comptages ont ensuite été comparés avec la liste de référence correspondant à la totalité du génome d'*Arabidopsis* (Tableau SII-1, colonne *REFLIST*), et notamment à une valeur statistique à partir de laquelle la voie correspondante à ce GO est considérée comme sur-représentée (Tableau SII-1, colonne *Expected*). Cette comparaison permet d'obtenir une valeur d'enrichissement ainsi qu'une *p-value* corrigée (Tableau SII-1, colonne FDR). Ce test permet donc de mettre en évidence que les voies de régulation positive du catabolisme des protéines par le protéasome (GO : 1901800), dépendantes de l'ubiquitine (GO : 2000060) et dépendantes du complexe ubiquitine-protéasome (GO : 0032436) présentent toutes les trois un enrichissement d'environ 2,5 fois par rapport au seuil statistique attendu. Il semblerait donc qu'une spécificité des AC-DC par rapport aux cellules de la columelle soit l'activation des voies de l'ubiquitination et des

protéasomes. Ces acteurs permettent notamment la dégradation des protéines dont la synthèse est anormale, mais aussi des régulateurs de certaines voies métaboliques (Sharma *et al.*, 2016). Ils jouent donc un rôle très important dans le métabolisme biologique des plantes ainsi que dans la signalisation hormonale ou encore lors des stress abiotiques ou biotiques. Il a en effet été montré que ceux-ci jouaient un rôle dans la réponse de défense du tabac à l'élicitation par des molécules issues des oomycètes du genre *Phytophthora* (Suty *et al.*, 2003).

Les 2242 gènes spécifiques aux cellules de la columelle par rapport aux AC-DC et ayant été détectés au moins une fois dans les trois réplicats (Figure SII-7) ont ensuite été analysés de la même manière avec Panther (Tableau SII-2). Cette analyse a permis de mettre en évidence 29 voies dont l'enrichissement statistique est d'au moins deux fois par rapport au seuil.

Tableau SII-2 : Voies représentées par les gènes détectés uniquement dans la columelle en comparaison avec les AC-DC. Seules les voies avec un enrichissement minimum de deux fois sont affichées. FDR = *False discovery rate*. Tableau généré par le test d'*overrepresentation* de Panther

Only in columella vs AC-DC					
GO biological process complete	REFLIST (27416)	Input (2208)	Expected	Enrichment	FDR
mitochondrial electron transport, NADH to ubiquinone (GO:0006120)	10	6	0,81	7,45	4,04E-02
ribosomal subunit export from nucleus (GO:0000054)	15	7	1,21	5,79	4,49E-02
rRNA-containing ribonucleoprotein complex export from nucleus (GO:0071428)	15	7	1,21	5,79	4,44E-02
ribosome localization (GO:0033750)	15	7	1,21	5,79	4,40E-02
mitochondrial ATP synthesis coupled electron transport (GO:0042775)	36	16	2,9	5,52	9,05E-05
mitochondrial respiratory chain complex I assembly (GO:0032981)	21	9	1,69	5,32	1,67E-02
ATP synthesis coupled electron transport (GO:0042773)	41	16	3,3	4,85	2,94E-04
respiratory electron transport chain (GO:0022904)	48	17	3,87	4,4	3,91E-04
oxidative phosphorylation (GO:0006119)	46	16	3,7	4,32	8,43E-04
NADH dehydrogenase complex assembly (GO:0010257)	26	9	2,09	4,3	4,53E-02
endoplasmic reticulum organization (GO:0007029)	32	11	2,58	4,27	1,66E-02
mitochondrial respiratory chain complex assembly (GO:0033108)	39	13	3,14	4,14	6,92E-03
cytoplasmic translation (GO:0002181)	71	17	5,72	2,97	1,65E-02
ribosome assembly (GO:0042255)	85	19	6,85	2,78	1,62E-02
electron transport chain (GO:0022900)	94	21	7,57	2,77	8,67E-03
ribosomal large subunit biogenesis (GO:0042273)	107	23	8,62	2,67	9,82E-03
protein folding (GO:0006457)	169	36	13,61	2,64	1,68E-04
ATP metabolic process (GO:0046034)	123	24	9,91	2,42	2,34E-02
mitochondrion organization (GO:0007005)	173	33	13,93	2,37	2,33E-03
peptide biosynthetic process (GO:0043043)	436	83	35,11	2,36	1,23E-08
translation (GO:0006412)	431	82	34,71	2,36	1,36E-08
response to cadmium ion (GO:0046686)	343	64	27,62	2,32	3,32E-06
peptide metabolic process (GO:0006518)	516	94	41,56	2,26	6,19E-09
amide biosynthetic process (GO:0043604)	502	89	40,43	2,2	5,17E-08
ribonucleoprotein complex assembly (GO:0022618)	154	27	12,4	2,18	3,27E-02
ribonucleoprotein complex subunit organization (GO:0071826)	157	27	12,64	2,14	3,80E-02
cellular amide metabolic process (GO:0043603)	642	110	51,7	2,13	5,40E-09
response to heat (GO:0009408)	205	34	16,51	2,06	1,80E-02
establishment of protein localization to organelle (GO:0072594)	266	43	21,42	2,01	5,97E-03

Sur les 2242 gènes entrés dans Panther, 2208 ont pu être assignés à leur annotation. La voie dont le GO présente l'enrichissement le plus fort est celle du transport des électrons dans la mitochondrie (GO : 0006120 ; enrichissement de 7,45 fois par rapport à la valeur seuil). Parmi les voies mises en évidence dans ce tableau, sept autres font référence à des processus biologiques de la mitochondrie, et notamment au transport d'électrons (GO : 0042775 ; GO : 0042773 ; GO : 0022904 ; GO : 0022900), à la chaîne respiratoire (GO : 0032981 ; GO : 0033108 ; GO : 0010257) et à l'organisation de la mitochondrie (GO : 0007005). Il semble donc que l'activité des mitochondries soit plus importante dans les cellules de la columelle que dans les AC-DC. Sept autres voies présentant un enrichissement supérieur à 2 font référence aux complexes ribosomiaux, notamment à leur localisation (GO : 0000054 ; GO : 0071428 ; GO : 0033750) ou encore à leur synthèse et assemblage et structure (GO : 0042255 ; GO : 0042273 ; GO : 0022618 ; GO : 0071826). Une spécificité des cellules de la columelle par rapport aux AC-DC serait donc une synthèse protéique plus importante grâce à la transcription de gènes de synthèse et mise en place des complexes ribosomiaux.

IV-4 Analyse Mapman de l'expression des gènes des AC-DC et des cellules de la columelle

Pour cette analyse, les fichiers de comptage des gènes obtenus avec *HTSeq-count* ont été utilisés. Dans cette partie, seuls les résultats pour le transcriptome des AC-DC et le réplicat 1 du transcriptome des cellules de la columelle (PET1) de Li *et al.* (2016) apparaissent. Ces deux échantillons ont 11 302 gènes en communs, 2765 spécifiques à PET 1 et 8021 spécifiques aux AC-DC (Figure SII-7).

Pour chacune des figures suivantes (Figures SII-9, SII-10, SII-11 et SII-12), les gènes ont été associés à leur annotation, permettant de leur assigner des GO (appelés *bins* sur le logiciel Mapman). Sur les figures, différentes catégories de *bins* apparaissent, et pour chacune d'elles des carrés correspondant chacun à un gène dont l'annotation contient ce *bin*. Un code couleur est ensuite associé à chaque gène, en fonction du comptage qui lui est associé. Ainsi, plus un gène est exprimé, plus celui-ci sera associé à une couleur rouge foncé. Un gène dont le *count* est de 0 sera donc affiché en blanc.

Les figures SII-9 et SII-10 correspondent à une représentation générale des voies métaboliques (*overview* des *bins* principaux) pour les transcriptomes des AC-DC (Figure SII-9) et de la columelle (Figure SII-10). On observe tout d'abord un plus grand nombre de gènes fortement exprimés dans la columelle par rapport aux AC-DC. On observe cependant la présence de gènes exprimés pour toutes les catégories de *bins*. Si l'on regarde en détail les *bins*, les gènes exprimés sont parfois semblables, comme par exemple pour la catégorie *hormones*, ou différents entre AC-DC et columelle, comme par exemple *metal handling* ou *proteins*.

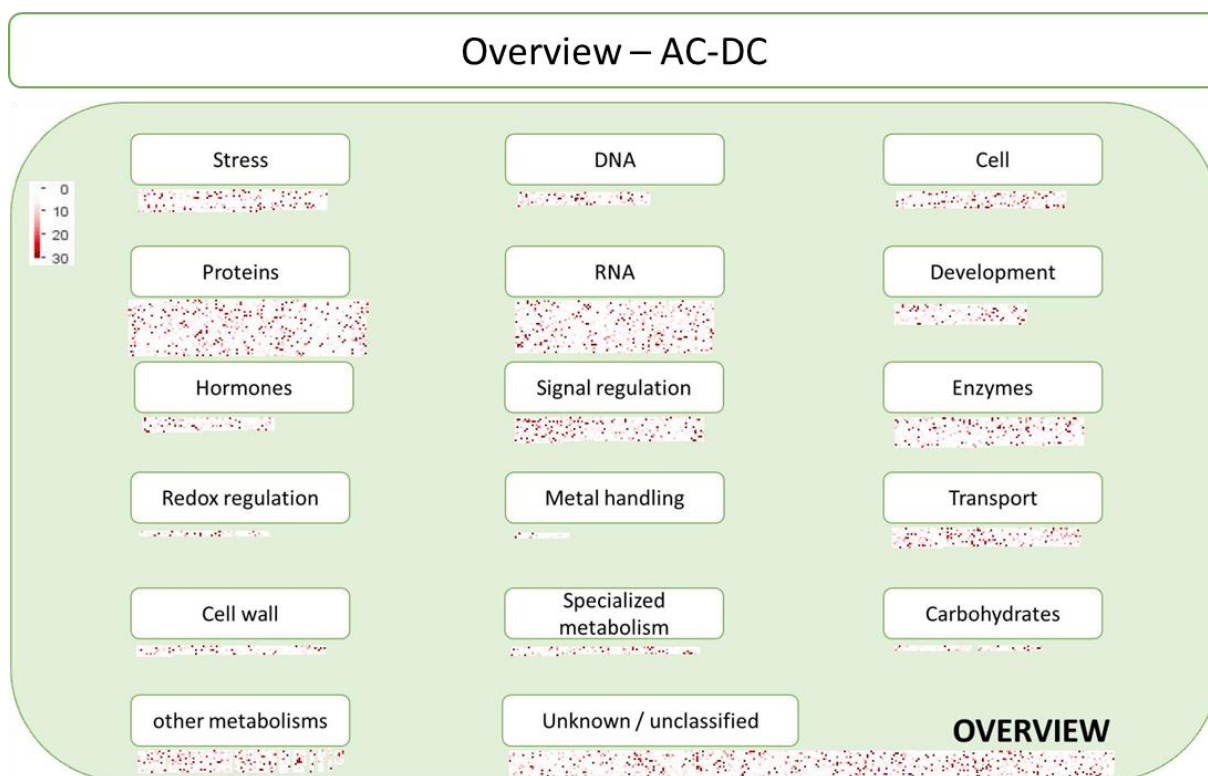


Figure SII-9 : *Overview* de l'expression des gènes des AC-DC

Les gènes sont représentés par des carrés. Ces carrés peuvent être de différentes couleurs, en fonction de leur *count*. Un gène n'ayant pas été détecté dans le transcriptome sera donc représenté en blanc, et un gène ayant été détecté 30 fois ou plus sera représenté en rouge foncé. Pour chaque catégorie, on observe la présence de gènes exprimés dans les AC-DC.

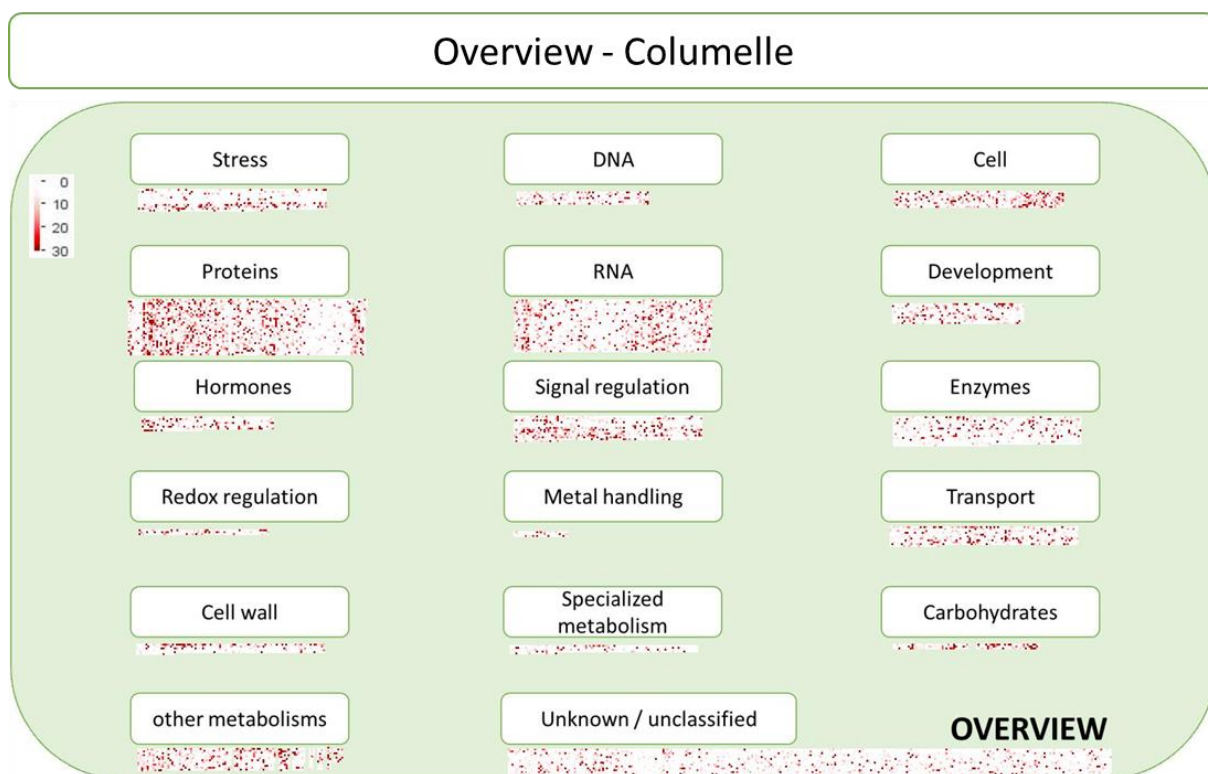


Figure SII-10 : *Overview* de l'expression des gènes de la columelle

Les gènes sont représentés par des carrés. Ces carrés peuvent être de différentes couleurs, en fonction de leur *count*. Un gène n'ayant pas été détecté dans le transcriptome sera donc représenté en blanc, et un gène ayant été détecté 30 fois ou plus sera représenté en rouge foncé. Pour chaque catégorie, on observe la présence de gènes exprimés dans la columelle.

Lorsque l'on se focalise sur l'expression des gènes impliqués dans le métabolisme des protéines, on observe des différences entre les gènes exprimés (Figures SII-11 et SII-12). Il y a par exemple très peu de gènes d'assemblage ou de ligation de cofacteurs qui sont exprimés dans les cellules de la columelle, contrairement aux AC-DC pour lesquelles on retrouve plus de gènes exprimés et avec des *counts* plus importants.

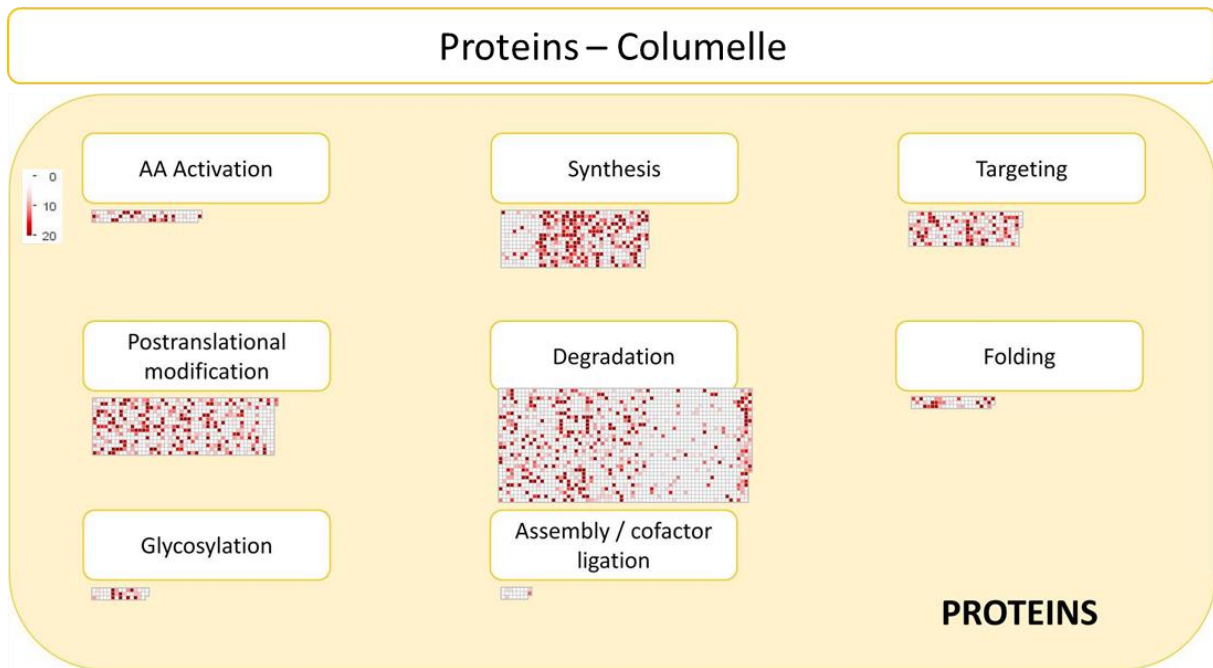


Figure SII-11 : Expression des gènes impliqués dans le métabolisme des protéines chez les cellules de la columelle

Les gènes sont représentés par des carrés. Ces carrés peuvent être de différentes couleurs, en fonction de leur *count*. Un gène n'ayant pas été détecté dans le transcriptome sera donc représenté en blanc, et un gène ayant été détecté 20 fois ou plus sera représenté en rouge foncé.

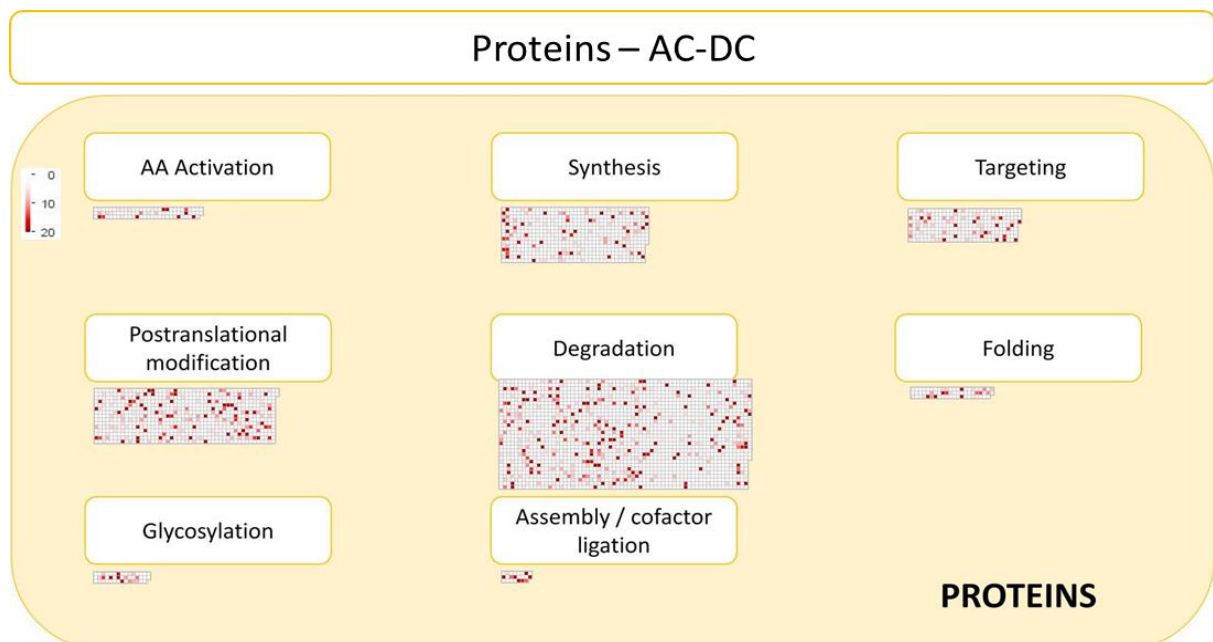


Figure SII-12 : Expression des gènes impliqués dans le métabolisme des protéines chez les AC-DC

Les gènes sont représentés par des carrés. Ces carrés peuvent être de différentes couleurs, en fonction de leur *count*. Un gène n'ayant pas été détecté dans le transcriptome sera donc représenté en blanc, et un gène ayant été détecté 20 fois ou plus sera représenté en rouge foncé.

V. Conclusion et perspectives

Cette analyse a tout d'abord permis d'établir pour la première fois le transcriptome des AC-DC d'*Arabidopsis thaliana*. Suite à cela nous avons réalisé une analyse bio-informatique préliminaire dans laquelle nous avons tenté de mettre en évidence les différences entre le transcriptome des AC-DC et celui d'autres types cellulaires de la racine et supposément proches des AC-DC : les cellules de la columelle et les cellules de la coiffe latérale. En effet les AC-DC sont issues de la maturation de ces deux types cellulaires.

Nous avons pu montrer dans un premier temps grâce à des ACP que le transcriptome des AC-DC semble très différent de tous les autres transcriptomes des différents types cellulaires réalisés par Li *et al.*, en 2016. Puis nous avons souhaité comparer les listes des gènes détectés dans ces différents transcriptomes grâce à des diagrammes de Venn. Cette analyse a été réalisée sur les AC-DC en comparaison avec les cellules de la columelle ainsi qu'avec les cellules latérales de la coiffe. Celle-ci nous a permis de mettre en évidence un nombre important de gènes communément détectés entre les AC-DC et ces deux types cellulaires, mais aussi des gènes qui semblent plus spécifiques aux AC-DC. Ces gènes ont été analysés avec le test d'*overrepresentation* de Panther, qui nous a permis de mettre en évidence les voies qui semblaient spécifiques aux AC-DC mais aussi celles qui seraient spécifiques aux cellules de la columelle car non détectées dans le transcriptome des AC-DC. Pour finir, nous avons visualisé plus globalement les transcriptomes des AC-DC et du réplicat 1 des cellules de la columelle ainsi que l'intensité de l'expression des gènes au sein de ces transcriptomes grâce à Mapman. Cette analyse a permis d'observer notamment des niveaux d'expression différents entre certaines catégories de gènes.

Dans un second temps, une analyse plus détaillée des gènes exprimés dans chacune des catégories de *bins* nous permettra d'observer plus finement ces différences de transcriptomes.

Pour la suite de cette étude, nous souhaiterions tout d'abord élargir celle-ci avec tous les types cellulaires dont le transcriptome a été réalisé par Li *et al.*, (2016) et étoffer l'analyse métatranscriptomique avec d'autres transcriptomes d'*Arabidopsis* en général et tissus de cellules racinaires en particulier afin de renforcer les spécificités et les ressemblances observées entre les AC-DC et les autres types cellulaires.

Annexe III : Figures supplémentaires

- I. Transcriptomique des AC-DC et des racines
- II. Analyse génomique de l'ADNex

Annexe III : Figures supplémentaires

Dans cette annexe sont présentées les figures supplémentaires à celles incluses dans les résultats de ce manuscrit.

I. Transcriptomique des AC-DC et des racines

I-1 Gènes sous régulés chez les AC-DC

Tableau III-1 : : Détail des gènes sous-exprimés chez les AC-DC de soja en réponse à PEP-13 classés en fonction de leur catégorie de bins

Une couleur est attribuée au Log2 du FC en fonction de sa valeur, plus celle-ci est faible plus elle apparaîtra en bleu foncé.

Stress			
ID		Transcript name	Log(FC)
glyma08g10960.1	moderately similar to (208)AT1G67360	rubber elongation factor (REF) family protein	-1,1837274
glyma10g00640.1	moderately similar to (470)AT1G04190	tetratricopeptide repeat (TPR)-containing protein	-1,1481248
glyma15g07940.1	weakly similar to (186)AT3G05500	rubber elongation factor (REF) family protein	-1,2584507
glyma17g11170.1	moderately similar to (256)AT4G19810	glycosyl hydrolase family 18 protein	-1,5338159
glyma11g08250.1	moderately similar to (275)AT4G24180	similar to thaumatin, putative [Arabidopsis thaliana] (TAIR:AT4G38660.2)	-1,5303427
glyma06g12900.1	moderately similar to (232)AT1G78780	pathogenesis-related family protein	-3,539566
glyma09g34100.1	weakly similar to (186)AT1G32210 ATDAD1	ATDAD1 (DEFENDER AGAINST APOPTOTIC DEATH 1)	-2,3353353
glyma01g43190.1	nearly identical (1439)AT1G64060 ATRBOHF	ATRBOHF (RESPIRATORY BURST OXIDASE PROTEIN F)	-1,8250742
glyma13g31640.1	moderately similar to (234)AT1G61105	disease resistance protein-related	-1,1671226
glyma09g29280.1	weakly similar to (138)AT1G17860	trypsin and protease inhibitor family protein / Kunitz family protein	-2,2737553
glyma08g44590.1	nearly identical (1056)AT5G56000	heat shock protein 81-4 (HSP81-4)	-1,3418236
glyma13g01400.1	highly similar to (890)AT5G57710	heat shock protein-related	-1,4233931
glyma20g24050.1	moderately similar to (266)AT5G23590	DNAJ heat shock N-terminal domain-containing protein	-4,8294687
glyma16g19300.1	moderately similar to (394)AT3G08970	DNAJ heat shock N-terminal domain-containing protein	-2,2812164
glyma16g23750.1	weakly similar to (118)AT4G36040	DNAJ heat shock N-terminal domain-containing protein (J11)	-5,7391706
glyma15g20400.1	highly similar to (998)AT1G79940	heat shock protein binding / unfolded protein binding	-4,4986615
glyma08g06950.1	nearly identical (1051)AT5G09590 HSC70-5	mtHSC70-2 (HEAT SHOCK PROTEIN 70)	-1,3661388
glyma18g01960.1	highly similar to (580)AT2G22360	DNAJ heat shock family protein	-1,2492815
glyma14g01530.1	nearly identical (1055)AT5G56000	heat shock protein 81-4 (HSP81-4)	-1,221662
glyma02g11890.1	nearly identical (1069)AT1G26850	dehydration-responsive family protein	-1,6528308
glyma07g09090.1	moderately similar to (348)AT3G18030 HAL3A	ATHAL3A (Arabidopsis thaliana Hal3-like protein A)	-5,0950694
glyma16g08410.1	nearly identical (1089)AT5G19690 STT3A	STT3A (STAUROSPOIN AND TEMPERATURE SENSITIVE 3-LIKE A)	-2,2058265
glyma02g05840.1	highly similar to (803)AT3G51070	dehydration-responsive protein-related	-1,9327314
DNA			
ID		Transcript name	Log(FC)
glyma17g11220.1	nearly identical (1123)AT2G07690	minichromosome maintenance family protein / MCM family protein	-1,6314868
glyma17g10770.1	moderately similar to (302)AT1G18800 NRP2	NRP2 (NAP1-RELATED PROTEIN 2)	-2,1597288
glyma06g46180.1	very weakly similar to (99.8)AT5G09600 SDH3-1	SDH3-1 (succinate dehydrogenase 3-1)	-3,1428607
glyma04g00340.1	weakly similar to (121)AT5G17690 LHP1	TFL2 (TERMINAL FLOWER 2)	-3,339524
glyma10g29360.1	highly similar to (646)AT4G34910	DEAD/DEAH box helicase, putative (RH16)	-1,063481
glyma04g05710.1	highly similar to (530)AT5G26180	NOL1/NOP2/sun family protein	-1,9745922
glyma17g06180.1	highly similar to (510)AT2G19490	recA family protein	-1,4642601
glyma18g01290.1	highly similar to (656)AT1G67630 POLA2	POLA2	-7,5678926
glyma03g39720.1	nearly identical (1354)AT1G67500 ATREV3	ATREV3 (Arabidopsis thaliana recovery protein 3)	-5,0395923
glyma06g00400.1	weakly similar to (108)AT5G17690 LHP1	TFL2 (TERMINAL FLOWER 2)	-1,6424499
glyma04g06010.1	highly similar to (527)AT2G26280 CID7	CID7	-2,636524
glyma15g04530.1	weakly similar to (144)AT5G02560 HTA12	HTA12	-1,0884
glyma18g50120.1	moderately similar to (216)AT4G40030	histone H3.2	-4,6261206
glyma15g04540.1	weakly similar to (148)AT5G02560 HTA12	HTA12	-1,1560894
glyma12g07030.1	weakly similar to (157)AT5G22880 H2B	H2B/HTB2 (HISTONE H2B)	-3,139133
glyma14g01510.1	weakly similar to (187)AT3G13226	regulatory protein RecX family protein	-1,7940255
glyma15g07380.1	nearly identical (1335)AT5G28740	transcription-coupled DNA repair protein-related	-1,4670538
glyma07g00710.1	highly similar to (953)AT1G79890	helicase-related	-1,7130955
glyma08g10780.1	highly similar to (844)AT1G27880	ATP-dependent DNA helicase, putative	-2,1286523

Cell			
ID		Transcript name	Log(FC)
glyma09g05950.1	weakly similar to (150)AT3G54070	ankyrin repeat family protein	-1,0746988
glyma03g42530.1	highly similar to (726)AT2G31820	ankyrin repeat family protein	-1,3682345
glyma08g03660.1	weakly similar to (190)AT2G19770 PRF5	PRF5 (PROFILIN5)	-1,2224864
glyma09g03430.1	moderately similar to (345)AT5G12380	annexin, putative	-1,5244777
glyma10g08480.1	nearly identical (1137)AT3G44730	kinesin motor protein-related	-1,1684923
glyma04g04580.1	very weakly similar to (92.0)AT1G47200 WPP2	WPP2 (WPP domain protein 2)	-5,8222523
glyma08g15940.1	weakly similar to (162)AT5G12320	ankyrin repeat family protein	-2,5029185
glyma17g29660.1	moderately similar to (253)AT5G19940	plastid-lipid associated protein PAP-related / fibrillin-related	-3,594144
glyma03g40780.1	highly similar to (548)AT3G04710	ankyrin repeat family protein	-5,4336886
glyma11g04750.1	highly similar to (858)AT1G50010 TUA2	TUA2 (tubulin alpha-2 chain)	-1,7407533
glyma05g28240.1	highly similar to (548)AT4G14150 KINESIN-12A	PAKRP1 (PHRAGMOPLAST-ASSOCIATED KINESIN-RELATED PROTEIN 1)	-1,5096791
glyma06g02650.1	highly similar to (832)AT1G75780 TUB1	TUB1 (tubulin beta-1 chain)	-1,4946306
glyma06g17920.1	moderately similar to (480)AT5G14170 CHC1	CHC1	-2,0325267
glyma02g02100.1	nearly identical (1007)AT3G53760	tubulin family protein	-1,3455262
glyma15g05490.1	highly similar to (572)AT1G76900 AtTLP1	AtTLP1 (TUBBY LIKE PROTEIN 1)	-3,8167534
glyma03g29860.1	highly similar to (798)AT3G57890	tubulin-specific chaperone C-related	-1,7572571
glyma01g40550.1	highly similar to (858)AT1G50010 TUA2	TUA2 (tubulin alpha-2 chain)	-4,6808195
glyma05g28420.1	moderately similar to (377)AT5G17900	similar to microfilament-associated protein-related [Arabidopsis thaliana] (TA	-1,2199751
glyma15g38010.1	moderately similar to (352)AT5G12380	annexin, putative	-3,2538457
glyma04g02610.1	highly similar to (833)AT1G75780 TUB1	TUB1 (tubulin beta-1 chain)	-2,5577085
glyma13g20680.1	nearly identical (1398)AT3G53230	cell division cycle protein 48, putative / CDC48, putative	-2,251424
glyma18g01540.1	moderately similar to (468)AT1G24490 ALB4	ALB4 (ALBINA 4)	-1,6055231
glyma02g09500.1	moderately similar to (283)AT1G14750 SDS	SDS (SOLO DANCERS)	-2,9018302
glyma05g01800.1	moderately similar to (306)AT5G50110	methyltransferase-related	-2,7674696
glyma13g10090.1	moderately similar to (471)AT1G44110 CYCA1	CYCA1	-2,2955604
glyma11g08960.1	moderately similar to (459)AT1G20610 CYCB2	CYCB2	-2,137426
glyma03g16440.1	highly similar to (842)AT3G54010 DE11	PAS1 (PASTICCINO 1)	-1,836453
glyma06g07380.1	moderately similar to (226)AT4G32420	peptidyl-prolyl cis-trans isomerase cyclophilin-type family protein	-1,1082486
glyma05g07830.1	moderately similar to (301)AT1G74070	peptidyl-prolyl cis-trans isomerase cyclophilin-type family protein	-1,3980632
glyma11g14770.1	highly similar to (773)AT5G59420	oxysterol-binding family protein	-1,4282625
glyma08g39930.1	nearly identical (1231)AT1G60070	clathrin binding	-2,1976314
glyma02g45580.1	highly similar to (987)AT1G67930	Golgi transport complex protein-related	-2,0560558
glyma20g24540.1	moderately similar to (342)AT3G54300 VAMP727	ATVAMP727 (Arabidopsis thaliana vesicle-associated membrane protein 727)	-1,6913631
glyma13g05680.1	moderately similar to (275)AT5G39510 VTI11	ATVT11/ATVT11A/SGR4/VTI11/VTI1A/ZIG (VESICLE TRANSPORT V-SNARE 11	-1,0572675
glyma08g43480.2	moderately similar to (275)AT1G16240 ATSYSP51	SYSP51 (SYNTAXIN OF PLANTS 51)	-5,0270796
glyma05g03310.1	highly similar to (679)AT5G50380 ATEXO70F1	ATEXO70F1 (exocyst subunit EXO70 family protein F1)	-3,0797143
Proteins			
ID		Transcript name	Log(FC)
glyma03g25750.1	highly similar to (832)AT3G13490 OVA5	OVAS (OVULE ABORTION 5)	-1,3544302
glyma20g22170.1	nearly identical (1116)AT5G26710	glutamate-tRNA ligase, putative / glutamyl-tRNA synthetase, putative / GluR	-2,2602885
glyma01g00960.1	highly similar to (854)AT4G26300 EMB1027	EMB1027 (EMBRYO DEFECTIVE 1027)	-1,263983
glyma08g03090.1	moderately similar to (433)AT5G35400	tRNA pseudouridine synthase family protein	-1,4725285
glyma16g08440.2	weakly similar to (155)AT3G18760	ribosomal protein S6 family protein	-2,2210674
glyma14g00900.2	weakly similar to (127)AT5G27820	ribosomal protein L18 family protein	-1,3017921
glyma18g02970.1	moderately similar to (286)AT1G08360	60S ribosomal protein L10A (RPL10aA)	-2,7389395
glyma08g02070.1	weakly similar to (186)AT5G52650	40S ribosomal protein S10 (RPS10C)	-1,5011224
glyma15g00610.1	weakly similar to (185)AT2G32060	40S ribosomal protein S12 (RPS12C)	-2,437589
glyma17g00820.1	moderately similar to (239)AT5G09500	40S ribosomal protein S15 (RPS15C)	-1,611985
glyma20g30730.1	moderately similar to (245)AT4G09800 RPS18C	RPS18C (S18 RIBOSOMAL PROTEIN)	-1,8977003
glyma02g08690.1	moderately similar to (244)AT4G09800 RPS18C	RPS18C (S18 RIBOSOMAL PROTEIN)	-4,3920875
glyma09g05030.1	weakly similar to (147)AT5G27700	40S ribosomal protein S21 (RPS21C)	-1,4176023
glyma12g04510.2	weakly similar to (120)AT2G21580	40S ribosomal protein S25 (RPS25B)	-1,6384895
glyma16g01460.1	moderately similar to (389)AT3G04770 RPSAB	RPSAB (40S RIBOSOMAL PROTEIN SA B)	-2,0013392
glyma09g00280.1	highly similar to (616)AT1G43170 EMB2207	ARP1 (ARABIDOPSIS RIBOSOMAL PROTEIN 1)	-1,3193291
glyma19g36450.1	moderately similar to (236)AT3G01790	ribosomal protein L13 family protein	-1,7994721
glyma08g10910.1	moderately similar to (257)AT1G67430	60S ribosomal protein L17 (RPL17B)	-1,4359912
glyma18g01110.1	moderately similar to (289)AT1G27400	60S ribosomal protein L17 (RPL17A)	-2,2655487
glyma15g02610.1	weakly similar to (179)AT4G29410	60S ribosomal protein L28 (RPL28C)	-2,2089171
glyma07g01540.1	weakly similar to (176)AT4G29410	60S ribosomal protein L28 (RPL28C)	-1,9260182
glyma10g00500.1	weakly similar to (196)AT1G36240	60S ribosomal protein L30 (RPL30A)	-1,9600227
glyma18g49340.1	weakly similar to (139)AT1G26880	60S ribosomal protein L34 (RPL34A)	-1,8753921
glyma15g01860.1	weakly similar to (117)AT1G52300	60S ribosomal protein L37 (RPL37B)	-2,1058135
glyma07g00700.2	weakly similar to (115)AT1G52300	60S ribosomal protein L37 (RPL37B)	-2,0916953
glyma15g03300.1	weakly similar to (110)AT3G59540	60S ribosomal protein L38 (RPL38B)	-1,5916462
glyma17g12430.1	weakly similar to (193)AT5G20160	ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	-1,1629878
glyma10g00890.1	moderately similar to (361)AT2G47610	60S ribosomal protein L7A (RPL7aA)	-2,0722792
glyma01g03180.1	moderately similar to (208)AT1G70600	60S ribosomal protein L27A (RPL27aC)	-1,038354
glyma17g09280.2	moderately similar to (208)AT1G70600	60S ribosomal protein L27A (RPL27aC)	-1,2734873
glyma20g30110.1	moderately similar to (282)AT5G15750	RNA-binding S4 domain-containing protein	-1,3979797

