



Université des Sciences et Technologies de Lille

- UMR1281 : Stress Abiotiques et Différenciation des Végétaux Cultivés -

THÈSE

pour l'obtention du grade de

Docteur en Sciences de l'Université de Lille 1

Stratégies d'Exploitation des Fonctions Biologiques

Etude transcriptomique au cours des phases précoces de l'embryogenèse somatique chez la chicorée (*Cichorium intybus* L.)

Par
Sylvain Legrand

Soutenue le 6 décembre 2006 devant la commission d'examen :

Kim BOUTILIER
Martin CRESPI
Sébastien DUPLESSIS
Theo HENDRIKS
Jean-Louis HILBERT
David HOT
Marie-Christine QUILLET

Wageningen UR, Pays-Bas
Directeur de Recherches, CNRS
Chargé de Recherches, INRA
Professeur, USTL
Professeur, USTL
Responsable de projets, Institut Pasteur
Maitre de conférences, USTL

Rapporteur
Rapporteur
Examineur
Examineur
Examineur
Examineur
Examineur

Merci...

Je remercie le professeur Jean-Louis Hilbert, directeur de ces travaux de thèse, pour son accueil au sein du laboratoire de Stress Abiotiques et Différenciation des Végétaux Cultivés (SADV) et pour avoir mis à disposition les moyens nécessaires à la bonne réalisation de ces travaux.

Je remercie Kim Boutilier et Martin Crespi d'avoir accepté d'être mes rapporteurs et de juger de cette thèse.

Je tiens à témoigner mes plus sincères remerciements à Marie-Christine Quillet pour son encadrement et ses multiples conseils pendant toute ma thèse.

Je remercie le professeur Yves Lemoine, qui nous a ouvert les portes de son laboratoire (Biopuces du Laboratoire d'Etudes Transcriptomiques et Génomiques Appliquées de l'Institut Pasteur de Lille) pour la réalisation des microarrays. Merci également à toute l'équipe des Biopuces : Christine, David, Erwan, Fabien, Hasna, Ludovic, pour leur accueil et leur disponibilité.

Je tiens à remercier le professeur Theo Hendriks pour ses multiples corrections d'anglais au cours de ma thèse et pour les nombreuses discussions que nous avons eues.

Je remercie l'ensemble des membres du laboratoire SADV pour leur aide apportée au cours de ces travaux. Je remercie particulièrement les étudiants qui se sont succédés dans le bureau 310 : Aline, Clara, David, Franck, Jérôme, Matthieu, Meriem, Sophie A., Sophie D.S. Un merci spécial à Estelle pour la correction de mes nombreuses fautes d'orthographe.

Enfin, je terminerai ces remerciements en témoignant ma reconnaissance à mes parents pour leur aide apportée au cours de ces longues années d'étude et en remerciant mon petit Amour pour sa patience et ses encouragements, particulièrement lors des dernières semaines de rédaction de thèse.

- ETUDE BIBLIOGRAPHIQUE -	11
Introduction générale	13
I. L'embryogenèse somatique	14
A. Caractéristiques générales	14
B. Les phases précoces de l'embryogenèse somatique	16
1) Description	16
a) L'induction	16
b) La réactivation cellulaire	17
c) La première division embryonnaire	17
2) Les fonctions mises en œuvre au cours des phases précoces	18
a) Gènes impliqués dans la transition de l'état somatique à l'état embryonnaire	20
b) Gènes intervenant dans la division cellulaire	21
c) Gènes induits par les phytohormones	21
d) Gènes impliqués dans la signalisation cellulaire	22
e) Gènes codant des protéines extra-cellulaires	22
II. Les techniques récentes d'identification de gènes exprimés différemment	24
A. Les banques d'ESTs	24
1) Création des banques d'ESTs	25
a) Banques d'ADNc	25
b) Banques soustractives d'ADNc	26
c) Séquençage des ESTs	27
2) Analyse des ESTs	28
a) Tri des séquences obtenues	28
b) Assemblage des ESTs	29
c) Annotation des EST	29
3) Le criblage des banques d'ADNc	30
4) Les ESTs et la mise en évidence de gènes exprimés différemment chez les végétaux : une alternative pour les espèces non séquencées ?	31
B. Les microarrays	32
1) Puces à façon ou puces à oligonucléotides ?	33
a) « Puces à façon »	33
b) Puces à oligonucléotides	34
c) Avantages et inconvénients	34
2) Les schémas expérimentaux des « puces à façon »	35
a) Les différents schémas	35
b) Les répétitions expérimentales et biologiques	37
3) Les analyses des résultats des « puces à façon »	38
a) Analyse des images	38
b) Normalisation des données	39
c) Analyses statistiques	40
d) Représentation des résultats	41
4) Validation technique des microarrays	42
5) Exemples d'études menées par microarray chez les plantes	43
a) Développements végétatifs et reproducteurs	45
b) Les signaux environnementaux	45
III. La chicorée, plante modèle de l'embryogenèse somatique	47
A. La chicorée	47
B. Contexte scientifique	47
C. Objectifs de la thèse	49
- MATERIEL ET METHODES -	51
I. Matériel végétal	53

II. Méthodes	53
A. Conditions de culture	53
1) Propagation clonale du matériel végétal	53
a) Maintien et multiplication du matériel par embryogenèse somatique	53
b) Maintien et multiplication par organogenèse	54
2) Induction de l'embryogenèse somatique à partir d'explants foliaires	55
B. Extraction d'ARN totaux	56
C. Construction des banques soustractives d'ADNc	56
1) Obtention des ADNc soustraits et normalisés	56
2) Clonage des ADNc	59
3) Transformation de bactéries compétentes	59
4) Séquençage	60
D. Analyse des ESTs	60
1) Traitement des séquences	60
2) Regroupement des EST identiques	60
3) Assemblage des ESTs en contigs	61
4) Analyse fonctionnelle	61
5) Soustraction <i>in silico</i> des ESTs	61
E. Microarrays	62
1) Préparation des produits PCR et dépôt des sondes sur les lames	62
2) Fabrication des cibles	63
3) Hybridation	64
4) Acquisition des données et tri des spots	64
5) Analyse des données	64
F. RT-PCR en temps réel	65
1) Définition des amorces	65
2) Transcription inverse	65
3) Amplification	65
4) Analyse des données de PCR en temps réel	66
- RESULTATS ET DISCUSSION -	67
Partie I : Fabrication et analyse des banques soustractives embryogène et non embryogène	69
A. Aspects cytologiques et choix du matériel végétal	69
B. Fabrication des banques SSH	71
1) Vérification des ADNc des deux banques	71
2) Clonage des ADNc des banques E et NE	71
C. Analyse des ESTs	73
1) Traitement des séquences	73
2) Informations sur les ESTs	74
3) Assemblage des ESTs identiques	74
4) Assemblage des ESTs en contigs	75
a) Stratégie	75
b) Résultats	75
5) Analyse fonctionnelle des ESTs	77
a) Annotation des séquences	77
b) Classification fonctionnelle	78
D. Soustraction <i>In silico</i> des ESTs des banques E et NE et validation par PCR en temps réel	81
1) Soustraction <i>in silico</i>	81
2) PCR en temps réel	82
E. Discussion	83
1) Choix des génotypes et du stade de développement pour la construction des banques	83
2) Comparaison des ESTs aux séquences d'autres espèces	84
3) Classification fonctionnelle des ESTs	85
4) Qualité des banques SSH	85
Partie II : Mise en œuvre de la technique de microarray pour l'étude de l'embryogenèse somatique chez la chicorée	89
A. Mise en place du schéma expérimental	89
B. Optimisation technique des microarray	91

Sommaire

1)	Choix de l'enzyme d'amplification pour la production des sondes	92
2)	Choix de la quantité d'ARN nécessaire à la fabrication des cibles	93
C.	Production des sondes et des lames	94
D.	Production des cibles	94
1)	Obtention de la référence	94
2)	Synthèse des cibles et hybridations	96
3)	Normalisation intra-lame	97
4)	Mise en évidence de gènes exprimés différemment	99
a)	Mise en évidence des gènes dont les ADNc présentent une incorporation préférentielle d'un fluorochrome	99
b)	Calcul des ratios d'expression	100
5)	Validation du protocole d'analyses	101
E.	Discussion	108
1)	Choix du schéma expérimental	108
2)	Validation du processus d'analyses par microarray	109
3)	Origine des gènes différemment exprimés au jour 4 de culture	111
4)	Gènes exprimés différemment entre les deux génotypes au jour 4 de culture	112
Partie III : Etude en cinétique, par microarray, des phases précoces de l'embryogenèse somatique chez la chicorée		115
A.	Différences transcriptomiques entre les deux génotypes au stade initial de la cinétique d'étude	115
B.	Gènes exprimés différemment au cours de la cinétique en conditions d'embryogenèse somatique	119
1)	Gènes montrant une modulation d'expression au cours de la cinétique d'étude	119
2)	Regroupement des profils d'expression	120
a)	Stratégie	120
b)	Résultats	121
3)	Classification fonctionnelle	130
C.	Discussion	138
1)	Différences transcriptomiques au stade initial de la cinétique	138
2)	Réponse commune des deux génotypes aux conditions de culture appliquées	139
3)	Réponse du génotype embryogène	140
4)	Réponse du génotype non-embryogène	144
- CONCLUSIONS ET PERSPECTIVES -		149
- REFERENCES BIBLIOGRAPHIQUES -		159
- ANNEXES -		179

Figures

Figure 1 : Les différentes phases de l'embryogenèse somatique.....14

Figure 2 : Phases précoces de l'embryogenèse somatique : aspects cellulaires (a) et parallèle avec les deux grands types d'embryogenèse somatique (b et c).....15

Figure 3 : Evolution du nombre d'ESTs depuis 2002 pour 6 espèces de plantes dont le génome n'est pas séquencé.32

Figure 4 : Principe général des « puces à façon » d'après Duggan *et al.* (1999).....33

Figure 5 : Schéma expérimental direct et schéma expérimental indirect.36

Figure 6 : Etude en cinétique par un modèle direct ou un modèle indirect.36

Figure 7: Normalisation intra lame39

Figure 8 : Normalisation inter-lame.40

Figure 9 : Regroupement des gènes d'intérêt par profils d'expression42

Figure 10 : Propagation clonale du génotype K59 par embryogenèse somatique54

Figure 11 : Propagation clonale du génotype C15 par organogenèse55

Figure 12 : Culture en conditions d'embryogenèse somatiques56

Figure 13 : Schéma général de la fabrication de la banque soustractive embryogène.....57

Figure 14 : Soustraction et normalisation au cours de la fabrication de la banque embryogène58

Figure 15 : Carte du vecteur pGEM-T de Promega59

Figure 16 : Equation mathématique de Audic and Claverie (1997) utilisée pour la soustraction *in silico* des ESTs62

Figure 17 : Observations cytologiques d'explants foliaires des génotypes K59 (a et b) et C15 (c et d) placés 4 jours en conditions de culture d'embryogenèse somatique.70

Figure 18 : Vérification de la taille des ADNc des banques embryogène (E) et non embryogène (NE) après migration sur gel d'agarose.....71

Figure 19 : Distribution des clones suivant leur taille pour 6 conditions de ligature.....72

Figure 20 : Composition des séquences brutes73

Figure 21 : Distribution du nombre d'ESTs qui représentent potentiellement un même gène 77

Figure 22: Distribution par classe de fonction des ESTs des banques E (a) et NE (b)79

Figure 23 : Schéma expérimental.....89

Figure 24 : Quatre lames composent une expérience.....90

Figure 25 : Plan de comparaison des expériences pour la mise en évidence de gènes présentant une variation d'expression aux cours de la cinétique de culture en conditions d'embryogenèse somatique.....91

Table des illustrations

Figure 26 : Représentation d'une distribution sous forme d'une boîte à moustaches (box plot)	92
Figure 27 : Effet de l'ADN polymérase utilisée pour la synthèse des sondes sur l'intensité des signaux enregistrés	93
Figure 28 : Effet de la quantité d'ARN utilisée lors de la fabrication des cibles sur l'intensité du signal enregistré.....	93
Figure 29 : Distribution par classe des quantités d'ADN produits à partir de 648 clones.	94
Figure 30 : Electrophorégramme des différents ARN utilisés pour l'obtention de la référence	96
Figure 31 : Validation d'une paire de cibles avant hybridation.....	96
Figure 32 : Effet de la normalisation intra-lame sur l'intensité des signaux enregistrés	98
Figure 33 : Observation de l'effet de la normalisation intra-lame à l'aide de représentations en boîtes à moustaches des données des Figure 32 C et D	99
Figure 34 : Graphiques en volcan représentant les gènes analysés par microarray pour la comparaison des expériences K59J2 et K59J0	101
Figure 35 : Corrélation entre les analyses transcriptomiques menées par PCR en temps réel et celles menées par microarray.....	109
Figure 36 : Stratégie mise en place pour le regroupement des profils d'expression des gènes montrant une modulation d'expression en cours de cinétique.....	121
Figure 37 : Courbe des valeurs FOM pour le regroupement des 834 profils d'expression observés chez les génotypes K59 et C15 pour les 417 gènes sélectionnés.	122
Figure 38 : Regroupement par la méthode de K-means des 834 profils d'expressions correspondant aux 417 gènes sélectionnés	122
Figure 39 : Regroupement par la méthode de K-means des gènes de la sous-population A..	124
Figure 40 : Distribution par classe de fonction des gènes montrant une hausse d'expression (a) ou une baisse d'expression (b) chez les deux génotypes au cours de la cinétique en conditions d'embryogenèse somatique.....	125
Figure 41 : Regroupement par la méthode de K-means des gènes de la sous-population B..	127
Figure 42 : profils d'expression de deux gènes issus du groupe 12.....	129
Figure 43 : Résultats de l'analyse transcriptionnelle et classification fonctionnelle des 249 gènes sélectionnés	132
Figure 44 : Profils d'expression des gènes codant des protéines ribosomales et montrant des variations d'expression observées uniquement chez le génotype embryogène	141
Figure 45 : Modèle synthétique de la réponse du génotype embryogène	154

Figure 46 : modèle synthétique de la réponse du génotype non-embryogène..... 155

Tableaux

Tableau 1 : Liste non exhaustive de gènes exprimés au cours des phases précoces de l'embryogenèse somatique chez différentes espèces 19

Tableau 2 : Les 5 espèces de plantes les plus représentées en terme de nombre d'EST en janvier 200 Putney, 1983 Nature, 302, p7186 dans dbEST (NCBI), et leur évolution depuis avril 2003. 25

Tableau 3 : Exemples d'études par microarray de différents processus de développement chez les plantes..... 44

Tableau 4 : Caractéristiques des ESTs des banques E et NE..... 74

Tableau 5 : Comparaison des ESTs de chicorée avec les séquences d'autres espèces de plantes. 76

Tableau 6 : Extrait de l'Annexe numérique 4, présentation des champs 80

Tableau 7 : Soustraction in silico et analyse transcriptomique par PCR en temps réel 82

Tableau 8 : Composition de la référence utilisée pour le schéma expérimental indirect de microarray 95

Tableau 9 : Liste des gènes identifiés comme étant exprimés différemment au quatrième jour de culture d'embryogenèse somatique par microarray..... 103

Tableau 10 : Validation par PCR en temps réel de résultats obtenus par microarray 107

Tableau 11 : Distribution des ESTs des banques E et NE suivant l'expression des gènes correspondants mesurée par microarray. 112

Tableau 12 : Liste des gènes préférentiellement exprimés chez les génotypes K59 (a) et C15 (b) au stade initial de la culture 118

Tableau 13 : Nombre de gènes exprimés différemment au cours de la cinétique d'étude chez les génotypes K59 et C15 119

Tableau 14 : Liste des gènes du groupe 19 issu de l'assemblage des profils d'expression des gènes de la sous-population B..... 128

Tableau 15 : Liste des gènes du groupe 11 issu de l'assemblage des profils d'expression des gènes de la sous-population B..... 129

- Etude bibliographique -

Introduction générale

Les cellules des organismes multicellulaires acquièrent leur structure et leur fonction par le mécanisme de différenciation. Ce processus correspond à une activation au sein de la cellule d'un programme génétique spécifique. Chez la plupart des organismes supérieurs, la différenciation est unidirectionnelle et irréversible. Cependant, chez les végétaux supérieurs, la majorité des cellules différenciées a la possibilité de se transdifférencier et donc de démarrer un nouveau programme génétique. C'est ce qu'on appelle la totipotence des cellules végétales.

L'embryogenèse somatique, qui correspond à la formation *in vitro* d'embryons à partir de cellules somatiques par un mécanisme qui ressemble à l'embryogenèse zygotique (Zimmerman, 1993), est une bonne illustration de la totipotence végétale et constitue de ce fait un modèle pour étudier les mécanismes de programmation cellulaire. Les premiers embryons somatiques ont été décrits à partir de cals (massifs de cellules indifférenciées) de carotte il y a bientôt 50 ans (Steward, 1958) et aujourd'hui, des embryons somatiques peuvent être obtenus pour de nombreuses plantes à partir d'une grande diversité de tissus tels que les feuilles, les racines, les cotylédons, les embryons zygotiques immatures, les protoplastes et les cultures de cellules (Zorinants *et al.*, 2003). En dépit d'efforts considérables, les processus qui permettent la transition de l'état somatique à l'état embryogène, c'est-à-dire l'induction, la dédifférenciation et la redifférenciation des cellules, restent à ce jour peu connus (Feher *et al.*, 2003).

Au laboratoire de Stress Abiotiques et Différenciation des Végétaux cultivés (SADV), les phases précoces du processus d'embryogenèse somatique chez la chicorée sont étudiées par différentes approches : cytologiques, physiologiques et moléculaires. Dans le cadre de travaux récents, plusieurs gènes impliqués dans la transition de l'état somatique à l'état embryogène ont pu être identifiés. L'objectif de cette thèse a été d'utiliser une approche globale et des outils actuels (banques soustractives et microarray) dans le but d'identifier un maximum de gènes exprimés différenciellement au cours des phases précoces de l'embryogenèse somatique et ainsi de mieux comprendre les mécanismes moléculaires qui permettent à certaines cellules somatiques de changer de programme génétique. Les microarrays ont été réalisés en étroite collaboration avec la plateforme Biopuces du Laboratoire d'Etudes Transcriptomiques et Génomiques Appliquées de l'Institut Pasteur de Lille.

I. L'embryogenèse somatique

A. Caractéristiques générales

Chez les plantes, l'embryogenèse somatique est un moyen de multiplication végétative et constitue au même titre que d'autres processus de développement *in vitro* tels que l'androgenèse ou la gynogenèse, ou *in vivo* tel que l'apomixie, une alternative à la reproduction sexuée.

Ce processus est un moyen rapide et efficace pour multiplier les plantes *in vitro*. C'est aussi un outil d'étude de l'embryogenèse zygotique, puisque les deux types d'embryogenèses suivent globalement les mêmes étapes de développement (Zimmerman, 1993) (Figure 1). De plus, les événements cellulaires qui ont lieu au cours de l'embryogenèse somatique, font de ce processus un bon outil d'étude de la programmation cellulaire. En effet, le processus d'embryogenèse somatique implique une dédifférenciation puis une redifférenciation des cellules somatiques. L'étape de dédifférenciation aboutit à l'obtention de cellules compétentes qui peuvent percevoir des stimuli, internes ou externes, qui vont induire le passage à un état embryogène (Toonen et al., 1994). Les cellules compétentes sont des cellules intermédiaires entre les cellules somatiques et les cellules dites embryogènes (Figure 2). Au cours de la phase de redifférenciation, certaines cellules embryogènes vont se développer en embryons somatiques par des divisions successives.

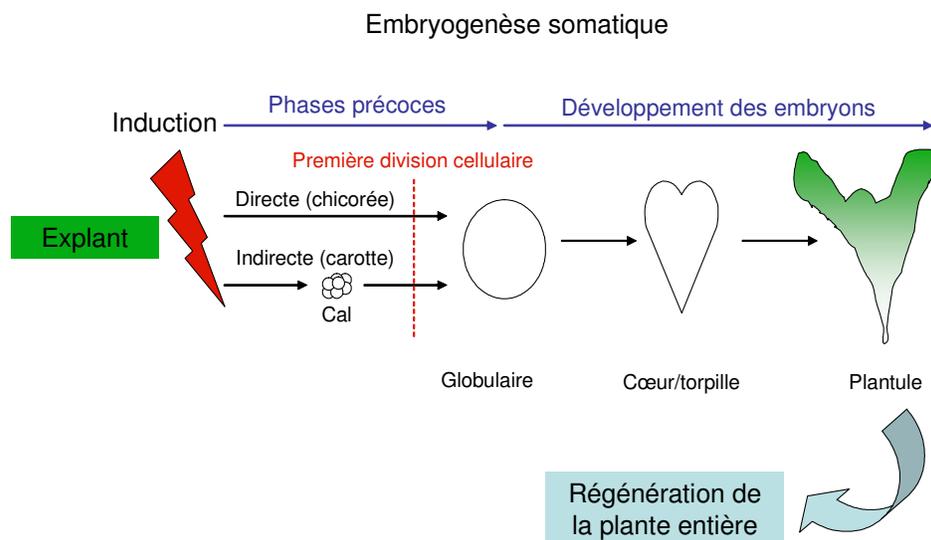
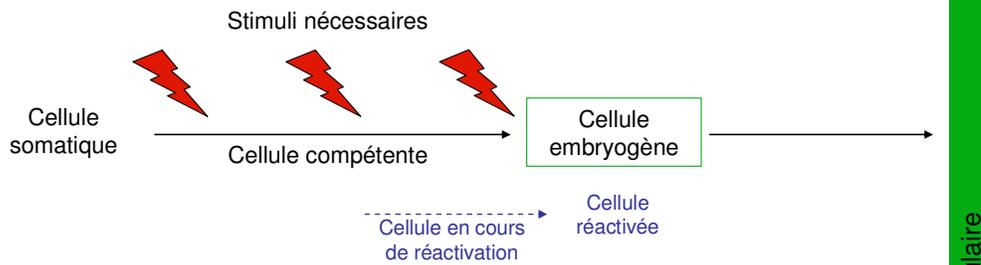


Figure 1 : Les différentes phases de l'embryogenèse somatique

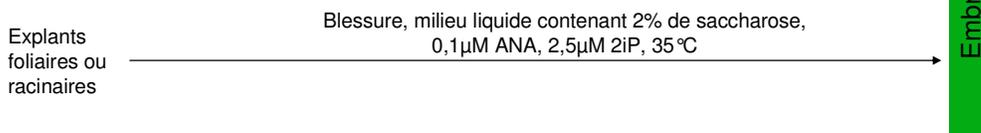
Embryogenèse somatique directe, indirecte, unicellulaire et pluricellulaire

Les embryons somatiques peuvent être obtenus de deux façons différentes : selon un mode direct ou selon un mode indirect (Figure 2b et c). L'embryogenèse somatique indirecte passe par une phase de multiplication cellulaire aboutissant à la formation d'un massif de cellules indifférenciées (callogenèse). C'est le cas, par exemple, chez la carotte (de Vries *et al.*, 1988). L'embryogenèse somatique directe se déroule, quant à elle, sans aucune phase préalable de formation de cal. La cellule somatique subit alors les étapes de dédifférenciation et de redifférenciation au sein même de l'explant initial. C'est ce type d'embryogenèse somatique que l'on trouve notamment chez la chicorée (Dubois *et al.*, 1990).

a : Embryogenèse somatique - aspects cellulaires des phases précoces



b : Embryogenèse somatique **directe**, exemple de la chicorée (Dubois *et al.*, 1991)



c : Embryogenèse somatique **indirecte**, exemple de la carotte (de Vries *et al.*, 1988)

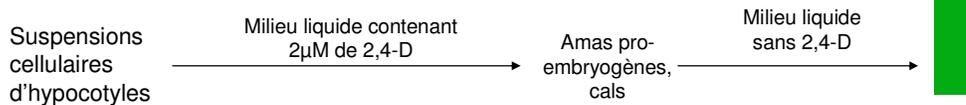


Figure 2 : Phases précoces de l'embryogenèse somatique : aspects cellulaires (a) et parallèle avec les deux grands types d'embryogenèse somatique (b et c).

Certaines cellules somatiques appelées cellules compétentes vont réagir aux stimuli appliqués tels que les régulateurs de croissance et vont entrer en phase de réactivation cellulaire. Les cellules réactivées sont alors qualifiées de cellules embryogènes et plus aucun stimulus, n'est à ce stade, nécessaire pour aboutir à la formation d'embryons somatiques (a).

On distingue deux types principaux d'embryogenèse somatique : l'ES directe (b) et l'embryogenèse somatique indirecte (c) : dans le premier cas, les cellules somatiques se dédifférencient et se redifférencient en embryons somatiques au sein du tissu en une seule phase de culture. Dans le second cas (c), une phase d'induction aboutit à la formation de structures cellulaires indifférenciées compétentes à devenir embryogènes (cals) qui nécessitent une seconde phase permettant leur différenciation en embryons somatiques.

Par ailleurs, les embryons somatiques peuvent être issus soit d'une cellule unique soit de plusieurs cellules appelées masses pro-embryogènes. Le premier cas, appelé embryogenèse somatique unicellulaire, est typique de la carotte ou de la chicorée. Le second cas, l'embryogenèse somatique pluricellulaire, est observé, par exemple, chez le trèfle (Maheswaran and Williams, 1985).

B. Les phases précoces de l'embryogenèse somatique

Les phases précoces de l'embryogenèse somatique, auxquelles nous allons nous intéresser, sont initiées par l'induction du processus et se terminent par l'apparition des premières divisions embryonnaires. C'est lors de ces phases précoces de l'embryogenèse somatique qu'ont lieu les événements de re-programmation génétique. Les phases tardives, correspondant au développement de l'embryon à partir d'un stade globulaire typique, ne seront pas décrites ici (Figure 1).

1) Description

a) L'induction

L'embryogenèse somatique peut être induite par des facteurs endogènes et/ou des facteurs exogènes. Les hormones endogènes, notamment les auxines et l'acide abscissique, impliquées dans la régulation de la division et de la différenciation cellulaire, sont considérées comme des facteurs déterminants pour l'induction de l'embryogenèse somatique. En effet, une concentration élevée en acide indole acétique (AIA) endogène est associée à l'augmentation de la réponse embryogène chez de nombreuses espèces (Ivanova *et al.*, 1994; Jimenez and Bangerth, 2001). Par exemple, chez le tournesol, un pic de concentration en AIA endogène est observé dans les embryons somatiques en cours de formation (Charrière *et al.*, 1999). Par ailleurs, la contribution directe de l'acide abscissique (ABA) endogène sur la phase d'induction de l'embryogenèse somatique a été montrée chez le tabac (Senger *et al.*, 2001). La déficience en ABA endogène perturbe fortement la formation des embryons.

Afin d'obtenir des embryons somatiques, les explants sont isolés de leur environnement tissulaire d'origine, puis sont mis en culture dans un milieu contenant des minéraux, des vitamines et des régulateurs de croissance en concentrations non physiologiques. Ces différentes étapes de mise en culture constituent autant de stress pouvant intervenir directement dans l'induction du processus. Les facteurs comme la blessure (Grosset *et al.*, 1990), le choc thermique (Decout *et al.*, 1994), le changement de pH (Smith and Krikorian, 1990), de fortes concentrations en sel, en ions ou en métaux lourds, ou un stress

osmotique (Dudits *et al.*, 1995; Pasternak *et al.*, 2002) peuvent participer à l'induction de la formation d'embryons somatiques.

L'induction peut également être obtenue par l'action de phytohormones exogènes. Parmi les différentes phytohormones utilisées, un analogue de l'auxine, l'acide 2,4-dichlorophenoxy acétique (2,4-D) est l'un des plus utilisés pour l'obtention d'embryons somatiques, notamment chez la carotte (Toonen *et al.*, 1994) et chez la luzerne (dos Santos *et al.*, 1983). L'embryogenèse somatique peut aussi être induite par l'application d'autres types d'hormones comme les cytokinines (Sagare *et al.*, 2000) ou l'acide abscissique (Nishiwaki *et al.*, 2000) qui agissent souvent en synergie avec d'autres hormones.

b) La réactivation cellulaire

La phase de réactivation a été particulièrement décrite chez l'hybride « 474 » de chicorée. Les embryons somatiques obtenus à partir d'explants foliaires se développent, après induction, sous une incubation à 35°C (Decout *et al.*, 1994) (Figure 2b). Les premiers événements observés pour les cellules en cours de réactivation sont la migration des chloroplastes vers le noyau. Ensuite, l'aboutissement à une cellule réactivée passe par un grossissement du noyau et sa migration en position centrale de la cellule. On observe, au sein du noyau, un nucléole volumineux, ce qui traduit une forte activité transcriptionnelle. Un réseau de microtubules part du noyau et forme des trabécules cytoplasmiques en forme d'étoile. En parallèle, la vacuole se fragmente et de la callose se dépose au niveau de la paroi des cellules embryonnaires (Blervacq *et al.*, 1995; Dubois *et al.*, 1991).

Les cellules réactivées sont alors prêtes à entrer en division et plus aucun stimulus n'est nécessaire à la poursuite du processus (Figure 2 a).

c) La première division embryonnaire

Au cours de l'embryogenèse zygotique, chez les plantes supérieures, la première division est asymétrique. Elle aboutit à la formation d'une petite cellule apicale qui donnera l'embryon proprement dit, et à une grande cellule basale qui formera le suspenseur, tissu qui relie l'embryon à la plante et assure sa nutrition (Cresti *et al.*, 1992). Cette division asymétrique établit une polarité qui est déterminante pour le développement ultérieur de l'embryon (Jurgens *et al.*, 1997).

La première division embryonnaire asymétrique n'est pas systématiquement retrouvée au cours de l'embryogenèse somatique. Par exemple, chez la luzerne, l'asymétrie de la première division est dépendante de la concentration en 2,4D apporté (Pasternak *et al.*, 2002)

ainsi que du génotype de la plante (Bögre *et al.*, 1990). Chez la carotte, lors de la formation d'embryons somatiques à partir de cultures cellulaires, la première division peut être asymétrique ou symétrique (Toonen *et al.*, 1994). Chez la chicorée, la première division de la cellule embryonnaire est symétrique et anticline par rapport à l'orientation des éléments vasculaires (Blervacq *et al.*, 1995). Ensuite, des divisions anticlines successives conduisent à la formation d'une structure embryonnaire qui aboutira à la future plantule (Blervacq *et al.*, 1995; Dubois *et al.*, 1991).

2) Les fonctions mises en œuvre au cours des phases précoces

Au cours des 20 dernières années, l'embryogenèse somatique a été étudiée par différentes approches, et des gènes, très divers, impliqués dans ce processus ont été caractérisés. La complexité de l'embryogenèse somatique suggère qu'une grande diversité de fonctions biologiques, et donc de gènes, est mise en œuvre, notamment au cours de l'initiation du mécanisme. Bien que la majorité des recherches ait été concentrée sur les phases tardives de l'embryogenèse, les quelques gènes identifiés comme étant essentiels aux phases précoces du processus (Tableau 1) permettent d'établir une corrélation entre les fonctions biologiques impliquées au cours des phases précoces et l'état cytologique observé au cours de ces mêmes phases.

Etude bibliographique

Classe de fonction	Gène	Nature/fonction des protéines déduites	Espèce	Référence
Transcription	C-LEC1	Facteur de transcription	Carotte	(Yazawa <i>et al.</i> , 2004)
	WUSCHEL	Facteur de transcription	Arabidopsis	(Zuo <i>et al.</i> , 2002)
	AGL15	Facteur de transcription	Luzerne	(Perry <i>et al.</i> , 1999)
	ZmMADS1, ZmMADS3	Facteur de transcription	Maïs	(Heuer <i>et al.</i> , 2001)
	Pavp1	Facteur de transcription induit par ABA	Epicea	(Footitt <i>et al.</i> , 2003)
	C-ABI3	Facteur de transcription en réponse à l'ABA	Carotte	(Shiota <i>et al.</i> , 1998)
	SCARECROW	Facteur de transcription	Coton	(Zeng <i>et al.</i> , 2006)
	Ethylene-responsive transcriptional coactivator-like protein	Elément de réponse à l'éthylène	Coton	(Zeng <i>et al.</i> , 2006)
Synthèse protéique	CEM1 elongation factor 1 alpha	Initiation de la traduction	Carotte	(Kawahara <i>et al.</i> , 1992)
	Elongation factor 1 alpha	Initiation de la traduction	Soja	(Thibaud-Nissen <i>et al.</i> , 2003)
	Pinhead	Initiation de la traduction	Soja	(Thibaud-Nissen <i>et al.</i> , 2003)
	40S ribosomal protein	Protéine ribosomale	Soja	(Thibaud-Nissen <i>et al.</i> , 2003)
	plastid ribosomal protein L11	Protéine ribosomale	Soja	(Thibaud-Nissen <i>et al.</i> , 2003)
Division cellulaire et synthèse de l'ADN	cdc2MS	protéine kinase dépendante de cyclines	Luzerne	(Hirt <i>et al.</i> , 1991) (Footitt <i>et al.</i> , 2003)
	CycA	Cycline	Carotte	(Hata <i>et al.</i> , 1991)
	Histones, H2A, H3, H4	Réplication de l'ADN	Soja	(Thibaud-Nissen <i>et al.</i> , 2003)
	Histones H3, H3-11	Réplication de l'ADN	Luzerne	(Kapro <i>et al.</i> , 1992)
	Cdc48	Formation des fuseaux de division	Soja	(Thibaud-Nissen <i>et al.</i> , 2003)
	CRK	Apparentée aux CDPK, phosphorylée par CDK	Carotte	(Lindzen and Choi, 1995)
Gènes induits par les phytohormones	Dchsp-I	Protéine de choc thermique induite par le 2,4-D	Carotte	(Kitamiya <i>et al.</i> , 2000)
	Dcarg1	Induit par l'auxine	Carotte	(Kitamiya <i>et al.</i> , 2000)
	cDNA clones 43, 93	Induits par l'auxine	Carotte	(Yasuda <i>et al.</i> , 2001)
	chi-GST1	Glutathione S transferase inductible par l'auxine	Chicorée	(Galland <i>et al.</i> , 2001)
Signalisation	ASET1, ASET2, ASET3	Signalisation	Luzerne	(Giroux and Pauls, 1997)
	MsCa1	Protéine fixatrice de calcium	Luzerne	(Dudits <i>et al.</i> , 1991)
	MsCPK3	Protéine kinase similaire aux calmodulines	Luzerne	(Davletova <i>et al.</i> , 2001)
	SERK	Récepteur kinase de l'embryogenèse somatique	Carotte	(Schmidt <i>et al.</i> , 1997)
	CH-GTP1	Protéine fixatrice de GTP	Chicorée	(Randoux <i>et al.</i> , 2002)
Protéines extra-cellulaires	EP3-1, EP3-2	Chitinases	Carotte	(van Hengel <i>et al.</i> , 1998)
	Chia4-Pa	Chitinase	Epicéa	(Wiweger <i>et al.</i> , 2003)
	PgChi-1	Endochitinase	Epicéa	(Dong and Dunstan, 1997)
	cg1, cg2, cg3	Glucanase beta 1-3	Chicorée	(Helleboed <i>et al.</i> , 2000a)
	PgGlu-1	Glucanase beta 1-3	Epicéa	(Dong and Dunstan, 1997)
	DcAGP1	AGP	Carotte	(Baldwin <i>et al.</i> , 2001)
	AGP	AGP	Chicorée	(Chapman <i>et al.</i> , 2000)
	EP2	Protéine de transfert de lipides	Carotte	(Sterk <i>et al.</i> , 1991)
	LTP	Protéine de transfert de lipides	Coton	(Zeng <i>et al.</i> , 2006)
Divers	DcCB1	Protéine avec domaine d'interaction à la chromatine	Carotte	(Kiyosue <i>et al.</i> , 1998)
	nshg	Hémoglobine non symbiotique	Chicorée	(Hendriks <i>et al.</i> , 1998)

Tableau 1 : Liste non exhaustive de gènes exprimés au cours des phases précoces de l'embryogenèse somatique chez différentes espèces

a) Gènes impliqués dans la transition de l'état somatique à l'état embryonnaire

L'embryogenèse somatique (ES) implique la mise en place d'un nouveau transcriptome. Ainsi il n'est pas surprenant qu'un certain nombre de gènes préférentiellement exprimés au cours des phases précoces du processus codent des protéines impliquées dans la synthèse des ARNm ou dans la synthèse de protéines telles que des facteurs de transcription ou des facteurs d'élongation de la traduction.

(i) Gènes régulant la transcription

Plusieurs facteurs de transcription impliqués dans l'ES ont été caractérisés. Chez la carotte, un homologue du gène LEC1 (leafy cotyledon 1) d'*Arabidopsis* qui régule la morphologie des embryons et la maturation des graines a été isolé et montre non seulement une expression dans les embryons somatiques et dans les graines, mais aussi dans les cellules embryogènes (Yazawa *et al.*, 2004). Par ailleurs, chez *Arabidopsis*, le gène WUSCHEL codant une protéine à homéodomaine impliquée dans la mise en place des méristèmes apicaux caulinaires et inflorescentiels, pourrait également jouer un rôle clé au cours de l'ES en permettant la transition des cellules d'un état somatique à un état embryonnaire et/ou en maintenant l'identité cellulaire des cellules embryogènes (Zuo *et al.*, 2002). Quelques autres facteurs de transcription ont également été identifiés au cours des phases précoces du processus chez des espèces très diverses telles que le maïs, l'épicéa, la luzerne ou le coton (Tableau 1). Citons, par exemple, le gène AGL15 (AGAMOUS-like 15) chez la luzerne, qui est exprimé dans les cellules avant ou juste après l'initiation des premières divisions cellulaires (Perry *et al.*, 1999). Notons aussi qu'une étude récente portant sur le gène codant le facteur de transcription SCARECROW, notamment essentiel à la mise en place de la morphologie radiale de la racine au cours du développement embryonnaire (Di Laurenzio *et al.*, 1996), a montré que ce gène est aussi impliqué dans les phases précoces de l'embryogenèse somatique chez le coton (Zeng *et al.*, 2006). Toujours chez le coton, un gène codant un élément de réponse à l'éthylène est aussi exprimé au cours des phases précoces de l'ES (Zeng *et al.*, 2006).

(ii) Gènes participant à la synthèse protéique

Comme la synthèse des ARNm, la synthèse des protéines est indispensable à la transition de l'état somatique à l'état embryogène. Le gène CEM1 codant le facteur d'élongation 1 α impliqué dans la traduction des messagers est exprimé dans les cellules se

divisant activement au stade globulaire de l'ES chez la carotte (Kawahara *et al.*, 1992). Chez le soja, des analyses par puce à ADN ont aussi mis en évidence l'expression préférentielle d'un gène codant un facteur d'élongation 1α au cours des phases précoces de l'ES ainsi que d'autres gènes en relation avec la synthèse protéique : le gène Pinhead et les gènes codant les protéines ribosomales 40S et L11 (Thibaud-Nissen *et al.*, 2003). Pinhead est, par ailleurs, requis pour la formation des méristèmes apicaux et axillaires des tiges (Lynn *et al.*, 1999).

b) Gènes intervenant dans la division cellulaire

Suite à la dédifférenciation des cellules somatiques, la formation des embryons nécessite une phase de divisions cellulaires actives. Quelques gènes reliés au cycle cellulaire ont été identifiés comme étant préférentiellement exprimés au cours des phases précoces de l'ES.

Chez la luzerne, un ADNc, *cdc2MS*, correspondant à une protéine kinase dépendante de cycline (CDK) a été isolé. Le gène codant cette CDK présente un niveau de transcrits élevé dans les suspensions cellulaires induites par l'auxine (Hirt *et al.*, 1991). Chez la carotte, un gène codant une protéine dépendante du calcium (CRK) est exprimé au cours des phases précoces de l'embryogenèse somatique. La CRK serait phosphorylée par les CDK et pourrait jouer un rôle dans la régulation du cycle cellulaire. Encore chez la carotte, un ADNc correspondant à une cycline (protéine régulatrice des CDK) a été isolé. Le gène correspondant montre une expression dans les cultures cellulaires lors de l'embryogenèse somatique. Chez le soja, *cdc48*, un gène impliqué dans le cycle cellulaire, présente une expression préférentielle au cours des phases précoces de l'ES (Thibaud-Nissen *et al.*, 2003). La protéine correspondante participe à la formation des fuseaux de division dans les cellules en prolifération chez les plantes (Feiler *et al.*, 1995).

D'autre part, la réplication de l'ADN se déroulant en parallèle à la division cellulaire, des gènes codant les histones H2A, H3 et H4 présentent le même profil d'expression que celui de *cdc48* au cours des phases précoces de l'ES chez le soja (Thibaud-Nissen *et al.*, 2003). De même chez la luzerne, les gènes codant les histones H1 et H3-11 sont préférentiellement exprimés au cours des phases précoces de l'ES (Kaproš *et al.*, 1992).

c) Gènes induits par les phytohormones

Quelques gènes préférentiellement exprimés au cours des phases précoces de l'ES sont induits par les phytohormones. Parmi les gènes identifiés, citons le gène *Dchsp-1* de carotte qui code une protéine de choc thermique et qui est inductible par le 2,4-D (Kitamiya *et al.*,

2000) ainsi que le gène de chicorée codant une GST (glutathione S-transferase), également inductible par l'auxine (Galland *et al.*, 2001). Parmi les facteurs de transcription exprimés au cours des phases précoces, le gène *Pavp1* est induit par l'ABA chez l'épicéa (Footitt *et al.*, 2003).

d) Gènes impliqués dans la signalisation cellulaire

La signalisation cellulaire est une classe fonctionnelle très représentée au sein des gènes préférentiellement exprimés au cours des phases précoces de l'ES. Chez la luzerne, trois ADNc, ASET1, ASET2 et ASET3 (alfalfa somatic embryogenesis-specific transcript) codant des protéines portant un domaine intra-membranaire et un site de phosphorylation ont été isolés, et présentent une expression préférentielle au cours des phases précoces de l'ES. La signalisation par la voie du calcium semble être importante. En effet, chez la luzerne, le gène *MsCa1* codant une protéine qui fixe le calcium (Dudits *et al.*, 1991) et le gène *MsCPK3* codant une protéine kinase similaire aux calmodulines (Davletova *et al.*, 2001) sont exprimés au cours des phases précoces de l'ES. D'autres gènes eux aussi impliqués dans la signalisation comme le gène *SERK* (somatic embryogenesis receptor kinase) chez la carotte (Schmidt *et al.*, 1997) et le gène *CH-GTP1* de chicorée codant une protéine fixatrice du GTP (Randoux *et al.*, 2002) sont exprimés au cours des phases précoces. Par ailleurs, le gène *SERK* est également exprimé dans les embryons zygotiques de carotte au début du stade globulaire (Schmidt *et al.*, 1997).

e) Gènes codant des protéines extra-cellulaires

Certaines protéines extra-cellulaires semblent jouer un rôle important dans la formation des embryons somatiques. En effet, chez la chicorée, l'accumulation de protéines extracellulaires au cours de l'ES a permis de les relier à l'induction du processus (Boyer *et al.*, 1993; Hilbert *et al.*, 1992). Chez la carotte, il a été montré que dans les suspensions cellulaires, des protéines sont libérées dans le milieu de culture. Ceci entraîne un conditionnement du milieu nécessaire à l'initiation de l'ES. Ces protéines peuvent aussi bien être libérées par les cellules embryogènes que par les cellules non embryogènes (Chug et Khurana, 2002).

Parmi ces protéines, on retrouve des chitinases, des glucanases, des AGPs (arabinogalactan proteins) et des protéines de transfert de lipides. Chez la carotte, les chitinases EP3-1 et EP3-2 sont capables de restaurer le développement des embryons somatiques d'un lignée cellulaire qui, à la température de culture, ne produisent normalement

pas d'embryons (van Hengel *et al.*, 1998). Chez la chicorée, le gène *cg1* code une glucanase β 1-3 et présente une forte expression dans les cellules embryogènes (Helleboid *et al.*, 2000a). Les glucanases pourraient avoir un rôle dans la dégradation de la callose au cours de l'embryogenèse somatique. Les AGPs sont des glycoprotéines composées à 90% de sucres liés à la partie protéique par des résidus sérine et hydroxyproline. On les retrouve au niveau de la membrane plasmique et de la paroi (Kreuger and van Holst, 1996). Les AGPs seraient impliquées dans la prolifération cellulaire (Nothnagel, 1997), l'expansion cellulaire (Willats and Knox, 1996) et la régulation du développement des embryons somatiques (Kreuger and van Holst, 1996). Chez la chicorée, l'addition dans le milieu de culture de réactif de Yariv (β D glycosyl [β GlcY]), réactif qui interagit avec les AGPs, bloque la formation des embryons somatiques (Chapman *et al.*, 2000), ce qui montre qu'effectivement, les AGPs semblent indispensables à la formation des embryons somatiques. Les protéines de transfert de lipides (LTPs) jouent un rôle dans le transport des phospholipides depuis le réticulum endoplasmique jusqu'à des localisations cellulaires diverses (Kader, 1996). Elles sont impliquées indirectement dans la défense des végétaux, et quelques LTPs sont induites par l'ABA. Chez la carotte, le gène *EP2* codant une LTP montre une expression corrélée avec l'apparition du premier tissu différencié, le protoderme aussi bien dans le cadre de l'embryogenèse somatique que zygotique (Sterk *et al.*, 1991). Chez le coton, un gène codant une LTP est exprimé dans les embryons somatiques au stade globulaire. Il montre également une forte expression dans les cals embryogènes mais n'est pas exprimé dans les cals non-embryogènes (Zeng *et al.*, 2006).

II. Les techniques récentes d'identification de gènes exprimés différemment

Depuis l'émergence du northern blot (Alwine *et al.*, 1977), une grande diversité de méthodes a été développée dans le but de mesurer la quantité des transcrits et d'identifier des gènes exprimés différemment. Citons notamment le criblage de banques d'ADNc par la méthode classique utilisant des sondes radioactives, et des méthodes plus récentes comme le « differential display » (Liang and Pardee, 1992), la technique « SAGE » (Serial Analysis of Gene Expression) (Velculescu *et al.*, 1995) et le cDNA-AFLP (Bachem *et al.*, 1996). Cependant, toutes ces méthodes présentent de nombreux inconvénients. Par exemple le « differential display » souffre du fait de ne pas être une technique quantitative et de mettre en évidence des séquences positives difficiles à cloner et à confirmer (Ding and Cantor, 2004). La SAGE est très complexe à réaliser et n'est pas une technique très sensible pour la détection des transcrits peu abondants.

Une autre technique apparue dans les années 1990, les microarrays (Schena *et al.*, 1995), s'est aujourd'hui largement répandue et a supplanté les autres techniques pour les mesures de l'expression des gènes. Cette technique offre l'immense avantage d'être relativement facile à mettre en œuvre et de mesurer l'expression d'un grand nombre de gènes en simultané. En amont des microarrays, les banques d'ESTs (Expressed Sequence Tag), qui se sont elles aussi largement développées, offrent la matière première pour l'analyse transcriptomique.

A. Les banques d'ESTs

Les ESTs sont de courtes séquences d'ADN dont la taille est habituellement comprise entre 300 et 500 pb et qui sont souvent obtenus à partir d'un seul séquençage d'un ADNc. C'est la forme de séquence la plus courante chez les plantes, aussi bien en terme de nombre de séquences qu'en terme de nombre de nucléotides séquencés. Ces ESTs fournissent aux chercheurs une ressource gigantesque qui peut être exploitée aussi bien pour la découverte de gènes, l'annotation de génomes, la génomique comparative que pour les analyses transcriptomiques. Les ESTs sont issus de banques d'ADNc qui fournissent chacune des milliers de clones représentatifs d'un transcriptome, c'est-à-dire de l'expression instantanée des gènes dans un type cellulaire, un tissu, un organe donné ou au cours d'un processus physiologique.

Le concept d'utiliser les ADNc pour mettre en évidence de nouveaux gènes est né dans les années 1980 (Putney *et al.*, 1983) et c'est Mark Adams, en 1991 qui utilise pour la première fois le terme d'EST dans le cadre du séquençage du génome humain. Depuis cette première publication, le nombre d'ESTs déposés dans les bases de données croît de façon exponentielle. Aujourd'hui, plus de 32 millions d'ESTs ont été obtenus à partir de plus de 1000 espèces de tous les règnes systématiques confondus (dbEST, le 20 janvier 2006), dont 662 884 pour le maïs qui est la plante la plus représentée en terme de nombre d'ESTs. A titre de comparaison, en avril 2003, il n'y avait qu'environ 16 millions d'ESTs déposés dans dbEST dont 206 002 pour le maïs (Tableau 2). Notons aussi que de nombreuses collections d'ESTs présentant un intérêt pour les industriels ne sont pas rendues publiques et s'ajoutent à ces chiffres.

Espèce	Nb 2006 (janvier)	Nb 2003 (avril)
Maïs	662 884	206 002
Blé	600 205	423 632
Arabidopsis	421 027	178 544
Riz	407 545	199 186
Orge	395 065	339 034

Tableau 2 : Les 5 espèces de plantes les plus représentées en terme de nombre d'EST en janvier 2006 dans dbEST (NCBI), et leur évolution depuis avril 2003.

1) Création des banques d'ESTs

a) Banques d'ADNc

Classiquement, le premier brin d'ADNc est obtenu par la transcription inverse d'ARNm en utilisant une amorce oligod(T). La synthèse du second brin est réalisée à partir d'un cocktail d'enzymes : une RNase H dégrade l'ARN du duplex ARN/premier brin d'ADNc, une ADN polymérase I synthétise le second brin à partir de la boucle formée en 3' du premier brin d'ADNc par la transcriptase inverse et une ligase, maintient la boucle. Cette boucle est ensuite dégradée par une nucléase spécifique de l'ADN simple brin et l'ADNc est prêt à être cloné. De nombreux kits commerciaux dérivés de ce principe de base existent afin d'obtenir des ADNc de grande taille et/ou afin de réaliser des clonages orientés.

Le clonage peut s'effectuer de deux façons : en plasmide ou en phage λ . Bien que les vecteurs plasmides aient une plus grande flexibilité de construction, le phage λ a longtemps

été préféré au plasmide puisqu'il bénéficie d'une grande efficacité d'empaquetage et de transfection qui permet la formation de banques d'ADNc composées d'un grand nombre de clones. Avec le développement de méthodes de transformation à grande efficacité pour des souches d'*E.coli*, il est devenu possible de réaliser des banques de grande taille avec des vecteurs plasmidiques. C'est pourquoi le plasmide est le vecteur le plus utilisé aujourd'hui pour la construction de banques d'ADNc.

Une banque d'ADNc est représentative d'un transcriptome donné. Cela implique que les transcrits des gènes fortement exprimés seront fortement redondants, mais qu'inversement les gènes faiblement exprimés, comme les facteurs de transcription, seront peu représentés. Ceci constitue la principale limite des banques d'ADNc. Par exemple, d'après Rudd (2003), les 178000 EST d'*Arabidopsis* ne représentaient que 16115 gènes uniques soit un peu plus de la moitié du nombre total de gènes, alors que les EST provenaient de 61 banques différentes.

b) Banques soustractives d'ADNc

Les banques SSH (Suppression Subtractive Hybridization) (Diatchenko *et al.*, 1996) permettent de comparer deux populations d'ARNm et d'obtenir les ADNc correspondants aux gènes qui sont exprimés dans une population mais pas dans l'autre (voir Matériel et méthodes).

La technique de SSH combine en une seule étape une normalisation et une soustraction des ADNc et doit, en principe, faciliter l'identification d'ESTs correspondant à des gènes faiblement exprimés. Les banques SSH sont, par ce fait, aujourd'hui fréquemment utilisées, et ont permis la mise en évidence de nombreux gènes exprimés différemment entre différents tissus ou génotypes, ou au cours d'un processus physiologique. La normalisation est réalisée par une étape de PCR, et c'est au cours de cette même réaction qu'a lieu l'étape de soustraction aboutissant à un enrichissement en ADNc spécifiques d'une des deux populations d'ARNm choisie.

Ces banques sont particulièrement bien adaptées aux espèces pour lesquelles peu d'informations de séquence sont disponibles puisqu'elles limitent, en principe, la redondance des ESTs et évitent le séquençage d'un grand nombre de clones. Par cette méthode, des gènes exprimés dans la zone d'élongation des racines primaires de maïs et non encore identifiés ont pu être isolés (Bassani *et al.*, 2004). Chez la pomme de terre, les banques SSH ont permis d'isoler des facteurs de réponse aux auxines dont le niveau d'expression est relié à l'activité du méristème (Favre-Rampant *et al.*, 2004) et chez la pomme, des gènes exprimés

différentiellement dans les feuilles de cultivars résistants à un pathogène ont pu être caractérisés (Degenhardt *et al.*, 2005).

Cependant, il ne faut pas oublier que les étapes de normalisation et de soustraction ne sont pas toujours optimales, et que comme les banques d'ADNc classiques, elles nécessitent un criblage pour identifier les gènes exprimés différenciellement. A titre d'exemple, chez le cacao, dans le cadre d'études de la réponse aux stress, des banques SSH ont permis d'isoler 1256 gènes dont seulement 330 (4%) correspondent à des gènes exprimés différenciellement au cours de la réaction de défense (Verica *et al.*, 2004). Un autre exemple, choisi toujours dans la thématique du stress, mais chez le peuplier, montre qu'à partir de 1344 ESTs issus de banques soustractives, 35 (3%) seulement étaient exprimés différenciellement dans les feuilles après 18h de traitement par une molécule qui préserve la plante des dégâts réalisés par les herbicides (Rishi *et al.*, 2004).

Le nombre de gènes distincts représentés dans la banque va dépendre en grande partie de la proximité des populations d'ARNm que l'on souhaite comparer. Il est bien évident que deux populations d'ARNm fortement différentes fourniront des banques avec un grand nombre de gènes représentés alors que des populations très proches donneront des banques avec peu de gènes représentés.

c) Séquençage des ESTs

Lorsque les clones sont orientés dans le vecteur, il est possible de choisir à partir de quelle extrémité (5' ou 3') la séquence des ADNc sera déterminée. Pendant longtemps, le séquençage par l'extrémité 5' était préféré à celui par l'extrémité 3'. En effet, la réaction de transcription inverse initiée par un oligo d(T) conduit à la formation d'ADNc généralement tronqués au niveau de l'extrémité 5'. La région non traduite en 5', UTR 5' (UnTranslated Region) est, par ce fait, plus courte que la région non traduite en 3' (UTR 3'). Les séquences initiées par l'extrémité 5' offrent donc une plus grande proportion de région codante. De plus, l'extrémité 3' poly-adénylée des ARNm ajoute une difficulté technique au séquençage et génère souvent des séquences de mauvaise qualité. C'est principalement par un séquençage de 5' vers 3', que Vettore (2001) a séquencé plus de 250000 EST de canne à sucre pour constituer la banque SUCEST.

Aujourd'hui, avec le progrès des techniques de séquençage, c'est le séquençage par l'extrémité 3' qui est préféré, car il offre davantage de variabilité (principalement dans l'UTR) et permet donc de distinguer les gènes paralogues. Ce fait constitue un intérêt pour les

végétaux, ceux-ci présentant fréquemment des duplications de gènes. Une solution qui commence à se répandre, grâce à la diminution des coûts de séquençage, est de séquencer les ADNc dans les deux sens. De cette manière, Pavy *et al.* (2005) ont séquencé 50000 clones de sapinette blanche (*Picea glauca*) pour obtenir 16578 séquences assemblées dont 45% de séquences complètes d'ADNc.

2) Analyse des ESTs

a) *Tri des séquences obtenues*

Il y a dix ans, on estimait jusqu'à 3%, le nombre de nucléotides incorrects lorsque la séquence d'un EST était comparée à la séquence génomique correspondante (Hillier *et al.*, 1996). Aujourd'hui, même s'ils restent encore présents, les nucléotides incorrects sont moins fréquents dans les séquences grâce aux progrès réalisés au niveau de la processivité des enzymes et à l'apparition des séquenceurs automatiques. Ces erreurs peuvent être des insertions, des délétions et des substitutions et sont causées en majorité par la transcriptase inverse, par les réactions PCR éventuelles, par la réaction de séquence et par la lecture des séquences. Ces erreurs ne peuvent pas être corrigées, mais d'autres types de malfaçon peuvent être détectés et éliminés.

Par exemple, on retrouve dans les banques des séquences de très mauvaise qualité sur la totalité de leur longueur. Ces séquences qui ne pourront être ni annotées, ni comparées aux bases de données devront être éliminées avant de commencer l'analyse. Les séquences qui comportent principalement des nucléotides indéfinis sont dues à un problème de séquençage parfois lié à la structure même de l'ADN à séquencer : extrémité 3' polyadénylée, structures en épingle...

Il faudra également filtrer des séquences contaminantes. Les contaminants sont principalement des artéfacts des processus de clonage ou de séquençage parmi lesquels, on retrouve des séquences issues du vecteur de clonage, de la bactérie hôte, des ARN ribosomiaux, des séquences mitochondriales ou chloroplastiques. Les approches pour identifier les contaminants et ainsi les éliminer sont basées sur une homologie des EST de la banque contre une liste de contaminants communs.

