

N° d'ordre : 2249

THESE

présentée

pour obtenir

LE TITRE DE DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE TOULOUSE

École doctorale : **BIOLOGIE, SANTE, BIOTECHNOLOGIES**

Spécialité : **BIOSCIENCES VEGETALES**

Par

Cécile BEN

Analyse du transcriptome lors de l'embryogenèse précoce chez le Tournesol

Soutenue le 15 septembre 2005 devant le jury composé de :

Dr. J.D. FAURE	INRA, Versailles, Rapporteur
Dr. D. JOB	CNRS/Bayer CropScience, Lyon, Rapporteur
Dr. J.P. GALAUD	UPS/CNRS, Toulouse, Examineur
Pr. L. GENTZBITTEL	INPT-ENSAT, Toulouse, Directeur de thèse

Ecole Nationale Supérieure Agronomique de Toulouse
Laboratoire de Biotechnologie et Amélioration des Plantes
Pôle de Biotechnologie Végétale - 18, Chemin de Borde Rouge, BP107 - 31326 CASTANET-TOLOSAN

**A mon fiancé,
Sébastien,
Avec tout mon amour.**

REMERCIEMENTS

Ces remerciements s'adressent à toutes les personnes qui, de près ou de loin, ont contribué à l'aboutissement de mon projet de thèse.

Je tiens tout d'abord à exprimer ma profonde reconnaissance au Professeur Laurent Gentsbittel pour m'avoir accordé sa confiance et m'avoir guidée dans mes réflexions scientifiques tout au long de ce travail. Que cette thèse soit le témoignage de toute ma gratitude pour ses précieux conseils, pour son soutien ainsi que pour les nombreuses discussions que nous avons eues, si riches en enseignements tant sur le plan scientifique qu'humain.

Je remercie Monsieur Dominique Job et Monsieur Jean-Denis Faure de m'avoir fait l'honneur d'accepter d'être rapporteurs de ce travail. Je remercie également Monsieur Jean-Philippe Galaud pour sa participation à mon jury de thèse. Qu'ils trouvent ici l'expression de ma plus haute considération.

Je remercie les Professeurs Gilbert Alibert et Michel Petitprez qui m'ont accueillie au sein du laboratoire de Biotechnologies et Améliorations des Plantes de l'ENSAT.

J'adresse également un immense merci à Tarek Hewezi pour toutes les connaissances qu'il m'a apportées, en particulier en ce qui concerne l'approche microarray, et pour notre agréable et efficace coopération. Qu'il sache que j'ai beaucoup appris à ses côtés.

Un merci tout particulier à Cecilia Tamborindeguy et Nathalie Ladouce qui, en me témoignant leur profonde amitié, m'ont soutenue à chaque instant.

Un grand merci également à Thierry Liboz pour ses encouragements et son aide précieuse, notamment pour l'annotation fonctionnelle des clones, ainsi qu'à Marie-Françoise Jardinaud pour ses nombreux conseils et nos discussions scientifiques des plus enrichissantes.

J'exprime toute ma reconnaissance à l'ensemble des personnes ayant très largement participé à la réalisation du projet Mago Nashi : Christian Brière, Laurent Gentsbittel, Marie-Françoise Jardinaud, Thierry Liboz, Cédric Muller, Michel Petitprez et Cecilia Tamborindeguy. Je les remercie pour la passion et l'énergie qui nous ont animés et unis au fil des découvertes.

Je tenais également à remercier très chaleureusement Cédric Muller, Sébastien Moretti et Nicolas Salatgé pour leur aide si précieuse dans le domaine informatique.

J'exprime toute ma gratitude à nos collègues de Génoplante pour m'avoir permis d'exploiter leur données ainsi qu'à ceux du CRGS Toulouse-Midi Pyrénées en particulier Cécile Tonon, Julien Sarry et Gérard Salin, pour m'avoir permis d'utiliser leur matériel ainsi que pour leur conseil technique concernant la gestion des banques d'ADNc et la réalisation des microarrays.

Merci également aux différentes stagiaires ayant aidé à l'accomplissement de mes travaux de thèse : Elodie, Marie-Josée, Gaby et Clara.