glyma18g02050.1	weakly similar to (180)AT4G39880	ribosomal protein L23 family protein	-1,7088593
glyma20g37480.1	very weakly similar to (80.5)AT1G54740	similar to structural constituent of ribosome [Arabidopsis thaliana] (TAIR:AT1G54740)	-2,4161234
glyma08g12110.1	moderately similar to (306)AT5G19830	aminoacyl-tRNA hydrolase	-1,6054467
glyma13g21500.1	moderately similar to (372)AT3G23620	brix domain-containing protein	-2,2050388
glyma18g52060.1	highly similar to (542)AT1G10840 TIF3H1	TIF3H1 (EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT H1)	-1,4885019
glyma05g00260.1	weakly similar to (108)AT4G11175	translation initiation factor IF-1, chloroplast, putative	-2,4921918
glyma14g39850.1	highly similar to (514)AT5G42220	ubiquitin family protein	-2,7471213
glyma16g05410.1	nearly identical (1097)AT2G13540 CBP80	ABH1 (ABA HYPERSENSITIVE 1)	-1,0732856
glyma13g31200.1	highly similar to (715)AT4G20980	eukaryotic translation initiation factor 3 subunit 7, putative / eIF-3 zeta, putative	-1,055931
glyma06g14690.1	moderately similar to (306)AT2G38560	transcription factor S-II (TFIS) domain-containing protein	-1,2767276
glyma06g14260.1	weakly similar to (177)AT3G49470 NACA2	NACA2 (NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX SUBUNIT ALPHA-LIKE)	-1,3877829
glyma16g00360.1	moderately similar to (318)AT1G57720	elongation factor 1B-gamma, putative / eEF-1B gamma, putative	-1,4118134
glyma16g04890.1	moderately similar to (496)AT3G26618 ERF1-3	ERF1-3 (EUKARYOTIC RELEASE FACTOR 1-3)	-4,9538813
glyma09g38380.1	highly similar to (609)AT5G36170 ATPRFB	HCF109 (HIGH CHLOROPHYLL FLUORESCENT 109)	-1,3262773
glyma11g01650.1	moderately similar to (232)AT1G71480	nuclear transport factor 2 (NTF2) family protein	-2,1488419
glyma05g34360.1	moderately similar to (457)AT3G20000 TOM40	TOM40 (translocase of the outer mitochondrial membrane 40)	-1,3419977
glyma13g34080.1	moderately similar to (216)AT4G03320 TIC20-IV	TIC20-IV (TRANSLOCAN AT THE INNER ENVELOPE MEMBRANE OF CHLOROPLAST)	-1,478144
glyma05g28300.1	moderately similar to (259)AT3G23710	chloroplast inner membrane import protein Tic22, putative	-2,2741778
glyma09g41070.1	highly similar to (901)AT3G52850 VSR1	A TELP/A TELP1/BP-80/BP80/BP80B/VSR-1/VSR1 (ARABIDOPSIS THALIANA EPITOPIC TAGGING)	-2,152063
glyma18g44580.1	highly similar to (915)AT3G52850 VSR1	A TELP/A TELP1/BP-80/BP80/BP80B/VSR-1/VSR1 (ARABIDOPSIS THALIANA EPITOPIC TAGGING)	-3,0997398
glyma01g38270.1	moderately similar to (312)AT2G18390 ATARLC1	TTN5 (TITAN 5)	-2,8550978
glyma08g20370.1	moderately similar to (237)AT1G60970	clathrin adaptor complex small chain family protein	-1,6837498
glyma06g05680.1	highly similar to (796)AT4G33080	protein kinase, putative	-2,062238
glyma20g35110.1	highly similar to (861)AT4G14350	protein kinase family protein	-2,0219195
glyma11g04180.1	moderately similar to (402)AT4G17640 CKB2	CKB2 (casein kinase II beta chain 2)	-2,1509094
glyma03g31820.1	nearly identical (1307)AT2G31320	poly (ADP-ribose) polymerase, putative / NAD(+) ADP-ribosyltransferase, putative	-2,1055706
glyma03g41180.1	highly similar to (860)AT5G28850	calcium-binding EF hand family protein	-1,4877771
glyma18g17100.1	highly similar to (635)AT1G10430 PP2A-2	PP2A-2 (protein phosphatase 2a-2)	-4,1437597
glyma17g33690.2	highly similar to (504)AT4G31750	protein phosphatase 2C, putative / PP2C, putative	-1,1642753
glyma08g23340.1	highly similar to (511)AT5G10930 SnRK3.24	CIPK5 (CBL-INTERACTING PROTEIN KINASE 5)	-1,6187835
glyma01g38270.1	moderately similar to (312)AT2G18390 ATARLC1	TTN5 (TITAN 5)	-2,8550978
glyma06g11410.2	moderately similar to (389)AT4G08500 ATMEKK1	MEKK1 (MYTOGEN ACTIVATED PROTEIN KINASE KINASE)	-1,4005487
glyma01g45030.1	highly similar to (605)AT5G19280 RAG1	KAPP (Kinase-associated protein phosphatase)	-1,4321104
glyma08g12370.1	highly similar to (638)AT4G18710 DWF12	BIN2 (BRASSINOSTEROID-INSENSITIVE 2)	-1,3670951
glyma01g02810.1	highly similar to (637)AT2G22560	kinase interacting protein-related	-2,2218938
glyma17g02580.1	highly similar to (713)AT1G54610	protein kinase family protein	-1,396178
glyma10g30030.1	highly similar to (704)AT1G54610	protein kinase family protein	-1,6654365
glyma10g42380.1	highly similar to (544)AT3G09880 ATB' BETA	ATB' BETA (Arabidopsis thaliana serine/threonine protein phosphatase 2A 55)	-2,4773605
glyma18g43950.1	moderately similar to (378)AT5G01700	protein phosphatase 2C, putative / PP2C, putative	-1,6050982
glyma14g06420.1	highly similar to (565)AT2G40120	protein kinase family protein	-2,3392484
glyma13g34990.1	moderately similar to (392)AT3G15260	protein phosphatase 2C, putative / PP2C, putative	-1,9989226
glyma19g00220.1	highly similar to (879)AT5G40440 MKK3	ATMKK3 (MITOGEN-ACTIVATED KINASE KINASE 3)	-1,415342
glyma12g06750.1	highly similar to (546)AT2G28940	protein kinase family protein	-1,7093868
glyma12g07870.1	highly similar to (581)AT1G07870	protein kinase family protein	-2,4935458
glyma05g28350.1	highly similar to (627)AT3G23750	leucine-rich repeat family protein / protein kinase family protein	-2,5103624
glyma13g40530.1	highly similar to (560)AT1G07870	protein kinase family protein	-2,0011647
glyma10g11840.1	highly similar to (660)AT5G06820 SRF2	SRF2 (STRUBBELIG-RECEPTOR FAMILY 2)	-1,8741525
glyma03g30260.1	highly similar to (556)AT2G41970	protein kinase, putative	-1,4867069
glyma19g43900.1	highly similar to (595)AT5G18460	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G23340.1)	-1,6215271
glyma12g30420.1	highly similar to (575)AT5G22860	serine carboxypeptidase S28 family protein	-3,4110384
glyma15g42010.1	highly similar to (610)AT1G55360	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G56530.1)	-1,3984041
glyma20g24740.1	highly similar to (758)AT3G10420	sporulation protein-related	-2,7055807
glyma03g41280.1	highly similar to (605)AT5G18460	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G23340.1)	-5,0333376
glyma20g23350.1	nearly identical (1123)AT1G50380	prolyl oligopeptidase family protein	-3,1017141
glyma14g06590.1	highly similar to (566)AT5G05740 ATEGY2	ATEGY2	-1,1157131
glyma19g44140.1	highly similar to (642)AT4G22720	glycoprotease M22 family protein	-1,4250301
glyma01g36130.1	highly similar to (840)AT5G67360 ARA12	ARA12	-3,3484144
glyma17g17850.1	nearly identical (1007)AT5G67360 ARA12	ARA12	-5,11178
glyma10g37900.1	nearly identical (1848)AT4G20850 TPP2	TPP2 (TRIPEPTIDYL PEPTIDASE II)	-1,3090649
glyma05g03760.1	highly similar to (532)AT1G04110 SDD1	SDD1 (STOMATAL DENSITY AND DISTRIBUTION)	-3,1678913
glyma10g04120.1	nearly identical (1023)AT2G31260 ATPG9	APG9 (AUTOPHAGY 9)	-4,703693
glyma12g15780.1	moderately similar to (363)AT5G50260	cysteine proteinase, putative	-1,2588873
glyma11g10600.1	highly similar to (685)AT5G23210 SCPL34	SCPL34	-1,7487775
glyma20g34110.1	highly similar to (664)AT4G17740	C-terminal processing protease, putative	-4,7042556
glyma17g04120.1	highly similar to (582)AT3G17180 SCPL33	SCPL33 (serine carboxypeptidase-like 33)	-1,3001842
glyma02g39040.1	highly similar to (966)AT5G58870 FTSH9	FTSH9 (FtsH protease 9)	-1,1913
glyma06g46540.1	highly similar to (963)AT5G22010 ATRFC1	ATRFC1 (REPLICATION FACTOR C 1)	-1,5938748
glyma09g40410.1	highly similar to (645)AT2G45500	similar to ATPase [Arabidopsis thaliana] (TAIR:AT3G27120.1)	-5,2735143
glyma19g38170.1	moderately similar to (234)AT3G52590 EMB2167	UBQ1 (EARLY-RESPONSIVE TO DEHYDRATION 16, UBIQUITIN EXTENSION PROTEIN 1)	-1,427502
glyma15g41230.1	moderately similar to (236)AT1G31340 NEDD8	RUB1 (RELATED TO UBIQUITIN 1)	-1,4807642
glyma06g13020.1	moderately similar to (291)AT4G27960 UBC9	UBC9 (UBIQUITIN CONJUGATING ENZYME 9)	-1,8836803
glyma09g41180.1	weakly similar to (144)AT5G01880	zinc finger (C3HC4-type RING finger) family protein	-3,0438344
glyma20g32670.1	highly similar to (710)AT1G68820	membrane protein, putative	-1,4962263
glyma04g36400.1	highly similar to (565)AT5G14420 RGLG2	RGLG2 (RING DOMAIN LIGASE2)	-3,1512544
glyma13g38720.1	moderately similar to (357)AT5G22000 RHF2A	CIC7E11	-1,1938317
glyma02g11960.1	moderately similar to (233)AT2G02960	zinc finger (C3HC4-type RING finger) family protein	-2,8942697

glyma17g13980.1	moderately similar to (390)AT4G11680	zinc finger (C3HC4-type RING finger) family protein	-2,1639798
glyma12g00330.1	highly similar to (692)AT1G57820 VIM1	ORTH2/VIM1 (VARIANT IN METHYLATION 1)	-1,6314944
glyma04g00490.1	weakly similar to (138)AT4G35070	similar to SBP1 (S-RIBONUCLEASE BINDING PROTEIN 1), protein binding / zinc	-3,5014396
glyma13g40150.1	moderately similar to (353)AT5G60580	zinc finger (C3HC4-type RING finger) family protein	-1,6863203
glyma05g28920.1	weakly similar to (172)AT3G25910	zinc ion binding	-1,1813926
glyma18g42230.1	moderately similar to (278)AT5G01160	e-cadherin binding protein-related	-4,962062
glyma01g03900.1	moderately similar to (288)AT5G40250	zinc finger (C3HC4-type RING finger) family protein	-2,0180063
glyma05g00270.1	moderately similar to (493)AT1G18610	kelch repeat-containing protein	-1,2595457
glyma18g50040.1	moderately similar to (342)AT5G39250	F-box family protein	-3,2573888
glyma07g39200.1	moderately similar to (248)AT4G22030	F-box family protein	-4,2097235
glyma15g09920.1	moderately similar to (495)AT5G46170	F-box family protein	-1,1990347
glyma02g45850.1	weakly similar to (169)AT5G54820	F-box family protein	-1,3744713
glyma06g13220.1	very weakly similar to (99.4)AT3G06240	F-box family protein	-1,138999
glyma15g05490.1	highly similar to (572)AT1G76900 AtTLP1	AtTLP1 (TUBBY LIKE PROTEIN 1)	-3,8167534
glyma08g41130.1	nearly identical (1192)AT1G26830 CUL3A	ATCUL3/ATCUL3A/CUL3/CUL3A (Cullin 3A)	-2,4127812
glyma02g16840.1	highly similar to (623)AT3G06190 ATBPM2	ATBPM2	-2,6832752
glyma02g43930.1	nearly identical (1782)AT5G06600 UBP12	UBP12 (UBIQUITIN-SPECIFIC PROTEASE 12)	-1,8516288
glyma01g02240.1	moderately similar to (345)AT4G17895 UBP20	UBP20 (UBIQUITIN-SPECIFIC PROTEASE 20)	-1,5870725
glyma13g20680.1	nearly identical (1398)AT3G53230	cell division cycle protein 48, putative / CDC48, putative	-2,251424
glyma06g35680.1	highly similar to (558)AT3G15180	proteasome-related	-1,380527
glyma11g13320.1	moderately similar to (275)AT2G43560	immunophilin / FKBP-type peptidyl-prolyl cis-trans isomerase family protein	-4,5160046
glyma08g12970.1	highly similar to (875)AT3G20050 ATTCP-1	ATTCP-1 (Arabidopsis thaliana T-complex protein 1 alpha subunit)	-2,629609
glyma18g01580.1	highly similar to (924)AT1G24510	T-complex protein 1 epsilon subunit, putative / TCP-1-epsilon, putative / chap	-1,0315149
glyma07g01190.1	highly similar to (783)AT3G13860	chaperonin, putative	-2,744505
glyma20g02110.1	weakly similar to (137)AT1G23100	10 kDa chaperonin, putative	-1,2951978
glyma06g08820.1	moderately similar to (348)AT1G27600	glycosyl transferase family 43 protein	-6,339938
glyma16g05340.1	highly similar to (763)AT1G08660	glycosyl transferase family 29 protein / sialyltransferase family protein	-1,8709258
glyma08g09480.1	weakly similar to (198)AT5G45680	FK506-binding protein 1 (FKBP13)	-2,410742
glyma01g38930.1	moderately similar to (305)AT4G37200 HCF164	HCF164 (High chlorophyll fluorescence 164)	-4,0488696
glyma05g36400.1	moderately similar to (357)AT3G10670 ATNAP7	ATNAP7 (Arabidopsis thaliana non-intrinsic ABC protein 7)	-1,8270038

RNA			
ID		Transcript name	Log(FC)
glyma08g45000.1	highly similar to (590)AT5G56130	transducin family protein / WD-40 repeat family protein	-4,1431055
glyma08g12850.1	highly similar to (538)AT5G46420	16S rRNA processing protein RimM family	-1,4170352
glyma10g26870.1	highly similar to (790)AT1G04510	transducin family protein / WD-40 repeat family protein	-2,1101663
glyma02g09380.1	highly similar to (798)AT2G01730 EDA26	ATCPSF73-II (CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR 73 KD	-1,0551257
glyma17g01320.1	highly similar to (776)AT1G62390	octicosapeptide/Phox/Bem1p (PB1) domain-containing protein / tetratricope	-1,4224426
glyma08g16210.1	moderately similar to (334)AT5G26880	tRNA/rRNA methyltransferase (SpoU) family protein	-1,9284989
glyma09g00310.1	moderately similar to (434)AT2G18510 EMB2444	EMB2444 (EMBRYO DEFECTIVE 2444)	-1,3444761
glyma13g01320.1	weakly similar to (144)AT4G30330	small nuclear ribonucleoprotein E, putative / snRNP-E, putative / Sm protein I	-1,6014682
glyma02g40900.1	moderately similar to (398)AT5G37370 ATSRL1	ATSRL1	-2,6125867
glyma17g06210.1	highly similar to (566)AT4G29750	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT3G23070.1)	-2,1423256
glyma10g43750.1	nearly identical (2245)AT3G55200	splicing factor, putative	-2,3339837
glyma19g39150.1	nearly identical (1599)AT3G11960	cleavage and polyadenylation specificity factor (CPSF) A subunit C-terminal d	-5,0414705
glyma19g40600.1	nearly identical (1270)AT3G62310	RNA helicase, putative	-2,368119
glyma19g38460.1	moderately similar to (319)AT3G44260	CCR4-NOT transcription complex protein, putative	-4,790013
glyma16g33820.1	very weakly similar to (87.8)AT5G41010	DNA-directed RNA polymerases I, II, and III 7 kDa subunit, putative	-3,5796273
glyma05g26040.1	moderately similar to (306)AT5G59180	DNA-directed RNA polymerase II	-4,353539
glyma02g44560.1	weakly similar to (188)AT1G54140 TAF9	TAFII21 (TATA BINDING PROTEIN ASSOCIATED FACTOR 21KDA SUBUNIT)	-3,3437724
glyma13g19420.1	highly similar to (958)AT3G53700 MEE40	MEE40 (maternal effect embryo arrest 40)	-1,179834
glyma08g02460.1	weakly similar to (155)AT4G23750 CRF2	CRF2 (CYTOKININ RESPONSE FACTOR 2)	-3,8640006
glyma19g44580.1	weakly similar to (155)AT1G01250	AP2 domain-containing transcription factor, putative	-3,2306077
glyma17g07010.1	moderately similar to (375)AT5G57390 AIL5	AIL5 (AINTEGUMENTA-LIKE 5)	-2,4005668
glyma08g43380.1	highly similar to (558)AT3G20300	extracellular ligand-gated ion channel	-2,369671
glyma05g37120.1	weakly similar to (154)AT4G23750 CRF2	CRF2 (CYTOKININ RESPONSE FACTOR 2)	-4,7633896
glyma02g45100.1	highly similar to (779)AT5G37020 ARF8	ARF8 (AUXIN RESPONSE FACTOR 8)	-2,7170603
glyma03g36710.1	moderately similar to (325)AT4G23980 ARF9	ARF9 (AUXIN RESPONSE FACTOR 9)	-3,114696
glyma03g28570.1	moderately similar to (241)AT3G57040 ATRR4	ARR9 (RESPONSE REACTOR 4)	-1,1927123
glyma17g06610.1	moderately similar to (204)AT5G57150	basic helix-loop-helix (bHLH) family protein	-1,564011
glyma19g30910.1	weakly similar to (197)AT2G41130	basic helix-loop-helix (bHLH) family protein	-1,3705435
glyma06g02970.1	weakly similar to (183)AT1G75540 STH2	STH2 (SALT TOLERANCE HOMOLOG2)	-4,9982667
glyma04g02960.1	weakly similar to (186)AT1G75540 STH2	STH2 (SALT TOLERANCE HOMOLOG2)	-2,4816148
glyma05g04260.1	moderately similar to (221)AT2G45190 FIL	AFO (ABNORMAL FLORAL ORGANS)	-1,8039511
glyma16g25550.1	moderately similar to (367)AT3G50700 ATIDD2	ATIDD2 (ARABIDOPSIS THALIANA INDETERMINATE(ID)-DOMAIN 2)	-3,3610315
glyma07g04870.1	highly similar to (663)AT1G72440 EDA25	EDA25 (embryo sac development arrest 25)	-1,8124322
glyma17g10130.1	moderately similar to (348)AT3G48160 E2L3	DEL1 (DP-E2F-like 1)	-1,6453791
glyma20g32770.1	moderately similar to (220)AT1G25550	myb family transcription factor	-3,2237482
glyma06g03210.1	moderately similar to (295)AT1G75430	transcription factor	-1,775322
glyma05g03650.1	moderately similar to (425)AT1G62990 IXR11	KNAT7 (Knotted-like Arabidopsis thaliana)	-2,8047593
glyma03g01000.1	weakly similar to (180)AT3G03660	similar to WOX12 (WUSCHEL RELATED HOMEBOX 12), DNA binding / transc	-3,1832592
glyma18g14910.1	weakly similar to (194)AT4G03250	homeobox-leucine zipper family protein	-1,307746
glyma12g00690.1	moderately similar to (467)AT4G25270	pentatricopeptide (PPR) repeat-containing protein	-1,3167231
glyma11g02400.1	moderately similar to (293)AT1G63910 ATMYB103	ATMYB103 (MYB DOMAIN PROTEIN 103)	-2,1116147
glyma20g34140.1	moderately similar to (232)AT3G24310 AtMYB71	MYB305 (myb domain protein 305)	-1,1425995
glyma13g01200.1	moderately similar to (256)AT5G57620 AtMYB36	MYB36 (myb domain protein 36)	-2,9855962
glyma14g00690.1	nearly identical (1170)AT5G09950	pentatricopeptide (PPR) repeat-containing protein	-2,6231177
glyma05g34010.1	nearly identical (1025)AT4G02750	pentatricopeptide (PPR) repeat-containing protein	-2,1297617
glyma15g14190.1	moderately similar to (201)AT3G24310 AtMYB71	MYB305 (myb domain protein 305)	-1,2635807

glyma17g17370.1	weakly similar to (200)AT5G67580 TRB2	ATTRB2/TRB2 (TELOMERE REPEAT BINDING FACTOR 2)	-1,6674325
glyma10g43190.1	moderately similar to (244)AT5G23280	TCP family transcription factor, putative	-3,3656902
glyma06g22240.1	weakly similar to (166)AT5G60970 TCP5	TCP5 (TEOSINTE BRANCHED1, CYCLOIDEA AND PCF TRANSCRIPTION FACTOR)	-3,3706944
glyma03g34960.1	moderately similar to (348)AT5G03680 PTL	PTL (PETAL LOSS)	-2,1823392
glyma04g33080.1	nearly identical (1779)AT1G68710	haloacid dehalogenase-like hydrolase family protein	-2,840487
glyma04g01210.1	weakly similar to (121)AT1G75390 ATBZIP44	ATBZIP44 (ARABIDOPSIS THALIANA BASIC LEUCINE-ZIPPER 44)	-1,0580178
glyma09g41040.1	nearly identical (1340)AT5G04930 ALA1	ALA1 (AMINOPHOSPHOLIPID ATPASE1)	-1,5203516
glyma09g02510.1	moderately similar to (220)AT1G31320 LBD4	LBD4 (LOB DOMAIN-CONTAINING PROTEIN 4)	-4,7552595
glyma07g01800.1	weakly similar to (189)AT3G16500 IAA26	PAP1 (PHYTOCHROME-ASSOCIATED PROTEIN 1)	-2,5084476
glyma06g17920.1	moderately similar to (480)AT5G14170 CHC1	CHC1	-2,0325267
glyma11g12180.1	weakly similar to (190)AT1G17880	nascent polypeptide-associated complex (NAC) domain-containing protein /	-2,3190913
glyma14g07140.1	moderately similar to (489)AT5G56740 HAG2	HAG2 (HISTONE ACETYLTRANSFERASE OF THE GNAT FAMILY 2)	-5,7591577
glyma04g16180.1	highly similar to (756)AT5G64610 HAM1	HAM1 (HISTONE ACETYLTRANSFERASE OF THE MYST FAMILY 1)	-2,8748133
glyma19g34340.1	highly similar to (695)AT3G54610 HAG1	GCN5 (HISTONE ACETYLTRANSFERASE OF THE GNAT FAMILY 1)	-1,6193736
glyma04g38730.1	highly similar to (523)AT5G09230 SRT2	SRT2	-2,4040203
glyma12g05890.1	highly similar to (521)AT3G20810	transcription factor jumonji (jmiC) domain-containing protein	-4,490048
glyma04g36630.1	highly similar to (946)AT3G48430 REF6	REF6 (RELATIVE OF EARLY FLOWERING 6)	-4,631757
glyma12g07080.1	moderately similar to (372)AT5G59830	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G13660.2)	-1,4760776
glyma17g35180.1	moderately similar to (456)AT3G49810	U-box domain-containing protein	-1,0963334
glyma02g00270.1	highly similar to (528)AT1G80270	DNA-binding protein, putative	-2,5838766
glyma11g12180.1	weakly similar to (190)AT1G17880	nascent polypeptide-associated complex (NAC) domain-containing protein /	-2,3190913
glyma13g41150.1	moderately similar to (376)AT1G13790	XH/XS domain-containing protein / XS zinc finger domain-containing protein	-2,596702
glyma08g40850.1	moderately similar to (315)AT1G14410 ATWHY1	ATWHY1/PTAC1 (A. THALIANA WHIRLY 1)	-1,5637203
glyma17g13770.1	highly similar to (718)AT1G56110 NOP56	NOP56 (ARABIDOPSIS HOMOLOG OF NUCLEOLAR PROTEIN NOP56)	-1,3887951
glyma02g41020.1	highly similar to (693)AT1G66170 MMD1	MMD1 (MALE MEIOCYTE DEATH 1)	-1,8561513
glyma15g07950.1	highly similar to (569)AT4G21705	pentatricopeptide (PPR) repeat-containing protein	-1,2559736
glyma01g38660.1	highly similar to (624)AT3G50780	protein binding	-2,6668713
glyma07g07970.1	very weakly similar to (99.8)AT3G61260	DNA-binding family protein / remorin family protein	-1,0067317
glyma20g35480.1	weakly similar to (161)AT5G62260	DNA binding	-3,920037
glyma05g03150.1	highly similar to (721)AT1G56110 NOP56	NOP56 (ARABIDOPSIS HOMOLOG OF NUCLEOLAR PROTEIN NOP56)	-1,5076336
glyma04g41500.1	nearly identical (1373)AT4G27910	PHD finger protein-related / SET domain-containing protein (TX4)	-3,2969952
glyma04g00380.1	moderately similar to (268)AT5G38110 SGA1	ASF1B/SGA01/SGA1 (ANTI- SILENCING FUNCTION 1B)	-2,4456766
glyma08g40850.1	moderately similar to (315)AT1G14410 ATWHY1	ATWHY1/PTAC1 (A. THALIANA WHIRLY 1)	-1,5637203
glyma12g07080.1	moderately similar to (372)AT5G59830	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G13660.2)	-1,4760776
glyma19g41360.1	moderately similar to (422)AT3G62200	similar to EDA32 (embryo sac development arrest 32) [Arabidopsis thaliana] (-1,6305814
glyma10g05550.1	moderately similar to (388)AT2G20280	zinc finger (CCCH-type) family protein	-1,8561248
glyma10g41050.1	highly similar to (619)AT1G56290	Cwfj-like family protein	-2,9423559
glyma10g36730.1	highly similar to (527)AT4G24630	receptor/ zinc ion binding	-1,6052382
glyma20g35670.1	highly similar to (630)AT1G33420	PHD finger family protein	-1,571831
glyma19g42780.1	highly similar to (510)AT3G04970	zinc finger (DHHC type) family protein	-2,573182
glyma07g07970.1	very weakly similar to (99.8)AT3G61260	DNA-binding family protein / remorin family protein	-1,0067317
glyma07g38940.1	moderately similar to (414)AT1G11650 ATRBP45B	ATRBP45B	-1,9401675
glyma03g02930.1	moderately similar to (473)AT3G60800	zinc finger (DHHC type) family protein	-2,6074674
glyma08g10530.1	moderately similar to (370)AT1G30320	remorin family protein	-2,5876014
glyma06g14030.1	moderately similar to (403)AT1G32790 CID11	CID11	-1,497919
glyma04g05070.1	moderately similar to (201)AT1G58470 RBP1	ATRBP1 (ARABIDOPSIS THALIANA RNA-BINDING PROTEIN 1)	-2,3294215
glyma07g33140.1	moderately similar to (356)AT3G19090	RNA-binding protein, putative	-1,584668
glyma17g08630.1	weakly similar to (145)AT5G61030 GR-RBP3	GR-RBP3 (glycine-rich RNA-binding protein 3)	-2,8149548
glyma03g32960.1	weakly similar to (141)AT2G37510	RNA-binding protein, putative	-2,234894
glyma02g05500.1	moderately similar to (286)AT2G17975	zinc finger (Ran-binding) family protein	-4,5201635
glyma08g18810.1	moderately similar to (383)AT3G14450 CID9	CID9 (CTC-Interacting Domain 9)	-1,2091112
glyma08g15370.1	moderately similar to (478)AT5G09880	RNA recognition motif (RRM)-containing protein	-1,7890087
glyma06g01670.1	moderately similar to (265)AT1G76460	RNA recognition motif (RRM)-containing protein	-1,3813871
glyma01g02100.1	moderately similar to (303)AT5G46840	RNA recognition motif (RRM)-containing protein	-1,2272161
glyma09g00310.1	moderately similar to (434)AT2G18510 EMB2444	EMB2444 (EMBRYO DEFECTIVE 2444)	-1,3444761
glyma03g42150.1	highly similar to (528)AT4G00830	RNA recognition motif (RRM)-containing protein	-2,0597208

Development			
ID		Transcript name	Log(FC)
glyma03g00380.1	moderately similar to (286)AT2G38600	acid phosphatase class B family protein	-1,4456639
glyma07g01740.1	moderately similar to (217)AT4G25150	acid phosphatase, putative	-5,544211
glyma10g39150.1	moderately similar to (219)AT3G22640	cupin family protein	-5,5594864
glyma06g14290.1	moderately similar to (290)AT4G35580	no apical meristem (NAM) family protein	-1,4414425
glyma08g08070.1	very weakly similar to (92.4)AT3G43660	nodulin, putative	-1,229933
glyma07g32010.1	highly similar to (750)AT1G28380 NSL1	NSL1 (NECROTIC SPOTTED LESIONS 1)	-1,0965304
glyma04g41410.1	highly similar to (551)AT3G61790	seven in absentia (SINA) family protein	-1,4176853
glyma12g07660.1	moderately similar to (410)AT3G45600 TET3	TET3 (TETRASPANIN3)	-2,637978
glyma20g23740.1	highly similar to (728)AT3G59040	pentatricopeptide (PPR) repeat-containing protein	-2,364325
glyma18g53680.1	moderately similar to (241)AT4G20370 TSF	TSF (TWIN SISTER OF FT)	-1,8606573
glyma07g29560.1	weakly similar to (182)AT1G49510 EMB1273	EMB1273 (EMBRYO DEFECTIVE 1273)	-7,3395476
glyma14g08500.1	highly similar to (516)AT5G42660	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT2G02910.1)	-2,762145
glyma06g02080.1	highly similar to (879)AT5G42310	pentatricopeptide (PPR) repeat-containing protein	-1,1108351
glyma07g37820.1	moderately similar to (481)AT3G49660	transducin family protein / WD-40 repeat family protein	-1,3940034
glyma19g37890.1	moderately similar to (272)AT5G03840 TFL-1	TFL1 (TERMINAL FLOWER 1)	-1,4704227
glyma02g44260.1	nearly identical (1380)AT2G27040 OCP11	AGO4 (ARGONAUTE 4)	-2,1297264
glyma17g07010.1	moderately similar to (375)AT5G57390 AIL5	AIL5 (AINTEGUMENTA-LIKE 5)	-2,4005668
glyma03g27120.1	moderately similar to (338)AT4G30420	nodulin MtN21 family protein	-1,1057489
glyma05g26050.1	highly similar to (531)AT5G46220	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G28240.1)	-1,2489284
glyma06g11090.1	weakly similar to (115)AT1G78020	senescence-associated protein-related	-2,9183946
glyma15g01160.1	weakly similar to (160)AT5G62200	embryo-specific protein-related	-1,2609601
glyma05g08110.1	highly similar to (612)AT4G32551 RON2	LUG (LEUNIG)	-2,238513
glyma06g41500.1	highly similar to (507)AT5G48150 PAT1	PAT1 (PHYTOCHROME A SIGNAL TRANSDUCTION 1)	-2,8676221
Hormones			
ID		Transcript name	Log(FC)
glyma14g40470.1	moderately similar to (291)AT5G42560	abscisic acid-responsive HVA22 family protein	-1,4539497
glyma08g41060.1	weakly similar to (198)AT1G69700 ATHVA22C	ATHVA22C (Arabidopsis thaliana HVA22 homologue C)	-3,6425006
glyma08g38710.1	moderately similar to (274)AT2G22475 GEM	GEM	-7,7509108
glyma20g23540.1	moderately similar to (223)AT5G08350	GRAM domain-containing protein / ABA-responsive protein-related	-1,9822065
glyma04g40440.1	moderately similar to (461)AT5G54140 ILL3	ILL3 (IAA-amino acid hydrolase ILR1-like 3)	-6,0755663
glyma17g06460.1	highly similar to (920)AT5G57090 WAV6	EIR1 (ETHYLENE INSENSITIVE ROOT 1)	-1,5032655
glyma06g00950.1	weakly similar to (176)AT4G34760	auxin-responsive family protein	-3,6840384
glyma09g06900.1	moderately similar to (420)AT1G51630	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT3G21190.1)	-2,8646243
glyma03g40870.1	moderately similar to (380)AT1G60690	aldo/keto reductase family protein	-1,0110835
glyma06g02810.1	weakly similar to (143)AT1G75590	auxin-responsive family protein	-1,9171997
glyma02g00780.1	highly similar to (600)AT1G04420	aldo/keto reductase family protein	-1,2424315
glyma19g44310.1	highly similar to (807)AT2G46370 JAR	JAR1 (JASMONATE RESISTANT 1)	-1,3590037
glyma12g01170.1	highly similar to (606)AT5G13710 CPH	SMT1 (STEROL METHYLTRANSFERASE 1)	-1,3356705
glyma06g02330.1	highly similar to (683)AT1G20330 CVP1	SMT2 (STEROL METHYLTRANSFERASE 2)	-1,204918
glyma17g06220.1	highly similar to (627)AT5G56970 ATCKX3	CKX3 (CYTOKININ OXIDASE 3)	-2,243401
glyma15g11930.1	moderately similar to (445)AT1G05010 ACO4	EFE (ETHYLENE FORMING ENZYME)	-1,1798514
glyma16g32220.1	moderately similar to (412)AT1G06650	2-oxoglutarate-dependent dioxygenase, putative	-2,2908134
glyma04g07520.1	moderately similar to (318)AT1G15550 ATGA3OX1	GA4 (GA REQUIRING 4)	-1,0182178
glyma09g26840.1	moderately similar to (392)AT1G06620	2-oxoglutarate-dependent dioxygenase, putative	-2,7210708
glyma07g13100.1	moderately similar to (404)AT5G59540	oxidoreductase, 2OG-Fe(II) oxygenase family protein	-1,6267105
glyma10g00980.1	weakly similar to (157)AT3G23240 ATERF1	ATERF1/ERF1 (ETHYLENE RESPONSE FACTOR 1)	-2,507195
glyma10g33070.1	weakly similar to (114)AT3G23230	ethylene-responsive factor, putative	-2,9635859
glyma10g33810.1	weakly similar to (116)AT5G07580	DNA binding / transcription factor	-2,6069062
glyma13g18340.1	weakly similar to (143)AT3G23240 ATERF1	ATERF1/ERF1 (ETHYLENE RESPONSE FACTOR 1)	-5,8286867
glyma02g01330.1	moderately similar to (383)AT1G02400 ATGA2OX6	ATGA2OX6/DTA1 (GIBBERELLIN 2-OXIDASE 6)	-3,1941478
glyma17g14030.1	highly similar to (613)AT1G63100	scarecrow transcription factor family protein	-2,1059308