Après avoir éliminé les mauvaises séquences et les contaminants, il est nécessaire de débarrasser les séquences des parties correspondantes aux amorces et au vecteur. De plus, la fin, et à moindre degré, le début des séquences comprennent un certain nombre de nucléotides indéfinis (N) ou erronés. L'utilisateur doit tronquer la séquence pour ne conserver que la

partie la moins riche en nucléotides indéfinis ou erronés. Les séquences sont généralement « nettoyées » par un programme informatique qui évite ainsi un fastidieux travail. La bio-informatique est donc devenue indispensable à toute équipe qui désire générer de nombreux EST aussi bien pour le tri des séquences que, comme nous allons le voir, pour l'assemblage des ESTs et leur annotation.

b) Assemblage des ESTs

Certains EST des banques étant redondants, il est possible de les assembler afin de former des contigs, c'est-à-dire des séquences théoriquement représentatives d'un ADNc unique. Le regroupement est basé sur un pourcentage d'identité pour une longueur minimale de chevauchement entre deux ESTs. Par exemple, Ranjan *et al.* (2004) ont assemblé des EST de peuplier à l'aide du programme CAP3 (Huang and Madan, 1999) avec les seuils de 95% d'identité pour un chevauchement minimal de 50 pb. D'autres programmes très complets pour assembler les ESTs existent comme PHRAP (Green, 2002) qui a été utilisé pour assembler environ 49000 ESTs de sapinette blanche en 16578 gènes uniques (Pavy *et al.*, 2005).

Plus simplement, la construction d'une base de données à partir de tous les ESTs de la banque et la soumission de tous ces ESTs contre cette même base de données à l'aide du programme BlastN utilisé « en local » suffit pour identifier les EST représentant un même gène. C'est ce procédé qui a été utilisé par Clarke *et al.* (2003) chez le blé pour identifier les séquences identiques parmi 4433 EST. Par contre, dans cette étude, les contigs ont quant à eux été formés à l'aide du programme PHRAP.

Des résultats d'assemblage d'EST sont accessibles dans certaines bases de données, comme PlantGDB (Plant Genome Database, <http://www.plantgdb.org/>), où il est possible d'obtenir, pour plus de 100 espèces de plantes, les ESTs sous forme de contigs théoriquement représentatifs d'ADNc uniques.

c) Annotation des EST

Pour obtenir une idée générale de la fonction des gènes représentés dans les banques, les ESTs sont annotés. Comme il est impossible de procéder manuellement lorsque l'on a des dizaines de milliers d'EST à annoter, la majorité des bases de données d'ESTs sont reliées à d'autres bases de données. Par exemple, dans la base de données PlantGDB, les contigs et les séquences uniques sont comparés à la banque de protéines GenPept de NCBI en utilisant le programme BlastX.

Un moyen d'annoter les ESTs de plante est d'utiliser les données de séquences d'Arabidopsis ou de riz. Le génome de ces plantes étant quasiment séquencé et de nombreux gènes ayant une fonction attribuée, les informations d'Arabidopsis ou du riz sont très utiles. Les informations de séquences d'Arabidopsis ont été utilisées pour l'annotation d'ESTs chez le blé (Clarke *et al.*, 2003), même si les deux espèces ne sont pas très proches d'un point de vue phylogénétique.

Grâce à de nombreuses ressources disponibles dans les sites Internet de Gene Ontology (<http://www.geneontology.org/>) ou de MIPS (Munich Information centre for Proteins Sequences : <http://mips.gsf.de/>), il est également possible de trier les protéines déduites des ESTs par classe de fonction.

3) Le criblage des banques d'ADNc

Afin d'identifier des gènes exprimés différemment, les banques d'ESTs, qu'elles soient soustractives ou non, nécessitent un crible. Une multitude de méthodes ont été développées dans les années 1980 et 1990 pour réaliser ce crible. Les plus anciennes, utilisées encore parfois aujourd'hui, se basent sur l'utilisation d'isotopes radioactifs pour construire des sondes marquées qui sont hybridées avec des clones d'ADNc fixés sur membrane (Sambrook, 1989). Elles peuvent également se baser sur l'utilisation d'anticorps monoclonaux et polyclonaux (French *et al.*, 1986) ou sur la PCR (Hovens and Wilks, 1989).

Il existe aussi d'autres possibilités, plus récemment apparues, pour cribler des banques d'ESTs. L'une d'elle est la soustraction *in silico* d'ESTs. Cette dénomination a été inventée par les bio-informaticiens par analogie au *in vitro* et au *in vivo* des biologistes pour désigner un travail informatisé. Cette approche est basée sur l'abondance différentielle d'ESTs dans des banques de données. Elle repose sur le principe que la banque d'ADNc est le reflet du transcriptome et que plus un gène est exprimé, plus les ADNc correspondants seront abondants dans la banque (Lee *et al.*, 1995; Okubo *et al.*, 1992). La soustraction *in silico* est également appelée « digital northern » en analogie à la technique de northern blot. Une approche statistique a été développée pour le digital northern (Audic and Claverie, 1997) et a été appliquée sur des banques non normalisées et non soustraites d'ADNc, en particulier par Rajkovic *et al.* (2001) et Miner and Rajkovic (2003) dans le but de mettre en évidence des gènes, respectivement différemment exprimés dans les cellules germinales de souris et dans le placenta, chez l'homme. Ranjan *et al.*, 2004 ont montré qu'il est également possible d'utiliser une soustraction *in silico* et le test statistique d'Audic et Claverie sur des banques

normalisées en identifiant des gènes exprimés préférentiellement dans différents tissus de plantes modifiées au niveau du métabolisme des phénylpropanoïdes. Cependant, il ne faut pas oublier que les soustractions *in silico* réalisées à partir de clones de banques SSH peuvent être faussées par l'étape de normalisation. Nous reviendrons sur ce point dans la discussion des résultats de ce mémoire.

Autre technique de grande importance apparue dans les années 1990 : les microarrays, qui offrent la possibilité de cribler de nombreux clones en simultané. Le criblage des banques d'ESTs par la technique de cDNA microarray était déjà pressenti il y a six ans (Richmond and Somerville, 2000) comme « la » solution pour l'identification de gènes exprimés différentiellement dans un type cellulaire donné ou dans une condition environnementale donnée. Les microarrays seront abordés dans le chapitre suivant de cette analyse bibliographique (II.B).

4) Les ESTs et la mise en évidence de gènes exprimés différentiellement chez les végétaux : une alternative pour les espèces non séquencées ?

La majorité des ESTs a été initialement séquencée dans le but de fournir une alternative au séquençage complet des génomes et de constituer une base de travail pour réaliser des analyses d'expression de gènes. En effet, à ce jour, seuls trois génomes sont quasiment séquencés et assemblés : celui d'*Arabidopsis* (2000), celui du riz (2005) et celui du peuplier (Tuskan *et al.*, 2006), et quatre autres sont en cours d'obtention : celui de *Medicago truncatula*, celui du lotier, et ceux de la tomate et du maïs. Les collections d'ESTs, parfois qualifiées de « poor man's genome » (Rudd, 2003), fournissent, pour les génomes non séquencés, un point de départ pour l'identification de gènes exprimés différentiellement. En effet, un grand intérêt est porté aux ESTs par les chercheurs désireux de réaliser une analyse transcriptomique sur des espèces de moindre intérêt agronomique ou n'étant pas des modèles biologiques. L'augmentation exponentielle du nombre d'ESTs déposés dans dbEST pour les espèces dont le génome n'est pas séquencé (Figure 3), ainsi que l'augmentation du nombre d'espèces représentées en témoignent.

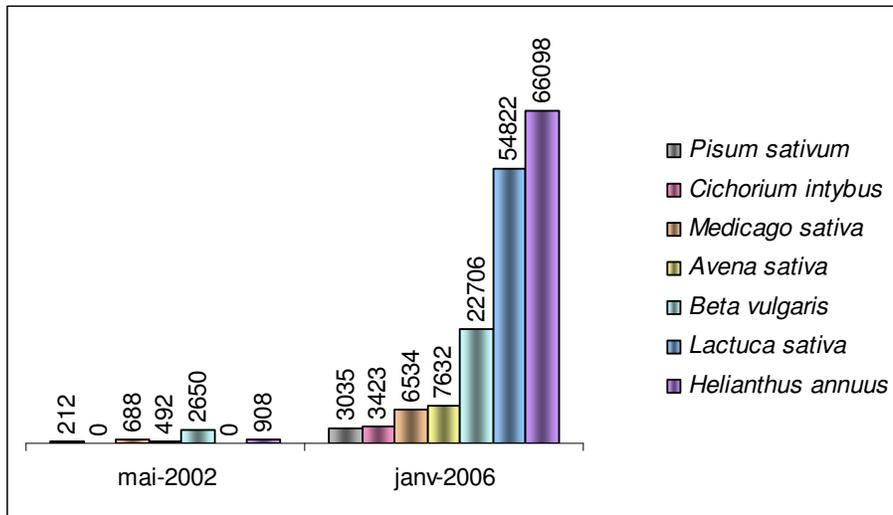


Figure 3 : Evolution du nombre d'ESTs depuis 2002 pour 6 espèces de plantes dont le génome n'est pas séquencé.

Les effectifs sont ceux de la base de données dbEST. Sur les 3423 ESTs soumis pour *Cichorium intybus* en janvier 2006, 3422 sont issus de ce travail.

Les utilisations des ESTs ne se limitent pas seulement aux analyses transcriptomiques. En effet, les outils issus de la bioinformatique, en exploitant la multitude d'EST obtenus chez les plantes, ont élargi le champ d'application des ESTs à l'annotation des génomes, à la protéomique et au développement de marqueurs moléculaires SSR (Single Sequence Repeats) ou SNP (Single Nucleotide Polymorphism) pour construire des cartes génétiques.

B. Les microarrays

Il y a plus de 30 ans, Ed Southern concevait le Southern Blot. Il démontrait alors qu'une molécule d'ADN marquée (sonde) pouvait détecter une autre molécule d'ADN fixée sur un support solide (Southern, 1975). Deux années plus tard, le northern blot, en se basant sur une hybridation ADN/ARNm ouvrait la voie de la détection et de la quantification de l'expression des gènes (Alwine *et al.*, 1977). Les mesures d'expression réalisées en hybridant des ADNc marqués radioactivement sur des clones d'ADNc fixés sur une membrane constituaient une avancée supplémentaire.

Il ne manquait alors plus qu'une étape pour donner naissance à la technique de biopuces également appelée *microarray* ou *DNA chips*. Cette étape repose principalement sur deux choses : l'utilisation de supports solides non poreux comme les lames de verre, ce qui a facilité la miniaturisation, et l'utilisation de fluorochromes en remplacement des isotopes radioactifs contraignants. Ainsi, plus de 20000 ADNc, qui constituent les sondes, peuvent être

déposés par un robot sur une lame de type microscope et hybridés avec des cibles doublement marquées (Schena *et al.*, 1995). Depuis, la technique n'a cessé d'évoluer. Elle est devenue la méthode la plus utilisée pour mesurer l'expression d'un grand nombre de gènes en simultané et est aujourd'hui une technique classique de la biologie moléculaire.

1) Puces à façon ou puces à oligonucléotides ?

Les progrès réalisés depuis dix ans au niveau de la technologie ont permis le développement de différents types de microarrays. Il en existe deux principaux : les « puces à façon » (cDNA microarray) et les puces à oligonucléotides.

a) « Puces à façon »

Les ADNc représentant les gènes d'intérêt sont amplifiés par PCR, purifiés et déposés à l'aide d'un robot sur des lames de verre (Figure 4). On peut déposer jusqu'à 20000 spots par lame.

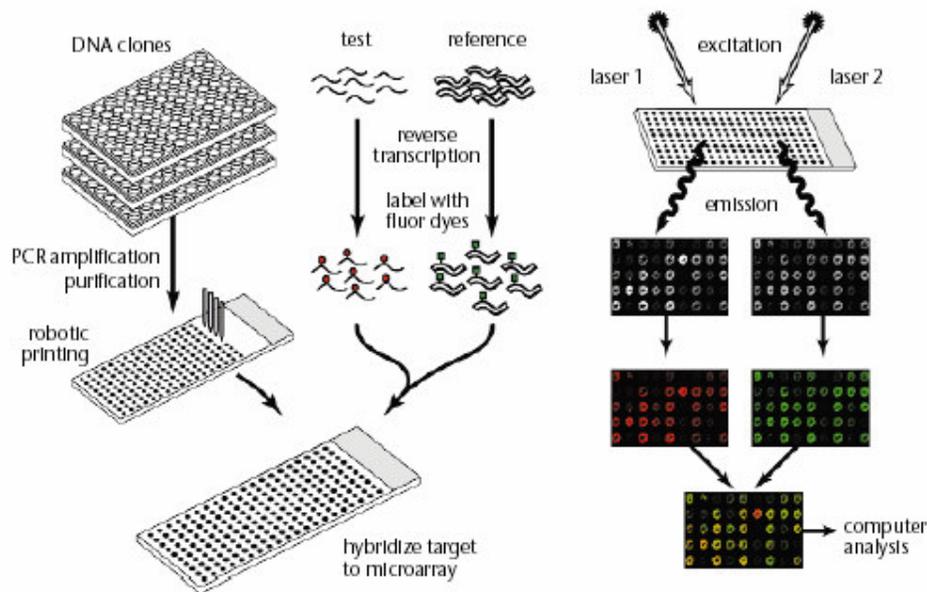


Figure 4 : Principe général des « puces à façon » d'après Duggan *et al.* (1999)

Les ARN totaux des deux conditions testées (ici test et référence) sont soumis à une réaction de transcription inverse en présence de dCTP marqués (se sont souvent des cyanines qui sont utilisées pour les marquages, cy3 pour une condition et cy5, pour l'autre condition). Les cibles fluorescentes obtenues sont mélangées, puis hybridées avec les clones déposés sur les lames. L'excitation par un laser des dCTP-cy3 ou -cy5 incorporés au sein des cibles hybridées entraîne une émission de photons à une longueur d'onde particulière et différente

pour chacune des deux cyanines (550 et 650 nm respectivement). La fluorescence émise est alors mesurée par le scanner. Les images monochromatiques obtenues pour chaque cyanine sont transformées en images pseudo-colorées puis superposées.

Les données d'une simple expérience avec la condition test marquée au cy5 et la condition référence marquée au cy3 fournissent un ratio cy5/cy3, calculé généralement en logarithme. Un ratio égal à zéro est obtenu pour un gène dont l'expression ne varie pas entre la référence et la condition test, un ratio supérieur à zéro est obtenu pour un gène préférentiellement exprimé dans la condition test et enfin un ratio inférieur à zéro est obtenu pour un gène préférentiellement exprimé dans la condition référence.

b) Puces à oligonucléotides

Les puces à oligonucléotides utilisent la synthèse d'oligonucléotides courts ou longs spécifiques des gènes d'intérêt. La technique de ce type la plus connue, mise au point par Fodor *et al.* (1991), utilise la synthèse directe d'oligonucléotides sur les lames de verre en se basant sur le principe de la photolithographie. La lame de verre porte des groupes protecteurs photo-labiles et au-dessus de ceux-ci, un photo-masque qui laisse passer la lumière sur des zones spécifiques de la lame. Ces zones sont successivement illuminées et le groupement photo-labile est alors détruit. La surface est ensuite incubée avec une solution contenant un nucléotide photo-protégé qui sera seulement fixé sur les zones activées par la lumière. Après avoir enlevé l'excès du nucléotide, un deuxième photo-masque est employé pour déprotéger d'autres zones de la lame où un autre type de nucléotide viendra s'ajouter. En répétant cette procédure, des oligonucléotides courts (25 bases) sont synthétisés à la surface de la lame. La synthèse des cibles passe par différentes étapes : à partir des ARN totaux, des ADNc, puis des ADN double brins sont synthétisés. Ces ADN servent de matrice à la production d'ARNc incorporant des nucléotides biotinylés. Après l'hybridation, la révélation se fait par incubation avec de la streptavidine couplée à la phycoérythrine. C'est cette technique qui est utilisée pour les puces à ADN Affymetrix.

c) Avantages et inconvénients

Les « puces à façon » présentent l'avantage de pouvoir être mises en place directement à partir de banques d'ADNc existantes, pas nécessairement séquencées. Dans ce cas il est possible, pour limiter le coût, de séquencer uniquement les clones correspondants aux gènes montrant une expression différentielle au cours de l'étude. Par ce fait, cette technique est

particulièrement adaptée aux espèces disposant de peu d'informations de séquence ainsi qu'aux études à budget plus limité.

Les puces à oligonucléotides permettent, quant à elles, la fabrication de microarray à haute densité (plus de 300000 spots pour 1,28 cm² (Lipshutz *et al.*, 1999)) et facilitent la production de grandes séries de lames. Cependant, une information de séquence pour chaque gène étudié est nécessaire. Les puces à oligonucléotides sont, par ce fait, plutôt utilisées pour les espèces disposant de nombreux ESTs et pour les génomes entièrement séquencés comme celui d'*Arabidopsis*. De plus, le gros inconvénient des puces à oligonucléotides est le coût de ce type de lame. Il est environ trois à cinq fois plus élevé que celui des puces à façon. A cela s'ajoute le fait qu'un énorme travail doit être réalisé en amont pour définir des oligonucléotides spécifiques de chacun des gènes. Il est possible également de synthétiser des oligonucléotides longs (50 à 70 bases) séparément et de les déposer ensuite sur lame à l'aide d'un robot, mais l'avantage des puces à haute densité est alors perdu.

Nous nous focaliserons, dans la suite de cette étude bibliographique, sur les « puces à façon » plus adaptées aux espèces dont le génome n'est pas séquencé, comme c'est le cas de la chicorée, notre modèle d'étude.

2) Les schémas expérimentaux des « puces à façon »

a) Les différents schémas

Contrairement aux puces à oligonucléotides à haute densité qui sont hybridées avec une seule population d'ADNc (voir plus haut), les « puces à façon » sont généralement hybridées avec deux populations d'ADNc. Le fait d'avoir deux conditions testées par lame permet, pour une étude donnée, de définir différents schémas expérimentaux. Il faudra donc choisir le schéma expérimental qui fournira une analyse des données et une interprétation des résultats simples et puissantes.

Il existe principalement deux types de schémas expérimentaux : le schéma direct et le schéma indirect. Pour le schéma direct (Figure 5A), les deux conditions (C₁ et C₂) marquées par des fluorochromes différents sont testées sur une même lame et le ratio d'expression entre les deux conditions est mesuré par la formule $\log_2(C_1/C_2)$. Pour le schéma indirect qui utilise une référence comme dénominateur commun (Figure 5B), les deux conditions sont testées sur deux lames différentes et le ratio d'expression est mesuré par la formule $\log_2(C_1/R) - \log_2(C_2/R)$ où R est la référence (Yang and Speed, 2002).

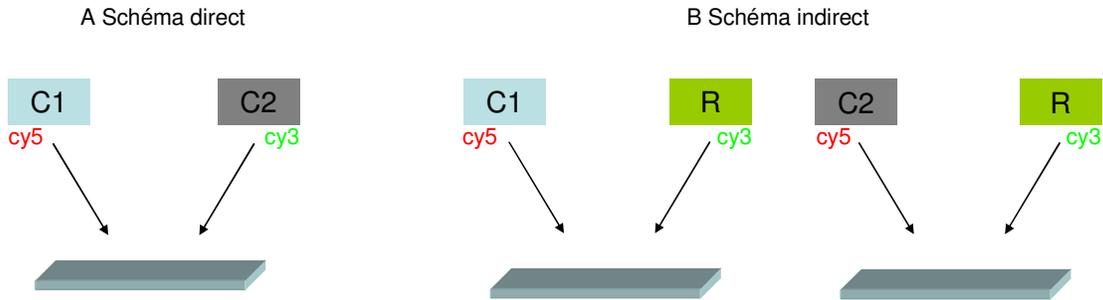


Figure 5 : Schéma expérimental direct et schéma expérimental indirect. C1 et C2 représentent les deux conditions testées et R la référence.

La précision du ratio d'expression dépend du nombre d'intermédiaires qui existent entre les deux conditions à comparer. Dans le cas présenté, le schéma direct est donc plus précis que le schéma indirect.

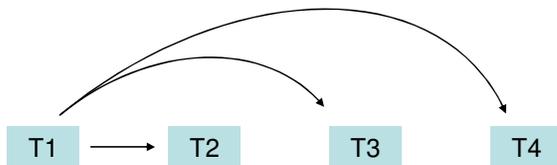
Prenons maintenant un autre exemple : dans le cas d'une étude en cinétique, le choix du schéma expérimental va dépendre de l'information souhaitée. Sur la Figure 6, le schéma direct A fait appel à une comparaison successive des points de la cinétique alors que le schéma B utilise le point initial de la cinétique comme référence contre laquelle seront comparés tous les autres points de la cinétique. Pour chacun des deux dispositifs, trois expériences sont nécessaires. Lorsque l'objectif de l'étude est de mettre en évidence des variations d'un point à l'autre, c'est le schéma direct séquentiel qui est le mieux adapté. Par contre si l'objectif est cette fois de mettre en évidence des variations entre les différents points de la cinétique et le point initial, c'est le schéma indirect qui est le plus approprié.

A Schéma direct, séquentiel



Figure 6 : Etude en cinétique par un modèle direct ou un modèle indirect. Les points de la cinétique sont indiqués T1 à T4.

B Schéma indirect, T1 comme référence commune



De façon plus générale, le schéma direct est utilisé lorsque l'étude nécessite la comparaison de peu de conditions différentes, alors que le schéma indirect est plus adapté lorsqu'au moins dix conditions sont étudiées (Novoradovskaya *et al.*, 2004). Il demandera en effet moins d'expériences à réaliser et sera plus facile à analyser puisque la précision des ratios d'expression sera identique pour toutes les conditions testées.

Il existe des variantes des schémas directs et indirects. Citons par exemple le schéma en boucle dérivé du schéma direct (Kerr and Churchill, 2001b), qui ajoute au modèle A de la Figure 6 une comparaison entre T1 et T4. Dans le modèle B présenté, c'est le point initial de la cinétique qui est utilisé comme référence mais il est également possible d'utiliser une référence externe. Cette référence peut être préparée en mélangeant les ARNm de différentes conditions biologiques, elle peut être un mélange des ADNc spottés sur les lames, ou de l'ADN génomique. L'essentiel étant qu'elle soit disponible en grande quantité, puisque comparée à chaque condition test, que les aliquots de cette référence soient homogènes et qu'un maximum de clones de la lame soient représentés (Novoradovskaya *et al.*, 2004).

b) Les répétitions expérimentales et biologiques

Le but des répétitions est d'augmenter la puissance des analyses statistiques ultérieures et de confirmer les résultats obtenus. La répétition la plus courante consiste à hybrider, sur une nouvelle lame, les mêmes cibles en échangeant les fluorochromes (dye swap). Cette répétition permet de détecter des différences éventuelles d'incorporation des deux fluorochromes dans les cibles.

Ces deux lames peuvent ensuite être à nouveau répétées. Deux types de répétitions peuvent alors être utilisées : la répétition technique, où les ARNm d'un seul échantillon ou d'un mélange d'échantillons biologiques sont utilisés pour plusieurs expériences, et la répétition biologique où différents échantillons d'ARNm, cette fois-ci non mélangés, sont utilisés pour les répétitions d'expériences. Les premières études de microarray utilisaient rarement les répétitions biologiques, mais il est indiscutable aujourd'hui qu'elles sont nécessaires pour la validité de l'expérience (Kerr, 2003a). En effet, les répétitions techniques permettent seulement d'estimer la variabilité liée à la mesure et de la corriger, alors que les répétitions biologiques permettent en plus d'estimer la variabilité liée à l'échantillon biologique.

3) Les analyses des résultats des « puces à façon »

Il y a six ans, Richmond and Somerville (2000) écrivaient ces mots : « The real heart of microarray experimentation and, ironically, the most underdeveloped part, is data analysis ». Depuis, de nombreux articles sont parus sur l'analyse des résultats de microarray et commencent à combler cette lacune.

La première méthode pour définir les gènes différentiels se basait uniquement sur l'application d'un seuil au ratio d'expression, souvent fixé empiriquement à deux fois. Cependant il est clair aujourd'hui que cette approche est inadéquate, puisqu'elle ne tient pas compte de la variabilité et n'associe pas au résultat un niveau de significativité (Budhraja *et al.*, 2003; Miller *et al.*, 2001). L'utilisation d'un seuil limite au ratio d'expression sans tenir compte de la variabilité et de la taille de l'échantillon pourrait entraîner l'identification de gènes qui ne seraient que des faux positifs et au contraire sous-estimer l'expression d'autres gènes. Une analyse statistique puissante est donc indispensable pour l'analyse des résultats de microarray. Avant de procéder à cette analyse statistique, diverses étapes dont la conversion des signaux fluorescents en données numériques et une normalisation de ces données sont nécessaires. Après l'analyse statistique, un outil de visualisation des données fournira une aide pour l'interprétation et la compréhension du sens biologique des résultats.

Afin d'assurer la vérification et l'échange des données entre les laboratoires, des standards comme les normes MIAME (Minimum Information About a Microarray Experiment) ont été définies. Les normes MIAME décrivent le minimum d'informations requises pour une expérience de microarray et sont parfois exigées dans le cadre d'une soumission d'article (Brazma *et al.*, 2001). Ces normes fixent les informations requises sur la production des échantillons biologiques et sur les méthodes utilisées pour générer les données de microarray.

a) Analyse des images

Une fois que les lames ont été hybridées, elles sont numérisées et converties en images (généralement au format TIF 16 bit). Le choix des paramètres du scanner peut affecter grandement les résultats expérimentaux (Armstrong and van de Wiel, 2004). Par exemple, il est important qu'il n'y ait pas de spots saturants, c'est-à-dire dépassant la valeur maximale d'intensité capable d'être mesurée. Les images sont ensuite converties en données numériques par un logiciel comme Imagene (BioDiscovery) ou GenePix (Molecular Devices). Pour chaque spot sur la lame, quatre informations sont fournies : le signal et le bruit de fond pour

chacun des deux fluorochromes utilisés. Les programmes d'analyse d'image vont définir les zones de signal et de bruit de fond pour chacun des spots suivant divers algorithmes. Les intensités sont généralement mesurées en calculant la moyenne ou la médiane des valeurs de chacun des pixels de la zone définie. Des critères de qualité des spots sont aussi choisis, afin d'éliminer de l'analyse ultérieure les spots non hybridés ainsi que les spots présentant un artéfact de « spotting ».

b) Normalisation des données

Les normalisations permettent de corriger différents biais expérimentaux comme des différences de quantités d'ARN utilisés pour fabriquer les deux cibles hybridées sur la lame ou des différences d'incorporation des deux fluorochromes dans les cibles. Elles corrigent aussi les biais dus aux paramètres du scanner, la variabilité qui existe entre les différentes aiguilles du robot, etc... Cette variabilité si elle n'est pas éliminée, risque de modifier la liste des gènes qui seront identifiés comme étant exprimés différemment, et peut engendrer la mise évidence de nombreux faux positifs. On distingue deux types de normalisation : la normalisation intra-lame et la normalisation inter-lame.

La normalisation intra-lame a pour but de corriger toutes les variations expérimentales qui peuvent être observées au sein d'une même lame. Par exemple, lorsque deux fluorochromes sont utilisés, la normalisation intra-lame ajuste la différence d'intensité enregistrée entre les deux fluorochromes. Notamment, le cy5 conduit à l'enregistrement d'intensités généralement plus faibles que le cy3. Une représentation des données par un graphique « MA » permet de détecter les biais d'incorporation des fluorochromes et de vérifier l'efficacité de la normalisation intra-lame. Sur la Figure 7, à gauche, le cy5 a été moins incorporé que le cy3, surtout pour les faibles valeurs d'intensité, et la normalisation intra-lame a permis la correction de ce biais (à droite).

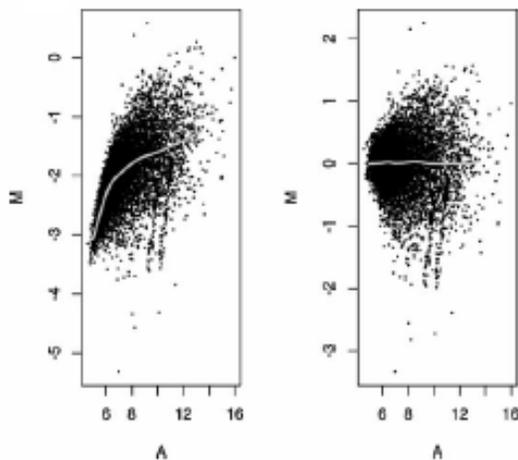


Figure 7: Normalisation intra lame
Détection (à gauche) et correction (à droite)
d'une incorporation différentielle des deux
fluorochromes à l'aide d'un graphique MA
 $M = \log_2(R/G)$; $A = \log_2(\sqrt{R \times G})$ avec R
représentant l'intensité du premier
fluorochrome (ex : cy5) et G l'intensité du
second fluorochrome (ex : cy3). La courbe est
une courbe « Lowess » utilisée pour la
normalisation de la distribution.

La méthode de normalisation intra-lame la plus répandue est la normalisation Lowess (Locally weighted least squares regression) qui est une normalisation non linéaire. Cette normalisation peut être appliquée pour la lame entière ou peut se faire par aiguille du robot afin de corriger les variations de dépôts entre les aiguilles (Yang *et al.*, 2002). Une méthode alternative pour normaliser est d'utiliser des contrôles internes qui peuvent être des gènes de ménage constitutivement exprimés ou des gènes de contrôle déposés sur les lames en combinaison avec les ARN correspondants ajoutés aux cibles avant hybridation.

La normalisation inter-lame, quant à elle, « égalise » pour chacune des lames de l'expérience les distributions des intensités des spots. Cette normalisation évite qu'une lame avec une distribution très différente de celles des autres ne modifie le résultat final de l'expérience.

Il est possible de vérifier l'homogénéité des dépôts entre les aiguilles du robot ou de vérifier l'homogénéité entre les lames d'une même expérience (Figure 8A) et éventuellement de s'assurer que la normalisation a éliminé les artéfacts en représentant les données par des graphiques de type « boîte à moustaches » ou *boxplots* (Figure 8B)

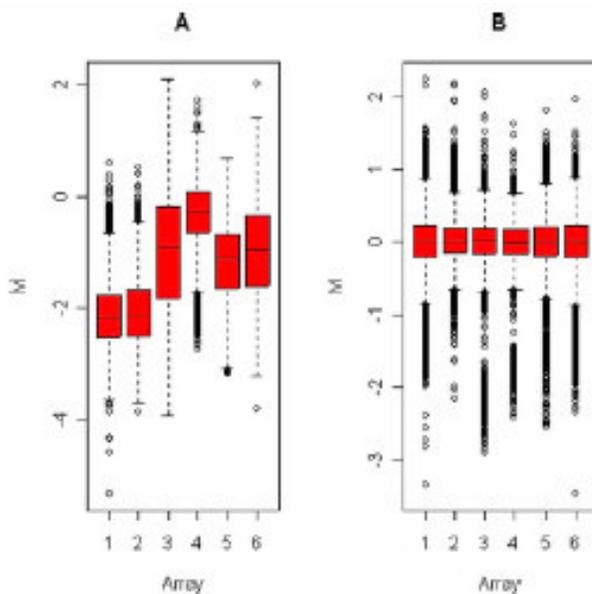


Figure 8 : Normalisation inter-lame. *Boxplots* montrant la distribution des valeurs d'intensité pour six lames d'une même expérience avant (A) et après (B) normalisation.

La boîte s'étend de sa limite inférieure définie comme le premier quartile à sa limite supérieure définie comme le troisième quartile. La ligne qui traverse la boîte représente la médiane. Les quarts supérieurs et inférieurs de la distribution sont définis par les « moustaches » qui partent de la boîte et les points représentent les valeurs extrêmes de la distribution.

D'après Armstrong and van de Wiel (2004).

c) Analyses statistiques

Afin de déterminer si les différences d'expression observées entre deux échantillons ont une signification biologique ou résultent seulement de variations aléatoires, des analyses

statistiques paramétriques ou non paramétriques sont utilisées. Le test statistique devant être adapté au schéma expérimental, il n'existe malheureusement pas de test universel.

La plupart des méthodes paramétriques assument que les données traitées ont une distribution qui suit une loi normale. Cependant, ce n'est pas le cas pour les gènes faiblement exprimés (Giles and Kipling, 2003). Les tests paramétriques les plus utilisés sont les tests-t de Student et de Welch qui comparent deux groupes et qui assument que la variance est égale ou inégale, respectivement. Des tests paramétriques basés sur les analyses de variance (ANOVA) ont également été développés (Wolfinger *et al.*, 2001) et permettent de comparer plus de deux groupes. Il existe des modèles d'ANOVA qui sont appropriés aux distributions ne suivant pas une loi normale (Kerr and Churchill, 2001a; Kerr *et al.*, 2000).

Les approches non-paramétriques les plus courantes sont les tests de Wilcoxon rank sum (Mann-Whitney) et de Kruskal-Wallis qui permettent de comparer respectivement deux et plus de deux groupes. Cependant, les méthodes non paramétriques manquent de puissance statistique pour les petits échantillons, ce qui est souvent le cas pour les expériences de microarray (Clarke and Zhu, 2006).

d) Représentation des résultats

La manière la plus simple pour représenter les résultats de microarray est le tableau de résultats qui va recenser des informations sur les gènes identifiés comme étant différentiellement exprimés. Le tableau rapporte généralement le numéro d'identifiant, le ratio d'expression, la significativité statistique et l'annotation de la séquence.

Lors d'une étude en cinétique, l'utilisation d'outils de regroupement, qui conservent toutes les données dans un seul ensemble (clustering hiérarchique, SOTA : self organizing tree algorithm...), ou d'outils de partitionnement, qui séparent les données en sous-ensembles (K-means, SOM : self organizing maps), permet d'identifier les gènes qui présentent les mêmes profils d'expression donc de mettre en évidence les fonctions biologiques mises en œuvre à chaque instant de la cinétique.

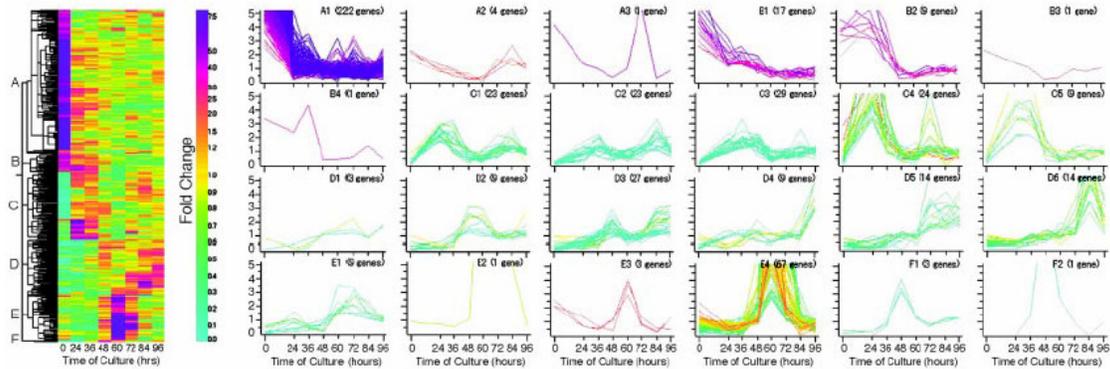


Figure 9 : Regroupement des gènes d'intérêt par profils d'expression

Figure de gauche : regroupement hiérarchique des gènes suivant leur expression au cours de la cinétique d'étude. La couleur représente une valeur d'expression relative. Les lignes représentent les différents gènes, alors que les colonnes représentent les différents temps étudiés au cours de la cinétique de 0 à 96 h. Figures de droite : les graphiques représentent les profils d'expression de gènes de certains groupes définis par le regroupement hiérarchique. D'après Demura *et al.* (2002).

4) Validation technique des microarrays

Une fois que les gènes différentiels ont été identifiés, certains résultats des microarray doivent être validés par une autre méthode indépendante comme le northern blot ou la PCR en temps réel. L'objectif est de confirmer que l'expression différentielle détectée peut être retrouvée par une autre technique et de montrer la robustesse de l'analyse microarray utilisée.

En général, les ratios d'expression mesurés par northern blot sont assez similaires à ceux mesurés par microarray. Cependant, Taniguchi *et al.* (2001) ont montré, lors d'une étude comparative, que pour 9% des gènes testés, les microarrays ne détectaient pas les changements d'expression observés par northern blot. La PCR en temps réel (RT-PCR quantitative) est, quant à elle, beaucoup plus sensible que les microarrays (Czechowski *et al.*, 2004). Par exemple, sur 256 gènes identifiés par microarray comme étant exprimés différemment au cours d'une symbiose mycorhizienne chez le riz, 224 ont pu être confirmés par PCR en temps réel. Cependant même si l'expression différentielle des gènes montrait le même profil, les ratios obtenus par la PCR en temps réel étaient beaucoup plus grands que ceux résultant des microarrays.

Ces résultats mettent en évidence les limites des microarrays, notamment pour la mise en évidence de transcrits peu abondants comme ceux des facteurs de transcription

(Czechowski *et al.*, 2004). Les limites de la technique seront abordées plus en détail dans la discussion des résultats du mémoire.

5) Exemples d'études menées par microarray chez les plantes

L'utilisation des microarrays pour explorer l'expression des gènes chez les végétaux a explosé depuis 1995, date de la première étude basée sur cette technique chez *Arabidopsis* (Schena *et al.*, 1995). Le nombre de publications basées sur une analyse par microarrays le montre : plus de 200 références d'articles portant sur les plantes et montrant le terme « microarray » dans le titre sont référencés dans PubMed (15 juin 2006).

Le développement chez les plantes est généralement étudié suivant trois schémas (Schnable *et al.*, 2004) : le premier consiste à comparer l'expression des gènes à un stade de développement donné entre deux génotypes. Le second consiste à comparer, pour un même génotype, différents stades d'un processus de développement ou différents organes ou tissus. Pour le troisième, afin d'étudier des facteurs environnementaux qui influencent le développement, des plantes ayant subies ou non un stimuli exogène sont analysées.

Les processus de développement sont contrôlés à la fois par une succession de programmes génétiques et par les réponses aux stimuli environnementaux. De nombreux aspects des développements végétatifs, reproducteurs ou liés à un stimulus environnemental ont été étudiés par microarray, cependant, nous nous limiterons ici simplement à la présentation de quelques exemples (voir Tableau 3).

Etude bibliographique

Type de développement	Espèce	Processus étudié	Référence
Végétatif	Soja	Embryogenèse somatique	(Thibaud-Nissen <i>et al.</i> , 2003)
	Sapinette blanche	Effets du polyéthylène glycol au cours de l'embryogenèse somatique	(Stasolla <i>et al.</i> , 2003)
	Pin	Formation de racines adventives	(Brinker <i>et al.</i> , 2004)
	Zinnia	Transdifférenciation de cellules du mésophyle en cellule du xylème	(Demura <i>et al.</i> , 2002)
	Peuplier	Sénescence automnale	(Andersson <i>et al.</i> , 2004)
Reproducteur	Arabidopsis	Identification de gènes régulés par <i>APETALA3</i> et <i>PISTILLATA</i>	(Zik and Irish, 2003)
	<i>Lotus Japonicus</i>	Développement des anthères et des pistils	(Endo <i>et al.</i> , 2002)
	Arabidopsis	Développement des fleurs	(Wellmer <i>et al.</i> , 2006)
Réponse à des stimuli environnementaux	Arabidopsis	Réponse aux auxines exogènes	(Sawa <i>et al.</i> , 2002)
	Arabidopsis	Réponse à un traitement par l'ABA	(Seki <i>et al.</i> , 2002)
	Arabidopsis	Réponse aux gibbérellines exogènes	(Ogawa <i>et al.</i> , 2003)
	Arabidopsis	Réponse aux brassinostéroïdes exogènes	(Goda <i>et al.</i> , 2002)
	Arabidopsis	Régulation de gènes des voies de synthèse de l'auxine, de l'éthylène et de l'acide abscissique en réponse au 2,4-D	(Raghavan <i>et al.</i> , 2006)
	Orge	Réponse à la sécheresse et au stress salin	(Oztur <i>et al.</i> , 2002)
	Orge	Stress osmotique	(Ueda <i>et al.</i> , 2004)
	Arabidopsis	Stress salin, osmotique, et froid	(Kreps <i>et al.</i> , 2002)
	Blé	Comparaison entre le blé d'hiver et le blé de printemps en réponse aux basses températures	(Gulick <i>et al.</i> , 2005)
	Arabidopsis	Carence en phosphate	(Wu <i>et al.</i> , 2003)
	Arabidopsis	Privation en oxygène des racines	(Klok <i>et al.</i> , 2002)
	Arabidopsis	Réponse aux fortes lumières	(Kimura <i>et al.</i> , 2003)
	Maïs	Réponse aux UV-B	(Casati and Walbot, 2004)
	Arabidopsis	Réponse au cadmium	(Herbette <i>et al.</i> , 2006)
	Peuplier	Effets d'agents protecteurs	(Rishi <i>et al.</i> , 2004)
	Ivraie vivace	Réponse à la vernalisation	(Ciannamea <i>et al.</i> , 2006)
	Sorgho	Défense contre un pathogène	(Zhu-Salzman <i>et al.</i> , 2004)
Luzerne	Symbiose avec <i>Sinorhizobium meliloti</i>	(Mitra <i>et al.</i> , 2004)	
Tournesol	Défense contre un champignon	(Alignan <i>et al.</i> , 2006)	

Tableau 3 : Exemples d'études par microarray de différents processus de développement chez les plantes

a) Développements végétatifs et reproducteurs

Différents aspects du développement végétatif, qui commence au moment de l'embryogenèse et qui s'arrête au début de la floraison, ont été étudiés chez les plantes. Par exemple chez le zinnia, une étude menée au cours de la transdifférenciation de cellules du mésophile en cellules du xylème a conduit à l'identification de nombreux gènes impliqués dans ce processus (Demura *et al.*, 2002). Les profils d'expression ont permis de réaliser une forte avancée des connaissances dans ce mécanisme. Les résultats montrent notamment une corrélation temporelle entre les variations d'expression des gènes et les évènements cytologiques observés. Chez le soja, l'embryogenèse somatique a été étudiée par le biais d'analyses microarray (Thibaud-Nissen *et al.*, 2003). Le regroupement des EST par profils d'expression similaires sur la cinétique d'étude a permis d'établir la chronologie des évènements moléculaires qui ont lieu au cours du mécanisme. Parmi les autres aspects du développement végétatif, les travaux menés sur la sénescence automnale chez le peuplier (Andersson *et al.*, 2004) ou les études sur la formation des racines adventives chez le pin en sont des exemples (Brinker *et al.*, 2004).

Le développement reproducteur débute au moment de la formation des organes reproducteurs et se termine avec la pollinisation. Chez *Arabidopsis*, les gènes *APETALA3* (*AP3*) et *PISTILLA* (*PI*) jouent un rôle déterminant lors de la formation des pétales et des étamines. L'analyse du transcriptome par microarray de mutants *AP3* et *PI* et de lignées qui surexpriment ces gènes, a montré, en comparaison à des lignées sauvages, que ces gènes ne contrôlent qu'un nombre limité de gènes directement impliqués dans la formation des pétales et des étamines (Zik and Irish, 2003). L'analyse par microarray menée à partir d'anthères mûres et immatures ainsi que de pistils, a permis de regrouper les profils de gènes exprimés différemment au cours de la formation de ces organes. Ces travaux sont un point de départ pour l'identification de gènes impliqués dans la formation des organes reproducteurs (Endo *et al.*, 2002). Plus récemment, le développement des fleurs a également été étudié chez *Arabidopsis* (Wellmer *et al.*, 2006).

b) Les signaux environnementaux

En plus des rôles joués par les programmes génétiques, l'influence des *stimuli* environnementaux sur le développement des végétaux a été largement étudiée. Des études par microarray impliquant l'application d'hormones exogènes telles que les auxines (Sawa *et al.*, 2002), l'acide abscissique (Seki *et al.*, 2002), l'acide gibberelique (Ogawa *et al.*, 2003) et les

brassinostéroïdes (Goda *et al.*, 2002) ont non seulement permis d'identifier des gènes spécifiquement régulés par ces hormones, mais ont permis également de mettre en évidence la synergie qui existe entre ces hormones. Plus récemment une étude menée chez *Arabidopsis* a montré l'influence du 2,4D sur les autres hormones (Raghavan *et al.*, 2006).

Une très grande variété de stress a également été étudiée par microarray, que ce soient des stress osmotiques, salins ou froid (Kreps *et al.*, 2002; Oztur *et al.*, 2002), la carence en phosphate (Wu *et al.*, 2003), la sécheresse (Oztur *et al.*, 2002), la privation d'oxygène au niveau des racines (Klok *et al.*, 2002), les fortes lumières (Kimura *et al.*, 2003), les UV (Casati and Walbot, 2004) ou les stress biotiques (Zhu-Salzman *et al.*, 2004).

III. La chicorée, plante modèle de l'embryogenèse somatique

A. La chicorée

La chicorée est une plante dicotylédone, de la famille des Astéracées qui se développe dans les régions à climat tempéré. La chicorée sauvage est une plante vivace connue depuis environ six mille ans. Elle était appréciée des plus grands médecins grecs et égyptiens pour ses vertus thérapeutiques. Elle était, par exemple, employée pour les maux de ventre, de foie ou des reins.

Le genre *Cichorium* est représenté par six espèces principalement retrouvées en Europe et en Asie : *C. intybus* L., *C. endivia* L., *C. spinosum* L., *C. glandulosum* Boiss, *C. bottae* Defl et *C. calvum* Schultz-Bip, les deux premières citées étant cultivées. *C. endivia* présentant un cycle de vie annuel est cultivée pour la consommation des feuilles en salade (chicorées scaroles et frisées). *C. intybus* peut présenter suivant les individus différents cycles de vie : annuel, bisannuel ou vivace. On distingue deux variétés, la variété *foliosum*, aussi appelée Witloof, cultivée pour la production d'endives (bourgeon étioilé) et la variété *sativum*, aussi appelée chicorée industrielle, qui est cultivée pour la récolte de ses racines utilisées pour la production d'inuline (polyfructosane) ou pouvant être torréfiées. L'espèce *C. intybus* représente un intérêt économique important pour le nord de la France, la Belgique ainsi que pour les Pays-Bas. En France (principalement dans le Nord-Pas-de Calais), la production d'endives correspond à une surface agraire de 12000 ha et la production de racines de chicorée pour l'année 2002 a été de 168750 tonnes. Cependant une baisse de 30% de cette production pour l'année 2005 a été observée.

Dans le nord de la France, *C. intybus* var *foliosum* est semée en avril/mai. La première année, la plante développe une rosette de feuilles, puis une racine pivotante tubérisée (septembre à octobre). Durant cette phase de croissance végétative, la racine accumule des réserves azotées et carbonées. A la fin de cette période, le cycle naturel peut être arrêté pour la production d'endives. Le cas échéant, après vernalisation, la plante érige au printemps une hampe florale ramifiée qui peut atteindre un mètre de haut. Cette hampe florale porte un grand nombre de capitules composés de fleurs bleu-violet.

B. Contexte scientifique

Depuis 20 ans, l'embryogenèse somatique chez la chicorée est étudiée au laboratoire à travers des aspects cytologiques, physiologiques et moléculaires en utilisant l'hybride interspécifique '474' (*Cichorium intybus* L. x *C. endivia* L.). A l'inverse des variétés

d'importance agronomique de chicorée, des embryons somatiques peuvent être obtenus facilement, rapidement et en grande quantité à partir de différents explants de l'hybride '474' lorsqu'ils sont placés sous agitation constante, à l'obscurité et à 35°C dans un milieu de culture de Murashige et Skoog contenant de faibles concentrations d'auxines et de cytokinines (Dubois *et al.*, 1991). L'embryogenèse somatique, dans ces conditions, est directe et unicellulaire (Blervacq *et al.*, 1995). En utilisant l'hybride '474' différentes méthodes ont été appliquées pour cloner des gènes exprimés au cours des phases précoces de l'embryogenèse somatique. Par le crible de banques d'ADNc, plusieurs clones correspondant à des gènes différentiellement exprimés au cours des phases précoces du processus ont été identifiés, le plus surprenant étant un ADNc codant une hémoglobine non symbiotique (Hendriks *et al.*, 1998). Plusieurs protéines ont été montrées comme étant spécifiquement accumulées dans le milieu de culture au cours de l'embryogenèse somatique (Boyer *et al.*, 1993; Hilbert *et al.*, 1992). Le séquençage partiel de quelques unes de ces protéines a permis de mettre en évidence l'expression différentielle de gènes codant une β -1,3 glucanase, une chitinase et une osmotin-like protein (Helleboid *et al.*, 2000b). L'expression spécifique d'au moins un gène codant une β -1,3 glucanase au cours de l'embryogenèse somatique a été ensuite confirmée après que les ADNc aient été obtenus par RT-PCR en utilisant des amorces dégénérées (Helleboid *et al.*, 2000a). Enfin, la technique de « mRNA differential display » a été utilisée pour isoler des ADNc codant une protéine de transfert de glutathion (Galland *et al.*, 2001) et une protéine fixatrice de GTP (Randoux *et al.*, 2002). Les gènes correspondants à ces ADNc sont exprimés différentiellement dans les explants de l'hybride '474' au cours de l'embryogenèse somatique et pas ou très faiblement exprimés dans les explants de variétés non embryogènes de *C. intybus* L. soumises aux mêmes conditions de culture, ce qui suggère que l'expression de ces gènes peut être reliée à l'embryogenèse somatique et non aux différents stress appliqués par les conditions de culture (Galland *et al.*, 2001; Helleboid *et al.*, 2000a; Hendriks *et al.*, 1998; Randoux *et al.*, 2002). Cependant ces résultats ne pouvaient être complètement concluants dans la mesure où le fond génétique de l'hybride '474' est très différent de celui des variétés non embryogènes utilisées. De plus, l'hybride '474' étant quasiment stérile il était difficile de transférer le caractère embryogène à d'autres variétés.

Plus récemment, une variabilité génétique pour la capacité des plantes à former des embryons somatiques a été découverte dans une population hongroise de *C. intybus* L. appelée 'Koospol' (M-C. Quillet, B. Delbreil and B. Deprez, résultats non-publiés). A partir du crible des plantes de cette population, des génotypes embryogènes et non embryogènes ont

été identifiés. La plante K59 a été sélectionnée comme étant un génotype fortement embryogène, et quelques graines ont été obtenues après avoir répété des autofécondations de ce génotype habituellement fortement auto-incompatible. Parmi les plantes issues de ces graines, la plante C15 a été sélectionnée comme étant un génotype non embryogène. Partageant un fond génétique commun, les génotypes K59 et C15 sont alors apparus comme étant un matériel de choix pour l'étude de l'embryogenèse somatique et notamment pour l'identification de gènes exprimés différemment au cours des phases précoces du processus.

C. Objectifs de la thèse

Dans cette étude, il s'agissait dans un premier temps d'utiliser les génotypes K59 (embryogène) et C15 (non embryogène) afin de construire des banques d'ADNc par la technique de banques soustractives (SSH) (Diatchenko *et al.*, 1996). Les banques devaient être construites à partir d'explants des deux génotypes prélevés au quatrième jour de culture d'embryogenèse somatique, c'est-à-dire au moment de l'apparition des premières divisions chez le génotype embryogène. Théoriquement, la banque embryogène devait être enrichie en gènes préférentiellement exprimés chez K59 et la banque non embryogène en gènes préférentiellement exprimés chez C15. Il s'agissait ensuite de séquencer un certain nombre de clones. Les ESTs obtenus devaient constituer une source d'information de séquences pour la chicorée qui ne disposait alors que de très peu de séquences déposées dans les bases de données.

Le second objectif de la thèse était de construire, à partir des ESTs des deux banques, les premières lames de cDNA microarray pour la chicorée et d'entreprendre des hybridations avec des cibles obtenues à partir d'ARNm issus des génotypes K59 et C15. Suite à ces tests, une étude en cinétique au cours des six premiers jours de culture d'embryogenèse somatique pour les deux génotypes K59 et C15 devait être réalisée. Ce travail devait permettre la mise en évidence de gènes exprimés différemment au cours du processus et l'identification des profils d'expression de gènes afin de relier les événements moléculaires aux études cytologiques déjà effectuées au laboratoire.

Certains gènes différemment exprimés devaient alors être validés par PCR en temps réel et un modèle expliquant la réponse aux conditions d'embryogenèse somatique chez les génotypes embryogène et non embryogène devait être construit. Ces résultats devaient alors participer à une meilleure compréhension du processus d'embryogenèse somatique et

Etude bibliographique

des mécanismes qui permettent la reprogrammation des cellules somatiques et qui confèrent aux cellules végétales leur totipotence.

- Matériel et méthodes -

I. Matériel végétal

Pour cette étude, ce sont les génotypes de chicorée K59 et C15, respectivement embryogène et non embryogène, qui ont été utilisés. Le génotype K59, issu de la population Hongroise de *C. intybus* L. appelée « Koospol », a été sélectionné pour sa forte capacité à produire des embryons somatiques (Etude bibliographique, III.B). Le génotype C15, issu de l'autofécondation de K59, se caractérise, quant à lui, par son incapacité à former des embryons lorsqu'il est placé dans les conditions de culture d'embryogenèse somatique.

II. Méthodes

A. Conditions de culture

1) Propagation clonale du matériel végétal

a) Maintien et multiplication du matériel par embryogenèse somatique

Le génotype embryogène K59 a été régénéré par embryogenèse somatique à partir de racines de plantes *in vitro*. L'explant racinaire, prélevé sur des plantes cultivées *in vitro*, présente l'avantage, par rapport à l'explant foliaire, de produire des embryons qui vont se détacher facilement de l'explant (Figure 10).

Après 8 jours d'induction dans le milieu M17S20 à 35°C sous agitation constante, à l'obscurité, les embryons, obtenus à partir des racines, ont été transférés et cultivés 10 à 15 jours dans un milieu liquide de Heller (H5) (Annexe 1) contenant 5 g.l⁻¹ de saccharose, sous agitation, à la lumière pendant 16 h à 24°C, et à l'obscurité pendant 8 h à 20°C. Les embryons somatiques ont ensuite été transférés en boîte de Pétri sur du milieu solide H5, en conservant la même photopériode, afin d'y poursuivre leur développement. Les plantules ont alors été repiquées en tubes contenant du milieu H5 solide et cultivées dans les mêmes conditions que précédemment. Après 6 à 8 semaines, les plantules peuvent être acclimatées en serre ou être utilisées pour la multiplication du génotype.

Matériel et méthodes

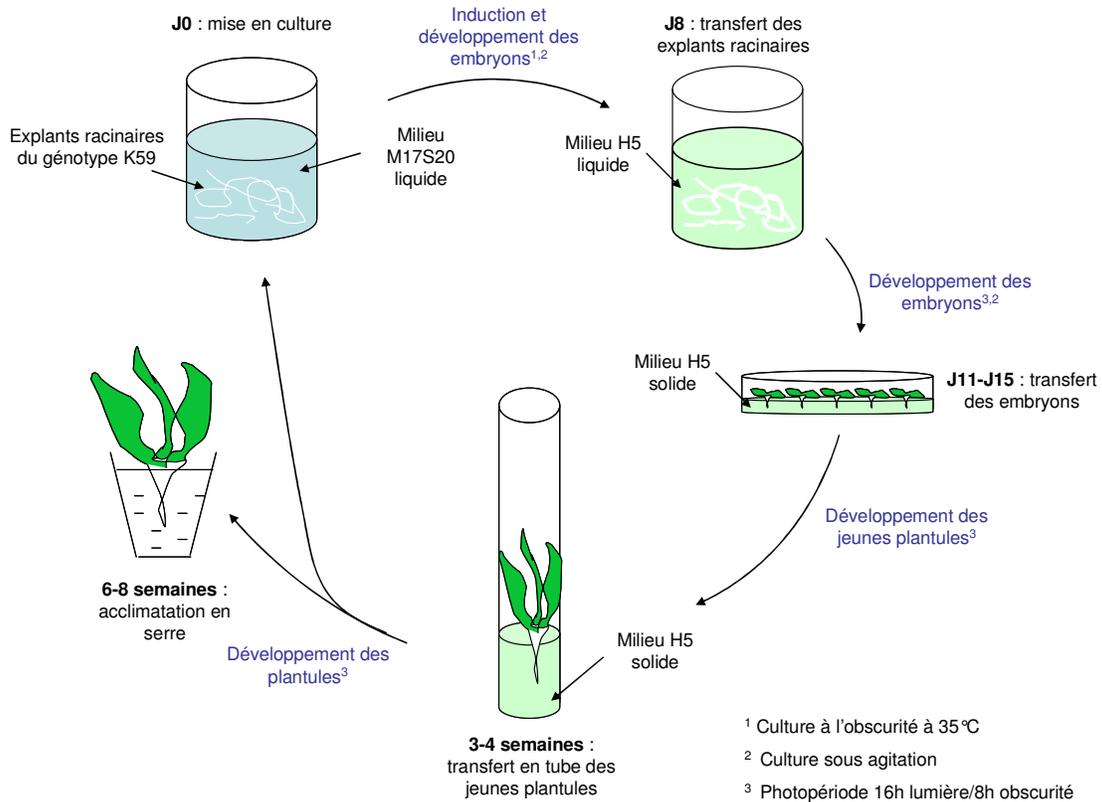


Figure 10 : Propagation clonale du génotype K59 par embryogenèse somatique

b) Maintien et multiplication par organogenèse

Le génotype non embryogène C15, ne pouvant être régénéré par embryogenèse somatique, a été multiplié par organogenèse à partir de feuilles de plantes cultivées en serre ou de hampes florales de plantes *in vitro* (Figure 11). Les feuilles de plantes cultivées en serre, prélevées et aseptisées comme décrit ultérieurement (II.A.2)), et les hampes florales ont été mises en culture en boîte de Pétri sur un milieu de bourgeonnement Bg 0,5 (Annexe 1) avec la même photopériode que les plantules II.A.1)a). Les bourgeons obtenus ont été transférés sur un milieu H5 pour leur enracinement en conservant la photopériode. Après 6 à 8 semaines de culture, les plantules pouvaient être acclimatées en serre.