Je voudrais vivement remercier tous les membres du BAP pour m'avoir accueillie si chaleureusement. Je garderai un merveilleux souvenir de ces années passées parmi eux. Tout particulièrement, j'adresse un grand merci à Annie, Cathy, Marie-Jo, Philippe, Patrick et Sylvie pour leur disponibilité et leur immense gentillesse.

Pour finir, j'adresse tout mon amour à mon fiancé, à mes parents, à mon frère, à toute ma famille et à mes amis qui m'ont toujours soutenue dans cette aventure. Vous comptez plus que tout pour moi.

Sébastien, mon Amour, sans ton soutien de chaque instant, sans ton aide si précieuse, sans ton amour, tout simplement, je n'aurais été capable de rien. Pour avoir tant contribué à transformer ce très beau défi en réussite...Merci !

ABSTRACT

In higher plants, embryogenesis is a process of main importance which corresponds to the establishment of the correct embryo pattern and to the accumulation of storage reserves. Thus, the knowledge of the molecular and physiological events of this process represents a major interest for agronomy to improve grains quality and yields.

However, the analysis of the early stages of development is often difficult because the embryo is small and embedded inside the maternal tissue. Sunflower (*Helianthus annuus L.*), which presents an inflorescence with numerous and relatively big embryos whose development is synchronized, has been proposed as a complementary model for the study of zygotic embryogenesis in dicotyledonous.

In order to analyze the transcriptional program induced during the embryo patterning, reference cDNA libraries were constructed from globular, heart-shaped and early cotyledonary embryos which represent critical stages of embryogenesis. A subtracted library from heart-shaped and cotyledonary embryos and enriched in transcripts from heart-shaped stage was also realized. A total of 23,800 clones was generated among which 4,118 were sequenced giving rise to 1,690 published EST. For *in silico* analysis, sequences from the reference libraries were coupled to EST from cDNA libraries prepared from unfertilized ovules and mid and late cotyledonary embryos. The global EST dataset represents 7,106 sequences and 3,064 potentially different genes. Functional annotation of the putative unique sequences allowed us to identify the major cell functions implicated at each stage of development revealing that embryos are engaged in active differentiation process. The EST collection analysis also led to the identification of some interesting genes such as putatively differentially expressed genes or genes with potential key roles in plant and animal embryogenesis such as Argonaute or Mago Nashi like genes.

These libraries, grouped with other sunflower cDNA libraries prepared from different tissues, were used to construct a microarray containing 8,185 unique sequences. This microarray was hybridized with radiolabelled cDNA from globular, heart-shaped and cotyledonary embryos as well as unfertilized ovules and leaves used as external controls. The statistical analysis of microarray data allowed us to identify 744 differentially expressed genes between at least two conditions. The analysis of the expression profiles of those genes permitted us to build hypothesis concerning the mechanisms participating in early phases of sunflower embryogenesis. Specific transcriptional and post-transcriptional regulatory pathways as well as diverse signal transduction processes seem to be involved at each stage of this crucial phase of plant life cycle. Functional plastids biogenesis and senescence also appear to play essential roles for normal embryo development.

Key words: *Helianthus annuus L.*, embryogenesis, EST, cDNA library, microarray, gene expression analysis.

ABREVIATIONS

ABA	Acide abscissique
ADN	Acide désoxyribonucléique
ADNc	Acide désoxyribonucléique complémentaire
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messenger
BAC	Bacterial artificial chromosome
BET	Bromure d'éthidium
bp	Paire de base
cDNA-AFLP	Complementary desoxyribonucleic acid-amplified fragment length
DNase	Désoxyribonucléase
dNTP	Désoxynucléotides triphosphate
EDTA	Acide Ethylène Diamine Tétra-acétique
EMS	Ethyl méthane sulfonate
EST	Expressed Sequence Tag
kb	kilobase
LCM	Laser Capture Microdissection
PCR	Polymerase Chain Reaction
QTL	Quantitative trait loci
eQTL	Expression quantitative trait loci
RNAi	RNA interference
RNase	Ribonucléase
RNP	Ribonucléoprotéine
RT-PCR	Reverse transcription-PCR
SDS	Sodium Dodécyl Sulfate
SSC	Salt Sodium Citrate
SSH	Suppressive subtractive hybridization
TAE	Tampon Acétate EDTA
TE	Tris-EDTA
Tris	Tris (hydroxyméthyl)-méthylamine