Signal regulation			
ID		Transcript name	Log(FC)
glyma17g00990.1	moderately similar to (323)AT4G16360	5'-AMP-activated protein kinase beta-2 subunit, putative	-1,9293052
glyma08g02450.1	highly similar to (732)AT4G23740	leucine-rich repeat transmembrane protein kinase, putative	-1,4040354
glyma17g13210.1	moderately similar to (348)AT1G68780	leucine-rich repeat family protein	-2,8064134
glyma18g48170.1	highly similar to (741)AT5G48380	leucine-rich repeat family protein / protein kinase family protein	-4,290465
glyma05g26520.1	nearly identical (1425)AT4G20140 GSO1	leucine-rich repeat transmembrane protein kinase, putative	-1,7364073
glyma17g10470.1	highly similar to (904)AT5G62710	leucine-rich repeat family protein / protein kinase family protein	-1,5631329
glyma11g07830.1	moderately similar to (384)AT5G66330	leucine-rich repeat family protein	-1,7117091
glyma08g10030.1	moderately similar to (486)AT1G70740	protein kinase family protein	-1,4517691
glyma17g06980.1	moderately similar to (474)AT2G18890	protein kinase family protein	-3,2245977
glyma17g12680.1	moderately similar to (462)AT5G20050	protein kinase family protein	-2,4461217
glyma12g32260.1	moderately similar to (201)AT3G22060	receptor protein kinase-related	-1,5376686
glyma17g10470.1	highly similar to (904)AT5G62710	leucine-rich repeat family protein / protein kinase family protein	-1,5631329
glyma19g34280.1	weakly similar to (180)AT2G27030 CAM5	CAM5 (CALMODULIN 5)	-2,3938162
glyma09g01160.1	highly similar to (808)AT3G20290	calcium-binding EF hand family protein	-1,3230942
glyma09g40290.1	highly similar to (670)AT3G61050 NTMC2TYPE4	NTMC2T4/NTMC2TYPE4	-1,2465925
glyma18g11030.1	highly similar to (802)AT3G20410 CPK9	CPK9 (CALMODULIN-DOMAIN PROTEIN KINASE 9)	-6,9525332
glyma10g28990.1	moderately similar to (248)AT4G25800	calmodulin-binding protein	-1,5262886
glyma02g00450.1	moderately similar to (203)AT2G27030 CAM5	CAM5 (CALMODULIN 5)	-1,4534746
glyma07g01760.1	moderately similar to (250)AT3G16490 IQD26	IQD26 (IQ-domain 26)	-2,8046317
glyma13g35950.1	weakly similar to (143)AT4G27280	calcium-binding EF hand family protein	-2,0840597
glyma14g11050.1	moderately similar to (293)AT2G26180 IQD6	IQD6 (IQ-domain 6)	-1,0530719
glyma09g32530.1	moderately similar to (387)AT4G28950 ARAC7	ARAC7/ATRACT7/ATROP9/RAC7/ROP9 (RHO-RELATED PROTEIN FROM PLANT	-1,3270277
glyma02g48200.1	nearly identical (2298)AT4G38200	guanine nucleotide exchange family protein	-1,5111114
glyma12g10130.1	highly similar to (655)AT1G08410	GTP-binding family protein	-1,5348614
glyma08g21940.1	moderately similar to (393)AT1G52280 AtRABG3d	AtRABG3d (Arabidopsis Rab GTPase homolog G3d)	-1,0579586
glyma10g03190.1	highly similar to (920)AT1G03160 FZL	FZL (FZO-LIKE)	-1,1302663
glyma18g53870.1	moderately similar to (332)AT2G31680 AtRABA5d	AtRABA5d (Arabidopsis Rab GTPase homolog A5d)	-1,2244697
glyma12g06150.1	highly similar to (778)AT3G12080 EMB2738	EMB2738 (EMBRYO DEFECTIVE 2738)	-2,6137981
glyma05g31020.1	moderately similar to (365)AT5G65270 AtRABA4a	AtRABA4a (Arabidopsis Rab GTPase homolog A4a)	-3,7317395
glyma13g07090.1	highly similar to (569)AT3G02460	plant adhesion molecule, putative	-1,434958
glyma18g12720.1	highly similar to (792)AT2G42880 ATMPK20	ATMPK20 (Arabidopsis thaliana MAP kinase 20)	-3,4036422
glyma13g28120.1	highly similar to (892)AT5G19010 MPK16	MPK16 (mitogen-activated protein kinase 16)	-1,7406952
glyma07g09090.1	moderately similar to (348)AT3G18030 HAL3A	ATHAL3A (Arabidopsis thaliana Hal3-like protein A)	-5,0950694
glyma17g08640.1	moderately similar to (248)AT2G29900	presenilin family protein	-1,8678044
glyma08g08940.1	moderately similar to (368)AT2G34460	flavin reductase-related	-4,0464005
glyma01g03100.1	highly similar to (705)AT1G67900	phototropic-responsive NPH3 family protein	-2,0335224
Redox regulation			
ID		Transcript name	Log(FC)
glyma08g09210.1	weakly similar to (153)AT2G35010 ATO1	ATO1 (ARABIDOPSIS THIOREDOXIN O1)	-1,7535536
glyma14g39940.1	moderately similar to (363)AT1G76080 ATCDSP32	ATCDSP32/CDSP32 (CHLOROPLASTIC DROUGHT-INDUCED STRESS PROTEIN O	-2,3883278
glyma01g04810.1	weakly similar to (110)AT5G39950 ATH2	ATRX2 (ARABIDOPSIS THIOREDOXIN H2)	-1,408361
glyma11g07750.1	weakly similar to (175)AT3G51030 ATTRX H1	ATRX1 (Arabidopsis thaliana thioredoxin H-type 1)	-1,584116
glyma15g06870.1	weakly similar to (154)AT1G11530 ATCXS1	ATCXS1 (C-TERMINAL CYSTEINE RESIDUE IS CHANGED TO A SERINE 1)	-1,3292778
glyma11g12960.1	moderately similar to (203)AT2G16060 ARATH GLB1	AHB1 (ARABIDOPSIS HEMOGLOBIN 1)	-1,2634456
glyma19g02500.1	weakly similar to (185)AT3G28850	glutaredoxin family protein	-2,9452026
glyma19g38820.1	weakly similar to (189)AT5G06470	glutaredoxin family protein	-1,9212524
glyma06g02040.1	highly similar to (919)AT4G35090 CAT2	CAT2 (CATALASE 2)	-4,6342835
Metal handling			
ID		Transcript name	Log(FC)
glyma04g00500.1	moderately similar to (236)AT4G38580 ATPF6	ATPF6 (FARNESYLATED PROTEIN 6)	-1,0516322
glyma09g01610.1	weakly similar to (130)AT5G50740	similar to ATPF3 (Arabidopsis thaliana farnesylated protein 3), metal ion bindi	-5,1288037
glyma08g01410.1	moderately similar to (203)AT5G50740	similar to ATPF3 (Arabidopsis thaliana farnesylated protein 3), metal ion bindi	-1,5101576

Transport			
ID		Transcript name	Log(FC)
glyma08g15250.1	moderately similar to (466)AT4G39390	transporter-related	-2,9372978
glyma05g27410.1	highly similar to (803)AT1G30220 ATINT2	ATINT2 (INOSITOL TRANSPORTER 2)	-2,3920248
glyma10g39500.1	highly similar to (707)AT5G26340 STP13	MSS1 (SUGAR TRANSPORT PROTEIN 13)	-1,1023525
glyma02g08260.1	highly similar to (533)AT1G71880 ATSUC1	SUC1 (SUCROSE-PROTON SYMPORTER 1)	-1,7716566
glyma13g39850.1	highly similar to (815)AT1G08090 NRT2	ATNRT2.1 (Arabidopsis thaliana high affinity nitrate transporter 2.1)	-1,282353
glyma17g09630.1	moderately similar to (436)AT5G33320 PPT	CUE1 (CAB UNDEREXPRESSED 1)	-2,4818532
glyma05g04140.1	moderately similar to (450)AT1G12500	phosphate translocator-related	-1,2520857
glyma10g36580.3	moderately similar to (365)AT4G39460 SAMC1	SAMC1/SAMT1 (S-ADENOSYLMETHIONINE CARRIER 1)	-3,07982
glyma16g26240.1	moderately similar to (458)AT5G14040	mitochondrial phosphate transporter	-1,8655839
glyma17g12450.1	moderately similar to (387)AT4G32400 SHS1	SHS1 (SODIUM HYPERSENSITIVE 1)	-1,7945256
glyma19g44250.1	moderately similar to (428)AT2G46320	mitochondrial substrate carrier family protein	-4,4054728
glyma20g02770.1	moderately similar to (391)AT1G31260 ZIP10	ZIP10 (ZINC TRANSPORTER 10 PRECURSOR)	-1,66338
glyma15g00590.1	highly similar to (717)AT1G15960 ATNRAMP6	NRAMP6 (NRAMP metal ion transporter 6)	-1,3610704
glyma11g27900.1	moderately similar to (390)AT2G30080 ZIP6	ZIP6 (ZINC TRANSPORTER 6 PRECURSOR)	-2,4838815
glyma10g02250.1	highly similar to (723)AT2G47600 MHX1	ATMHX (MAGNESIUM/PROTON EXCHANGER)	-2,6855087
glyma08g40730.1	highly similar to (734)AT1G69850 NTL1	ATNRT1.2 (NITRATE TRANSPORTER 1:2)	-3,075739
glyma17g01000.1	highly similar to (941)AT4G16370 OPT3	ATOPT3 (OLIGOPEPTIDE TRANSPORTER)	-2,3200707
glyma13g23460.1	moderately similar to (425)AT1G02630	equilibrative nucleoside transporter, putative (ENT8)	-2,3361104
glyma13g33940.1	moderately similar to (411)AT1G78560	bile acid:sodium symporter family protein	-2,2030878
glyma10g30020.1	highly similar to (738)AT3G05030 ATNHX2	NHX2 (sodium proton exchanger 2)	-3,5298605
glyma02g13080.1	moderately similar to (419)AT2G03620	magnesium transporter CorA-like family protein (MRS2-5)	-4,4432206
glyma16g05060.1	nearly identical (1217)AT3G02050 ATKUP3	KUP3 (K+ uptake permease 3)	-4,1334844
glyma10g41110.1	nearly identical (1072)AT2G01320	ABC transporter family protein	-3,2601595
glyma19g36820.1	nearly identical (2076)AT2G36910 PGP1	ATPGP1 (ARABIDOPSIS THALIANA P GLYCOPROTEIN1)	-3,4308457
glyma17g12910.1	nearly identical (2234)AT2G26910 PDR4	ATPDR4/PDR4 (PLEIOTROPIC DRUG RESISTANCE 4)	-5,603938
glyma13g43870.1	nearly identical (2126)AT1G15520 PDR12	ATPDR12/PDR12 (PLEIOTROPIC DRUG RESISTANCE 12)	-1,442078
glyma16g33470.1	nearly identical (1152)AT1G17840 WBC11	ABCG11/COF1/DSO/WBC11 (DESPERADO)	-1,7409024
glyma16g08480.1	nearly identical (1147)AT3G28345	ABC transporter family protein	-2,1688256
glyma17g10670.1	nearly identical (1040)AT3G47780 ATATH6	ATATH6 (ABC2 homolog 6)	-2,8814964
glyma17g04590.1	nearly identical (1585)AT1G02520 PGP11	PGP11 (P-GLYCOPROTEIN 11)	-2,045647
glyma20g38380.1	nearly identical (2272)AT3G55320 PGP20	PGP20 (P-GLYCOPROTEIN 20)	-2,2255142
glyma09g03600.1	highly similar to (938)AT3G62270	anion exchange family protein	-1,0180242
glyma19g40720.1	nearly identical (1124)AT2G47160 BOR1	BOR1 (REQUIRES HIGH BORON 1)	-1,3411338
glyma17g09660.1	highly similar to (944)AT2G47160 BOR1	BOR1 (REQUIRES HIGH BORON 1)	-4,3515964
glyma11g15200.1	moderately similar to (335)AT4G01470 GAMMA-TIP	GAMMA-TIP3/TIP1	-1,9086206
glyma02g07680.1	moderately similar to (204)AT5G18290 SIP1	SIP1	-1,3705425
glyma12g08160.1	highly similar to (995)AT2G28260 CNGC15	ATCNGC15 (cyclic nucleotide gated channel 15)	-1,3058002
glyma03g00770.1	highly similar to (536)AT1G33110	MATE efflux family protein	-1,4464958
glyma12g00690.1	moderately similar to (467)AT4G25270	pentatricopeptide (PPR) repeat-containing protein	-1,3167231
glyma01g00750.1	highly similar to (528)AT3G14330	pentatricopeptide (PPR) repeat-containing protein	-1,7592685
glyma07g39890.1	moderately similar to (498)AT5G47730	SEC14 cytosolic factor, putative / polyphosphoinositide-binding protein, puta	-1,9151087
glyma06g04990.1	highly similar to (669)AT1G65550	xanthine/uracil permease family protein	-2,5815327
glyma14g00690.1	nearly identical (1170)AT5G09950	pentatricopeptide (PPR) repeat-containing protein	-2,6231177
glyma09g41040.1	nearly identical (1340)AT5G04930 ALA1	ALA1 (AMINOPHOSPHOLIPID ATPASE1)	-1,5203516
Specialized metabolism			
ID		Transcript name	Log(FC)
glyma08g10960.1	moderately similar to (208)AT1G67360	rubber elongation factor (REF) family protein	-1,1837274
glyma07g38260.1	nearly identical (1184)AT4G15560 DEF	CLA1 (CLOROPLASTOS ALTERADOS 1)	-1,8159534
glyma13g40140.2	nearly identical (1297)AT5G60600 ISPG	GcpE (CHLOROPLAST BIOGENESIS 4)	-1,4215888
glyma11g12870.1	highly similar to (706)AT4G34350 ISPH	CLB6 (CHLOROPLAST BIOGENESIS 6)	-1,2393893
glyma17g18160.1	highly similar to (525)AT2G34630 GPPS	GPPS/GPS1 (GERANYLPYROPHOSPHATE SYNTHASE)	-2,4220164
glyma02g44070.1	highly similar to (834)AT1G76490 HMGR1	HMG1 (3-HYDROXY-3-METHYLGLUTARYL COA REDUCTASE)	-2,846195
glyma10g42630.1	highly similar to (637)AT3G54250	mevalonate diphosphate decarboxylase, putative	-1,9459244
glyma17g18160.1	highly similar to (525)AT2G34630 GPPS	GPPS/GPS1 (GERANYLPYROPHOSPHATE SYNTHASE)	-2,4220164
glyma15g12780.1	highly similar to (583)AT5G47770 FPS1	FPS1 (FARNESYL DIPHOSPHATE SYNTHASE 1)	-1,6213259
glyma14g03410.1	highly similar to (558)AT1G06570 HPD	PDS1 (PHYTOENE DESATURATION 1)	-1,7608833
glyma17g06940.1	moderately similar to (493)AT2G18950 TPT1	HPT1 (HOMOGENITISATE PHYTYLTRANSFERASE 1)	-1,3608401
glyma08g42440.1	moderately similar to (311)AT2G19070	transferase family protein	-3,5327394
glyma14g06710.1	highly similar to (647)AT2G39980	transferase family protein	-1,3634245
glyma16g26650.1	moderately similar to (370)AT1G65450	transferase family protein	-1,97729
glyma08g03440.1	moderately similar to (273)AT4G26220	caffeoyl-CoA 3-O-methyltransferase, putative	-1,2322384
glyma10g40870.2	moderately similar to (459)AT3G19450 ATCAD4	CAD4 (CINNAMYL ALCOHOL DEHYDROGENASE 4)	-6,73989
glyma06g46550.1	highly similar to (514)AT1G08470	strictosidine synthase family protein	-2,0812175
glyma17g03650.1	nearly identical (1033)AT3G66658 ALDH22a1	ALDH22a1 (ALDEHYDE DEHYDROGENASE 22A1)	-1,0654826
glyma13g36500.1	moderately similar to (356)AT5G54010	glycosyltransferase family protein	-1,5826654
glyma02g39750.1	highly similar to (766)AT2G30210 LAC3	LAC3 (laccase 3)	-2,3899949
glyma10g36310.1	highly similar to (627)AT5G48100 LAC15	TT10 (TRANSPARENT TESTA 10)	-2,2442205
glyma14g37040.1	highly similar to (856)AT2G38080 IRX12	IRX12/LAC4 (laccase 4)	-1,1411992
glyma18g06450.1	highly similar to (753)AT2G30210 LAC3	LAC3 (laccase 3)	-1,2291081
glyma07g16060.1	highly similar to (791)AT5G60020 LAC17	LAC17 (laccase 17)	-1,371068

Carbohydrates			
ID		Transcript name	Log(FC)
glyma05g36100.1	moderately similar to (495)AT4G26260 MIOX4	MIOX4 (MYO-INOSITOL OXYGENASE 4)	-2,4841516
glyma10g07520.1	highly similar to (533)AT5G03770	3-deoxy-D-manno-octulosonic acid transferase-related	-3,7509534
glyma09g30070.1	highly similar to (793)AT1G48430	dihydroxyacetone kinase family protein	-1,1321855
glyma02g00780.1	highly similar to (600)AT1G04420	aldo/keto reductase family protein	-1,2424315
glyma09g40990.1	highly similar to (653)AT5G08380 ATAGAL1	ATAGAL1 (ARABIDOPSIS THALIANA ALPHA-GALACTOSIDASE 1)	-1,8010552
glyma17g01900.1	highly similar to (852)AT5G19220 APL1	ADG2 (ADPG PYROPHOSPHORYLASE 2)	-5,10591
glyma13g37140.1	highly similar to (542)AT4G28390 ATAAC3	AAC3 (ADP/ATP CARRIER 3)	-1,6453803
glyma20g22780.1	highly similar to (951)AT1G56560	beta-fructofuranosidase, putative / invertase, putative / saccharase, putative	-1,4311537
glyma01g41990.1	highly similar to (818)AT1G12240 ATBETAFRUCT4	ATBETAFRUCT4/VAC-INV (VACUOLAR INVERTASE)	-1,2458812
glyma18g46630.1	highly similar to (802)AT2G45880 BMY4	BAM7/BMY4 (BETA-AMYLASE 7)	-2,888672
Cell wall			
ID		Transcript name	Log(FC)
glyma07g05200.1	highly similar to (674)AT1G17270	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G50420.1)	-1,0044544
glyma01g43540.1	highly similar to (602)AT4G10960 UGE5	UGE5 (UDP-D-GLUCOSE/UDP-D-GALACTOSE 4-EPIMERASE 5)	-1,6018212
glyma19g03500.1	highly similar to (910)AT3G29360	UDP-glucose 6-dehydrogenase, putative	-1,5463464
glyma11g15020.1	highly similar to (621)AT3G46440 UXS5	UXS5 (UDP-Xyl synthase 5)	-2,4296267
glyma14g10230.1	highly similar to (919)AT4G07960 CSLC12	ATCSLC12 (Cellulose synthase-like C12)	-1,3506855
glyma05g32100.1	nearly identical (1863)AT5G64740 CESA6	CESA6 (CELLULOSE SYNTHASE 6)	-1,0207179
glyma06g07320.1	nearly identical (1950)AT4G32410 CESA1	CESA1 (CELLULOSE SYNTHASE 1)	-9,618326
glyma12g07440.1	weakly similar to (142)AT5G03170 FLA11	FLA11 (fasciclin-like arabinogalactan protein 11)	-1,6758168
glyma19g04240.1	highly similar to (655)AT5G15650 RGP2	RGP2 (REVERSIBLY GLYCOSYLATED POLYPEPTIDE 2)	-1,8852113
glyma12g12770.1	highly similar to (610)AT2G20680	glycosyl hydrolase family 5 protein / cellulase family protein	-1,082746
glyma04g36520.1	highly similar to (630)AT5G63180	pectate lyase family protein	-3,6199124
glyma19g00230.1	highly similar to (550)AT5G14650	polygalacturonase, putative / pectinase, putative	-5,689046
glyma03g38140.1	highly similar to (597)AT1G02460	glycoside hydrolase family 28 protein / polygalacturonase (pectinase) family	-1,9895445
glyma04g02380.1	moderately similar to (392)AT1G20190 EXP11	ATEXA11 (ARABIDOPSIS THALIANA EXPANSIN A11)	-1,6696278
glyma15g04240.1	moderately similar to (422)AT2G39700 ATEXP4	ATEXA4 (ARABIDOPSIS THALIANA EXPANSIN A4)	-4,505645
glyma02g46890.1	moderately similar to (397)AT5G55590 QRT1	QRT1 (QUARTET 1)	-1,7865
glyma08g03700.1	highly similar to (583)AT5G19730	pectinesterase family protein	-3,746991
glyma11g03560.1	highly similar to (531)AT5G47500	pectinesterase family protein	-1,0269092
glyma05g34810.1	moderately similar to (476)AT1G23200	pectinesterase family protein	-5,5132875
Other metabolisms			
ID		Transcript name	Log(FC)
glyma10g35490.1	highly similar to (781)AT5G17530	phosphoglucosamine mutase family protein	-1,1439574
glyma01g03040.1	highly similar to (801)AT2G22480	phosphofructokinase family protein	-1,2556441
glyma04g09180.1	highly similar to (823)AT4G26270	phosphofructokinase family protein	-1,4281634
glyma07g39380.1	highly similar to (950)AT1G12000	pyrophosphate--fructose-6-phosphate 1-phosphotransferase beta subunit, p	-1,8053714
glyma02g38730.1	highly similar to (605)AT2G36460	fructose-bisphosphate aldolase, putative	-2,0159698
glyma18g01330.1	highly similar to (546)AT3G04120 GAPC-1	GAPC (GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C SUBUNIT)	-1,5291816
glyma13g17790.1	moderately similar to (278)AT5G28050	cytidine/deoxycytidylate deaminase family protein	-1,535228
glyma11g03400.1	moderately similar to (298)AT4G22570 APT3	APT3 (ADENINE PHOSPHORIBOSYL TRANSFERASE 3)	-1,5852232
glyma09g36680.3	moderately similar to (290)AT4G25280	adenylate kinase family protein	-1,0617548
glyma04g07790.1	moderately similar to (408)AT3G57560	aspartate/glutamate/uridylate kinase family protein	-1,5070937
glyma06g07870.1	moderately similar to (420)AT3G57560	aspartate/glutamate/uridylate kinase family protein	-1,6768533
glyma19g25930.1	highly similar to (574)AT3G27060 TSO2	TSO2 (TSO MEANING 'UGLY' IN CHINESE)	-1,1528589
glyma18g04940.1	highly similar to (872)AT2G02010 GAD4	GAD4 (GLUTAMATE DECARBOXYLASE 4)	-2,378282
glyma02g41850.1	nearly identical (1035)AT3G55610 P5CS2	P5CS2 (DELTA 1-PYRROLINE-5-CARBOXYLATE SYNTHASE 2)	-1,6309904
glyma18g02060.1	nearly identical (1041)AT5G10240 ASN3	ASN3 (ASPARAGINE SYNTHETASE 3)	-2,2750716
glyma08g05250.1	highly similar to (570)AT1G14810	semialdehyde dehydrogenase family protein	-2,1457043
glyma20g26530.1	highly similar to (807)AT4G34200 EDA9	EDA9 (embryo sac development arrest 9)	-4,5180006
glyma16g03080.1	moderately similar to (431)AT5G56760 SAT5	AtSerat1	-3,0231245
glyma19g32730.1	nearly identical (1422)AT3G57880	C2 domain-containing protein	-3,7612062
glyma17g14040.1	highly similar to (586)AT4G08900	arginase	-2,1074128
glyma05g36450.1	highly similar to (885)AT5G35360 CAC2	CAC2 (acetyl co-enzyme A carboxylase biotin carboxylase subunit)	-3,2387748
glyma15g11590.1	highly similar to (731)AT1G11860	aminomethyltransferase, putative	-1,0061157
glyma02g44160.1	highly similar to (553)AT5G42250	alcohol dehydrogenase, putative	-3,5341473
glyma02g44170.1	highly similar to (560)AT5G42250	alcohol dehydrogenase, putative	-1,9474211
glyma05g36450.1	highly similar to (885)AT5G35360 CAC2	CAC2 (acetyl co-enzyme A carboxylase biotin carboxylase subunit)	-3,2387748
glyma08g45990.1	moderately similar to (491)AT2G05990 ENR1	MOD1 (MOSAIC DEATH 1)	-4,11441
glyma17g03500.1	highly similar to (898)AT3G16910 AAE7	AAE7/ACN1 (ACYL-ACTIVATING ENZYME 7)	-1,1549981
glyma13g11700.2	nearly identical (1038)AT2G04350	long-chain-fatty-acid--CoA ligase family protein / long-chain acyl-CoA synthetase	-4,313326
glyma19g42860.1	weakly similar to (102)AT1G54630 ACP3	ACP3 (ACYL CARRIER PROTEIN 3)	-1,2802365
glyma09g01630.1	weakly similar to (116)AT5G47630 MTACP3	MTACP3 (MITOCHONDRIAL ACYL CARRIER PROTEIN 3)	-3,0930746
glyma04g41220.1	moderately similar to (391)AT5G53470 ACBP	ACBP1 (ACYL-COA BINDING PROTEIN)	-1,6691542
glyma15g11810.1	highly similar to (706)AT4G04320	malonyl-CoA decarboxylase family protein	-2,2081835
glyma14g08400.2	highly similar to (507)AT1G75020 LPAT4	LPAT4	-2,33308
glyma03g29600.1	highly similar to (560)AT3G57650 LPAT2	LPAT2 (Lysophosphatidyl acyltransferase 2)	-1,4587399
glyma03g04960.1	very weakly similar to (80.5)AT5G59310 LTP4	LTP4 (LIPID TRANSFER PROTEIN 4)	-4,6988525
glyma20g31200.1	moderately similar to (300)AT1G69640	acid phosphatase, putative	-1,605947
glyma07g39500.1	moderately similar to (399)AT4G22330 ATCES1	ATCES1	-1,6565313
glyma01g01130.1	highly similar to (770)AT2G19880	ceramide glucosyltransferase, putative	-2,14757
glyma02g43480.1	moderately similar to (384)AT5G14180 MPL1	MPL1 (MYZUS PERSICAE-INDUCED LIPASE 1)	-2,0239813
glyma15g08780.1	highly similar to (519)AT4G18550	lipase class 3 family protein	-4,6119943
glyma07g32610.1	moderately similar to (303)AT4G18550	lipase class 3 family protein	-1,771238
glyma05g04940.1	nearly identical (1120)AT4G16760 ATACX1	ACX1 (ACYL-COA OXIDASE 1)	-1,5746183
glyma01g43300.1	weakly similar to (188)AT1G64185	lactoylglutathione lyase family protein / glyoxalase I family protein	-1,6449075
glyma08g09260.1	moderately similar to (271)AT2G44050 COS1	COS1 (COI1 SUPPRESSOR1)	-1,0262488
glyma17g15190.1	moderately similar to (493)AT5G47435	formyltetrahydrofolate deformylase, putative	-1,4887681
glyma15g42290.1	highly similar to (616)AT2G31955 CNX2	CNX2 (COFACTOR OF NITRATE REDUCTASE AND XANTHINE DEHYDROGENAS	-1,913023

ID	Enzymes	Transcript name	Log(FC)
glyma11g12800.1	highly similar to (616)AT5G66680 DGL1	DGL1 (defective glycosylation 1)	-5,839565
glyma08g44740.1	moderately similar to (438)AT4G01070 GT72B1	GT72B1	-2,0312016
glyma05g28340.1	moderately similar to (390)AT4G14090	UDP-glucuronosyl/UDP-glucosyl transferase family protein	-2,121883
glyma13g36500.1	moderately similar to (356)AT5G54010	glycosyltransferase family protein	-1,5826654
glyma18g48230.1	moderately similar to (370)AT1G05680	UDP-glucuronosyl/UDP-glucosyl transferase family protein	-1,6664909
glyma02g40480.1	highly similar to (847)AT1G77130 PGSIP2	PGSIP2 (PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN 2)	-2,97616
glyma03g34410.1	highly similar to (571)AT2G36760 UGT73C2	UGT73C2 (UDP-GLUCOSYL TRANSFERASE 73C2)	-2,9104733
glyma12g36870.1	highly similar to (604)AT5G44640	glycosyl hydrolase family 1 protein	-1,419362
glyma12g12770.1	highly similar to (610)AT2G20680	glycosyl hydrolase family 5 protein / cellulase family protein	-1,082746
glyma08g04840.1	highly similar to (815)AT1G70710 CEL1	AtGH9B1 (ARABIDOPSIS THALIANA GLYCOSYL HYDROLASE 9B1)	-3,5229285
glyma07g18410.1	highly similar to (583)AT1G02850	glycosyl hydrolase family 1 protein	-1,3332117
glyma08g20650.1	nearly identical (1422)AT3G13750 BGAL1	BGAL1 (BETA GALACTOSIDASE 1)	-3,5836117
glyma13g42560.1	highly similar to (881)AT1G72990 BGAL17	BGAL17 (BETA-GALACTOSIDASE 17)	-2,0290322
glyma13g29000.1	moderately similar to (438)AT1G30080	glycosyl hydrolase family 17 protein	-5,347852
glyma05g31860.1	highly similar to (511)AT5G64790	glycosyl hydrolase family 17 protein	-2,4687371
glyma05g28870.1	highly similar to (619)AT3G13560	glycosyl hydrolase family 17 protein	-1,4716307
glyma16g08040.1	moderately similar to (374)AT1G23740	oxidoreductase, zinc-binding dehydrogenase family protein	-2,4136822
glyma01g38980.1	highly similar to (790)AT5G66920 SKS17	SKS17 (SKUS Similar 17)	-1,9666247
glyma08g18500.1	highly similar to (505)AT1G56000	amine oxidase-related	-1,1282469
glyma18g03950.1	moderately similar to (359)AT5G06060	tropinone reductase, putative / tropine dehydrogenase, putative	-1,9627869
glyma05g30870.1	highly similar to (688)AT1G12990	glycosyl transferase family 17 protein	-2,9903662
glyma05g03580.1	highly similar to (750)AT1G12570	glucose-methanol-choline (GMC) oxidoreductase family protein	-5,1288576
glyma11g06320.1	moderately similar to (233)AT2G23610	esterase, putative	-3,2196522
glyma02g40760.1	moderately similar to (248)AT1G78380 GST8	ATGSTU19 (GLUTATHIONE TRANSFERASE 8)	-1,1488801
glyma13g35230.1	highly similar to (618)AT3G14690 CYP72A15	CYP72A15 (cytochrome P450, family 72, subfamily A, polypeptide 15)	-1,9912542
glyma13g33700.1	highly similar to (546)AT3G14610 CYP72A7	CYP72A7 (cytochrome P450, family 72, subfamily A, polypeptide 7)	-1,1654354
glyma07g09170.1	highly similar to (582)AT2G45510 CYP704A2	CYP704A2 (cytochrome P450, family 704, subfamily A, polypeptide 2)	-3,5775483
glyma13g33690.1	highly similar to (570)AT3G14610 CYP72A7	CYP72A7 (cytochrome P450, family 72, subfamily A, polypeptide 7)	-1,3881335
glyma15g10180.1	highly similar to (677)AT2G34500 CYP710A1	CYP710A1 (cytochrome P450, family 710, subfamily A, polypeptide 1)	-1,4504973
glyma15g03250.1	moderately similar to (451)AT2G43480	peroxidase, putative	-1,7315335
glyma09g02670.1	moderately similar to (296)AT3G49120 ATPERX34	ATPCB/ATPERX34/PERX34/PRXCB (PEROXIDASE 34)	-3,4683802
glyma14g40150.1	moderately similar to (471)AT5G42180	peroxidase 64 (PER64) (P64) (PRXR4)	-4,0969157
glyma18g06210.1	moderately similar to (404)AT5G58390	peroxidase, putative	-6,332616
glyma07g38500.1	highly similar to (956)AT5G18980	binding	-2,0795267
glyma08g20800.1	moderately similar to (312)AT1G17710	similar to phosphoric monoester hydrolase [Arabidopsis thaliana] (TAIR:AT1G17710)	-1,529529
glyma03g00380.1	moderately similar to (286)AT2G38600	acid phosphatase class B family protein	-1,4456639
glyma15g10920.1	highly similar to (951)AT5G18980	binding	-1,5845124
glyma13g40050.1	highly similar to (941)AT1G07990	SIT4 phosphatase-associated family protein	-2,4970622
glyma16g01860.1	moderately similar to (496)AT5G63140 ATPAP29	ATPAP29/PAP29 (purple acid phosphatase 29)	-2,063049
glyma12g08100.1	highly similar to (966)AT1G07990	SIT4 phosphatase-associated family protein	-1,5214047
glyma08g45380.1	highly similar to (997)AT3G60190 ADL4	ADL4/ADLP2/DRP1E/EDR3 (DYNAMIN-LIKE PROTEIN 4)	-1,1453569
glyma03g00880.1	moderately similar to (287)AT1G10310	short-chain dehydrogenase/reductase (SDR) family protein	-1,1113954
glyma08g01390.1	moderately similar to (332)AT5G50700	short-chain dehydrogenase/reductase (SDR) family protein	-5,423911
glyma20g30080.1	moderately similar to (451)AT4G23420	short-chain dehydrogenase/reductase (SDR) family protein	-1,3746841
glyma16g30060.1	moderately similar to (412)AT4G11410	short-chain dehydrogenase/reductase (SDR) family protein	-5,2874055
glyma03g38150.1	moderately similar to (278)AT2G47140	short-chain dehydrogenase/reductase (SDR) family protein	-1,5887158
glyma05g38260.1	moderately similar to (292)AT5G50700	short-chain dehydrogenase/reductase (SDR) family protein	-1,9808315
glyma18g51360.1	moderately similar to (252)AT3G42960 ASD	ATA1 (ARABIDOPSIS TAPETUM 1)	-2,8824563
glyma10g37760.1	moderately similar to (408)AT4G23420	short-chain dehydrogenase/reductase (SDR) family protein	-3,6839633
glyma07g06300.1	moderately similar to (342)AT4G01050	hydroxyproline-rich glycoprotein family protein	-1,0075707
glyma01g44720.1	weakly similar to (194)AT3G22560	GCN5-related N-acetyltransferase (GNAT) family protein	-1,874612
glyma14g00630.1	moderately similar to (266)AT5G11340	GCN5-related N-acetyltransferase (GNAT) family protein	-1,1878512
glyma11g06360.1	highly similar to (512)AT2G23540	GDLSL-motif lipase/hydrolase family protein	-1,9503233
glyma15g08590.1	moderately similar to (412)AT5G45960	GDLSL-motif lipase/hydrolase family protein	-1,8918464

I-2 Gènes sous régulés chez les racines

Tableau III-2 : Détail des gènes sous-exprimés chez les racines de soja en réponse à PEP-13 classés en fonction de leur catégorie de *bins*

Une couleur est attribuée au Log2 du FC en fonction de sa valeur, plus celle-ci est faible plus elle apparaîtra en bleu foncé.