Matériel et méthodes

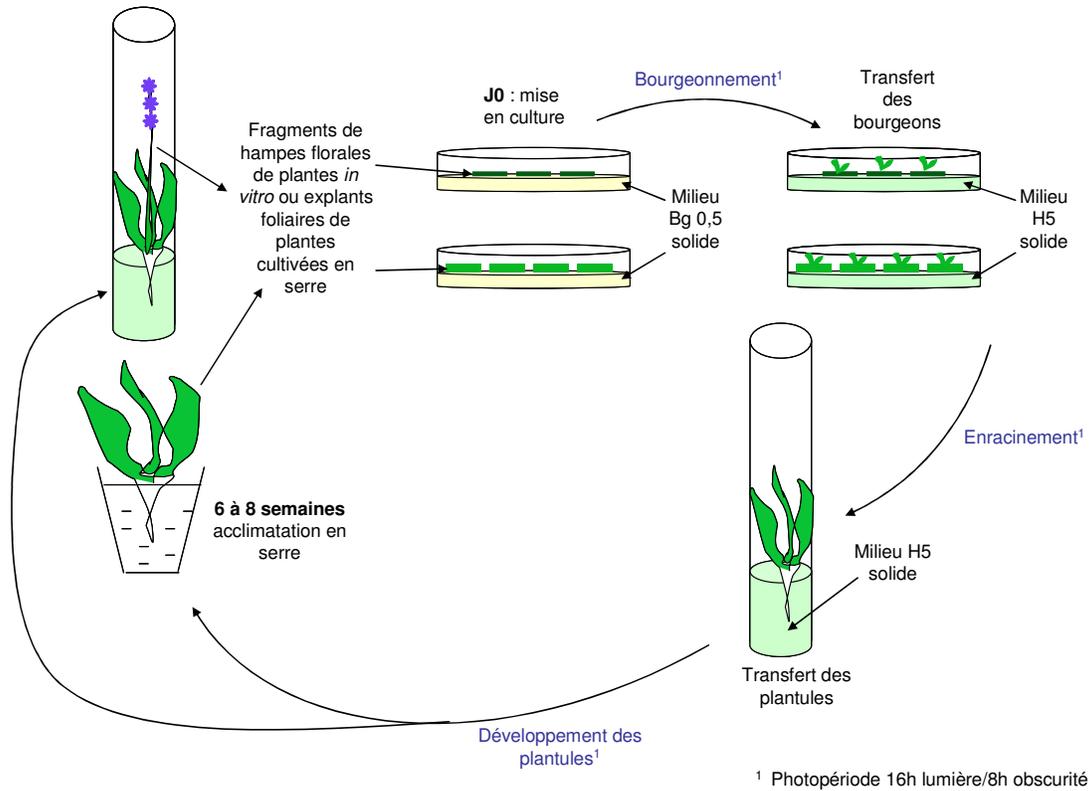


Figure 11 : Propagation clonale du génotype C15 par organogénèse

2) Induction de l'embryogenèse somatique à partir d'explants foliaires

Des chicorées de génotype embryogène (K59) et non embryogène (C15) issues de la multiplication clonale *in vitro* (II.A.1) ont été cultivées en serre. Les feuilles des plantes des deux génotypes ayant atteint un stade six feuilles ont été collectées puis stérilisées à l'aide d'un bain de Domestos 12% pendant 20 min, suivi de trois rinçages de 1 min à l'eau stérilisée. Après aseptisation, les feuilles ont été coupées en fines lamelles (2 cm × 0,2 cm) en ôtant la nervure centrale et l'apex (Figure 12). Une culture est constituée de quinze explants placés dans un milieu liquide M17S20 (Annexe 1) et cultivés à 35°C sous agitation constante, à l'obscurité. Les cultures ont ensuite été collectées entre 0 et 6 jours d'induction afin de construire les banques soustractives, fabriquer les cibles de microarray et réaliser les analyses de PCR en temps réel. Les explants ont été conservés à -80°C jusqu'à l'extraction des ARN.

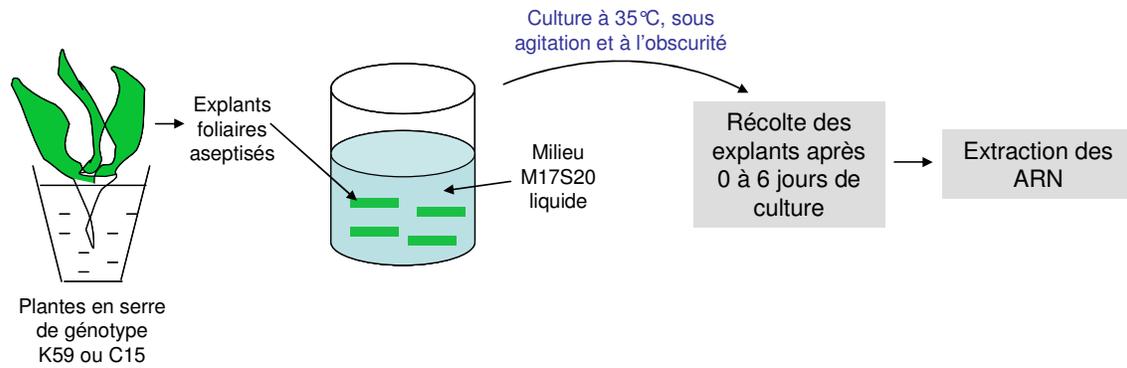


Figure 12 : Culture en conditions d'embryogenèse somatiques

B. Extraction d'ARN totaux.

Les ARN totaux des deux génotypes ont été extraits à l'aide du kit Tri Reagent d'Euromedex. Environ 100 mg de tissus foliaires broyés dans l'azote liquide ont été homogénéisés dans 1 ml de Tri Reagent. Après 5 min d'incubation à température ambiante, 200 µl de chloroforme ont été ajoutés et le mélange a été vigoureusement agité pendant 15 sec. Après 15 min d'incubation à température ambiante, le mélange a été centrifugé 15 min à 12000 g et à 4°C. La phase aqueuse, contenant les ARN, transférée dans un nouveau tube, ont été ajoutés 500 µl d'isopropanol. Après 10 min d'incubation à température ambiante, le mélange a été centrifugé 10 min à 12000 g et à 4°C. Le surnageant a été éliminé et 1 ml d'éthanol 70% a été ajouté au culot. Après 5 min de centrifugation à 7500 g, l'éthanol a été éliminé, le culot séché à l'air et les ARN repris dans 20 µl d'eau traitée au DMPC.

La qualité des ARN a été vérifiée par électrophorèse capillaire en utilisant le Bioanalyser 2100 d'Agilent et la quantité a été déterminée par une mesure de densité optique à 260 nm. Les ARN ont été conservés à -80°C jusqu'à leur utilisation.

C. Construction des banques soustractives d'ADNc.

1) Obtention des ADNc soustraits et normalisés

Pour la construction des banques, les explants ont été prélevés après quatre jours de culture en conditions d'embryogenèse somatique. Les ARN polyadénylés ont été extraits à l'aide du kit Quick Messenger RNA de Talent en suivant les instructions du fabricant.

Deux banques soustractives ont été construites à l'aide du kit PCR Select cDNA Subtraction de Clontech en suivant le protocole fourni. Pour la synthèse des ADNc simples brins, 4 µg d'ARN polyadénylés ont été utilisés. La banque embryogène (E) a été obtenue en

utilisant, comme condition « test », les ADNc provenant du génotype embryogène K59 et comme condition « référence », ceux du génotype non embryogène C15 (Figure 13 et Figure 14). Afin de fabriquer la banque non embryogène (NE), les conditions « test » et « référence » ont été inversées.

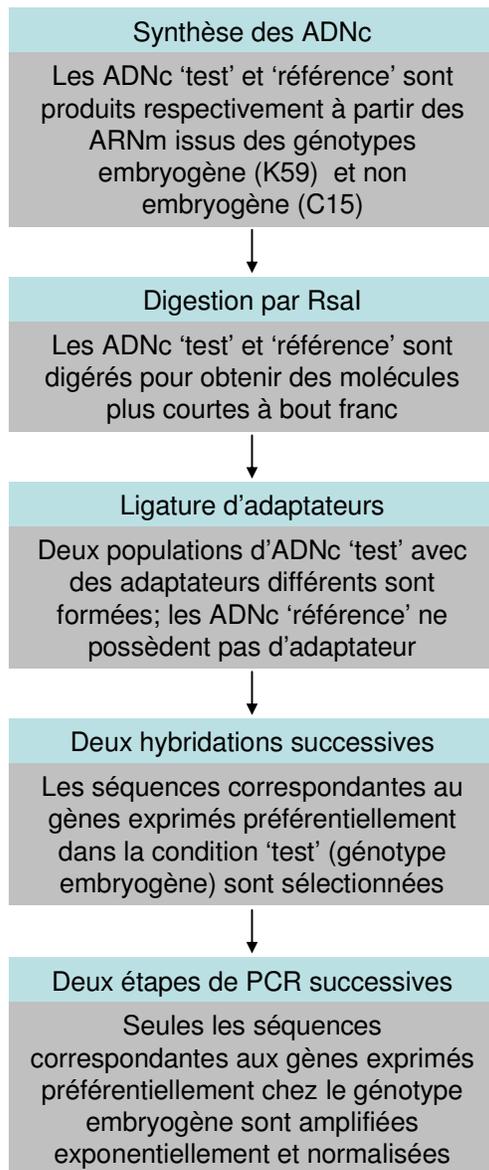


Figure 13 : Schéma général de la fabrication de la banque soustractive embryogène

La construction procède en 7 étapes : une étape de synthèse d'ADNc, une digestion totale par l'enzyme de restriction RsaI, une ligature d'adaptateurs aux extrémités des ADNc issus du génotype embryogène, deux hybridations successives entre les ADNc « test » issus du génotype embryogène et les ADNc « référence » issus du génotype non embryogène, et enfin deux étapes de PCR successives afin d'enrichir et de normaliser les séquences représentatives des gènes préférentiellement exprimés chez le génotype embryogène (K59).

Matériel et méthodes

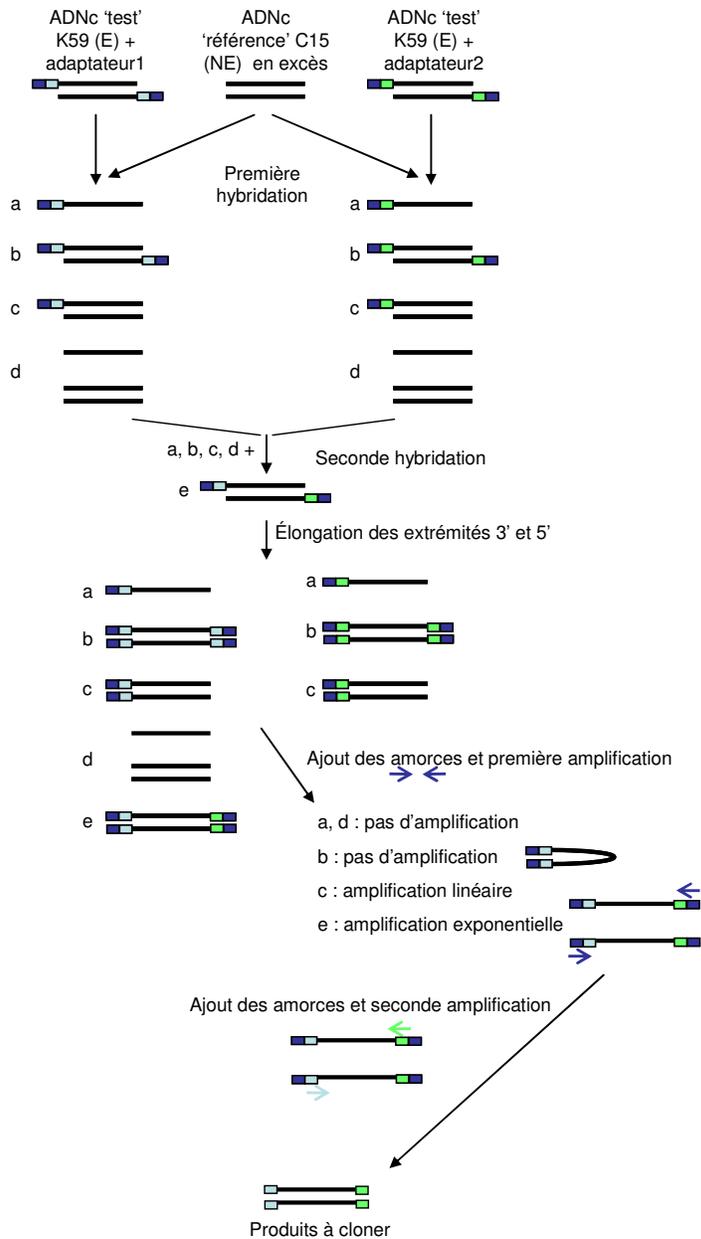


Figure 14 : Soustraction et normalisation au cours de la fabrication de la banque embryogène
 Au cours d'une première hybridation, les deux lots d'ADNc 'test' obtenus à partir du génotype embryogène, K59 (E) et qui présentent des adaptateurs différents mais avec une partie terminale commune, sont mélangés avec un excès de 30 fois d'ADNc 'référence' obtenus à partir du génotype non embryogène, C15 (NE). Les ADNc 'a' simples brins représentent les gènes préférentiellement exprimés chez K59. La seconde hybridation aboutit à la formation des ADNc doubles brins 'e' formés par l'hybridation des ADNc 'a'. Après élongation des extrémités 5' et 3', seuls les ADNc 'e' vont pouvoir être amplifiés exponentiellement. Ils seront enrichis par rapport aux autres ADNc, correspondant à des gènes non exprimés différenciellement, et normalisés, par deux étapes successives de PCR. La première PCR utilise des amorces complémentaires de la partie commune externe des deux types d'adaptateurs, alors que les amorces utilisées au cours de la seconde PCR (Nested PCR) sont complémentaires de la région la plus interne des adaptateurs.

2) Clonage des ADNc

Les ADNc amplifiés des deux banques ont été clonés dans un vecteur pGEM-T (Promega) (Figure 15). Six conditions de ligation ont été réalisées : 50 ng de vecteur plasmidique ont été mis en présence de 25, 50, 75, 100, 150 ou 300 ng d'ADN. Au mélange contenant les plasmides et l'ADN à cloner ont été ajoutés 1 µl d'ADN ligase du phage T4, 5 µl de tampon 2×, et de l'eau afin d'obtenir un volume final de 10 µl. La ligation a été réalisée pendant une nuit à 4°C.

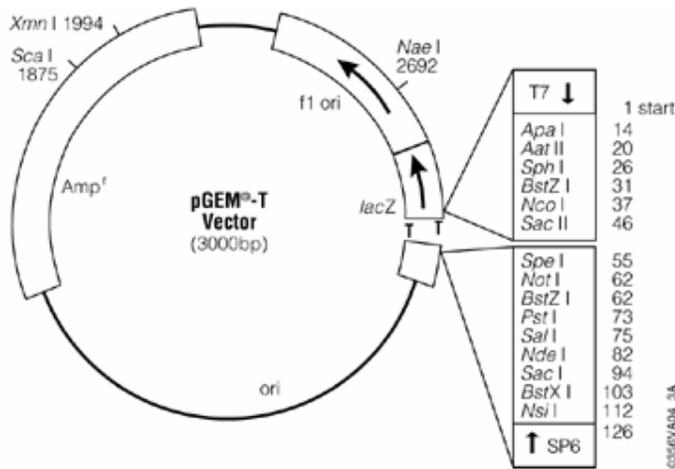


Figure 15 : Carte du vecteur pGEM-T de Promega

La présence du gène *Amp^r* confère à la bactérie une résistance à l'ampicilline ce qui autorise la discrimination entre les bactéries transformées et celles qui ne les sont pas, tandis que la présence du point d'insertion au sein du gène *lacZ* permet la discrimination des bactéries contenant un insert de celles qui n'en possèdent pas. Le vecteur présente un site d'hybridation de l'amorce universelle T7 qui a été utilisée pour le séquençage des clones.

3) Transformation de bactéries compétentes

Cinquante microlitres de bactéries compétentes *E. coli* JM109 (Promega) ont été transformées par un choc thermique à 42°C pendant 45 sec avec 5 µl de produits de ligation. Les bactéries ont ensuite été cultivées 1 h à 37°C dans 1 ml de milieu SOC (2% de tryptone ; 0,5% d'extraits de levure ; 10 mM de NaCl ; 2,5 mM de KCl ; 20 mM de Mg²⁺ ; 20 mM de glucose, pH 7,0), étalées sur un milieu LB 1× (1% de tryptone ; 0,5% d'extraits de levure, 10% de NaCl ; 15% d'agar, pH 7,0) contenant 100 µg.ml⁻¹ d'ampicilline, 80 µg.ml⁻¹ de X-gal et 0,5 mM d'IPTG puis cultivées une nuit à 37°C.

Les colonies blanches ont été mises en culture dans des plaques à 96 puits contenant chacun 200 µl de LB 1× et 50 µg.ml⁻¹ d'ampicilline. Les plaques ont été placées à 37°C et

sous agitation à 180 tr.min⁻¹. Du glycérol, 30% v/v, a alors été ajouté dans chaque puit et les plaques ont été agitées 1 h à 180 tr.min⁻¹ avant d'être stockées à -80°C.

Avant d'entreprendre le séquençage des clones, une préculture a été réalisée afin d'augmenter le nombre de bactéries par culture. Pour cela, quelques microlitres des cultures bactériennes ont été transférés dans des plaques dont chaque puit contenait 200 µl de LB 1× et 50µg.ml⁻¹ d'ampicilline. Les plaques ont été placées une nuit à 37°C, sous agitation à 180 tr.min⁻¹. Pour la culture bactérienne proprement dite, les clones ont à nouveau été repiqués dans des plaques à 96 puits profonds contenant 1,5 ml de LB 2× et 50µg.ml⁻¹ d'ampicilline puis placés en culture une nuit à 37°C et sous agitation à 180 tr.min⁻¹.

4) Séquençage

Les plasmides ont été extraits des cultures bactériennes à l'aide du kit Plasmid Miniprep96 de Millipore. Le séquençage des clones a été réalisé à l'aide d'un séquenceur ABI à 96 capillaires, par l'entreprise Genoscreen, plateforme de séquençage, sur le site de l'Institut Pasteur de Lille. L'amorce universelle T7 a été utilisée pour initier les réactions de séquence (Figure 15).

D. Analyse des ESTs

1) Traitement des séquences

A partir des ESTs obtenus, trois traitements successifs des séquences ont été réalisés. Un premier tri a consisté en l'élimination des séquences de mauvaise qualité, c'est-à-dire des séquences qui comportent un grand nombre de nucléotides indéfinis. Au cours d'un second traitement, les parties correspondant au vecteur et aux amorces de PCR ont été enlevées. Enfin, un second tri a consisté en l'élimination des séquences ribosomales.

2) Regroupement des EST identiques

Les ESTs identiques des banques E et/ou NE ont été regroupés en OC (Original Cluster) à l'aide de deux critères successifs. Dans un premier temps, une base de données composée de tous les ESTs a été construite. Ces mêmes ESTs ont ensuite été soumis à une comparaison contre la base de données formée en utilisant le programme BlastN inclus dans le logiciel BioEdit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). La probabilité associée à la qualité de l'alignement des séquences (E-value) devait être inférieure ou égale à 1.10⁻³⁰ pour que des séquences soient considérées comme appartenant à un même OC. Les résultats de cette recherche ont été analysés à l'aide d'un programme codé en langage Perl (Annexe

numérique 1). Dans un second temps, la validité de chaque OC « primaire » comportant au moins deux séquences a été vérifiée par l'intermédiaire d'un alignement de séquences réalisé avec le logiciel BioEdit. Les séquences devaient comporter 95% d'identité pour un chevauchement minimal de 150 paires de bases pour appartenir au même OC. Dans le cas contraire, les OC « primaires » ont été divisés en groupes distincts.

3) Assemblage des ESTs en contigs

Afin d'assembler les ESTs en contigs, les ESTs ont été comparés, en utilisant les programmes BlastN ou BlastX, aux ESTs assemblés en contigs de différentes plantes de la famille des Astéracées (BlastN), et après traduction, aux séquences protéiques de la base de données non redondantes de GenBank et aux séquences codantes traduites d'*Arabidopsis* (BlastX) (voir résultats). Un programme codé en Perl a été construit pour analyser les fichiers de sortie de Blast (Annexe numérique 2).

4) Analyse fonctionnelle

Les séquences ont été annotées en utilisant les résultats des comparaisons effectuées à l'aide du programme BlastX, de tous les ESTs des deux banques contre les séquences protéiques non redondantes de la base de donnée GenBank. A chaque EST soumis a été associé la protéine dont la séquence montrait la plus forte probabilité d'alignement avec l'EST traduit. La probabilité de l'alignement devait être inférieure ou égale à 1.10^{-5} pour qu'une correspondance puisse être établie.

De la même façon que précédemment, en utilisant le programme BlastX, des comparaisons des ESTs ont été menées cette fois contre les séquences codantes traduites d'*Arabidopsis*. Les résultats ont permis d'associer à chaque EST, un gène d'*Arabidopsis*. La classe fonctionnelle du gène d'*Arabidopsis* était alors retrouvée à l'aide du catalogue fonctionnel de MIPS (Munich Information Center for Proteins Sequences) (<http://mips.gsf.de/projects/funcat>) et attribuée à l'EST de chicorée correspondant.

5) Soustraction *in silico* des ESTs

La soustraction *in silico* d'ESTs appelée aussi « digital northern » est basée sur le fait que plus un gène est exprimé, plus les ADNc correspondants vont être abondants et ont donc plus de chance d'être clonés. Le test de probabilité de Audic and Claverie (1997) (Figure 16) a été utilisé afin de prédire les gènes différentiellement exprimés entre les deux génotypes utilisés pour la fabrication des banques soustractives.

$$p(y|x) = \left(\frac{N_2}{N_1} \right)^y \frac{(x+y)!}{x!y! \left(1 + \frac{N_2}{N_1} \right)^{(x+y+1)}}$$

Figure 16 : Equation mathématique de Audic and Claverie (1997) utilisée pour la soustraction *in silico* des ESTs

Cette formule donne la probabilité P(y/x) pour un gène donné représenté x fois dans la banque 1 d'être représenté y fois dans la banque 2. N1 et N2 représentent les nombres totaux d'ESTs de chacune des deux banques.

E. Microarrays

1) Préparation des produits PCR et dépôt des sondes sur les lames

Six microlitres des cultures bactériennes conservées dans du glycérol ont été transférés dans 100 µl de tampon TTE (1% de triton ×100, 20 mM de Tris-HCl, 2 mM d'EDTA, pH 8.0) contenu dans des plaques à 96 puits. Les plaques ont été incubées 10 min à 95°C et centrifugées 5 min à 1200 g. Les inserts contenus dans les plasmides ont été amplifiés à l'aide des amorces NP1¹ et NP2² (Nested primer 1 et 2), complémentaires des adaptateurs utilisés au cours de la fabrication des banques soustractives et situés de part et d'autre des ADNc.

Pour chaque clone, trois amplifications indépendantes ont été réalisées dans un volume de 100 µl contenant le tampon de l'enzyme ThermoPol Detergent Free Reaction Buffer 1× (Biolabs), 2 mM MgCl₂, 0,25 mM de chaque dNTP, 0,25 µM de chacune des amorces, 2,5 U de Taq DNA Polymerase (Biolabs) et 7 µl de la matrice d'ADN contenue dans le tampon TTE. Le programme d'amplification débutait par 3 min à 94°C suivie de 35 cycles composés de trois étapes : 30 sec à 94°C, 30 sec à 60°C et 90 sec à 70°C. La réaction se terminait par une étape de 10 min à 70°C.

Les produits des 3 PCR réalisées pour chaque clone ont été mélangés, puis purifiés à l'aide du kit Multiscreen-PCR de Millipore. La qualité de l'amplification a été vérifiée par une électrophorèse sur gel d'agarose et l'ADN produit a été quantifié par une mesure de densité optique à 260 nm. L'ADN a ensuite été réduit à sec à l'aide d'un Speed Vac (Savant Instruments), puis repris dans du tampon de « spotting » contenant 0,1 M de MES et 20% de DMSO afin d'obtenir une concentration finale moyenne de 0,5 µg.µl⁻¹.

¹ 5'—TCGAGCGGCCCGCCCGGGCAGGT—3'

² 5'—AGCGTGGTCGCGGCCGAGGT—3'

Les produits PCR purifiés ont été déposés trois fois sur des lames Codelink Activated Slides (Amersham) de 25×75 mm par un robot Affymetrix 417 Arrayer. Le dépôt des produits PCR a été réalisé à 21-22°C et à 60-70% d'humidité.

Après dépôt, les lames ont été conservées à 95% d'humidité pendant une nuit, puis traitées avec une solution de blocage (50 mM de Trizmahydrochloride, 50 mM de Tris(hydroxyméthyl)aminomethane et 0,3% d'éthanolamine, pH 9) contenant 0,1% de SDS pendant 30 min à 50°C. Les lames ont ensuite été lavées par deux bains successifs d'eau distillée à température ambiante, suivis d'un bain de SSC 4× contenant 0,1% de SDS pendant 60 min à 50°C, puis de différents bains d'eau distillée : 1 min à température ambiante, 2 min à 95°C et deux bains successifs de 2 min dans de l'eau distillée à température ambiante. A la suite du dernier bain, les lames ont été séchées par une étape de 5 min de centrifugation à 180 g.

2) Fabrication des cibles

Pour chaque cible, 10 µg d'ARN total, extraits comme précédemment (II.B), ont été soumis à une réaction de transcription inverse en présence de dCTP marqués avec les cyanines 3 ou 5 (cy3 ou cy5) (Amersham). Les ARN et 3 µg d'oligo d(T) (Roche) ont été placés dans un volume total de 16 µl puis dénaturés pendant 5 min à 70°C, et refroidis sur glace. Aux ARN et oligo d(T) ont été ajoutés 10 µl d'un mélange contenant le tampon 5× de la transcriptase inverse (Invitrogen) (250 mM de Tris-HCl (pH 8,3), 375 mM de KCl, et 15mM de MgCl₂), ainsi que 30 mM de dithiothreitol, 0,5 mM de dATP, dGTP et dTTP (Roche), 0,2 mM de dCTP (Roche) et 16 U de RNAsine (Promega). Au volume total ont été ajoutés 2µl de dCTP-cy3 ou de dCTP-cy5 suivant les cibles. Le mélange a été incubé 2 min à 42°C avant d'ajouter 400 U de SuperscriptII (Invitrogen). La réaction de transcription inverse a été réalisée par une incubation de 2 h à 42°C. Les ARN ont ensuite été dégradés par un ajout de 5 µl de NaOH 1 N, suivi par une incubation de 15 min à 65°C. L'arrêt de la réaction a été réalisé par ajout de 12,5 µl de Tris-HCl 1 M, pH 7,4.

Les cibles marquées au cy3 ou au cy5 ont alors été mélangées suivant les plans d'expérience et purifiées à l'aide du kit PCR Cleaning Kit (Qiagen). Le volume total obtenu a été réduit jusqu'à environ 8 µl à l'aide d'un Speed Vac (Savant Instruments). La qualité des sondes obtenues a été contrôlée par un spectre d'absorption de 220 à 700 nm.

3) Hybridation

A chaque paire de cibles ont été ajoutés 25 µl de tampon d'hybridation contenant 40% de formamide, du Denhardt's 5×, 50 mM de Tris-HCl (pH 7,4), 0,1% de SDS, 1 mM de pyrophosphate de sodium, et du SSC 5×. Les ADNc marqués ont été dénaturés pendant 3 min à 95°C avant d'être déposés sur les lames, recouvertes ensuite d'une lamelle. Les lames ont été placées dans des chambres d'hybridation (Corning), contenant dans des cupules, 25µl de tampon d'hybridation, afin de maintenir une humidité constante au sein de la chambre. Les chambres d'hybridation, une fois fermées, ont été incubées 14-16 h dans un bain-marie à 42°C.

Après hybridation, les lamelles ont été retirées dans un bain de SSC 4×. Les lames ont ensuite été lavées par deux bains successifs de 5 min à 42°C dans du SSC 2× contenant 0,1% de SDS, par un bain de 1 min à température ambiante dans du SSC 0,2×, et par un bain d'1 min à température ambiante dans du SSC 0,1×. Après ces lavages, qui ont pour but d'éliminer les hybridations aspécifiques et limiter ainsi le bruit de fond, les lames ont été séchées par 1 min de centrifugation à 180 g.

4) Acquisition des données et tri des spots

Les lames hybridées ont été numérisées par un scanner GMS418 (Affimetrix) et une pseudo-image représentant la fluorescence émise par les fluorochromes a été générée pour chacun des deux fluorochromes. Pour tous les spots, la moyenne et la médiane des intensités du signal ont été déterminées par le logiciel ImaGene (BioDiscovery). La moyenne et la médiane du bruit de fond autour de chaque spot ont également été mesurées.

5) Analyse des données

Les résultats obtenus ont été enregistrés dans la base de données BASE, et pour chaque lame, une normalisation intra-lame 'LIMMA LOWESS' intégrée dans R (Yang *et al.*, 2002) a été réalisée par aiguille de dépôt afin de corriger les biais éventuels entre les aiguilles du robot (voir résultats).

Des analyses de variance, réalisées à partir du \log_2 du ratio calculé à partir du signal enregistré pour la condition test et du signal enregistré pour la condition référence, ont été utilisées afin de détecter les signaux engendrés par des biais de fluorochromes et de calculer un ratio d'expression entre les conditions testées (voir résultats).

F. RT-PCR en temps réel

1) Définition des amorces

Les amorces sens et anti-sens spécifiques de chacun des gènes étudiés ont été définies à l'aide du logiciel Beacon Designer de Biosoft. Les paramètres fixés pour la définition des amorces ont été : la température d'hybridation ($60^{\circ}\text{C} \pm 0,5^{\circ}\text{C}$), la longueur des amorces (de 18 à 30 pb) et la taille de l'amplicon (de 75 à 200 pb). Les séquences de toutes les amorces définies et utilisées au cours de cette étude sont données dans l'Annexe 2.

2) Transcription inverse

La transcription inverse a été réalisée à l'aide du kit iScript cDNA synthesis de Bio-Rad. Un microgramme d'ARN a été ajouté à 4 μl de tampon (comprenant un mélange d'oligo d(T) et d'hexamères de séquences aléatoires) et à 1 μl de transcriptase inverse. Les ADNc simples brins ont été synthétisés dans un volume total de 20 μl placé à 25°C pendant 5 min puis à 42°C pendant 30 min. La réaction a ensuite été inactivée par une incubation à 85°C pendant 5 minutes.

3) Amplification

L'amplification a été réalisée en plaque de qualité optique à 96 puits (ABgene) dans un volume de 20 μl comprenant 375 nM de chacune des amorces, 10 nM de fluorescéine (référence interne), 10 μl du mélange Quantitech SYBR green concentré deux fois (Qiagen), contenant du SYBR Green I, 5 mM de MgCl_2 et de la Taq polymérase Hot Start, et 5 μl d'une dilution au $1/5^{\text{ème}}$ des produits de la transcription inverse.

L'amplification a été réalisé dans un thermocycleur iCycler iQ de Bio-Rad et comportait une étape de 10 min à 95°C afin d'activer l'ADN polymérase, suivie par 50 cycles constitués d'une étape de dénaturation de 30 sec à 95°C et d'une étape d'hybridation et d'élongation des amorces de 45 sec à 60°C . L'intensité de la fluorescence émise par le SYBR green a été mesurée à la fin des phases d'élongation. A la fin de l'amplification, pour s'assurer que le signal de fluorescence n'était pas causé par des dimères d'amorces, mais bien par un seul produit d'amplification, une courbe de fusion a été réalisée par étapes de $0,5^{\circ}\text{C}$ de 55°C à 95°C .

4) Analyse des données de PCR en temps réel

Les données de PCR en temps réel ont été analysées en suivant la méthode du $\Delta\Delta C_T$ précédemment décrite par (Livak and Schmittgen, 2001) et adaptée au laboratoire par Matthieu Duban (Duban, 2004).

La méthode du $\Delta\Delta C_T$ n'est applicable que si l'efficacité de l'amplification est proche de 100%. De ce fait, afin de pouvoir justifier de son utilisation, l'amplification de chacun des couples d'amorces définis a été vérifiée. Pour cela, une gamme de dilution de matrice d'ADNc (1/30, 1/60, 1/120 et 1/240) a été amplifiée dans les mêmes conditions que précédemment (II.F.3)). Cette amplification a permis de tracer une droite de régression déduite d'un graphique exprimant le C_T en fonction du logarithme de la quantité relative initiale. Chacun des couples était validé si l'efficacité de l'amplification³ (E) était comprise entre 0,85 et 1,15 et si le coefficient de détermination (R^2) de la droite de régression était supérieur à 0,985.

Pour la mesure de l'expression des gènes cibles, le ΔC_T a été calculé pour chaque génotype comme la différence entre le C_T attribué à l'ADNc cible et le C_T attribué à l'ADNc de contrôle. Le gène de contrôle utilisé code une actine de chicorée et l'ADNc correspondant a été isolé au laboratoire (Poulain, 2003) (GenBank DY800534). Le $\Delta\Delta C_T$ a été défini comme la différence entre le ΔC_T obtenu pour le génotype embryogène et le ΔC_T obtenu pour le génotype non embryogène.

³ $E=10^{(-1/p)}-1$ où p est la pente de la droite de régression déduite du graphique exprimant le C_t en fonction du logarithme de la quantité relative initiale

- Résultats et discussion -

Partie I : Fabrication et analyse des banques soustractives embryogène et non embryogène

A. Aspects cytologiques et choix du matériel végétal

Lorsque les explants foliaires des génotypes K59 et C15 sont cultivés dans les conditions d'embryogenèse somatique, on observe la formation d'embryons à partir des explants du génotype K59, mais pas à partir de ceux du génotype C15. Des cultures d'embryogenèse somatique réalisées au laboratoire par Matthieu Duban (2004), et les coupes cytologiques des explants réalisées par Sophie Da Silva (2004) montrent qu'au bout de quatre jours de culture, les explants du génotype K59 présentent de nombreuses cellules en cours de réactivation, des cellules réactivées ainsi que quelques cellules en division (Figure 17). Les cellules en cours de réactivation (Figure 17a) sont caractérisées par un noyau présentant un nucléole volumineux et des chloroplastes en position voisine du noyau. Les cellules réactivées (Figure 17a et b) présentent quant à elles un noyau en position centrale duquel part un réseau de microtubules formant des trabécules cytoplasmiques en forme d'étoile. Après 4 jours de culture, la proportion de cellules en divisions par rapport à celles en cours de réactivation et celles réactivées augmente nettement. Ainsi, nous intéressant principalement aux phases les plus précoces de l'embryogenèse somatique, le jour 4 de culture est apparu comme l'instant où les explants du génotype K59 devaient présenter l'expression maximale de gènes qui sont impliqués dans les phases précoces du processus et qui permettent la transition des cellules de l'état somatique à l'état embryonnaire.

Au bout de 4 jours de culture en conditions d'embryogenèse somatique, les explants du génotype C15 montrent, quant à eux, uniquement des cellules qui paraissent en cours de réactivation et, ceci, en bien plus petite quantité que celles observées chez le génotype embryogène (Figure 17c et d).

A partir des résultats de cette étude cytologique, il a été conclu que les ARN issus des explants des génotypes K59 et C15 prélevés au jour 4 de culture d'embryogenèse somatique devaient refléter l'expression différentielle des gènes qui conduit les deux génotypes à réagir de façon différente aux conditions de culture d'embryogenèse somatique. C'est pourquoi, ce sont les ARN issus des explants prélevés au jour 4 de culture qui ont été choisis pour la fabrication des banques soustractives embryogène et non embryogène.

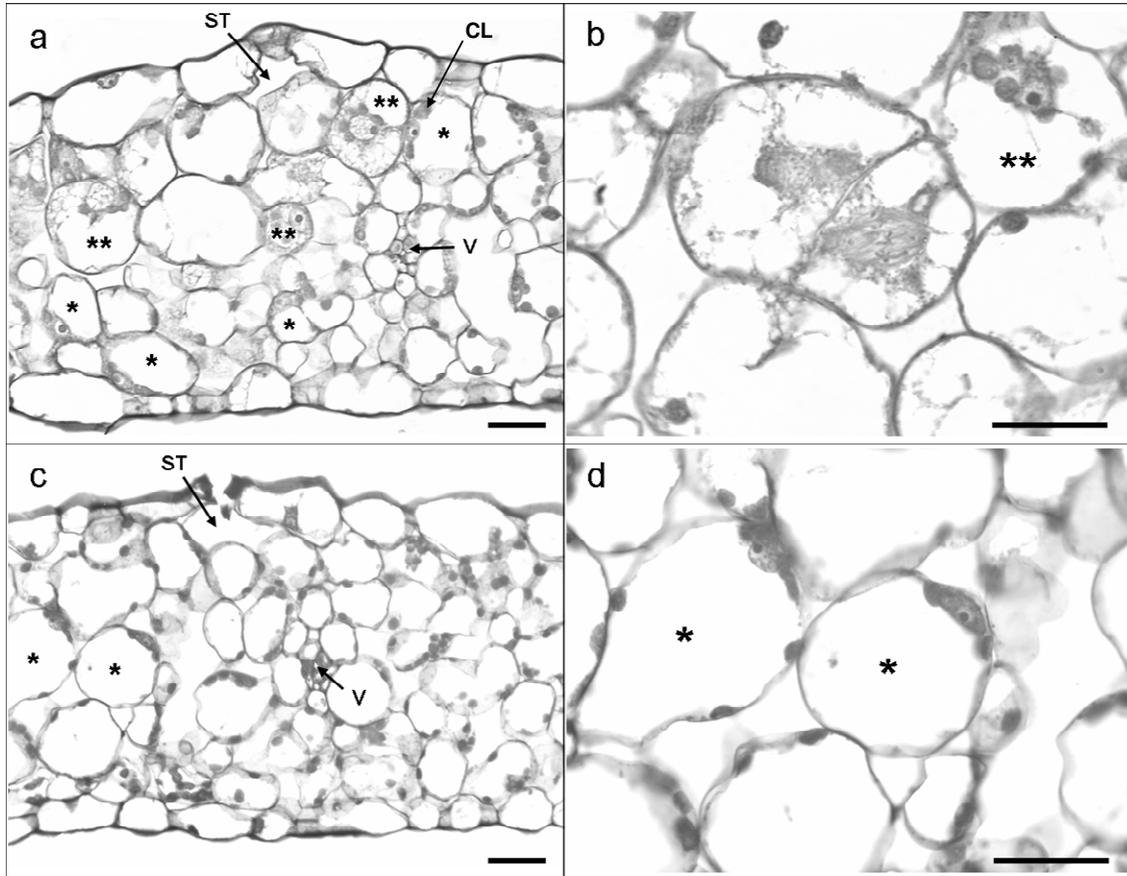


Figure 17 : Observations cytologiques d'explants foliaires des géotypes K59 (a et b) et C15 (c et d) placés 4 jours en conditions de culture d'embryogenèse somatique. Des cellules en cours de réactivation (*) et réactivées (***) sont observées chez le géotype K59, ainsi que des cellules en division (b), alors que chez le géotype non-embryogène, ce sont uniquement des cellules en cours de réactivation qui sont observées (c et d). ST : stomate ; CL : chloroplaste ; V : vaisseaux. Les barres représentent 30 μ m.

Pour le géotype embryogène, six cultures indépendantes en conditions d'embryogenèse somatique, contenant chacune 15 explants, ont été mélangées afin d'obtenir la quantité d'ARN suffisante pour la fabrication des banques. Pour chaque culture, un contrôle a été réalisé afin de vérifier la réactivité des explants aux conditions de culture. Pour cela, quelques explants ont été laissés dans les conditions de culture d'embryogenèse somatique pendant 8 jours puis transférés en milieu H5 (voir matériel et méthodes) afin de poursuivre le développement des embryons somatiques. Ce sont les cultures dont les contrôles présentaient le plus d'embryons somatiques visibles qui ont été choisies pour la fabrication des banques. Pour le géotype non-embryogène, les explants à jour 4 contenant généralement moins d'ARN, ce sont 9 cultures qui ont été nécessaires pour atteindre la quantité d'ARN requise.

B. Fabrication des banques SSH

1) Vérification des ADNc des deux banques

A partir des ARN issus des explants des génotypes K59 et C15, deux banques soustractives ont été construites : une banque embryogène (E) théoriquement enrichie en ADNc correspondant à des gènes préférentiellement exprimés chez le génotype embryogène, et une banque non embryogène (NE) théoriquement enrichie en ADNc représentant des gènes préférentiellement exprimés chez le génotype non embryogène.

A l'issue de la seconde PCR (Nested PCR), lors de la fabrication des deux banques (Figure 14), les produits de l'amplification ont été soumis à une électrophorèse sur gel d'agarose. Pour les deux banques, les « smears » observés sur la photo du gel après migration des ADNc (Figure 18), montrent qu'une grande diversité de taille d'ADNc (de 200 à 1500 pb) a été amplifiée au cours de la fabrication des banques. On note également la présence d'une bande à 700 pb qui pourrait représenter un (des) ADNc fortement abondant(s) dans les banques E et NE.

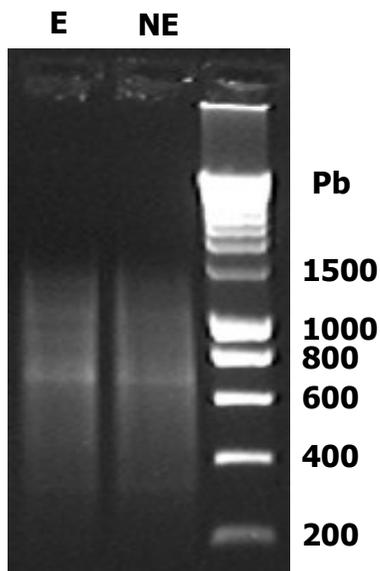


Figure 18 : Vérification de la taille des ADNc des banques embryogène (E) et non embryogène (NE) après migration sur gel d'agarose

2) Clonage des ADNc des banques E et NE

Pour la ligation, la quantité de produits de PCR nécessaire est dépendante de la taille du fragment à cloner. Les fragments à cloner étant de tailles différentes, nous avons décidé d'utiliser pour le clonage une gamme de produits de PCR allant de 25 à 300 ng de produits de PCR afin de cloner un maximum de fragments de tailles différentes. A partir de toutes les conditions de ligation (25, 50, 75, 100, 150 et 300 ng de produits PCR), les bactéries

compétentes ont été transformées. Au total, 5500 colonies bactériennes représentatives des 6 conditions de ligature ont été repiquées dans les plaques de culture, puis ont été conservées à -80°C, après addition de glycérol.

Afin de vérifier que les différentes ligatures réalisées avaient permis le clonage de toute la diversité de taille des ADNc des banques, des PCR ont été réalisées sur deux plaques de 96 puits pour chacune des conditions de ligature, en utilisant les amorces NP1 et NP2 complémentaires des adaptateurs situés de part et d'autre des inserts. L'électrophorèse sur gel d'agarose des produits PCR en présence d'un marqueur de taille a permis, après déduction de la longueur des amorces NP1 et NP2, de déterminer la longueur des inserts. La distribution par classe de taille des produits PCR montre que la taille des ADNc clonés ne semble pas différente pour les six conditions de ligature (Figure 19). Par ailleurs, les différentes conditions de ligatures ne semblent pas avoir entraîné le clonage préférentiel d'ADNc de certaines tailles, la majorité des ADNc ayant une taille comprise entre 200 et 600 pb.

Le séquençage a été réalisé pour un total de 2800 clones de la banque E et de 2200 clones de la banque NE, représentatifs des six conditions de ligature utilisées.

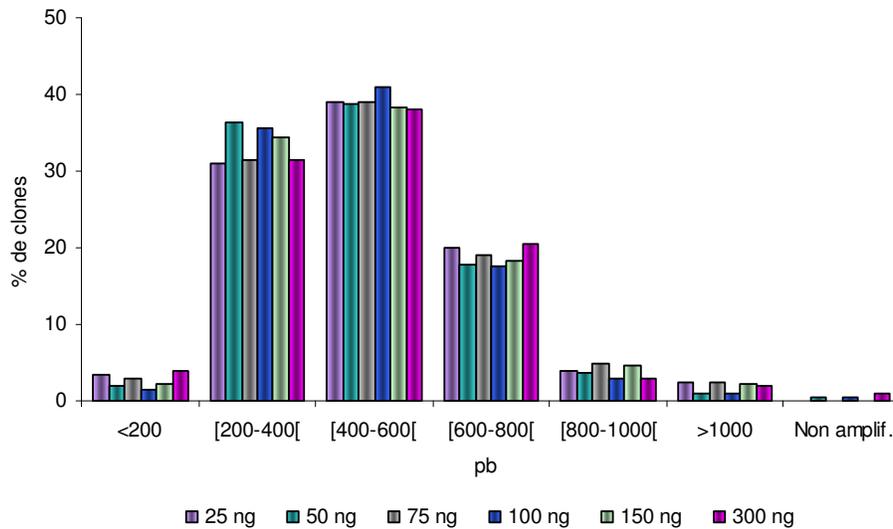


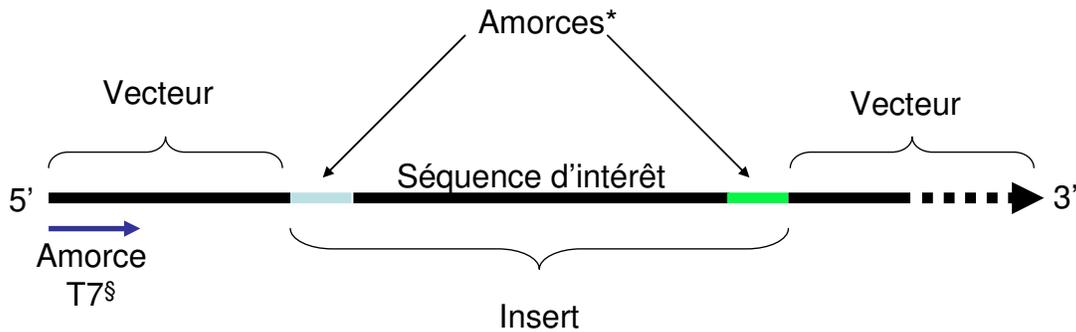
Figure 19 : Distribution des clones suivant leur taille pour 6 conditions de ligature. Vingt cinq à 300 ng de produits PCR ont été utilisés pour la ligature et 192 clones ont été amplifiés pour chaque condition de ligature.

C. Analyse des ESTs

1) Traitement des séquences

Au cours d'un premier tri, les séquences de mauvaise qualité ont été éliminées. Ce tri a été réalisé manuellement et a permis d'éliminer environ 300 séquences présentant un nombre important de nucléotides indéfinis qui rendent la séquence inutilisable lors de processus de comparaison de séquences.

Un second tri a consisté à enlever les parties correspondant au vecteur et aux amorces de PCR à l'aide d'un programme codé en Perl (Annexe numérique 3) qui permet de retrouver les amorces au sein des séquences (Figure 20). Dans le cas où les inserts sont longs (>500pb), les séquences obtenues se terminent souvent avant la fin du clone, et on ne trouve donc pas les parties correspondantes à l'amorce et au vecteur du côté 3'.



* Parties correspondant aux amorces de la seconde PCR, utilisées au cours de la fabrication des banques soustractives

§ Amorce universelle utilisée pour le séquençage des clones

Figure 20 : Composition des séquences brutes

A l'issue du séquençage des premiers clones, nous avons retrouvé dans les banques E et NE un pourcentage assez élevé (~20%) de séquences codant des ARNr 26S nucléaires et 23S chloroplastiques. Ceci peut s'expliquer par le fait que la réaction de transcription inverse qui a lieu au cours de la fabrication des banques soustractives a été initiée par un oligo d(T) et que ces séquences ribosomales de plantes présentent des régions riches en résidus adénine. Ces ARNr correspondent probablement à la bande de 700 pb observée lors de l'électrophorèse des produits issus de la seconde PCR au cours du processus de fabrication des banques (B.1)). En effet, la séquence la plus abondante dans les banques, représentant l'ARNr 26S nucléaire, a une longueur de 661 nucléotides avec les amorces NP1 et NP2 comprises. De ce fait, un troisième traitement des séquences a été réalisé dans le but

d'éliminer ces ARN ribosomiaux. L'identification des séquences ribosomiales a été réalisée à l'aide d'une comparaison des ESTs avec une séquence ribosomiale 26S de laitue, effectuée avec le programme BlastN. Au total, ce tri a permis d'éliminer environ 1350 EST représentant des ARNr. Les ESTs restants (3348) ont été enregistrés dans la base de données dbEST de GenBank le 19 août 2005 sous les numéros d'accès DT210770 à DT214189, DT317741 et DT317742.

2) Informations sur les ESTs

Après les différents tris réalisés, il reste 3348 ESTs dont 1944 proviennent de la banque embryogène et 1404 de la banque non embryogène. La longueur moyenne des séquences est de 456 pb et le pourcentage GC moyen est de 45%. On dénombre environ 180 séquences polyadénylées qui présentent au moins 10 résidus adénine au niveau de l'extrémité 3' (Tableau 4).

Nombre total d'ESTs	3348
Nombre d'ESTs dans la banque E	1944
Nombre d'ESTs dans la banque NE	1404
Longueur moyenne des ESTs (pb)	456
Pourcentage GC (%)	45
Nombre de séquences se terminant par un poly A	177
Nombre d'OCs regroupant des ESTs identiques	2302
Nombre de contigs	189
Nombre maximal de gènes	2077
Nombre de gènes représentés par un EST unique	1461
Nombre de gènes représentés uniquement dans la banque E	1061
Nombre de gènes représentés uniquement dans la banque NE	730
Nombre de gènes représentés dans les deux banques	286

Tableau 4 : Caractéristiques des ESTs des banques E et NE

3) Assemblage des ESTs identiques

Afin d'assembler les ESTs identiques, deux critères successifs ont été appliqués. Premièrement, à partir de tous les ESTs des deux banques, une base de données nucléotidiques a été créée à l'aide du logiciel BioEdit. Puis, tous les ESTs des banques E et NE ont été comparés, à l'aide du programme BlastN, aux séquences de la base de données créée. Les séquences présentant entre elles une probabilité d'alignement (E-value) inférieure ou égale à 1.10^{-30} ont été groupées dans un même OC (Original Cluster) « primaire ». En additionnant les ESTs ainsi assemblés et les ESTs uniques, nous avons dénombré 2174 OCs « primaires » numérotés de OC0001 à OC2174.

Deuxièmement, les séquences des OCs « primaires » contenant plus de 1 EST ont été alignées, et nous avons vérifié qu'elles présentaient au moins 95% d'identité entre elles sur

150 pb de chevauchement. Le cas échéant, les OCs « primaires » ont été divisés en sous-groupes (exemple : OC0603_a et OC0603_b). Cette analyse a permis la définition de 2302 OCs, dont 1764 (52,7%) sont des OC composés d'un seul EST et 538 sont composés de 2 à 26 ESTs.

4) Assemblage des ESTs en contigs

a) Stratégie

La détermination directe du nombre de gènes représentés par les 2302 OCs identifiés n'a pas été possible, puisque la digestion totale par l'enzyme de restriction RsaI, lors de la fabrication des banques soustractives, entraîne la perte de séquences chevauchantes et empêche ainsi la formation de contigs. De ce fait, afin d'identifier les OCs qui représentent potentiellement un même gène, nous avons comparé les ESTs des banques E et NE aux séquences nucléotidiques d'espèces voisines de la chicorée, cette dernière ne présentant que très peu de séquences déposées dans les bases de données rendues publiques.

Dans la famille des Astéracées, de nombreuses séquences sont disponibles pour la laitue, le tournesol et le zinnia. D'un point de vue phylogénétique, parmi ces espèces citées, la plus proche de la chicorée est la laitue, toutes deux appartenant à la sous-famille des *Cichorioideae*. Le tournesol et le zinnia, qui appartiennent à l'autre sous-famille des Astéracées, les *Asteroideae*, sont des espèces plus éloignées de notre modèle d'étude (Bremer 1994). En suivant cet ordre de proximité botanique, les ESTs des banques E et NE ont été comparés successivement, à l'aide du programme BlastN, aux ESTs assemblés de laitue, de tournesol puis de zinnia. Les ESTs de laitue et de tournesol, utilisés pour cette comparaison, proviennent de la base de données CGPD (Compositae Genome Project Database, <http://cgpdb.ucdavis.edu/>) et ceux du zinnia proviennent de la base de données PGDB (Plant Genome Database, <http://www.plantgdb.org/>).

En parallèle, afin d'établir des homologues entre un maximum d'ESTs des banques E et NE et des séquences déjà connues, les ESTs de chicorée ont été comparés, en utilisant le programme BlastX, aux séquences de la base de données non-redondantes de GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>) ainsi qu'aux séquences codantes traduites d'*Arabidopsis* (<http://www.plantgdb.org/AtGDB/>).

b) Résultats

La totalité des ESTs des deux banques (3348) a été comparée aux 19523 ESTs assemblés de laitue. Cette recherche a montré que 2040 ESTs présentent des homologues

significatives ($E\text{-value} \leq 1.10^{-30}$) avec des séquences contenues dans la base de données de laitue (Tableau 5). Les 1308 ESTs restants ont été confrontés aux ESTs de tournesol. La recherche a permis d'identifier 202 ESTs de chicorée qui présentent des homologies significatives avec les séquences de tournesol. Les ESTs des banques E et NE qui restaient à nouveau sans homologue, ont alors été soumis contre les ESTs de zinnia, ce qui a permis l'identification de 71 autres ESTs présentant des homologies avec ces séquences.

Base de données		Nombre de séquences	Nombre d'ESTs de chicorée soumis	Type de Blast (E-value)	Nombre de séquences reconnues	% de séquences reconnues
Astéracées	Laitue (CGPD)	19523 ESTs assemblés	3348	N (1E-30)	2040	69
	Tournesol (CGPD)	18031 ESTs assemblés	1308		202	
	Zinnia (PGDB)	15859 ESTs assemblés	1106		71	
Arabidopsis (AtGDB)		26719 Séquences codantes traduites	3348	X (1E-5)	2905	87
GenBank non-redondantes		>100 gigabases	3348	X (1E-5)	2794	84

Tableau 5 : Comparaison des ESTs de chicorée avec les séquences d'autres espèces de plantes.

En parallèle, les comparaisons, utilisant le programme BlastX, montrent que 2905 ESTs des banques E et NE présentent des homologies significatives ($E\text{-value} \leq 1.10^{-5}$) avec les séquences codantes traduites d'Arabidopsis et 2794 avec des séquences de la base de données non-redondantes de GenBank (Tableau 5).

Le résultat des comparaisons des ESTs de chicorée avec les séquences d'Astéracées montre qu'une grande proportion d'ESTs des deux banques (1035 ESTs soit 31% du nombre total d'ESTs) n'est pas représentée dans les bases de données des Astéracées. A titre indicatif, les comparaisons avec les séquences de la base de données non redondantes de GenBank et avec les séquences traduites d'Arabidopsis montrent respectivement 17 et 13% d'ESTs non représentés dans ces deux bases de données (Tableau 5). De plus, on remarque, que souvent, des OCs qui étaient clairement définis comme différents avec les critères appliqués (95% d'identité entre les séquences sur 150 pb de chevauchement), montrent leur plus forte homologie avec un même contig d'Astéracée alors que lorsque qu'ils sont comparés aux séquences traduites d'Arabidopsis, ils présentent leur plus forte homologie avec des gènes différents. Par exemple, les OC 1531_b et OC 1531_a, qui avaient été dissociés, révèlent leur plus forte homologie avec le même contig de laitue (QG_CA_Contig4215) mais lorsqu'ils sont comparés aux séquences codantes traduites d'Arabidopsis, ils présentent leur plus forte homologie avec des gènes différents (respectivement At4g25740 et At5g52650). Ces résultats

indiquent clairement que les bases de données d'Astéracées ne sont pas, à ce jour, assez exhaustives pour représenter l'ensemble des transcriptomes des Astéracées. De ce fait, nous avons décidé d'utiliser les résultats des comparaisons avec les séquences traduites d'*Arabidopsis* pour assembler les OCs en contigs. Les OCs qui présentaient la plus forte homologie ($E\text{-value} \leq 1.10^{-5}$) avec la même séquence codante d'*Arabidopsis* ont été considérés comme étant les représentants d'un même gène.

Cette analyse a conduit à la formation de 189 contigs qui regroupent 111 OCs représentés par plus d'un EST, et 303 OCs représentés par des séquences uniques. En additionnant les OCs assemblés en contigs aux 1888 OCs non assemblés, nous avons estimé que les 3348 ESTs des banques E et NE représentent 2077 gènes. Parmi ces 2077 gènes, 1061 (51%) sont composés d'ESTs représentés exclusivement dans la banque E, 730 (35%) d'ESTs représentés uniquement dans la banque NE, et 286 (14%) d'ESTs représentés dans les deux banques.

La distribution, du nombre d'ESTs qui représentent potentiellement un même gène, montre qu'une grande proportion de gènes sont représentés par une séquence unique (1461 sur 2077, soit 70%) (Figure 21).