Sommaire Détaillé

RESUME.....	1
PARTIE 1 : ETUDE BIBLIOGRAPHIQUE	
Chapitre I - L'embryogenèse chez les végétaux supérieurs.....	2
I. L'embryogenèse : une étape clé du développement des plantes.....	2
I.1. Une phase essentielle du développement.....	2
I.2. La graine : un organe végétal d'un intérêt agronomique capital.....	2
I.3. La séquence de développement de la graine comporte trois phases clés.	3
II. Description histologique et morphologique de l'embryogenèse des Dicotylédones.....	4
II.1. La fécondation: étape d'initiation de l'embryogenèse.....	4
<i>II.1.1. Les acteurs de la fécondation.....</i>	<i>4</i>
<i>II.1.2. La double fécondation : une originalité des Angiospermes</i>	<i>5</i>
II.2. Mise en place des axes apico-basal et radial de l'embryon.....	6
<i>II.2.1. La première division asymétrique traduit la polarisation du zygote.....</i>	<i>6</i>
<i>II.2.1.1. Déroulement de la première division du zygote.....</i>	<i>6</i>
<i>II.2.1.2. Hypothèses sur l'origine de la polarité du zygote.....</i>	<i>7</i>
<i>II.2.2. De l'embryon deux cellules jusqu'à l'embryon globulaire : mise en place des axes apico-basal et radial.....</i>	<i>9</i>
<i>II.2.3. Transition globulaire-coeur : acquisition de la symétrie bilatérale et organisation définitive de la plante.....</i>	<i>10</i>
<i>II.2.4. Stade cotylédonaire : croissance et accumulation des réserves.....</i>	<i>10</i>
II.3. Autres modèles de développements embryonnaires.....	10
<i>II.3.1. Comparaison avec l'embryogenèse zygotique chez les Monocotylédones.</i>	<i>10</i>
<i>II.3.2. Principales différences entre embryogenèse végétale et animale.....</i>	<i>11</i>
III. Méthodes d'étude des mécanismes moléculaires impliqués au cours de l'embryogenèse végétale.....	12
III.1. L'analyse par mutagenèse, une source d'informations abondantes mais insuffisantes.....	12
<i>III.1.1. Criblage de siliques immatures.....</i>	<i>13</i>
<i>III.1.2. Criblage de plantules en germination.....</i>	<i>13</i>

<i>III.1.3. Limites de l'approche par mutagenèse.....</i>	14
III.2. L'analyse par les méthodes de la génomique fonctionnelle, une approche prometteuse.....	14
<i>III.2.1. Les techniques de transcriptomique à faible débit.....</i>	14
<i>III.2.2. Les approches ESTs et microarrays ou l'étude globale de l'expression du génome.....</i>	15
IV. Modèle génétique et fonctionnel de l'établissement du patron embryonnaire chez les dicotylédones.....	17
IV.1. L'élaboration du plan d'organisation de l'embryon végétal implique une régulation génétique très fine.....	17
<i>IV.1.1. Inégalité des sexes lors de l'embryogenèse précoce : rôle prédominant du génome maternel.....</i>	17
<i>IV.1.2. Gènes impliqués dans la mise en place de l'axe apico-basal.....</i>	18
<i>IV.1.2.1. Suspenseur vs embryon : acquisition de devenirs cellulaires distincts dès la première division asymétrique.....</i>	18
<i>IV.1.2.2. Mise en place du méristème racinaire.....</i>	19
<i>IV.1.2.3. Mise en place du méristème apical.....</i>	20
<i>IV.1.3. Gènes intervenant dans la mise en place de l'axe radial.....</i>	20
<i>IV.1.3.1. Gènes impliqués dans la spécification et le maintien du protoderme.....</i>	20
<i>IV.1.3.2. Gènes impliqués dans la différenciation des tissus internes.....</i>	21
<i>IV.1.4. Gènes intervenant dans l'établissement de la symétrie bilatérale : Mise en place des organes latéraux.....</i>	22
IV.2. La différenciation cellulaire au sein de l'embryon végétal nécessite l'échange d'indices de position.....	22
<i>IV.2.1. La paroi cellulaire : un composant actif de la transduction du signal au sein de l'embryon.....</i>	23
<i>IV.2.1.1. Propriétés chimiques de la paroi végétale et régulation du développement embryonnaire.....</i>	23
<i>IV.2.1.2. Les plasmodesmes, douanes régulant les échanges moléculaires entre cellules.....</i>	23
<i>IV.2.2. Les hormones végétales, médiateurs indispensables à la formation d'un patron embryonnaire correct.....</i>	24
<i>IV.2.2.1. L'auxine, molécule coordinatrice de l'établissement du patron embryonnaire.....</i>	24
<i>IV.2.2.2. Les stérols : régulateurs de la division et de l'expansion cellulaire lors de l'embryogenèse précoce?</i>	26
<i>IV.2.2.3. Le rôle des cytokinines au cours de l'embryogenèse précoce reste encore à définir.....</i>	27
IV.3. Les récentes analyses transcriptomiques offrent un premier aperçu des principales fonctions cellulaires et voies de régulation impliquées au cours de l'embryogenèse végétale.....	27
<i>IV.3.1. Un large éventail de fonctions cellulaires indispensables au développement embryonnaire.....</i>	27
<i>IV.3.2. Comparaison 'tissus maternels vs. embryonnaires' et identification des particularités du programme embryonnaire.....</i>	28