Stress			
ID		Transcript name	Log(FC)
glyma15g08450.1	moderately similar to (397)AT2G20560	DNAJ heat shock family protein	-1,3616817
glyma06g12360.1	moderately similar to (388)AT1G21610	wound-responsive family protein	-1,4459374
glyma01g04070.1	weakly similar to (179)AT1G24020	MLP423 (MLP-LIKE PROTEIN 423)	-2,0329213
Cell			
ID		Transcript name	Log(FC)
glyma16g26790.1	weakly similar to (185)AT1G23220	dynein light chain type 1 family protein	-1,6706796
glyma10g29050.1	highly similar to (866)AT5G27000	ATK4 (ARABIDOPSIS THALIANA KINESIN 4)	-1,0299951
glyma05g27500.1	moderately similar to (201)AT1G30280	heat shock protein binding	-1,2541041
DNA			
ID		Transcript name	Log(FC)
glyma19g31030.1	nearly identical (1025)AT5G04130	DNA topoisomerase, ATP-hydrolyzing, putative / DNA topoisomerase II, putative	-1,0433012
glyma02g00330.1	nearly identical (1313)AT5G35970	DNA-binding protein, putative	-1,3569105
Proteins			
ID		Transcript name	Log(FC)
glyma04g41130.1	moderately similar to (295)AT1G09150	pseudouridine synthase and archaeosine transglycosylase (PUA) domain-containing protein	-1,3939755
glyma07g03000.1	weakly similar to (194)AT1G80620	ribosomal protein S15 family protein	-1,2019569
glyma10g02050.1	moderately similar to (208)AT3G16780	60S ribosomal protein L19 (RPL19B)	-1,5544467
glyma03g33430.1	weakly similar to (119)AT3G53470	similar to unnamed protein product [Vitis vinifera] (GB:CAO18045.1)	-1,3631108
glyma13g03920.1	moderately similar to (479)AT5G53530	vacuolar protein sorting-associated protein 26, putative / VPS26, putative	-1,3955802
glyma05g09120.1	highly similar to (533)AT5G50180	protein kinase, putative	-1,538603
glyma05g35830.1	moderately similar to (287)AT4G26080	ABI1 (ABA INSENSITIVE 1)	-5,357992
glyma03g39260.1	highly similar to (618)AT1G09160	protein phosphatase 2C-related / PP2C-related	-2,05911
glyma06g11410.2	moderately similar to (389)AT4G08500	MEKK1 (MYTOGEN ACTIVATED PROTEIN KINASE KINASE)	-2,5156968
glyma06g13920.1	highly similar to (750)AT5G24430	calcium-dependent protein kinase, putative / CDPK, putative	-2,577779
glyma17g36910.1	highly similar to (684)AT2G16250	leucine-rich repeat transmembrane protein kinase, putative	-2,3342798
glyma11g04200.1	moderately similar to (431)AT5G47070	protein kinase, putative	-1,8931028
glyma19g40500.1	highly similar to (882)AT4G02010	protein kinase family protein	-1,9146696
glyma10g31660.1	moderately similar to (379)AT4G10790	UBX domain-containing protein	-1,2371322
glyma20g34110.1	highly similar to (664)AT4G17740	C-terminal processing protease, putative	-2,428904
glyma09g05820.1	highly similar to (963)AT1G06430	FTSH8 (FtsH protease 8)	-3,9972494
glyma08g44530.1	moderately similar to (414)AT5G55970	zinc finger (C3HC4-type RING finger) family protein	-2,945985
glyma15g11600.1	moderately similar to (325)AT4G22030	F-box family protein	-3,319338
RNA			
ID		Transcript name	Log(FC)
glyma08g43380.1	highly similar to (558)AT3G20300	extracellular ligand-gated ion channel	-1,4756292
glyma12g08110.1	highly similar to (732)AT2G28350	ARF10 (AUXIN RESPONSE FACTOR 10)	-1,8795288
glyma17g10170.1	moderately similar to (208)AT3G48100	ARR5 (ARABIDOPSIS RESPONSE REGULATOR 5)	-1,5414329
glyma17g35180.1	moderately similar to (456)AT3G49810	U-box domain-containing protein	-1,1742028
glyma13g19380.1	moderately similar to (449)AT3G53710	AGD6	-1,2649902
glyma07g05250.1	weakly similar to (136)AT3G54770	RNA recognition motif (RRM)-containing protein	-1,4731082
glyma09g33790.1	moderately similar to (213)AT2G35410	33 kDa ribonucleoprotein, chloroplast, putative / RNA-binding protein cp33, putative	-1,0876963
Development			
ID		Transcript name	Log(FC)
glyma07g29560.1	weakly similar to (182)AT1G49510	EMB1273 (EMBRYO DEFECTIVE 1273)	-4,0998063
Hormones			
ID		Transcript name	Log(FC)
glyma20g01400.1	moderately similar to (223)AT5G08350	GRAM domain-containing protein / ABA-responsive protein-related	-1,4340905
glyma03g42250.2	moderately similar to (461)AT4G10500	oxidoreductase, 2OG-Fe(II) oxygenase family protein	-1,3884554
glyma10g00980.1	weakly similar to (157)AT3G23240	ATERF1/ERF1 (ETHYLENE RESPONSE FACTOR 1)	-1,9391391
glyma11g03880.1	moderately similar to (294)AT5G47310	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT4G17486.1)	-1,0300725
Signal regulation			
ID		Transcript name	Log(FC)
glyma09g36460.1	nearly identical (1157)AT5G61480	leucine-rich repeat transmembrane protein kinase, putative	-1,3677999
glyma12g01410.1	moderately similar to (258)AT5G13460	IQD11 (IQ-domain 11)	-1,3802853
glyma07g01760.1	moderately similar to (250)AT3G16490	IQD26 (IQ-domain 26)	-1,5843486
glyma12g33550.1	moderately similar to (332)AT5G60860	AtRABA1f (Arabidopsis Rab GTPase homolog A1f)	-1,8473591
glyma06g24770.1	weakly similar to (164)AT2G26695	binding	-4,7167215
glyma05g31020.1	moderately similar to (365)AT5G65270	AtRABA4a (Arabidopsis Rab GTPase homolog A4a)	-2,0684488
glyma01g39480.1	highly similar to (582)AT5G67130	phospholipase C	-1,2468892
glyma07g37690.1	weakly similar to (124)AT4G15800	RALFL33 (RALF-LIKE 33)	-1,1732126

Transport			
ID		Transcript name	Log(FC)
glyma09g37570.1	moderately similar to (387)AT3G01280	porin, putative	-1,4323416
glyma02g42800.1	moderately similar to (411)AT3G56200	amino acid transporter family protein	-2,5431714
glyma13g39850.1	highly similar to (815)AT1G08090	ATNRT2:1 (Arabidopsis thaliana high affinity nitrate transporter 2.1)	-1,1981598
glyma05g37810.1	moderately similar to (422)AT4G11440	binding	-1,1151502
glyma04g41140.1	highly similar to (987)AT4G27730	ATOPT6 (oligopeptide transporter 6)	-1,8122624
glyma14g01610.1	highly similar to (991)AT5G55930	ATOPT1 (oligopeptide transporter 1)	-3,5913963
glyma05g08230.1	nearly identical (1249)AT2G26650	AKT1 (ARABIDOPSIS K TRANSPORTER 1)	-1,35865
glyma15g01490.1	nearly identical (2112)AT1G15520	ATPDR12/PDR12 (PLEIOTROPIC DRUG RESISTANCE 12)	-2,0745056
glyma17g09660.1	highly similar to (944)AT2G47160	BOR1 (REQUIRES HIGH BORON 1)	-2,7467802
glyma11g03690.1	moderately similar to (273)AT4G17340	DELTA-TIP2/TIP2	-1,3328745
glyma15g05190.1	moderately similar to (265)AT1G09580	emp24/gp25L/p24 family protein	-1,8343254
Enzymes			
ID		Transcript name	Log(FC)
glyma18g43980.1	moderately similar to (395)AT2G15480	UGT73B5 (UDP-GLUCOSYL TRANSFERASE 73B5)	-1,2092091
glyma04g03430.1	moderately similar to (456)AT5G42760	similar to unknown [Populus trichocarpa] (GB:ABK95091.1)	-1,365997
glyma09g41570.1	moderately similar to (371)AT3G26300	CYP71B34 (cytochrome P450, family 71, subfamily B, polypeptide 34)	-1,4374844
glyma15g00450.1	highly similar to (586)AT5G25900	GA3 (GA REQUIRING 3)	-1,9402403
glyma09g42160.1	moderately similar to (265)AT4G11290	peroxidase, putative	-1,8877459
glyma08g42670.1	highly similar to (681)AT3G20500	ATPAP18/PAP18 (purple acid phosphatase 18)	-3,0738146
glyma07g06130.1	nearly identical (1009)AT3G61760	dynamain-like protein B (DL1B)	-1,7163105
glyma08g25810.1	moderately similar to (445)AT3G12800	short-chain dehydrogenase/reductase (SDR) family protein	-1,3011248
glyma18g51670.1	moderately similar to (295)AT2G38130	ATMAK3 (Arabidopsis thaliana MAK3 homologue)	-4,7110944
Cell wall			
ID		Transcript name	Log(FC)
glyma15g43040.1	nearly identical (1898)AT5G05170	CESA3 (CELLULOSE SYNTHASE 3)	-1,153026
glyma06g07320.1	nearly identical (1950)AT4G32410	CESA1 (CELLULOSE SYNTHASE 1)	-3,086822
glyma08g19280.1	nearly identical (1173)AT5G64570	XYL4 (beta-xylosidase 4)	-3,8888621
glyma10g29150.1	highly similar to (524)AT4G02330	ATPMEPCRB	-3,686737
Specialized metabolism			
ID		Transcript name	Log(FC)
glyma14g40170.1	moderately similar to (426)AT4G39330	mannitol dehydrogenase, putative	-1,4127247
Carbohydrates			
ID		Transcript name	Log(FC)
glyma05g02510.1	highly similar to (996)AT5G40390	SIP1 (SEED IMBIBITION 1-LIKE)	-2,5903356
Other metabolism			
ID		Transcript name	Log(FC)
glyma18g47760.1	moderately similar to (292)AT3G02770	dimethylmenaquinone methyltransferase family protein	-2,5959818
glyma14g02260.1	highly similar to (568)AT5G55130	CNX5 (SIRTINOL RESISTANT 1)	-1,1202729
glyma08g20080.1	highly similar to (682)AT1G22940	TH1 (THIAMINE REQUIRING 1)	-1,0603186
glyma17g15740.1	highly similar to (754)AT5G07440	GDH2 (GLUTAMATE DEHYDROGENASE 2)	-1,7605915
glyma03g33830.1	highly similar to (805)AT5G11880	diaminopimelate decarboxylase, putative / DAP carboxylase, putative	-1,044199
glyma18g12090.1	highly similar to (761)AT4G27070	TSB2 (TRYPTOPHAN SYNTHASE BETA-SUBUNIT)	-2,158763
glyma03g27150.1	moderately similar to (479)AT3G60510	enoyl-CoA hydratase/isomerase family protein	-1,5211747
glyma06g11860.1	nearly identical (1087)AT1G77590	LACS9 (LONG CHAIN ACYL-COA SYNTHETASE 9)	-2,5434408
glyma14g07290.1	highly similar to (674)AT3G11430	ATGPAT5/GPAT5 (GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 5)	-1,6100144
glyma07g07580.1	highly similar to (768)AT4G00400	GPAT8 (GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 8)	-4,2262893
glyma11g03730.1	moderately similar to (278)AT4G16700	PSD1 (PHOSPHATIDYLSERINE DECARBOXYLASE 1)	-2,7870417
glyma19g35340.1	highly similar to (694)AT5G43940	ADH2 (ALCOHOL DEHYDROGENASE 2)	-1,0497478

I. Analyse génomique de l'ADNex

II-1 Alignement des séquences

Tableau III-3 : Analyse du nombre de séquences de l'ADNex alignées (*counts*) sur le génome complet (chromosomes, mitochondrie et plaste) du soja pour le réplicat 8 NE (ADNex en condition non élicitée)

Pour les chromosomes, plastes et mitochondrie, un ratio *count*/Mb a été calculé. Ce ratio correspond au nombre de *counts* alignés sur le génome, divisé par sa taille en mégabases (Mb). Enfin, les trois dernières lignes de ce tableau correspondent à la somme totale des *counts*, de la taille ou des ratios pour l'ensemble des chromosomes (ligne ADNchr), la mitochondrie et le plaste (ligne ADNm et ADNp) et enfin pour le génome complet (ligne ADN total). ADNchr = ADN chromosomique ; ADNm = ADN mitochondrial ; ADNp = ADN plastidial.

		8 NE		
	Chromosome / Organite	Somme des counts	Taille (Mb)	Count/Mb
ADNchr	GM01	65 013	56,83	1 143,99
	GM02	61 347	48,58	1 262,80
	GM03	51 334	45,78	1 121,32
	GM04	57 279	52,39	1 093,32
	GM05	48 574	42,23	1 150,22
	GM06	59 922	51,42	1 165,34
	GM07	51 210	44,63	1 147,43
	GM08	56 887	47,84	1 189,11
	GM09	57 080	50,19	1 137,28
	GM10	61 264	51,59	1 187,52
	GM11	46 762	34,77	1 344,90
	GM12	47 876	40,09	1 194,21
	GM13	54 607	45,87	1 190,47
	GM14	57 211	49,04	1 166,62
	GM15	56 732	51,76	1 096,06
	GM16	43 429	37,89	1 146,19
	GM17	47 729	41,64	1 146,23
	GM18	63 856	58,02	1 100,59
	GM19	58 913	50,75	1 160,85
	GM20	57 852	47,90	1 207,77
Organites	Mitochondrie	31 326	0,40	78 315,00
	Plastes	9 314	0,15	62 093,33
Total	ADNchr	1 104 877	949,21	23 352,22
	Mitochondrie et plastes	40 640	0,55	140 408,33
	ADN total (ADNchr + organites)	1 145 517	949,76	163 760,55

Tableau III-4 : Analyse du nombre de séquences de l'ADNex alignées (*counts*) sur le génome complet (chromosomes, mitochondrie et plaste) du soja pour le réplicat 9 NE (ADNex en condition non élicitée)

Pour les chromosomes, plastes et mitochondrie, un ratio *count*/Mb a été calculé. Ce ratio correspond au nombre de *counts* alignés sur le génome, divisé par sa taille en mégabases (Mb). Enfin, les trois dernières lignes de ce tableau correspondent à la somme totale des *counts*, de la taille ou des ratios pour l'ensemble des chromosomes (ligne ADNchr), la mitochondrie et le plaste (ligne ADNm et ADNp) et enfin pour le génome complet (ligne ADN total). ADNchr = ADN chromosomique ; ADNm = ADN mitochondrial ; ADNp = ADN plastidial.

9 NE				
	Chromosome / Organite	Somme des counts	Taille (Mb)	Count/Mb
ADNchr	GM01	72 805	56,83	1 281,10
	GM02	66 702	48,58	1 373,03
	GM03	55 654	45,78	1 215,68
	GM04	61 559	52,39	1 175,01
	GM05	53 088	42,23	1 257,12
	GM06	64 219	51,42	1 248,91
	GM07	55 533	44,63	1 244,30
	GM08	60 193	47,84	1 258,21
	GM09	62 774	50,19	1 250,73
	GM10	66 821	51,59	1 295,23
	GM11	49 643	34,77	1 427,75
	GM12	51 859	40,09	1 293,56
	GM13	57 677	45,87	1 257,40
	GM14	62 893	49,04	1 282,48
	GM15	61 331	51,76	1 184,91
	GM16	47 584	37,89	1 255,85
	GM17	51 627	41,64	1 239,84
	GM18	69 532	58,02	1 198,41
	GM19	65 067	50,75	1 282,11
	GM20	63 331	47,90	1 322,15
Organites	Mitochondrie	17 099	0,40	42 747,50
	Plastes	7 009	0,15	46 726,67
Total	ADNchr	1 199 892	949,21	25 343,81
	Mitochondrie et plastes	24 108	0,55	89 474,17
	ADN total (ADNchr + organites)	1 224 000	949,76	114 817,97

Tableau III-5 : Analyse du nombre de séquences de l'ADNex alignées (*counts*) sur le génome complet (chromosomes, mitochondrie et plaste) du soja pour le réplicat 8 E (ADNex en condition élicitée)

Pour les chromosomes, plastes et mitochondrie, un ratio *count/Mb* a été calculé. Ce ratio correspond au nombre de *counts* alignés sur le génome, divisé par sa taille en mégabases (Mb). Enfin, les trois dernières lignes de ce tableau correspondent à la somme totale des *counts*, de la taille ou des ratios pour l'ensemble des chromosomes (ligne ADNchr), la mitochondrie et le plaste (ligne ADNm et ADNp) et enfin pour le génome complet (ligne ADN total). ADNchr = ADN chromosomique ; ADNm = ADN mitochondrial ; ADNp = ADN plastidial.

8 E				
	Chromosome / Organite	Somme des count	Taille (Mb)	Count/Mb
ADNchr	GM01	68 999	56,83	1 214,13
	GM02	64 339	48,58	1 324,39
	GM03	53 968	45,78	1 178,86
	GM04	60 404	52,39	1 152,97
	GM05	51 399	42,23	1 217,12
	GM06	62 612	51,42	1 217,66
	GM07	54 338	44,63	1 217,52
	GM08	59 595	47,84	1 245,71
	GM09	59 769	50,19	1 190,85
	GM10	65 502	51,59	1 269,66
	GM11	48 761	34,77	1 402,39
	GM12	50 352	40,09	1 255,97
	GM13	58 165	45,87	1 268,04
	GM14	59 858	49,04	1 220,60
	GM15	59 098	51,76	1 141,77
	GM16	45 400	37,89	1 198,21
	GM17	50 046	41,64	1 201,87
	GM18	66 795	58,02	1 151,24
	GM19	61 801	50,75	1 217,75
	GM20	60 118	47,90	1 255,07
Organites	Mitochondrie	26 883	0,40	67 207,50
	Plastes	9 971	0,15	66 473,33
Total	ADNchr	1 161 319	949,21	24 541,79
	Mitochondrie et plastes	36 854	0,55	133 680,83
	ADN total (ADNchr + organites)	1 198 173	949,76	158 222,63

Tableau III-6 : Analyse du nombre de séquences de l'ADNex alignées (*counts*) sur le génome complet (chromosomes, mitochondrie et plaste) du soja pour le réplicat 9 E (ADNex en condition élicitée)

Pour les chromosomes, plastes et mitochondrie, un ratio *count/Mb* a été calculé. Ce ratio correspond au nombre de *counts* alignés sur le génome, divisé par sa taille en mégabases (Mb). Enfin, les trois dernières lignes de ce tableau correspondent à la somme totale des *counts*, de la taille ou des ratios pour l'ensemble des chromosomes (ligne ADNchr), la mitochondrie et le plaste (ligne ADNm et ADNp) et enfin pour le génome complet (ligne ADN total). ADNchr = ADN chromosomique ; ADNm = ADN mitochondrial ; ADNp = ADN plastidial.

9 E				
	Chromosome / Organite	Somme des counts	Taille (Mb)	Count/Mb
ADNchr	GM01	50 042	56,83	880,56
	GM02	46 554	48,58	958,30
	GM03	39 536	45,78	863,61
	GM04	43 911	52,39	838,16
	GM05	37 815	42,23	895,45
	GM06	37 815	51,42	735,41
	GM07	39 277	44,63	880,06
	GM08	43 394	47,84	907,07
	GM09	43 683	50,19	870,35
	GM10	47 015	51,59	911,32
	GM11	35 718	34,77	1 027,26
	GM12	36 392	40,09	907,76
	GM13	41 976	45,87	915,11
	GM14	44 017	49,04	897,57
	GM15	42 512	51,76	821,33
	GM16	33 777	37,89	891,45
	GM17	35 964	41,64	863,69
	GM18	49 075	58,02	845,83
	GM19	45 682	50,75	900,14
	GM20	44 390	47,90	926,72
Organites	Mitochondrie	16 517	0,40	41 292,50
	Plastes	8 450	0,15	56 333,33
Total	ADNchr	838 545	949,21	17 737,14
	Mitochondrie et plastes	24 967	0,55	97 625,83
	ADN total (ADNchr + organites)	863 512	949,76	115 362,97