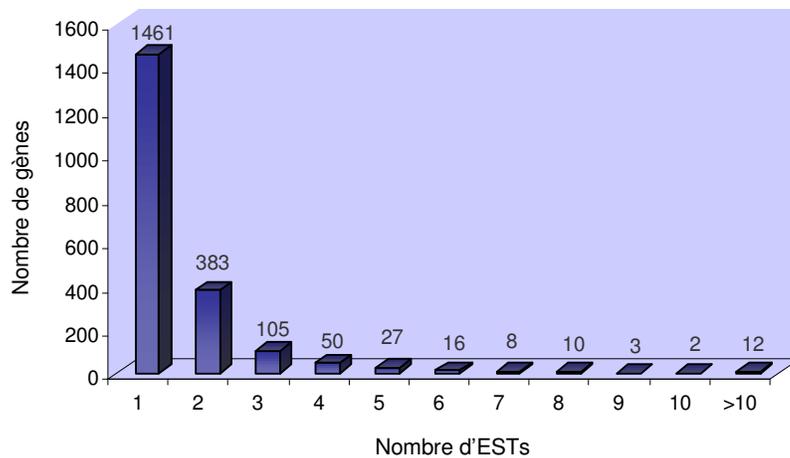


Figure 21 : Distribution du nombre d'ESTs qui représentent potentiellement un même gène

5) Analyse fonctionnelle des ESTs

a) Annotation des séquences

Les ESTs ont été annotés après comparaison avec les séquences de la base de données non-redondantes de GenBank et avec les séquences codantes traduites d'*Arabidopsis*. L'annotation des ESTs indique que 55% des ESTs présentent une forte homologie ($E\text{-value} \leq$

1.10^{-30}) avec les séquences de la base de données non-redondantes de GenBank, et 32% une homologie moyenne (1.10^{-30} à 1.10^{-5}). Les ESTs restants (13%) ne présentent pas d'homologies ou présentent des homologies non significatives avec les séquences de la base de données ($E\text{-value} \geq 1.10^{-5}$). Ces séquences ont été classées comme n'ayant 'pas d'homologie' (Figure 22).

b) Classification fonctionnelle

Les ESTs ont été classés suivant leur fonction à l'aide des résultats de comparaisons des ESTs avec les séquences codantes d'Arabidopsis et les données du catalogue fonctionnel de MIPS (Munich Information centre for Proteins Sequences ; <http://mips.gsf.de/projects/funcat>) qui attribue à la plupart des séquences codantes d'Arabidopsis une classe de fonction biologique. Les gènes d'Arabidopsis codant des protéines non classées ont été soumis manuellement à une recherche dans le catalogue fonctionnel afin d'établir des homologies avec d'autres gènes d'Arabidopsis (BlastX , $E\text{value} \leq 1.10^{-5}$), présentant une fonction établie. Cette fonction a alors été attribuée au gène soumis et aussi au gène de chicorée correspondant.

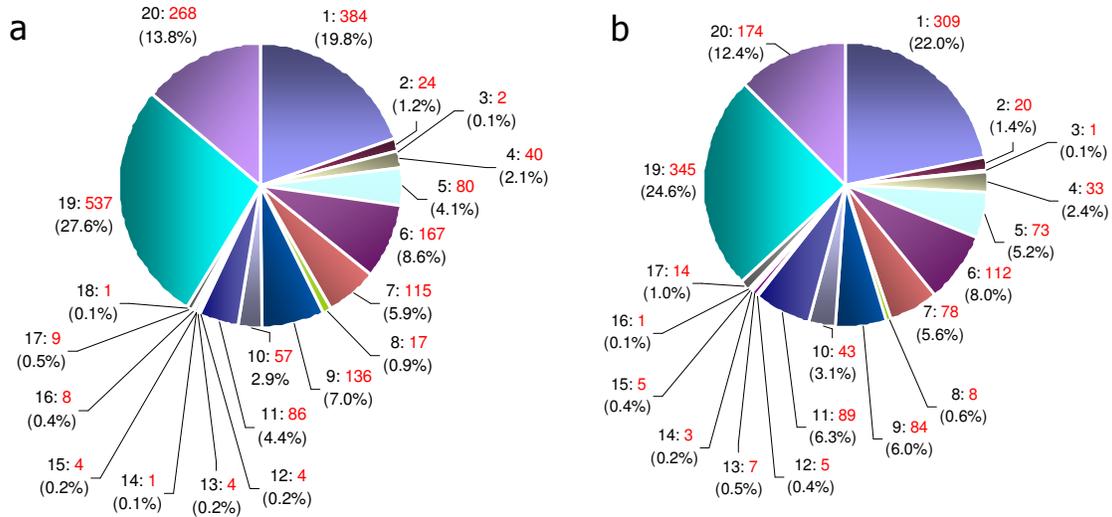


Figure 22: Distribution par classe de fonction des ESTs des banques E (a) et NE (b) Pour chaque classe, le 1^{er} chiffre indique la classe et le chiffre en rouge indique l'effectif. 1: Métabolisme ; 2: Energie ; 3: Stockage ; 4: Cycle cellulaire et synthèse de l'ADN ; 5: Transcription ; 6: Synthèse protéique ; 7: Devenir des protéines (repliement, modification, destination) ; 8: Protéines fixatrices ou nécessitant un cofacteur; 9: Transport cellulaire ; 10: Communication cellulaire ; 11: Défense ; 12: Interaction avec l'environnement cellulaire ; 13: Interaction avec l'environnement ; 14: Eléments transposables, protéines virales et plasmidiques ; 15: Devenir cellulaire ; 16: Développement ; 17: Biosynthèse de composants cellulaires ; 18: Différenciation cellulaire ; 19: Protéines inconnues (regroupant localisation cellulaire, classification non certaine, et protéines non classées) ; 20: Aucune homologie

La distribution par classe de fonction des ESTs est quasiment identique pour les deux banques (Figure 22). Les classes les plus représentées sont les fonctions inconnues (26%), le métabolisme (21%), la synthèse protéique (8%), le transport cellulaire (7%), le devenir cellulaire (6%) et la défense (5%). Sur les 209 ESTs qui correspondent à des gènes liés à la synthèse protéique, 202 (6% du nombre total d'ESTs) sont des gènes codant des protéines ribosomales ; 122 proviennent de la banque E et 80 de la banque NE.

Les gènes les plus représentés dans les banques E et NE sont un gène codant une métallothionéine (OC3039), protéine liée à la défense, représenté par 40 ESTs dont 14 proviennent de la banque E et 26 de la banque NE et un gène codant une 3-hydroxy-3-methylglutaryl coenzyme A réductase (Cont0060) liée au métabolisme général et représentée par 8 ESTs dans la banque E et 15 dans la banque NE.

L'ensemble des informations concernant le regroupement des ESTs identiques, l'assemblage en contigs, ainsi que l'annotation et la classification fonctionnelle des ESTs sont données dans l'Annexe numérique 4. Le Tableau 6 est un extrait de l'Annexe numérique 4 et présente le détail des différents champs de cette annexe.

Résultats et discussion

Information sur les clones			Original Cluster (OC)			Contig			Comparaison des ESTs, à l'aide du programme BlastX, avec les séquences de la base de données non redondante de GenBank										Comparaisons des ESTs, à l'aide du programme BlastX, avec les séquences codantes traduites d' <i>Arabidopsis thaliana</i> de la base de données AtGDB				
Num d'accension	Identifiant du clone	Longueur (bp)	Identifiant de l'OC	Nb d'ESTs provenant de la banque E	Nb d'ESTs provenant de la banque NE	Identifiant du contig	Nb d'ESTs provenant de la banque E	Nb d'ESTs provenant de la banque NE	Identifiant	Fonction de la protéine déduite	Longueur (AA)	Score	E-value	Identité	Homologie	Gap	Cadre de lecture	Alignement EST (pb)	Alignement séquence homologue (AA)	Identifiant	E-value	Classification fonctionnelle selon le catalogue fonctionnel de MIPS	
DT210806	MMD019F4	533	OC0026_a	0	1	Cont0001			gb AAM92008.1	glyceraldehyde 3-phosphate dehydrogenase [Solanum tuberosum]	338	275bits(702)	5E-73	141/177(79%)	146/177(82%)	0	-2	532-2	62-238	At1g13440	1,00E-91	Metabolism	01.05 C-compound and carbohydrate metabolism
DT212554	E001F9	577	OC0026_b			Cont0001														At1g13440	8,00E-59	Metabolism	01.05 C-compound and carbohydrate metabolism
DT212714	E003G8	565	OC0026_b			Cont0001			emb CAA27844.1	glyceraldehyde-3-phosphate dehydrogenase [Sinapis alba]	338	362bits(928)	4,00E-99	177/187(94%)	180/187(96%)	0	-3	563-3	52-238	At3g04120	1,00E-100	Metabolism	01.05 C-compound and carbohydrate metabolism
DT212825	E005C9	505	OC0026_b			Cont0001														At1g13440	3,00E-89	Metabolism	01.05 C-compound and carbohydrate metabolism
DT213087	E009C5	561	OC0026_b			Cont0001														At3g04120	3,00E-99	Metabolism	01.05 C-compound and carbohydrate metabolism
DT213859	E020B8	429	OC0026_b			Cont0001														At3g04120	3,00E-71	Metabolism	01.05 C-compound and carbohydrate metabolism
DT211226	N107A9	261	OC0026_b	5	1	Cont0001														At3g04120	5,00E-36	Metabolism	01.05 C-compound and carbohydrate metabolism
DT212526	E001C4	430	OC2148			Cont0001														At1g13440	2,00E-20	Metabolism	01.05 C-compound and carbohydrate metabolism
DT213042	E008E10	419	OC2148			Cont0001														At3g04120	3,00E-36	Metabolism	01.05 C-compound and carbohydrate metabolism
DT213919	E021A9	409	OC2148			Cont0001			emb CAA42103.1	glycolytic glyceraldehyde 3-phosphate dehydrogenase [Antirrhinum majus]	336	178bits(452)	3,00E-44	89/99(89%)	95/99(95%)	0	1	7-303	238-336	At3g04120	3,00E-41	Metabolism	01.05 C-compound and carbohydrate metabolism
DT214173	E024G7	444	OC2148	4	0	Cont0001	9	2												At3g04120	1,00E-34	Metabolism	01.05 C-compound and carbohydrate metabolism

Tableau 6 : Extrait de l'Annexe numérique 4, présentation des champs

L'Annexe numérique 4 compile les données sur le regroupement des ESTs ainsi que sur leur annotation et leur classification.

Le tableau se découpe en 5 champs majeurs. Le premier composé de trois colonnes fournit des informations sur les clones, telles que le numéro d'accension des séquences dans dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>), le numéro d'identifiant issu des plaques de culture bactérienne et la longueur des ESTs. Les numéros d'identifiant commençant par A, B, C, et E caractérisent les clones de la banque embryogène (E), alors que ceux débutant par M ou N représentent les clones de la banque non embryogène (NE). Le champ Original Cluster (OC) regroupe les informations sur l'assemblage des EST identiques. Ce champ est découpé en trois colonnes : l'identifiant de l'OC puis, pour chaque OC, les colonnes suivantes indiquent le nombre d'ESTs représentant cet OC provenant de la banque E et de la banque NE. Le champ Contig regroupe les informations sur l'assemblage des ESTs en contigs et se découpe en trois colonnes comme le champ Original Cluster. Le quatrième champ regroupe les données des comparaisons des ESTs avec les séquences de la base de données non redondantes de GenBank. Les résultats, donnés pour chacun des OC, sont ceux de la séquence présentant la plus forte homologie avec un des EST de l'OC. Les données fournies sont l'identifiant de la séquence homologue, la fonction de la protéine, sa longueur en acides aminés, le « score » de l'alignement de la séquence protéique déduite de l'EST de chicorée avec la séquence protéique de la base de données GenBank, la probabilité (E-value) associée à l'alignement des séquences, le nombre et le pourcentage d'identité, le nombre et le pourcentage d'homologie, le nombre et le pourcentage de « gap », le cadre de lecture de l'EST et les alignements des deux séquences. Le dernier champ fournit des informations sur les comparaisons des ESTs avec les séquences traduites d'*Arabidopsis* réalisées à l'aide du programme BlastX. Les résultats, donnés pour chaque EST, sont ceux du gène d'*Arabidopsis* présentant la plus forte homologie avec l'EST correspondant. Ces résultats comprennent l'identifiant du gène d'*Arabidopsis*, la probabilité associée à l'alignement de la séquence protéique déduite de l'EST et de celle du gène d'*Arabidopsis* et les deux dernières colonnes représentent la classification fonctionnelle réalisée à partir du catalogue fonctionnel de MIPS.

D. Soustraction *In silico* des ESTs des banques E et NE et validation par PCR en temps réel

1) Soustraction *in silico*

Dans le but de vérifier l'efficacité des étapes de normalisation et de soustraction lors de la fabrication des banques soustractives, nous avons réalisé une soustraction *in silico*, méthode appelée aussi « digital northern » à partir de tous les ESTs des deux banques. Cette analyse est basée sur la relation qui existe entre l'abondance des ESTs dans les banques d'ADNc et l'expression différentielle des gènes (Lee *et al.*, 1995; Okubo *et al.*, 1992). Habituellement les soustractions *in silico* sont réalisées à partir d'ESTs provenant de banques d'ADNc non soustraites (Miner and Rajkovic, 2003; Rajkovic *et al.*, 2001), mais puisque la normalisation réalisée lors de la fabrication des banques SSH n'est pas toujours efficace, nous avons décidé de réaliser cette soustraction à partir des ESTs de nos banques soustractives. Si cette analyse n'était pas efficace, c'est-à-dire qu'elle ne permettait pas la mise en évidence de gènes exprimés différentiellement, cela indiquerait que les banques SSH sont effectivement normalisées et soustraites. Dans le cas contraire, cette analyse nous permettrait de mettre en évidence des gènes exprimés différentiellement au cours de l'embryogenèse somatique qui pourraient être validés par PCR en temps réel.

En appliquant le test de probabilité de Audic and Claverie (1997), nous avons identifié à partir des 2077 gènes des deux banques, 33 gènes (1,6%) différentiellement représentés dans les deux banques, avec 6 gènes préférentiellement représentés dans la banque embryogène et 27 dans la banque non embryogène (Tableau 7).

Résultats et discussion

Num acc	Contig/OC	Fonction déduite	Analyses <i>in silico</i>				PCR en temps réel	
			E	NE	P	Pred	ER 1	ER 2
DT212490	Cont0080	vegetative storage protein, VSP [<i>C. intybus</i>]	5	0]0.08;0.07[E		
DT212771	Cont4942	elongation factor EF-2 [<i>P. sativum</i>]	10	0]0.006;0.005[E	-0.18	0.38
DT212374	Cont9008	Shaggy-like kinase tetha [<i>A. thaliana</i>]	5	0]0.08;0.07[E		
DT212650	OC0603_a	methionine synthase [<i>S. tuberosum</i>]	6	0]0.05;0.04[E		
DT212300	OC1814	No hits found	5	0]0.08;0.07[E	2.00	1.90
DT212502	OC1929	G protein beta subunit [<i>N. plumbaginifolia</i>]	5	0]0.08;0.07[E	1.25	1.24
DT213897	Cont0011	catalase 3 [<i>H. annuus</i>]	4	8]0.09;0.08[NE	-3.85	-3.88
DT211305	Cont0015	L-asparaginase [<i>G. max</i>]	0	3]0.07;0.06[NE	0.18	0.40
DT211671	Cont0026	ss-1,3-glucanase [<i>C. intybus</i> x <i>C. endivia</i>]	0	3]0.07;0.06[NE	-2.87	-4.10
DT213305	Cont0060	3-hydroxy-3-methylglutaryl coenzyme A reductase [<i>S. tuberosum</i>]	8	15]0.03;0.02[NE	-2.30	-2.03
DT212144	Cont0082	putative argininosuccinate synthase [<i>A. thaliana</i>]	0	4]0.03;0.02[NE	1.23	0.70
DT210848	Cont0112	heat shock protein 70 like protein [<i>A. thaliana</i>]	0	3]0.07;0.06[NE		
DT211362	Cont0201	putative auxin-induced protein [<i>A. thaliana</i>]	0	5]0.02;0.01[NE	-1.88	-2.03
DT211152	Cont0206	fasciclin-like AGP 12 [<i>P. alba</i> x <i>P. tremula</i>]	0	4]0.03;0.02[NE	-0.25	-0.40
DT211657	Cont1237	pre-pro-cysteine proteinase [<i>L. esculentum</i>]	0	4]0.03;0.02[NE	-2.32	-3.00
DT211016	Cont1242	unknown protein [<i>A. thaliana</i>]	0	3]0.07;0.06[NE		
DT210870	Cont9022	unknown yeast pheromone receptor-like protein AR781 [<i>A. thaliana</i>]	0	3]0.07;0.06[NE		
DT213289	Cont9039	metallothionein 1 [<i>A. tripolium</i>]	14	26]0.004;0.003[NE	-2.45	-2.45
DT210772	OC0003_b	glutathione S-transferase GST 13 [<i>G. max</i>]	0	4]0.03;0.02[NE		
DT212623	OC0023_a	glutathione transferase [<i>H. muticus</i>]	3	11]0.007;0.006[NE	-2.78	-2.85
DT211055	OC0168	1-aminocyclopropane-1-carboxylate oxidase [<i>L. esculentum</i>]	0	9	<0.001	NE	-3.87	-3.80
DT210897	OC0178	putative ribosomal S29 protein [<i>A. thaliana</i>]	0	3]0.07;0.06[NE	1.15	0.70
DT210805	OC0344	unknown protein [<i>A. thaliana</i>]	0	3]0.07;0.06[NE	-1.35	-1.68
DT211689	OC0378_b	putative 60S ribosomal protein L5 [<i>O. sativa</i>]	0	3]0.07;0.06[NE		
DT214125	OC0533b	lipid transfer protein precursor [<i>M. domestica</i>]	2	7]0.04;0.03[NE	-0.18	-0.25
DT211562	OC0559	putative phosphatase type 2C [<i>A. thaliana</i>]	0	3]0.07;0.06[NE	-0.25	-0.25
DT211372	OC0631	No hits found	0	3]0.07;0.06[NE	0.20	0.20
DT211599	OC0686	unknown protein [<i>A. thaliana</i>]	0	4]0.03;0.02[NE	-3.37	-3.23
DT211633	OC0696	aquaporin [<i>H. annuus</i>]	0	3]0.07;0.06[NE	-3.72	-3.52
DT211922	OC0724	cellulose synthase family protein [<i>A. thaliana</i>]	0	3]0.07;0.06[NE	-2.60	-2.47
DT211365	OC0727_b	G protein beta subunit-like [<i>M. sativa</i>]	0	5]0.02;0.01[NE		
DT210839	OC0758	putative RNA-binding protein [<i>A. thaliana</i>]	0	5]0.02;0.01[NE	0.08	0.18
DT212897	OC1229	sucrose synthase isoform I [<i>D. carota</i>]	1	5]0.05;0.04[NE	0.08	-0.32

Tableau 7 : Soustraction *in silico* et analyse transcriptomique par PCR en temps réel
Contig/OC: ID du contig ou de l'OC (Original Cluster regroupant les ESTs identiques). E et NE: nombre d'ESTs provenant respectivement de la banque E et de la banque NE; P : probabilité associée à la prédiction *in silico* suivant Audic and Claverie (1997). Pred (prediction) indique dans quelle banque (E: embryogène ou NE: non-embryogène) le gène est préférentiellement représenté. ER1, ER2 : ratio d'expression (\log_2) calculé à partir des résultats de PCR en temps réel entre le génotype embryogène et le génotype non-embryogène au jour 4 de culture d'embryogenèse somatique pour les deux répétitions. Les ratios d'expressions sur fond gris sont ceux des gènes différentiellement exprimés.

2) PCR en temps réel

L'abondance des transcrits a été mesurée par PCR en temps réel pour 24 gènes cibles, prédits comme étant différentiellement exprimés par l'analyse *in silico* (Tableau 7).

Les ADNc correspondant à chaque gène cible ont été placés à trois endroits distincts sur la plaque et l'expérience a été répétée. Cette expérience a permis de calculer pour chaque cible deux ratios d'expression, ER1 et ER2. Pour le génotype embryogène, la répétition a été biologique, c'est-à-dire que les ARN utilisés pour la synthèse des ADNc proviennent de deux

cultures indépendantes en conditions d'embryogenèse somatique. Pour le génotype non embryogène, la répétition a été technique, c'est-à-dire que les ADNc ont été synthétisées à partir d'un mélange d'ARN obtenu à partir de trois cultures indépendantes. Un gène a été défini comme étant différentiellement exprimé lorsque les deux répétitions de PCR conduisaient à l'obtention d'un ratio d'expression (\log_2) supérieur ou égal à 0,6 pour un gène préférentiellement exprimé chez K59 et inférieur ou égal à -0,6 pour un gène préférentiellement exprimé chez C15, soit une différence d'abondance de transcrits entre les deux génotypes de 1,5 fois.

Les résultats des analyses par PCR en temps réel confirment l'expression différentielle entre les deux génotypes pour 14 gènes, avec 2 préférentiellement exprimés chez le génotype embryogène et 12 chez le génotype non embryogène (Tableau 7). Pour deux gènes qui étaient prédits comme étant préférentiellement représentés dans la banque non embryogène, les résultats de PCR en temps réel ont montré qu'ils étaient au contraire préférentiellement exprimés chez l'autre génotype (Cont0082 et OC0178). Les 8 autres gènes prédits n'ont pas été identifiés comme étant différentiellement exprimés.

E. Discussion

1) Choix des génotypes et du stade de développement pour la construction des banques

Pour la fabrication des banques E et NE, nous avons décidé d'utiliser des ARNm isolés d'explants foliaires de chicorées de génotypes K59 et C15. Les deux génotypes diffèrent quant à leur compétence à former des embryons somatiques. Précédemment au laboratoire, c'est l'hybride interspécifique '474' qui était utilisé dans le but d'étudier l'embryogenèse somatique. Cependant l'hybride '474' est quasiment stérile, ce qui obligeait à comparer les résultats avec ceux de génotypes non embryogènes, au fond génétique éloigné de celui de l'hybride '474'. Plus récemment, le génotype K59 a été sélectionné comme fortement embryogène au sein de la population « Koospol » de *C. intybus*. Les aspects cytologiques observés dans les explants issus du génotype K59 au cours d'une cinétique de culture en conditions d'embryogenèse somatique sont similaires à ceux observés pour les explants issus de l'hybride interspécifique '474' (Da Silva, 2004). Le génotype K59 présente l'avantage par rapport à '474' d'être fertile et d'avoir une capacité relative d'autofécondation. C'est à partir de plantes issues de l'autofécondation du génotype K59 qu'a été obtenu le génotype C15, sélectionné pour son incapacité à former des embryons somatiques. Les deux génotypes, partageant un fond génétique commun, devraient montrer des différences

génétiques minimales pour tout autre caractère que l'aptitude à former des embryons somatiques *in vitro*. Les aspects morphologiques des plantes de génotypes K59 et C15, à l'état végétatif et cultivées en serre, sont par ailleurs très similaires.

Dans cette étude nous avons décidé de nous focaliser sur les stades les plus précoces de l'embryogenèse somatique. Ainsi, le jour 4 de culture en conditions d'embryogenèse somatique a été choisi pour extraire les ARN des explants des deux génotypes, puisque c'est à ce jour que les explants issus du génotype K59 présentent un maximum de cellules en cours de transition de l'état somatique à l'état embryonnaire. Les banques soustractives E et NE ont ainsi été obtenues par le clonage d'ADNc qui représentent en théorie, respectivement des gènes préférentiellement exprimés dans les feuilles des génotypes K59 et C15 cultivés dans les conditions choisies. En effet, la technique de SSH utilisée pour la construction des banques, présente l'avantage, par rapport à des banques d'ADNc classiques, de soustraire les transcrits communs aux explants des deux génotypes qui pourraient, par exemple, représenter des gènes reliés aux stress engendrés par les conditions de culture ou des gènes de ménage.

2) Comparaison des ESTs aux séquences d'autres espèces

Après le séquençage de 5000 clones des banques E et NE, et l'obtention de 3348 ESTs, nous avons estimé, à la suite de comparaisons avec les séquences codantes traduites d'*Arabidopsis*, à l'aide du programme BlastX, que l'ensemble des ESTs représentait 2077 gènes. Ce nombre total de gènes est probablement surestimé puisqu'une partie des OCs et des contigs, qui ne présentent pas d'homologie avec les séquences des bases de données, pourraient correspondre à des régions non traduites en 5' ou en 3' de gènes déjà identifiés. En effet, on dénombre dans les banques 180 ESTs poly adénylés.

Nous avons vu que certains OC qui avaient été dissociés, puisque les séquences les représentant ne montraient pas 95% d'identité de séquence sur 150 pb de chevauchement, ont révélé leur plus forte homologie avec le même contig de laitue. Nous avons conclu que les banques d'ESTs d'Astéracées, malgré un nombre important de séquences assemblées proche du nombre de gènes d'*Arabidopsis*, devaient être trop peu exhaustives. Leur utilisation afin d'assembler les ESTs représentant un même gène aurait conduit à une sous-estimation du nombre de gènes. De la même façon, parfois, les ESTs classés dans différents OCs ont montré leur plus forte homologie avec la même séquence codante d'*Arabidopsis*. Cependant, dans la plupart des cas, ces ESTs ont présenté des homologies différentes lorsqu'ils étaient comparés à la base de données non redondantes de GenBank. Ce résultat pourrait refléter une plus forte

duplication des gènes chez la chicorée, et les Astéracées en général, en comparaison à *Arabidopsis*.

3) Classification fonctionnelle des ESTs

La distribution par classe de fonctions des ESTs des deux banques ressemble à celles des gènes d'*Arabidopsis* (2000) ainsi qu'aux classifications fonctionnelles d'ADNc de *Zinnia* isolés lors de la transdifférenciation de cellules du mésophylle en cellules du xylème (Demura *et al.*, 2002) et des ADNc d'*Antirrhinum* obtenus lors de la formation des pétales (Bey *et al.*, 2004). Comme celles des banques E et NE, ces distributions présentent toutes environ 25% de séquences aux fonctions indéterminées. Ce résultat témoigne que chacune des banques est représentative de la diversité des fonctions observées dans un transcriptome. On note cependant une exception puisque les ESTs correspondant à des gènes codant des protéines ribosomales sont assez fortement représentés dans nos banques. En effet, environ 200 ESTs (6% du nombre total d'ESTs) représentent 82 gènes codant environ 61 protéines ribosomales. A titre de comparaison, il existe chez *Arabidopsis* 249 gènes codant 80 protéines ribosomales (32 pour la petite sous-unité, 48 pour la grande sous-unité) (Barakat *et al.*, 2001). Cette abondance de gènes représentés dans les banques E et NE et codant des protéines ribosomales pourrait refléter le fait que la construction des banques soustractives a été réalisée au cours de la mise en place de nouveaux transcriptomes dans les explants des génotypes K59 et C15.

4) Qualité des banques SSH

A l'exception des séquences ribosomales contaminantes, on observe qu'une grande proportion des gènes (1461 sur un total de 2077, soit 70%) sont représentés par un EST unique et que peu de gènes sont fortement représentés (3% des gènes sont représentés par plus de 5 ESTs). La redondance de nos banques, calculée comme le ratio du nombre d'ESTs associés en contig rapporté au nombre total d'ESTs est de 56%. A titre de comparaison, une redondance de 69% a été obtenue à partir de banques SSH composées de 1344 ESTs chez le peuplier dans le cadre d'une étude portant sur les modulations d'expression suite à l'application d'un agent protecteur contre des herbicides (Rishi *et al.*, 2004). A nouveau chez le peuplier, cette fois-ci lors de travaux portant sur l'analyse de différents tissus présentant une déficience au niveau du métabolisme des phénylpropanoïdes, des redondances de l'ordre de 49 à 84% ont été obtenues pour des banques SSH composées de 1500 ESTs (Ranjan *et al.*, 2004). Notre résultat indique que les banques semblent relativement normalisées et est confirmé par la soustraction *in silico*, qui n'a permis la prédiction que de 33 gènes

différentiellement représentés, soit seulement 1,6% du total des 2077 gènes annotés. De plus, une expression différentielle pour seulement 14 de ces 33 gènes a pu être montrée par PCR en temps réel. Le faible nombre de gènes prédits ainsi que la faible proportion de gènes validés comme différentiellement exprimés par PCR en temps réel traduisent le fait que la soustraction *in silico* est une méthode très limitée pour la mise en évidence de gènes exprimés différentiellement à partir des banques E et NE. Ces résultats renforcent l'idée que les ADNc des banques E et NE ont été effectivement normalisés.

Le résultat, montrant que nos deux banques présentent la même distribution des ESTs par classe de fonction, révèle que la fabrication des banques soustractives n'a pas entraîné un enrichissement d'ESTs représentant une classe de fonction particulière. Cependant, bien que les banques aient la même distribution, les gènes présents dans la banque E sont en grande majorité différents de ceux présents dans la banque NE. En effet, après annotation des gènes à l'aide de comparaisons des ESTs avec les séquences de la base de données non redondantes de GenBank, on remarque que seulement 14% des gènes annotés sont représentés par des ESTs présents dans les deux banques d'ADNc, E et NE. Cette information pourrait témoigner d'une soustraction efficace lors de la fabrication des banques soustractives.

Les analyses de PCR en temps réel ont permis l'identification de 16 gènes exprimés différentiellement entre les génotypes embryogène et non embryogène après 4 jours en conditions de culture d'embryogenèse somatique. Bien que les gènes cibles n'aient pas été sélectionnés au hasard, les résultats de l'analyse transcriptomique par PCR en temps réel semblent montrer que les banques E et NE contiennent de nombreux gènes exprimés différentiellement. Ce résultat tendrait, à nouveau, à montrer qu'il y a eu un enrichissement des ADNc représentant des gènes exprimés différentiellement au cours de la fabrication des banques soustractives.

L'analyse transcriptomique par PCR en temps réel a permis de mettre en évidence des gènes différentiellement exprimés chez les génotypes C15 ou K59 qui présentent un intérêt particulier (Tableau 7). Par exemple, on remarque qu'un gène codant une protéine G (OC1929) est préférentiellement exprimé chez le génotype embryogène K59. Ce gène pourrait être impliqué dans la signalisation lors de la formation des embryons somatiques. Chez le génotype non-embryogène, plusieurs gènes, qui peuvent être reliés à une réaction de défense, sont préférentiellement exprimés. Ils codent une glutathione S-transferase (OC0003_b), une métallothionéine (Cont9039) et une ACC oxydase (OC0168), enzyme impliquée dans la synthèse de l'éthylène. La défense pourrait être une classe de fonction

particulièrement représentée par les gènes impliqués dans la réaction du génotype non-embryogène aux conditions de culture appliquées.

Les résultats obtenus au cours de la partie I ont fait l'objet d'un article soumis à la revue BMC Plant Biology, donné en annexe (Annexe 3).

Partie II : Mise en œuvre de la technique de microarray pour l'étude de l'embryogenèse somatique chez la chicorée

A. Mise en place du schéma expérimental

Pour ce travail, nous avons choisi de cibler notre étude transcriptomique sur les phases précoces de l'embryogenèse somatique, qui correspondent à la transition des cellules d'un état somatique à un état embryonnaire. Nous avons décidé d'étudier en cinétique et pour les deux génotypes, K59 et C15, l'expression d'un certain nombre de gènes des banques E et NE. Nous avons choisi d'étendre la cinétique de 0 à 6 jours (noté J0 à J6) avec un point d'étude par jour. Cet intervalle de temps couvre, pour le génotype embryogène K59, l'induction de l'embryogenèse somatique, la réactivation des cellules et les premières divisions cellulaires.

Pour la mise en place de l'étude transcriptomique par microarray, nous avons choisi un schéma expérimental indirect utilisant une référence constituée par un mélange d'ARN (Partie II D.1). En comparant les signaux d'hybridation des conditions « test », (génotype embryogène ou non embryogène à un point donné de la cinétique (notés K59J0 à K59J6 et C15J0 à C15J6) avec ceux de la « référence », cela nous a permis de comparer l'expression des gènes entre toutes les conditions testées. Notre schéma a nécessité la réalisation de 14 expériences, et à titre d'exemple, a permis de comparer aussi bien la condition C15J0 à la condition C15J4 qu'à la condition K59J0.

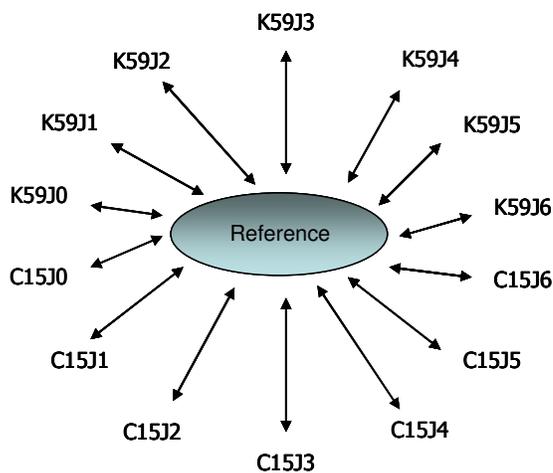


Figure 23 : Schéma expérimental
Toutes les conditions « test » (génotype/jour de collecte des explants) ont été comparées au mélange d'ARN référence, pour constituer un total de 14 expériences.

Pour chaque expérience, nous avons réalisé des hybridations sur quatre lames distinctes (Figure 24). Dans une première combinaison de fluorochromes, une lame ainsi que sa répétition ont été hybridées avec la condition « test » marquée avec le cy3 et la condition

« référence » avec le cy5. Dans une seconde combinaison de fluorochromes, comprenant également une lame et sa répétition, les fluorochromes ont été inversés entre les conditions « test » et « référence ». Pour le génotype embryogène, la répétition a été biologique, c'est-à-dire que les ARN utilisés pour la fabrication des cibles de la condition « test » provenaient de deux cultures indépendantes d'explants prélevés sur différentes plantes. Pour le génotype non embryogène, en raison d'une faible quantité d'ARN contenue dans les explants pour certains points de la cinétique, la répétition a été technique, c'est-à-dire que les cibles de la condition « test » ont toutes été synthétisées à partir d'un mélange d'ARN obtenu à partir de trois cultures indépendantes.

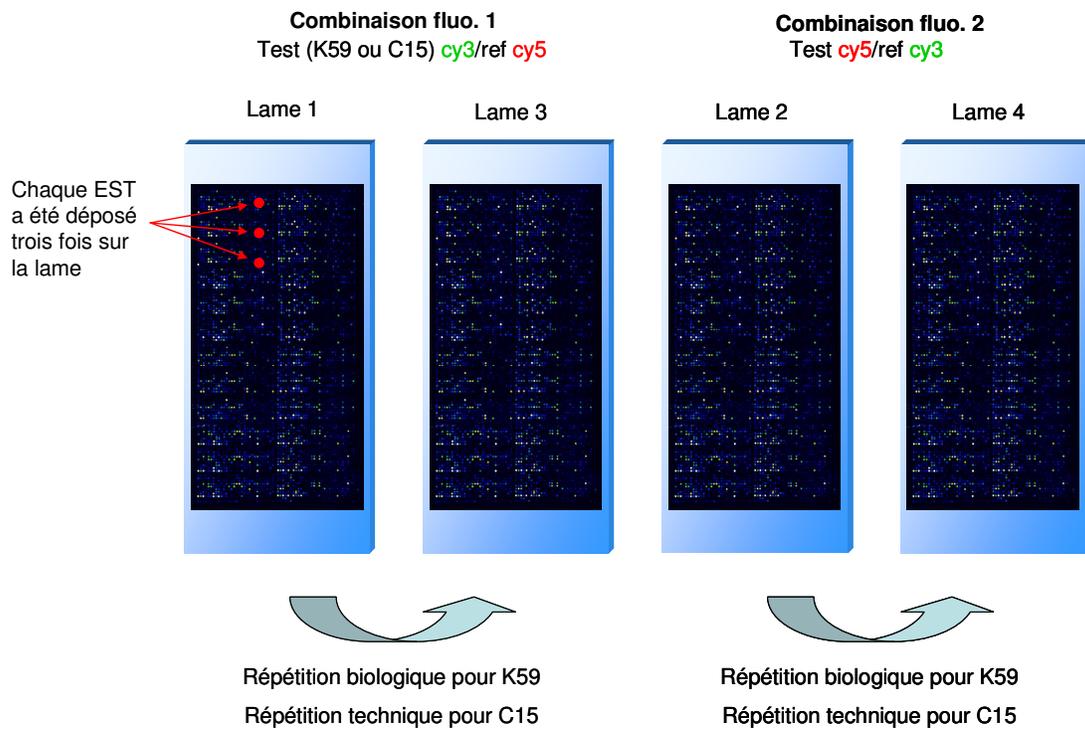


Figure 24 : Quatre lames composent une expérience

Pour chaque expérience, et pour chacun des spots, le ratio de l'intensité enregistrée pour la condition « test » et de celle enregistrée pour la condition « référence » a été calculé. Ensuite, pour les deux génotypes, afin d'observer les profils d'expression des gènes au cours du temps, les ratios test/référence obtenus pour chaque point de la cinétique ont été comparés à ceux obtenus pour le jour 0 (Figure 25). Les gènes présentant une variation significative par rapport au jour 0 pour l'un et/ou l'autre génotype ont été conservés pour réaliser des regroupements des profils d'expression, et déterminer ainsi les gènes essentiels aux réponses

des deux génotypes aux conditions de culture appliquées. De plus, une comparaison des conditions K59J0 et C15J0 a été réalisée afin d'étudier l'expression des gènes chez les deux génotypes au stade initial de la cinétique. Par ailleurs, une comparaison des conditions K59J4 et C15J4 a permis notamment la mise au point du processus d'analyses des résultats de microarray.

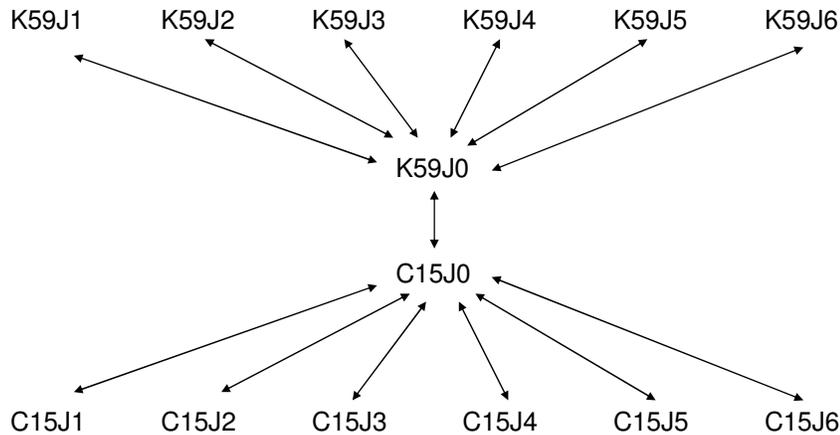


Figure 25 : Plan de comparaison des expériences pour la mise en évidence de gènes présentant une variation d'expression aux cours de la cinétique de culture en conditions d'embryogenèse somatique

Pour chaque génotype, les ratios test/référence obtenus pour chaque point d'expérience ont été comparés au ratio test/référence obtenu pour le jour 0. Les gènes conservés pour les analyses de regroupement des profils sont ceux présentant pour l'un et/ou l'autre génotype une variation significative au cours de la cinétique. Les ratios test/référence obtenus pour chacun des deux génotypes au jour zéro ont également été comparés l'un avec l'autre.

B. Optimisation technique des microarray

Pour la Plateforme Biopuces de l'Institut Pasteur de Lille, où ont été réalisés les microarrays, ce travail a constitué la première expérience avec du matériel végétal et des sondes obtenues à partir d'ADNc issus de banques soustractives. Ainsi une optimisation du protocole a été nécessaire. Afin d'entreprendre divers essais, des lames de microarray, appelées lames « essai » et ne comportant que 32 ADNc différents ont été fabriquées. Ces lames ont permis d'optimiser la production des sondes et la production des cibles.

1) Choix de l'enzyme d'amplification pour la production des sondes

Chaque clone a été amplifié à l'aide de trois enzymes de marques différentes (Biolabs, Applied, Qiagen) en suivant le protocole expérimental fourni. Puis, les produits obtenus pour chaque amplification ont été purifiés et déposés six fois sur lame. Le robot utilisé est un appareil à huit aiguilles munies d'un anneau (système « pin and ring »). La tête du robot se positionne au dessus des plaques contenant les produits à déposer, plonge dans les puits et emporte un faible volume d'échantillon (0,5 µl). La tête vient ensuite se positionner au dessus du plateau contenant les lames et en descendant, chaque aiguille traverse l'anneau qui lui est associé et dépose quelques dizaines de picolitres sur la lame. A la suite d'essais, il s'est avéré que deux passages successifs de l'aiguille dans l'anneau conduisaient à l'obtention de dépôts de meilleure qualité, aussi bien morphologiquement qu'au niveau de l'homogénéité des quantités de liquide déposées. Ainsi, les dépôts des 32 clones des lames ont été obtenus par deux passages de l'aiguille dans l'anneau.

Afin de choisir l'ADN polymérase la mieux adaptée pour la production des ADNc à déposer sur lames (sondes), les lames « essai » ont été hybridées avec des cibles marquées au cy3 et obtenues à partir de 20 µg d'ARN issus d'explants foliaires du génotype K59, prélevés au jour 10 de culture d'embryogenèse somatique. Nous avons mesuré alors pour chaque spot, la différence, exprimée en \log_2 , entre la moyenne du signal émis par le cy3 et le bruit de fond autour du spot. Pour les trois ADN polymérases utilisées pour la production des sondes, une représentation en boîtes à moustaches (Figure 26) des distributions des intensités enregistrées a été construite (Figure 27). On remarque que les trois enzymes montrent des distributions similaires, avec une médiane des intensités proche pour les trois enzymes utilisées. En raison de son coût avantageux, le choix s'est reporté sur l'ADN polymérase Biolabs.

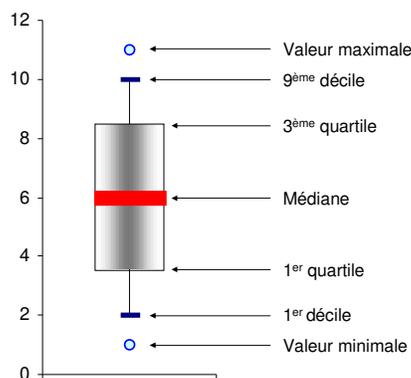


Figure 26 : Représentation d'une distribution sous forme d'une boîte à moustaches (box plot)
Tous les box plots représentés suivent ce modèle.

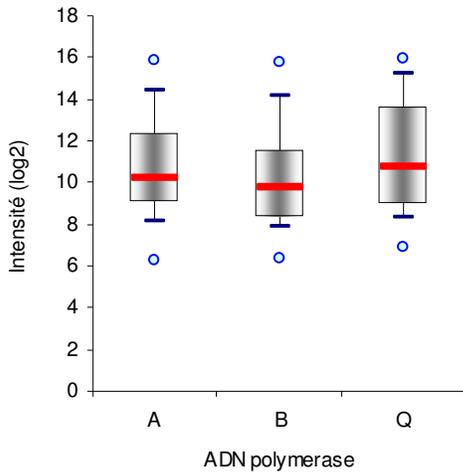


Figure 27 : Effet de l'ADN polymérase utilisée pour la synthèse des sondes sur l'intensité des signaux enregistrés
 Les boîtes à moustaches représentent la distribution des intensités enregistrées pour les 32 ADNc déposés six fois sur lame. Les clones ont été amplifiés par trois enzymes différentes : A : Ampli Taq polymérase de Applied, B : Taq polymérase de Biolabs ; Q : Hot Start Taq polymérase de Qiagen

2) Choix de la quantité d'ARN nécessaire à la fabrication des cibles

Afin d'estimer s'il était possible de limiter la consommation d'ARN nécessaire à la fabrication des cibles, nous avons entrepris trois hybridations sur des lames « essai » distinctes avec des cibles marquées au cy3 et obtenues à partir de 5, 10 ou 20 μg d'ARN. L'intensité du signal d'hybridation obtenu pour chaque spot a été calculée comme précédemment. Nous remarquons que la distribution des intensités des spots ne varie quasiment pas pour les différentes quantités d'ARN utilisées lors de la fabrication des cibles (Figure 28). Les résultats de chaque condition étant obtenus à partir de lames différentes, les faibles variations observées pourraient simplement être dues à la variabilité qui existe entre les lames. Pour la suite de ce travail, nous avons choisi d'utiliser 10 μg d'ARN pour la fabrication des cibles.

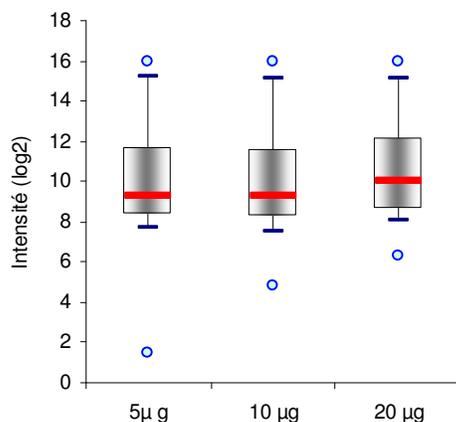


Figure 28 : Effet de la quantité d'ARN utilisée lors de la fabrication des cibles sur l'intensité du signal enregistré
 Les boîtes à moustaches représentent la distribution des intensités enregistrées pour les 32 ADNc déposés 18 fois sur la lame (en combinant les produits obtenus par les 3 enzymes utilisées).

C. Production des sondes et des lames

Au total, 1470 clones, 1229 provenant de la banque E et 241 de la banque NE, ont été amplifiés par PCR. D'après le regroupement des séquences représentant des gènes identiques effectué auparavant, les clones choisis représentent un total de 1098 gènes, c'est-à-dire 53% du nombre total de gènes des deux banques.

Les clones ont été amplifiés chacun trois fois, et les produits obtenus ont été mélangés, et purifiés, afin principalement d'éliminer les amorces de PCR. Un échantillon représentant 648 des 1470 produits PCR purifiés, a été dosé par spectrophotométrie.

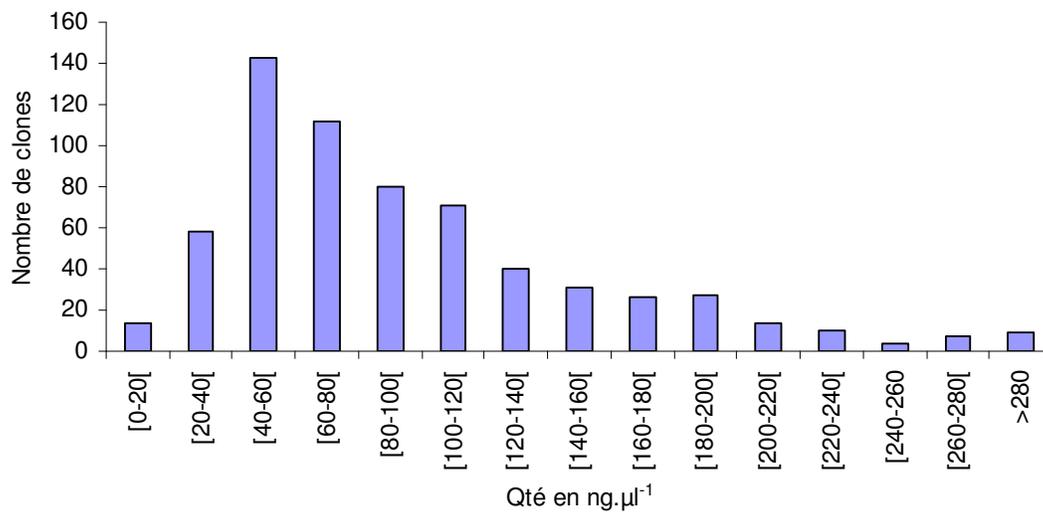


Figure 29 : Distribution par classe des quantités d'ADN produits à partir de 648 clones.

Le but était de déposer sur les lames des quantités égales d'ADN pour chacun des clones. Cependant, au regard de la diversité des quantités d'ADN obtenues après PCR (Figure 29), nous avons considéré la classe médiane des quantités d'ADN produites ([40-60[ng.µl⁻¹), comme représentative de l'ensemble des quantités. Les clones ont été réduits à sec et repris dans le tampon de « spotting » en considérant que la concentration initiale de chacun des ADNc était de 50 ng.µl⁻¹. Les produits PCR ont ensuite été déposés sur lame.

D. Production des cibles

1) Obtention de la référence

La « référence » utilisée dans le schéma expérimental indirect se devait d'être disponible en grande quantité afin de pouvoir être comparée à toutes les conditions « test », que les aliquots de cette « référence » soient homogènes et qu'un maximum des ADNc

déposés sur les lames puissent être hybridés avec les cibles obtenues à partir de la référence. Nous avons choisi comme condition « référence » un mélange d'ARN issus d'explants foliaires des génotypes K59 et C15 prélevés entre les jours 0 et 10 de culture d'embryogenèse somatique (Tableau 8). La quantité totale d'ARN du mélange est de 3553 µg, ce qui permet en théorie, à raison de 10µg d'ARN consommés par cible synthétisée, la comparaison de 88 conditions biologiques.

Conditions	Quantité d'ARN (en µg)
K59J0	247
K59J2	184
K59J4	557
K59J6	329
K59J8	570
K59J10	781
C15J0	185
C15J2	208
C15J4	128
C15J6	101
C15J8	119
C15J10	146
Total	3553

Tableau 8 : Composition de la référence utilisée pour le schéma expérimental indirect de microarray. Pour chaque condition, le premier argument représente le génotype de chicorée (K59 ou C15), et le second le jour de collecte des explants au cours de la culture en conditions d'embryogenèse somatique (J0 à J10).

La qualité de tous les ARN utilisés pour l'obtention de la référence a été validée par une électrophorèse capillaire (Figure 30). On remarque sur l'électrophorégramme, la présence de bandes comprises entre 1000 et 1500 pb qui disparaissent au fur et à mesure de la culture pour les deux génotypes. La culture d'embryogenèse somatique se déroulant à l'obscurité, ces bandes correspondent probablement à des ARN chloroplastiques, ou à des ARNm liés à l'activité photosynthétique. Par exemple, les bandes à 500 et à 1500 pb représentent probablement les ARNm codant respectivement les petites et grandes sous-unités de la rubisco.

La qualité des ARN des conditions « test » qui ont été hybridés sur les lames en simultané avec la référence a également été vérifiée par électrophorèse capillaire.

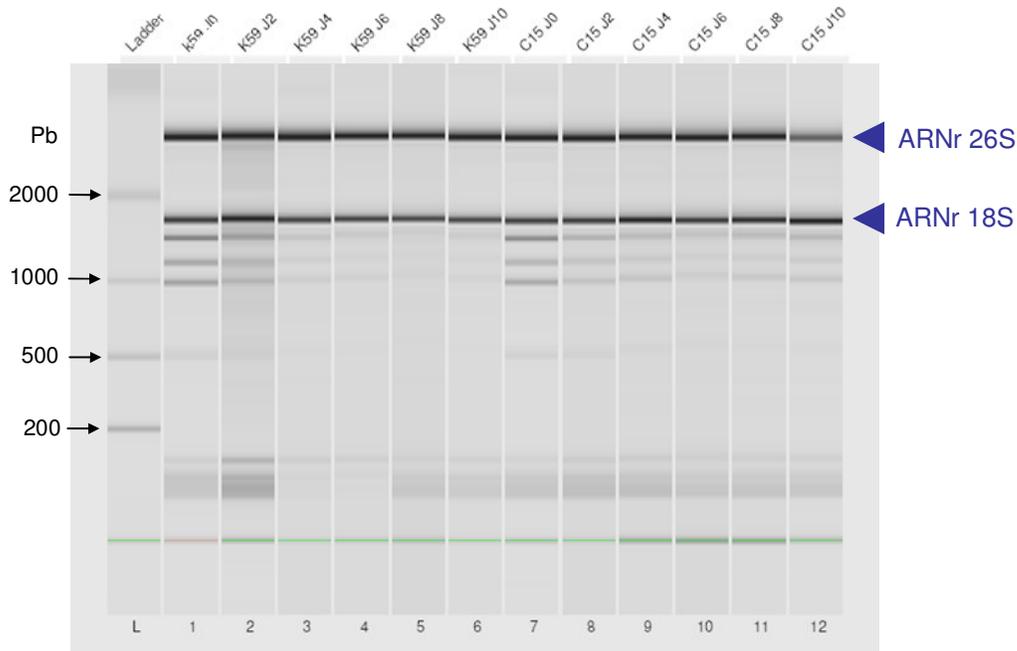


Figure 30 : Electrophorégramme des différents ARN utilisés pour l'obtention de la référence. On note la qualité des ARN par l'intégrité des ARNr 26S et 18S.

2) Synthèse des cibles et hybridations

Après leur synthèse, les cibles ont été couplées suivant l'expérience, purifiées, et l'incorporation des fluorochromes dans le mélange a été vérifiée par un spectre d'absorption de 220 à 700 nm (Figure 31).

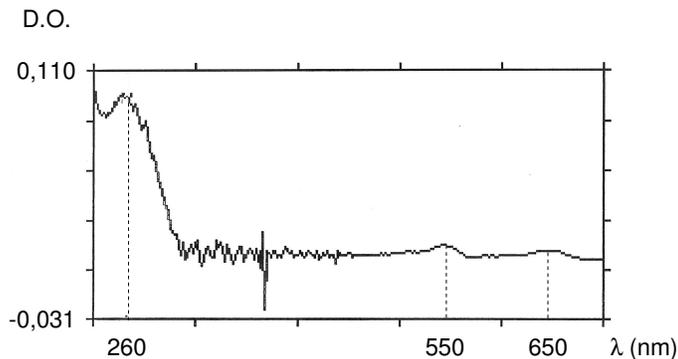


Figure 31 : Validation d'une paire de cibles avant hybridation

Le graphique représente le spectre d'absorption d'une dilution au 1/50^{ème} d'une paire de cibles purifiées, l'une étant marquée au cy3 et l'autre au cy5. La valeur d'absorbance à 260 nm est de 0,095. En utilisant la formule 1 unité de DO (260 nm)=50 ng d'ADN, on en déduit la concentration en ADNc synthétisé qui est de 237,5 ng.μl⁻¹. La masse de 1 mole de dNTP étant de 330 g, la concentration molaire de dNTP du mélange est donc $C_{dNTP}=7,20E^{-4}$ M. Les densités optiques mesurées à 550 et à 650 nm correspondent respectivement à la quantité de dCTP-cy3 et de dCTP-cy5 dans le mélange des cibles. A l'aide de la formule $A=\epsilon[C]L$ où A

est l'absorbance, ϵ le coefficient d'extinction molaire de la cyanine, C la concentration et L la largeur de la cuve du spectrophotomètre, on en déduit les concentrations molaires en dCTP-cy3 et dCTP-cy5. Respectivement, à partir de DO mesurées de 0,011 et de 0,08, on calcule donc $C_{\text{dCTP-cy3}}=3,66E^{-6}$ M et $C_{\text{dCTP-cy5}}=1,6E^{-6}$ M. Par le ratio de C_{dNTP} sur $C_{\text{dCTP-cy3}}$ ou sur $C_{\text{dCTP-cy5}}$, on obtient les fréquences de dCTP-cy3 et de dCTP-cy5 dans le mélange des cibles qui sont respectivement de 1 dCTP-cy3 pour 196 dNTP et de 1 dCTP-cy5 pour 450 dNTP.

Le spectre d'absorption permet de vérifier si la synthèse des cibles s'est faite correctement. Un premier pic à 260 nm correspond à l'ADNc marqué synthétisé, et deux pics, respectivement à 550 et à 650 nm, représentent la fluorescence émise par les cy3 et les cy5. Le spectre d'absorption permet aussi de calculer une fréquence théorique d'incorporation des cyanines. Sur l'exemple donné, Figure 31, on mesure qu'il y a dans le mélange des cibles, un résidu dCTP-cy3 pour 196 résidus dNTP, alors qu'il y a un résidu dCTP-cy5 pour 450 résidus dNTP. Les marquages réalisés ont permis de calculer des fréquences d'incorporation comprises entre 1 dCTP pour 150 à 500 dNTP. Cette variation de la fréquence d'incorporation n'a pas montré d'incidence majeure sur les valeurs moyennes d'intensités enregistrées, en raison de la sensibilité du scanner utilisé.

3) Normalisation intra-lame

L'intensité de chaque signal d'hybridation, calculée ici comme la médiane du signal moins la médiane du bruit de fond autour du dépôt, a été normalisée par régression non linéaire de type Lowess (Local Weighted Scatter Smoothing), réalisée entre les 8 zones de la lame obtenues par les dépôts respectifs des 8 aiguilles du robot. Cette normalisation permet la détection et la correction d'une éventuelle incorporation différentielle des fluorochromes cy5 et cy3 lors de la fabrication des cibles. Sur un graphique représentant chaque spot, avec en abscisse l'intensité enregistrée pour le cy3 et en ordonnée celle enregistrée pour le cy5, l'ensemble des points forme un nuage. Sur la Figure 32, on remarque que dans le premier cas (Figure 32 A), le nuage de points des intensités des spots est quasiment rectiligne et ne montre pratiquement pas de fléchissement vers l'un ou l'autre des fluorochromes. Ce résultat signifie qu'il n'y a pas de biais d'incorporation des fluorochromes et que les distributions des intensités enregistrées pour les deux fluorochromes sont comparables. La normalisation intra-lame, dans ce cas, ne modifie quasiment pas les données initiales et donc l'allure générale du nuage de points. On remarque simplement que la base du nuage de points, correspondant aux spots présentant des valeurs de faible intensité, est ici recentrée (Figure 32 B). Cet exemple est représentatif de la majorité des lames, cependant dans certains cas, la normalisation intra-lame s'avère indispensable. Par exemple, dans le second cas (Figure 32 C), on observe un

nuage de point en forme de « banane » qui traduit une incorporation préférentielle du cy3 dans les cibles en comparaison au cy5. Dans cet exemple, la normalisation intra-lame corrige fortement l'intensité des signaux émis par les deux cyanines (Figure 32 D). Après normalisation, le nuage de points a retrouvé une apparence rectiligne, comme observée précédemment (Figure 32 B). Pour ce dernier exemple, on peut vérifier également l'efficacité de cette normalisation à l'aide d'une représentation en boîtes à moustaches qui montre effectivement la correction des distributions des intensités des cy3 et cy5 (Figure 33).