IV.3.3. Des programmes génétiques spécifiques à chaque type de graines : exemple des graines vertes et non vertes.....29

IV.4. Elaboration du patron embryonnaire chez les dicotylédones : organisation modulaire de l’embryon ou croissance différentielle ?.....29

Chapitre II - Le tournesol : un modèle d’étude complémentaire de l’embryogenèse zygotique chez les Dicotylédones.....31

I. Le tournesol : une plante d’intérêt agronomique.....31

I.1. Le tournesol : une plante ayant subi les effets de la domestication.....31

I.2. Historique de la culture du tournesol.....31

I.3. Quelques généralités sur le tournesol.....32

I.4. Intérêt économique et production.....32

I.4.1. Le tournesol, une culture aux débouchés très diversifiés.32

I.4.1.1. Production d’huile alimentaire.....32

I.4.1.2. Utilisation des protéines en alimentation humaine.....33

I.4.1.3. Utilisation des tourteaux en alimentation animale.....33

I.4.1.4. Valorisation non alimentaire.....33

I.4.2. Importance économique.34

II. Une plante présentant de nombreux atouts pour l’étude de l’embryogenèse.....35

II.1. Un appareil reproducteur à l’anatomie avantageuse pour l’étude de l’embryogenèse.....35

II.2. L’embryogenèse somatique du tournesol : un modèle d’étude des phases d’initiation de l’embryogenèse.....35

II.3. Des outils disponibles en constant développement.....36

III. Etat actuel des connaissances sur le développement embryonnaire chez le tournesol.....36

III.1. Embryogenèse somatique et analyse du transcriptome lors de l’acquisition de la polarité embryonnaire.....36

III.2. Description de QTLs impliqués dans les processus du développement embryonnaire du tournesol.....37

III.3. Contrôle moléculaire de la mise en place des réserves au sein de l’embryon de tournesol.....38

III.3.1. Cinétique de l’accumulation des réserves au sein de l’embryon.....38

III.3.2. Modèle de synthèse des acides gras au sein de l’embryon de tournesol.....38

III.3.3. Les hélianthinines, principales protéines de réserves du tournesol.....39

Chapitre III – Intérêts et objectifs du projet de thèse.....40

PARTIE 2 - RESULTATS ET DISCUSSION

Chapitre I – Approche EST : Création de ressources et analyses d’expression génique *in silico*.....42

I. Choix méthodologiques.....42

I.1. Choix du matériel végétal: une stratégie « stades de développement » préférée à « jours après pollinisation ».....42

I.2. La technologie SMART est utilisée pour pallier aux faibles quantités de matériel végétal.....43

I.3. Les banques d’ADNc de référence : un échantillonnage de l’ensemble des gènes exprimés.....44

I.4. Les banques SSH : enrichissement en gènes faiblement et préférentiellement exprimés pour un stade de développement.....45