II-2 Observation de la couverture des chromosomes avec IGV

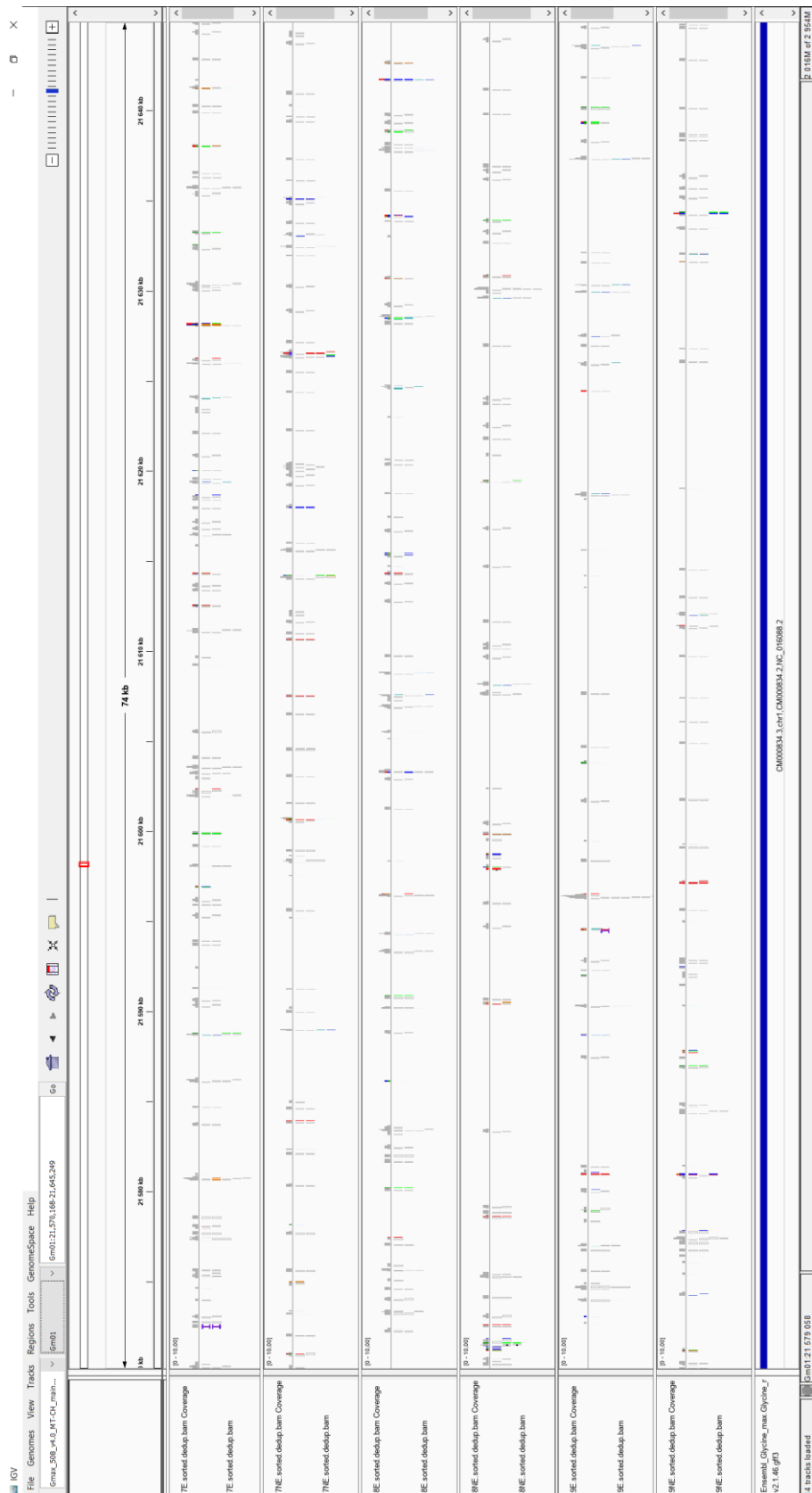


Figure III-1 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM01

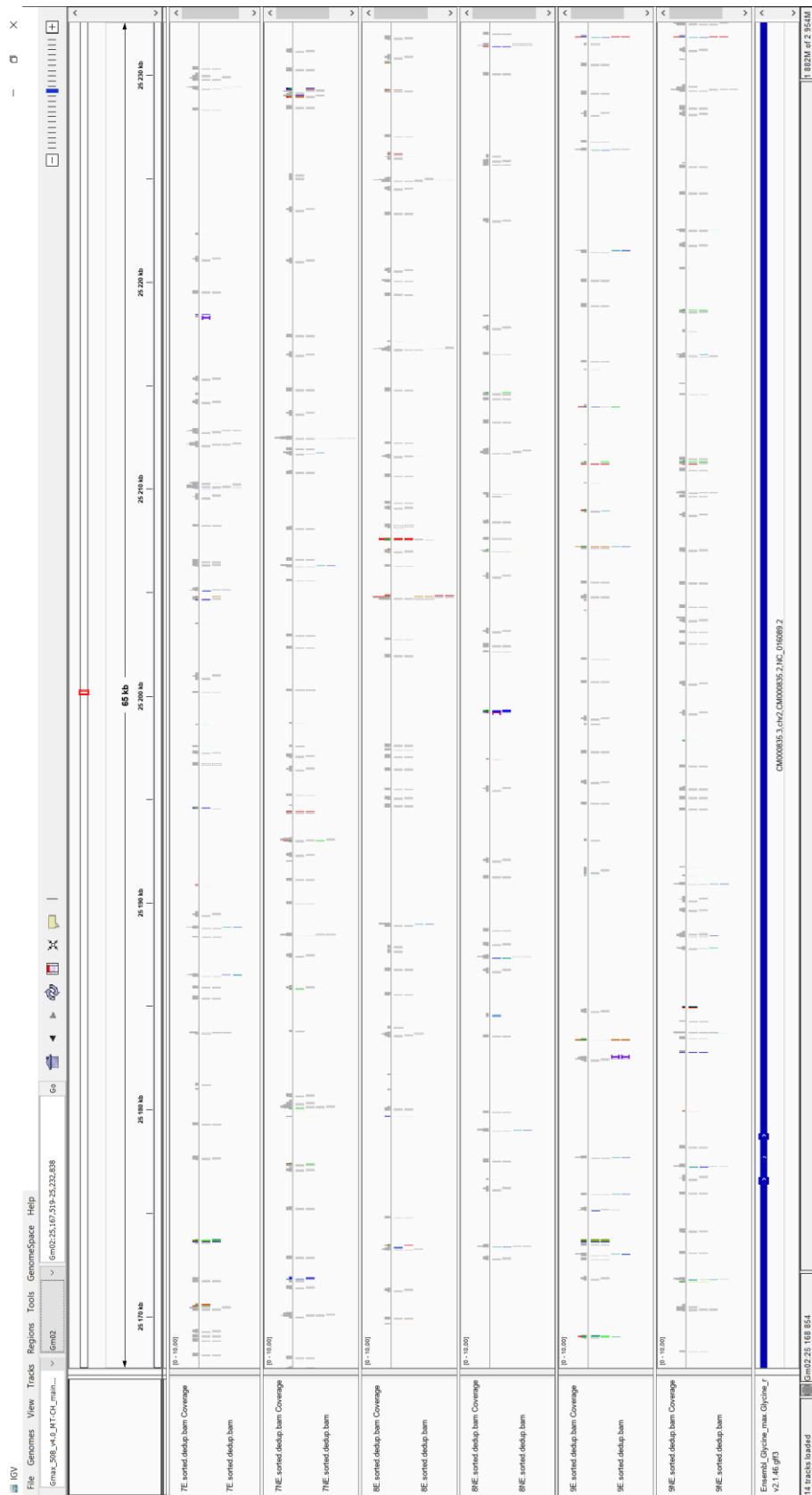


Figure III-2 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM02

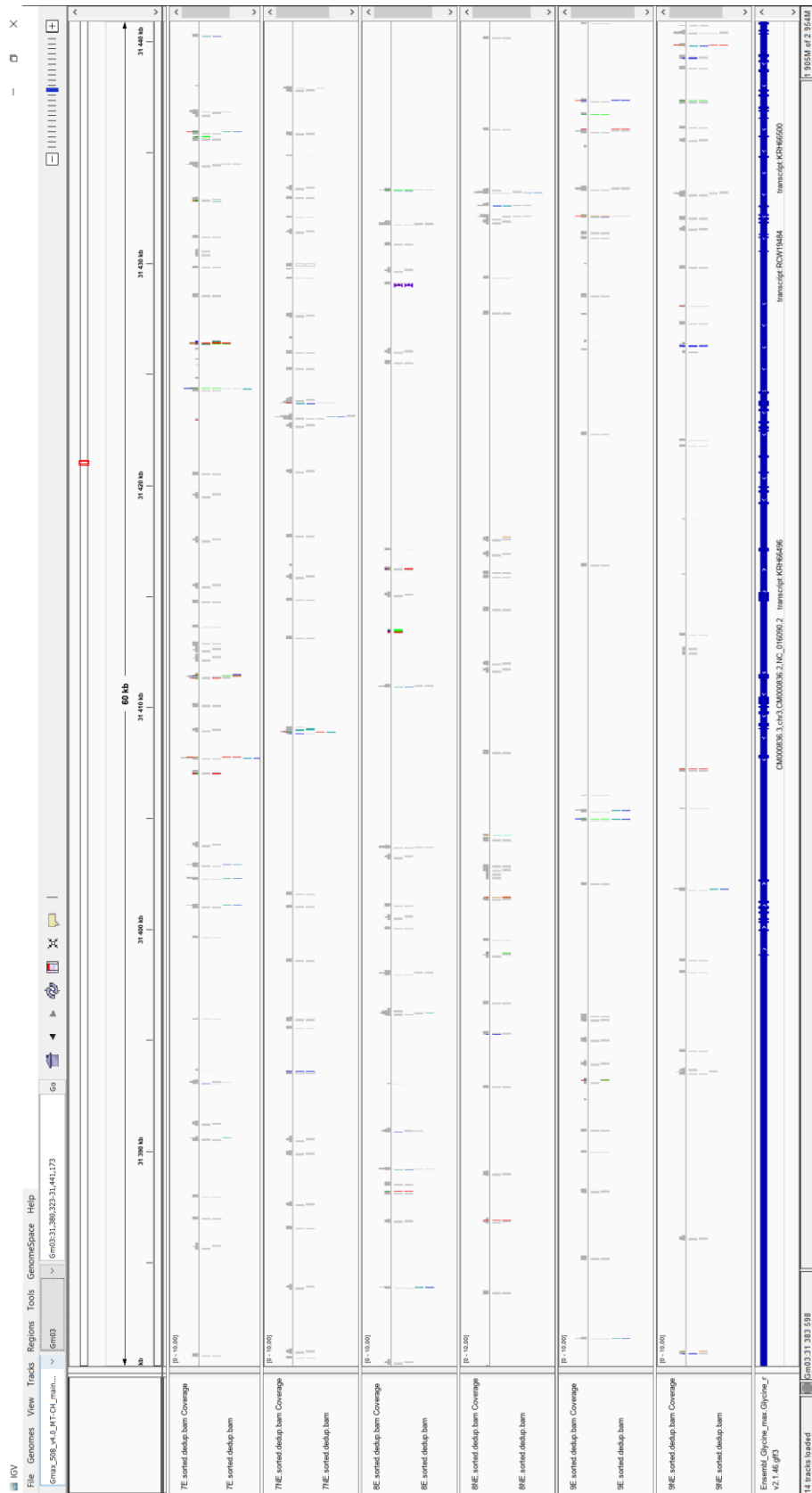


Figure III-3 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM03

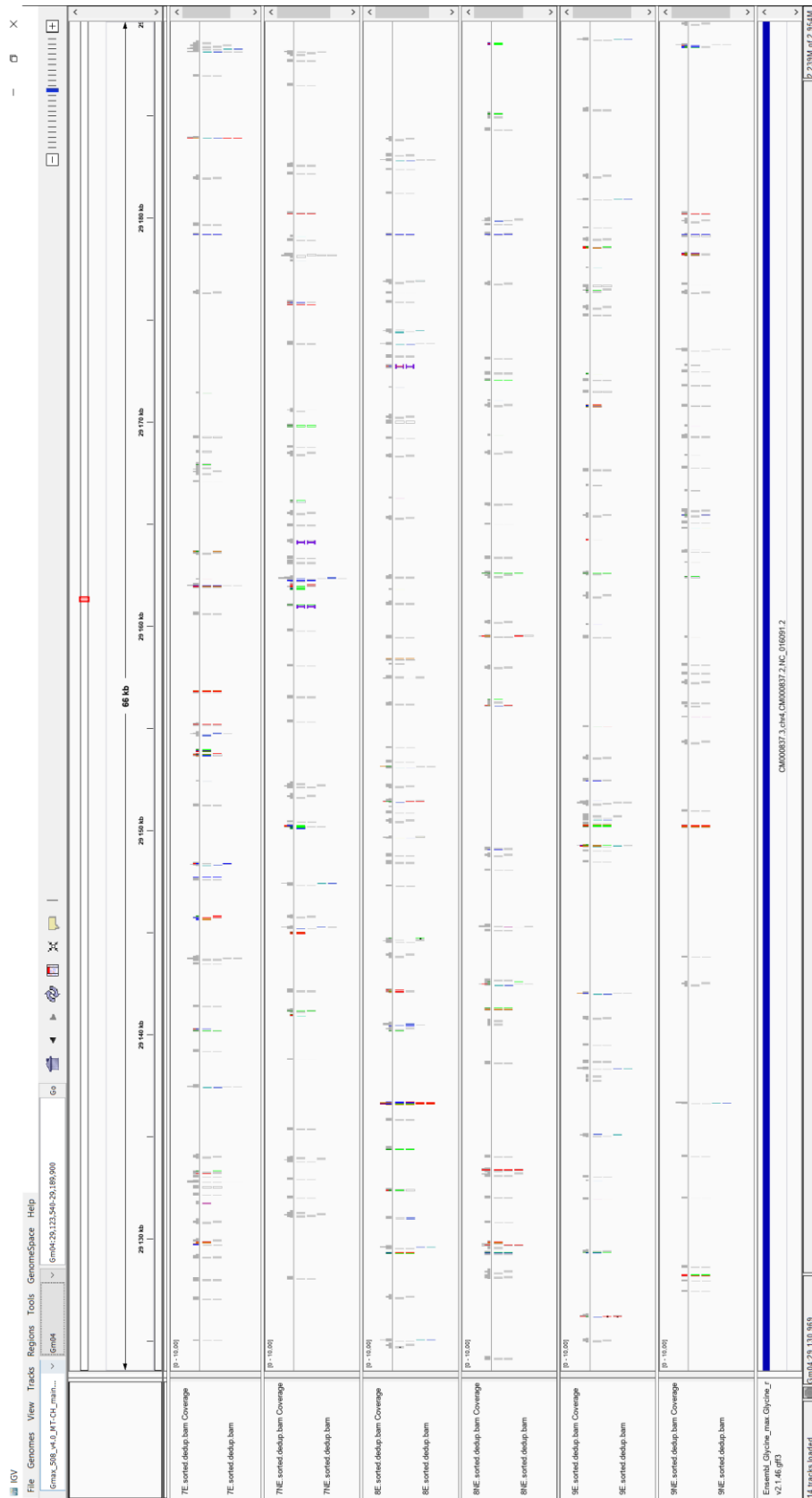


Figure III-4 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM04

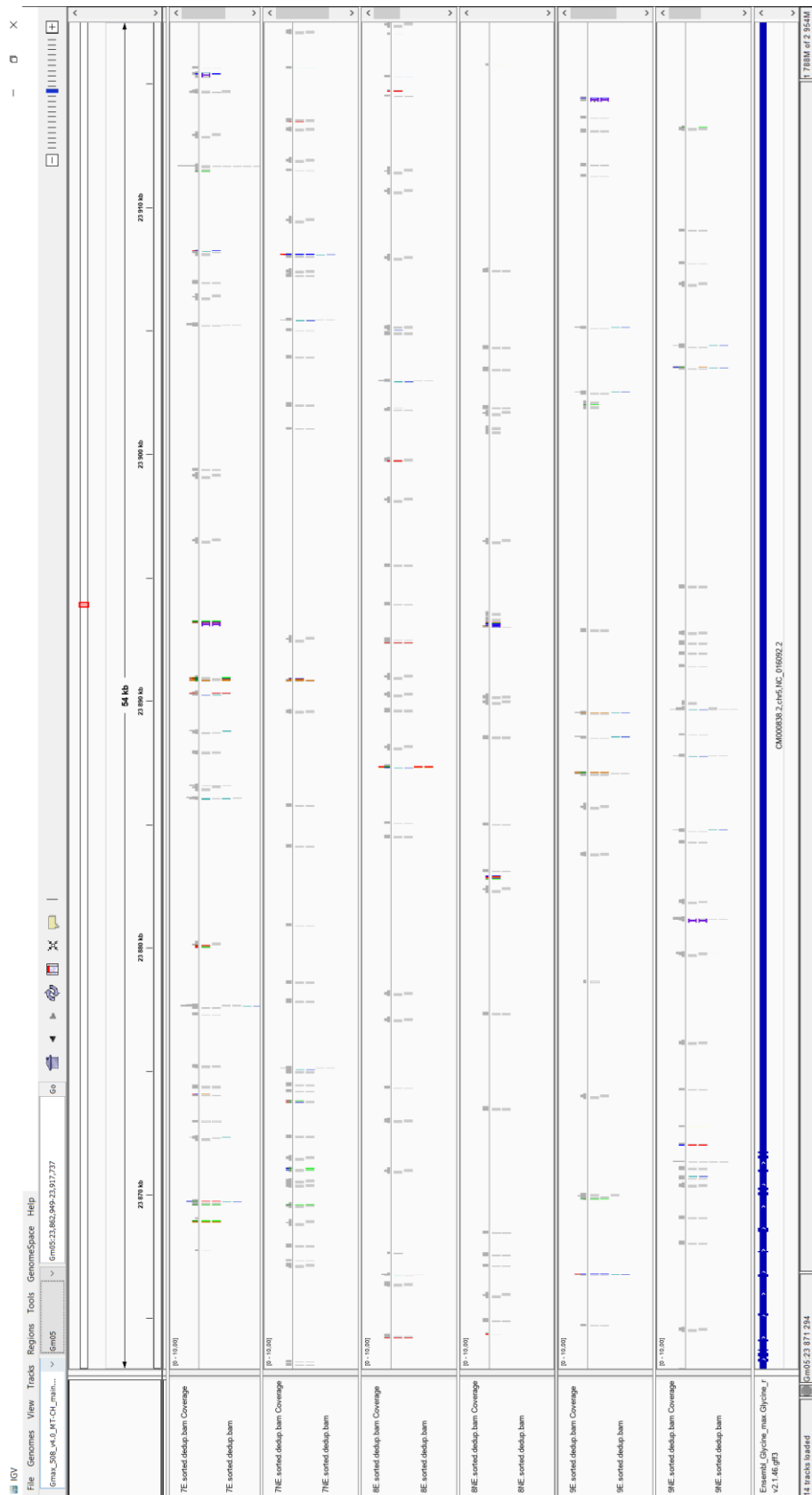


Figure III-5 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM05

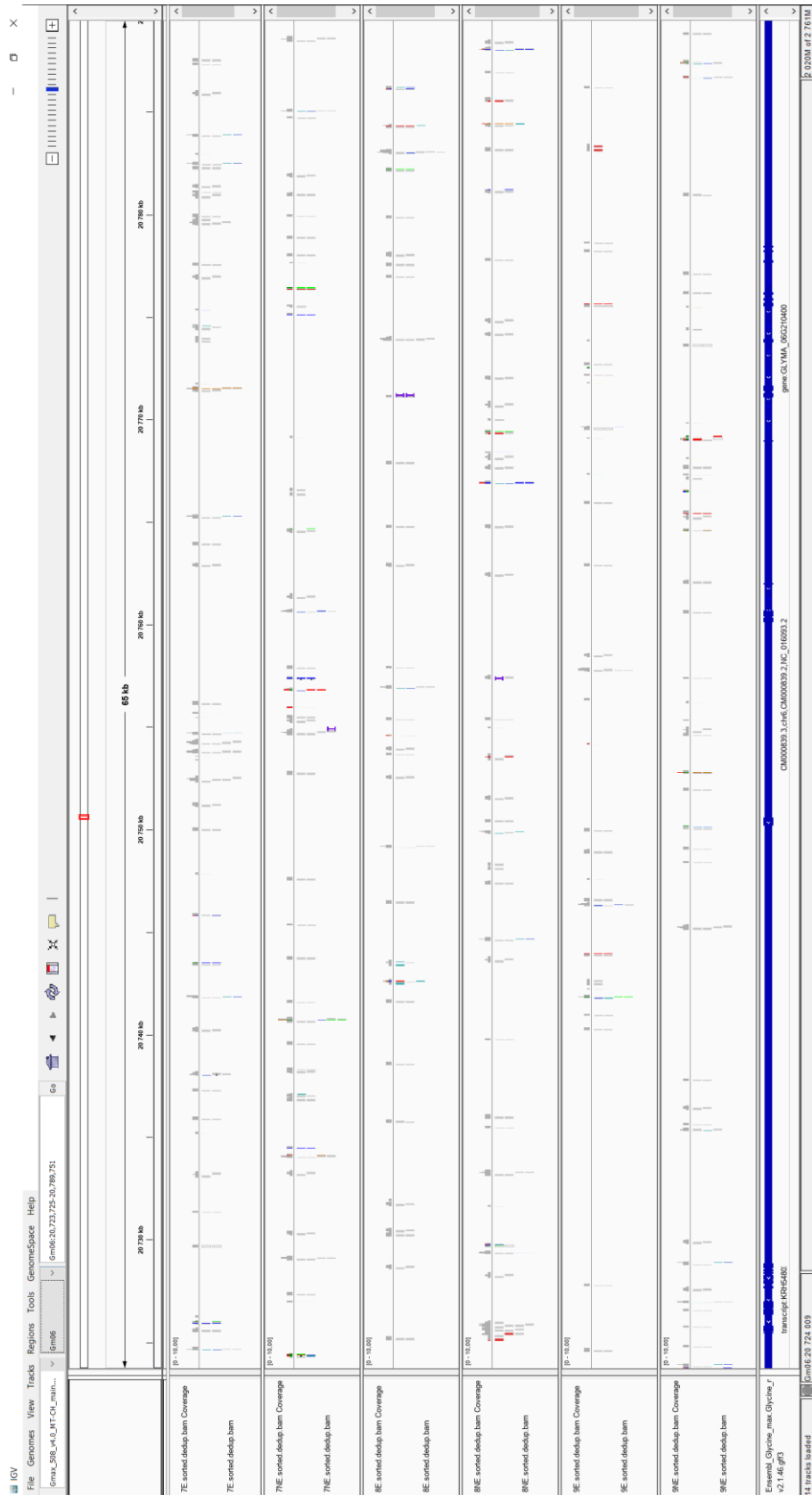


Figure III-6 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM06

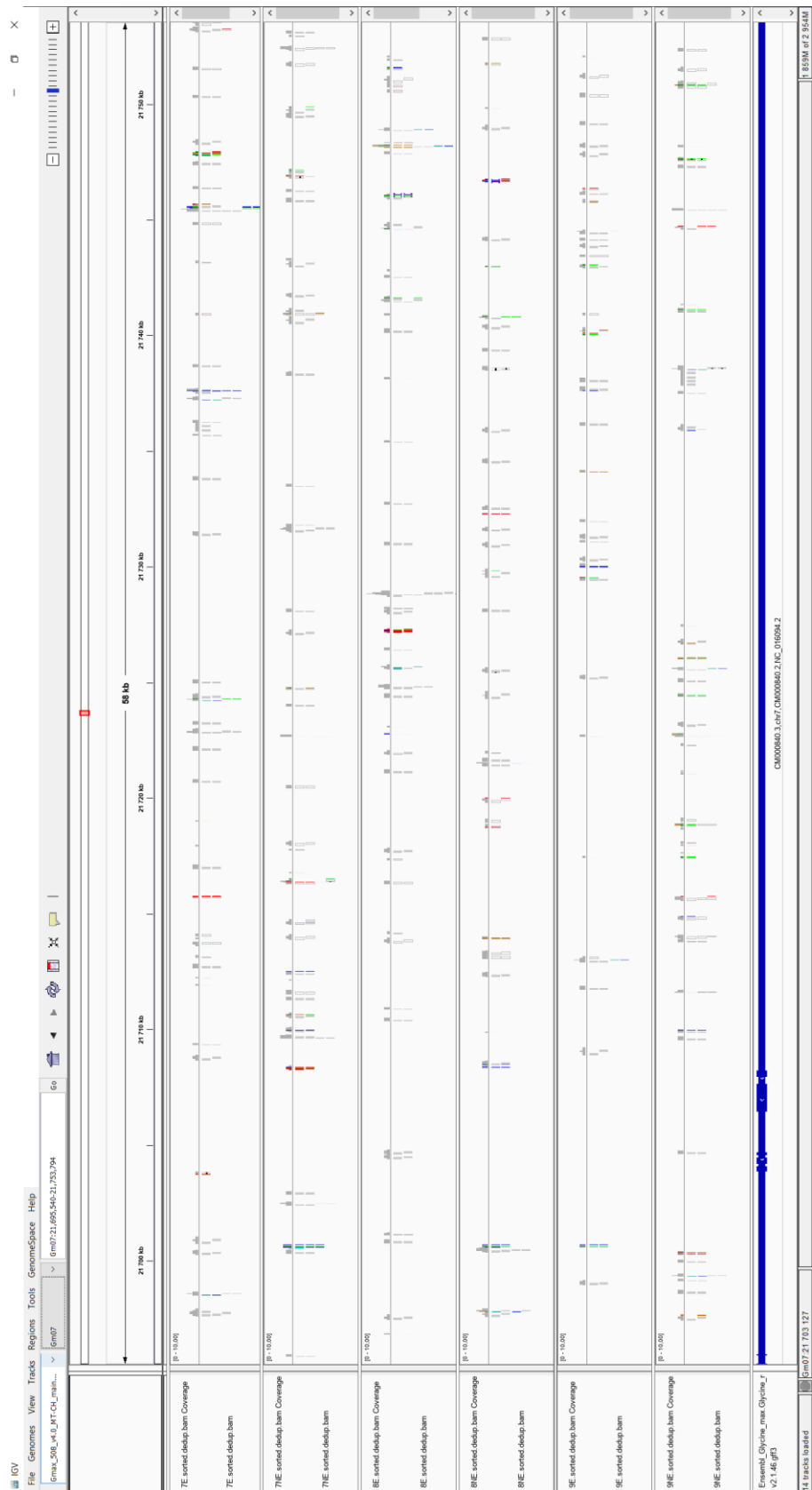


Figure III-7 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM07

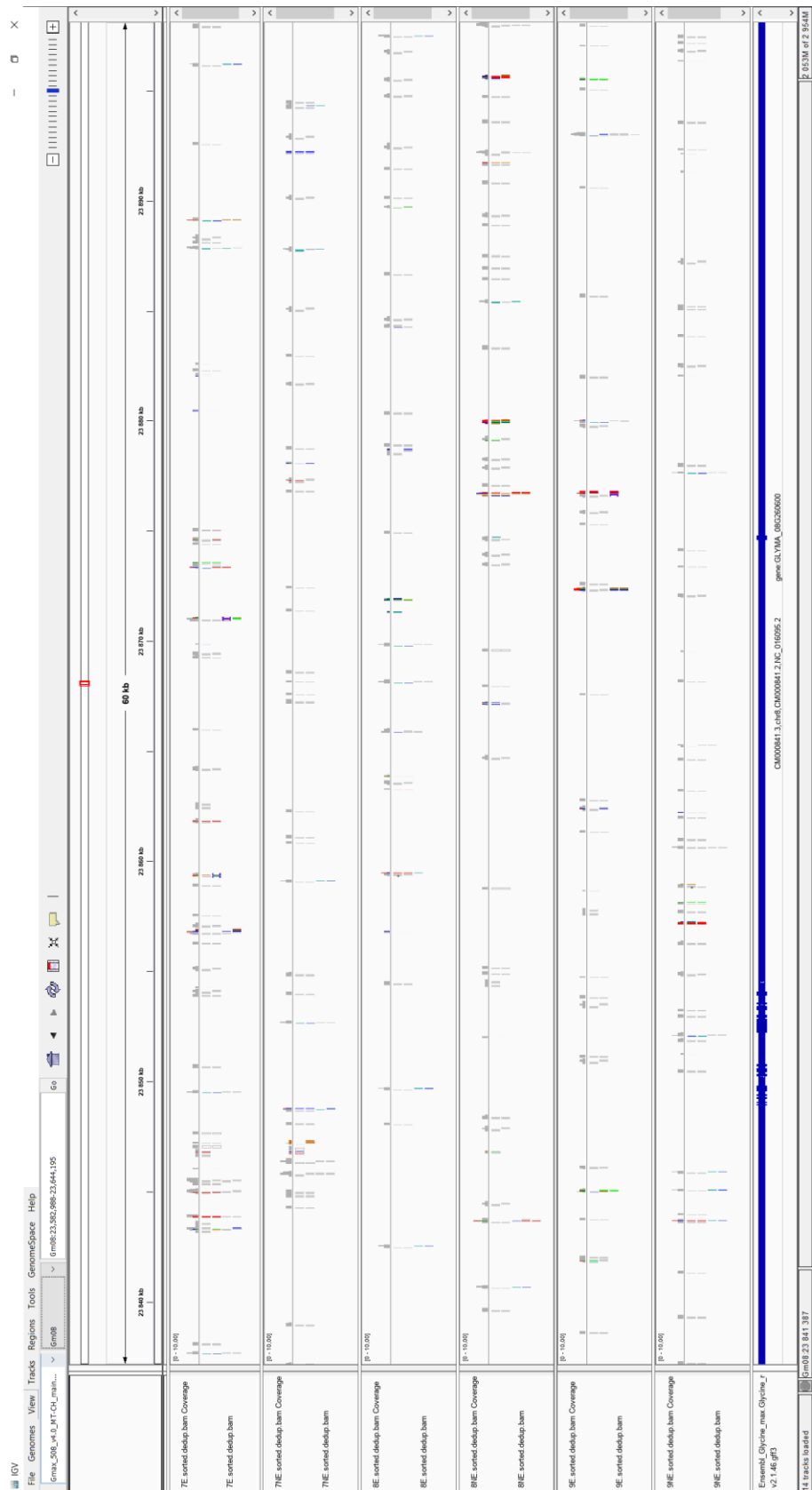


Figure III-8 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM08

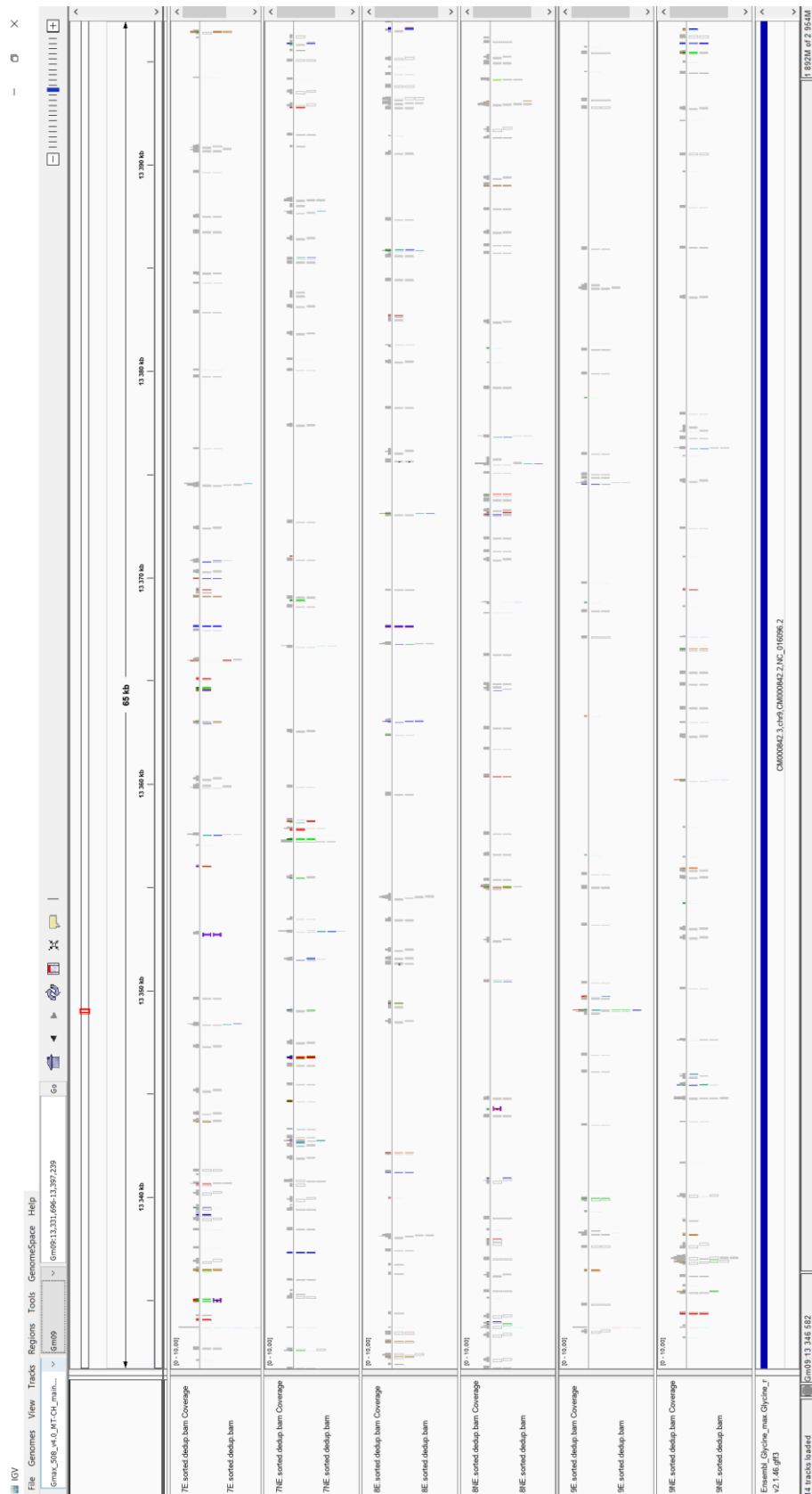


Figure III-9 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM09

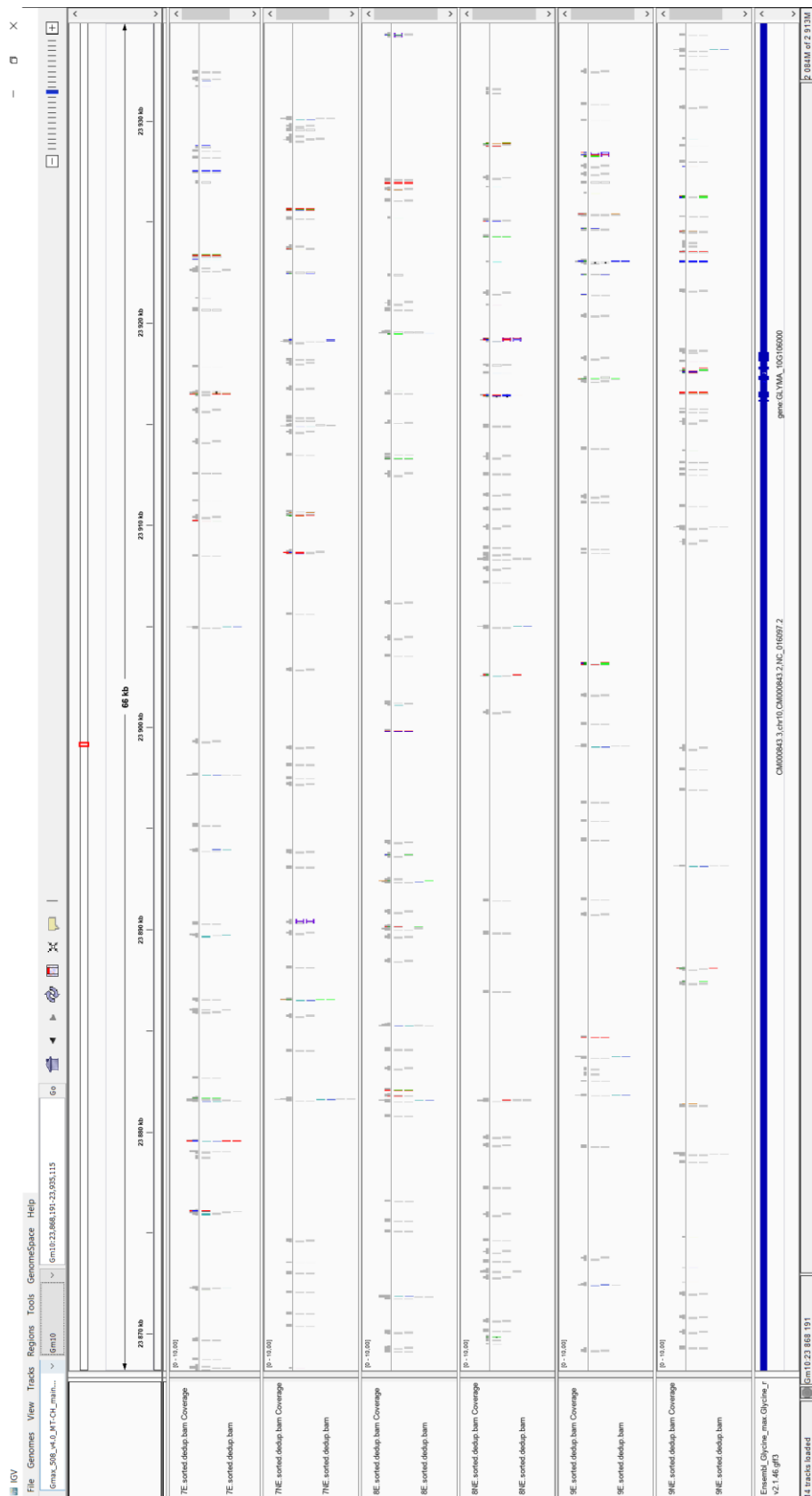


Figure III-10 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM10

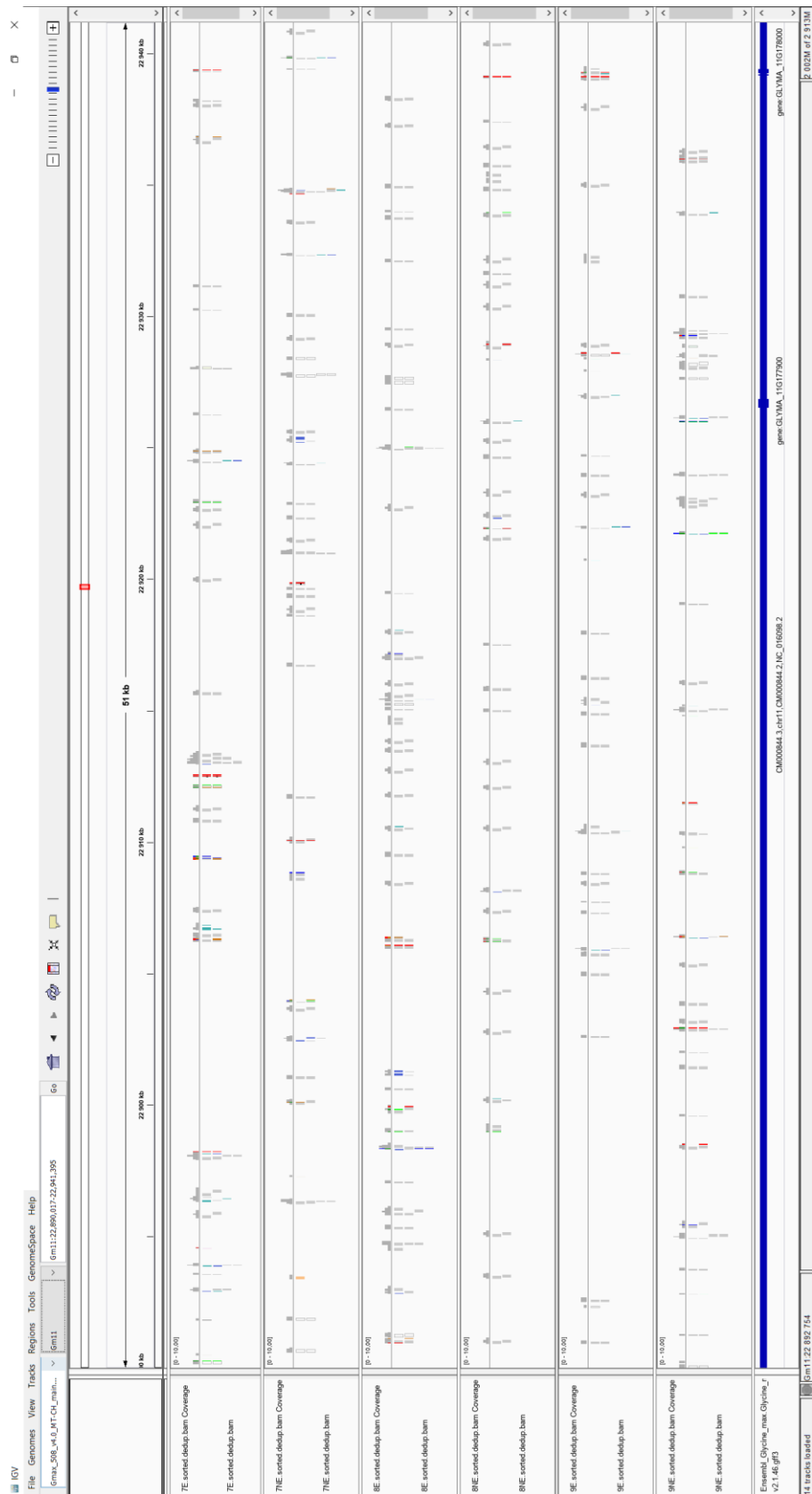


Figure III-11 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM11

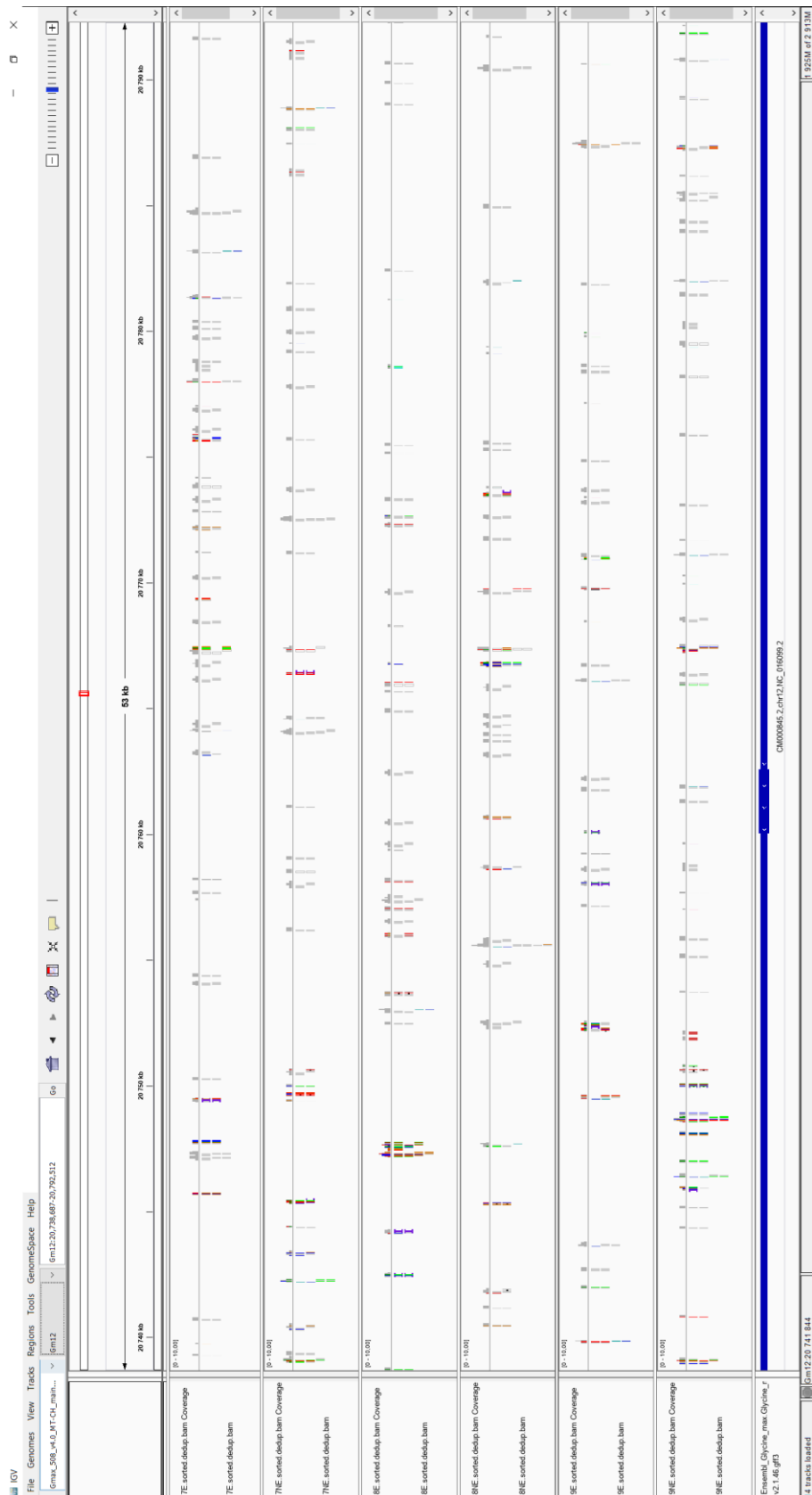


Figure III-12 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM12

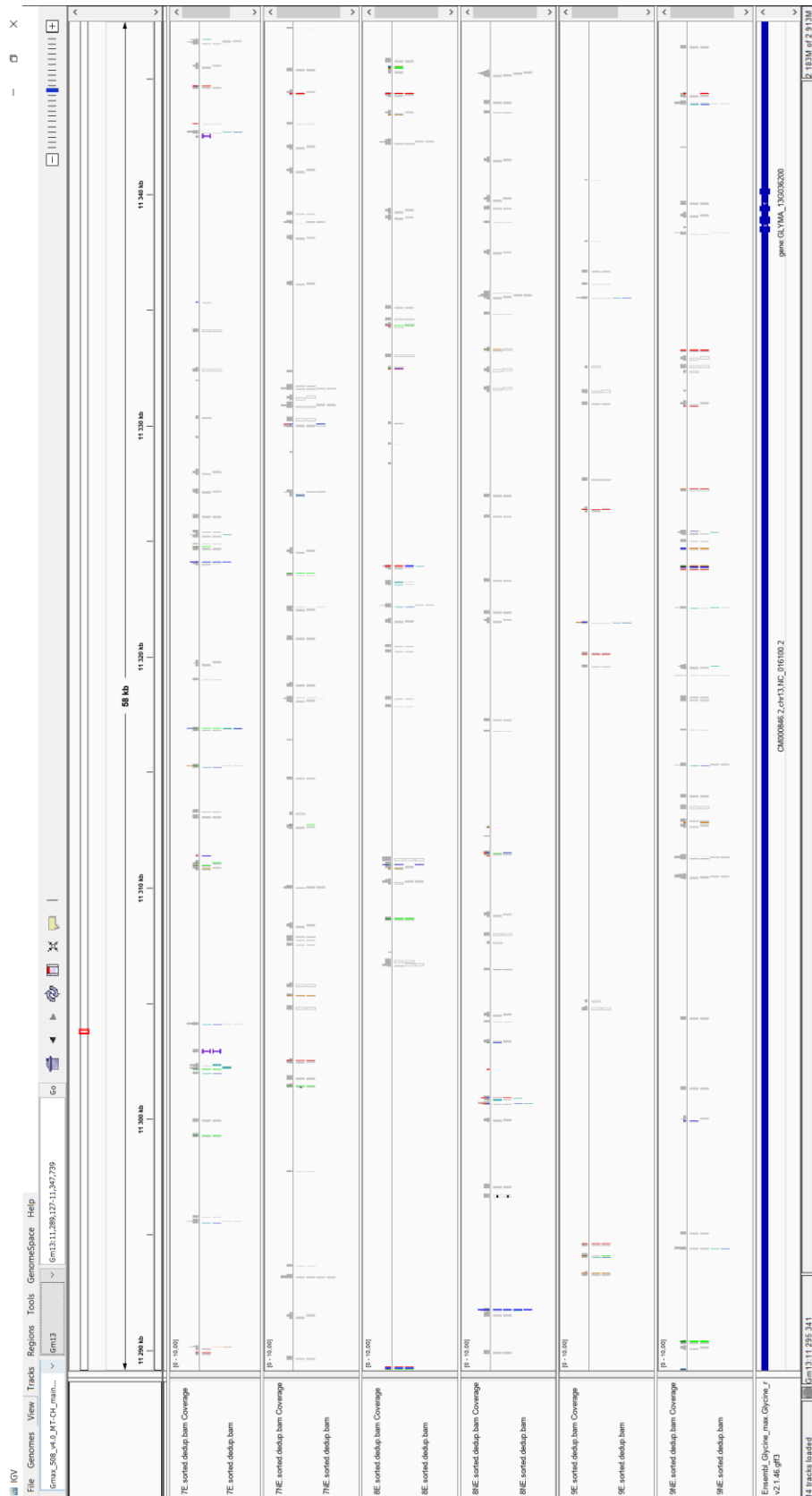


Figure III-13 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM13

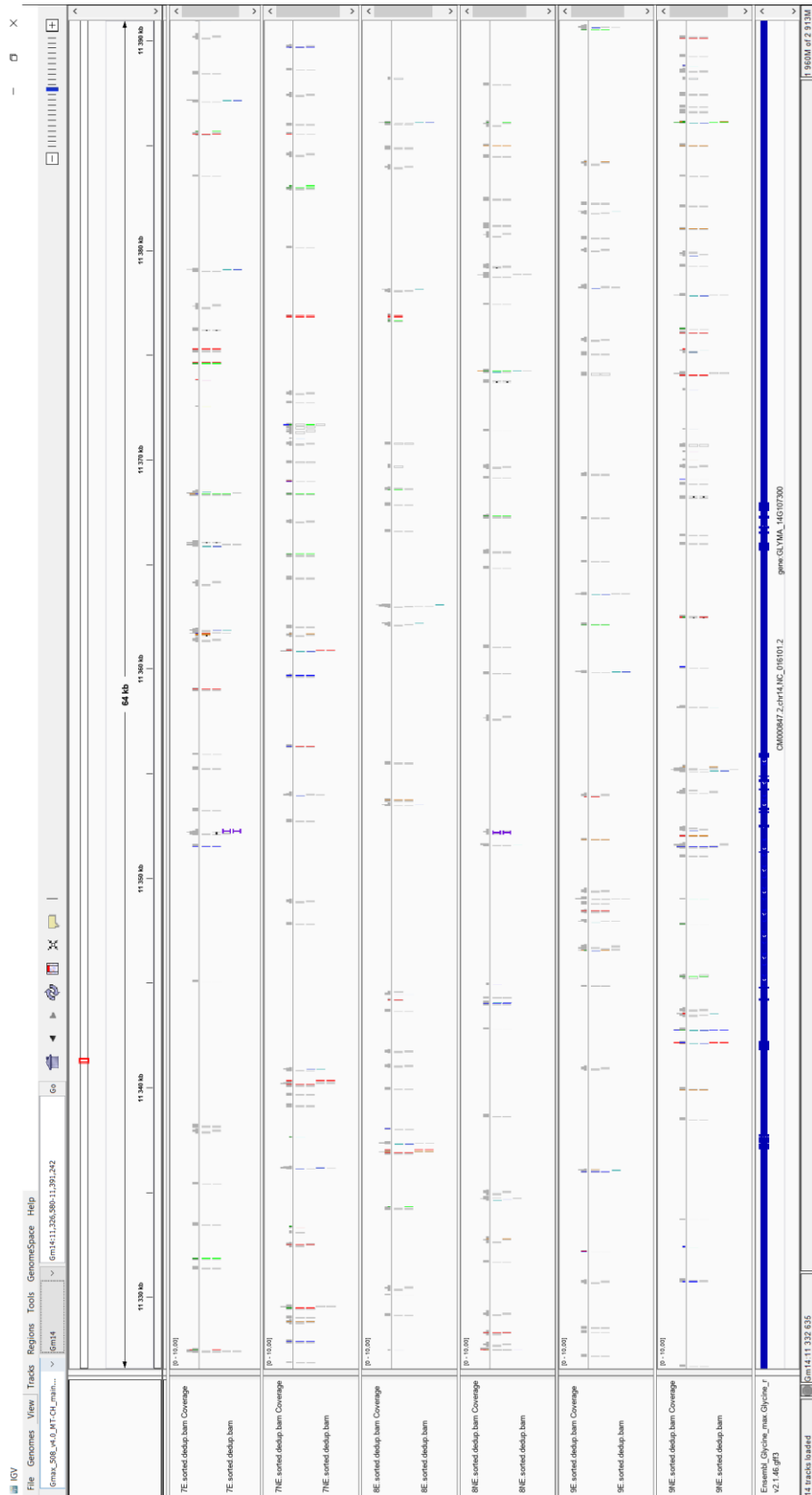


Figure III-14 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM14

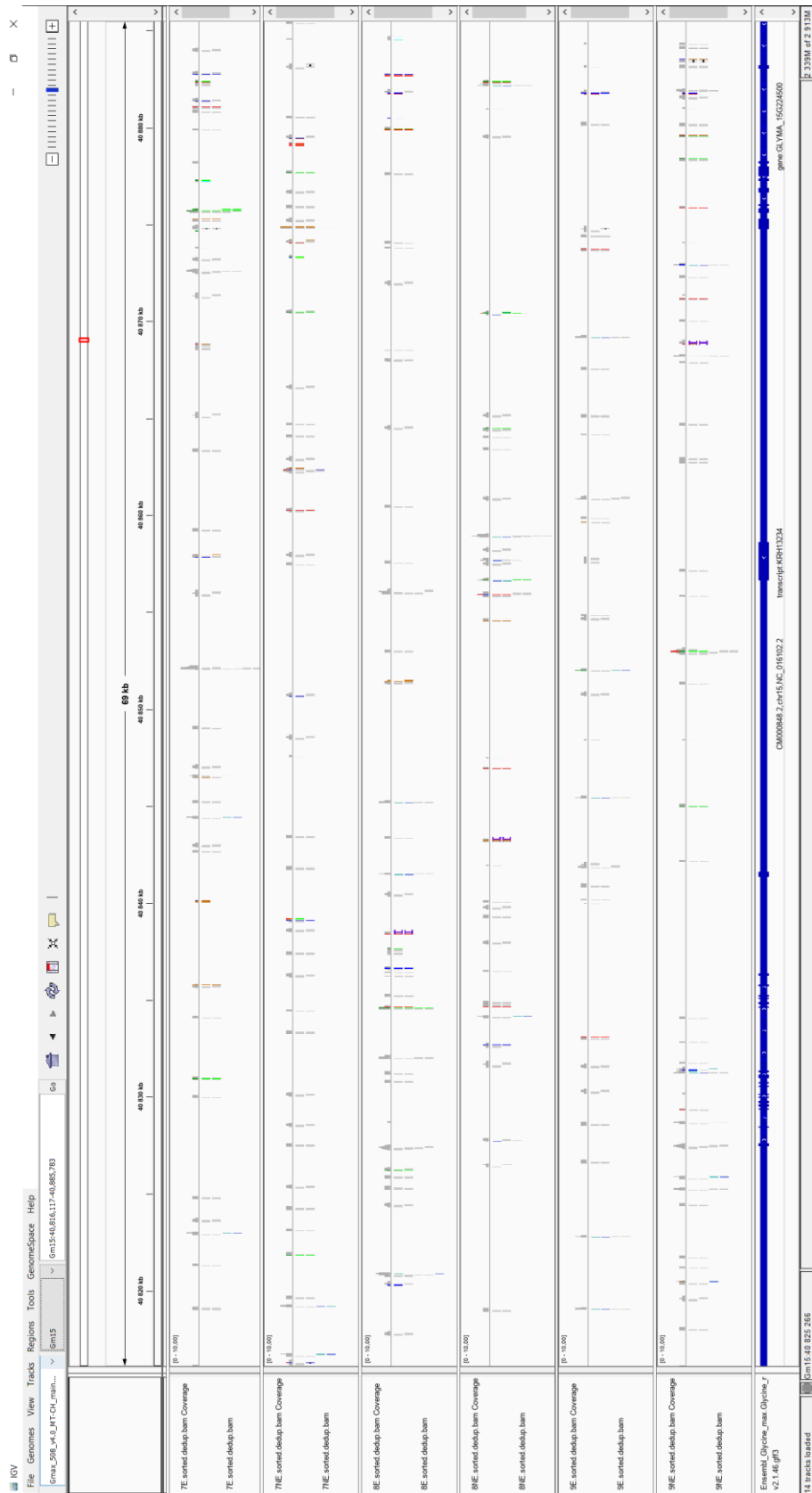


Figure III-15 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM15

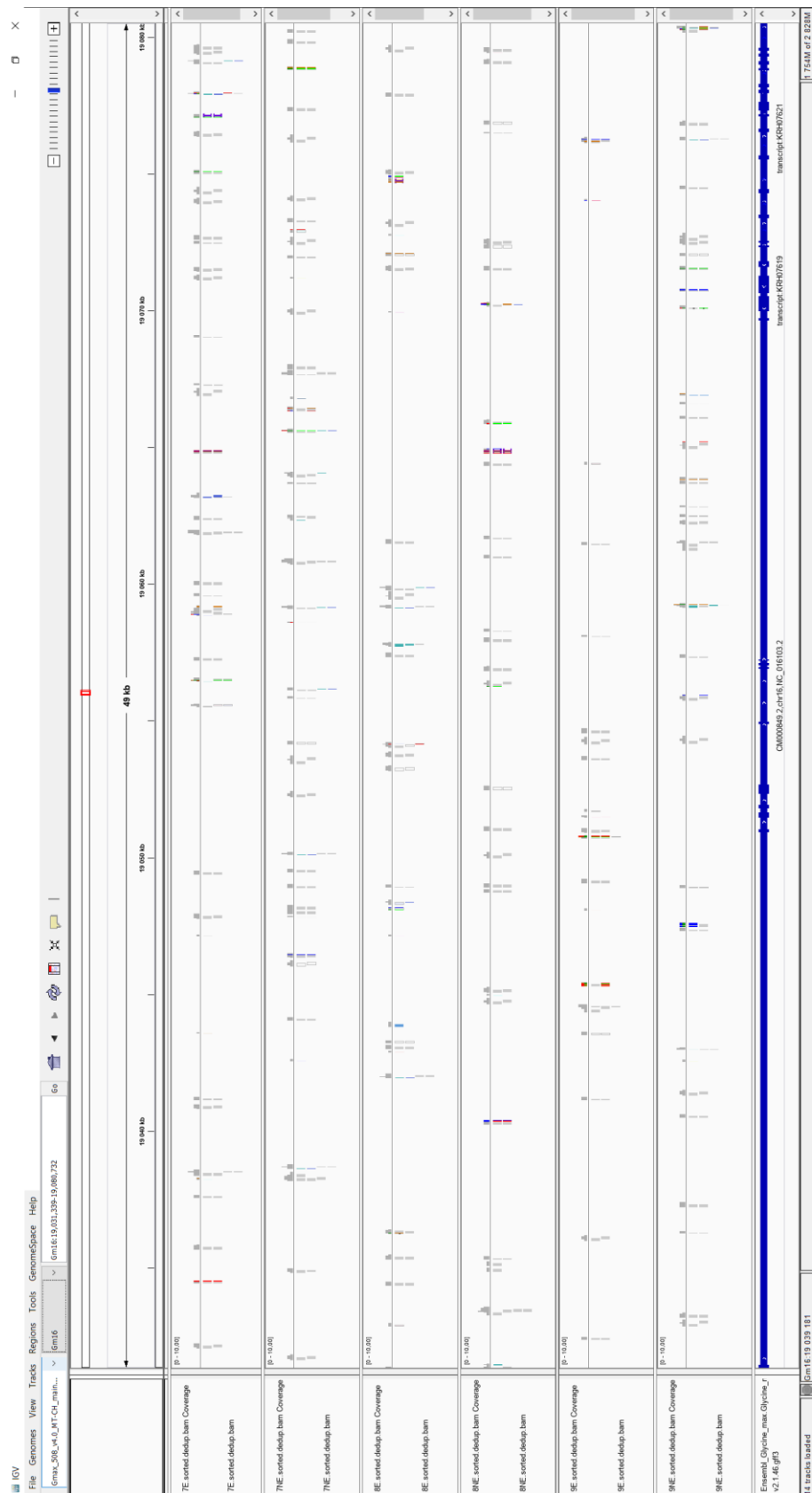


Figure III-16 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM16

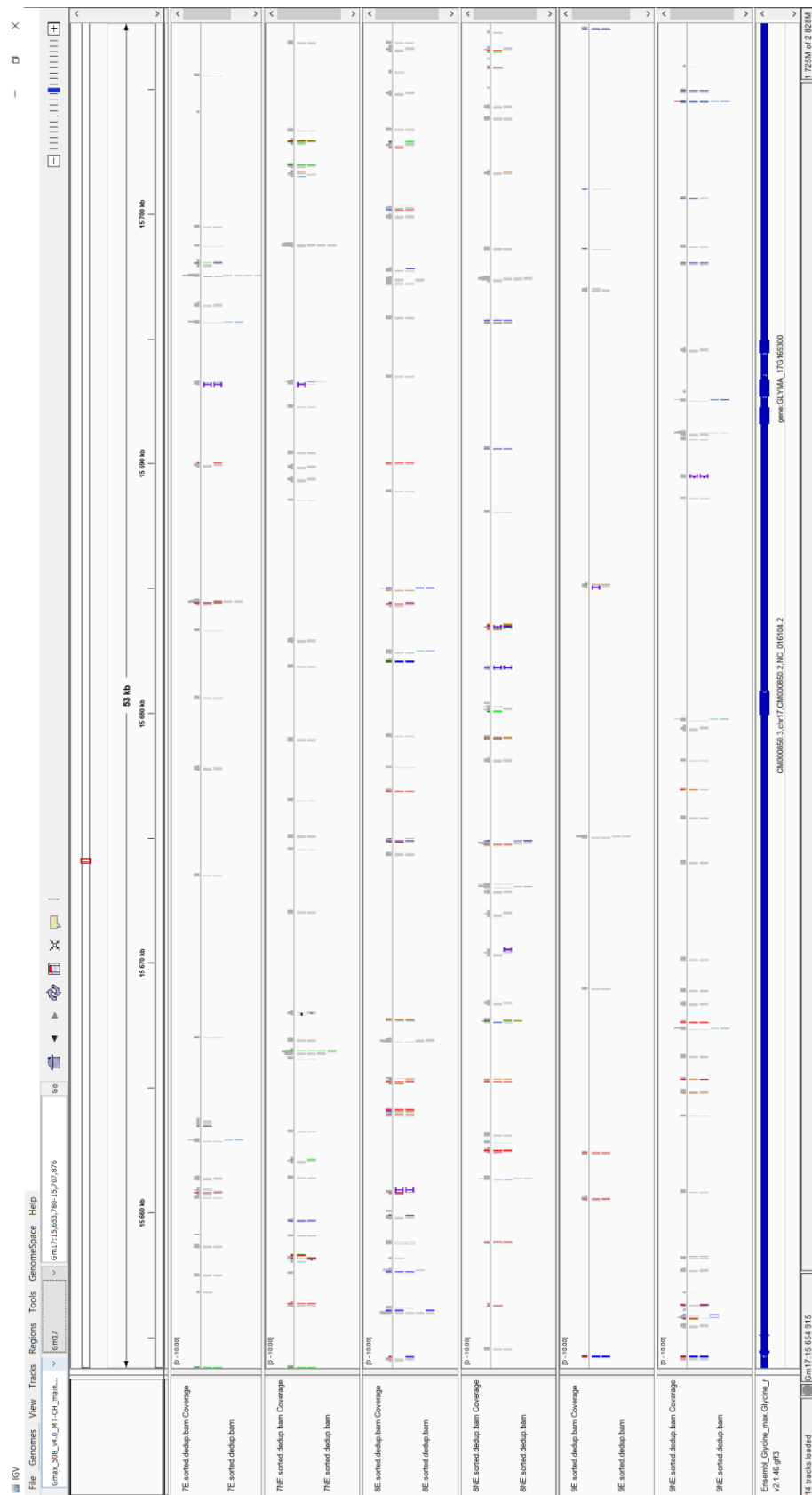


Figure III-17 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM17

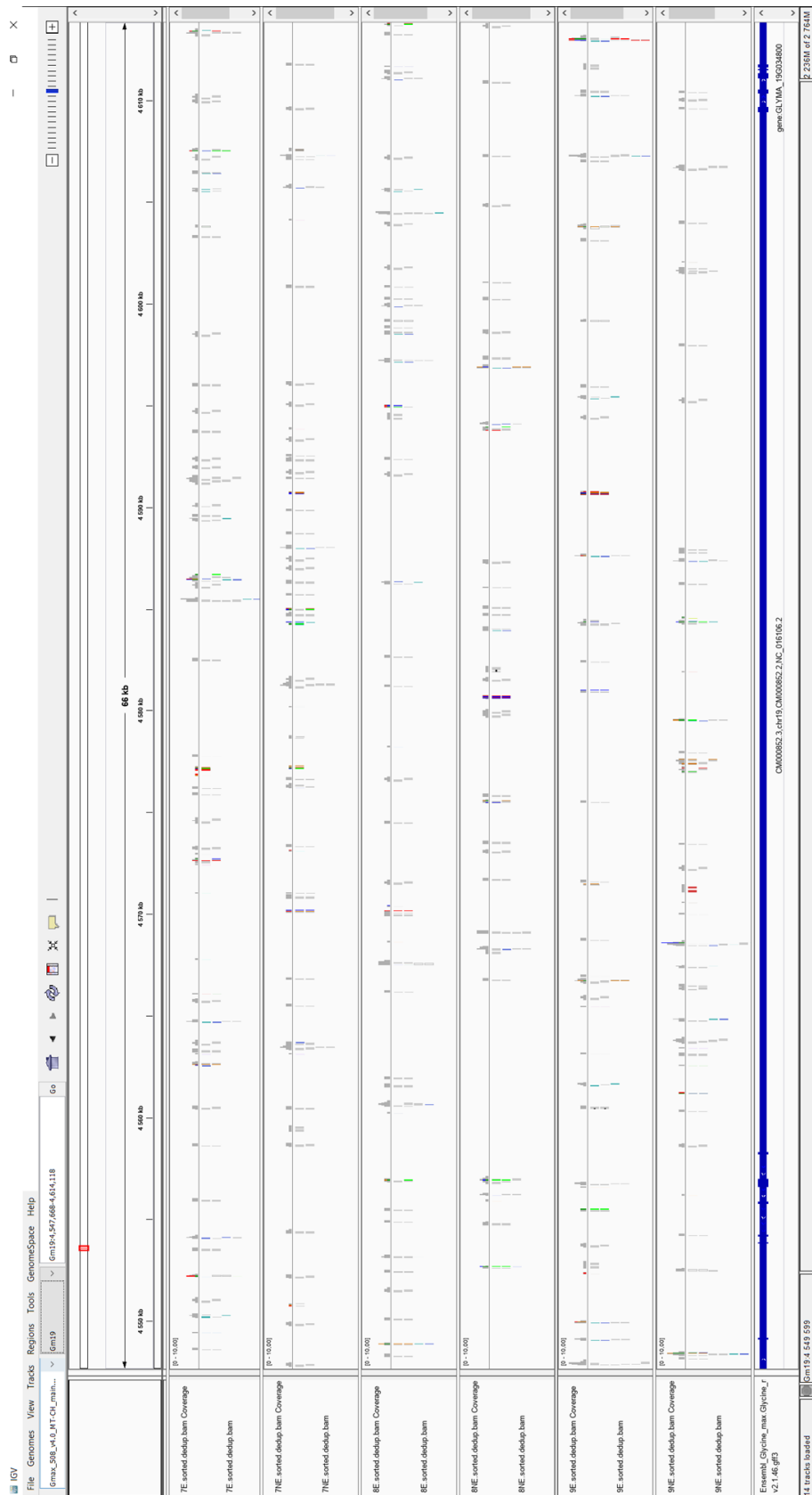


Figure III-18 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM19