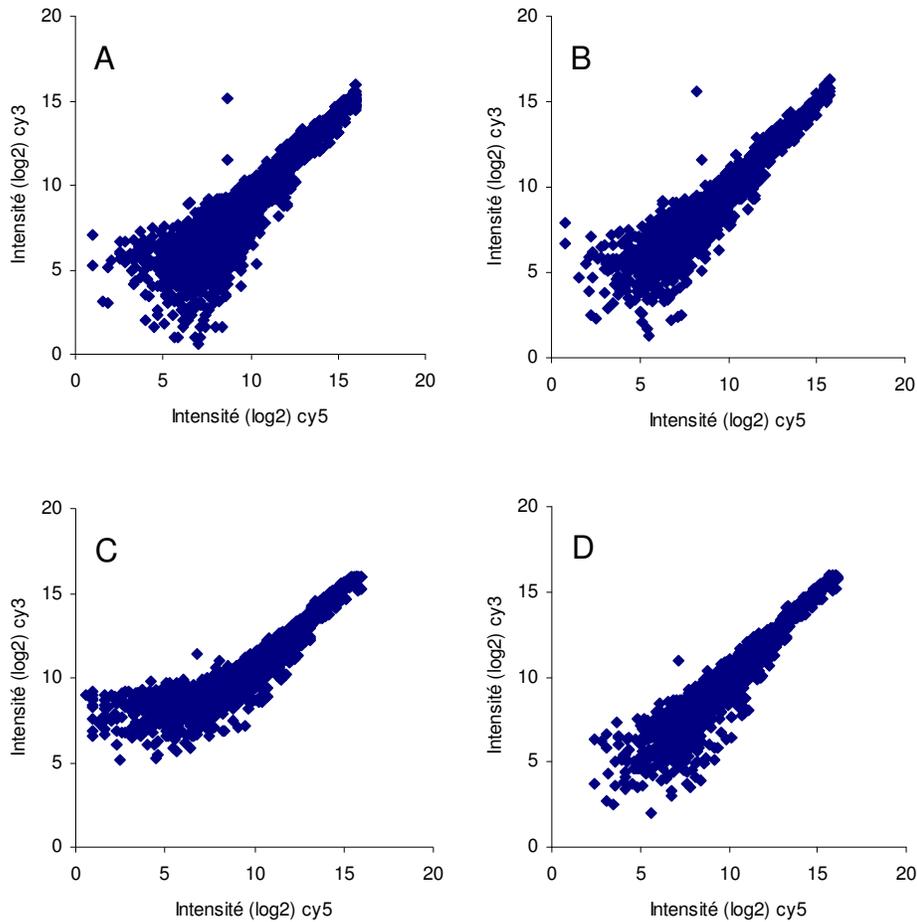


Figure 32 : Effet de la normalisation intra-lame sur l'intensité des signaux enregistrés
 Lame 1 de la condition K59J4 avant (A) et après normalisation (B) et lame 1 de la condition K59J6 avant (C) et après normalisation (D). Pour chaque spot, les coordonnées du point correspondant sont données par la valeur d'intensité enregistrée pour le cy5 en abscisse et la valeur d'intensité enregistrée pour le cy3 en ordonnée.

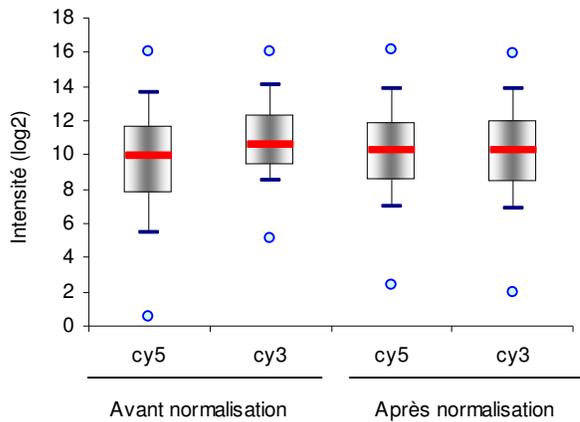


Figure 33 : Observation de l'effet de la normalisation intra-lame à l'aide de représentations en boîtes à moustaches des données des Figure 32 C et D

4) Mise en évidence de gènes exprimés différemment

Après la normalisation intra-lame, pour chaque expérience (génotype/jour), le ratio en logarithme à base 2 entre la condition « test » et la condition « référence » a été calculé pour chaque spot. Comme décrit précédemment (Partie II A, Figure 25), les ratios test/référence, obtenus pour chaque expérience, ont été comparés à celui obtenu pour le jour 0. Ces comparaisons ont permis d'une part de détecter les gènes dont les cibles sont préférentiellement marquées par l'un des deux fluorochromes, et d'autre part de mettre en évidence les gènes présentant une variation significative au cours de la cinétique en conditions d'embryogenèse somatique chez l'un et/ou l'autre des deux génotypes.

a) Mise en évidence des gènes dont les ADNc présentent une incorporation préférentielle d'un fluorochrome

Pour chaque couple d'expérience, génotype/jour X (1 à 6) comparé à génotype/jour 0, les données des combinaisons de fluorochromes 1 et 2 ont été soumises indépendamment à une analyse de variance en utilisant la procédure GLM (General Linear Model) du logiciel de statistique SAS (SAS Institute Inc. 1989). Le modèle d'analyse utilisé a été le suivant : $Y_{ij} = \mu + J_i + S_j + e_{ij}$ où Y_{ij} est le \log_2 du ratio des intensités, μ est la moyenne générale, J_i est l'effet jour, S_j est l'effet lame (effet de la répétition) et e_{ij} est la résiduelle du modèle.

La différence d'expression des gènes entre le jour X (J1 à J6) et le jour 0 a été estimée par le ratio d'expression $ER(\log_2) = \log_2(\text{jour X/référence}) - \log_2(\text{jour 0/référence})$, et la significativité associée à l'effet jour a été déterminée à partir de la probabilité associée à la valeur de F (Fisher-Snedecor) de l'effet jour. Les gènes préférentiellement exprimés au jour X

ont donc un ratio qui est positif, alors que ceux préférentiellement exprimés au jour 0 ont un ratio négatif. Les gènes montrant pour les 2 combinaisons de fluorochromes un effet jour significatif ($p < 0,01$), mais des ratio d'expression avec des signes opposés, ont été considérés comme des gènes présentant un biais d'incorporation d'un des deux fluorochromes et ont été éliminés des analyses suivantes.

b) Calcul des ratios d'expression

Pour les gènes représentés par un seul EST sur les lames de microarray, les données ont ensuite été analysées par le modèle d'analyse de variance suivant : $Y_{ij} = \mu + J_i + D_j + JD_{ij} + e_{ij}$, avec Y_{ij} , μ , G_i , et e_{ij} décrits plus haut et où D_j est l'effet de la combinaison des fluorochromes, et JD_{ij} est l'interaction jour \times combinaison de fluorochromes (Kerr, 2003b). Lorsque les gènes étaient représentés par plusieurs ESTs sur les lames, le modèle a été : $Y_{ij} = \mu + E_i + J_j + D_k + JD_{jk} + e_{ijk}$, E_i étant l'effet EST.

Dans le but de limiter le nombre de faux positifs, une mesure de significativité pour chacun des gènes a été obtenue par le calcul de la valeur q , basé sur la probabilité (p -value) associée au test F de l'effet jour. Une mesure de valeur q est particulièrement adaptée lorsque de nombreux tests sont réalisés en simultané comme c'est le cas sur une lame de microarray. La valeur q indique la proportion attendue du nombre de faux positifs sur le nombre de tests significatifs identifiés par l'analyse (Storey and Tibshirani, 2003). Le calcul a été réalisé à l'aide du programme QVALUE, intégré au logiciel R.

Les gènes représentés par un EST unique étaient considérés comme différentiellement exprimés lorsque la valeur q était inférieure à 0,01 et lorsque que le ratio d'expression (ER) était supérieur ou égal, ou inférieur ou égal, aux seuils fixés. Le seuil appliqué au ratio d'expression a été déterminé empiriquement en calculant la moyenne (μ) et l'écart type (sd) de tous les ratios d'expression correspondants à des gènes qui présentent un effet jour non significatif (Figure 34). Nous avons considéré que les ratios d'expression en dehors de l'intervalle $\mu \pm 1.96 \text{ sd}$ (95% de confiance) devaient correspondre à des gènes différentiellement exprimés. Les gènes, représentés par au moins deux ESTs, ont été considérés comme étant différentiellement exprimés lorsque la valeur q était inférieure à 0,001 et lorsque le ER (\log_2) était supérieur ou égal au seuil fixé de la même façon que pour les gènes représentés par un EST unique (Figure 34).

Par exemple, lors de la comparaison de l'expérience K59J2 à K59J0 (Figure 34), pour les ESTs représentés par un seul EST (a), la valeur seuil de q de 0,01 correspond à une valeur

p de 0,0048. Les seuils fixés au ratio d'expression sont de 0,47 pour un gène préférentiellement exprimé au jour 2 et de -0,52 pour un gène préférentiellement exprimé au jour 0. Pour les ESTs représentés par plusieurs ESTs (b), la valeur seuil q de 0,001 correspond à une probabilité de 0,0004 et les seuils fixés, comme précédemment, au ratio d'expression sont de 0,47 et de -0,52.

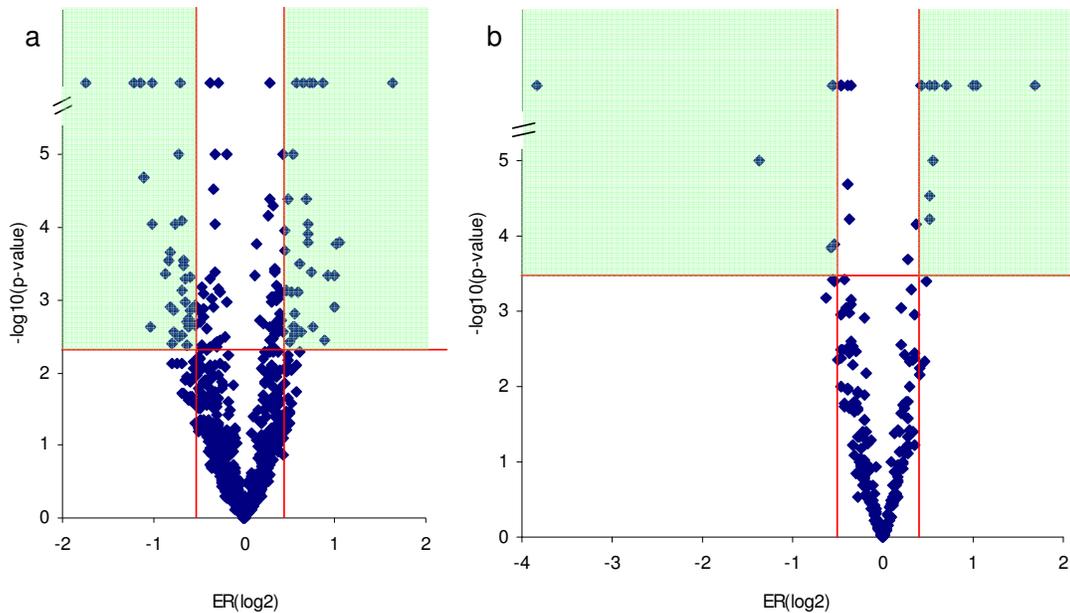


Figure 34 : Graphiques en volcan représentant les gènes analysés par microarray pour la comparaison des expériences K59J2 et K59J0

a : gènes représentés par un seul EST, b : gènes représentés par au moins deux ESTs.

ER(log₂) : ratio d'expression entre les explants foliaires issus du génotype embryogène, prélevés au deuxième jour (J2) et au point initial (J0) de culture en conditions d'embryogenèse somatique. La ligne horizontale indique le $-\log_{10}$ du seuil de probabilité ($p=0,0048$ pour a; $p=0,0004$ pour b correspondant respectivement à $q=0,01$ et $q=0,001$) associé à l'effet jour. Les lignes verticales indiquent les seuils appliqués au ratio d'expression et calculés comme $\mu \pm 1,96 \times sd$ où μ est la moyenne et sd l'écart-type des valeurs des ratios d'expression situés sous le seuil de $-\log_{10}(p\text{-value})$ appliqué. Les gènes exprimés différemment sont situés dans les zones colorées qui sont délimitées par l'intersection des valeurs des deux seuils déterminés (valeur p et ratio d'expression).

5) Validation du protocole d'analyses

Dans un premier temps, avant de débiter l'analyse en cinétique, nous avons décidé de nous intéresser à l'expression différentielle des gènes entre les deux génotypes au jour 4 de culture en conditions d'embryogenèse somatique. Cette première expérience avait pour but de mettre au point la totalité du processus d'analyse de microarray et de le valider. D'autre part,

cette analyse devait permettre d'identifier, au sein des banques soustractives qui ont également été fabriquées à partir d'explants prélevés au jour 4 de culture, des gènes exprimés différemment. C'est donc ce point de la cinétique qui devait théoriquement présenter le maximum de gènes différemment exprimés entre les deux génotypes. Cette analyse devait donner une première idée de la réponse des deux génotypes aux conditions de culture appliquées.

Les modèles d'analyse de variance utilisés pour cette comparaison sont les mêmes que ceux présentés précédemment, en remplaçant l'effet jour par un effet génotype. Les analyses par microarray ont mis en évidence 105 gènes (9,6% du nombre total de gènes représentés sur les lames de microarray) exprimés différemment entre les deux génotypes au jour 4 de culture, avec respectivement, 37 gènes exprimés préférentiellement chez le génotype embryogène et 68 chez le génotype non-embryogène (Tableau 9).

Les classifications fonctionnelles des gènes préférentiellement exprimés chez chacun des deux génotypes montrent de fortes différences. Parmi les gènes préférentiellement exprimés chez le génotype embryogène (Tableau 9a), 14 codent des protéines liées à la synthèse protéique. Ces 14 protéines sont, en fait, toutes des protéines ribosomales. Six gènes ne présentent pas d'homologie avec les bases de données (16%), et 5 gènes codent des protéines liées au métabolisme (14%). Les gènes préférentiellement exprimés chez le génotype non-embryogène (Tableau 9b) codent principalement des protéines liées au métabolisme (19 gènes, 28%) et des protéines aux fonctions inconnues (19 gènes, 28%). Dix gènes (15%) ne présentent pas d'homologies avec les bases de données et 7 gènes (10%) sont reliés au transport cellulaire. Les autres classes sont représentées par moins de 10% des gènes préférentiellement exprimés chez C15. La classification révèle que seulement deux gènes (3%) sont liés à la synthèse protéique et qu'aucun ne code une protéine ribosomale. Pour les deux classifications, les contigs ou OC qui ne présentent pas d'homologies montrent rarement un cadre de lecture (ORF), et dans quelques cas, une queue polyA a pu être retrouvée dans la séquence. Ce résultat suggère qu'au moins une partie des séquences qui ne présentent pas d'homologie correspond à des régions non traduites.

Tableau 9 : Liste des gènes identifiés comme étant exprimés différemment au quatrième jour de culture d'embryogenèse somatique par microarray

a : gènes préférentiellement exprimés chez le génotype embryogène, b : gènes préférentiellement exprimés chez le génotype non embryogène

Les gènes ont été classés par fonction suivant le catalogue fonctionnel de MIPS (Munich Information center for Proteins Sequences : <http://mips.gsf.de/>). La classe 'protéines inconnues' regroupe les sous-classes 'localisation cellulaire', 'classification non certaine' et 'protéines non classées'. Num acc: numéro d'accèsion d'un EST représentant le contig ou l'OC (groupe réunissant les ESTs identiques) ; Contig/OC: numéro du contig ou de l'OC ; N-EST: nombre d'ESTs représentant le contig ou l'OC; recherche BlastX NR: recherché BlastX contre la base de données non redondantes de GenBank (<http://www.ncbi.nlm.nih.gov/>) ; BlastX At: recherché BlastX contre les séquences codantes traduites d'Arabidopsis provenant de AtGDB (*Arabidopsis thaliana* Genome Database, <http://www.plantgdb.org/AtGDB/>) ; N: nombre d'ESTs du contig ou de l'OC analysés par microarray ; ER: ratio d'expression (\log_2) entre les génotypes embryogène et non embryogène au jour 4 de culture d'embryogenèse somatique.

Résultats et discussion

Tableau 9 a

Classe de fonction	Num acc	Contig/OC	N-EST	Recherche BlastX contre la base de données GenBank NR		Recherche Blast X contre AtGDB		Analyses pas microarray		
				Fonction [<i>Genre espèce</i>]	E-value	ID	E-value	N	q-value	ER (log ₂)
Metabolism	DT213258	OC0003_a	5	glutathione S-transferase GST 13 [<i>G. max</i>]	2.00E-38	At3g09270	8.00E-34	3	0.00027	0.50
	DT212580	OC1044	2	CTP synthase-like protein [<i>A. thaliana</i>]	1.00E-75	At4g02120	4.00E-90	1	0.00000	0.57
	DT212740	OC1143	2	ribonucleotide reductase R2 [<i>N. tabacum</i>]	2.00E-36	At3g27060	1.00E-44	2	0.00000	0.86
	DT213673	OC1752	1	lactate dehydrogenase-2 [<i>S. tuberosum</i>]	2.00E-37	At4g17260	8.00E-37	1	0.00000	0.59
	DT213399	OC0316	11	alcohol dehydrogenase [<i>L. sativa</i>]	1.00E-100	At1g77120	2.00E-93	2	0.00084	0.48
Cell cycle and DNA processing	DT211083	Cont0077	9	salt tolerance protein 5 [<i>B. vulgaris</i>]	6.00E-49	At5g53400	3.00E-49	4	0.00000	0.60
Protein synthesis	DT213363	Cont0010	7	putative ribosomal protein S10 [<i>A. thaliana</i>]	2.00E-53	At4g25740	2.00E-55	3	0.00000	0.51
	DT213435	Cont0023	4	40S ribosomal protein S6 [<i>C. arietinum</i>]	2.00E-54	At4g31700	6.00E-81	2	0.00015	0.50
	DT213022	Cont0030	4	ribosomal protein L7 [<i>S. tuberosum</i>]	6.00E-36	At3g13580	6.00E-38	2	0.00058	0.50
	DT213255	OC0082_a	8	60s acidic ribosomal protein [<i>H. orientalis</i>]	9.00E-24	At2g27710	1.00E-39	4	0.00000	0.55
	DT212642	OC0378_a	1	60S ribosomal protein L5 [<i>S. melongena</i>]	2.00E-35	At3g25520	2.00E-40	1	0.00068	0.57
	DT213516	OC0390	6	40S ribosomal protein S2 [<i>A. thaliana</i>]	1.00E-94	At2g41840	4.00E-97	3	0.00000	0.72
	DT213097	OC1359_b	1	ribosomal protein L17 [<i>C. sativa</i>]	3.00E-10	At2g33370	9.00E-17	1	0.00300	0.98
	DT211061	OC1549_a	3	40S ribosomal protein S17 [<i>L. esculentum</i>]	4.00E-55	At5g04800	3.00E-56	2	0.00000	0.71
	DT212942	OC1821	2	40S ribosomal protein S11 [<i>E. esula</i>]	3.00E-63	At3g48930	1.00E-80	1	0.00000	0.89
	DT213592	OC1822_b	1	putative ribosomal protein L10a [<i>O. sativa</i>]	3.00E-89	At1g08360	1.00E-100	1	0.00000	0.62
	DT210916	OC1833_b	4	ribosomal protein L3A [<i>N. tabacum</i>]	9.00E-70	At1g61580	1.00E-82	2	0.00062	0.45
	DT211087	OC2012_b	2	putative ribosomal protein S14 [<i>O. sativa</i>]	1.00E-58	At3g52580	9.00E-69	1	0.00049	0.65
	DT212711	OC2022	6	ribosomal protein S3a [<i>C. roseus</i>]	1.00E-92	At4g34670	4.00E-89	2	0.00058	0.47
	DT213478	OC2153_b	3	ribosomal Pr 117 [<i>T. aestivum</i>]	2.00E-49	At3g04400	3.00E-50	1	0.00000	0.69
	Protein fate (folding, modification, destination)	DT212506	OC1817_b	2	mitochondrial chaperonin HSP60 [<i>A. thaliana</i>]	2.00E-49	At2g33210	2.00E-51	1	0.00171
Cellular transport, transport facilitation and transport routes	DT211032	OC1225_b	1	34 kDa porin [<i>S. tuberosum</i>]	3.00E-60	At3g01280	3.00E-56	1	0.00000	0.94
Cellular communication/ signal transduction mechanism	DT212523	OC1011	1	GTP-binding protein-related [<i>A. thaliana</i>]	4.00E-12	At3g63150	6.00E-19	1	0.00302	0.59
	DT213317	OC1495	1	ZEITLUPE [<i>M. crystallinum</i>]	2.00E-24	At5g57360	1.00E-29	1	0.00095	0.60
	DT213352	OC1929	5	G protein beta subunit [<i>N. plumbaginifolia</i>]	3.00E-33	At1g48630	2.00E-30	4	0.00000	0.68
Cell rescue, defense and virulence	DT212721	Cont0019	6	putative heat shock protein 81-2 [<i>A. thaliana</i>]	5.00E-59	At5g56030	8.00E-66	3	0.00000	0.51
Development	DT212718	OC1129	1	F-box protein family, AtFBL20 [<i>A. thaliana</i>]	3.00E-34	At4g05460	4.00E-39	1	0.00940	0.73
Unknown proteins	DT212234	OC0687	8	arabinogalactan protein [<i>D. carota</i>]	1.00E-16	At2g33790	7.00E-16	4	0.00000	1.85
	DT213606	OC1702	1	putative histone H2B [<i>A. thaliana</i>]	7.00E-22	At3g46030	7.00E-22	1	0.00018	0.86
	DT213275	OC1471	1	unknown protein [<i>A. thaliana</i>]	2.00E-06	At2g02880	5.00E-09	1	0.00427	0.68
	DT213573	OC1775	2	phosphatidylinositol-4-phosphate 5-kinase family protein [<i>A. thaliana</i>]	6.00E-52	At3g14270	7.00E-54	2	0.00092	0.55
No hits found	DT212279	OC0859	1	putative gag-pol polyprotein [<i>O. sativa</i>]	6.00E-25	No hits found		1	0.00345	1.17
	DT212270	OC0852	1	No hits found		No hits found		1	0.00053	0.57
	DT212539	OC1020	2	No hits found		No hits found		2	0.00000	1.10
	DT212839	OC1202	2	No hits found		No hits found		1	0.00007	0.82
	DT213562	OC1678	2	No hits found		No hits found		1	0.00086	0.77
	DT213640	OC1839_b	2	No hits found		No hits found		1	0.00000	1.11

Résultats et discussion

Tableau 9 b

Classe de fonction	Num acc	Contig/OC	N-EST	Recherche BlastX contre la base de données GenBank NR		Recherche Blast X contre AtGDB		Analyses pas microarray		
				Fonction [<i>Genre espèce</i>]	E-value	ID	E-value	N	q-value	ER (log ₂)
Metabolism	DT212690	Cont0009	2	P protein [<i>F. pringlei</i>]	3.00E-95	At4g33010	1.00E-94	1	0.00961	-0.79
	DT213594	OC1696	3	P protein [<i>F. pringlei</i>]	1.00E-144	At2g26080	1.00E-150	2	0.00055	-0.57
	DT213116	OC0603_a	6	methionine synthase [<i>S. tuberosum</i>]	1.00E-107	At3g03780	1.00E-106	4	0.00000	-0.98
	DT212513	OC0603_b	8	cobalamine-independent methionine synthase [<i>S. scutellarioides</i>]	1.00E-103	At3g03780	1.00E-102	4	0.00000	-0.81
	DT212209	OC0818	2	glutamate synthase (ferredoxin) [<i>N. tabacum</i>]	2.00E-48	At5g04140	7.00E-49	1	0.00053	-0.65
	DT213235	OC1447	1	aminoacylase, putative [<i>A. thaliana</i>]	1.00E-105	At4g38215	1.00E-112	1	0.00045	-0.86
	DT213358	OC1528	1	NADH-glutamate dehydrogenase [<i>N. plumbaginifolia</i>]	1.00E-106	At5g18170	1.00E-107	1	0.00426	-0.84
	DT210918	Cont0201	5	putative auxin-induced protein [<i>A. thaliana</i>]	8.00E-59	At1g60730	2.00E-61	1	0.00000	-1.12
	DT212742	Cont0052	14	delta12-fatty acid acetylenase [<i>D. sinuata</i>]	1.00E-111	At3g12120	7.00E-82	5	0.00000	-0.56
	DT210875	Cont0060	23	3-hydroxy-3-methylglutaryl coenzyme A reductase [<i>S. tuberosum</i>]	1.00E-133	At1g76490	1.00E-136	9	0.00000	-1.34
	DT213526	Cont0073	2	putative glycerol-3-phosphate dehydrogenase [<i>A. thaliana</i>]	6.00E-92	At3g10370	1.00E-104	1	0.00000	-1.44
	DT212244	Cont0090	2	farnesyl diphosphate synthase [<i>P. argentatum</i>]	5.00E-50	At5g47770	2.00E-45	1	0.00000	-0.89
	DT212708	OC1121	1	isopentenyl pyrophosphate:dimethylallyl pyrophosphate isomerase [<i>L. sativa</i>]	1.00E-87	At5g16440	7.00E-68	1	0.00824	-0.98
	DT213158	OC1393	2	isopentenyl pyrophosphate:dimethylallyl pyrophosphate isomerase [<i>L. sativa</i>]	1.00E-48	At3g02780	2.00E-45	1	0.00056	-0.82
	DT212217	Cont0006	7	1-aminocyclopropane-1-carboxylate oxidase [<i>P. hortorum</i>]	1.00E-47	At1g05010	5.00E-46	1	0.00000	-1.32
	DT211055	OC0168	9	1-aminocyclopropane-1-carboxylate oxidase [<i>L. esculentum</i>]	2.00E-60	At1g62380	4.00E-58	1	0.00000	-1.45
	DT211036	OC0023_a	14	glutathione transferase [<i>H. muticus</i>]	7.00E-68	At1g02920	1.00E-59	3	0.00000	-1.17
	DT210895	OC0075	1	4-coumarate:coenzyme A ligase [<i>N. tabacum</i>]	3.00E-31	At3g21240	5.00E-28	1	0.00060	-0.60
	DT213525	OC1648	1	trans-cinnamate 4-monooxygenase [<i>Z. elegans</i>]	8.00E-28	At2g30490	2.00E-38	1	0.00145	-0.63
	Energy	DT213193	OC1802	4	photosystem I chain II precursor [<i>S. oleracea</i>]	2.00E-77	At1g03130	7.00E-81	2	0.00058
Transcription	DT213494	Cont0078	5	RNA Binding Protein 47 [<i>N. plumbaginifolia</i>]	2.00E-77	At3g19130	2.00E-82	2	0.00000	-0.71
	DT212682	OC1104	1	zinc finger protein [<i>P. radiata</i>]	3.00E-26	At5g57660	7.00E-28	1	0.00980	-0.62
	DT211017	OC1623_b	2	DNA/RNA binding protein-like [<i>A. thaliana</i>]	1.00E-48	At1g49600	3.00E-52	2	0.00000	-0.69
Protein synthesis	DT213474	Cont9010	3	translation initiation factor B04 [<i>H. annuus</i>]	2.00E-57	At1g54290	1.00E-48	1	0.00551	-0.70
	DT210871	Cont0071	8	poly(A)-binding protein [<i>N. tabacum</i>]	2.00E-41	At4g34110	1.00E-51	3	0.00008	-0.71
Protein fate (folding, modification, destination)	DT210904	OC0081	1	putative chloroplast outer membrane protein [<i>O. sativa</i>]	3.00E-41	At4g02510	4.00E-40	1	0.00000	-1.05
Cellular transport, transport facilitation and transport routes	DT213069	Cont0096	12	lipid transfer protein [<i>H. annuus</i>]	5.00E-37	At5g59320	2.00E-26	3	0.00000	-1.26
	DT213417	OC1568	1	putative sugar transporter protein [<i>A. thaliana</i>]	8.00E-43	At4g35300	2.00E-48	1	0.00245	-0.60
	DT210994	Cont0031	3	putative MATE efflux family protein [<i>A. thaliana</i>]	2.00E-49	At1g12950	2.00E-51	1	0.00023	-0.65
	DT212558	OC1030	1	plasma membrane intrinsic protein 2 [<i>T. baeoticum</i>]	3.00E-19	At1g01620	3.00E-17	1	0.00000	-0.88
	DT212748	OC1149	4	sucrose transporter [<i>G. max</i>]	2.00E-63	At1g22710	7.00E-53	2	0.00000	-0.92
	DT213421	OC1590	4	plasma intrinsic protein 2,2 [<i>J. regia</i>]	6.00E-81	At5g60660	2.00E-81	3	0.00000	-0.65
	DT211073	Cont0007	5	synaptobrevin-related family protein [<i>A. thaliana</i>]	5.00E-73	At4g15780	2.00E-73	1	0.00483	-0.74
Cellular communication/signal transduction mechanism	DT212770	OC1165	1	shaggy kinase 4 [<i>P. hybrida</i>]	3.00E-72	At5g26751	5.00E-71	1	0.00028	-0.82

Résultats et discussion

Tableau 9 b (suite)

Classe de fonction	Num acc	Contig/OC	N-EST	Recherche BlastX contre la base de données GenBank NR		Recherche Blast X contre AtGDB		Analyses pas microarray		
				Fonction [<i>Genre espèce</i>]	E-value	ID	E-value	N	q-value	ER (log ₂)
Cell rescue, defense and virulence	DT212691	Cont0080	5	vegetative storage protein, VSP [<i>C. intybus</i>]	2.00E-45	At1g24020	5.00E-10	3	0.00000	-1.14
	DT213128	Cont9039	40	metallothionein 1 [<i>A. tripolium</i>]	2.00E-22	At5g02380	1.00E-21	13	0.00000	-1.04
	DT212949	Cont9071	14	peroxidase 3 [<i>S. baicalensis</i>]	1.00E-89	At5g05340	2.00E-93	5	0.00000	-2.58
	DT213375	OC1539	2	bacterial-induced class III peroxidase [<i>G. hirsutum</i>]	7.00E-15	At4g11290	3.00E-15	2	0.00000	-0.93
Interaction with the environment	DT211019	OC1582	7	snakin2 [<i>S. tuberosum</i>]	1.00E-27	At1g75750	2.00E-27	4	0.00000	-0.75
Unkonwn proteins	DT212510	OC1658	5	leucine-rich repeat family protein [<i>A. thaliana</i>]	2.00E-10	At3g22800	2.00E-10	4	0.00000	-0.93
	DT212755	Cont0085	5	NADPH-ferrihemoprotein reductase [<i>H. tuberosus</i>]	1.00E-109	At4g30210	1.00E-98	2	0.00000	-0.66
	DT212954	OC1635_a	5	arachidonic acid-induced DEA1 [<i>L. esculentum</i>]	3.00E-22	At2g45180	2.00E-36	2	0.00000	-1.25
	DT213509	OC1635_b	2	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein [<i>A. thaliana</i>]	4.00E-11	At2g45180	1.00E-19	1	0.00000	-1.27
	DT212233	OC1796	18	ring domain containing protein [<i>C. annuum</i>]	5.00E-49	At4g03510	7.00E-34	10	0.00000	-1.13
	DT212973	OC1824	3	HyPRP [<i>F. ananassa</i>]	1.00E-08	At2g45180	2.00E-15	2	0.00000	-0.93
	DT213590	OC1932	3	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein [<i>A. thaliana</i>]	1.00E-15	At1g12090	1.00E-15	2	0.00000	-1.24
	DT213425	OC2013_c	1	cationic peroxidase 2 precursor [<i>A. hypogaea</i>]	1.00E-19	At5g39580	8.00E-17	1	0.00007	-0.56
	DT213213	Cont0095	11	dermal glycoprotein precursor, extracellular [<i>D. carota</i>]	2.00E-89	At1g03220	1.00E-77	3	0.00000	-1.16
	DT210933	OC0098	1	putative protein [<i>A. thaliana</i>]	4.00E-08	At4g05400	1.00E-20	1	0.00730	-0.66
	DT212243	OC0839	1	dormancy-associated protein [<i>C. lanceolata</i>]	2.00E-32	At1g28330	2.00E-39	1	0.00079	-1.04
	DT212577	OC1043	1	unknown [<i>A. thaliana</i>]	2.00E-08	At4g32460	6.00E-10	1	0.00000	-0.77
	DT212681	OC1191	4	protease inhibitor 2 [<i>Z. elegans</i>]	9.00E-21	At2g38870	2.00E-17	3	0.00000	-1.09
	DT213021	OC1307	1	putative spermine/spermidine synthase [<i>G. max</i>]	3.00E-58	At2g31740	1.00E-63	1	0.00714	-0.72
	DT213182	OC1412	1	unknown [<i>A. thaliana</i>]	3.00E-29	At5g65650	7.00E-39	1	0.00603	-0.74
	DT213196	OC1419	2	proline-rich family protein [<i>A. thaliana</i>]	5.00E-12	At2g41420	5.00E-12	1	0.00045	-0.85
	DT213345	OC1517	2	PDI-like protein [<i>A. thaliana</i>]	6.00E-60	At1g60420	5.00E-62	2	0.00000	-0.64
	DT210833	OC0044	1	putative dipeptidyl peptidase IV [<i>O. sativa</i>]	8.00E-44	At5g24260	2.00E-28	1	0.00360	-0.58
	DT213388	OC1546	1	auxin transport protein (BIG) [<i>A. thaliana</i>]	2.00E-85	At3g02260	1.00E-96	1	0.00000	-1.11
	No hits found	DT210947	OC0108	1	No hits found		No hits found		1	0.00000
DT210989		OC0134	1	No hits found		No hits found		1	0.00000	-0.99
DT210997		OC0138	1	No hits found		No hits found		1	0.00000	-1.47
DT211040		OC0163	1	No hits found		No hits found		1	0.00160	-0.71
DT211054		OC0502	2	No hits found		No hits found		1	0.00000	-1.15
DT212527		OC1013	1	No hits found		No hits found		1	0.00045	-0.66
DT213157		OC1392	1	No hits found		No hits found		1	0.00914	-0.60
DT213397		OC1554_a	1	No hits found		No hits found		1	0.00142	-0.80
DT213404		OC1768	4	No hits found		No hits found		3	0.00000	-0.58
DT210860		OC3004	1	No hits found		No hits found		1	0.00000	-1.03

Résultats et discussion

Suite à l'analyse par microarray, l'expression de 33 gènes, représentés sur les lames de microarray, a été mesurée par PCR en temps réel afin de vérifier les résultats obtenus. Ces gènes ont été choisis parmi les gènes, identifiés comme différentiellement exprimés chez l'un ou l'autre des génotypes par microarray, mais aussi parmi les gènes non différentiellement exprimés, ceci afin d'estimer la puissance d'analyse des microarray.

Acc. nb	Contig/OC	Fonction [<i>Genre espèce</i>]	Analyses par microarray			PCR en temps réel	
			N	q-value	ER (log ₂)	ER (log ₂) 1	ER (log ₂) 2
DT212234	OC0687	arabinogalactan protein [<i>D. carota</i>]	4	0.00000	1.85	6.73	6.82
DT212300	OC1814	No hits found	2	0.00071	0.36	2.00	1.90
DT212813	OC0311_a	acidic ribosomal protein P0 [<i>G. max</i>]	4	0.11539	0.25	1.62	1.68
DT213187	OC1549_a	40S ribosomal protein S17 [<i>L. esculentum</i>]	2	0.00000	0.71	1.72	1.55
DT211067	Cont6402	putative leaf development protein Argonaute [<i>A. thaliana</i>]	1	0.18545	0.30	1.37	1.37
DT212502	OC1929	G protein beta subunit [<i>N. plumbaginifolia</i>]	4	0.00000	0.68	1.25	1.24
DT212655	OC1088	ribosomal protein L12 [<i>C. intybus</i>]	2	0.01347	0.40	1.25	0.95
DT212554	Cont0001	glyceraldehyde-3-phosphate dehydrogenase [<i>S. alba</i>]	6	0.00538	0.19	1.20	0.97
DT210897	OC0178	putative ribosomal S29 protein [<i>A. thaliana</i>]	3	0.29288	0.10	1.15	0.70
DT212250	OC2155_a	P40-like ribosomal protein [<i>D. carota</i>]	2	0.00926	0.45	0.58	0.82
DT212789	OC1427_b	Cell division control protein 48 homolog E [<i>A. thaliana</i>]	2	0.51528	0.01	0.60	0.60
DT212237	OC0834	chromatin remodeling factor CHD3 [<i>A. thaliana</i>]	1	0.33644	-0.24	0.50	0.50
DT212803	OC1211	BTB/POZ domain-containing protein [<i>A. thaliana</i>]	2	0.26776	0.22	0.38	0.42
DT210839	OC0758	putative RNA-binding protein [<i>A. thaliana</i>]	2	0.22522	0.15	0.08	0.18
DT212771	Cont4942	elongation factor EF-2 [<i>P. sativum</i>]	7	0.27455	0.06	-0.18	0.38
DT211255	Cont0024	seed specific protein Bn15D17A [<i>B. napus</i>]	3	0.50195	-0.02	0.00	0.13
DT210862	OC0172	embryogenic callus protein 181 [<i>D. carota</i>]	2	0.19735	0.24	-0.28	0.05
DT212897	OC1229	sucrose synthase isoform I [<i>D. carota</i>]	1	0.05978	0.27	0.08	-0.32
DT212829	OC1960	tuber-specific /sucrose-responsive element binding factor [<i>S. tuberosum</i>]	3	0.22701	0.17	-0.15	-0.65
DT212596	OC1155	xyloglucan endotransglycosylase [<i>M. domestica</i>]	2	0.28741	-0.09	-0.57	-0.52
DT213276	OC1785_a	H(+)-transporting ATPase [<i>P. vulgaris</i>]	1	0.01333	-0.32	-0.87	-0.93
DT213079	OC1347	putative ethylene response element binding protein [<i>N. tabacum</i>]	2	0.53209	0.01	-1.45	-1.22
DT212556	OC1068	leaf senescence protein-related (YLS7) [<i>A. thaliana</i>]	2	0.06047	-0.31	-1.55	-1.38
DT210805	OC0344	unknown protein [<i>A. thaliana</i>]	1	0.50346	-0.03	-1.35	-1.68
DT212452	OC0603_b	cobalamine-independent methionine synthase [<i>S. scutellarioides</i>]	4	0.00000	-0.81	-1.90	-1.57
DT210888	OC0071	14-3-3 protein [<i>S. tuberosum</i>]	1	0.00119	-0.41	-1.78	-2.05
DT211362	Cont0201	putative auxin-induced protein [<i>A. thaliana</i>]	1	0.00000	-1.12	-1.88	-2.03
DT213305	Cont0060	3-hydroxy-3-methylglutaryl coenzyme A reductase [<i>S. tuberosum</i>]	9	0.00000	-1.34	-2.30	-2.03
DT212233	OC1796	ring domain containing protein [<i>C. annuum</i>]	10	0.00000	-1.13	-2.18	-2.35
DT213289	Cont9039	metallothionein 1 [<i>A. tripolium</i>]	13	0.00000	-1.04	-2.45	-2.45
DT212623	OC0023_a	glutathione transferase [<i>H. muticus</i>]	3	0.00000	-1.17	-2.78	-2.85
DT211055	OC0168	1-aminocyclopropane-1-carboxylate oxidase [<i>L. esculentum</i>]	1	0.00000	-1.45	-3.87	-3.80
DT210799	Cont0006	1-aminocyclopropane-1-carboxylate oxidase [<i>P. hortorum</i>]	1	0.00000	-1.32	-3.82	-4.28

Tableau 10 : Validation par PCR en temps réel de résultats obtenus par microarray
Contig/OC: ID du contig ou de l'OC (Original Cluster regroupant les ESTs identiques). N : nombre d'ESTs, représentant le contig ou l'OC, analysés par microarray. ER, ER1, ER2 : ratios d'expression (log₂) calculés par microarray ou par PCR en temps réel entre le génotype embryogène et le génotype non-embryogène au jour 4 de culture en conditions d'embryogenèse somatique. Les fonds gris indiquent les gènes différentiellement exprimés.

Par microarray, 3 des 33 gènes analysés ont été définis comme préférentiellement exprimés chez le génotype embryogène, 8 comme préférentiellement exprimés chez le génotype non embryogène et 22 comme non différentiellement exprimés. L'expression

différentielle des 11 gènes a été confirmée par les analyses de PCR en temps réel. Parmi les 22 gènes restants, la PCR en temps réel a montré que 13 sont différentiellement exprimés.

Les PCR en temps réel ont montré que les 5 gènes codant des protéines ribosomales sont préférentiellement exprimés chez le génotype embryogène. D'autres gènes préférentiellement exprimés chez le génotype K59 semblent présenter un intérêt particulier dans la réponse du génotype embryogène aux conditions d'induction de l'embryogenèse somatique : ces gènes codent, par exemple, une protéine de développement de type Argonaute (Cont6402), une protéine CDC48 potentiellement impliquée dans les divisions cellulaires (OC1427_b), une protéine G (OC1929) et une AGP (arabinogalactan protein) (OC0687). Les gènes préférentiellement exprimés chez le génotype non embryogènes semblent en majorité en lien avec la défense. Ces gènes codent 2 ACC oxydases, enzyme de la voie de synthèse de l'éthylène (Cont0006 et OC0168), un élément de réponse à l'éthylène (OC1347), une glutathione transférase (OC0023), une métallothionéine (Cont9039) et une protéine reliée à la sénescence de la feuille (OC1068). D'autres gènes présentent également un intérêt, comme celui codant une protéine 14-3-3 (OC0071), potentiellement impliquée dans la signalisation.

E. Discussion

1) Choix du schéma expérimental

Le but était d'une part d'observer les profils d'expression des gènes, pour un génotype donné, au cours de la cinétique, et d'autre part, de comparer ces profils d'expression entre les deux génotypes. Une autre possibilité aurait été de comparer directement tous les points de la cinétique à la condition jour zéro pour les deux génotypes et ensuite comparer les conditions C15J0 et K59J0 entre elles. La mise en place d'un tel schéma aurait présenté l'avantage de réduire le nombre d'expériences à 13 au lieu de 14 et aurait permis d'augmenter la précision du ratio d'expression. C'est, par exemple, ce type de schéma qui a été mis en place pour étudier l'embryogenèse somatique chez le soja (Thibaud-Nissen *et al.*, 2003).

Cependant, le choix d'un schéma expérimental indirect utilisant une référence commune, nous permettra ultérieurement de comparer d'autres conditions biologiques à toutes celles déjà testées. Il sera, par exemple, possible de compléter la cinétique sur une fenêtre de temps présentant un intérêt particulier. Il sera également possible de comparer l'embryogenèse somatique obtenue à partir de racines à celle obtenue à partir de feuilles ou de comparer l'embryogenèse zygotique à l'embryogenèse somatique. En effet nous disposons

d'une grande quantité du mélange d'ARN référence, ce qui laisse la possibilité de réaliser de nombreuses autres expérimentations.

2) Validation du processus d'analyses par microarray

A partir de la Figure 35, construite en utilisant les résultats issus du Tableau 10, on observe, une forte corrélation entre les résultats obtenus par microarray et ceux par PCR en temps réel ($r=0,92$). De plus, une expression différentielle a pu être confirmée par PCR en temps réel pour l'ensemble des gènes identifiés comme différentiellement exprimés par microarray. Ce résultat montre la robustesse du dispositif d'analyse des résultats de microarray mis en place.

Comme cela a été montré précédemment (Czechowski *et al.*, 2004), on remarque également que la PCR en temps réel est une technique beaucoup plus sensible que les microarrays. Cela se traduit par l'obtention de ratios d'expression beaucoup plus élevés. Par exemple, un ratio de 1 mesuré par microarray correspond, selon la droite de régression, à un ratio de plus de 3 mesuré par PCR en temps réel. Cette différence de sensibilité se traduit également par davantage de gènes identifiés comme différentiellement exprimés puisque parmi les 33 gènes analysés par les deux techniques, 24 ont été définis comme différentiellement exprimés par PCR en temps réel alors qu'ils n'étaient que 11 par microarray.

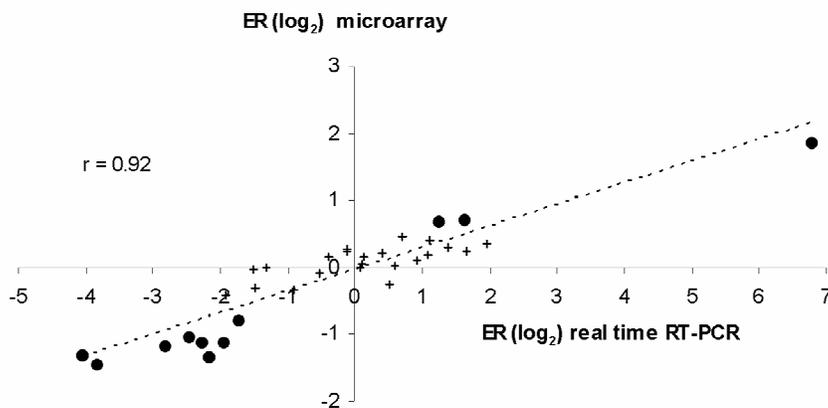


Figure 35 : Corrélation entre les analyses transcriptomiques menées par PCR en temps réel et celles menées par microarray.

Les ARN utilisés ont été isolés d'explants de chicorées de génotypes embryogène et non embryogène prélevés au jour 4 de culture d'embryogenèse. D'après les résultats de microarray, 11 gènes ont été définis comme différentiellement exprimés (●), et 22 comme non différentiellement exprimés (+). ER(log₂) : ratios d'expression mesurés par microarray ou par PCR en temps réel entre le génotype embryogène et le génotype non embryogène ; r : coefficient de corrélation.

Les gènes préférentiellement exprimés chez K59 plus difficiles à détecter que ceux exprimés chez C15 par microarray ?

Le faible nombre de gènes identifiés comme préférentiellement exprimés par microarray chez K59 pourrait être dû à un plus faible niveau d'expression de ces gènes en comparaison aux gènes préférentiellement exprimés chez C15. La moyenne des ratios d'expression des gènes préférentiellement exprimés chez K59 est de 0,71 alors qu'elle est de 0,91 pour les gènes préférentiellement exprimés chez C15 au jour 4 de culture en conditions d'embryogenèse somatique. De plus, les analyses réalisées par PCR en temps réel pour 33 gènes également analysés par microarray, augmentent sensiblement le nombre de gènes préférentiellement exprimés chez K59. Ces deux résultats suggèrent que pour les gènes préférentiellement exprimés chez K59, les différences d'abondance de transcrits entre les deux génotypes pour les gènes pourraient être faibles, et plus difficiles à détecter par des méthodes basées sur des hybridations (banques soustractives, microarray), que par PCR en temps réel (Czechowski *et al.*, 2004). Une des explications pourrait être qu'une faible proportion de cellules au sein de l'explant entre dans le processus d'embryogenèse somatique. En effet, bien que plus nombreuses que les environ 1 cellule sur 10000 qui entrent dans le processus d'embryogenèse somatique dans les suspensions cellulaires de carottes (Toonen *et al.*, 1994), ces cellules ne représentent au maximum que quelques pourcents (~3%-10%) des cellules de l'explant chez le génotype K59 (Da Silva, 2004). Si l'on considère que la majorité des gènes préférentiellement exprimés chez K59, le sont dans les cellules qui entrent dans le processus d'embryogenèse somatique ou dans les cellules environnantes participant à la mise en place du processus, et ne le sont pas dans la majorité des cellules de l'explant, ceci pourrait expliquer la faible représentation des gènes préférentiellement exprimés chez le génotype K59 dans les banques soustractives et la difficulté à mettre en évidence une expression différentielle par microarray pour ces gènes. Au contraire, du génotype K59, les cellules du génotype C15 semblent présenter une réponse homogène, du moins d'un point de vue cytologique. Une grande majorité des cellules du génotype non-embryogène exprimerait donc les mêmes gènes.

Bien que la technique de microarray soit une technique moyennement sensible qui ne permet pas toujours la mise en évidence de gènes montrant une variation d'expression très fine, un échantillon des gènes définis comme différentiellement exprimés par microarray a été confirmé par PCR en temps réel. Ce résultat indique que l'ensemble du processus d'analyse

de microarray que nous avons défini ne génère pas de faux positifs. L'étude préliminaire menée au jour 4 de culture témoigne de la validité et de la robustesse du processus d'analyse utilisé pour l'étude en cinétique présentée dans la partie III.

3) Origine des gènes différentiellement exprimés au jour 4 de culture

Afin de vérifier les hypothèses concernant l'efficacité de la soustraction émises au cours de la première partie des résultats, nous avons décidé, à l'aide des analyses de microarray réalisées au jour 4 de culture pour les deux génotypes, d'observer si les gènes préférentiellement exprimés chez K59 sont effectivement représentés dans la banque E et si les gènes préférentiellement exprimés chez C15 sont effectivement représentés dans la banque NE. Pour cela, nous avons réalisé un test de χ^2 (Tableau 11). L'hypothèse H_0 émise pour ce test est que les gènes exprimés différentiellement chez les deux génotypes proviennent indifféremment des banques E et NE. Le test du χ^2 rejette l'hypothèse H_0 avec un seuil inférieur à 0,0005 et montre ainsi que la distribution des ESTs, représentant des gènes exprimés différentiellement, est significativement différente pour les banques E et NE. On constate que 160 ESTs provenant de la banque E et 146 ESTs provenant de la banque NE représentent des gènes préférentiellement exprimés chez C15. Ces effectifs observés correspondent respectivement à 73,5% et à 160% des effectifs théoriques attendus. Ces différences significatives montrent qu'il y a eu une soustraction des gènes préférentiellement exprimés chez C15 dans la banque E, alors que ces gènes ont été enrichis dans la banque NE. Par contre, les effectifs observés de 85 ESTs de la banque E et de 35 ESTs de la banque NE correspondant à des gènes préférentiellement exprimés chez K59 ne diffèrent pas des effectifs théoriques calculés. Ce dernier résultat renforce l'hypothèse émise précédemment, qui suppose que la technique de SSH a une faible emprise sur les gènes préférentiellement exprimés chez le génotype K59, en raison d'une faible représentation de ces transcrits.

Expression	Nombre d'ESTs représentant des gènes analysés par microarray						Total (%)
	Provenant de la banque E			Provenant de la banque NE			
	Observés (%)	Attendus	χ^2	Observés (%)	Attendus	χ^2	
Préf. chez K59	85 (5.6)	85.2	0.001	36 (5.7)	35.8	0.001	121 (5.6)
Préf. chez C15	160 (10.6)	217.6	15.26	146 (23.5)	91.4	36.34	309 (14.4)
Non différentiels	1265 (83.8)	1207.2	2.77	449 (70.8)	506.8	6.60	1714 (79.9)
total	1510 (100)		18.03	634 (100)		42.95	2144 (100)

Tableau 11 : Distribution des ESTs des banques E et NE suivant l'expression des gènes correspondants mesurée par microarray.

L'ensemble des ESTs des banques E et NE qui représentent des gènes dont l'expression a été mesurée par microarray ont été classés dans 3 catégories : gènes préférentiellement exprimés chez K59, chez C15, ou gènes non différentiellement exprimés. La distribution des ESTs a été comparée avec une distribution théorique basée sur la somme des effectifs des deux banques pour les trois catégories.

L'hypothèse H_0 émise pour le test de χ^2 est que les deux banques présentent la même distribution d'ESTs suivant qu'ils représentent des gènes préférentiellement exprimés chez K59, préférentiellement exprimés chez C15 ou non différentiellement exprimés. La somme des χ^2 est de 60,98 et permet de rejeter l'hypothèse H_0 avec un seuil inférieur à 0,0005.

4) Gènes exprimés différentiellement entre les deux génotypes au jour 4 de culture

Les distributions par classe de fonction, des gènes différentiellement exprimés chez les génotypes K59 et C15, indiquent clairement qu'au jour 4 de culture en conditions d'embryogenèse somatique, les cellules des explants des deux génotypes sont entrées dans des programmes de développement différents (Tableau 9).

Sur les 45 gènes exprimés préférentiellement chez K59, 18 (40%) codent des protéines ribosomales. Ce résultat reflète probablement la mise en place d'un nouveau transcriptome et la synthèse de ribosomes dans les cellules se préparant à entrer dans une voie de transdifférenciation et à réinitier les divisions cellulaires. La relation entre une augmentation de l'expression de gènes codant des protéines ribosomales et la division cellulaire, a par ailleurs, été montrée dans diverses études (Barakat *et al.*, 2001 ; Dresselhaus *et al.*, 1999). Chez le zinnia, plusieurs gènes codant des protéines ribosomales ont également montré une expression différentielle au cours de la transdifférenciation de cellules du mésophylle en cellules du xylème (Demura *et al.*, 2002) et chez le peuplier de nombreux ESTs représentant des protéines ribosomales ont été obtenus à partir de banques d'ADNc construites à partir de tissus méristématiques (Matsubara *et al.*, 2005 ; Ranjan *et al.*, 2004). Du fait que les banques

et les analyses transcriptomiques ont été réalisées au jour 4 de culture, c'est à dire à l'instant où sont initiées les premières divisions cellulaires chez le génotype K59, il n'est pas surprenant d'identifier des gènes liés à la division cellulaire. Ainsi des gènes codant une protéine CDC48-like (Feiler *et al.*, 1995), une protéine associée au déplacement du noyau, une protéine G (McKhann *et al.*, 1997), une histone H2B (van den Heuvel *et al.*, 1999), et une ribonucléotide réductase (Chaboute *et al.*, 1998) ont été identifiés comme préférentiellement exprimés chez le génotype K59.

Les aspects morphologiques des cellules, observés dans les explants foliaires du génotype C15, montrent que les cellules de ce génotype ne réagissent que très peu aux conditions de culture d'induction de l'embryogenèse somatique et présentent seulement un noyau qui augmente de volume. Nous avons alors supposé que les cellules du génotype C15 pouvaient être bloquées dans la phase de réactivation qui conduit à la formation de cellules embryogènes. Bien que limitées, à ce stade, à un seul jour d'étude (J4), les résultats de l'analyse transcriptomique pourraient être contradictoires avec cette hypothèse. En effet, le nombre important de gènes préférentiellement exprimés chez le génotype C15 (68) semble indiquer que l'observation de noyaux volumineux dans les cellules des explants du génotype non embryogènes pourrait être simplement due à une forte activité transcriptionnelle.

Les cellules du génotype C15 semblent répondre aux stress appliqués en installant un état de défense et de mise en veille cellulaire. En effet, plusieurs gènes préférentiellement exprimés chez C15 codent des protéines impliquées dans le transport et la mise en réserve, alors que d'autres sont impliquées dans la défense. Contrairement au génotype K59, nous n'avons pas identifié de gènes préférentiellement exprimés chez le génotype C15 codant des protéines ribosomales alors qu'il semble y avoir au sein des cellules, une activité de synthèse protéique importante. Nous supposons que chez le génotype non-embryogène, les protéines ribosomales présentes initialement dans les feuilles pourraient être les mêmes que celles qui permettent la réponse du génotype aux conditions de culture.

Cette étude réalisée au jour 4 de culture en conditions d'embryogenèse somatique n'est qu'une fenêtre au cours de la cinétique d'embryogenèse somatique. Elle ne permet pas, par exemple, de distinguer les gènes présentant une expression différentielle induite ou constitutive. L'étude en cinétique, présentée au cours de la partie III, conduira à étoffer cette discussion sur les gènes liés aux réponses des deux génotypes aux conditions de culture appliquées.

Partie III : Etude en cinétique, par microarray, des phases précoces de l'embryogenèse somatique chez la chicorée

A. Différences transcriptomiques entre les deux génotypes au stade initial de la cinétique d'étude

Avant d'étudier, pour les deux génotypes, l'expression des gènes au cours de la cinétique en conditions d'embryogenèse somatique, nous avons décidé de nous intéresser aux différences transcriptomiques entre les deux génotypes au stade initial de la cinétique (J0). A ce stade, les explants ont subi l'aseptisation et une incubation d'une demi-heure dans le milieu d'induction à température ambiante et à la lumière avant d'être congelés et conservés à -80°C. Les différences transcriptionnelles observées cumuleront donc les différences d'expression intrinsèques aux deux génotypes ainsi que les différences induites de façon très rapide par les stress occasionnés par le prélèvement de l'explant, l'aseptisation, et l'incubation dans le milieu de culture.

Le schéma expérimental mis en œuvre au cours de la partie II nous a permis d'effectuer cette comparaison au jour 0. Les modèles d'analyses de variance utilisés, comme pour le cas de la comparaison à J4, ont été les mêmes que ceux décrits dans la partie II, en remplaçant l'effet jour par un effet génotype.

Cette analyse a permis de mettre en évidence 81 gènes exprimés différemment entre les deux génotypes au jour 0 de culture, avec 20 préférentiellement exprimés chez K59 et 61 chez C15 (Tableau 12). Chez le génotype K59, après les gènes ne présentant pas d'homologie avec les séquences des bases de données, les classes fonctionnelles les plus représentées sont les protéines aux fonctions inconnues (3 gènes) et le métabolisme (3 gènes). Nous notons que deux gènes (OC0018_a et OC0018_b) codant des protéines potentiellement induites par la blessure sont préférentiellement exprimés chez le génotype K59. Ces deux gènes sont issus d'un même OC « primaire » (OC0018) (voir partie I), ce qui indique que les EST représentant les deux gènes présentent entre eux des homologies assez fortes. De ce fait, une hybridation croisée des cibles correspondantes aux deux gènes sur les lames de microarray n'est pas à exclure. Ce résultat indique néanmoins qu'au moins un des deux gènes est préférentiellement exprimé chez le génotype embryogène. Parmi les autres gènes préférentiellement exprimés chez K59, nous remarquons la présence d'un gène codant une protéine induite par l'auxine (Cont0125) et d'un gène codant un transporteur ABC potentiellement associé à la défense (OC1588). Ces trois gènes pourraient être induits très rapidement dès le prélèvement de la feuille sur le plant de chicorée.

Résultats et discussion

Chez le génotype C15 non embryogène, comme chez le génotype K59, après les gènes ne présentant pas d'homologie avec les séquences des bases de données, les classes fonctionnelles les plus représentées sont le métabolisme (13 gènes), les protéines aux fonctions inconnues (13 gènes), le métabolisme des protéines (synthèse et devenir) (10 gènes) et le transport cellulaire (4 gènes). Le nombre significatif de gènes codant des protéines liées au métabolisme des protéines pourrait suggérer qu'au stade initial de la cinétique, les cellules du génotype C15 amorcent un renouvellement des protéines. Ce renouvellement pourrait être induit très précocement par les conditions d'induction. Les gènes potentiellement liés à la défense font partie des gènes qui pourraient être induits de façon très rapide : un gène code une glutathione S-transferase (OC0003_a) et un gène code une protéine de choc thermique (OC1363).

Tableau 12a

Classe de fonction	Contig/OC	Fonction	E-value	q-value	ER
Métabolisme	Cont9073	alpha-glucosidase precursor (Maltase) [<i>Spinacia oleracea</i>]	1,00E-33	0,001053	-0,57
	OC1732	myo-inositol-1-phosphate synthase [<i>Aster tripolium</i>]	2,00E-23	0,003969	-0,75
	Cont0060	3-hydroxy-3-methylglutaryl coenzyme A reductase [<i>Solanum tuberosum</i>]	1,00E-133	0,000000	-0,72
Cycle cellulaire et synthèse d'ADN	OC2119	RAD23 like protein [<i>Arabidopsis thaliana</i>]	4,00E-09	0,000000	-0,59
	OC1729	putative T-complex protein 1, ETA subunit [<i>Arabidopsis thaliana</i>]	8,00E-72	0,000000	-0,71
Devenir des protéines	Cont0128	26S proteasome regulatory particle triple-A ATPase subunit6 [<i>Oryza sativa</i>]	1,00E-102	0,000978	-0,54
	Cont3602	putative ubiquitin-conjugating enzyme [<i>Oryza sativa</i>]	1,00E-50	0,000171	-0,55
Cytosquelette	OC0129_a	beta tubulin 1 [<i>Lupinus albus</i>]	1,00E-137	0,000449	-0,48
Transport cellulaire	OC1588	PDR-like (multi drug resistance) ABC (ATP binding cassette) transporter [<i>Oryza sativa</i>]	2,00E-09	0,008451	-0,75
	OC1626_a	dicarboxylate/tricarboxylate carrier [<i>Citrus junos</i>]	4,00E-37	0,008901	-0,55
Défense	OC0018_b	putative wound-induced protein [<i>Medicago sativa</i> subsp. <i>x varia</i>]	4,00E-18	0,000000	-0,75
	OC0018_a	putative wound-induced protein [<i>Medicago sativa</i> subsp. <i>x varia</i>]	5,00E-22	0,000000	-0,80
Autres	Cont0125	IN2-2 protein, auxin induced [<i>Zea mays</i>]	5,00E-46	0,000302	-0,58
	OC2100	Pot1-like protein [<i>Arabidopsis thaliana</i>]	6,00E-12	0,005094	-0,50
Protéines inconnues	OC1772	unknown [<i>Arabidopsis thaliana</i>]	2,00E-10	0,000000	-0,81
	OC1471	unknown protein [<i>Arabidopsis thaliana</i>]	2,00E-06	0,003778	-0,60
	OC1367	hypothetical protein [<i>Cicer arietinum</i>]	3,00E-20	0,000593	-0,62
	OC1757			0,004538	-0,56
Aucune homologie	OC1342			0,000000	-0,59
	OC1013			0,002822	-0,60
	OC1839_b			0,000000	-1,57

Résultats et discussion

Tableau 12b

Classe de fonction	Contig/OC	Fonction	E-value	q-value	ER
Métabolisme	OC0603_a	methionine synthase [<i>Solanum tuberosum</i>]	1,00E-107	0,000000	1,30
	OC0603_b	cobalamine-independent methionine synthase [<i>Solenostemon scutellarioides</i>]	1,00E-103	0,000000	1,02
	Cont0108	phenylalanine ammonia-lyase2; PAL [<i>Lactuca sativa</i>]	4,00E-44	0,000000	0,91
	OC1216	cyanate hydratase [<i>Oryza sativa</i>]	2,00E-41	0,000000	1,65
	OC1052	putative gluconokinase [<i>Oryza sativa</i>]	5,00E-11	0,000000	1,00
	Cont0777	ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit [<i>Lactuca sativa</i>]	5,00E-36	0,000000	0,82
	Cont0119	pectin Methylesterase, chain A [<i>Daucus carota</i>]	2,00E-38	0,000259	0,35
	OC1311	putative beta-ketoacyl-CoA synthase [<i>Oryza sativa</i>]	3,00E-90	0,000259	0,61
	Cont0006	1-aminocyclopropane-1-carboxylate oxidase [<i>Pelargonium x hortorum</i>]	1,00E-47	0,002744	0,72
	Cont0072	phenylcoumaran benzylic ether reductase homolog Fi1 [<i>Forsythia x intermedia</i>]	1,00E-62	0,003321	0,71
	OC1139	thioredoxin H3 [<i>Ipomoea batatas</i>]	1,00E-42	0,000978	0,95
	OC1124	putative 3-beta hydroxysteroid dehydrogenase/isomerase protein [<i>Arabidopsis thaliana</i>]	1,00E-15	0,000848	0,94
	OC1370	deoxyribodipyrimidine photolyase family protein-like [<i>Oryza sativa</i>]	6,00E-33	0,005070	0,59
Cycle cellulaire et synthèse de l'ADN	OC1036	60S ribosomal protein L10 (EQM) [<i>Solanum melongena</i>]	9,00E-83	0,009954	0,74
Transcription	OC1153	putative small nuclear ribonucleoprotein D1 [<i>Oryza sativa</i>]	2,00E-09	0,000000	1,02
Synthèse protéique	OC1371	putative 60S ribosomal protein L7 [<i>Oryza sativa</i>]	5,00E-43	0,000000	0,97
	OC1422_a	60S ribosomal protein L13a [<i>Arabidopsis thaliana</i>]	5,00E-19	0,000000	0,79
	OC1544	ubiquitin extension protein [<i>Solanum tuberosum</i>]	2,00E-42	0,003921	0,57
	OC0533_b	ubiquitin/ribosomal protein [<i>Brassica rapa</i>]	3,00E-33	0,002284	0,48
	OC0846	alpha NAC-like protein [<i>Arabidopsis thaliana</i>]	5,00E-23	0,000302	0,79
	OC2094	putative protein disulfide-isomerase [<i>Prunus avium</i>]	1,00E-35	0,002744	0,73
Devenir des protéines	OC1110_a	putative polyubiquitin [<i>Arabidopsis thaliana</i>]	8,00E-33	0,002871	0,55
	OC1717_a	Probable 26S proteasome non-ATPase regulatory subunit 3 [<i>Nicotiana tabacum</i>]	1,00E-103	0,000000	0,76
	OC1110_c	polyubiquitin (UBQ14) [<i>Arabidopsis thaliana</i>]	3,00E-49	0,000393	0,67
Transport cellulaire	OC1301	multidrug resistance-associated protein MRP1 [<i>Triticum aestivum</i>]	1,00E-17	0,000171	0,86
	OC0827	transmembrane transport protein-like [<i>Arabidopsis thaliana</i>]	7,00E-6	0,003105	0,68
	OC1785_a	H ⁽⁺⁾ -transporting ATPase [<i>Phaseolus vulgaris</i>]	3,00E-48	0,003237	0,47
	OC1356	Beta-COP-like protein [<i>Arabidopsis thaliana</i>]	8,00E-62	0,000000	0,97
Signalisation cellulaire	OC0116	putative receptor-like kinase Xa21-binding protein 3 [<i>Oryza sativa</i>]	3,00E-27	0,001865	0,60
Défense	OC0023_a	glutathione transferase [<i>Hyoscyamus muticus</i>]	7,00E-68	0,000000	0,55
	OC0003_a	glutathione S-transferase GST 13 [<i>Glycine max</i>]	2,00E-38	0,000168	0,56
	OC1363	heat-shock protein 80 [<i>Euphorbia esula</i>]	3,00E-32	0,000000	0,80
Autres	OC1480	putative RING zinc finger protein [<i>Arabidopsis thaliana</i>]	1,00E-09	0,000000	0,61
	OC0120	cytochrome P450 [<i>Nepeta racemosa</i>]	3,00E-36	0,000000	0,87
	Cont7599	high mobility group protein HMG-1 [<i>Ipomoea nil</i>]	1,00E-07	0,007839	0,55

Résultats et discussion

Tableau 12b (suite)

Classe de fonction	Contig/OC	Fonction	E-value	q-value	ER	
Protéines inconnues	OC1200	unknown protein [<i>Arabidopsis thaliana</i>]	1,00E-37	0,000171	0,87	
	OC0151	hypothetical protein [<i>Oryza sativa</i>]	6,90E-02	0,000000	1,19	
	OC1395	unknown protein [<i>Arabidopsis thaliana</i>]	2,00E-30	0,000171	1,06	
	OC1060	unknown [<i>Oryza sativa</i>]	1,00E-45	0,000000	1,03	
	OC1458	unknown protein [<i>Arabidopsis thaliana</i>]	1,00E-64	0,000000	0,97	
	OC1161	unknown protein [<i>Arabidopsis thaliana</i>]	1,40E-06	0,000000	0,95	
	OC1118	unknown [<i>Arabidopsis thaliana</i>]	3,00E-14	0,000000	0,85	
	OC1619	unknown protein [<i>Arabidopsis thaliana</i>]	8,00E-42	0,000393	0,77	
	OC0856	unknown protein [<i>Oryza sativa</i>]	1,00E-20	0,002739	0,73	
	OC1108	unnamed protein product [<i>Arabidopsis thaliana</i>]	7,00E-12	0,004766	0,69	
	OC1095	unknown protein [<i>Oryza sativa</i>]	9,00E-16	0,003886	0,62	
	OC1275	unknown [<i>Arabidopsis thaliana</i>]	3,00E-64	0,005252	0,58	
	OC1330	unknown [<i>Arabidopsis thaliana</i>]	2,00E-39	0,000171	0,51	
	Aucune homologie	OC1860			0,000000	1,01
		OC1171			0,000000	0,86
OC1300				0,000848	0,82	
OC0836				0,000000	0,82	
OC1354				0,000593	0,80	
OC1391				0,000978	0,74	
OC1639				0,000000	0,73	
OC0078				0,000302	0,71	
OC1472				0,000000	0,70	
OC0502				0,001124	0,69	
OC0829				0,001362	0,60	
OC1028				0,000978	0,58	

Tableau 12 : Liste des gènes préférentiellement exprimés chez les génotypes K59 (a) et C15 (b) au stade initial de la culture

La fonction des gènes, la probabilité associée à la qualité de l'alignement (E-value) et la classification fonctionnelle sont issues des comparaisons des ESTs avec les séquences de la base de données non redondantes de GenBank et avec les séquences codantes traduites d'*Arabidopsis* (voir partie I). Les deux colonnes de droite fournissent le \log_2 du ratio d'expression (ER) C15J0/ref rapporté à K59J0/ref et la valeur q associée à la significativité du ratio.

B. Gènes exprimés différemment au cours de la cinétique en conditions d'embryogenèse somatique

1) Gènes montrant une modulation d'expression au cours de la cinétique d'étude

Les analyses de variance décrites au cours de la partie II ont permis la mise en évidence de 417 gènes présentant une modulation d'expression significative, sur au moins un des points de la cinétique d'étude, par rapport au jour 0, chez l'un et/ou l'autre des deux génotypes. Parmi ces 417 gènes, 119 présentent une modulation d'expression significative observée uniquement chez le génotype K59, 187 présentent une variation d'expression observée seulement chez le génotype C15 et 111 montrent une variation d'expression significative chez les deux génotypes à la fois (Tableau 13).

	K59	C15
J1 contre J0	44	47
J2 contre J0	87	62
J3 contre J0	49	157
J4 contre J0	117	165
J5 contre J0	45	132
J6 contre J0	50	136
Nombre de gènes spécifiques	119	187
Nombre de gènes communs	111	

Tableau 13 : Nombre de gènes exprimés différemment au cours de la cinétique d'étude chez les génotypes K59 et C15

Pour chaque génotype, les effectifs correspondent au nombre de gènes exprimés différemment par rapport à J0 (Figure 25). Le nombre de gènes spécifiques a été obtenu en dénombrant, une seule fois, les gènes présentant une expression différentielle sur au moins un point de la cinétique. Le nombre de gènes communs représente les gènes montrant une expression différentielle, par rapport à J0, chez les deux génotypes.

Nous mettons en évidence qu'aux jours 1 et 2 de la cinétique, en comparaison aux jours plus tardifs, peu de gènes présentant une modulation d'expression ont été détectés. Nous remarquons qu'aux jours 3, 5 et 6 de la cinétique, nous identifions 2,5 à 3 fois plus de gènes préférentiellement exprimés chez C15 que chez K59, ce qui pourrait suggérer une réponse plus rapide et/ou plus intense du génotype non embryogène aux conditions de culture. Comme suggéré précédemment (Partie II D.5), c'est effectivement au jour 4 de culture, jour de construction des banques soustractives, que l'on a identifié le maximum de gènes différemment exprimés, avec toujours plus de gènes différemment exprimés chez le génotype C15.

2) Regroupement des profils d'expression

a) Stratégie

Pour les 417 gènes montrant une modulation d'expression, nous avons représenté les profils d'expression chez le génotype C15 en utilisant la condition K59J0 comme référence commune lors du calcul des ratios d'expression. Les profils d'expression obtenus pour les génotypes K59 et C15 ont ensuite été soumis à des analyses de regroupement à l'aide de la méthode de K-means (Soukas *et al.*, 2000) en utilisant le programme inclu au logiciel Mev4 (MultiExperiment viewer, <http://www.tm4.org/mev.html>).

Au cours d'une première étape, nous avons choisi de distinguer les gènes présentant un profil d'expression similaire chez les deux génotypes de ceux présentant des profils différents. Nous voulions ainsi discriminer les gènes qui montrent la même réponse chez les deux génotypes de ceux qui montrent des réponses différentes. Nous avons pour cela, à l'aide de la méthode de K-means, classé dans différents groupes, suivant leur similitude, les profils d'expression obtenus chez les génotypes K59 et C15. Les gènes dont les profils d'expressions chez K59 et C15 se trouvaient dans un même groupe ont constitué une première sous-population (A) et les gènes dont les profils obtenus chez K59 et C15 se trouvaient dans des groupes différents ont constitué une seconde sous-population (B) (Figure 36).

Au cours d'une seconde étape, il s'est agit de classer dans différents groupes, suivant leur profil d'expression, premièrement les gènes de la sous-population A et deuxièmement ceux de la sous-population B.

Résultats et discussion

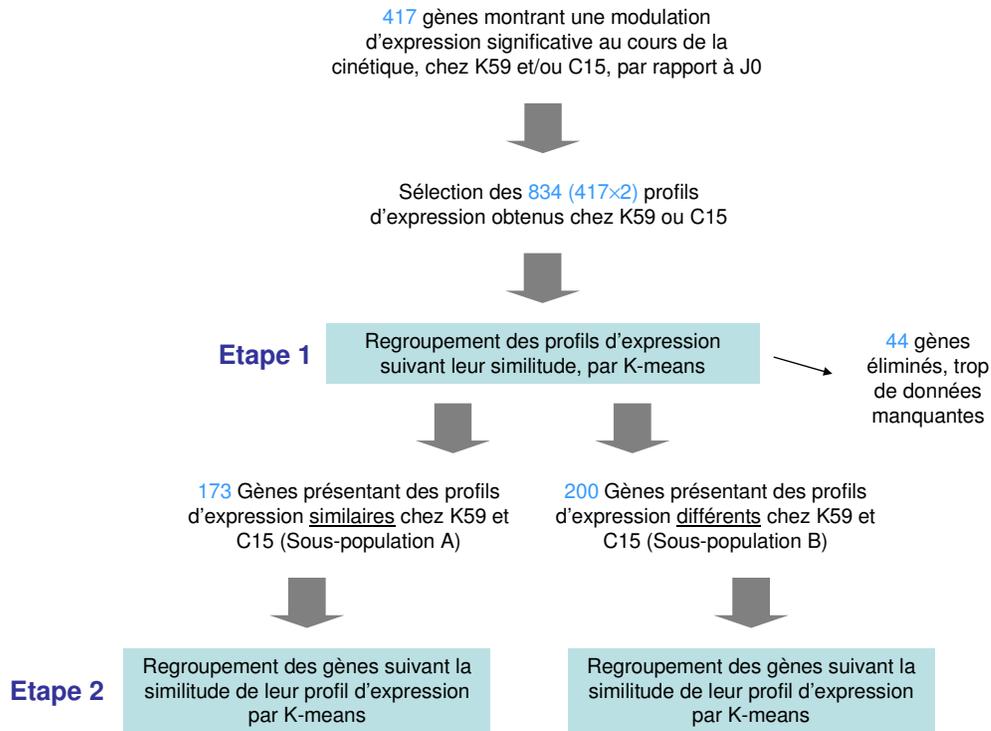


Figure 36 : Stratégie mise en place pour le regroupement des profils d’expression des gènes montrant une modulation d’expression en cours de cinétique

b) Résultats

(i) Etape 1

Le regroupement des profils d’expression par la méthode de K-means laisse le libre choix du nombre de groupes dans lesquels les profils seront classés. Ainsi, afin de déterminer un nombre de groupes appropriés, nous avons utilisé des calculs de valeur FOM (Figure Of Merit) (Yeung *et al.*, 2001). Cette mesure permet d’estimer, pour différents nombres de groupes, la puissance de l’algorithme de regroupement. Une faible valeur de FOM signifie que le nombre de groupes choisi fournira une grande puissance de regroupement des gènes. La génération d’une courbe de FOM, présentant en abscisse le nombre de groupes et en ordonnée les valeurs FOM correspondantes, fournit une aide dans le choix d’un nombre approprié de groupes.

Le jeu de données des 834 profils d’expression a permis de tracer une première courbe de valeurs FOM. Nous observons sur la courbe qu’au-delà de 10 groupes définis, la puissance de regroupement n’augmente que très faiblement (Figure 37). Ainsi, augmenter le nombre de groupes au-delà de 10, ne conduirait pas à une meilleure qualité de regroupement. Nous avons donc choisi de définir 10 groupes pour le regroupement de ces profils (Figure 38).

Résultats et discussion

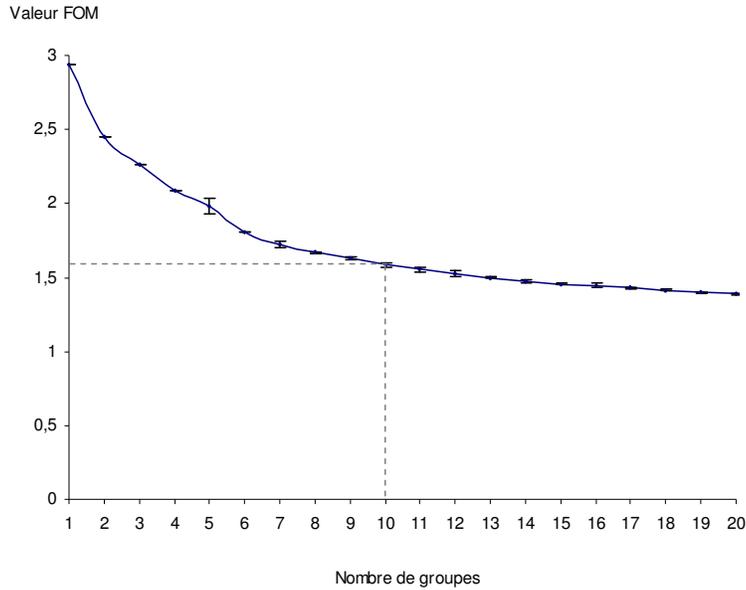


Figure 37 : Courbe des valeurs FOM pour le regroupement des 834 profils d'expression observés chez les génotypes K59 et C15 pour les 417 gènes sélectionnés. Cinq mesures de valeur FOM par nombre de groupes ont permis de calculer une moyenne et un écart-type.

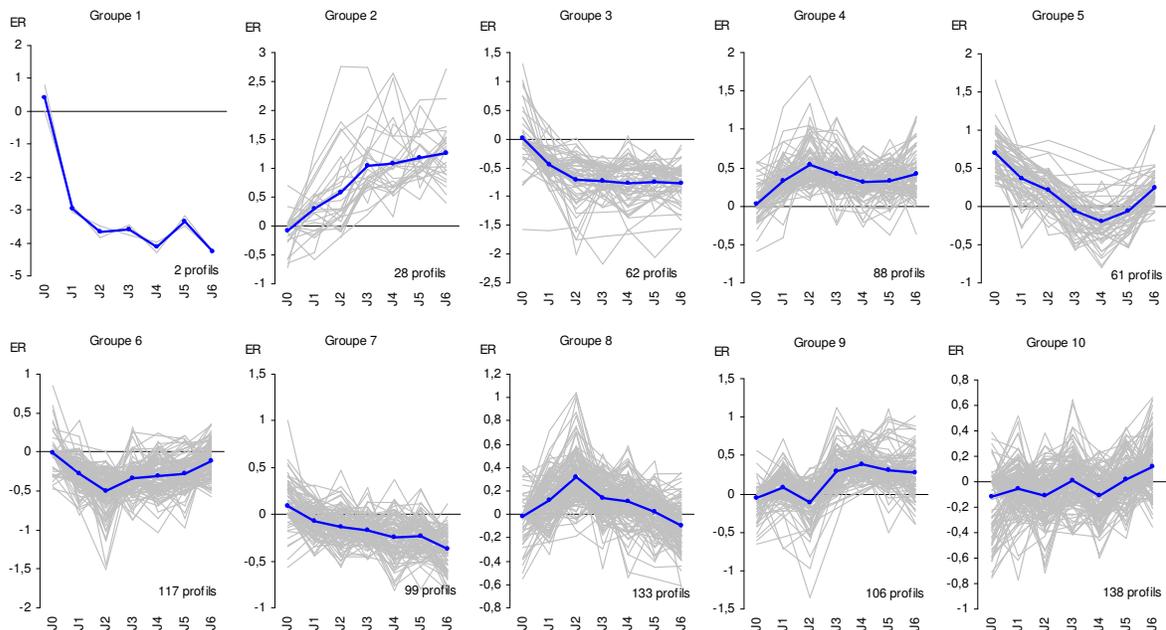


Figure 38 : Regroupement par la méthode de K-means des 834 profils d'expressions correspondant aux 417 gènes sélectionnés. Le trait bleu indique le profil moyen obtenu à partir de tous les profils du groupe (traits gris).

Pour 310 gènes, sur les 417 gènes présentant une modulation d'expression au cours de la cinétique chez l'un et/ou l'autre des génotypes, les profils d'expression obtenus chez C15 et K59 ont été classés dans des groupes différents. Pour 107 gènes, les profils ont été classés dans un même groupe. Nous avons ensuite vérifié manuellement gène par gène, si les profils obtenus pour K59 et C15 étaient effectivement similaires ou au contraire différents et nous avons éliminé les profils pour lesquels il y avait trop de données absentes. Parmi les 310 gènes, 43 gènes ont été éliminés puisque les cinétiques présentaient trop de données manquantes, et qu'il était de ce fait difficile d'interpréter les profils d'expression. Soixante sept autres gènes ont été reclassés avec ceux présentant un même profil d'expression chez les deux génotypes. Parmi les 107 gènes dont les profils d'expression ont été classés dans un même groupe, seulement 3 gènes ont été reclassés comme ayant des profils différents chez K59 et C15 et 1 gène a été éliminé de l'analyse, en raisons de données absentes trop nombreuses. Au final nous avons donc 200 gènes présentant des profils d'expression différents chez les deux génotypes (sous-population A) et 173 des profils similaires (sous-population B) (Figure 36).

(ii) Etape 2

Au cours de la seconde étape, pour les deux sous-populations A et B, nous avons réalisé des regroupements des gènes suivant leur profil global d'expression chez les deux génotypes.

Pour l'échantillon A, en se basant à nouveau sur une courbe de valeurs FOM, nous avons défini 7 groupes pour dissocier les profils d'expression des gènes (Figure 39).

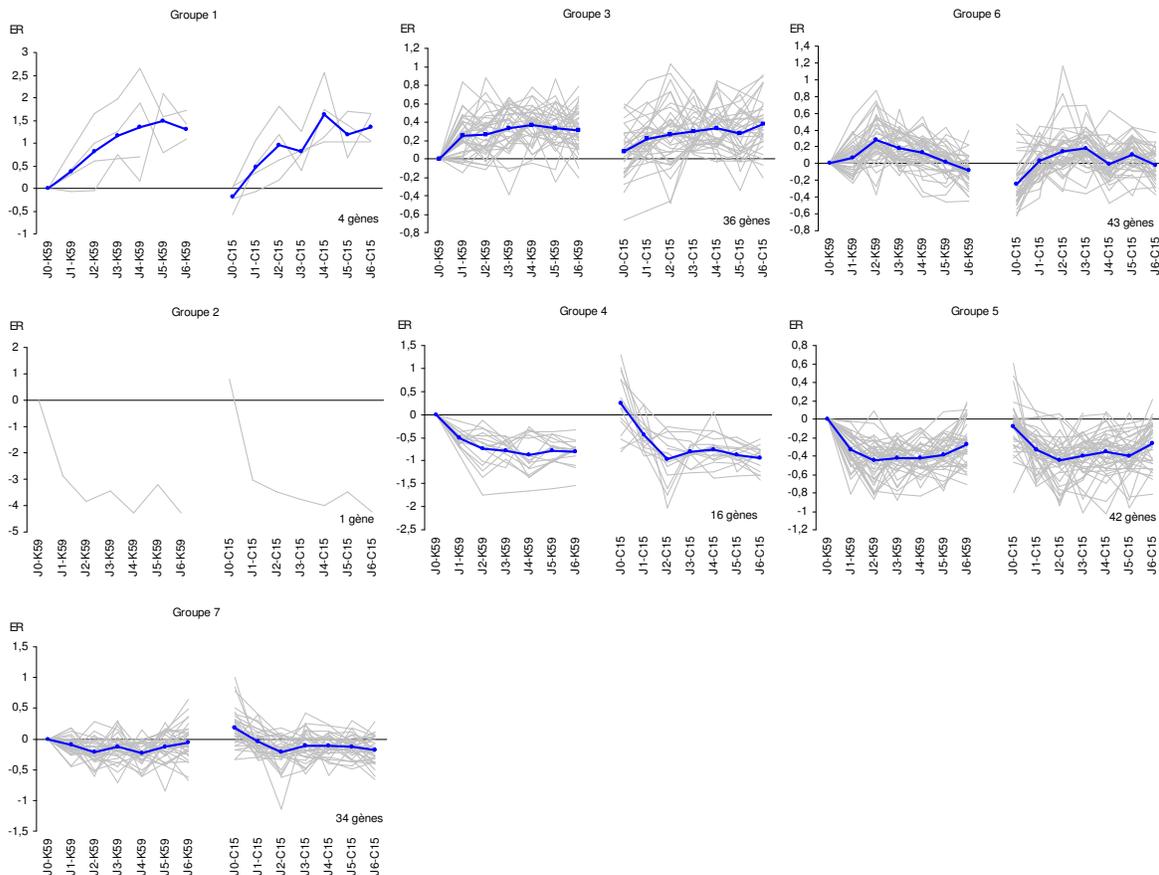


Figure 39 : Regroupement par la méthode de K-means des gènes de la sous-population A

Le groupe 1 représente les gènes montrant la plus forte hausse d'expression au cours de la cinétique. Parmi les 4 gènes représentés, un seul montre une fonction connue. Il code une protéine de transfert de lipide (LTP) (Cont0096). Le groupe 2 compte un seul gène, et celui-ci code la petite sous-unité de la rubisco (Cont0777). C'est le gène qui montre la plus forte baisse d'expression chez les deux génotypes au cours de la cinétique.

Parmi les 7 groupes définis, on observe deux types majeurs de profils : ceux correspondant aux gènes présentant une hausse d'expression par rapport au jour 0 chez les deux génotypes et classés dans les groupes 1, 3 et 6, et ceux correspondants aux gènes présentant une baisse d'expression et classés dans les groupes 2, 4, 5 et 7. La classification fonctionnelle des gènes des groupes 1, 3 et 6 (Figure 40 a) semble assez différente, pour certaines classes, de celle des gènes des groupes 2, 4, 5 et 7 (Figure 40 b). Par exemple, le métabolisme général est plus représenté par les gènes montrant une baisse d'expression et la synthèse protéique ainsi que le transport cellulaire sont des classes de fonction plus représentées par les gènes montrant une hausse d'expression chez les deux génotypes. Toutefois, les proportions de gènes liés à la défense, de gènes ne présentant pas d'homologie avec les bases de données et de gènes codant des protéines aux fonctions inconnues sont similaires pour les deux distributions.

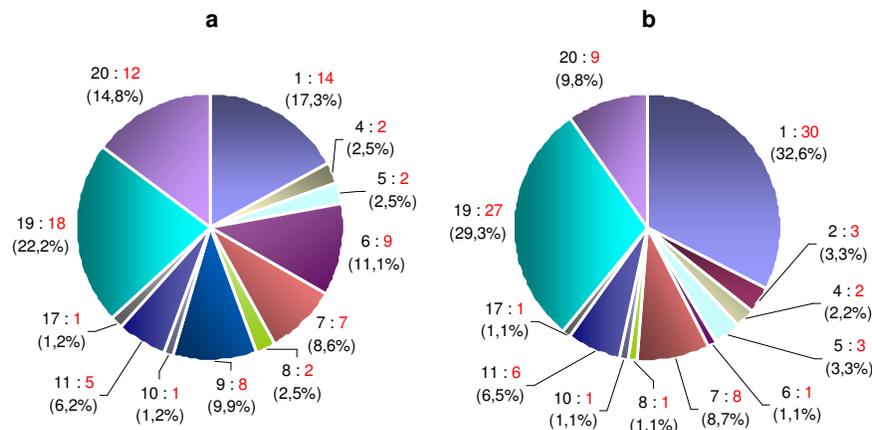


Figure 40 : Distribution par classe de fonction des gènes montrant une hausse d'expression (a) ou une baisse d'expression (b) chez les deux génotypes au cours de la cinétique en conditions d'embryogenèse somatique

Les ESTs ont été classés suivant leur fonction à l'aide des résultats des comparaisons des ESTs avec les séquences codantes d'Arabidopsis et les données du catalogue fonctionnel de MIPS (Munich Information centre for Proteins Sequences ; <http://mips.gsf.de/projects/funecat>) Pour chaque classe, le 1^{er} chiffre indique la classe de fonction et le chiffre en rouge indique l'effectif.

1 : Métabolisme ; 2 : Energie ; 3 : Stockage ; 4 : Cycle cellulaire et synthèse de l'ADN ; 5 : Transcription ; 6 : Synthèse protéique ; 7 : Devenir des protéines (repliement, modification, destination) ; 8 : Protéines fixatrices ou nécessitant un cofacteur ; 9 : Transport cellulaire ; 10 : Communication cellulaire ; 11 : Défense ; 12 : Interaction avec l'environnement cellulaire ; 13 : Interaction avec l'environnement ; 14 : Eléments transposables, protéines virales et plasmidiques ; 15 : Devenir cellulaire ; 16 : Développement ; 17 : Biosynthèse de composants cellulaires ; 18 : Différenciation cellulaire ; 19 : Protéines inconnues (regroupant localisation cellulaire, classification non certaine, et protéines non classées) ; 20 : Aucune homologie.

Pour la sous-population B, représentant les gènes montrant des profils d'expression différents chez K59 et C15, nous avons défini 21 groupes pour dissocier les profils d'expression par la méthode de K-means. Comme précédemment, le nombre de groupes a été défini après construction d'une courbe de valeurs FOM. Nous remarquons que le nombre de groupes définis pour le regroupement des gènes de la sous-population B est plus important, que celui défini pour la sous-population A, en raison ici de profils d'expression différents chez les deux génotypes, ce qui a pour conséquence d'augmenter la variété de profils considérés.

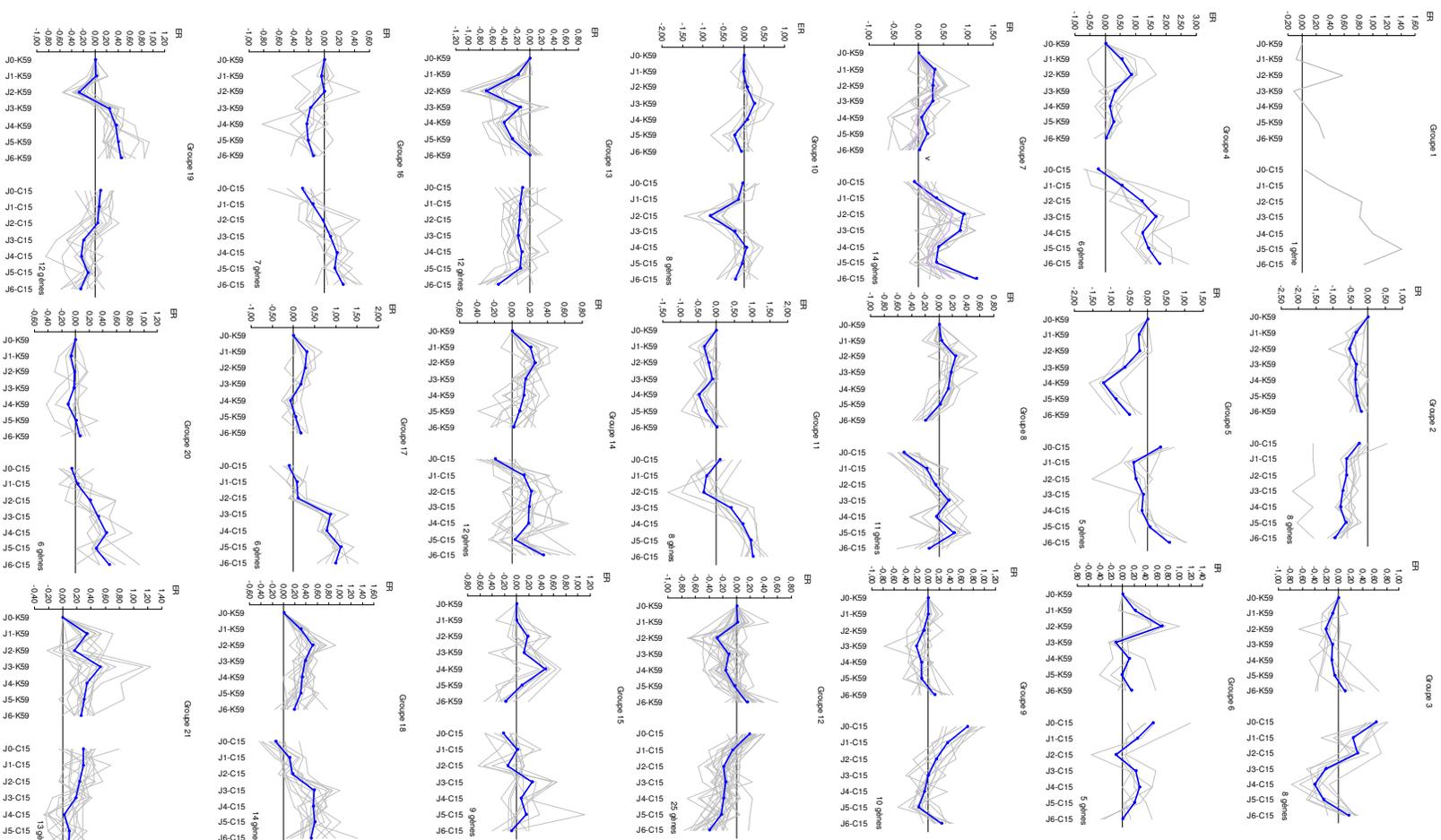


Figure 41 : Regroupement par la méthode de K-means des gènes de la sous-population B

Le regroupement par la méthode de K-means, semble avoir dans certains cas une assez bonne efficacité. Par exemple, le groupe 19 rassemble des gènes qui montrent une hausse d'expression significative à partir du jour 3 de cinétique chez le génotype embryogène et dont l'expression ne varie pas ou alors diminue significativement chez le génotype non embryogène. On retrouve dans ce groupe une certaine homogénéité d'un point de vue fonctionnel, puisque sur 12 gènes assemblés, 6 codent des protéines ribosomales, et que 5 des 6 autres gènes pourraient être en lien avec les divisions cellulaires observées dans les explants du génotype embryogène à partir du jour 4 de culture (Cont0077, OC0687, OC1929, OC0831 et OC1702) (Tableau 14).

Contig	Fonction
Cont0077	nuclear movement protein [<i>Medicago truncatula</i>]
OC1549_a	40S ribosomal protein S17 [<i>Lycopersicon esculentum</i>]
OC0390	40S ribosomal protein S2 [<i>Arabidopsis thaliana</i>]
Cont0023	40S ribosomal protein S6 [<i>Cicer arietinum</i>]
OC0082_a	60s acidic ribosomal protein [<i>Hyacinthus orientalis</i>]
OC0082_b	60s acidic ribosomal protein [<i>Hyacinthus orientalis</i>]
OC1833_b	60S ribosomal protein L3A [<i>Nicotiana tabacum</i>]
OC0687	arabinogalactan protein [<i>Daucus carota</i>]
OC1929	G protein beta subunit [<i>Nicotiana plumbaginifolia</i>]
OC0831	histone H2A [<i>Petroselinum crispum</i>]
OC1702	hypothetical protein, similar to H2B histone [<i>Arabidopsis thaliana</i>]
OC1814	Aucune homologie

Tableau 14 : Liste des gènes du groupe 19 issu de l'assemblage des profils d'expression des gènes de la sous-population B

On retrouve cette même homogénéité au sein du groupe 11, qui assemble des gènes présentant une hausse d'expression significative chez le génotype C15 à partir du jour 4 de culture et qui, au contraire, ne présentent pas de modulation d'expression ou qui présentent une baisse d'expression significative chez le génotype embryogène. Sur les 8 gènes assemblés dans le groupe 11, deux ne présentent pas de fonction connue et 3 représentent des gènes qui peuvent être reliés à la défense (Cont9039, OC1539, OC0839) (Tableau 15).

Contig	Fonction
OC1546	auxin transport protein [<i>Arabidopsis thaliana</i>]
Cont9039	metallothionein 1 [<i>Aster tripolium</i>]
OC1539	bacterial-induced class III peroxidase [<i>Gossypium hirsutum</i>]
OC1635_a	arachidonic acid-induced DEA1 [<i>Lycopersicon esculentum</i>]
OC1635_b	lipid binding [<i>Arabidopsis thaliana</i>]
OC0839	dormancy-associated protein [<i>Codonopsis lanceolata</i>]
OC1658	hypothetical protein [<i>Arabidopsis thaliana</i>]
OC0502	Aucune homologie

Tableau 15 : Liste des gènes du groupe 11 issu de l'assemblage des profils d'expression des gènes de la sous-population B

Dans d'autres cas, par exemple celui du groupe 12, la méthode de K-means a assemblé au sein d'un même groupe, des gènes aux profils assez différents, au moins pour l'un des deux génotypes. L'exemple donné présente les profils de l'OC1274_a et de l'OC0088 assemblés dans ce groupe 12 (Figure 42). On remarque que les deux gènes bien que présentant des profils d'expression assez similaires chez le génotype C15, présentent des profils très différents chez le génotype K59. Les analyses de variance ont montré, par ailleurs, que l'expression du gène représenté par l'OC1274_a ne varie pas significativement en cours de cinétique chez le génotype K59 alors que l'expression du gène correspondant à l'OC0088 montre une baisse significative chez le génotype K59 au cours de la cinétique.

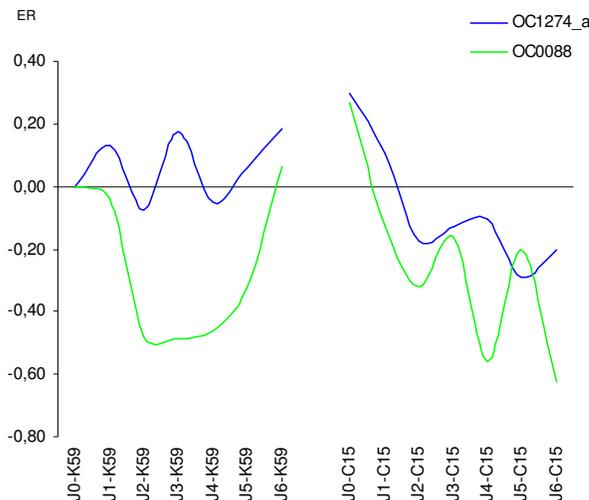


Figure 42 : profils d'expression de deux gènes issus du groupe 12
Le groupe a été formé à partir des gènes de la sous-population B. OC1274_a : NADH glutamate synthase isoform 2 ; OC0088 : DNA binding/protein dimerization.

La méthode de K-means semble ici montrer ses limites, peut-être à cause des petites valeurs des ratios d'expression enregistrées, notamment pour le génotype embryogène. En

effet, comme suggéré au cours de la partie II, en raison de la faible proportion dans l'explant de cellules qui entrent dans le processus d'embryogenèse somatique ou de cellules environnantes qui participent à la mise en place du processus chez le génotype K59, les différences d'abondance de transcrits entre les deux génotypes pour les gènes préférentiellement exprimés chez K59 pourraient être faibles, et engendrer ainsi des ratios d'expression limités. Une autre explication peut aussi se trouver dans le nombre de points de la cinétique. Pour l'étape 1 de regroupement, nous avons soumis les profils obtenus pour les génotypes K59 et C15 de façon indépendante, ce qui limitait alors les profils d'expression à 7 points. Au cours de l'étape 2, nous avons soumis au logiciel de regroupement des profils d'expression à 13 points (profil d'expression chez le génotype K59 et profil d'expression chez le génotype C15 cumulés). Cela se traduit par une baisse de l'efficacité de l'algorithme de regroupement et par l'obtention de valeurs FOM relativement élevées en comparaison à celles obtenues au cours de l'étape 1. Pourtant, au cours de l'étape 2, en dépit d'une valeur FOM assez élevée pour la définition des 21 groupes (2,8), le regroupement opéré semble d'assez bonne qualité pour les gènes présentant des ratios d'expression élevés (notamment, groupes 4, 11, 19 et 21).

Enfin, nous pouvons remarquer que pour la majorité des 36 gènes présentant une expression différentielle au stade initial de la culture et montrant une modulation d'expression significative au cours de la cinétique, la modulation d'expression conduit généralement à l'égalisation des quantités de transcrits entre les deux génotypes. Ces gènes sont présents dans les groupes 3, 8 9 et 12 (Figure 41).

En complément des regroupements effectués, qui présentent, parfois, une efficacité limitée, nous nous sommes intéressés à la classification fonctionnelle des gènes qui permet, dans certains cas, de mieux comprendre les fonctions mises en œuvre dans les explants des génotypes K59 et C15 au cours de la culture en conditions d'embryogenèse somatique.

3) Classification fonctionnelle

Une sélection de 249 gènes a été réalisée à partir de l'ensemble des gènes de la sous-population B (200 gènes présentant des profils d'expression différents chez les deux génotypes), ainsi qu'à partir de 49 gènes de la sous-population A (gènes aux profils d'expression similaires chez les deux génotypes) pour lesquels les analyses de variance ont montré qu'ils présentaient une hausse ou une baisse d'expression significative chez les deux génotypes.

Les 249 gènes sélectionnés ont été soumis à une classification fonctionnelle suivant le catalogue fonctionnel de Kegg (Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.jp/kegg/>). Cette classification a eu pour but de compléter celle effectuée précédemment à l'aide du catalogue fonctionnel de MIPS. Le regroupement des données des deux classifications a permis d'attribuer à 192 des 249 gènes une fonction. La Figure 43 donne la liste des 249 gènes par classe de fonction avec 26 gènes présentant des fonctions potentielles multiples.

Le métabolisme a été découpé en 6 sous-classes : métabolisme carboné et énergie, métabolisme des nucléotides, métabolisme des acides aminés, métabolisme lipidique, métabolisme secondaire et autres métabolismes. Au sein du métabolisme carboné et de l'énergie, nous mettons en évidence qu'un gène codant la petite sous-unité de la rubisco (Cont0777) montre une baisse d'expression significative au cours de la cinétique chez les deux génotypes. Deux autres gènes peuvent être liés à l'activité photosynthétique et codent un précurseur d'une protéine du photosystème I (OC1802) et un précurseur d'une protéine fixatrice de la chlorophylle (OC1326). Ces deux gènes montrent également une diminution de leur expression au cours de la cinétique pour les deux génotypes. Toujours au niveau du métabolisme carboné, plusieurs gènes peuvent être liés, selon le catalogue fonctionnel de Kegg, à la glycolyse ou à la néoglucogenèse. Deux gènes codent une glyceraldéhyde-3-phosphate déshydrogénase (Cont0001) et une émolase (Cont0035), enzymes associées à la glycolyse. Les deux gènes montrent une hausse d'expression significative chez le génotype embryogène alors qu'ils ne montrent pas de modulation d'expression significative chez l'autre génotype. Quatre gènes, codant une pyruvate décarboxylase (OC1896), une aldéhyde déshydrogénase (OC1008), une dihydrolipoamide acétyltransférase (OC1565) et une alcool déshydrogénase (OC0316), quatre enzymes annotées comme étant reliées à la néoglucogenèse, montrent tous une baisse d'expression chez le génotype non embryogène. Deux de ces quatre gènes (OC1008 et OC0316) montrent également une baisse d'expression chez le génotype embryogène.

Résultats et discussion

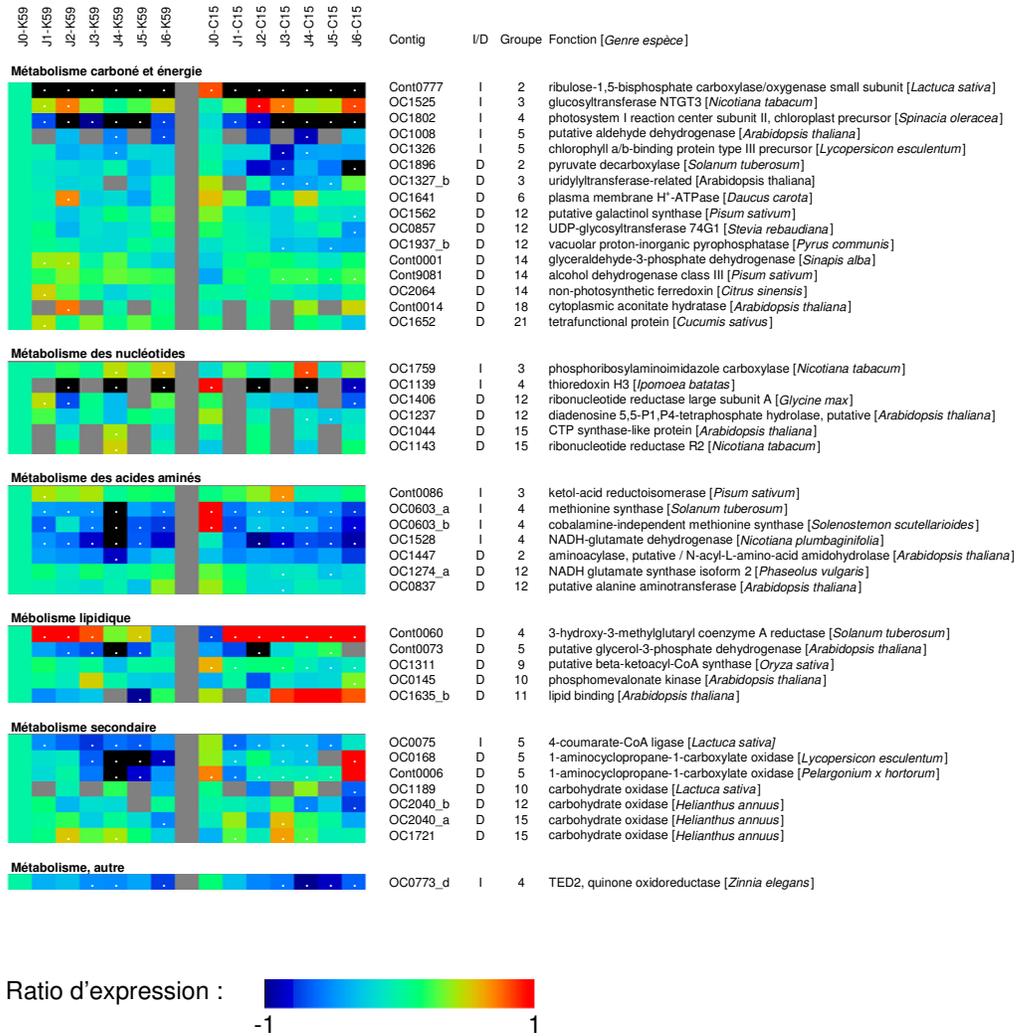


Figure 43 : Résultats de l'analyse transcriptionnelle et classification fonctionnelle des 249 gènes sélectionnés

Chaque case de couleur correspond à une intensité du ratio d'expression (voir échelle). Les ratios d'expression ont été calculés en utilisant la condition K59J0 comme référence commune. Les points blancs, au centre des cases colorées, représentent les ratios d'expression significatifs. I/D, I : gènes présentant des profils d'expression identiques pour les deux génotypes et issus de la sous-population A ; D : gènes présentant des profils d'expression différents pour les deux génotypes et provenant de la sous-population B. Groupe : indique le numéro du groupe suite à l'assemblage des gènes par la méthode de K-means (Figure 39 et Figure 41).

Résultats et discussion

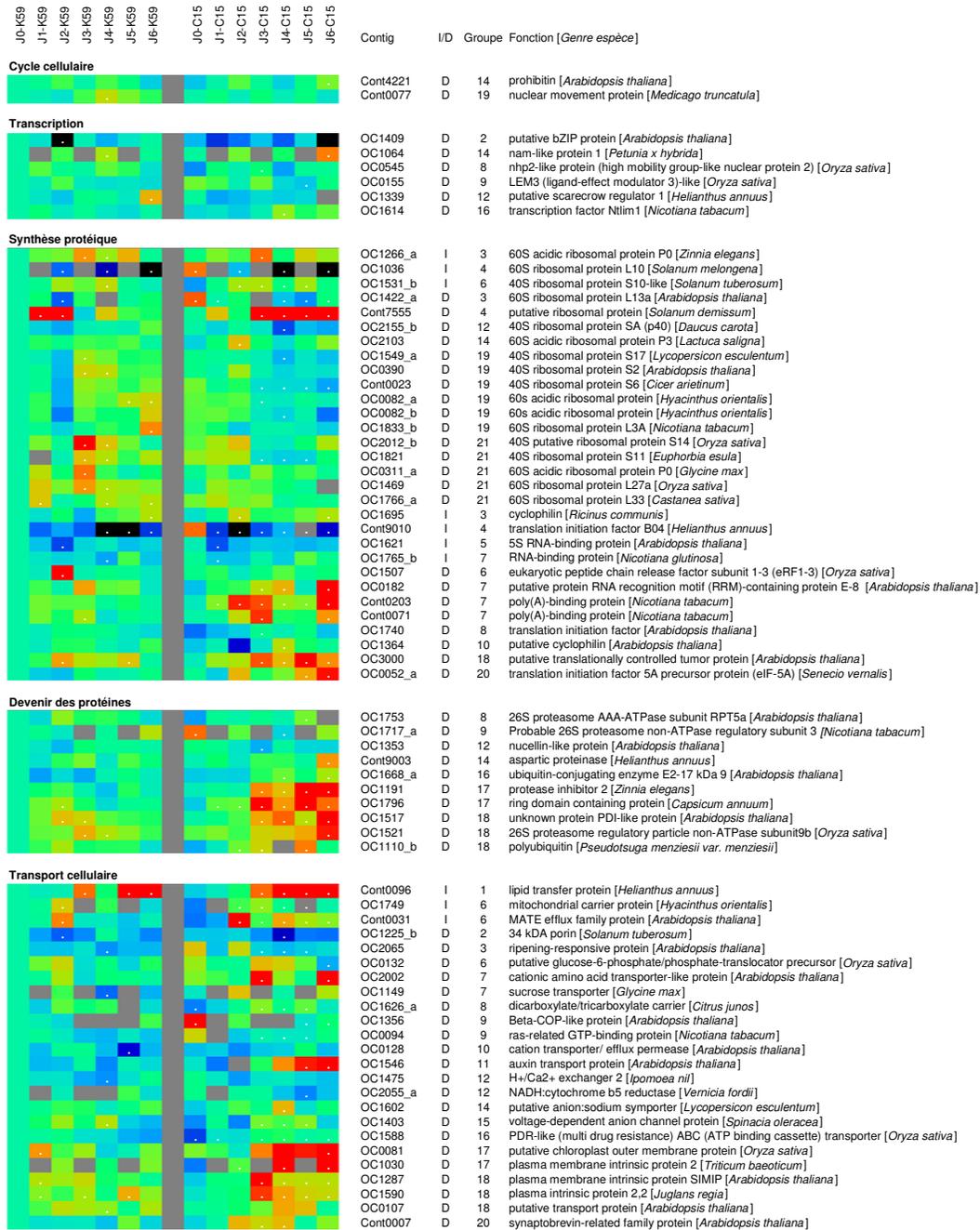


Figure 43 (suite)

Résultats et discussion

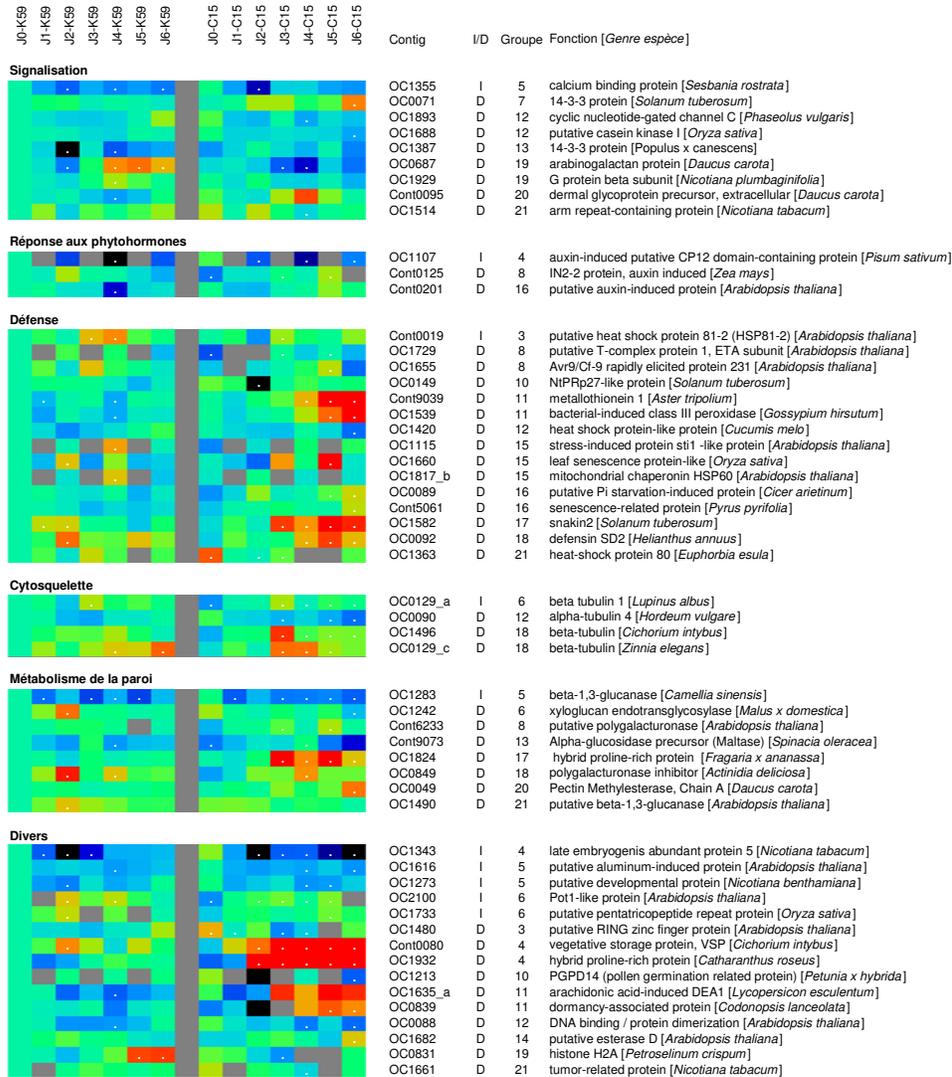


Figure 43 (suite)

Résultats et discussion

Fonctions multiples

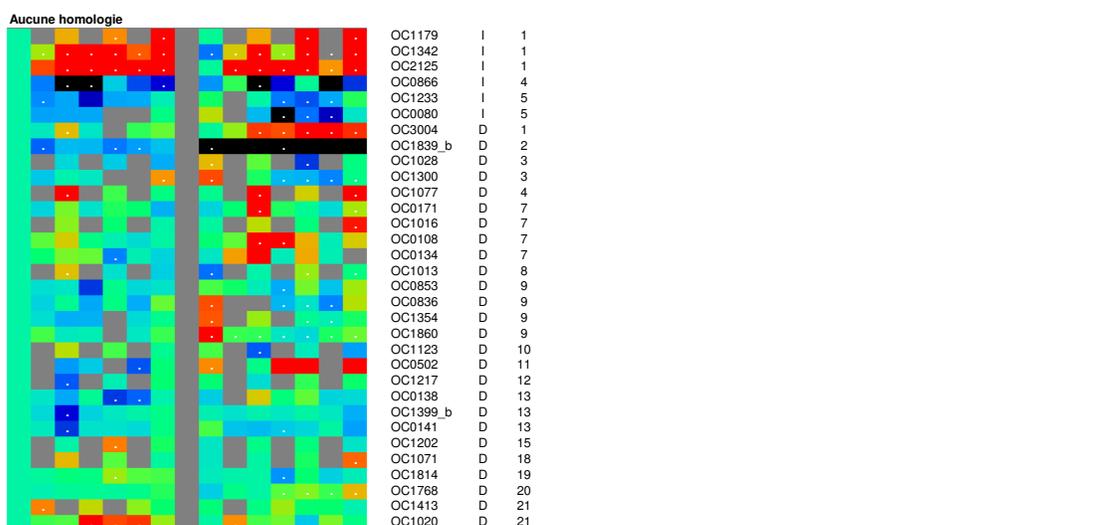
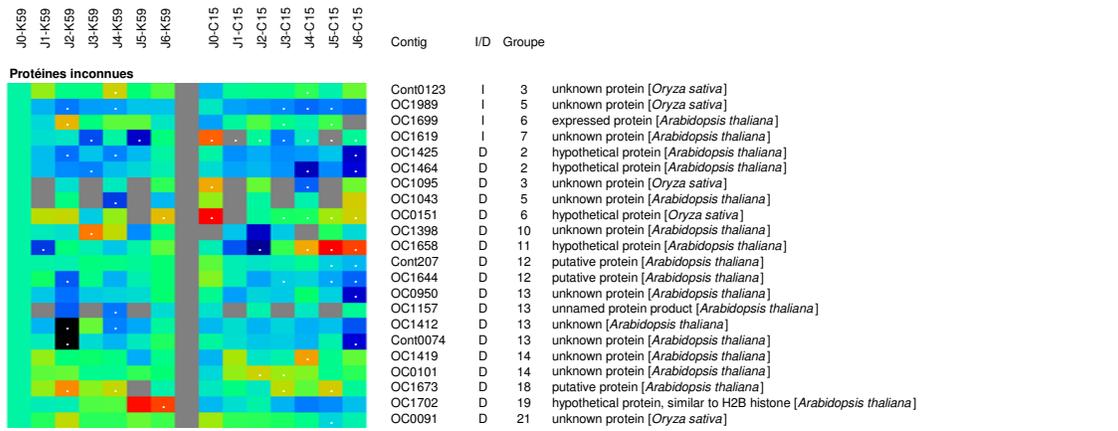
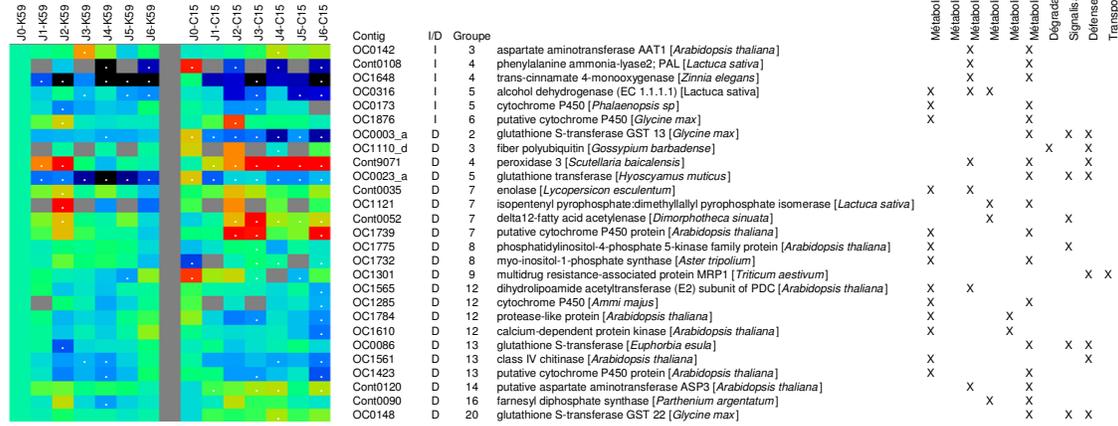


Figure 43 (suite et fin)

Au niveau du métabolisme des nucléotides, 2 gènes montrent des profils similaires chez les deux génotypes (OC1759 et OC1139) et quatre des profils différents. Sur ces 4 gènes, 3 montrent une hausse d'expression significative chez le génotype embryogène et ne montrent pas de modulation ou montrent une baisse d'expression chez le génotype non-embryogène. Ces gènes codent une ribonucléotide réductase (OC1406), une CTP synthase (OC1044) et une ribonucléotide réductase R2 (OC1143). Les deux derniers gènes ont des profils d'expression classés dans le groupe 15.