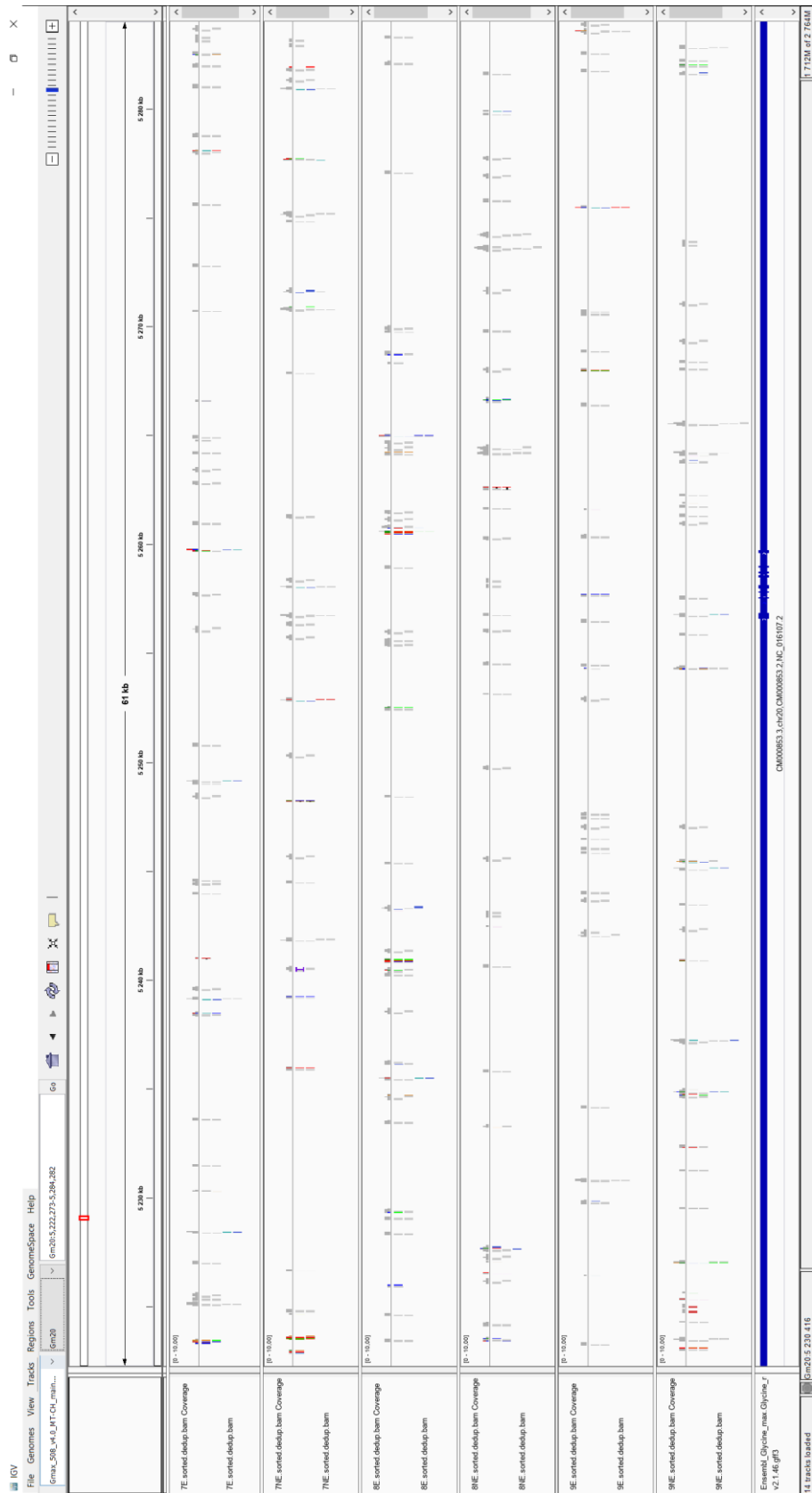


Figure III-19 : Visualisation IGV de l'alignement des *reads* des différents échantillons d'ADNex sur le chromosome GM20

II-3 Graphiques d'alignement des *reads* sur les chromosomes et les organites (réplicat 7)

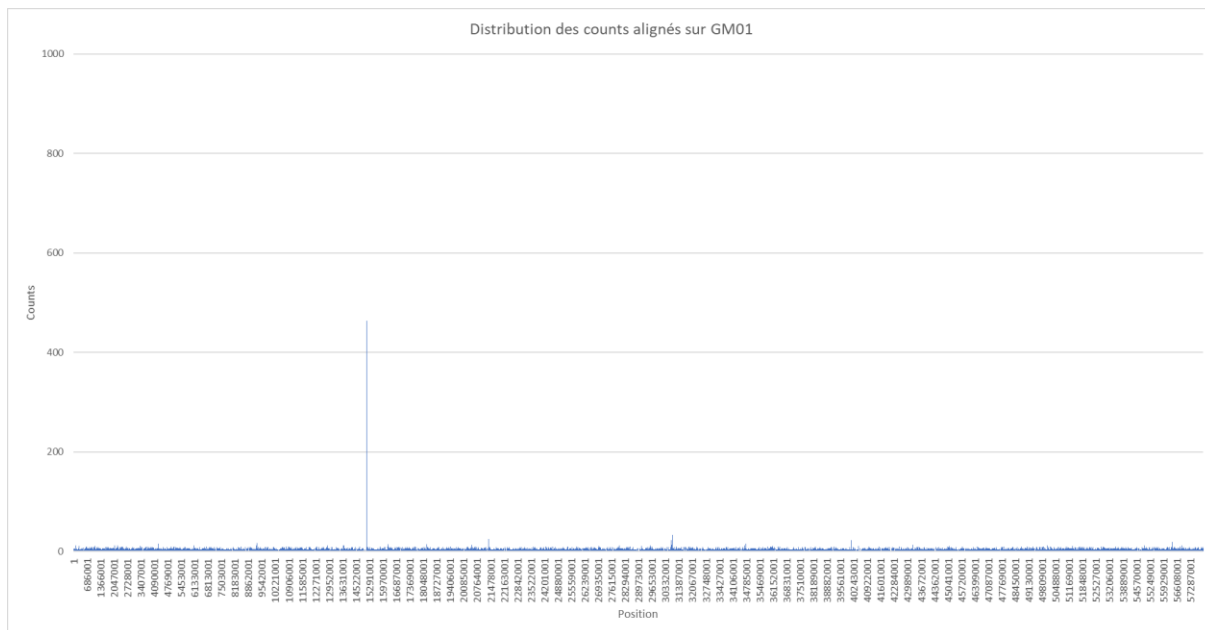


Figure III-20 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM01 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 1, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

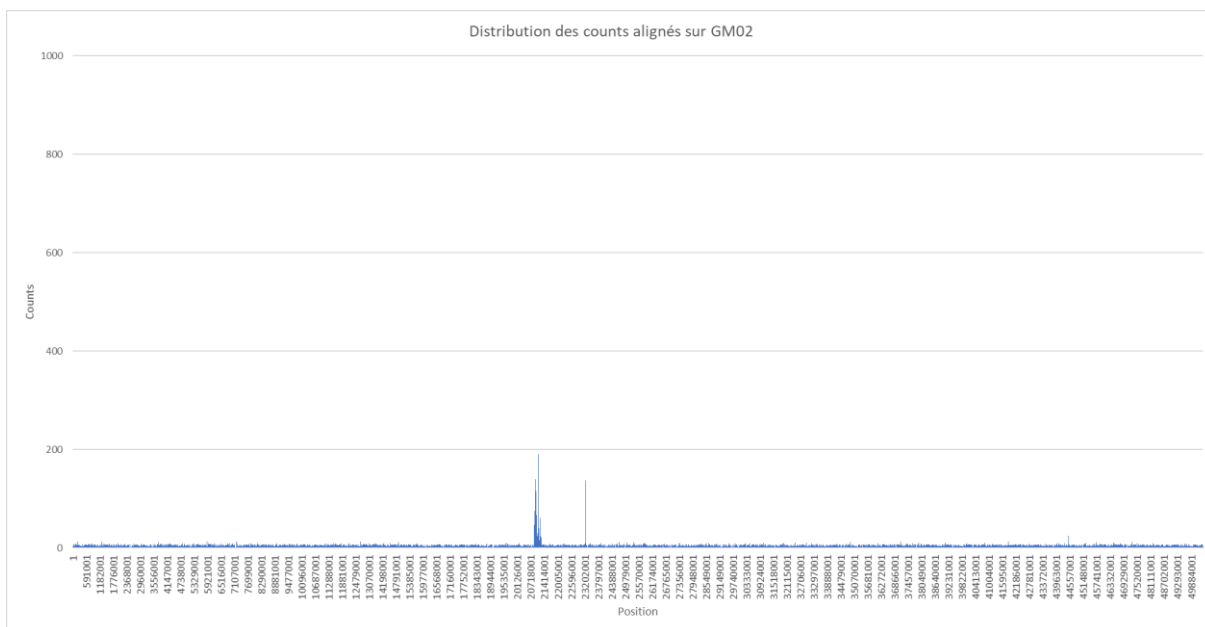


Figure III-21 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM02 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 2, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

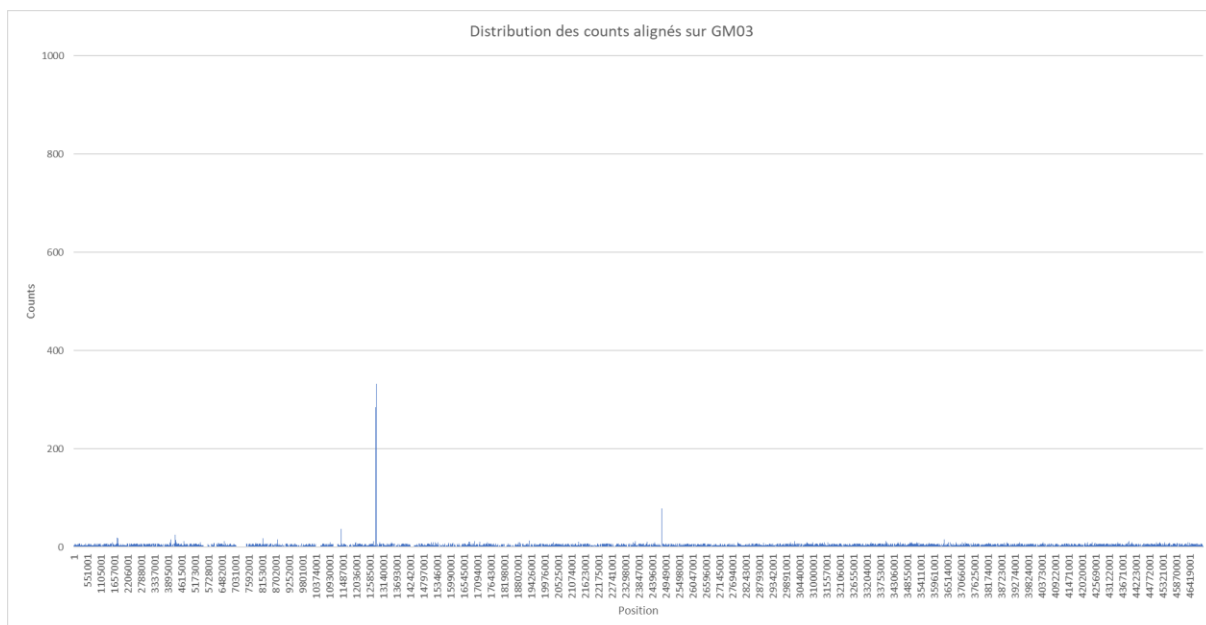


Figure III-22 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM03 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 3, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

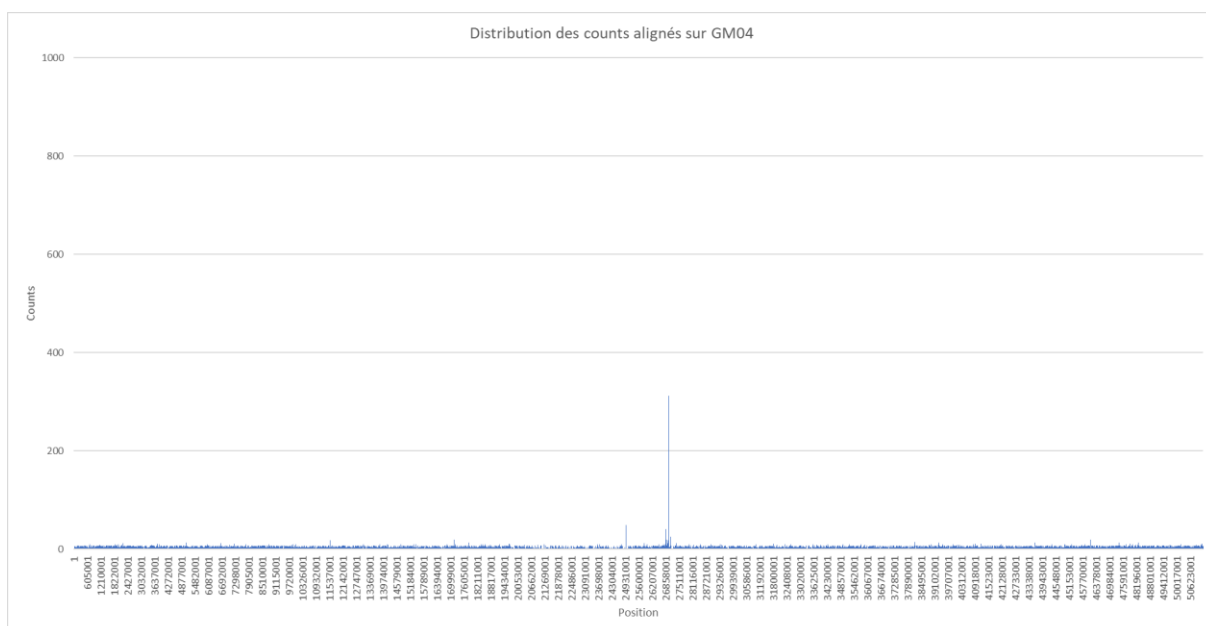


Figure III-23 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM04 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 4, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

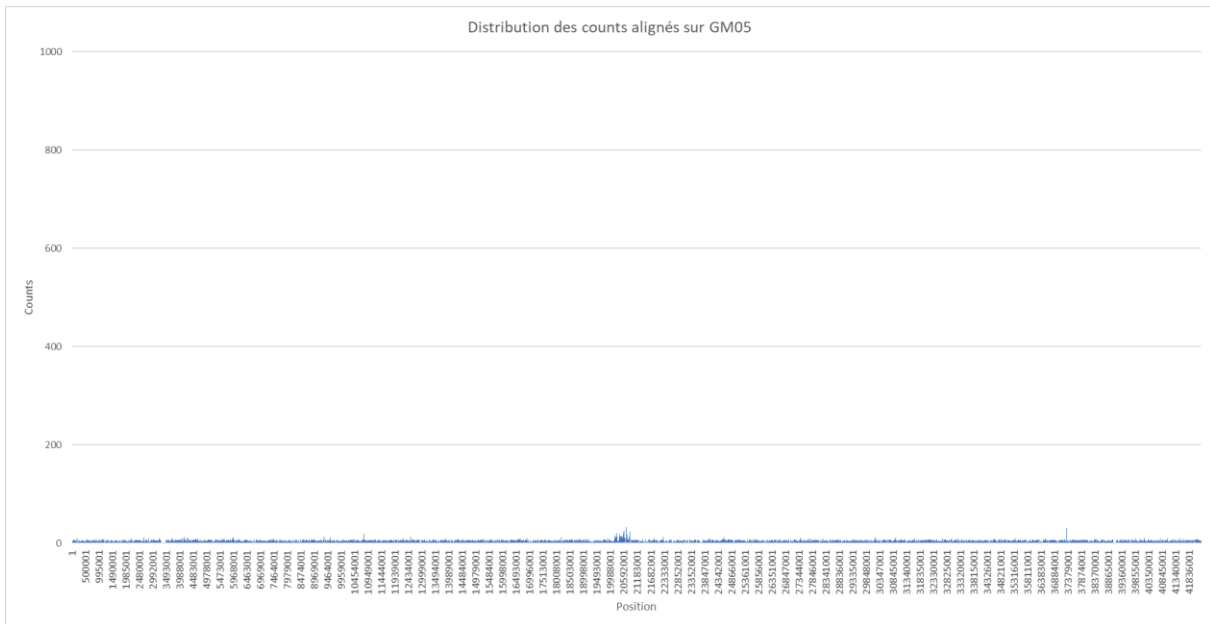


Figure III-24 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM05 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 5, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

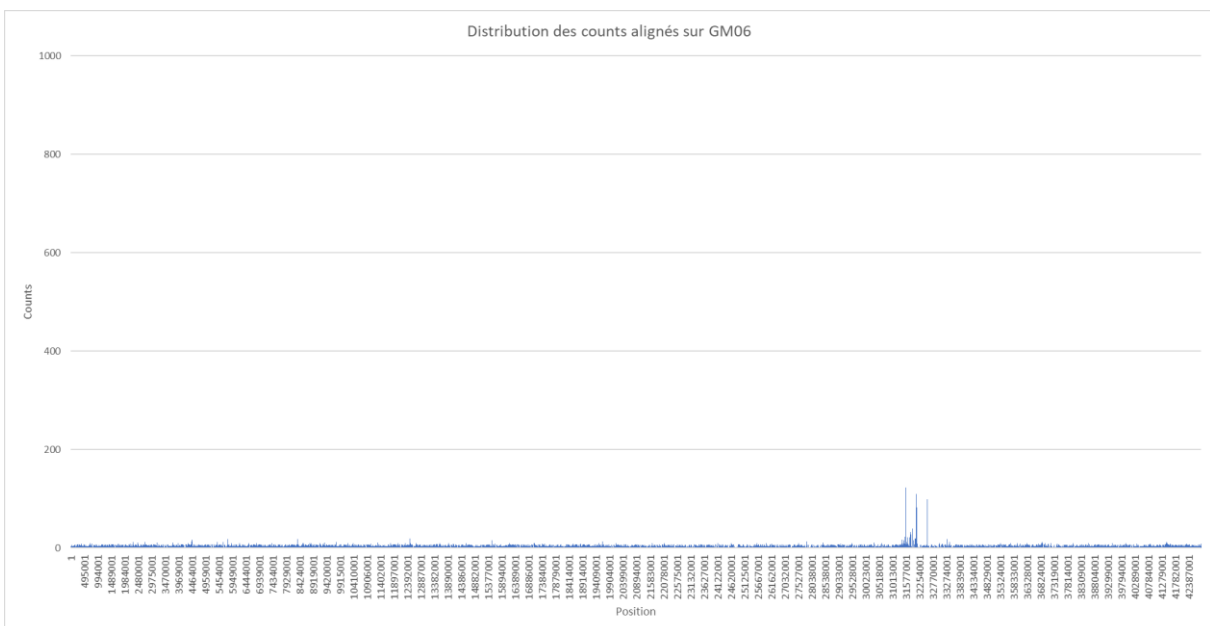
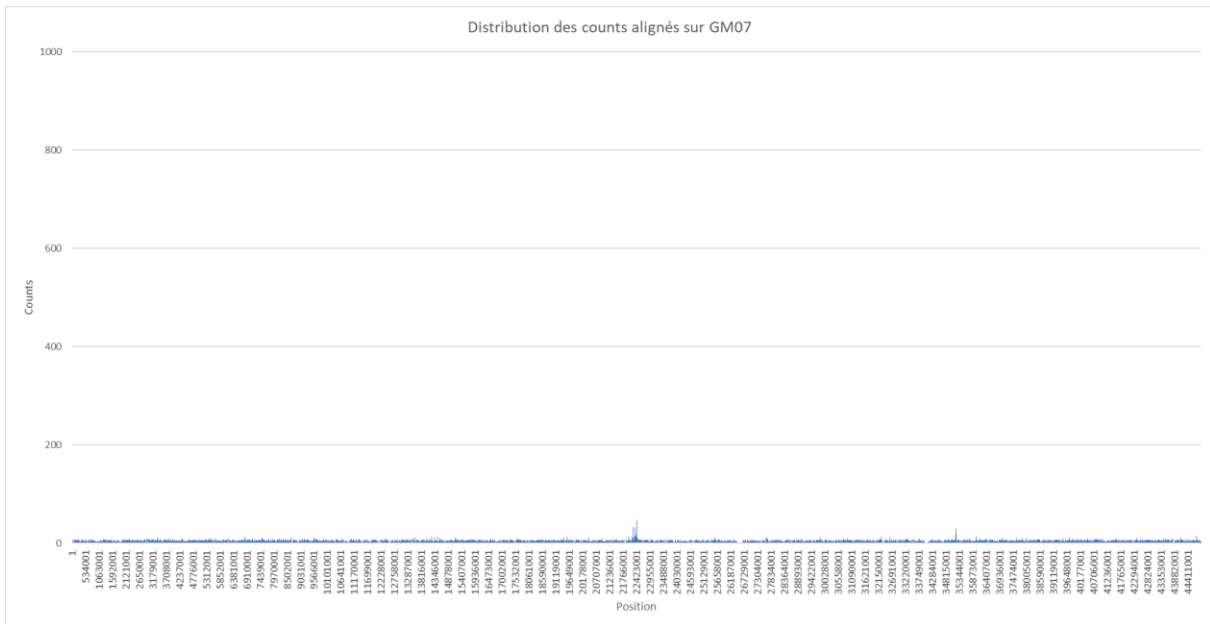


Figure III-25 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM06 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 6, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).



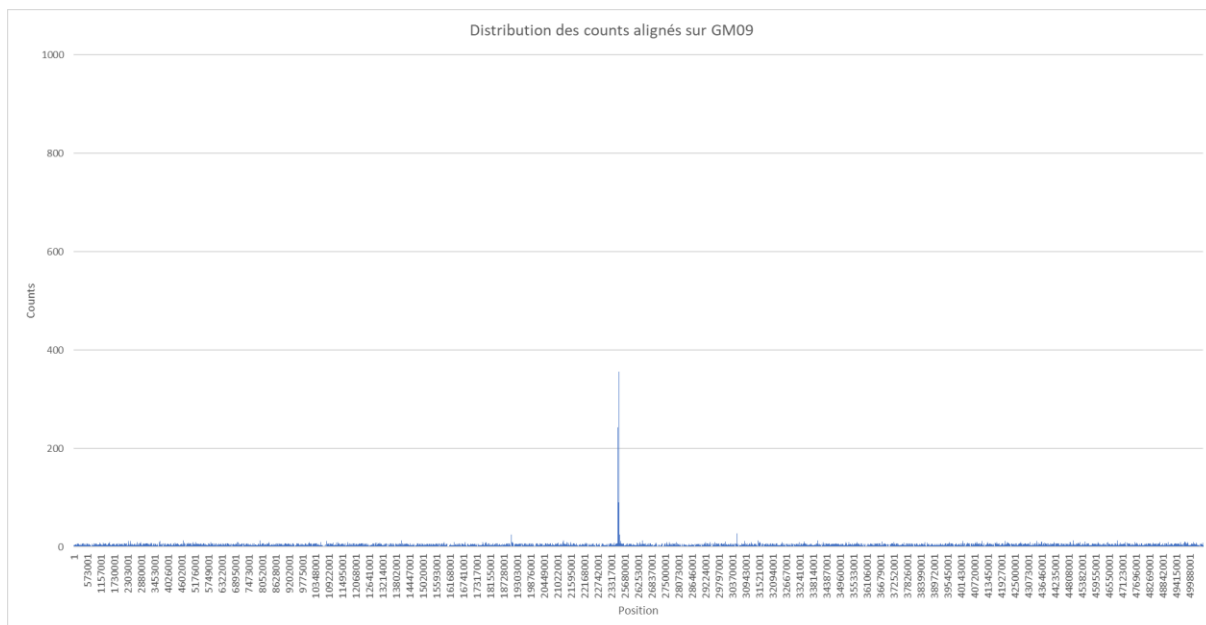


Figure III-28 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM09 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 9, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

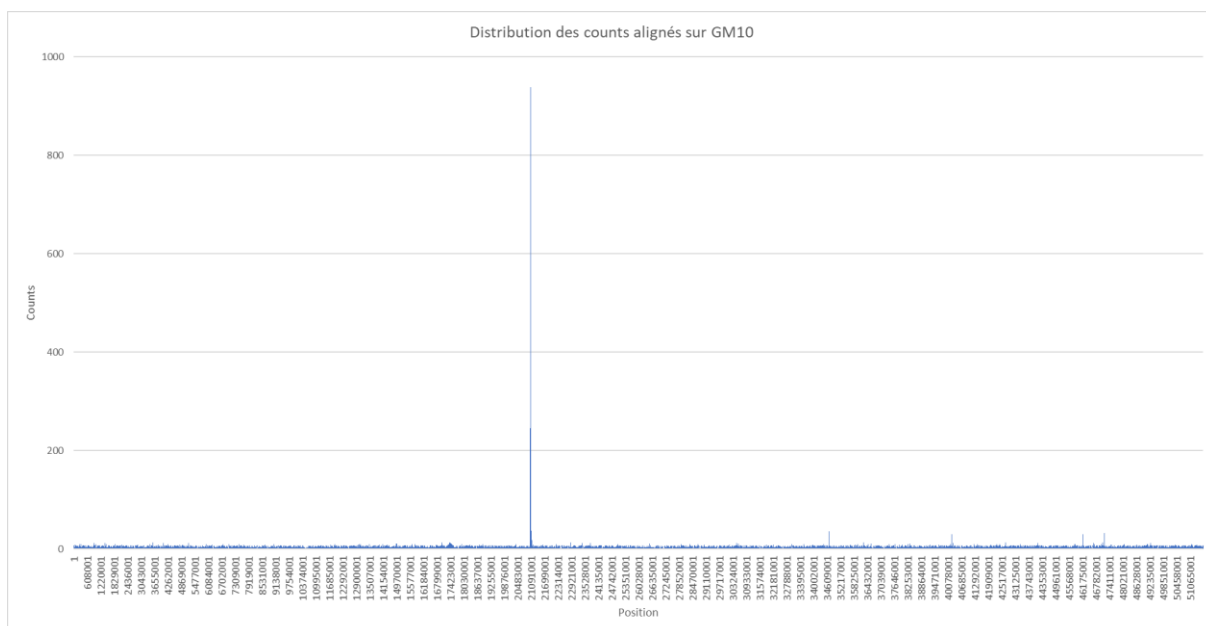


Figure III-29 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM10 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 10, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

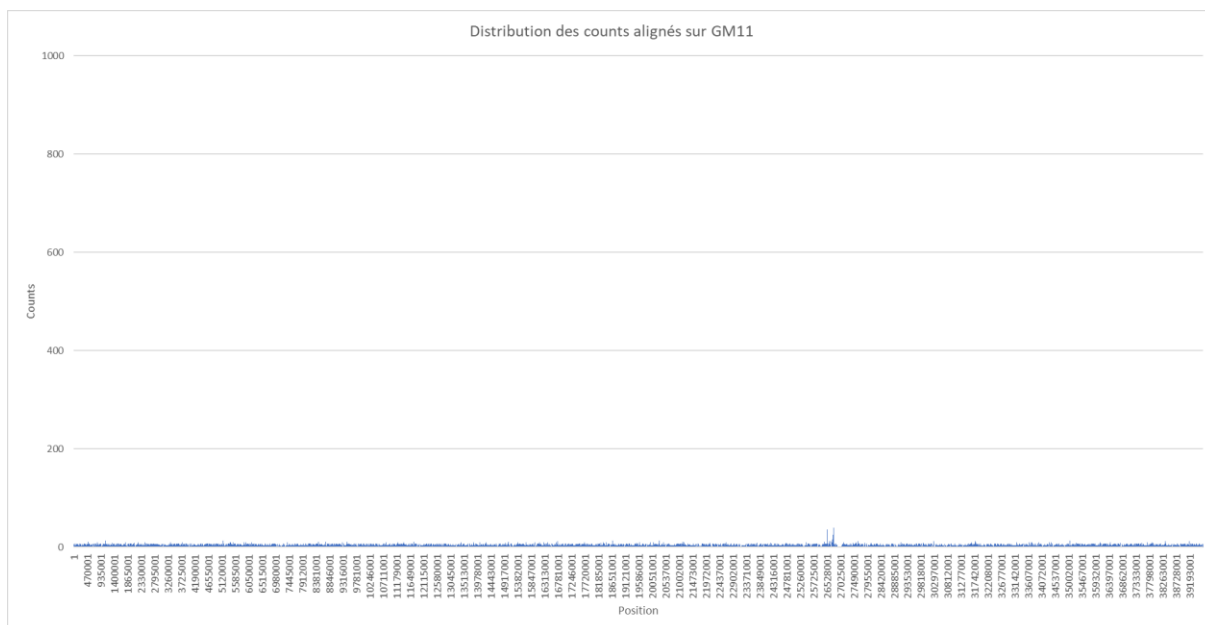


Figure III-30 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM11 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 11, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

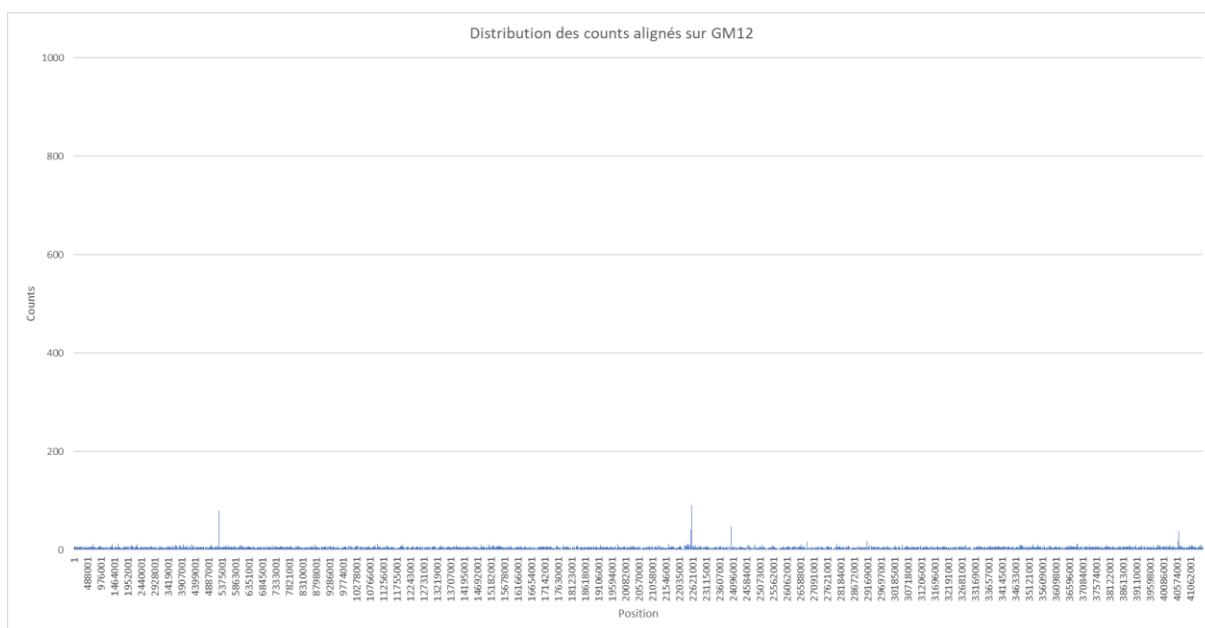


Figure III-31 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM12 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 12, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

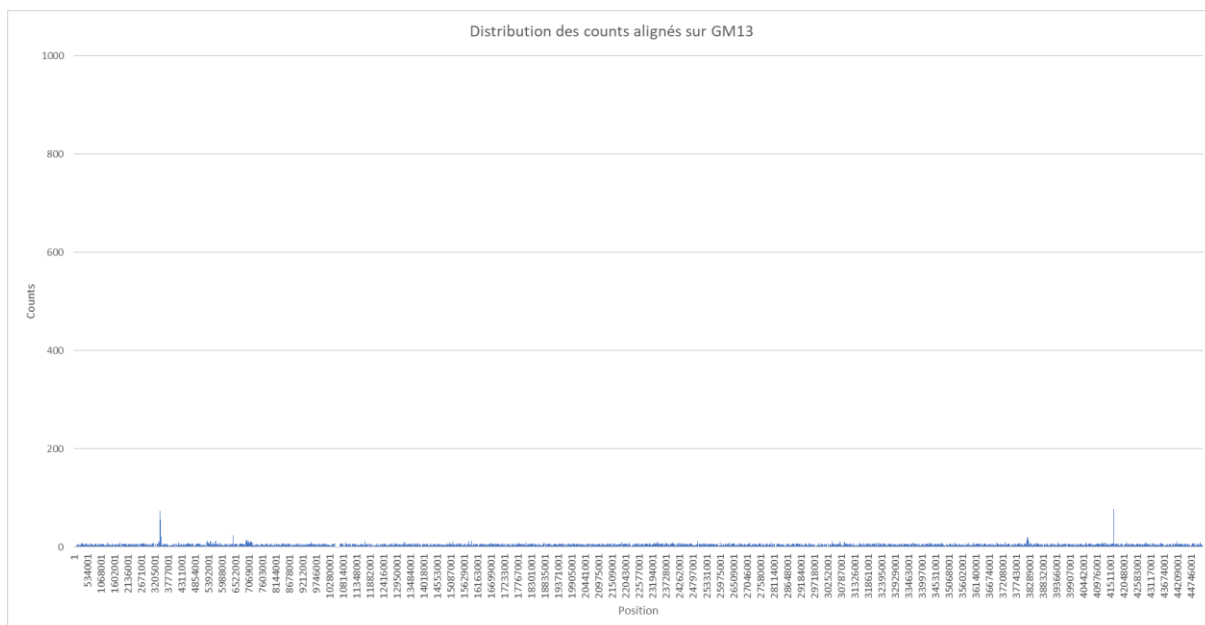


Figure III-31 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM13 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 13, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

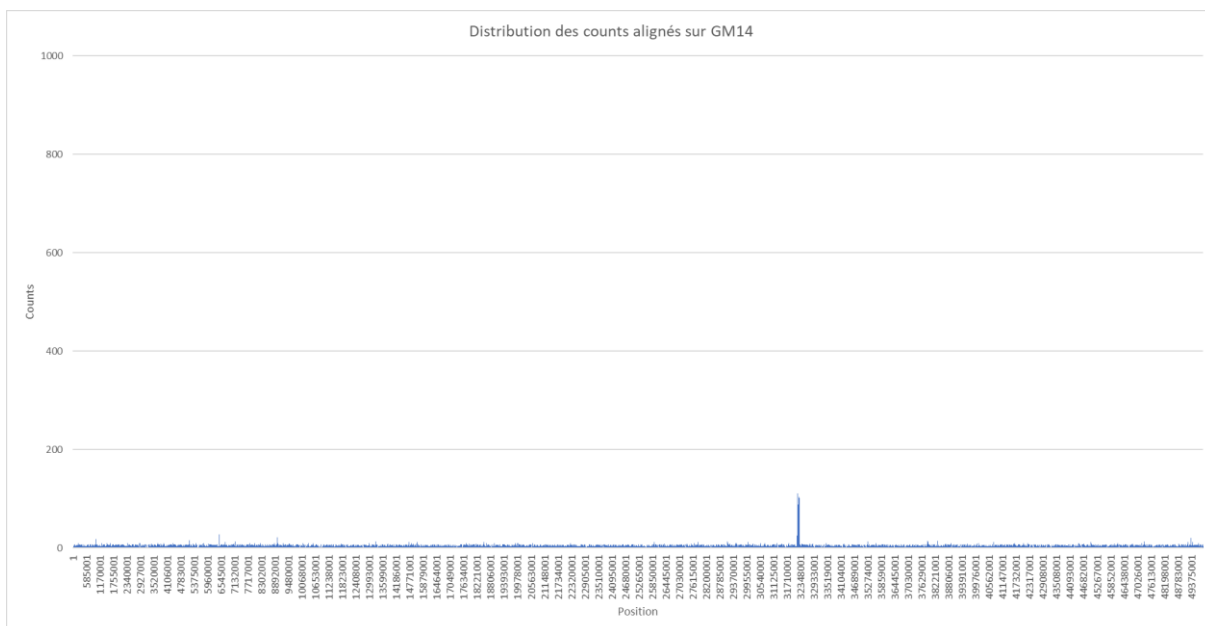


Figure III-32 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM14 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 14, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

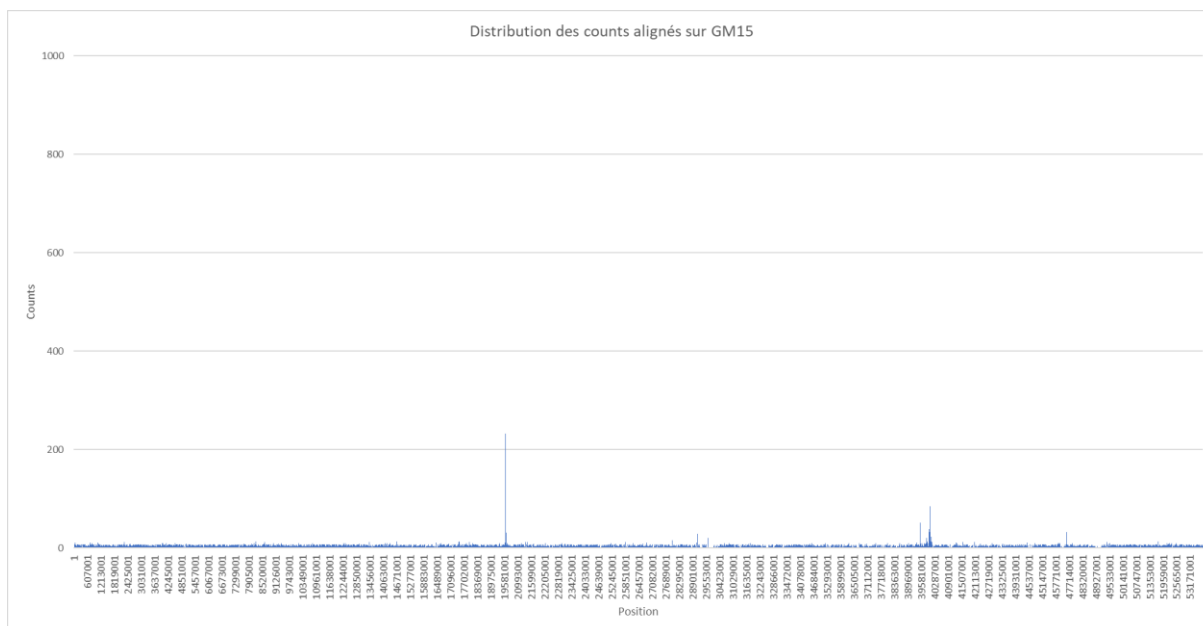


Figure III-34 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM15 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 15, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

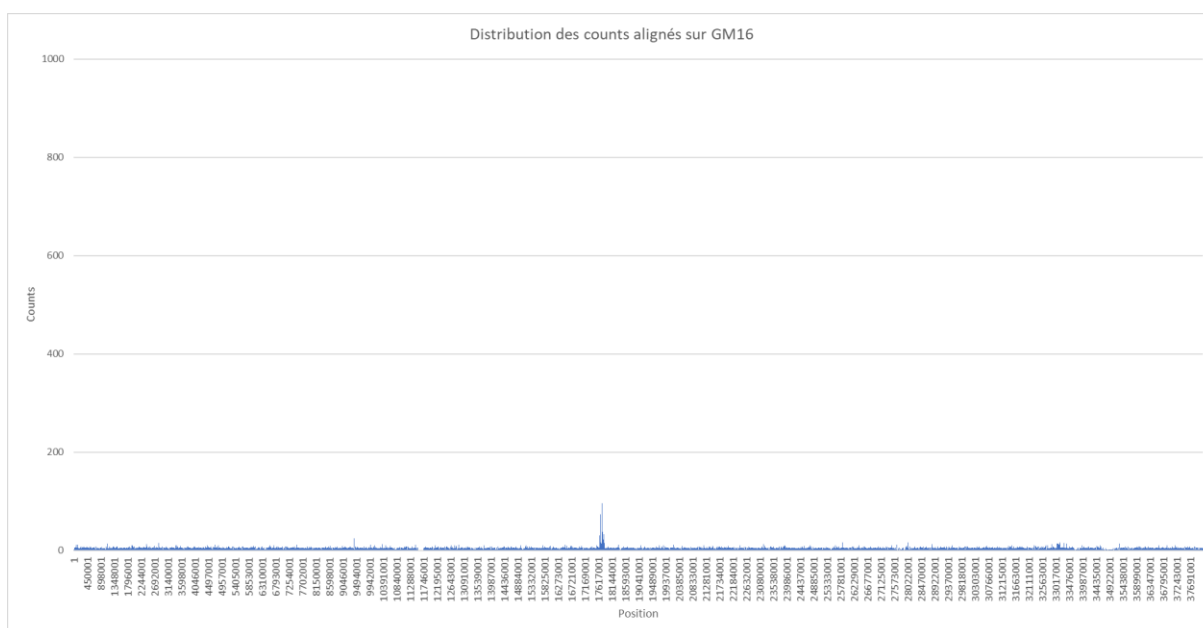


Figure III-35 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM16 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 16, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

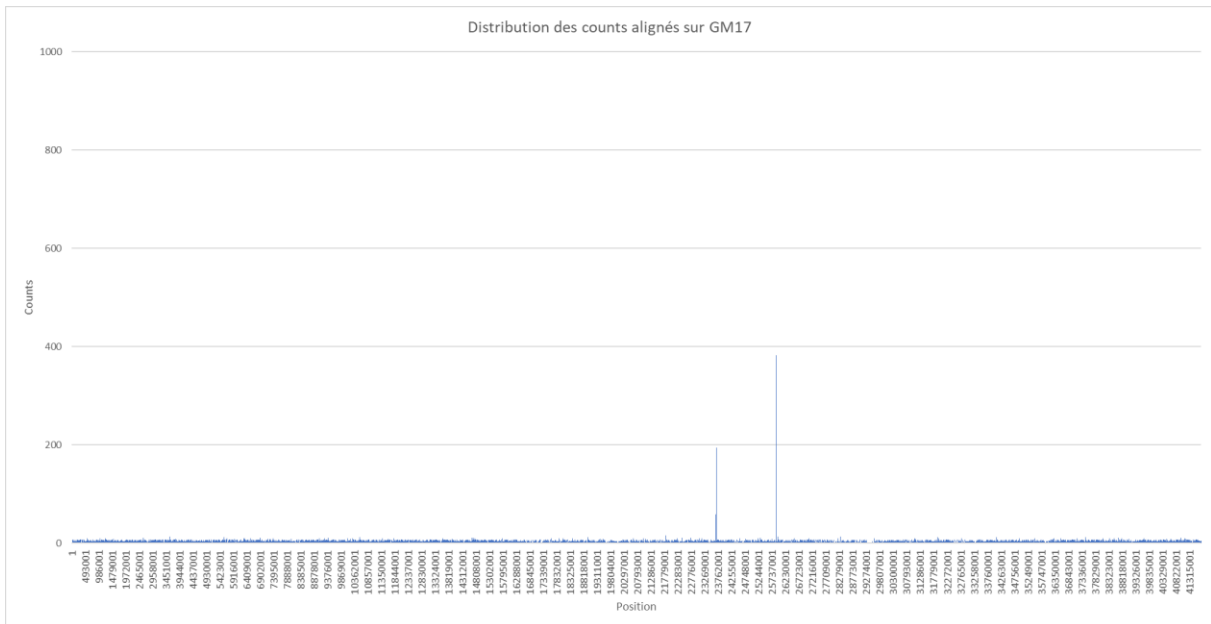


Figure III-36 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM17 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 17, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

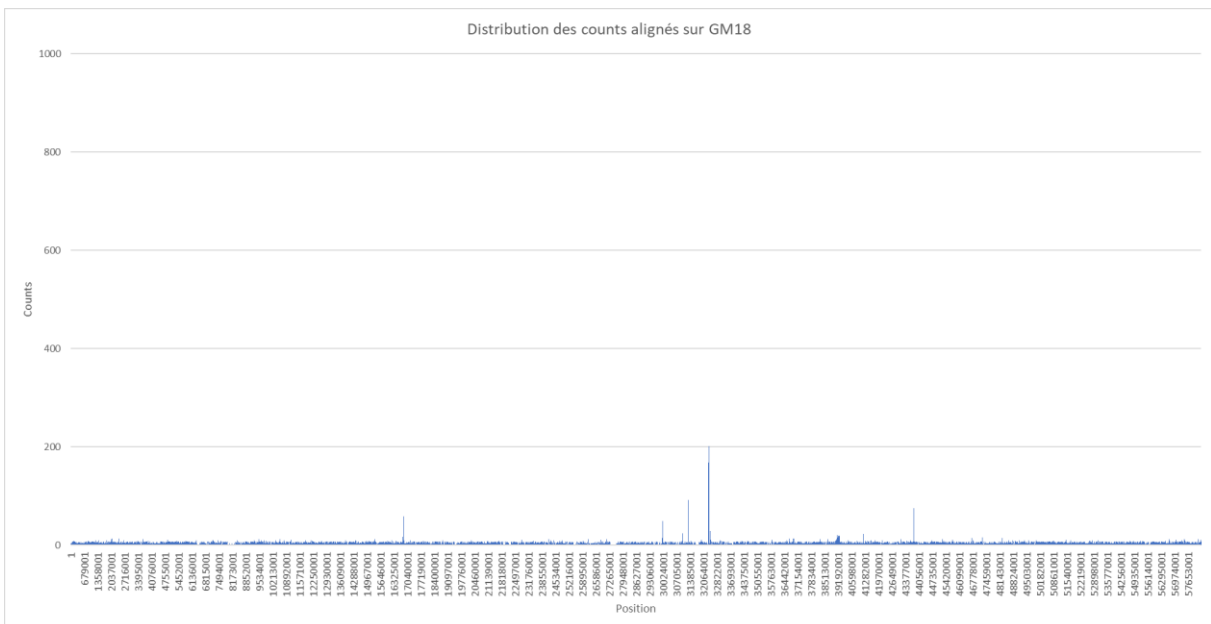


Figure III-37 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM18 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 18, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

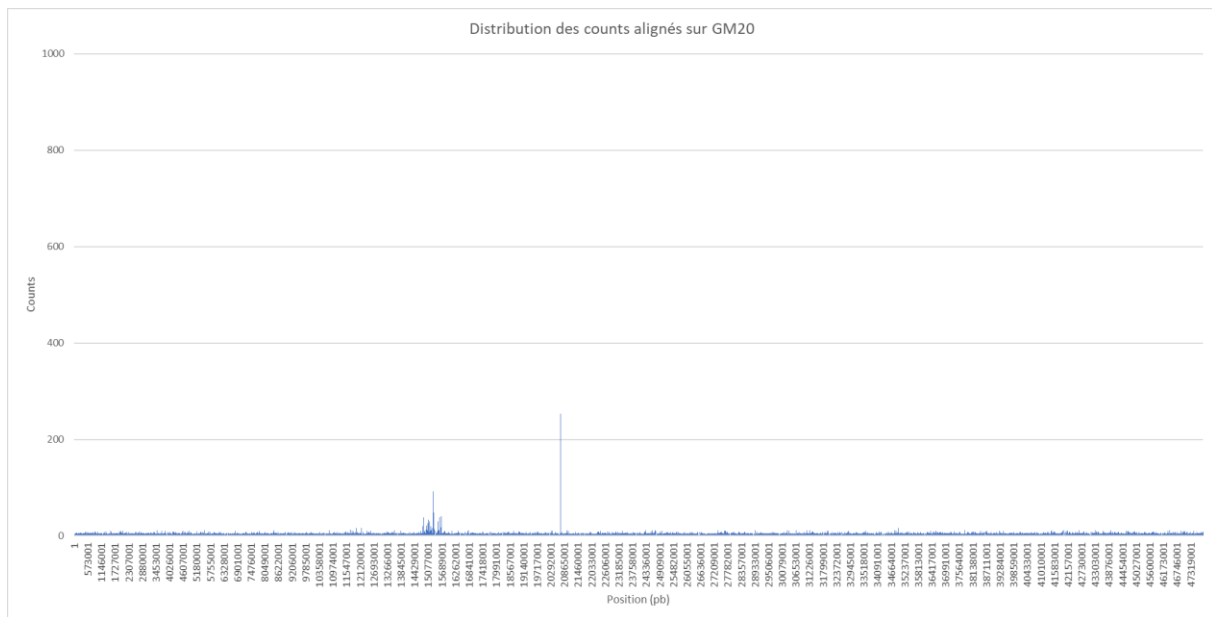


Figure III-38 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM20 en condition non élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 20, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

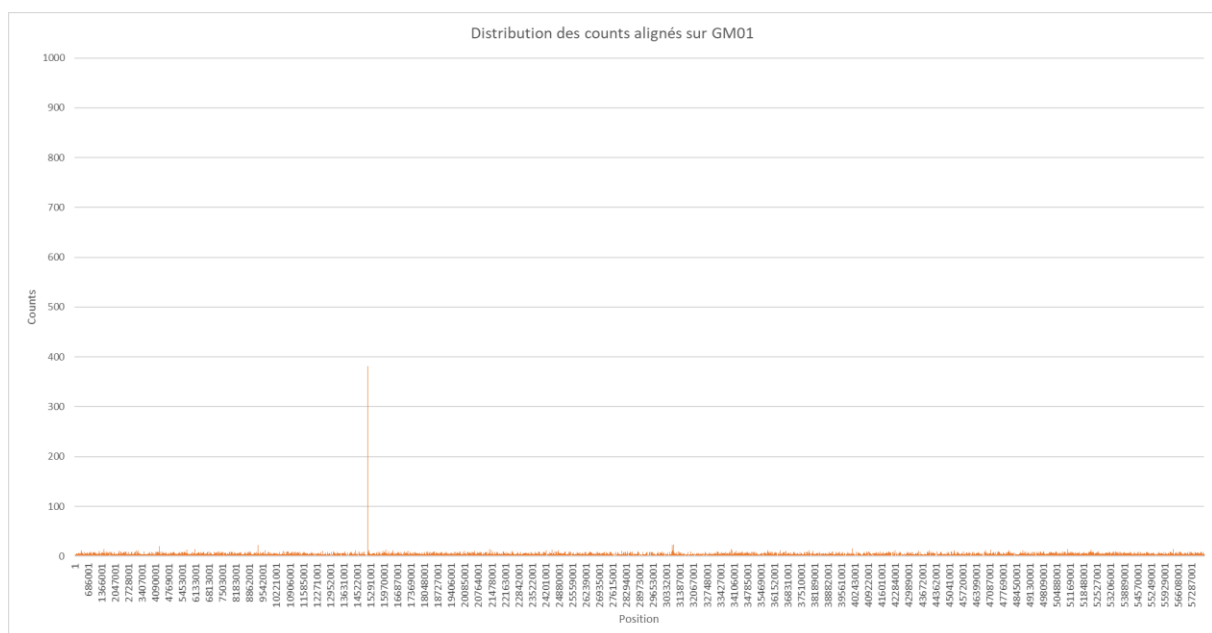


Figure III-39 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM01 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 1, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

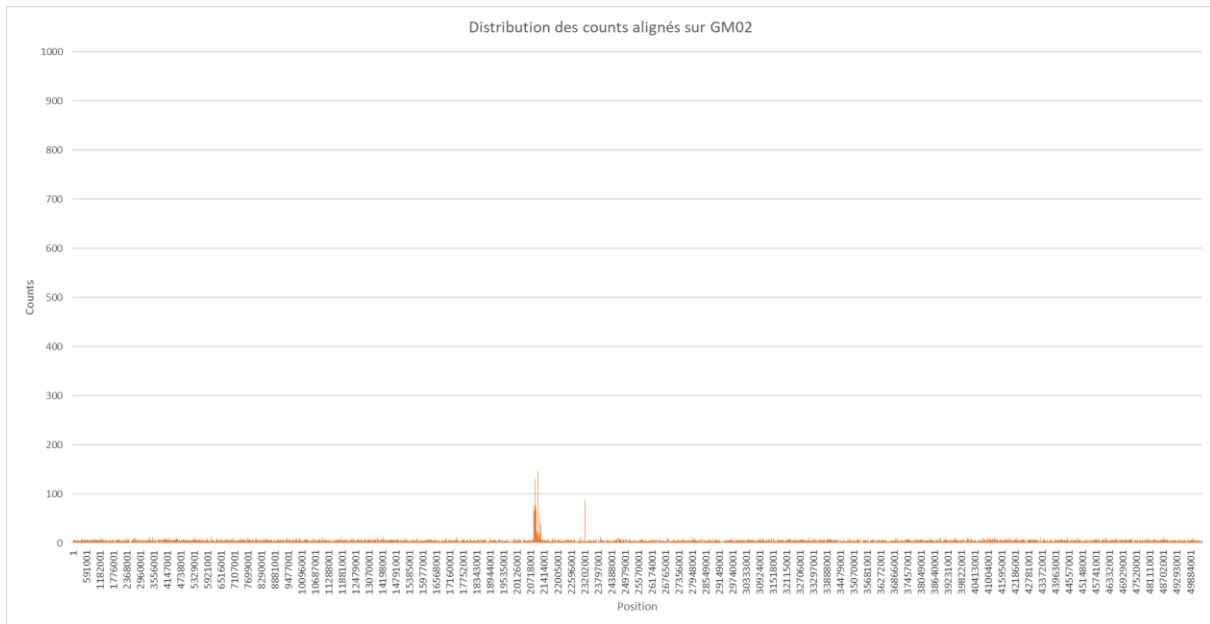


Figure III-40 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM02 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 2, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

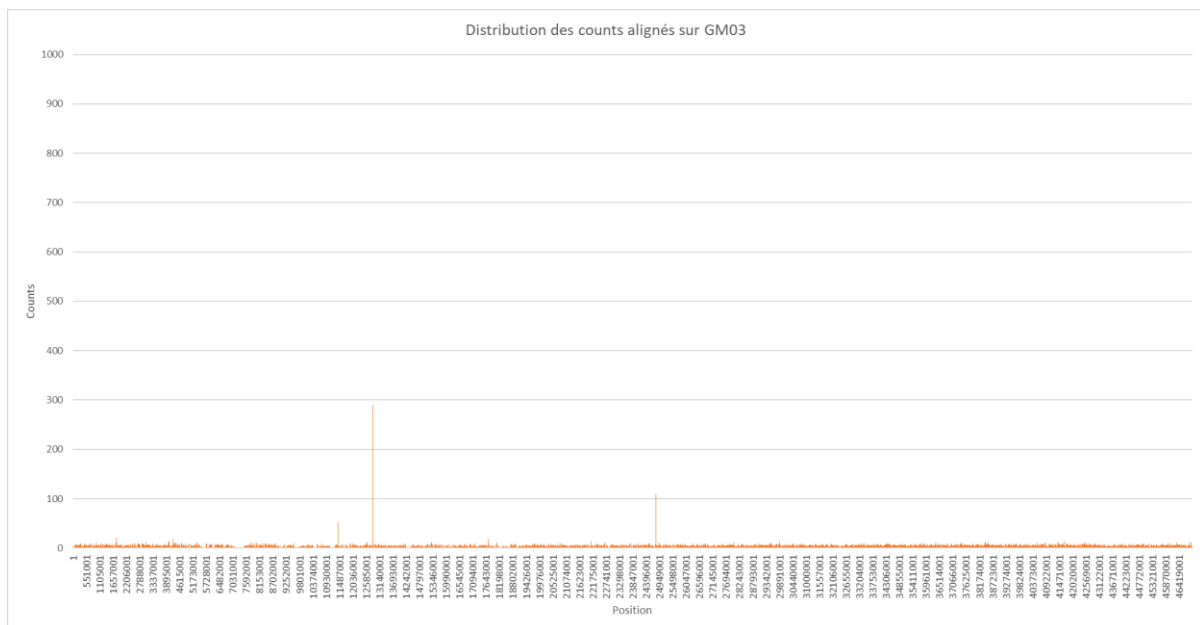


Figure III-41 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM03 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 3, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

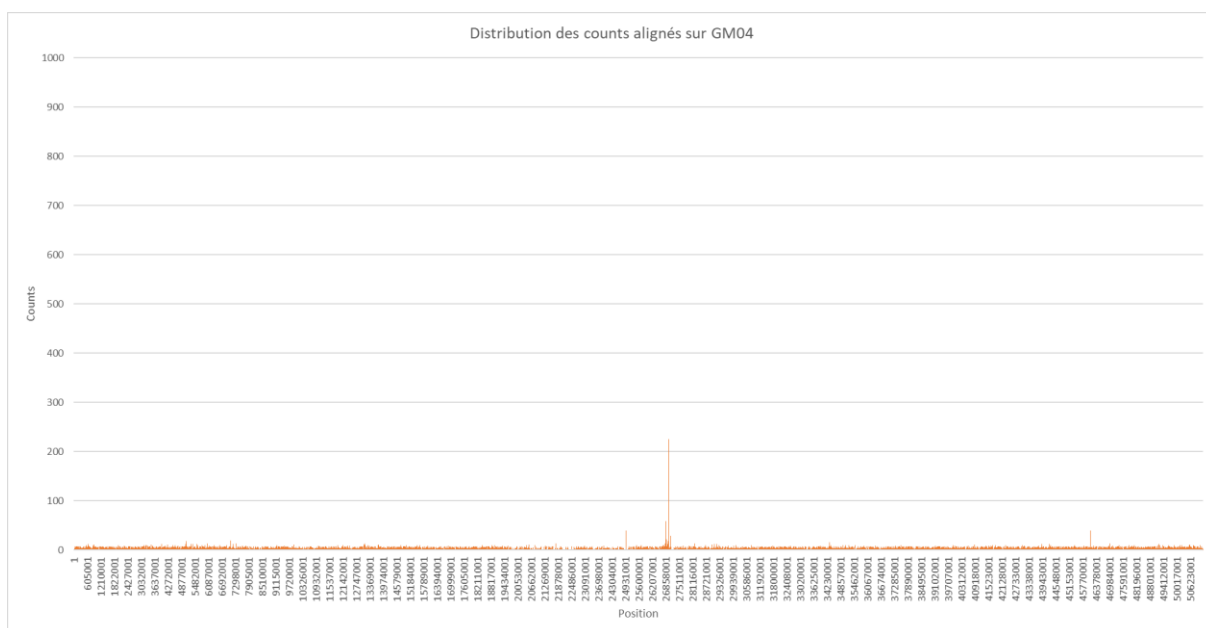


Figure III-42 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM04 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 4, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

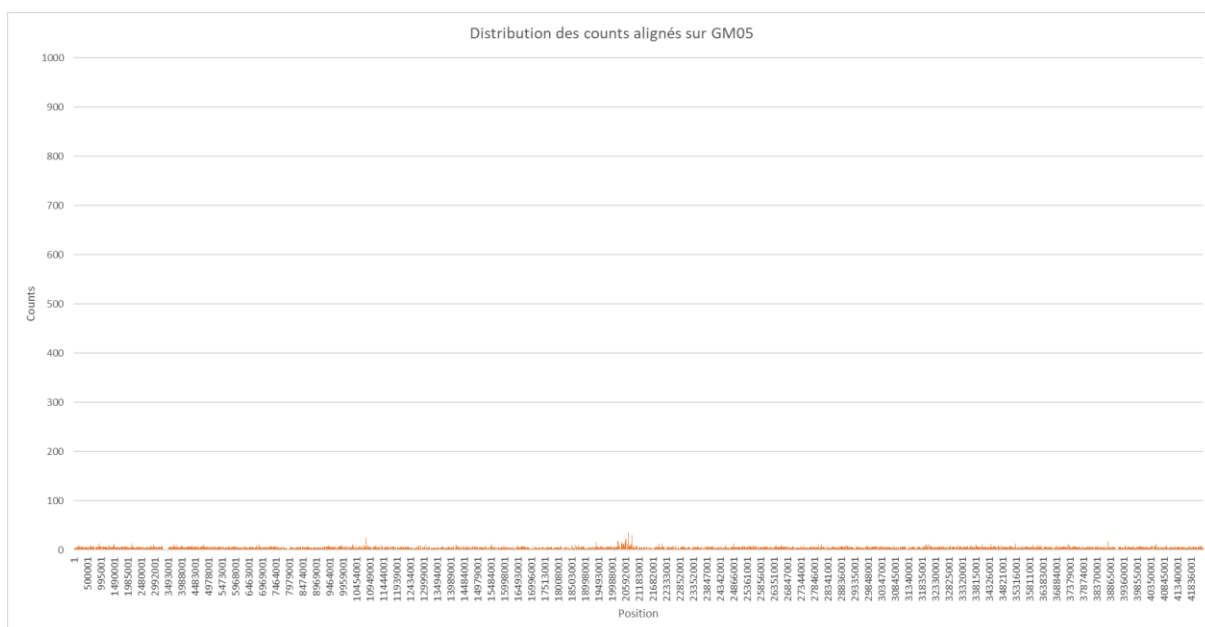


Figure III-42 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM05 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 5, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

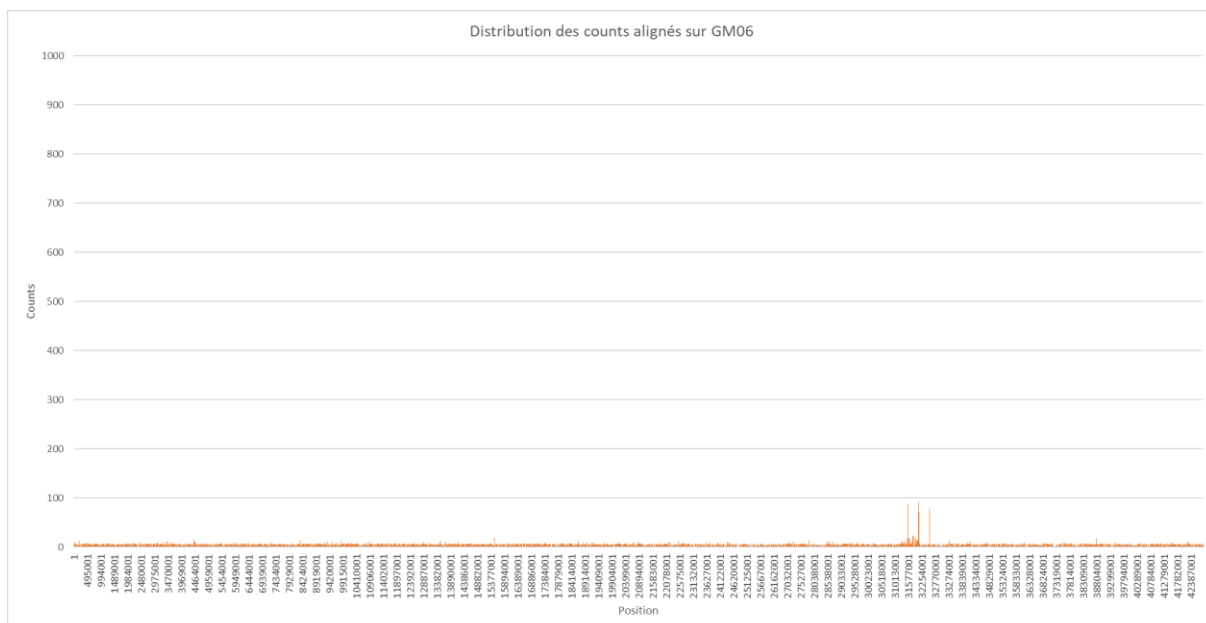


Figure III-43 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM06 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 6, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

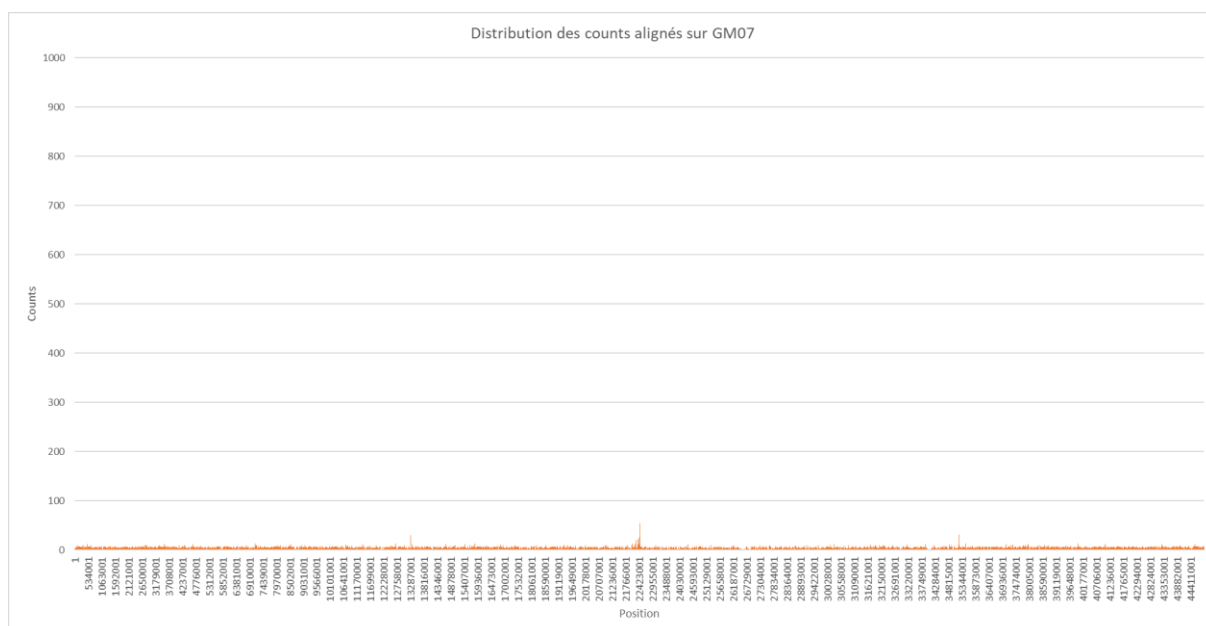


Figure III-44 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM07 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 7, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