Sur un total de 15 gènes liés au métabolisme des acides aminés, en considérant également les gènes aux fonctions multiples, 8 ont montré des profils d'expression similaires chez les deux génotypes et 6 de ces 8 gènes montrent une baisse d'expression au cours de la cinétique d'étude chez les génotypes K59 et C15. Parmi ces 6 gènes, 4 ont été classés dans le groupe 4 et codent une méthionine synthase (OC0603_a), une méthionine synthase indépendante de cobalamine (OC0603_b), une NADH-glutamate déshydrogenase (OC1528) et une phénylalanine ammonia-lyase2 (Cont0086). Cependant, la phénylalanine ammonia-lyase2 est une enzyme impliquée dans de nombreuses voies du métabolisme (Figure 43) et son implication dans le métabolisme des acides aminés n'est qu'une possibilité parmi d'autres.

Un nombre important de gènes (43) est impliqué dans le métabolisme des protéines (synthèse et devenir). La majorité des gènes impliqués dans la synthèse protéique codent des protéines ribosomales (18) et montrent globalement une hausse d'expression chez le génotype embryogène au cours de la cinétique alors qu'aucune modulation, ou une baisse d'expression, est constatée chez l'autre génotype. Au contraire, nous remarquons que la majorité des gènes impliqués dans le devenir des protéines montrent une hausse d'expression chez le génotype C15 et ne montrent pas de modulation d'expression chez K59. La plupart des gènes impliqués dans le devenir des protéines sont, en fait, impliqués dans leur dégradation et codent par exemple une ubiquitine (OC1668_a) ou une protéine du protéasome 26S (OC1521).

Parmi les 5 gènes liés à la transcription, nous remarquons la présence d'un facteur de transcription de type Scarecrow (OC1339), impliqué dans diverses étapes du développement des plantes ; celui-ci montre une hausse d'expression chez le génotype embryogène au jour 6 de la cinétique et aucune modulation d'expression chez le génotype C15. Nous remarquons également la présence d'un facteur de transcription lim (OC1614) impliqué potentiellement

dans la voie de synthèse des lignines qui présente une hausse d'expression chez le génotype C15 non observée chez le génotype K59.

Le transport et la défense, qui sont des classes de fonctions fortement représentées avec respectivement 22 et 17 gènes, semblent plutôt constituées par des gènes montrant une hausse d'expression chez le génotype non-embryogène, non observée chez le génotype embryogène. Ainsi le transport et la défense pourraient être des fonctions particulièrement importantes pour la réponse du génotype non-embryogène aux conditions de culture.

Au total, 9 gènes pourraient être impliqués dans des voies de signalisation chez les deux génotypes. Parmi ces gènes, notons la présence d'un gène codant une protéine G (OC1929) et montrant une hausse d'expression chez le génotype K59 non observée chez le génotype C15, et de deux gènes codant des protéines 14-3-3 (OC0071 et OC1387) dont l'un montre une hausse d'expression spécifique au génotype C15 et l'autre une baisse d'expression spécifique au génotype K59. Nous observons aussi qu'un gène codant une protéine ribosomale 40 S6 (Cont0023) potentiellement impliquée dans la voie de signalisation via la protéine TOR (target of rapamycin), montre une baisse d'expression significative chez C15, à partir du jour 3 de la cinétique.

Au sein des 32 gènes ne présentant pas d'homologie avec les bases de données, nous remarquons la présence d'un gène au profil d'expression particulier (OC1839_b) (Figure 41, groupe 2). Ce gène montre, en effet, un niveau de transcrits bien plus bas chez le génotype non embryogène que chez le génotype embryogène et ceci dès le stade initial de la cinétique d'étude. Un profil aussi marqué n'a été retrouvé ni parmi les 417 gènes présentant une modulation d'expression significative au cours de la cinétique chez l'un et/ou l'autre des génotypes, ni parmi les 681 gènes restants ne présentant pas de modulations d'expression significatives.

C. Discussion

1) Différences transcriptomiques au stade initial de la cinétique

Un plus grand nombre de gènes est préférentiellement exprimés chez C15 que chez K59 au stade initial de la culture (respectivement 61 et 20). Ceci pourrait indiquer une réactivité plus rapide du génotype C15 aux conditions de stress appliquées. Cependant, il ne faut pas perdre de vue que l'analyse transcriptomique menée entre les deux génotypes au stade initial de la culture ne permet pas de dissocier les différences intrinsèques des deux génotypes, des réponses rapides qui auraient pu avoir lieu depuis le prélèvement de la feuille sur la plante jusqu'à la congélation des explants. Toutefois, un certain nombre de gènes codant des protéines en lien avec la défense peuvent potentiellement être reliés à une réponse rapide des génotypes. En effet, au moins un gène (OC0018) codant une protéine identifiée comme étant induite par la blessure chez le fruit de la tomate (Parsons and Mattoo, 1991), et un gène codant un transporteur ABC potentiellement lié à la défense (OC1588) sont préférentiellement exprimés chez K59, alors qu'un gène codant une protéine MRP1 associée à la défense contre les drogues (OC1301) et un gène codant une protéine de choc thermique (OC1363) sont préférentiellement exprimés chez C15.

Pour les autres gènes, il est plus délicat de juger s'ils sont induits très précocement par les conditions de culture, ou si les différences d'expression observées chez les deux génotypes sont intrinsèques. Cependant, nous avons privilégié le choix d'un tel J0 comme point de référence car à ce stade, les explants ont subi l'aseptisation comme tous les autres points de la cinétique. Nous remarquons, par ailleurs, que sur les 80 gènes exprimés différemment chez les deux génotypes au jour 0 de culture, seulement 36 montrent une modulation d'expression significative au cours de la cinétique chez l'un et/ou l'autre des deux génotypes, et pour la majorité d'entre eux, les profils d'expression en cours de cinétique sont assez semblables. Il sera nécessaire de réaliser une comparaison entre une condition obtenue à partir d'explants congelés immédiatement après prélèvement et la condition J0, utilisée comme point initial de la cinétique au cours de cette étude. Cette expérience permettra de distinguer les différences intrinsèques des deux génotypes, des réponses rapides liées aux stress appliqués.

2) Réponse commune des deux génotypes aux conditions de culture appliquées

Les trois gènes liés à l'activité photosynthétique et codant, la petite sous-unité de la rubisco (Cont0777), un précurseur d'une protéine du photosystème I (OC1802) et un précurseur d'une protéine fixatrice de la chlorophylle (OC1326), montrent une forte baisse d'expression chez les deux génotypes au cours de la cinétique (Figure 43). L'induction de l'embryogenèse somatique se déroulant à l'obscurité, cette observation témoigne de l'arrêt de l'activité photosynthétique dans les explants des génotypes K59 et C15.

D'après la classification fonctionnelle entreprise, 4 gènes codant une pyruvate décarboxylase (OC1896), une aldéhyde déshydrogénase (OC1008), une dihydrolipoamide acétyltransférase (OC1565) et une alcool déshydrogénase (OC0316), ont été annotés par le catalogue fonctionnel de Kegg comme étant impliquées dans la néoglucogenèse. De façon globale, ces 4 gènes montrent une baisse d'expression chez les deux génotypes au cours de la cinétique. Si nous observons une baisse des quantités de transcrits au cours de la cinétique pour ces gènes, cela signifie que ces gènes devaient présenter un niveau d'expression assez élevé au stade initial de la culture. Cependant, il est difficile d'imaginer que la néoglucogenèse fonctionne de façon intense dans les feuilles des deux génotypes cultivés en serre. En effet, les feuilles choisies pour préparer les cultures sont des feuilles dont l'expansion est juste terminée mais qui ne sont pas visiblement sénescentes. Une explication possible du résultat pourrait se trouver dans le fait que certaines de ces enzymes peuvent avoir une activité réversible (alcool déshydrogénase) ou peuvent être liées à d'autres voies du métabolisme carboné. Ce résultat nous montre qu'il faut rester prudent, pour certains gènes, sur les fonctions potentielles que nous leur avons attribués en se basant sur les catalogues fonctionnels de Mips et de Kegg.

Au niveau du métabolisme des acides aminés, les résultats indiquent que 8 gènes présentent la même réponse chez les deux génotypes et que pour 6 de ces 8 gènes, la réponse est une baisse d'expression (Figure 43).

Les conditions d'induction de l'embryogenèse somatique provoquent des modifications du métabolisme, observées pour les deux génotypes, qui conduisent à l'arrêt de certaines fonctions. De la même façon, certains gènes présentent une hausse d'expression chez les deux génotypes. Ces gènes sont liés à diverses classes de fonction définies : par exemple, le métabolisme (3-hydroxy-3-méthylglutaryl coenzyme A réductase, Cont0060), le transport (protéine de transfert de lipides, Cont0096), le cytosquelette (beta-tubuline,

OC0129_c). Une partie des gènes liés à la défense (OC1582 et OC0092) présente également une hausse d'expression chez les deux génotypes et indique qu'il existe une réponse de défense commune aux deux génotypes. Enfin deux gènes (OC1342 et OC2125), ne présentant pas d'homologie avec les bases de données, montrent une très forte hausse d'expression chez les deux génotypes dès le premier jour de culture (Figure 39, groupe 1).

3) Réponse du génotype embryogène

Dès le début de la cinétique (J1 à J2), nous mettons en évidence que des gènes, impliqués dans des métabolismes qui permettent la production d'ATP, présentent une hausse d'expression significative chez le génotype embryogène, non observée chez le génotype non embryogène. Deux gènes peuvent être liés à la glycolyse : l'un code une émolase (Cont0035) et l'autre une glyceraldéhyde-3-phosphate déshydrogénase (Cont0001). Un gène peut être associé au cycle de Krebs et code une aconitate hydratase (Cont0014), un autre, lié à la phosphorylation oxydative, code une H⁺-ATPase (OC1641). Le génotype K59 semble utiliser le saccharose contenu dans le milieu de culture pour produire de l'énergie qui pourrait être utilisée pour la transdifférenciation des cellules et leur entrée en division. Cette hausse d'expression de gènes liés à la glycolyse ou au cycle de Krebs pourrait être directement induite, chez le génotype K59, par le saccharose contenu dans le milieu. D'autres gènes pourraient également être impliqués dans la synthèse d'ATP, puisqu'un gène associé à l'oxydation d'acides gras (OC1652) présente une hausse d'expression exclusive au génotype K59 au jour 1 de culture et qu'un gène codant une ferredoxine non-photosynthétique a été classé dans le même groupe (groupe 14) que le gène codant la glyceraldéhyde-3-phosphate déshydrogénase (Cont0001) évoquée plus haut.

Un fait marquant chez le génotype embryogène a été d'identifier un nombre conséquent de gènes (9) codant des protéines ribosomales et montrant une hausse d'expression significative chez le génotype K59, non observée chez le génotype C15. Ce résultat confirme les résultats préliminaires avancés dans les parties II et III qui ont montré l'implication des gènes codant des protéines ribosomales dans la réponse du génotype embryogène. Ces protéines ribosomales pourraient jouer un rôle, dans la mise en place de nouveaux transcriptomes et dans la synthèse de ribosomes, dans les cellules qui entrent dans le processus d'embryogenèse somatique ou dans les cellules environnantes qui peuvent participer à la mise en place du processus. Les gènes codant des protéines ribosomales ont été classés par la méthode de K-means dans deux groupes différents : 19 et 21, ce qui montre une certaine diversité de leurs profils d'expression. Les gènes classés dans le groupe 19 montrent

une hausse d'expression à partir du jour 3 qui reste plus ou moins stable jusqu'au jour 6 alors que les gènes classés dans le groupe 21 montrent des pics d'expression au jours 1 et 3 de la cinétique. Nous remarquons que les 9 gènes montrent tous une baisse d'expression au jour 2 de la cinétique. La diversité des profils d'expression témoigne de la complexité des mécanismes se déroulant au cours des phases précoces du processus. Chez le zinnia, au cours de la transdifférenciation de cellules du mésophylle en cellules du xylème, plusieurs vagues d'expression de gènes codant des protéines ribosomales ont pu également être observées et ceci sur une courte période d'étude, s'étendant de 0 à 4 jours (Demura *et al.*, 2002).

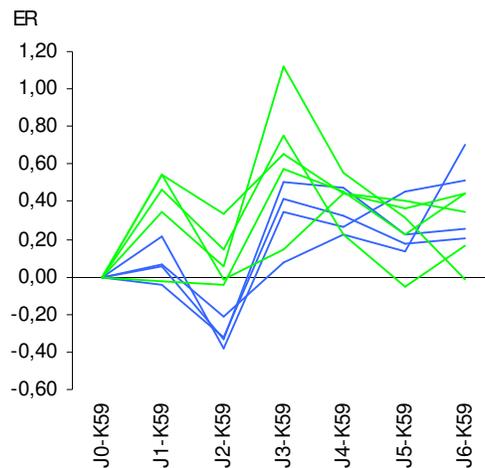


Figure 44 : Profils d'expression des gènes codant des protéines ribosomales et montrant des variations d'expression observées uniquement chez le génotype embryogène. Les profils d'expression des gènes classés dans le groupe 19 sont représentés en bleu et ceux classés dans le groupe 21 en vert.

Parmi ces 9 gènes codant des protéines ribosomales, nous remarquons la présence d'un gène codant une protéine ribosomale P0 (OC1266_a). Chez le maïs, il a été montré qu'un gène codant une protéine ribosomale P0 est fortement induit au cours de la phase G1 du cycle cellulaire lors de la formation d'embryons somatiques (Dresselhaus *et al.*, 1999). Ainsi, le gène identifié chez la chicorée pourrait potentiellement être impliqué dans la préparation des divisions cellulaires conduisant à la formation des embryons chez le génotype K59.

D'autres gènes semblent être en lien avec les divisions cellulaires. En effet, un gène (Cont0077) code une protéine présentant de fortes homologues avec une protéine identifiée chez *Medicago truncatula* et annotée comme étant impliquée dans le déplacement du noyau. Cette protéine pourrait interagir avec les protéines du cytosquelette. Le gène identifié montre une hausse d'expression significative au jour 4 de culture chez le génotype K59 et ne montre pas de modulations d'expression significatives chez le génotype C15. Le jour 4 de culture

correspond d'un point de vue cytologique à l'apparition des premières divisions dans les explants du génotype embryogène. Ainsi, le produit de ce gène pourrait être impliqué dans l'entrée en division des cellules embryogènes. Il est intéressant de remarquer que deux gènes codant une protéine G (OC1929) et une AGP (arabinogalactan protein) (OC0687) présentent des profils d'expression similaires à celui du gène codant l'hypothétique protéine de déplacement du noyau (groupe 19). L'OC1929 montre de fortes homologies avec un gène codant une protéine G de luzerne qui pourrait jouer un rôle dans les divisions cellulaires induites par les auxines et les cytokinines (McKhann *et al.*, 1997). Le gène codant l'AGP montre, quant à lui, des homologies avec le gène DcAGP1 de carotte dont l'ADNc a été isolé dans des suspensions cellulaires. DcAGP1 code une AGP non classique qui présente de fortes similitudes avec une famille de protéines riches en proline (Baldwin *et al.*, 2001). Les AGP sont supposées être impliquées dans de nombreuses voies de signalisation. Elles sont, par ailleurs, connues chez la chicorée pour être essentielles à la formation d'embryons somatiques (Chapman *et al.*, 2000) et une corrélation entre le nombre d'embryons produits et la quantité d'AGP libérée dans le milieu de culture a pu être établie (Windels, 2005). Il faut noter que ces travaux sur la chicorée portaient sur des AGP dites classiques alors que le gène identifié au cours de cette étude code potentiellement une AGP non classique. Les AGP « classiques » et « non classiques » montrent quelques différences au niveau de leur ossature protéique. Ces différences entraînent, par exemple, l'impossibilité pour les AGP « non classiques », contrairement aux AGP « classiques », d'être ancrées au niveau membranaire (Windels, 2005). Ainsi, il est raisonnable d'imaginer que ces deux gènes, codant une protéine G et une AGP non classique, et présentant une hausse d'expression à l'instant où les premières divisions sont initiées dans les explants du génotype K59, pourraient être impliqués dans une voie de signalisation qui conduirait à la division cellulaire.

Parmi les gènes codant des protéines impliquées dans la synthèse des nucléotides, 3 montrent une hausse d'expression significative chez le génotype embryogène et ne montrent pas de modulation ou montrent une baisse d'expression chez le génotype non-embryogène. Ces gènes codent une ribonucléotide réductase (OC1406), une CTP synthase (OC1044), une ribonucléotide réductase R2 (OC1143). Il a été montré chez le tabac, qu'un gène codant une ribonucléotide reductase R2 montre un pic d'expression à la transition des phases G1/S et peut donc être directement associé à la division cellulaire (Chaboute *et al.*, 1998). En plus des gènes liés au métabolisme des nucléotides, un gène codant une histone H2A (OC0831) présente également une hausse d'expression significative et exclusive au génotype

embryogène. Par ailleurs, un autre gène codant, cette fois-ci, une histone H2B (OC1702) a montré une expression différentielle entre les deux génotypes au jour 4 de la cinétique. Cependant, ce gène n'a montré de modulation d'expression significative chez aucun des deux génotypes au cours de la cinétique d'étude. Ces résultats semblent traduire une synthèse intense d'ADN en vue de préparer les divisions cellulaires nécessaires à la formation des embryons somatiques.

Un gène codant une protéine CDC48 (Cell Division Cycle) a, lui aussi, montré une expression différentielle entre les deux génotypes au jour 4 de la cinétique et n'a pourtant pas montré de modulation d'expression significative au cours de la cinétique d'étude. Le gène *AtCDC48* d'*Arabidopsis* est fortement exprimé dans les cellules en prolifération et dans les cellules présentant une croissance rapide (Feiler *et al.*, 1995). L'implication potentielle de ce gène au cours des divisions cellulaires conduisant à la formation d'embryons somatiques restera à valider par PCR en temps réel.

Deux gènes liés au métabolisme pariétal et codant une xyloglucane endotransglycosylase (OC1242) et une β -1,3 glucanase (OC1490) montrent une expression transitoire avec un pic au jour 2 de culture chez le génotype embryogène. Cette modulation d'expression pourrait témoigner d'un remaniement de la paroi en vue de l'entrée en division des cellules. Par ailleurs, nous remarquons, que l'ADNc de β -1,3 glucanase identifié au cours de cette étude est différent de celui isolé précédemment au laboratoire et dont les transcrits correspondant ne sont détectables qu'à partir du jour 3 de cinétique en conditions d'embryogenèse somatique (Helleboid *et al.*, 2000a).

Un autre résultat a montré qu'un gène codant un facteur de transcription de type Scarecrow montre une hausse d'expression significative chez le génotype embryogène au jour 6 de culture. Les rôles des gènes Scarecrow sont multiples au cours du développement des végétaux. Scarecrow participe, en général, aux processus qui impliquent une division asymétrique. Il participe, par exemple, à la formation des cellules du stomate, au cours de laquelle il permet la mise en place d'une polarité au sein du cytoplasme (Kamiya *et al.*, 2003). Des analyses moléculaires menées chez *Arabidopsis* ont montré l'importance du gène Scarecrow, au niveau de la racine, pour la régulation de la division asymétrique de la cellule initiale qui permet la formation de l'endoderme et du cortex. Scarecrow joue ainsi un rôle dans la mise en place de l'organisation radiale de la racine (Di Laurenzio *et al.*, 1996). Chez les mutants *scr-1*, l'absence de division asymétrique de l'initiale conduit à la formation d'une seule couche de cellules aux identités mélangées (Di Laurenzio *et al.*, 1996). De plus, il a été

montré que Scarecrow joue également un rôle dans la mise en place de l'organisation radiale des tiges. Par ailleurs, il est aussi impliqué dans l'organisation radiale des pôles apico-radicaux au cours de l'embryogenèse (Wysocka-Diller *et al.*, 2000). Chez le coton et chez la sapinette blanche, Scarecrow joue un rôle potentiel dans les phases précoces de l'embryogenèse somatique, sans, toutefois, connaître son rôle exact au cours de ce processus (Stasolla *et al.*, 2003; Zeng *et al.*, 2006). Chez la chicorée, le gène de type Scarecrow montre une hausse d'expression significative au jour 6 de culture chez le génotype K59, c'est-à-dire à un stade où l'on observe déjà au sein des explants du génotype embryogène, des embryons multicellulaires. L'expression du gène identifié n'est peut-être pas liée à la première division de la cellule embryogène, puisque les premières divisions cellulaires ont lieu au jour 4 de culture. Toutefois, il ne faut pas oublier que l'expression différentielle de Scarecrow pourrait ne pas être détectée avant le jour 6 en raison de la sensibilité modérée de la technique de microarray. Le produit du gène Scarecrow identifié ne participe probablement pas à la mise en place de la polarité apico-basale de l'embryon puisque cet événement se met en place bien plus tard au cours de la cinétique. Par contre ce gène pourrait réguler les modalités des divisions cellulaires et empêcher un développement anarchique des structures embryonnaires.

4) Réponse du génotype non-embryogène

Le génotype C15 a été choisi comme contrôle négatif de l'embryogenèse somatique, puisque placé dans les conditions d'induction de l'embryogenèse somatique, il n'a pas la capacité à former des embryons. Un autre élément important dans le choix de ce génotype a été qu'il ne possède pas non plus les capacités de callogenèse et d'organogenèse, du moins au cours de la cinétique d'étude ciblant les phases précoces du processus et dans les conditions de cultures appliquées. Sur les six premiers jours de culture, nous n'observons en effet que très peu de changements morphologiques des cellules. Cependant, nos résultats montrent que le génotype C15 ne semble pas du tout passif aux conditions de culture. Le nombre important de gènes (298) montrant une modulation d'expression au cours de la cinétique d'étude en témoigne. Ce résultat pourrait indiquer qu'à l'instar des cellules des explants du génotype embryogène, un nouveau transcriptome se met en place dans les cellules du génotype non-embryogène.

Parmi les gènes présentant une hausse d'expression exclusive au génotype C15, on trouve un gène présentant de fortes homologues avec le facteur de transcription Ntlim1 de tabac (OC1614). Ce facteur de transcription est impliqué notamment dans la synthèse de lignine. En effet, des plants de tabac transformés avec un gène antisens Ntlim1 présentent une

réduction de 27% des teneurs en lignine (Kawaoka *et al.*, 2000). On remarque également, qu'un gène codant une peroxydase 3 (Cont9071), enzyme impliquée, selon le catalogue fonctionnel de Kegg, dans la voie de synthèse des lignines, montre également une forte hausse d'expression chez C15 au cours de la cinétique. Malgré ces résultats, il est très difficile de conclure à une quelconque synthèse de lignine dans les cellules des explants obtenus à partir du génotype C15. En effet, d'un point de vue cytologique, aucun épaissement pariétal n'est observé. Les peroxydases jouent des rôles multiples au sein de la cellule ; le facteur de transcription *lim1* cible une diversité de gènes pouvant conduire à la synthèse d'autres composés que la lignine ; enfin, un gène codant une 4-coumarate-CoA ligase (Cont0075), enzyme en amont de la peroxydase montre, quant à lui, une baisse d'expression significative au cours de la cinétique chez les deux génotypes.

Les résultats ont mis en évidence qu'un transporteur d'auxine (OC1546) et qu'un gène codant une protéine induite par l'auxine (Cont0125) montrent une hausse d'expression chez C15 au cours de la cinétique et aucune modulation chez K59. Cependant, l'implication de l'auxine dans la réponse du génotype non-embryogène reste à démontrer puisqu'un autre gène codant une protéine induite par les auxines (OC1107) montre, quant à lui, une baisse d'expression significative chez C15.

Malgré une activité métabolique importante et la mise en place d'un nouveau transcriptome, un seul gène codant une protéine ribosomale (OC2103) montre une hausse d'expression au cours de la cinétique chez le génotype C15, qui n'est pas observée chez le génotype K59. Ce résultat suggère que, chez le génotype non embryogène, la mise en place d'un nouveau transcriptome pourrait se faire globalement par l'intermédiaire des mêmes protéines ribosomales que celles initialement présentes dans les cellules.

Nous avons identifié plusieurs gènes codant des protéines impliquées dans la dégradation des protéines montrant une forte hausse d'expression chez le génotype non embryogène. La plupart de ces gènes jouent un rôle dans l'ubiquitination des protéines ou dans la structure du protéasome 26S. Ces gènes codent une protéine interagissant avec une ubiquitine (OC1668_a), une protéine contenant un domaine « ring » (OC1796), une polyubiquitine (OC1110_b), une sous-unité RPT5a de l'ATPase du protéasome 26S (OC1753), une sous-unité 9b régulatrice du protéasome 26S (OC1521), et une « aspartic proteinase » (Cont9003). Les protéines dégradées par l'intermédiaire des protéines citées pourraient servir de substrat pour la synthèse de nouvelles protéines impliquées dans la réponse du génotype C15.

La défense est une classe de fonction fortement représentée par les gènes montrant une hausse d'expression au cours de la cinétique chez le génotype non-embryogène et non observée, ou observée modérément, chez le génotype embryogène. Les réactions de défense semblent se mettre progressivement en place au cours de la cinétique. Ces gènes codent par exemple, une métallothionéine 1 (Cont9039), une peroxydase de classe III, annotée comme étant induite par les bactéries chez le coton (OC1539), une peroxydase 3 (Cont9071), une protéine reliée à la sénescence chez le poirier (Cont5061), une protéine annotée comme impliquée dans la sénescence des feuilles chez le riz (OC1660), une glutathion S-transférase (OC0148) et une ACC oxydase (OC0168), enzyme impliquée dans la voie de synthèse de l'éthylène. Il a été montré que des homologues de tous les gènes cités sont potentiellement impliqués au cours de la sénescence (Buchanan-Wollaston *et al.*, 2003; Kunieda *et al.*, 2005). D'autres gènes potentiellement associés à la défense présentent un intérêt particulier. Un gène code une protéine 14-3-3 (OC0071) et montre une hausse d'expression significative chez C15 au jour 6 de culture et pas de modulation d'expression chez K59. Du fait que les protéines 14-3-3 peuvent avoir un rôle dans les réactions de défense (Roberts, 2003), cette protéine 14-3-3 pourrait être impliquée dans la signalisation de la réponse de défense du génotype non-embryogène. En effet, il est connu que. Un autre gène code une protéine associée à la dormance des bourgeons axillaires chez le pois (Stafstrom *et al.*, 1998) (OC0839).

Un gène codant un précurseur d'un facteur d'initiation de la traduction eIF-5A (OC0052_a) montre une hausse d'expression chez le génotype C15 aux jours 5 et 6, non détectée chez le génotype K59. Il existe 3 isoformes de eIF-5A chez Arabidopsis : une régule la division cellulaire et deux régulent la sénescence (Thompson *et al.*, 2004). Des études menées chez la levure et les mammifères ont montré que eIF-5A est synthétisée sous forme inactive et qu'elle nécessite une activation post-traductionnelle par l'action séquentielle de deux enzymes, la déoxyhypusine synthase (DHS) et la déoxyhypusine hydroxylase (Park *et al.*, 1997). Chez Arabidopsis, les gènes AteIF5-A1 et DHS sont surexprimés au cours de la sénescence des feuilles, et des plantes transformées avec une construction antisens DHS montrent un retard important dans la sénescence des feuilles (Wang *et al.*, 2003). Les gènes eIF-5A ne semblent pas impliqués dans la synthèse de l'ensemble des protéines. En effet, chez la levure, lorsque les deux isoformes de eIF-5A, impliquées dans la sénescence, sont mutées, peu de conséquences au niveau de la cellule sont observées. eIF-5A pourrait permettre le transport, du noyau vers le cytoplasme, d'ARNm spécifiques en lien avec la sénescence et la mort cellulaire (Kang and Hershey, 1994). Ainsi, la protéine eIF-5A codée

par le gène identifié au cours de cette étude pourrait réguler la traduction d'autres gènes observés comme étant potentiellement impliqués dans la sénescence.

Le transport est une classe de fonction particulièrement active dans les cellules sénescences qui évacuent leurs réserves vers d'autres organes et qui tentent d'éliminer les déchets engendrés par la forte activité métabolique de ces cellules (Zimmermann and Zentgraf, 2005). Le transport paraît être fortement représenté par les gènes montrant une hausse d'expression significative et exclusive au génotype C15. Nous trouvons au sein de cette classe fonctionnelle un symporteur anion:sodium (OC1602), un transporteur ABC PDR-like (OC1588), une protéine intrinsèque à la membrane plasmique (OC1030), une « synaptobrevin-related family protein » (Cont0007), un transporteur dicarboxylate/tricarboxylate (OC1626_a) et un transporteur d'auxine (OC1546).

La réaction du génotype C15 aux conditions de culture présente de fortes similitudes avec la sénescence. Au cours des six premiers jours de culture, cette réaction ne semble pas conduire à la mort cellulaire, même si nous remarquons de la nécrose sur le pourtour des explants. Dans certains cas même, les explants sont capables, une fois transférés dans un milieu de Heller, de former des bourgeons après environ 4 semaines de culture. Il faut noter que la sénescence ne conduit pas toujours forcément à la mort cellulaire et peut même être réversible suivant l'intensité de stress appliqué (Thompson *et al.*, 2004).

Les résultats ont mis en évidence qu'un gène (Cont0080), montrant une hausse d'expression exclusive au génotype C15, présente 87% d'identité avec la séquence d'un ADNc codant une protéine de stockage (VSP) de chicorée (Richard-Molard, 1999). Le gène mis en évidence au cours de cette étude est donc probablement un autre gène que celui déjà identifié. On sait, par ailleurs, qu'il existe au moins trois gènes codant des VSP chez la chicorée (Richard-Molard, 1999). Les deux ADNc présentent de fortes homologues avec des gènes codant des protéines PR (Pathogenesis Related). Les VSP pourraient, en plus de leur rôle de protéine de stockage, être impliquées dans la défense. Le gène déjà identifié est fortement exprimé au niveau racinaire et les protéines sont accumulées dans la racine tubérisée lors de la croissance végétative de la plante au cours de la première année. Les VSP pourraient servir de réserve en azote nécessaire à l'élaboration de la hampe florale la seconde année (Limami *et al.*, 1993; Limami *et al.*, 1996). Au niveau foliaire, la protéine est peu abondante, en comparaison des racines. Des analyses par northern blot, ont montré que les transcrits sont détectés dans le bouquet foliaire sénescence à la fin de la phase végétative de la première année, principalement au niveau des nervures (Richard-Molard, 1999). La

localisation des transcrits de VSP au niveau du système vasculaire suggère que les VSP produites au niveau des feuilles sénescences pourraient être remobilisées vers les racines à la fin de la phase végétative. Ce résultat montre, à nouveau, que la réponse du génotype non-embryogène présente de fortes similitudes avec la mise en place de la sénescence.

Les résultats ont mis en évidence qu'un gène codant une protéine ribosomale 40S S6 potentiellement impliquée dans la voie de signalisation via la protéine TOR (Target Of Rapamycin), montre une baisse d'expression significative chez C15 dès le jour 3 de culture non observée chez K59. La protéine TOR contrôle de nombreux processus qui contribuent à la croissance des cellules tels que la traduction, la transcription, la biogenèse des ribosomes ou l'organisation de l'actine (Schmelzle and Hall, 2000). Il a par ailleurs été montré que l'inactivation du gène AtTOR d'*Arabidopsis* par l'insertion d'un T-DNA conduisait à une embryogenèse qui est arrêtée très précocement (Menand *et al.*, 2002). Les explants du génotype C15 pourraient ainsi présenter un blocage de certaines fonctions liées au développement. Cela pourrait permettre alors, la mise en place du programme de sénescence que nous observons chez le génotype non-embryogène.

- Conclusions et perspectives -

Ce travail avait pour but, l'étude des phases précoces du processus d'embryogenèse somatique chez la chicorée. Ce type d'analyse nécessite des témoins afin d'éliminer, au maximum, les gènes qui ne sont pas directement impliqués dans le processus. Certains gènes peuvent être, par exemple, induits par les stress occasionnés par les conditions de culture et ne pas participer à la réponse étudiée. Chez le soja, au cours de l'étude de la formation d'embryons somatiques obtenus à partir de la face supérieure des cotylédons, c'est la face inférieure de cet organe, ne réagissant pas aux conditions d'induction, qui a été utilisée comme témoin négatif du processus de développement (Thibaud-Nissen *et al.*, 2003).

Au cours de cette étude, nous avons disposé du génotype C15 non-embryogène obtenu à partir d'une autofécondation du génotype K59. Ce génotype non-embryogène a été utilisé comme témoin négatif de l'embryogenèse somatique à diverses étapes de ce travail. Au cours de la fabrication des banques soustractives, il a permis, en théorie, d'éliminer un maximum de transcrits correspondants aux gènes qui ne sont pas impliqués directement dans le processus. Pour l'analyse transcriptomique par microarray, le génotype non-embryogène a permis d'identifier les gènes présentant la même réponse chez les deux génotypes. En effet, sur les 373 gènes conservés parce qu'ils présentaient des modulations d'expression chez l'un ou l'autre des génotypes au cours de la cinétique d'étude, nous avons identifié 173 gènes montrant, chez les deux génotypes, le même profil d'expression. En l'absence du témoin non-embryogène C15, nous n'aurions pas pu distinguer les gènes potentiellement impliqués au cours de l'embryogenèse somatique, des gènes associés à une réponse généraliste des explants mis en culture.

Nous avons montré dans les parties I et II que les banques soustractives obtenues à partir des explants des génotypes K59 et C15 sont effectivement normalisées et soustraites. Nous nous sommes alors appuyés sur le fait que les banques présentent relativement peu de redondance et que l'analyse *in silico* ne permet pas la prédiction de gènes différenciellement exprimés. D'autre part, le fait que seulement 14% des gènes soient représentés par des ESTs provenant des deux banques, suggère que les banques sont effectivement soustraites. Les résultats obtenus au cours de la partie III confirment une certaine efficacité de soustraction, puisque sur 1098 gènes analysés par microarray, 200 (18%) montrent des profils d'expression différents chez les deux génotypes, avec des modulations d'expression chez au moins un des deux génotypes. Ces 200 gènes sont potentiellement impliqués dans la réponse de chacun des deux génotypes aux conditions d'induction de l'embryogenèse somatique. A titre de comparaison, chez le soja, seulement 6% des gènes représentés par des ADNc obtenus à partir

de banques non soustraites ont montré une modulation au cours d'une cinétique en conditions d'embryogenèse somatique (Thibaud-Nissen *et al.*, 2003) et chez le zinnia, lors de la transdifférenciation des cellules du mésophylle en cellules de xylème, 6,5% des gènes représentés par des ADNc issus également de banques non soustraites ont montré une expression différentielle (Demura *et al.*, 2002). Chez le peuplier, lors d'une étude portant sur les modifications engendrées par un agent protecteur contre des herbicides, environ 5% des gènes représentés par des ADNc issus de banques SSH ont montré une expression différentielle (Rishi *et al.*, 2004). La proportion de gènes différentiellement exprimés au cours de notre étude pourrait paraître importante. Dans les études citées précédemment, la mise en évidence de gènes exprimés différentiellement s'est faite uniquement sur la base d'un seuil de ratio d'expression. Or, les analyses de variance, réalisées au cours de cette étude, ont montré qu'un ratio d'expression élevé n'était pas forcément systématique pour l'obtention d'un effet biologique significatif.

D'autres arguments montrent également la robustesse des résultats mis en évidence au cours de cette étude. Ainsi, alors que les cultures en conditions d'embryogenèse somatique se déroulent à l'obscurité, les trois gènes liés à la photosynthèse et analysés par microarray, ont montré une baisse d'expression significative chez les deux génotypes avec des quantités de transcrits assez semblables. De plus, nous avons pu faire le lien, pour le génotype embryogène, entre les observations cytologiques et certains gènes exprimés différentiellement. Nous avons, par exemple, détecté de nombreux gènes liés aux divisions cellulaires à l'instant où l'on observe au sein des explants les premières divisions des cellules embryogènes. Par ailleurs, l'analyse par microarray semble très robuste. En effet, l'ensemble des gènes identifiés comme différentiellement exprimés au jour 4 de culture en conditions d'embryogenèse somatique et testés par PCR en temps réel ont pu être validés.

Nous avons vu, à plusieurs reprises, que les gènes préférentiellement exprimés chez le génotype embryogène sont plus difficiles à détecter que ceux préférentiellement exprimés chez le génotype non embryogène, surtout par des techniques basées sur des hybridations comme les microarrays. Nous avons proposé une hypothèse basée sur le fait que les cellules qui participent à la formation d'embryons somatiques ne sont qu'une minorité au sein des explants du génotype K59. Ainsi, les transcrits des gènes exprimés dans les cellules qui entrent dans le processus d'embryogenèse somatique, et peut-être, dans les cellules environnantes participant à la mise en place du processus, doivent être faiblement représentés au sein de l'ensemble des transcriptomes de l'explant. Cette faible représentation des

transcrits pourrait expliquer, en effet, pourquoi les analyses transcriptomiques menées au jour 4 de culture ont mis en évidence davantage de gènes préférentiellement exprimés chez le génotype C15 que chez le génotype K59. De même, l'étude en cinétique a montré plus de gènes présentant une modulation d'expression par rapport au jour 0, chez le génotype non-embryogène que chez le génotype embryogène. Cette hypothèse pourrait aussi expliquer le fait que la moyenne des ratios d'expression des gènes préférentiellement exprimés chez K59, ou montrant une modulation d'expression chez ce génotype, est plus faible que celle observée pour le génotype C15. Il est certain que la sensibilité modérée des microarrays, pour les gènes représentés par des transcrits peu abondants, entraîne une sous-estimation du nombre de gènes exprimés différemment. Par exemple, l'étude par PCR en temps réel, menée au jour 4 de la cinétique, a mis en évidence une expression préférentielle du gène codant une protéine de développement Argonaute1 (AGO1) chez le génotype K59 alors que l'analyse par microarray, sur ce même point de la cinétique, n'avait pas détecté d'expression différentielle pour ce gène.

L'étude en cinétique, menée sur 7 points et pour deux génotypes, a permis la mise en évidence de nombreux gènes impliqués dans les réponses des génotypes aux conditions de culture appliquées aux explants foliaires. Parmi les gènes déjà reliés à l'embryogenèse somatique et mis en évidence au cours de cette étude, citons les gènes codant un facteur de transcription Scarecrow chez le coton et la sapinette blanche (Stasolla *et al.*, 2003; Zeng *et al.*, 2006), une AGP chez la carotte (Baldwin *et al.*, 2001), une histone H2A et une protéine ribosomale 40S S11 chez le soja (Thibaud-Nissen *et al.*, 2003). On remarque surtout, qu'aucun des gènes identifiés par les études antérieures au laboratoire et codant une hémoglobine non symbiotique (Hendriks *et al.*, 1998), une β 1-3 glucanase (Helleboid *et al.*, 2000a), une glutathion S-transférase (Galland *et al.*, 2001) et une protéine RAB fixatrice de GTP (Randoux *et al.*, 2002) n'a été identifié au cours de cette étude. Toutefois, nous avons mis en évidence quelques gènes des mêmes familles que ceux déjà isolés au laboratoire (β 1-3 glucanase, GST, protéine RAB). Le fait de ne pas avoir identifié les gènes déjà connus est probablement dû au nombre limité de séquences dans nos banques, et/ou à l'utilisation d'un matériel végétal et de techniques de mise en évidence différents. Nous remarquons également que peu de facteurs de transcription ont été identifiés au cours de cette étude dans les réponses de l'un ou l'autre des génotypes, ceci sûrement en raison d'un faible nombre de transcrits représentant, en général, ce type de gènes.

Les gènes mis en évidence au cours de cette étude ont permis de construire un modèle de réponse du génotype embryogène aux conditions d'induction de l'embryogenèse somatique (Figure 45).

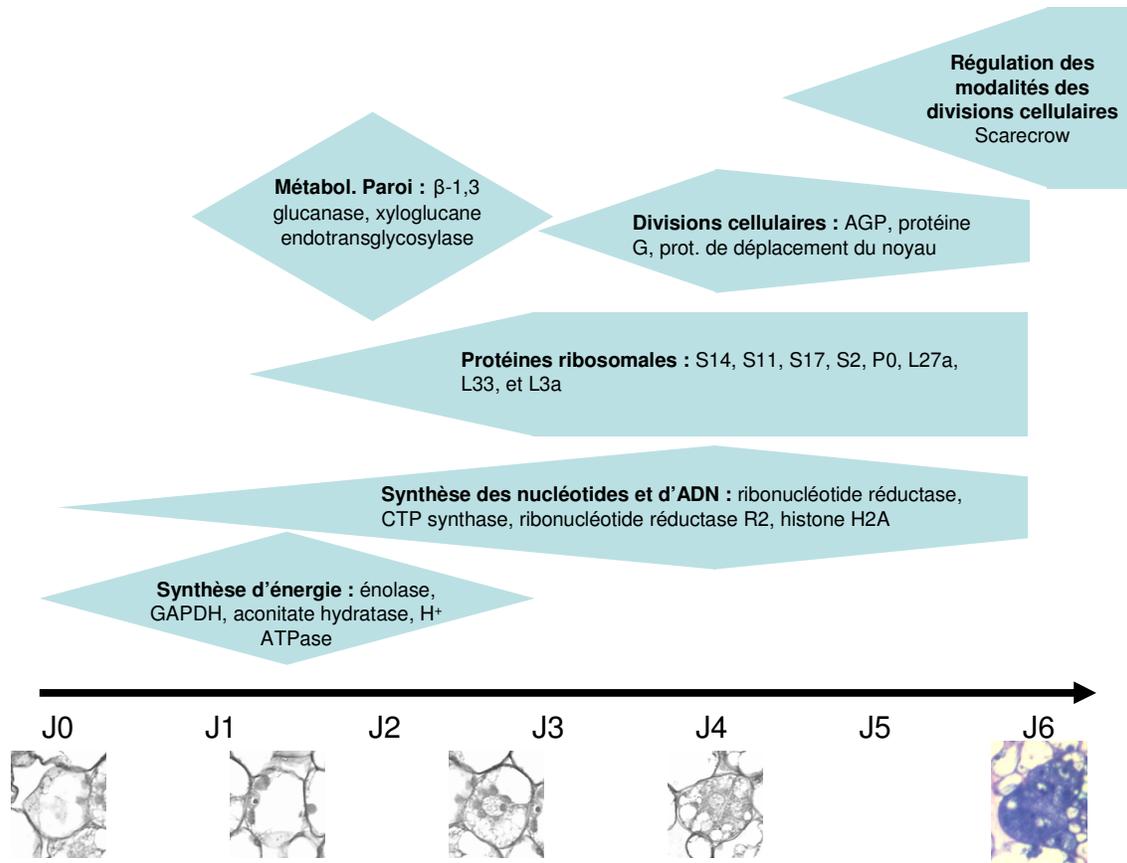


Figure 45 : Modèle synthétique de la réponse du génotype embryogène

Du fait de la hausse d'expression de gènes liés à la glycolyse et au cycle de Krebs, dès les premiers jours de culture, le saccharose du milieu de culture pourrait être utilisé pour la production d'énergie. Cette énergie, qui pourrait également être produite par l'oxydation d'acides gras, pourrait servir à la transdifférenciation et aux divisions cellulaires. D'autres types de fonctions biologiques semblent également conduire à la division cellulaire. En effet, nous supposons, à la vue des gènes différenciellement exprimés, que très précocement, un remaniement pariétal ainsi qu'une synthèse active d'ADN pourraient avoir lieu. De plus, l'expression de nombreux gènes codant des protéines ribosomales pourrait participer à la synthèse de nouveaux ribosomes et à la mise en place des nouveaux transcriptomes. Plusieurs autres gènes semblent être également en lien avec la division cellulaire ou la signalisation pour la division cellulaire. Citons notamment des gènes codant une protéine potentielle de

déplacement du noyau, une protéine G et une AGP. A cela s'ajoute d'autres gènes mis en évidence par l'analyse transcriptomique menée au jour 4, mais ne montrant pas de modulation d'expression significative selon l'étude en cinétique. Ainsi, nous pourrions ajouter l'implication d'un gène codant une protéine CDC48 dans la division des cellules. Ce gène est déjà connu pour être impliqué dans l'embryogenèse somatique chez le soja où il participerait à la formation des fuseaux de division (Thibaud-Nissen *et al.*, 2003). Enfin, l'expression d'un gène codant un facteur de transcription de type Scarecrow, qui pourrait être impliqué dans la régulation des modalités des divisions cellulaires, est détectée après l'apparition des premières divisions dans les explants du génotype embryogène. Scarecrow pourrait empêcher le développement anarchique des embryons.

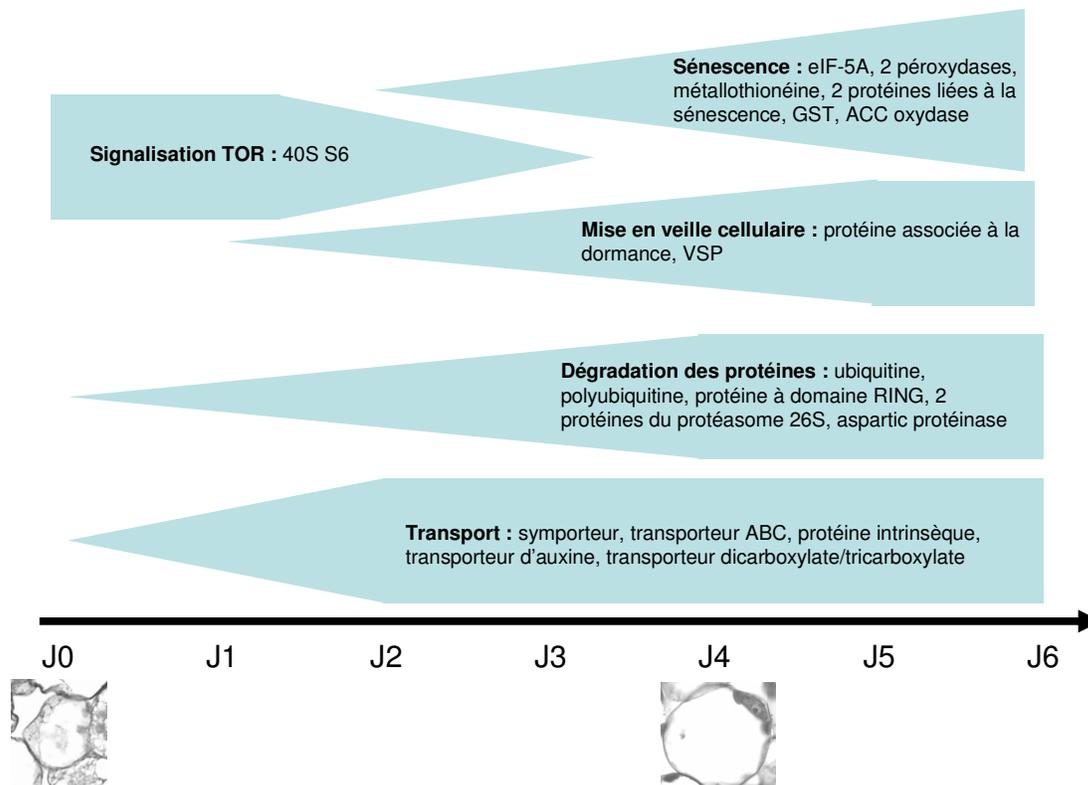


Figure 46 : modèle synthétique de la réponse du génotype non-embryogène

De façon générale, la réponse du génotype non embryogène est très différente de celle observée pour le génotype embryogène (Figure 46). Ici, la plupart des fonctions mises en œuvre, semblent conduire les cellules à une réponse qui ressemble fortement à la sénescence. On observe l'expression de nombreux gènes en relation avec le transport, la mise en réserve, la dégradation des protéines ainsi qu'une expression croissante au cours de la cinétique de

gènes potentiellement liés à la détoxification et à la sénescence. On observe une baisse d'expression d'un gène potentiellement impliqué dans la signalisation via la protéine TOR qui régule, entre autres, la transcription, la traduction et l'élongation cellulaire. Le blocage de ces fonctions liées au développement pourrait permettre l'installation du programme de sénescence au sein des explants du génotype non-embryogène.

Il sera important de vérifier, par la technique de PCR en temps réel, un certain nombre des résultats obtenus par microarray, afin de valider les modèles exposés. Nous avons déjà validé, par PCR en temps réel, le dispositif d'analyses de microarray par une étude comparative menée au jour 4 de culture. Il sera nécessaire de vérifier l'expression en cinétique de certains gènes, en particulier ceux montrant une expression transitoire. Il sera aussi possible d'observer l'expression de gènes, liés aux fonctions biologiques mises en évidence comme étant impliquées dans la réponse des génotypes, mais non déposés sur les lames de microarray.

Dans le but de pallier la puissance modérée des microarrays, nous pourrions envisager d'utiliser la microdissection au laser (LCM), pour sélectionner au sein des explants du génotype K59 les cellules participant à l'embryogenèse somatique. Cette approche pourrait, par ailleurs, conduire à l'identification de gènes spécifiquement exprimés dans les cellules en cours de réactivation et dans les cellules embryogènes. La LCM a déjà été utilisée, chez les plantes, pour sélectionner quelques cellules au cours d'une étude portant sur la défense du soja contre l'attaque d'un pathogène et a permis une analyse transcriptomique fine (Klink *et al.*, 2005).

Dans le but d'identifier des gènes clés du processus d'embryogenèse somatique, il sera nécessaire de distinguer les gènes spécifiquement impliqués dans l'embryogenèse somatique de ceux qui pourraient également être impliqués dans d'autres processus de développement faisant intervenir, par exemple, une phase de divisions cellulaires actives. Pour cela, il sera possible de réaliser des études transcriptomiques à partir de cellules issues de suspensions cellulaires ou de cals. Ces études transcriptomiques pourront être comparées à celle réalisées, ici, au cours de la formation d'embryons somatiques. Le schéma expérimental de microarray, mis en œuvre au cours de cette étude, permettra d'ajouter de nouvelles conditions biologiques qui pourront être facilement comparées à toutes les conditions déjà testées. De la même façon, il serait intéressant de comparer les transcriptomes observés au cours de l'embryogenèse somatique à ceux observés au cours du processus d'organogenèse. Cette comparaison

permettrait d'identifier des gènes potentiellement impliqués dans les deux processus et des gènes spécifiques.

Une autre perspective intéressante sera d'identifier, par hybridation *in situ*, dans quels types cellulaires, au sein des explants du génotype K59, s'expriment certains gènes choisis. Par exemple, il serait intéressant de savoir si le gène codant une AGP est exprimé dans les cellules qui entrent dans le processus de transdifférenciation ou alors s'il est exprimé dans les cellules environnantes qui pourraient participer à l'embryogenèse somatique par l'envoi de signaux. L'hypothèse que Scarecrow régule les modalités des divisions cellulaires lors de la formation des embryons somatiques pourrait être également vérifiée par hybridation *in situ*, afin de préciser l'expression spatio-temporelle de ce gène.