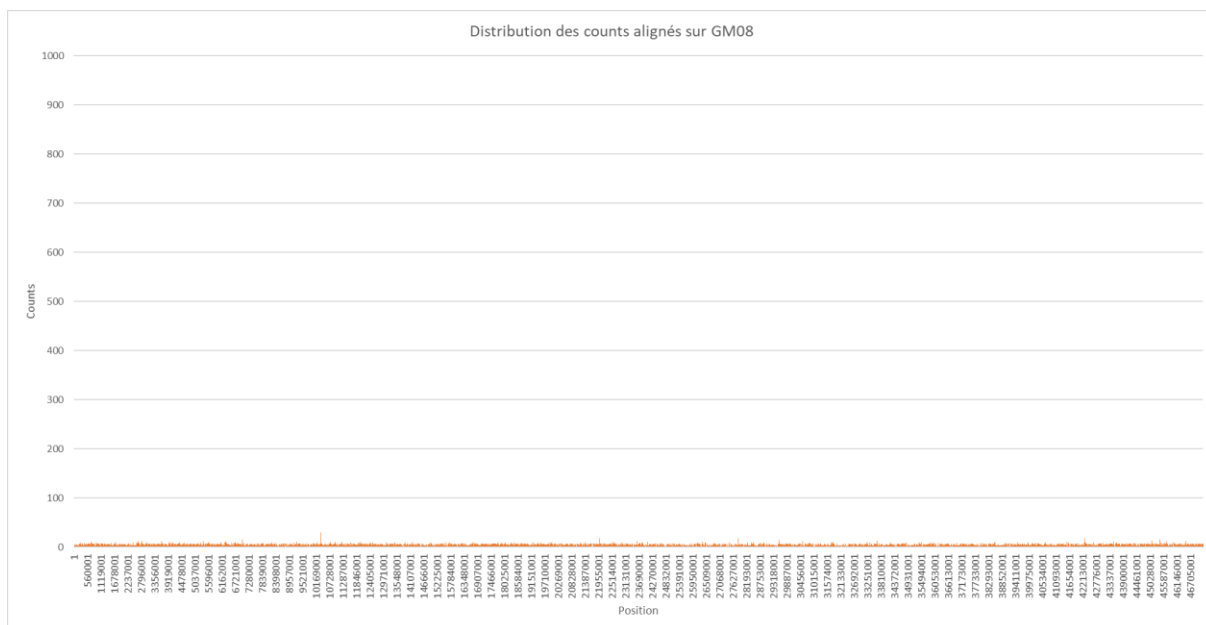


Figure III-45 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM08 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 8, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

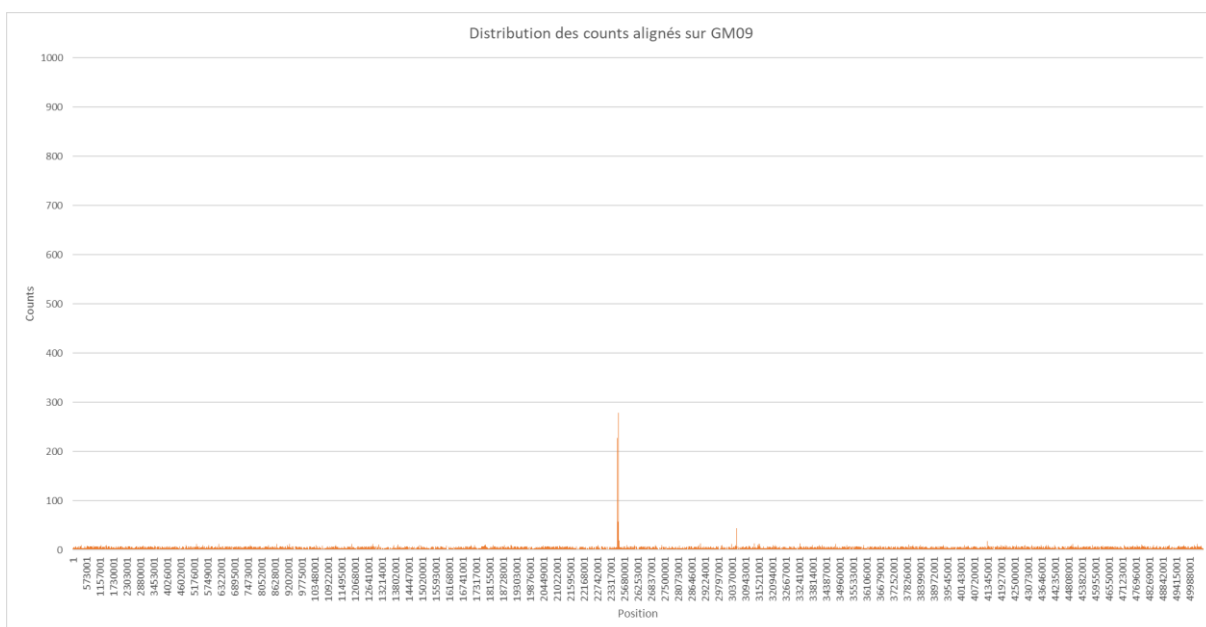


Figure III-46 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM09 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 9, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

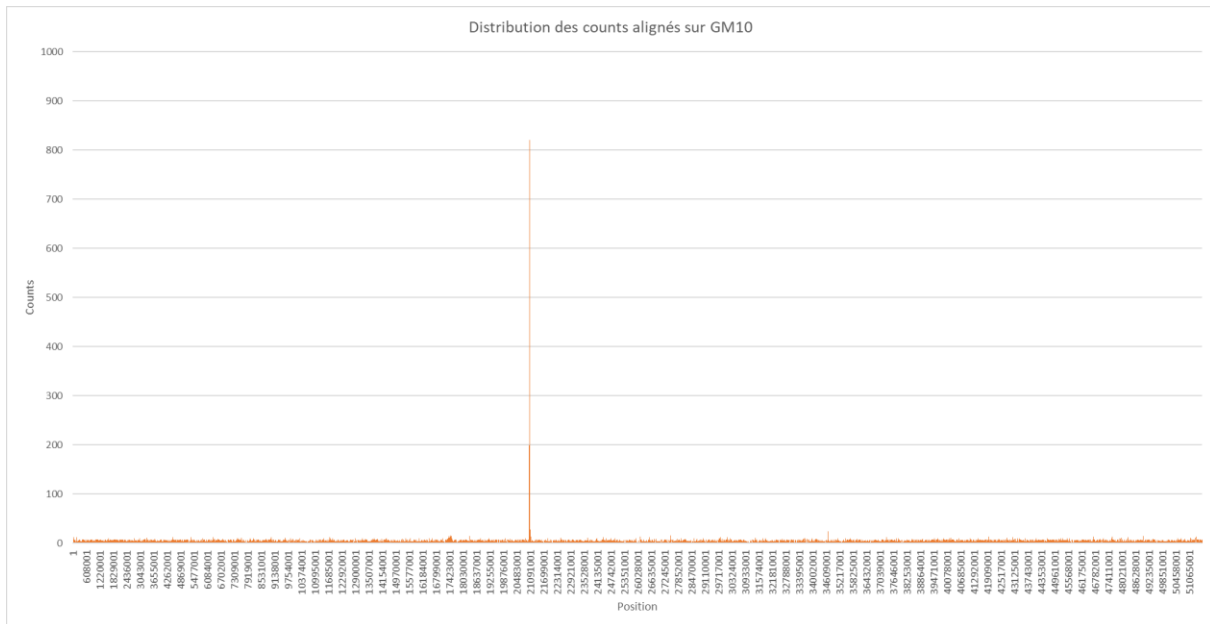


Figure III-47 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM10 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 10, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

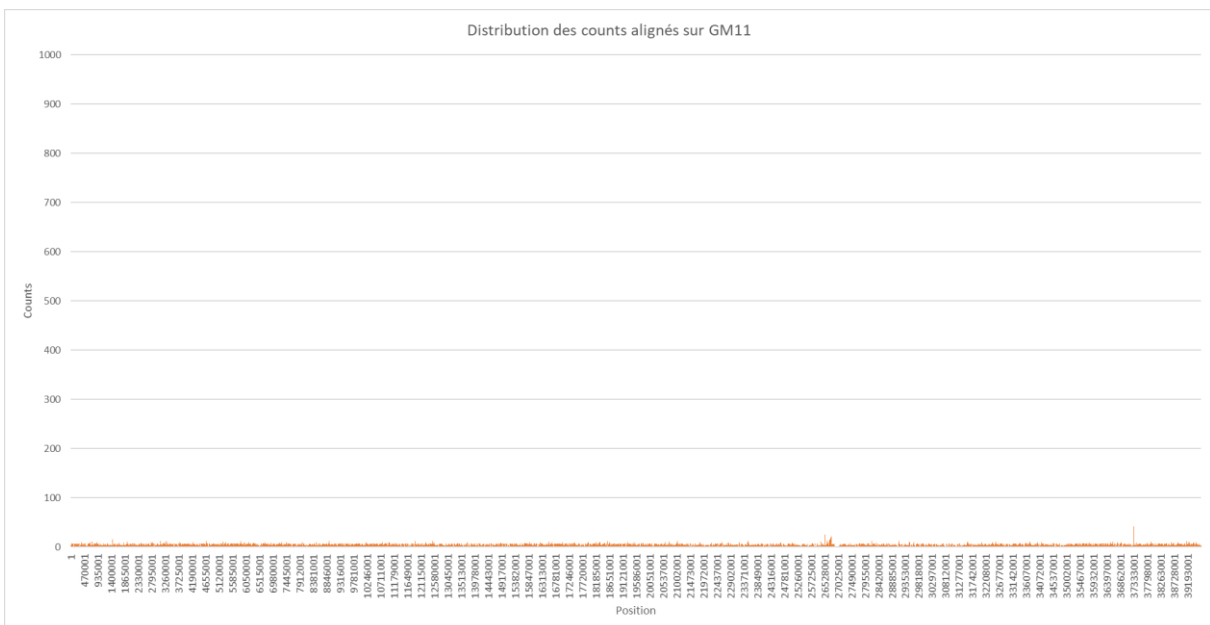


Figure III-48 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM11 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 11, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).



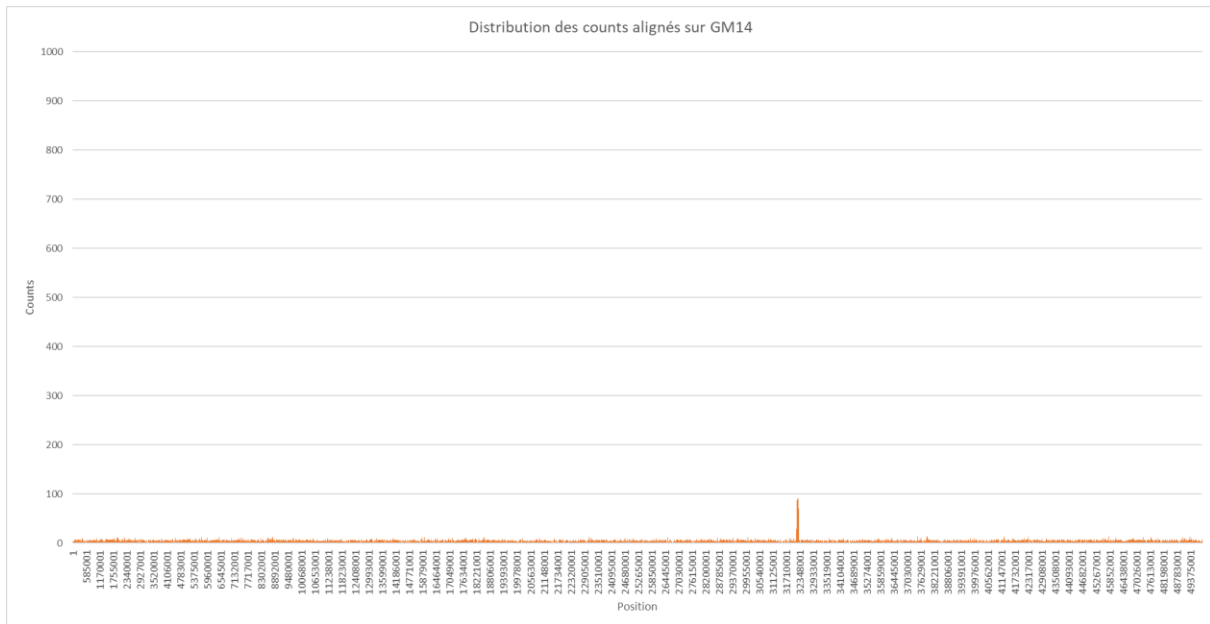


Figure III-51 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM14 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 14, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

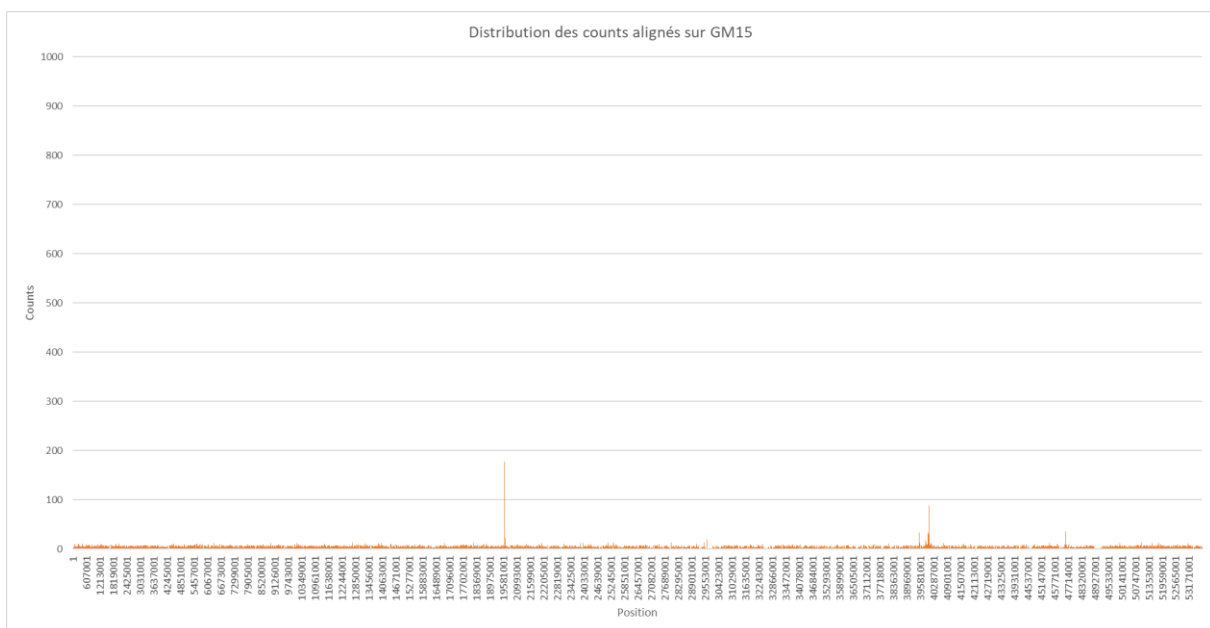


Figure III-52 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM15 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 15, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

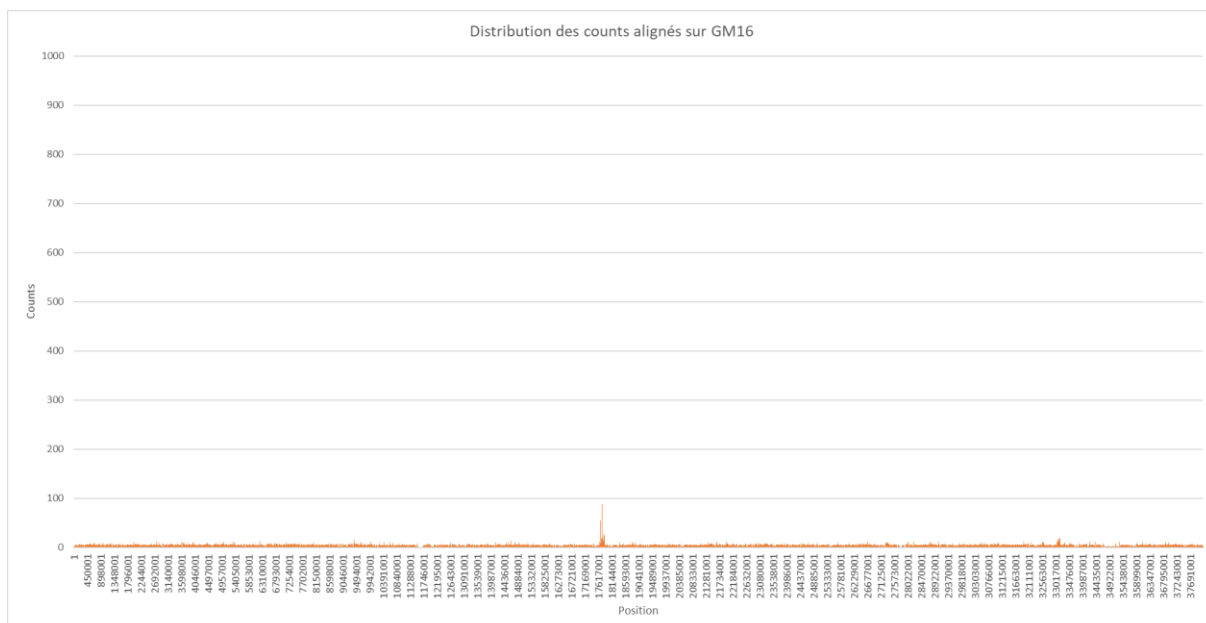


Figure III-53 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM16 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 16, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

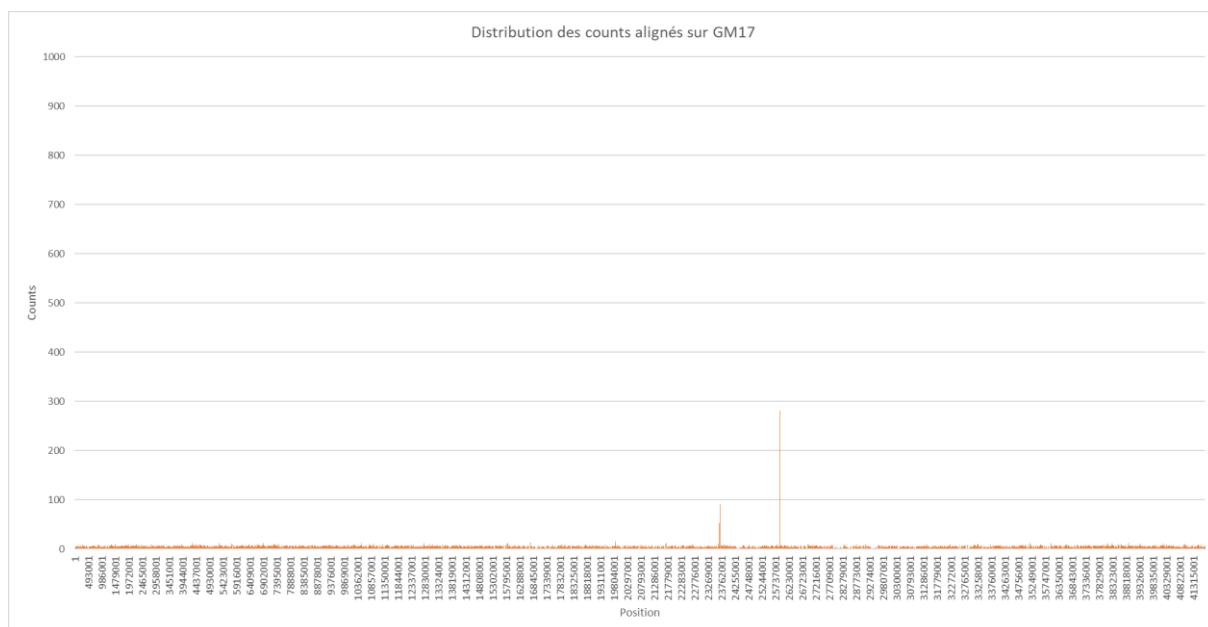


Figure III-54 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM17 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 17, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

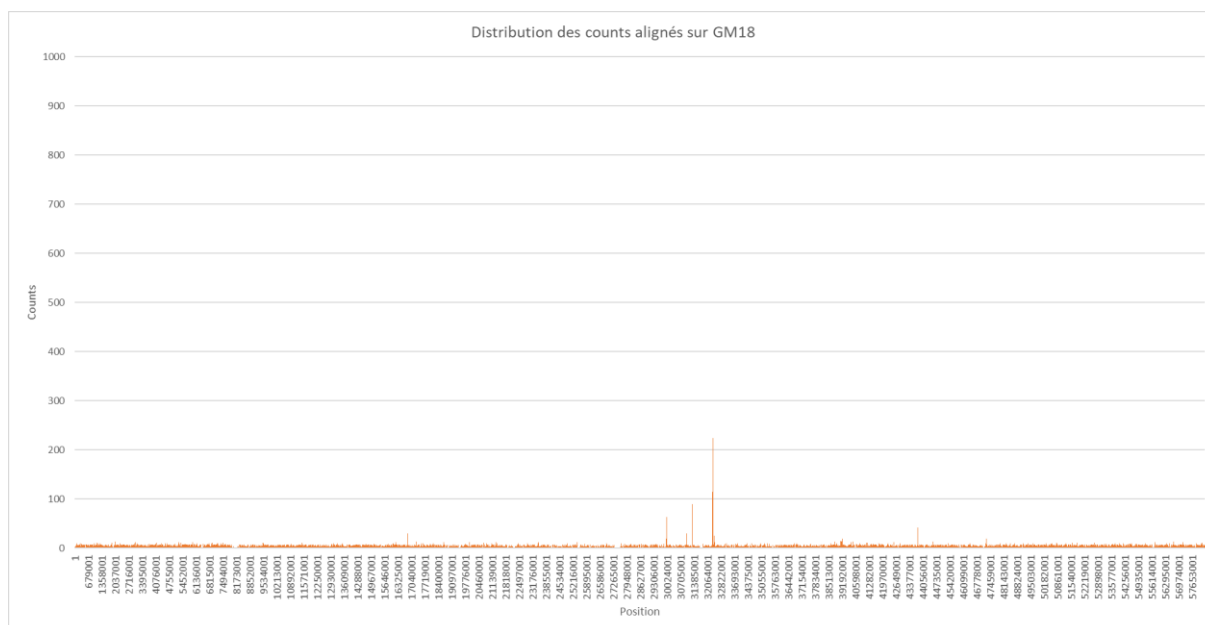


Figure III-55 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM18 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 18, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

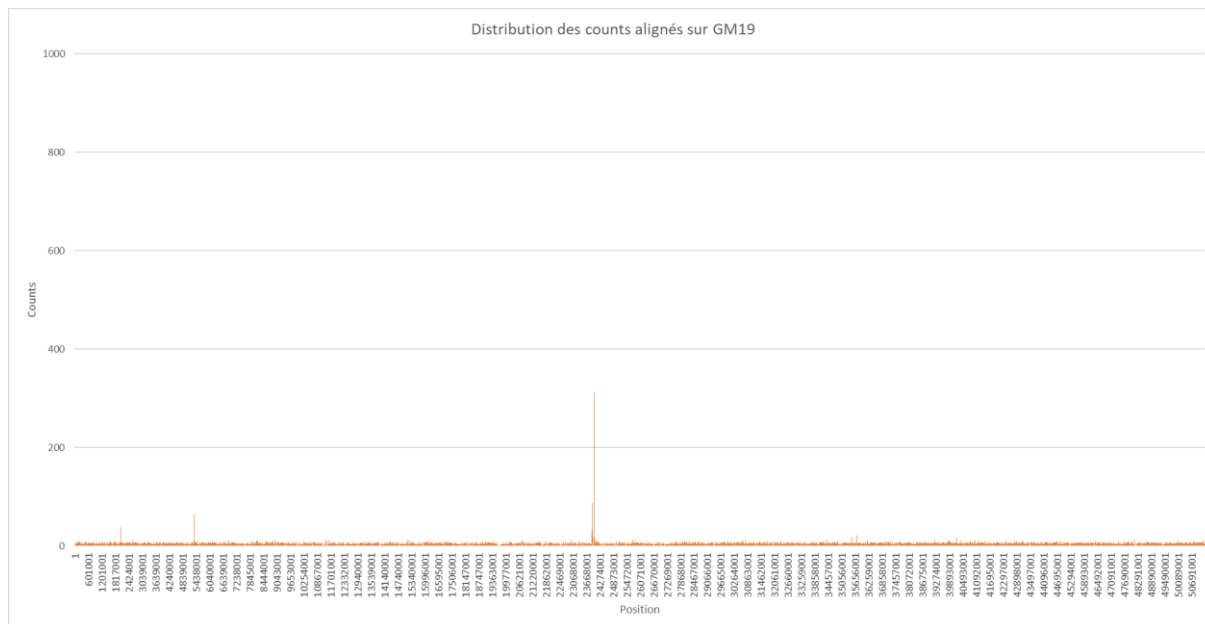


Figure III-56 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM19 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 19, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

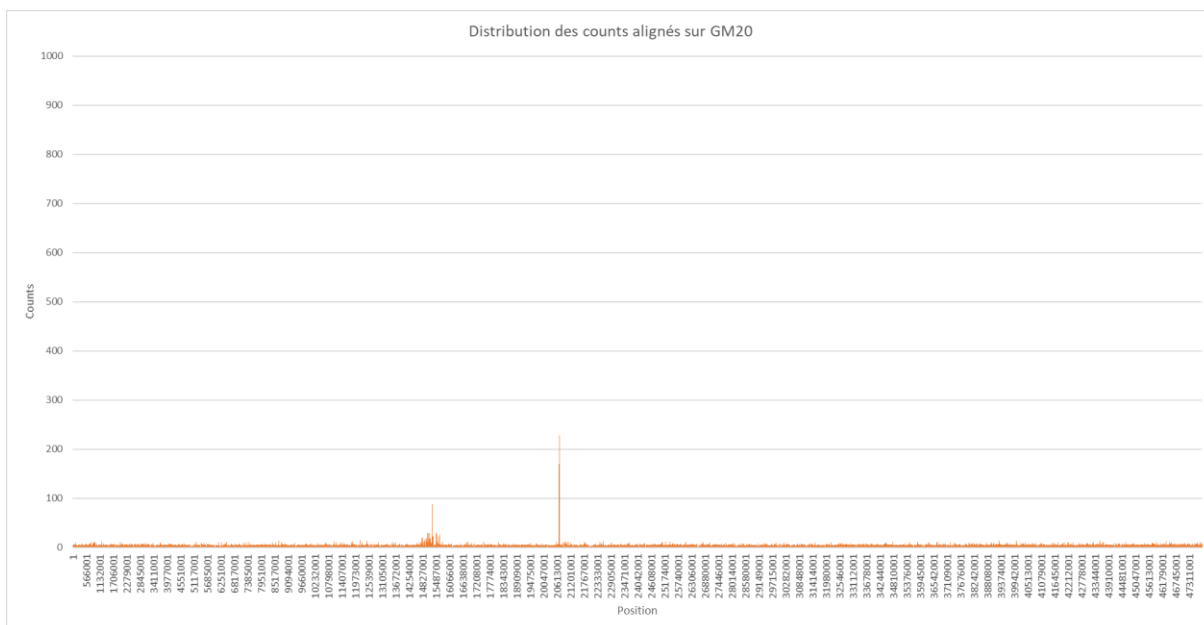


Figure III-57 : Distribution des *counts* d'ADNex de soja alignés sur le chromosome GM20 en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le chromosome 20, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

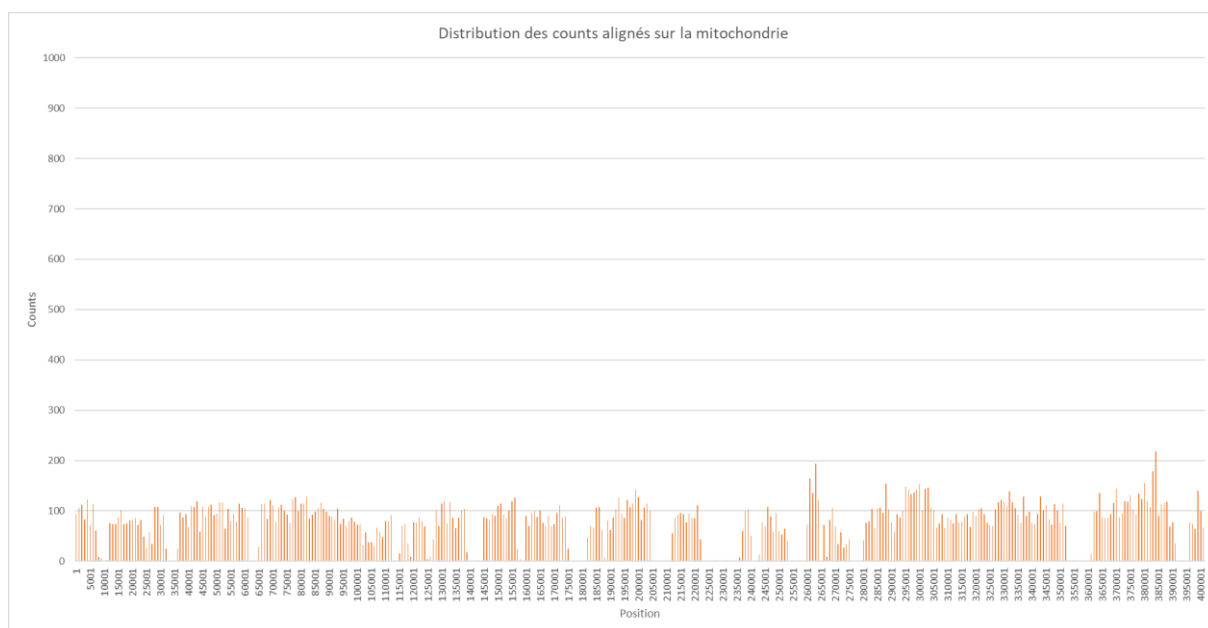


Figure III-58 : Distribution des *counts* d'ADNex de soja alignés sur la mitochondrie en condition élicitée

Sur cette figure on observe la distribution des *counts* sur la mitochondrie, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

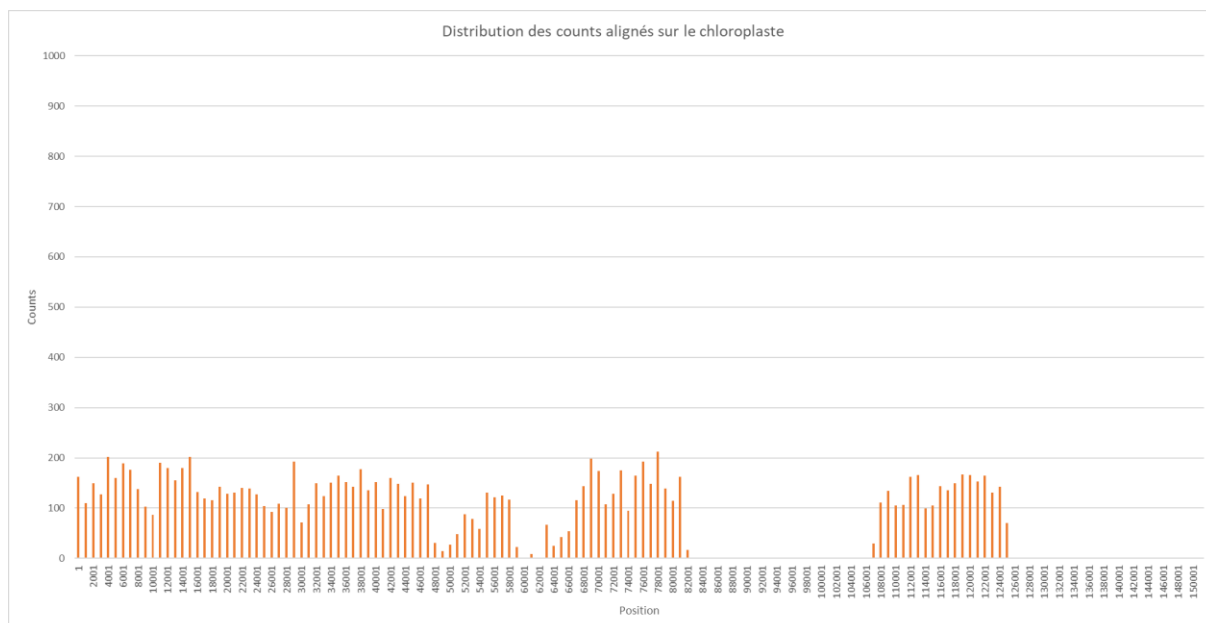


Figure III-59 : Distribution des *counts* d'ADN_{ex} de soja alignés sur le plaste en condition élicitée

Sur cette figure on observe la distribution des *counts* sur le plaste, c'est-à-dire le nombre de *counts* alignés en fonction de leur position. La position est présentée en paires de bases (pb).

Résumé

Le soja, culture d'intérêt agronomique mondiale et bientôt Normande, doit faire face à l'attaque de nombreux phytopathogènes et notamment à l'oomycète *Phytophthora sojae* Kaufm. & Gerd., qui engendrent chaque année d'importantes pertes économiques. Le RET (*Root Extracellular Trap*) est situé à l'apex racinaire, est constitué de cellules détachées de la racine appelées cellules frontières ou AC-DC (*root Associated Cap-Derived Cells*) et de leur mucilage associé. On retrouve au sein de ce mucilage des glycomolécules, des protéines, ou encore de l'ADN extracellulaire (ADNex). Cet ensemble, formant un territoire racinaire particulier, va permettre la protection de la racine et notamment de son apex contre les stress biotiques et abiotiques. Afin de mieux comprendre les mécanismes de défense racinaires, et notamment de certains acteurs du RET, une analyse transcriptomique et protéomique des AC-DC et de la racine et une analyse génomique de l'ADNex ont été réalisées en conditions témoin et élicitée avec PEP-13, un éliciteur provenant des oomycètes du genre *Phytophthora sp.* Les analyses transcriptomique et protéomique ont montré une spécificité des AC-DC par rapport aux racines et une réponse à PEP-13 qui semble différente entre ces deux zones. Le séquençage génomique de l'ADNex du RET a été réalisé afin de déterminer l'origine de celui-ci et savoir s'il existe une spécificité de séquences entre les deux conditions. Il semblerait que l'ADNex ait une meilleure couverture lors de l'alignement sur l'ADN mitochondrial et l'ADN plastidial comparé à l'ADN chromosomique. Cette différence de couverture peut indiquer une différence de persistance de l'ADNex dans le RET en fonction de son origine (chromosomique ou mitochondrial et plastidial), ou une libération dans le RET par ces organites. Il semblerait également qu'il n'y ait pas de différences entre les séquences d'ADNex du RET en condition élicitée avec PEP-13 ou témoin.

Mots clés : *Glycine max* (L.) Merr. (soja), éliciteur, PEP-13, *root extracellular trap*, AC-DC, ADN extracellulaire, transcriptomique, protéomique, séquençage ADN.

Abstract

Soybean, a crop of Normand and world agronomic interest, is threatened by numerous phytopathogens like the oomycete *Phytophthora sojae* Kaufm. & Gerd., wich generate high levels of economical losses. The RET (root extracellular trap) is located at the root apex and is composed of border cells or AC-DC (root associated cap-derived cells) and their mucilage. This mucilage is made up of glycomolecules, proteins or also extracellular DNA (exDNA). The RET play a role in root protection against biotic stresses. In order to better understand the role of the RET in root protection, a transcriptomic and a proteomic analysis where done on AC-DC and roots in controle condition and in elicited condition with PEP-13 (an elicitor from *Phytophthora sp.*). The results show a specificity of AC-DC compared to the root, and an answer to PEP-13 wich seems to be different between these two tissues. An other experiment was to sequence RET exDNA in controle and elicited conditions, in order to define the origin of this exDNA. We show that the coverage of mitochondrial and plastidial DNA where much better than the coverage of chromosomic DNA. It could mean that chromosomic DNA isn't conserved as well as organelles DNA, or exDNA could originate from organelles. Furthermore, results seems to show no differences between the sequences of elicited or control exDNA.

Key words : *Glycine max* (L.) Merr. (soybean), elicitor, PEP-13, *root extracellular trap*, AC-DC, extracellular DNA, transcriptomic, proteomic, DNA sequencing.