En parallèle, les outils de la génétique pourront apporter une perspective importante à ce travail. En effet, nous disposons au laboratoire d'une carte génétique obtenue à partir de la descendance du génotype K59 et d'un génotype peu embryogène (K28), issu lui aussi de la population « Koospol ». Par ailleurs, la capacité des individus de la descendance, à former des embryons somatiques, est en cours de caractérisation. Ce travail permettra de localiser sur la carte génétique de la chicorée, des QTL liés à la formation des embryons somatiques. Il sera intéressant de rechercher du polymorphisme au sein des gènes identifiés au cours de notre étude comme étant impliqués dans les réponses de chacun des deux génotypes. Cela permettra d'obtenir des marqueurs moléculaires qui seront ajoutés à la carte. Il faudra ensuite vérifier si ces marqueurs seront, ou non, co-localisés avec les QTL d'embryogenèse somatique. Nous pourrons alors conclure, que les gènes, correspondant aux marqueurs qui ne seront pas co-localisés avec les QTL, ne sont pas des gènes régulateurs de l'embryogenèse somatique. Cependant, les gènes colocalisant avec des QTL ne seront pas, pour autant, des gènes régulateurs, une grande distance physique pouvant séparer les gènes marqueurs des gènes impliqués dans l'expression du caractère.

Le recoupement des informations génétiques avec celles obtenues par les comparaisons des transcriptomes observés au cours de différents processus faisant intervenir des divisions cellulaires (embryogenèse somatique, organogenèse, callogenèse, culture cellulaire...) permettra d'établir une liste restreinte de gènes candidats potentiellement spécifiques, voire régulateurs, du processus d'embryogenèse somatique.

Il sera alors intéressant de préciser la fonction et de déterminer le rôle ainsi que le mode d'action de ces gènes candidats. Différentes approches pourront alors être explorées

Conclusions et perspectives

suivant les gènes à caractériser : recherches de gènes homologues mutés dans les banques de mutants d'Arabidopsis, complémentation fonctionnelle dans la levure, transformation génétique de la chicorée.

- Références bibliographiques -

- (2000) The Arabidopsis Genome Initiative; Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- (2005) International Rice Genome Sequencing Project; The map-based sequence of the rice genome. *Nature*, 436, 793–800.
- Alignan, M., Hewezi, T., Petitprez, M., Dechamp-Guillaume, G. and Gentzbittel, L. (2006) A cDNA microarray approach to decipher sunflower (*Helianthus annuus*) responses to the necrotrophic fungus *Phoma macdonaldii*. *New Phytol*, 170, 523-536.
- Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A*, 74, 5350-5354.
- Andersson, A., Keskitalo, J., Sjodin, A., Bhalerao, R., Sterky, F., Wissel, K., Tandre, K., Aspeborg, H., Moyle, R., Ohmiya, Y., Brunner, A., Gustafsson, P., Karlsson, J., Lundeberg, J., Nilsson, O., Sandberg, G., Strauss, S., Sundberg, B., Uhlen, M., Jansson, S. and Nilsson, P. (2004) A transcriptional timetable of autumn senescence. *Genome Biol*, 5, R24. Epub 2004 Mar 2010.
- Armstrong, N.J. and van de Wiel, M.A. (2004) Microarray data analysis: from hypotheses to conclusions using gene expression data. *Cell Oncol*, 26, 279-290.
- Audic, S. and Claverie, J.M. (1997) The significance of digital gene expression profiles. *Genome Res*, 7, 986-995.
- Bachem, C.W., van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M. and Visser, R.G. (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J*, 9, 745-753.
- Baldwin, T.C., Domingo, C., Schindler, T., Seetharaman, G., Stacey, N. and Roberts, K. (2001) DcAGP1, a secreted arabinogalactan protein, is related to a family of basic proline-rich proteins. *Plant Mol Biol*, 45, 421-435.
- Barakat, A., Szick-Miranda, K., Chang, I.F., Guyot, R., Blanc, G., Cooke, R., Delseny, M. and Bailey-Serres, J. (2001) The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome. *Plant Physiol*, 127, 398-415.

Références bibliographiques

- Bassani, M., Neumann, P.M. and Gepstein, S. (2004) Differential expression profiles of growth-related genes in the elongation zone of maize primary roots. *Plant Mol Biol*, 56, 367-380.
- Bey, M., Stuber, K., Fellenberg, K., Schwarz-Sommer, Z., Sommer, H., Saedler, H. and Zachgo, S. (2004) Characterization of *antirrhinum* petal development and identification of target genes of the class B MADS box gene DEFICIENS. *Plant Cell*, 16, 3197-3215.
- Blervacq, A., Dubois, T., Dubois, J. and Vasseur, J. (1995) First divisions of somatic embryogenic cells in *Cichorium* hybrid "474". *Protoplasma*, 186, 163-168.
- Bögre, L., Stefanov, I., Ábrahám, M., Somogyi, I. and Dudits, D. (1990) Differences in responses to 2,4-D-dichlorophenoxy acetic acid (2,4-D) treatment between embryogenic and non-embryogenic lines of alfalfa. In *Progress in Plant Cellular and Molecular Biology*: J. eds. Kluwer Academic Publishers, Dordrecht, pp. 427-436.
- Boyer, C., Hilbert, J.L. and Vasseur, J. (1993) Embryogenesis-related protein synthesis and accumulation during early acquisition of somatic embryogenesis competence in *Cichorium*. *Plant Science*, 93, 41-53.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J. and Vingron, M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*, 29, 365-371.
- Brinker, M., van Zyl, L., Liu, W., Craig, D., Sederoff, R.R., Clapham, D.H. and von Arnold, S. (2004) Microarray analyses of gene expression during adventitious root development in *Pinus contorta*. *Plant Physiol*, 135, 1526-1539. Epub 2004 Jul 1529.
- Buchanan-Wollaston, V., Earl, S., Harrison, E., Mathas, E., Navabpour, S., Page, T. and Pink, D. (2003) The molecular analysis of leaf senescence - a genomics approach. *Plant Biotechnology Journal*, 1, 3-22.
- Budhreja, V., Spitznagel, E., Schaiff, W.T. and Sadovsky, Y. (2003) Incorporation of gene-specific variability improves expression analysis using high-density DNA microarrays. *BMC Biol*, 1, 1.

Références bibliographiques

- Casati, P. and Walbot, V. (2004) Rapid transcriptome responses of maize (*Zea mays*) to UV-B in irradiated and shielded tissues. *Genome Biol*, 5, R16. Epub 2004 Mar 2001.
- Chaboute, M.E., Combettes, B., Clement, B., Gigot, C. and Philipps, G. (1998) Molecular characterization of tobacco ribonucleotide reductase RNR1 and RNR2 cDNAs and cell cycle-regulated expression in synchronized plant cells. *Plant Mol Biol*, 38, 797-806.
- Chapman, A., Blervacq, A.S., Vasseur, J. and Hilbert, J.L. (2000) Arabinogalactan-proteins in *Cichorium* somatic embryogenesis: effect of beta-glucosyl Yariv reagent and epitope localisation during embryo development. *Planta*, 211, 305-314.
- Charrière, F., Sotta, B., Miginiac, E. and Hahne, G. (1999) Induction of adventitious shoots or somatic embryos on in vitro cultured zygotic embryos of *Helianthus annuus*: Variation of endogenous hormone levels. *Plant Physiol. Biochem*, 37, 751-757.
- Ciannamea, S., Busscher-Lange, J., de Folter, S., Angenent, G.C. and Immink, R.G. (2006) Characterization of the vernalization response in *Lolium perenne* by a cDNA microarray approach. *Plant Cell Physiol*, 47, 481-492.
- Clarke, B., Lambrecht, M. and Rhee, S.Y. (2003) Arabidopsis genomic information for interpreting wheat EST sequences. *Funct Integr Genomics*, 3, 33-38.
- Clarke, J.D. and Zhu, T. (2006) Microarray analysis of the transcriptome as a stepping stone towards understanding biological systems: practical considerations and perspectives. *Plant J*, 45, 630-650.
- Cresti, M., Blackmore, S. and van Went, J.L. (1992) *Atlas of sexual reproduction in flowering plants*. Berlin and New York: Springer-Verlag.
- Czechowski, T., Bari, R.P., Stitt, M., Scheible, W.R. and Udvardi, M.K. (2004) Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J*, 38, 366-379.
- Da Silva, S. (2004) Identification des gènes CiSTM et CiCLV1 chez *Cichorium intybus*. Implication potentielle dans les modifications d'identité cellulaire au cours des phases précoces de l'embryogenèse somatique et du développement des embryons zygotiques et de la graine. Thèse de l'Université de Lille1.
- Davletova, S., Meszaros, T., Miskolczi, P., Oberschall, A., Torok, K., Magyar, Z., Dudits, D. and Deak, M. (2001) Auxin and heat shock activation of a novel member of the

Références bibliographiques

- calmodulin like domain protein kinase gene family in cultured alfalfa cells. *J Exp Bot*, 52, 215-221.
- de Vries, S.C., Booij, H., Janssens, R., Vogels, R., Saris, L., Lo Schiavo, F., Terzi, M. and Van Kammen, A. (1988) Carrot somatic embryogenesis depends on the phytohormone-controlled presence of correctly glycosylated extracellular protein. *Genes Dev*, 2, 462-476.
- Decout, E., Dubois, T., Guerida, M., Dubois, J., Audran, J.C. and Vasseur, J. (1994) Role of temperature as triggering signal for organogenesis or somatic embryogenesis in wounded leaves of chicory cultured *in vitro*. *J Exp Bot*, 45, 1859-1865.
- Degenhardt, J., Al-Masri, A.N., Kurkcuoglu, S., Szankowski, I. and Gau, A.E. (2005) Characterization by suppression subtractive hybridization of transcripts that are differentially expressed in leaves of apple scab-resistant and susceptible cultivars of *Malus domestica*. *Mol Genet Genomics*, 273, 326-335.
- Demura, T., Tashiro, G., Horiguchi, G., Kishimoto, N., Kubo, M., Matsuoka, N., Minami, A., Nagata-Hiwatashi, M., Nakamura, K., Okamura, Y., Sassa, N., Suzuki, S., Yazaki, J., Kikuchi, S. and Fukuda, H. (2002) Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells. *Proc Natl Acad Sci U S A*, 99, 15794-15799.
- Di Laurenzio, L., Wysocka-Diller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M.G., Feldmann, K.A. and Benfey, P.N. (1996) The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. *Cell*, 86, 423-433.
- Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. and Siebert, P.D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A*, 93, 6025-6030.
- Ding, C. and Cantor, C.R. (2004) Quantitative analysis of nucleic acids--the last few years of progress. *J Biochem Mol Biol*, 37, 1-10.
- Dong, J.Z. and Dunstan, D.I. (1997) Endochitinase and beta-1,3-glucanase genes are developmentally regulated during somatic embryogenesis in *Picea glauca*. *Planta*, 201, 189-194.

Références bibliographiques

- dos Santos, A.V.P., Cutter, E.G. and Davey, M.R. (1983) Origin and development of somatic embryos in *Medicago sativa* L. (alfalfa). *Protoplasma*, 117, 107-115.
- Dresselhaus, T., Cordts, S., Heuer, S., Sauter, M., Lorz, H. and Kranz, E. (1999) Novel ribosomal genes from maize are differentially expressed in the zygotic and somatic cell cycles. *Mol Gen Genet*, 261, 416-427.
- Duban, M. (2004) Clonage et caractérisation de deux gènes codant des récepteurs transmembranaires à activité kinase chez *Cichorium intybus* L. Expression au cours des phases précoces de l'embryogenèse somatique et du développement de la graine. Thèse de l'Université de Lille1.
- Dubois, T., Guerida, M., Dubois, J. and Vasseur, J. (1990) Direct somatic embryogenesis in roots of *Cichorium*: is callose an early marker? *Ann Bot*, 65, 539-545.
- Dubois, T., Guerida, M., Dubois, J. and Vasseur, J. (1991) Direct somatic embryogenesis in leaves of cichorium. An histological and SEM study of early stages. *Protoplasma*, 162, 120-127.
- Dudits, D., Bögre, L. and Györgyey, J. (1991) Molecular and cellular approaches to the analysis of plant embryo development from somatic cells in vitro. *J Cell Sci*, 99, 475-484.
- Dudits, D., Györgyey, J., Bögre, L. and Bakó, L. (1995) Molecular biology of somatic embryogenesis. In *InVitro Embryogenesis in Plants*. Dordrecht: Kluwer AP, pp. 267-308.
- Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. and Trent, J.M. (1999) Expression profiling using cDNA microarrays. *Nat Genet*, 21, 10-14.
- Endo, M., Matsubara, H., Kokubun, T., Masuko, H., Takahata, Y., Tsuchiya, T., Fukuda, H., Demura, T. and Watanabe, M. (2002) The advantages of cDNA microarray as an effective tool for identification of reproductive organ-specific genes in a model legume, *Lotus japonicus*. *FEBS Lett*, 514, 229-237.
- Faivre-Rampant, O., Cardle, L., Marshall, D., Viola, R. and Taylor, M.A. (2004) Changes in gene expression during meristem activation processes in *Solanum tuberosum* with a focus on the regulation of an auxin response factor gene. *J Exp Bot*, 55, 613-622.
- Feher, A., Pasternak, T.P. and Dudits, D. (2003) Transition of somatic plant cells to an embryogenic state. *Plant Cell Tissue and Organ Culture*, 201-228.

Références bibliographiques

- Feiler, H.S., Desprez, T., Santoni, V., Kronenberger, J., Caboche, M. and Traas, J. (1995) The higher plant *Arabidopsis thaliana* encodes a functional CDC48 homologue which is highly expressed in dividing and expanding cells. *Embo J*, 14, 5626-5637.
- Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T. and Solas, D. (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science*, 251, 767-773.
- Footitt, S., Ingouff, M., Clapham, D. and von Arnold, S. (2003) Expression of the viviparous 1 (Pavp1) and p34cdc2 protein kinase (cdc2Pa) genes during somatic embryogenesis in Norway spruce (*Picea abies* [L.] Karst). *J Exp Bot*, 54, 1711-1719.
- French, B.T., Maul, H.M. and Maul, G.G. (1986) Screening cDNA expression libraries with monoclonal and polyclonal antibodies using an amplified biotin-avidin-peroxidase technique. *Anal Biochem*, 156, 417-423.
- Galland, R., Randoux, B., Vasseur, J. and Hilbert, J.L. (2001) A glutathione S-transferase cDNA identified by mRNA differential display is upregulated during somatic embryogenesis in *Cichorium*. *Biochim Biophys Acta*, 1522, 212-216.
- Giles, P.J. and Kipling, D. (2003) Normality of oligonucleotide microarray data and implications for parametric statistical analyses. *Bioinformatics*, 19, 2254-2262.
- Giroux, R.W. and Pauls, K.P. (1997) Characterization of somatic embryogenesis-related cDNAs from alfalfa (*Medicago sativa* L.). *Plant Mol Biol*, 33, 393-404.
- Goda, H., Shimada, Y., Asami, T., Fujioka, S. and Yoshida, S. (2002) Microarray analysis of brassinosteroid-regulated genes in *Arabidopsis*. *Plant Physiol*, 130, 1319-1334.
- Green, P. (2002) Documentation for phrap, University of Washington. <http://www.phrap.org/>.
- Grosset, J., Marty, I., Chartier, Y. and Meyer, Y. (1990) mRNAs newly synthesized by tobacco mesophyll protoplasts are wound-inducible. *Plant Mol Biol*, 15, 485-496.
- Gulick, P.J., Drouin, S., Yu, Z., Danyluk, J., Poisson, G., Monroy, A.F. and Sarhan, F. (2005) Transcriptome comparison of winter and spring wheat responding to low temperature. *Genome*, 48, 913-923.
- Hata, S., Kouchi, H., Suzuka, I. and Ishii, T. (1991) Isolation and characterization of cDNA clones for plant cyclins. *Embo J*, 10, 2681-2688.

Références bibliographiques

- Helleboid, S., Chapman, A., Hendriks, T., Inze, D., Vasseur, J. and Hilbert, J.L. (2000a) Cloning of beta-1,3-glucanases expressed during *Cichorium* somatic embryogenesis. *Plant Mol Biol*, 42, 377-386.
- Helleboid, S., Hendriks, T., Bauw, G., Inze, D., Vasseur, J. and Hilbert, J.L. (2000b) Three major somatic embryogenesis related proteins in *Cichorium* identified as PR proteins. *J Exp Bot*, 51, 1189-1200.
- Hendriks, T., Scheer, I., Quillet, M.C., Randoux, B., Delbreil, B., Vasseur, J. and Hilbert, J.L. (1998) A nonsymbiotic hemoglobin gene is expressed during somatic embryogenesis in *Cichorium*. *Biochim Biophys Acta*, 1443, 193-197.
- Herbette, S., Taconnat, L., Hugouvieux, V., Piette, L., Magniette, M.L., Cuine, S., Auroy, P., Richaud, P., Forestier, C., Bourguignon, J., Renou, J.P., Vavasseur, A. and Leonhardt, N. (2006) Genome-wide transcriptome profiling of the early cadmium response of Arabidopsis roots and shoots. *Biochimie*.
- Heuer, S., Hansen, S., Bantin, J., Brettschneider, R., Kranz, E., Lorz, H. and Dresselhaus, T. (2001) The maize MADS box gene ZmMADS3 affects node number and spikelet development and is co-expressed with ZmMADS1 during flower development, in egg cells, and early embryogenesis. *Plant Physiol*, 127, 33-45.
- Hilbert, J.L., Dubois, T. and Vasseur, J. (1992) Detection of embryogenesis-related proteins during somatic embryo formation in *Cichorium*. *Plant Phys Biochem* 30, 733-741.
- Hillier, L.D., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Marra, M. and et al. (1996) Generation and analysis of 280,000 human expressed sequence tags. *Genome Res*, 6, 807-828.
- Hirt, H., Pay, A., Gyorgyey, J., Bako, L., Nemeth, K., Bogre, L., Schweyen, R.J., Heberle-Bors, E. and Dudits, D. (1991) Complementation of a yeast cell cycle mutant by an alfalfa cDNA encoding a protein kinase homologous to p34cdc2. *Proc Natl Acad Sci U S A*, 88, 1636-1640.
- Hovens, C.M. and Wilks, A.F. (1989) Rapid screening of highly complex cDNA libraries using the polymerase chain reaction. *Nucleic Acids Res*, 17, 4415-4416.

Références bibliographiques

- Huang, X. and Madan, A. (1999) CAP3: A DNA sequence assembly program. *Genome Res*, 9, 868-877.
- Ivanova, A., Velcheva, M., Denchev, P., Atanassov, A. and Van Onckelen, H.A. (1994) Endogenous hormone levels during direct somatic embryogenesis in *Medicago falkata*. *Physiol Plant*, 92, 85–89.
- Jimenez, V.M. and Bangerth, F. (2001) In vitro culture and endogenous hormone levels in immature zygotic embryos, endosperm and callus cultures of normal and high-lysine barley genotypes. *J Appl Bot*, 75, 1-7.
- Jurgens, G., Grebe, M. and Steinmann, T. (1997) Establishment of cell polarity during early plant development. *Curr Opin Cell Biol*, 9, 849-852.
- Kader, J.C. (1996) Lipid-Transfer Proteins in Plants. *Annu Rev Plant Physiol Plant Mol Biol*, 47, 627-654.
- Kamiya, N., Itoh, J., Morikami, A., Nagato, Y. and Matsuoka, M. (2003) The SCARECROW gene's role in asymmetric cell divisions in rice plants. *Plant J*, 36, 45-54.
- Kang, H.A. and Hershey, J.W. (1994) Effect of initiation factor eIF-5A depletion on protein synthesis and proliferation of *Saccharomyces cerevisiae*. *J Biol Chem*, 269, 3934-3940.
- Kapros, T., Bogre, L., Nemeth, K., Bako, L., Gyorgyey, J., Wu, S.C. and Dudits, D. (1992) Differential Expression of Histone H3 Gene Variants during Cell Cycle and Somatic Embryogenesis in Alfalfa. *Plant Physiol*, 98, 621-625.
- Kawahara, R., Sunabori, S., Fukuda, H. and Komamine, A. (1992) A gene expressed preferentially in the globular stage of somatic embryogenesis encodes elongation-factor 1 alpha in carrot. *Eur J Biochem*, 209, 157-162.
- Kawaoka, A., Kaothien, P., Yoshida, K., Endo, S., Yamada, K. and Ebinuma, H. (2000) Functional analysis of tobacco LIM protein Ntlm1 involved in lignin biosynthesis. *Plant J*, 22, 289-301.
- Kerr, M.K. (2003a) Design considerations for efficient and effective microarray studies. *Biometrics*, 59, 822-828.

- Kerr, M.K. (2003b) Linear models for microarray data analysis: hidden similarities and differences. *J Comput Biol*, 10, 891-901.
- Kerr, M.K. and Churchill, G.A. (2001a) Experimental design for gene expression microarrays. *Biostatistics*, 2, 183-201.
- Kerr, M.K. and Churchill, G.A. (2001b) Statistical design and the analysis of gene expression microarray data. *Genet Res*, 77, 123-128.
- Kerr, M.K., Martin, M. and Churchill, G.A. (2000) Analysis of variance for gene expression microarray data. *J Comput Biol*, 7, 819-837.
- Kimura, M., Yamamoto, Y.Y., Seki, M., Sakurai, T., Sato, M., Abe, T., Yoshida, S., Manabe, K., Shinozaki, K. and Matsui, M. (2003) Identification of Arabidopsis genes regulated by high light-stress using cDNA microarray. *Photochem Photobiol*, 77, 226-233.
- Kitamiya, E., Suzuki, S., Sano, T. and Nagata, T. (2000) Isolation of two genes that were induced upon the initiation of somatic embryogenesis on carrot hypocotyls by high concentrations of 2,4-D. *Plant Cell Reports*, 19, 551-557.
- Kiyosue, T., Shiota, H., Higashi, K., Kamada, H. and Shinozaki, K. (1998) A chromo box gene from carrot (*Daucus carota* L.): its cDNA structure and expression during somatic and zygotic embryogenesis. *Biochim Biophys Acta*, 1398, 42-46.
- Klink, V.P., Alkharouf, N., MacDonald, M. and Matthews, B. (2005) Laser capture microdissection (LCM) and expression analyses of *Glycine max* (soybean) syncytium containing root regions formed by the plant pathogen *Heterodera glycines* (soybean cyst nematode). *Plant Mol Biol*, 59, 965-979.
- Klok, E.J., Wilson, I.W., Wilson, D., Chapman, S.C., Ewing, R.M., Somerville, S.C., Peacock, W.J., Dolferus, R. and Dennis, E.S. (2002) Expression profile analysis of the low-oxygen response in Arabidopsis root cultures. *Plant Cell*, 14, 2481-2494.
- Kreps, J.A., Wu, Y., Chang, H.S., Zhu, T., Wang, X. and Harper, J.F. (2002) Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. *Plant Physiol*, 130, 2129-2141.
- Kreuger, M. and van Holst, G.J. (1996) Arabinogalactan proteins and plant differentiation. *Plant Mol Biol*, 30, 1077-1086.

Références bibliographiques

- Kunieda, T., Fujiwara, T., Amano, T. and Shioi, Y. (2005) Molecular cloning and characterization of a senescence-induced tau-class Glutathione S-transferase from barley leaves. *Plant Cell Physiol*, 46, 1540-1548.
- Lee, N.H., Weinstock, K.G., Kirkness, E.F., Earle-Hughes, J.A., Fuldner, R.A., Marmaros, S., Glodek, A., Gocayne, J.D., Adams, M.D., Kerlavage, A.R. and et al. (1995) Comparative expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment. *Proc Natl Acad Sci U S A*, 92, 8303-8307.
- Liang, P. and Pardee, A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 257, 967-971.
- Limami, A., Roux, L., Laville, J. and Roux, Y. (1993) Dynamics of nitrogen compounds in the chicory (*Cichorium intybus* L.) tap root during the growing season and cold storage period. *J Plant Physiol*, 141, 263–268.
- Limami, M.A., Dufossé, C., Richard-Molard, C., Fouldrin, K., Roux, L. and J.F., M.-G. (1996) Effect of exogenous nitrogen ($^{15}\text{NO}_3$) on utilization of Vegetative Storage Proteins (VSP) during regrowth in chicory (*Cichorium intybus*). *J Plant Physiol* 149, 564–572.
- Lindzen, E. and Choi, J.H. (1995) A carrot cDNA encoding an atypical protein kinase homologous to plant calcium-dependent protein kinases. *Plant Mol Biol*, 28, 785-797.
- Lipshutz, R.J., Fodor, S.P., Gingeras, T.R. and Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nat Genet*, 21, 20-24.
- Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-(\Delta\Delta C(T))}$ Method. *Methods*, 25, 402-408.
- Lynn, K., Fernandez, A., Aida, M., Sedbrook, J., Tasaka, M., Masson, P. and Barton, M.K. (1999) The PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis development and has overlapping functions with the ARGONAUTE1 gene. *Development*, 126, 469-481.
- Maheswaran, G. and Williams, E.G. (1985) Origin and development of somatic embryos formed directly on immature embryos of *Trifolium repens* *in vitro*. *Ann Bot*, 56, 619-630.

Références bibliographiques

- Matsubara, S., Hurry, V., Druart, N., Benedict, C., Janzik, I., Chavarria-Krauser, A., Walter, A. and Schurr, U. (2005) Nocturnal changes in leaf growth of *Populus deltoides* are controlled by cytoplasmic growth. *Planta*, 1-14.
- McKhann, H.I., Frugier, F., Petrovics, G., de la Pena, T.C., Jurkevitch, E., Brown, S., Kondorosi, E., Kondorosi, A. and Crespi, M. (1997) Cloning of a WD-repeat-containing gene from alfalfa (*Medicago sativa*): a role in hormone-mediated cell division? *Plant Mol Biol*, 34, 771-780.
- Menand, B., Desnos, T., Nussaume, L., Berger, F., Bouchez, D., Meyer, C. and Robaglia, C. (2002) Expression and disruption of the Arabidopsis TOR (target of rapamycin) gene. *Proc Natl Acad Sci U S A*, 99, 6422-6427.
- Miller, R.A., Galecki, A. and Shmookler-Reis, R.J. (2001) Interpretation, design, and analysis of gene array expression experiments. *J Gerontol A Biol Sci Med Sci*, 56, B52-57.
- Miner, D. and Rajkovic, A. (2003) Identification of expressed sequence tags preferentially expressed in human placentas by in silico subtraction. *Prenat Diagn*, 23, 410-419.
- Mitra, R.M., Shaw, S.L. and Long, S.R. (2004) Six nonnodulating plant mutants defective for Nod factor-induced transcriptional changes associated with the legume-rhizobia symbiosis. *Proc Natl Acad Sci U S A*, 101, 10217-10222. Epub 12004 Jun 10225.
- Nishiwaki, M., Fujino, K., Koda, Y., Masuda, K. and Kikuta, Y. (2000) Somatic embryogenesis induced by the simple application of abscisic acid to carrot (*Daucus carota* L.) seedlings in culture. *Planta*, 211, 756-759.
- Nothnagel, E.A. (1997) Proteoglycans and related components in plant cells. *Int Rev Cytol*, 174, 195-291.
- Novoradovskaya, N., Whitfield, M.L., Basehore, L.S., Novoradovsky, A., Pesich, R., Usary, J., Karaca, M., Wong, W.K., Aprelikova, O., Fero, M., Perou, C.M., Botstein, D. and Braman, J. (2004) Universal Reference RNA as a standard for microarray experiments. *BMC Genomics*, 5, 20.
- Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y. and Yamaguchi, S. (2003) Gibberellin biosynthesis and response during Arabidopsis seed germination. *Plant Cell*, 15, 1591-1604.

Références bibliographiques

- Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y. and Matsubara, K. (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat Genet*, 2, 173-179.
- Oztur, Z.N., Talame, V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., Tuberosa, R. and Bohnert, H.J. (2002) Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol Biol*, 48, 551-573.
- Park, M.H., Lee, Y.B. and Joe, Y.A. (1997) Hypusine is essential for eukaryotic cell proliferation. *Biol Signals*, 6, 115-123.
- Parsons, B.L. and Mattoo, A.K. (1991) Wound-regulated accumulation of specific transcripts in tomato fruit: interactions with fruit development, ethylene and light. *Plant Mol Biol*, 17, 453-464.
- Pasternak, T.P., Prinsen, E., Ayaydin, F., Miskolczi, P., Potters, G., Asard, H., Van Onckelen, H.A., Dudits, D. and Feher, A. (2002) The Role of auxin, pH, and stress in the activation of embryogenic cell division in leaf protoplast-derived cells of alfalfa. *Plant Physiol*, 129, 1807-1819.
- Pavy, N., Paule, C., Parsons, L., Crow, J.A., Morency, M.J., Cooke, J., Johnson, J.E., Noumen, E., Guillet-Claude, C., Butterfield, Y., Barber, S., Yang, G., Liu, J., Stott, J., Kirkpatrick, R., Siddiqui, A., Holt, R., Marra, M., Seguin, A., Retzel, E., Bousquet, J. and MacKay, J. (2005) Generation, annotation, analysis and database integration of 16,500 white spruce EST clusters. *BMC Genomics*, 6, 144.
- Perry, S.E., Lehti, M.D. and Fernandez, D.E. (1999) The MADS-domain protein AGAMOUS-like 15 accumulates in embryonic tissues with diverse origins. *Plant Physiol*, 120, 121-130.
- Poulain, J. (2003) Etude du rôle de gènes contrôlant le cycle cellulaire au cours du développement racinaire de *Cichorium intybus* L. Isolement et caractérisation d'une cycline mitotique de type B de chicorée. Thèse de l'Université de Lille1.
- Putney, S.D., Herlihy, W.C. and Schimmel, P. (1983) A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. *Nature*, 302, 718-721.
- Raghavan, C., Ong, E.K., Dalling, M.J. and Stevenson, T.W. (2006) Regulation of genes associated with auxin, ethylene and ABA pathways by 2,4-dichlorophenoxyacetic acid in Arabidopsis. *Funct Integr Genomics*, 6, 60-70.

- Rajkovic, A., Yan, M.S.C., Klysik, M. and Matzuk, M. (2001) Discovery of germ cell-specific transcripts by expressed sequence tag database analysis. *Fertil Steril*, 76, 550-554.
- Randoux, B., Quillet, M.C., Rambaud, C., Vasseur, J. and Hilbert, J.L. (2002) Identification of cDNAs encoding Rab-related GTP-binding proteins expressed during somatic embryogenesis in *Cichorium*. *Plant Science*, 162, 413-422.
- Ranjan, P., Kao, Y.Y., Jiang, H., Joshi, C.P., Harding, S.A. and Tsai, C.J. (2004) Suppression subtractive hybridization-mediated transcriptome analysis from multiple tissues of aspen (*Populus tremuloides*) altered in phenylpropanoid metabolism. *Planta*, 219, 694-704. Epub 2004 May 2014.
- Richard-Molard, C. (1999) Caractérisation physiologique, génétique et moléculaire de la mise en réserve de l'azote (VSP) chez *Cichorium intybus* (endive). Conséquences sur la qualité morphologique et sanitaire du chicon. *Ecole Nationale Supérieure Agronomique de Rennes*, Thesis.
- Richmond, T. and Somerville, S. (2000) Chasing the dream: plant EST microarrays. *Curr Opin Plant Biol*, 3, 108-116.
- Rishi, A.S., Munir, S., Kapur, V., Nelson, N.D. and Goyal, A. (2004) Identification and analysis of safener-inducible expressed sequence tags in *Populus* using a cDNA microarray. *Planta*, 18, 18.
- Roberts, M.R. (2003) 14-3-3 proteins find new partners in plant cell signalling. *Trends Plant Sci*, 8, 218-223.
- Rudd, S. (2003) Expressed sequence tags: alternative or complement to whole genome sequences? *Trends Plant Sci*, 8, 321-329.
- Sagare, A.P., Lee, Y.L., Lin, T.C., Chen, C.C. and Tsay, H.S. (2000) Cytokinin-induced somatic embryogenesis and plant regeneration in *Corydalis yanhusuo* (*Fumariaceae*) - a medicinal plant. *Plant Sci*, 160, 139-147.
- Sambrook, J., Fritsch, E.F and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd. ed. *Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press*.
- Sawa, S., Ohgishi, M., Goda, H., Higuchi, K., Shimada, Y., Yoshida, S. and Koshiba, T. (2002) The HAT2 gene, a member of the HD-Zip gene family, isolated as an auxin

- inducible gene by DNA microarray screening, affects auxin response in Arabidopsis. *Plant J*, 32, 1011-1022.
- Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270, 467-470.
- Schmelzle, T. and Hall, M.N. (2000) TOR, a central controller of cell growth. *Cell*, 103, 253-262.
- Schmidt, E.D., Guzzo, F., Toonen, M.A. and de Vries, S.C. (1997) A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *Development*, 124, 2049-2062.
- Schnable, P.S., Hochholdinger, F. and Nakazono, M. (2004) Global expression profiling applied to plant development. *Curr Opin Plant Biol*, 7, 50-56.
- Seki, M., Ishida, J., Narusaka, M., Fujita, M., Nanjo, T., Umezawa, T., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K. (2002) Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. *Funct Integr Genomics*, 2, 282-291.
- Senger, S., Mock, H.P., Conrad, U. and Manteuffel, R. (2001) Immunomodulation of ABA function affects early events in somatic embryo development. *Plant Cell Reports*, 20, 112-120.
- Shiota, H., Satoh, R., Watabe, K., Harada, H. and Kamada, H. (1998) C-ABI3, the carrot homologue of the Arabidopsis ABI3, is expressed during both zygotic and somatic embryogenesis and functions in the regulation of embryo-specific ABA-inducible genes. *Plant Cell Physiol*, 39, 1184-1193.
- Smith, D.L. and Krikorian, A.D. (1990) Somatic embryogenesis of carrot in hormone-free medium: external pH control over morphogenesis. *Am J Bot*, 77, 1634-1647.
- Soukas, A., Cohen, P., Socci, N.D. and Friedman, J.M. (2000) Leptin-specific patterns of gene expression in white adipose tissue. *Genes Dev*, 14, 963-980.
- Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*, 98, 503-517.

- Stafstrom, J.P., Ripley, B.D., Devitt, M.L. and Drake, B. (1998) Dormancy-associated gene expression in pea axillary buds. Cloning and expression of PsDRM1 and PsDRM2. *Planta*, 205, 547-552.
- Stasolla, C., van Zyl, L., Egertsdotter, U., Craig, D., Liu, W. and Sederoff, R.R. (2003) The effects of polyethylene glycol on gene expression of developing white spruce somatic embryos. *Plant Physiol*, 131, 49-60.
- Sterk, P., Booij, H., Schellekens, G.A., Van Kammen, A. and De Vries, S.C. (1991) Cell-specific expression of the carrot EP2 lipid transfer protein gene. *Plant Cell*, 3, 907-921.
- Steward, F., Mapes, MO, Smith, J (1958) Growth and organized development of cultured cells. I. Growth and division of freely suspended cells. *Am. J. Bot.*, 45, 693-703.
- Storey, J.D. and Tibshirani, R. (2003) Statistical methods for identifying differentially expressed genes in DNA microarrays. *Methods Mol Biol*, 224, 149-157.
- Taniguchi, M., Miura, K., Iwao, H. and Yamanaka, S. (2001) Quantitative assessment of DNA microarrays--comparison with Northern blot analyses. *Genomics*, 71, 34-39.
- Thibaud-Nissen, F., Shealy, R.T., Khanna, A. and Vodkin, L.O. (2003) Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean. *Plant Physiol*, 132, 118-136.
- Thompson, J.E., Hopkins, M.T., Taylor, C. and Wang, T.W. (2004) Regulation of senescence by eukaryotic translation initiation factor 5A: implications for plant growth and development. *Trends Plant Sci*, 9, 174-179.
- Toonen, M.A.J., Hendriks, T., Schmidt, E.D.L., Verhoeven, H.A., Van Kammen, A. and De Vries, S.C. (1994) Description of somatic-embryo-forming single cells in carrot suspension cultures employing video cell tracking. *Planta*, 194, 565-572.
- Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt,

Références bibliographiques

- R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J.C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y. and Rokhsar, D. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 313, 1596-1604.
- Ueda, A., Kathiresan, A., Inada, M., Narita, Y., Nakamura, T., Shi, W., Takabe, T. and Bennett, J. (2004) Osmotic stress in barley regulates expression of a different set of genes than salt stress does. *J Exp Bot*, 55, 2213-2218. Epub 2004 Sep 2210.
- van den Heuvel, K.J., van Esch, R.J., Barendse, G.W. and Wullems, G.J. (1999) Isolation and molecular characterization of gibberellin-regulated H1 and H2B histone cDNAs in the leaf of the gibberellin-deficient tomato. *Plant Mol Biol*, 39, 883-890.
- van Hengel, A.J., Guzzo, F., van Kammen, A. and de Vries, S.C. (1998) Expression pattern of the carrot EP3 endochitinase genes in suspension cultures and in developing seeds. *Plant Physiol*, 117, 43-53.
- Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science*, 270, 484-487.
- Verica, J.A., Maximova, S.N., Strem, M.D., Carlson, J.E., Bailey, B.A. and Gultinan, M.J. (2004) Isolation of ESTs from cacao (*Theobroma cacao* L.) leaves treated with inducers of the defense response. *Plant Cell Rep*, 23, 404-413
- Vettore, A., da Silva, FR, Kemper, EL and Arruda, P (2001) The libraries that made SUCEST. *Genetics and Molecular Biology*, 24, 1-7.
- Wang, T.W., Lu, L., Zhang, C.G., Taylor, C. and Thompson, J.E. (2003) Pleiotropic effects of suppressing deoxyhypusine synthase expression in *Arabidopsis thaliana*. *Plant Mol Biol*, 52, 1223-1235.

Références bibliographiques

- Wellmer, F., Alves-Ferreira, M., Dubois, A., Riechmann, J.L. and Meyerowitz, E.M. (2006) Genome-Wide Analysis of Gene Expression during Early Arabidopsis Flower Development. *PLoS Genetics*, preprint.
- Willats, W.G. and Knox, J.P. (1996) A role for arabinogalactan-proteins in plant cell expansion: evidence from studies on the interaction of beta-glucosyl Yariv reagent with seedlings of *Arabidopsis thaliana*. *Plant J*, 9, 919-925.
- Windels, D. (2005) Etude de l'implication des arabinogalactanes protéines (AGPs) au cours de l'embryogenèse somatique de différents génotypes de chicorée. Thèse de l'Université de Lille 1.
- Wiweger, M., Farbos, I., Ingouff, M., Lagercrantz, U. and Von Arnold, S. (2003) Expression of Chia4-Pa chitinase genes during somatic and zygotic embryo development in Norway spruce (*Picea abies*): similarities and differences between gymnosperm and angiosperm class IV chitinases. *J Exp Bot*, 54, 2691-2699.
- Wolfinger, R.D., Gibson, G., Wolfinger, E.D., Bennett, L., Hamadeh, H., Bushel, P., Afshari, C. and Paules, R.S. (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol*, 8, 625-637.
- Wu, P., Ma, L., Hou, X., Wang, M., Wu, Y., Liu, F. and Deng, X.W. (2003) Phosphate starvation triggers distinct alterations of genome expression in *Arabidopsis* roots and leaves. *Plant Physiol*, 132, 1260-1271.
- Wysocka-Diller, J.W., Helariutta, Y., Fukaki, H., Malamy, J.E. and Benfey, P.N. (2000) Molecular analysis of SCARECROW function reveals a radial patterning mechanism common to root and shoot. *Development*, 127, 595-603.
- Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J. and Speed, T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*, 30, e15.
- Yang, Y.H. and Speed, T. (2002) Design issues for cDNA microarray experiments. *Nat Rev Genet*, 3, 579-588.
- Yasuda, H., Nakajima, M., Ito, T., Ohwada, T. and Masuda, H. (2001) Partial characterization of genes whose transcripts accumulate preferentially in cell clusters at the earliest stage of carrot somatic embryogenesis. *Plant Mol Biol*, 45, 705-712.

Références bibliographiques

- Yazawa, K., Takahata, K. and Kamada, H. (2004) Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis. *Plant Physiol Biochem*, 42, 215-223.
- Yeung, K.Y., Haynor, D.R. and Ruzzo, W.L. (2001) Validating clustering for gene expression data. *Bioinformatics*, 17, 309-318.
- Zeng, F., Zhang, X., Zhu, L., Tu, L., Guo, X. and Nie, Y. (2006) Isolation and characterization of genes associated to cotton somatic embryogenesis by suppression subtractive hybridization and macroarray. *Plant Mol Biol*, 60, 167-183.
- Zhu-Salzman, K., Salzman, R.A., Ahn, J.E. and Koiwa, H. (2004) Transcriptional regulation of sorghum defense determinants against a phloem-feeding aphid. *Plant Physiol*, 134, 420-431.
- Zik, M. and Irish, V.F. (2003) Global identification of target genes regulated by APETALA3 and PISTILLATA floral homeotic gene action. *Plant Cell*, 15, 207-222.
- Zimmerman, J.L. (1993) Somatic Embryogenesis: A Model for Early Development in Higher Plants. *Plant Cell*, 5, 1411-1423.
- Zimmermann, P. and Zentgraf, U. (2005) The correlation between oxidative stress and leaf senescence during plant development. *Cell Mol Biol Lett*, 10, 515-534.
- Zorinants, S.E., Nosov, A.V., Monforte-Gonzalez, M., Mendes-Zeel, M. and Loyola-Vargas, V.M. (2003) Variation of nuclear DNA content during somatic embryogenesis and plant regeneration of *Coffea arabica* L. using. *Plant Science*, 164, 141-146.
- Zuo, J., Niu, Q.W., Frugis, G. and Chua, N.H. (2002) The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. *Plant J*, 30, 349-359.

- Annexes -

Annexes

Annexe 1 : Milieux de culture utilisés pour la multiplication clonale des génotypes K59 et C15 de chicorées et la production d'embryons somatiques

Composition des milieux de culture	M17S20	H5	H5 solide
Solution de macroéléments concentrée 10 x	100 ml	100 ml	100 ml
Microéléments de Heller	1 ml	1 ml	1 ml
Glutamine	250 mg	-	-
Vitamines de Morel et Wetmore	10 ml	10 ml	10 ml
Inositol	100 mg	-	-
2iP	0,5 mg	-	-
ANA	0,02 mg	-	-
KH2P04	85 mg	-	-
Solution de Fe-EDT A	5 ml	5 ml	5 ml
Saccharose	20 g	5 g	5 g
Agar	-	-	6 g
H2O	qsp 1 l	qsp 1 l	Qsp 1 l

Solutions stock macro éléments x10 (mg/l)	M17S20	Heller
MgS04, 7 H2O	1850	2500
CaCl2, 2H2O	2200	750
NH4N03	8250	
KCl	7500	7500
NaN03		6000
NaH2P04, 2H2O		1413

Microéléments Heller (1953) (mg/ml)	
ZnS04, 7H2O	1
H3B04	1
MnS04	0,075
CuS04,5H2O	0,03
AlCl3	0,03
NiCL2, H2O	0,01
KCl	0,01

Vitamines de Morel et Wetmore (1951) (mg/ml)	
Panthoténate de Ca	1
Thiamine	1
Acide Nicotinique	1
Pyridoxine	1
Biotine	0,01

Microéléments de Murashige et Skoog (MS) (1962) (mg/l)	
H3B03	620
CoCl2	2,5
CUS04	2,5
MnS04	1690
KI	83
ZnS04, 7H2O	860
Na2Mo04, 2H2O	25

Milieu de Bourgeonnement Bg 0,5	
CaCl2, H2O	4,4 mg
MgS04, 7H2O	3,7 mg
NH4N03	16,5 mg
KN03	19 mg
KH2P04	1,7 mg
Solution Fe-EDTA	5 ml
Microéléments (MS)	1 ml
Glutamine	100 mg
Vitamines de Morel et Wetmore	10 ml
Inositol	100 mg
BAP	0,5 mg
Saccharose	20 g
pH	5,7

Solution de Fe-EDTA (g/l)	
Na2EDTA	7,46
FeS04, 7H2O	5,56

Annexes

Annexe 2 : Amorces utilisées pour les analyses de PCR en temps réel

Numéro d'accession	Contig/OC ID	Nature de la protéine déduite	Amorce sens (5' - 3')	Amorce Anti-sens (5' - 3')
DT212234	OC0687	arabinogalactan protein [D. carota]	CCCAGATGCTCACCCACTTCC	TCAAACACCACCCCTCAACTACC
DT212300	OC1814	No hits found	TGGCGATGGTGGCTCATGC	CAAATTC AACACAAACCCTCAACTCAG
DT212813	OC0311_a	acidic ribosomal protein P0 [G. max]	GGAAGTTTGCCCTTGGTGTCTC	AGCATTCTGTAGCCATTGATGACC
DT213187	OC1549_a	40S ribosomal protein S17 [L. esculentum]	AATCCTTGAAGAAGTTGCCATCATCC	TGCCTCTGACGGGTCCCTTC
DT211067	Cont6402	putative leaf development protein Argonaute [A. thaliana]	GAGCAAGCTCGTGGCAGAAAC	AGTAGGTCAATGGAACATGATGAACAAG
DT212502	OC1929	G protein beta subunit [N. plumbaginifolia]	TCTCAAGCAGGAGTCCGAAATGG	CAACATACITTTCTAATAACGCCCAATTC
DT212655	OC1088	ribosomal protein L12 [C. intybus]	AGCCTCATTTACCTCTGTTCTTTG	CCGACCCGACCTCACCTC
DT212554	Cont0001	glyceraldehyde-3-phosphate dehydrogenase [S. alba]	ACGACAGGTTTGGCATTGTTGAG	TTCCACCTCTCCAGTCCCTTCATTG
DT212144	Cont0082	putative argininosuccinate synthase [A. thaliana]	GATACCTCAGTATTGTTCCTTGGC	TGCTTCCAAACCTTCCAATTCGAG
DT210897	OC0178	putative ribosomal S29 protein [A. thaliana]	ACGAGATTAGCTGCGGCTTCTTC	GGACCGTAAGTCTTAGGATGAGAGTTC
DT212250	OC2155_a	P40-like ribosomal protein [D. carota]	AACAAGGGAAAGCACAGCATTGG	AGGGGCAGCAGCATCATCTTC
DT212789	OC1427_b	Cell division control protein 48 homolog E [A. thaliana]	GATTTGGAAGATGATTCTATTGATGCTGAG	CTCACGCTTAACATTTCTCAAGACCTC
DT212237	OC0834	chromatin remodeling factor CHD3 [A. thaliana]	AGGATTGCAGTCTTGCTTCTTGTAAG	ACTATTGCTTGCCATCTTCCATAACC
DT212803	OC1211	BTB/POZ domain-containing protein [A. thaliana]	GCATTGTAACATGGATCAGCAAAGC	GCGGGCAGCAAACCTCGTAATC
DT211305	Cont0015	L-asparaginase [G. max]	ATCAGCATCTACGCCACAGAGAC	TGAGTCACCTATTCGCGCCATTC
DT211372	OC0631	No hits found	AGCCTGCTAACGATACTAGTGGTG	AATTTAACATCCGAAAGTCTTTGAGTAAGC
DT210839	OC0758	putative RNA-binding protein [A. thaliana]	GATCGCAGTGTATCGGGATATTCCG	GGGAAGCCTTGTGGAGATGGG
DT212771	Cont4942	elongation factor EF-2 [P. sativum]	CACTGTCTACCTGAGTCCGGTTTG	CTGAGAACCTCATGGCAGCGATTG
DT211255	Cont0024	seed specific protein Bn15D17A [B. napus]	CGAGTCTACCTGAGTACCAATC	TTCAATGTGCTGCTTCTTATCCTGTATC
DT210862	OC0172	embryogenic callus protein 181 [D. carota]	GGTGAAGACAAGTAAACAGGTTTG	TTCACGTAACCTCGCAGCATCAAG
DT212897	OC1229	sucrose synthase isoform I [D. carota]	AATGGCGAGATTGGACAATGTGAAG	CTCCTTCTTCTGTCCACCACTAC
DT214125	OC3001	lipid transfer protein precursor [M. domestica]	GTAAGTTGACACCACATTTCCAGG	CACCGCCGACCGTAAGCTG
DT211562	OC0559	putative phosphatase type 2C [A. thaliana]	TCTTGGAAGTGACGGTCTCTG	TCTCTCTTCAACGCTTCGGCTAC
DT211152	Cont0206	fasciclin-like AGP 12 [P. alba x P. tremula]	AATGGGTTTGCAGGATTGGTGTTC	TTGGCGGTGACTCGGCTAATG
DT212829	OC1960	tuber-specific and sucrose-responsive element binding factor [S. tuberosum]	CTCCACCCTACCCGAAACC	ACCTCCATCCTAACCATCTCTGC
DT212596	OC1155	xyloglucan endotransglycosylase [M. domestica]	CGATGTTCCAATCAGAGTGTTCAG	ATCGGCGTTCCTAACTAGAATATATC
DT213276	OC1785_a	H(+)-transporting ATPase [P. vulgaris]	AGTGTCAATTACAGTGTGTGCTTAC	TGTTCCGTCGTTAGAATTGCTATG
DT213079	OC1347	putative ethylene response element binding protein [N. tabacum]	GGTGTGGCTTGGGACTTATGAAAC	CCATCACTGACTTTCGTCGTTGAG
DT212556	OC1068	leaf senescence protein-related (YLS7) [A. thaliana]	TGAAGCCCGCAGAGGATATAGTTG	GCCTATACGCAAACGCCTTAGTTATATC
DT210805	OC0344	unknown protein [A. thaliana]	ATGTCGTCGCTCTTCTTCTTCC	ATCTTGCTTTCATTGTATGTGGATCGTC
DT212452	OC0603_b	cobalamine-independent methionine synthase [S. scutellarioides]	CGGTTCTCGGTGCGTGAAG	GGGCGTTTGGTCAITTTCTTGAG
DT210888	OC0071	14-3-3 protein [S. tuberosum]	TGGTTCCTCGGCTTCTTTCAG	GTTTCATAGTGTGCTCTGCTGCTTC
DT211362	Cont0201	putative auxin-induced protein [A. thaliana]	GTTGGTCCATAAATTGGCAGAAAGTTG	CAGGTATTGGAACCACATCGCTTC
DT213305	Cont0060	3-hydroxy-3-methylglutaryl coenzyme A reductase [S. tuberosum]	TCCGACCATCACCGCCATAG	GGTAGTCTTGAACATCTGAACATCATA
DT212233	OC1796	ring domain containing protein [C. annuum]	TTATTAACCCTAGTCCGACGATTGG	CGGCAAGATTATAAGAATTCGGAGTAG
DT213289	Cont9039	metallothionein 1 [A. tripolium]	AGTCCGGATGCGGCTCAAG	CGACGATCACGGTGGCGG
DT211922	OC0724	cellulose synthase family protein [A. thaliana]	TGTAACCATAAGCACTGTAGCGATTG	CACCCATTTACCACGCCAAACC
DT211657	Cont1237	pre-pro-cysteine proteinase [L. esculentum]	AACTACCGATGCTGTCTTCC	CACATGATCTTGGTCTTGTACTGG
DT212623	OC0023_a	glutathione transferase [H. muticus]	CGACGAGGCGATAGTGAAGAG	TTCACCTGTGTTCCCATCAAATACTTC
DT211599	OC0686	unknown protein [A. thaliana]	TTAGCGGTGTTGGATACGAATGTG	ACAGAAGATAAGGTGGGGAACAGTC
DT211671	Cont0026	ss-1,3-glucanase [C. intybus x C. endivia]	TGTTGGGCAGTTCCTTCTTCC	GGGTAGATGTTGGCAAGCATTGG
DT211633	OC0696	aquaporin [H. annuus]	GATCAATCGGTGCCAATCTTCC	ACCCGCAACCCGTTGATGG
DT211055	OC0168	1-aminocyclopropane-1-carboxylate oxidase [L. esculentum]	GGGATTAGAGAAGGGCTACCTCAAG	GCAATATGACACCCTGAGCATC
DT210799	Cont0006	1-aminocyclopropane-1-carboxylate oxidase [P. hortorum]	GCATGGAGAAACTGACCGGAGAAG	CCCTTTGTTCCGCCACCAATTCC
DT213897	Cont0011	catalase 3 [H. annuus]	CCAATCAATTCGGCTAGGGTACTG	TCTGTATCTATCTCTGGTCTGTTG

Annexe 3 : Article soumis à BMC Plant Biology : Characterization of expressed sequence tags obtained by SSH during somatic embryogenesis in *Cichorium intybus* L.

Characterization of expressed sequence tags obtained by SSH during somatic embryogenesis in *Cichorium intybus* L.

Sylvain Legrand, Theo Hendriks, Jean-Louis Hilbert, Marie-Christine Quillet[§]

UMR USTL-INRA 1281 Laboratoire de Stress Abiotiques et de Différenciation des Végétaux Cultivés, Université de Sciences et Technologies de LILLE, Bâtiment SN2- 59655 Villeneuve d'Ascq, France

[§]Corresponding author

Email addresses:

SL: sylvain.legrand@ed.univ-lille1.fr

TH: theo.hendriks@univ-lille1.fr

JLH: jean-louis.hilbert@univ-lille1.fr

MCQ: marie-christine.quillet@univ-lille1.fr

Summary

Somatic embryogenesis (SE) is an asexual propagation pathway that requires a somatic-to-embryonic transition of differentiated somatic cells toward embryogenic cells capable of producing embryos in a process that resembles zygotic embryogenesis. Processes related to the induction of somatic embryogenesis have been, and still are, studied in our laboratory by physiological, cytological, and, more recently, molecular and genetic approaches.

In chicory, variability with respect to the formation of somatic embryos was detected between plants from a population of *Cichorium intybus* L. landrace Koospol. This variability was exploited to generate an embryogenic (E) and a non-embryogenic (NE) cDNA library by suppression subtractive hybridisation (SSH), using mRNAs isolated from explants from an embryogenic (K59) and a non-embryogenic (C15) genotype cultured for 4 days under SE-inducing conditions. A total of 5,000 ESTs from both libraries were sequenced, and 3,348 thereof turned out to represent a maximum of 2,077 genes.

The low redundancy in our libraries, the low efficiency of an *in silico* subtraction from all ESTs, and the high number of differentially expressed genes in the libraries, indicated a high efficiency of both normalisation and subtraction steps during library construction.

cDNA microarray analyses during the first six days of the SE culture led to the identification of 81 differentially expressed genes in the E or the NE-genotype, and 417 genes with a modulated expression. Clustering of the expression profiles and functional classification analyses of the genes allowed the construction of a model explaining the response of both genotypes to the SE culture conditions.

Functions of the genes with an increased expression during the time course study indicate that cells in explants from the embryogenic genotype react with the expression of genes involved in energy generation followed by the expression of genes involved in cell division. Moreover, a gene encoding a Scarecrow transcription factor may be involved in the control of how cells divide.

Rather than the opportunistic response observed for cells in explants from the embryogenic genotype, functions of the genes with an increased expression in the non-embryogenic genotype pointed to a defensive reaction of cells, leading possibly to the initiation of a senescence program.

Keywords : *Cichorium intybus*, somatic embryogenesis, SSH, EST, microarray, Real time RT-PCR

Résumé

L'embryogenèse somatique est un moyen de propagation végétative qui nécessite, à partir de cellules différenciées, une transition d'un état somatique à un état embryonnaire. Au laboratoire, différentes approches sont utilisées pour étudier ce processus chez la chicorée : physiologie, cytologie, protéomique et transcriptomique. C'est dans le cadre de cette dernière approche que se situe ce travail.

Récemment au laboratoire, une variabilité génétique pour la capacité à former des embryons somatiques *in vitro* a été mise en évidence chez la chicorée et un génotype embryogène (K59), ainsi qu'un génotype non embryogène (C15) obtenu par autofécondation du génotype embryogène, ont été sélectionnés.

L'utilisation d'explants foliaires des deux génotypes, placés 4 jours en conditions d'embryogenèse somatique, a permis de construire deux banques d'ADNc soustractives (SSH) : une banque dite « embryogène » et une banque dite « non embryogène ». A partir de ces banques, environ 5000 clones ont été séquencés. Après avoir trié les séquences de mauvaise qualité et les contaminants ribosomiaux, 3348 ESTs (Expressed Sequence Tag) ont été conservés pour la suite de l'étude et des travaux d'alignement de séquences ont établi que ces ESTs représentent 2077 gènes.

La faible redondance des ESTs dans les banques et la faible puissance d'une soustraction *in silico* des ESTs témoignent de l'efficacité de l'étape de normalisation des transcrits lors de la fabrication des banques. L'identité différente des gènes représentés dans les deux banques et la forte proportion de gènes différentiellement exprimés suggère également une soustraction efficace des transcrits.

Une analyse transcriptomique par microarray sur les six premiers jours de culture en conditions d'embryogenèse somatique a permis d'identifier 81 gènes différentiellement exprimés au stade initial de la cinétique entre les deux génotypes K59 et C15 et 417 gènes montrant une modulation d'expression significative sur au moins un point de la cinétique chez l'un et/ou l'autre des génotypes. Le regroupement des gènes par profils d'expression et la classification fonctionnelle des gènes a permis de proposer un modèle expliquant les réponses des génotypes embryogènes et non embryogènes aux conditions de culture.

Les fonctions des gènes présentant une hausse d'expression chez le génotype K59 au cours de la cinétique montrent que les cellules du génotype embryogène réagissent aux conditions de culture par l'expression de gènes conduisant à la synthèse d'énergie puis de gènes potentiellement impliqués dans la division des cellules. De plus, un gène codant un facteur de transcription de type Scarecrow pourrait être impliqué dans la régulation des modalités des divisions cellulaires.

Contrairement à la réaction opportuniste du génotype embryogène, le génotype non embryogène présente une réaction très différente aux conditions de culture. En effet, les fonctions des gènes montrant une hausse d'expression significative chez le génotype C15 semblent traduire la mise en place d'un programme de sénescence au sein des explants du génotype non-embryogène.

Mots clés : *Cichorium intybus*, embryogenèse somatique, banques soustractives, EST, microarray, PCR en temps réel