II. Approche fonctionnelle de l’embryogenèse précoce du tournesol par analyse.....46

II.1. Article: Comparative analysis of early embryonic sunflower cDNA libraries.....46

II.2. Analyse de la banque SSH enrichie en transcrits spécifiques d’embryons au stade coeur.....64

III. Difficultés rencontrées pour l’obtention des données EST.65

III.1. Un fort taux de redondance des banques HaGlbR, HaHeaR et HaHeaS.....65

III.2. Existence de plusieurs facteurs expérimentaux limitant le rendement du séquençage.....66

Chapitre II – Approche ‘Microarrays’ : Analyse simultanée de l’expression de 8 185 gènes lors de l’embryogenèse précoce du tournesol.....68

I. Choix méthodologiques et mise en œuvre technique.....69

I.1. Construction d’un microarray ADNc sur membrane nylon.....69

I.1.1. Raisons du choix de la technique employée.....69

I.1.2. Construction du microarrays.....70

I.1.2.1. Construction d’une puce générale plutôt qu’une puce thématique.....70

I.1.2.2. Sélection de l’EST la plus représentative pour chaque contig.....71

I.1.2.3. Choix des témoins.....71

I.2. Elaboration d’un plan d’expérience performant pour faciliter l’analyse des données.....72

I.2.1. Sélection des échantillons à tester : Stades embryonnaires d’intérêt et témoins.....72

I.2.2. Importance des répétitions des mesures.....72

I.2.3. Carré Latin Incomplet : Intérêt et puissance du dispositif expérimental utilisé.73

I.3. L’analyse statistique des données est basée sur deux ANOVA interconnectées.....73

I.3.1. Contrôle de la qualité des microarrays et de la fiabilité des résultats des hybridations.....	74
<i>I.3.1.1. Contrôle de la qualité des microarrays grâce à l' hybridation oligos-vecteurs.....</i>	<i>74</i>
<i>I.3.1.2. Evaluation de la reproductibilité entre hybridations.....</i>	<i>74</i>
I.3.2. La normalisation des données : Identification et retrait de la variabilité expérimentale.....	75
<i>I.3.2.1. Normalisation à l'aide de gènes de ménage.....</i>	<i>75</i>
<i>I.3.2.2. Normalisation à l'aide de témoins externes (spikes).....</i>	<i>75</i>
<i>I.3.2.3. Normalisation globale à l'aide de tous les gènes présents sur la puce.....</i>	<i>76</i>
<i>I.3.2.4. Normalisation par régression et normalisation non linéaire.....</i>	<i>76</i>
I.3.3. Identification des gènes différenciellement exprimés.....	76
<i>I.3.3.1. Définition d'un fold change, une méthode peu spécifique et peu sensible.....</i>	<i>76</i>
<i>I.3.3.2. Visualisation de ratios inhabituels, une méthode très utilisée mais biaisée.....</i>	<i>77</i>
<i>I.3.3.3. Analyse de variance.....</i>	<i>77</i>
I.3.4. Méthode d'analyse statistique choisie pour notre étude.....	77
I.3.5. Classification des gènes différentiels en fonction des profils d'expression.....	79
<i>I.3.5.1. Méthodes de mesures de similarité.</i>	<i>79</i>
<i>I.3.5.2. Méthodes de clustering.....</i>	<i>80</i>
I.4. Validation des gènes différentiels par RT-PCR quantitative, une méthode rapide et très sensible.	80
II. Vers l'élaboration d'un modèle fonctionnel de l'embryogenèse précoce chez le tournesol.	81
Article – Towards a functional analysis of early embryogenesis in exalbuminous dicotyledonous : Comparison of transcriptional profiles of developing sunflower embryos using a 10K cDNA microarray.....	82
Chapitre III – ‘Collection EST’ – ‘Analyse Microarrays’ : Atouts et contraintes des deux approches transcriptomiques utilisées.....	109
I. La collection d'EST : une ressource précieuse mais ne donnant qu'un premier aperçu du profil fonctionnel.....	109
I.1. Une avancée vers une meilleure connaissance du génome du tournesol.....	109
I.2. Une ressource précieuse pour l'étude de l'embryogenèse chez les plantes dicotylédones exalbuminées.....	110
I.3. Les limites de l'approche EST.....	110
II. L'approche microarrays fournit une information plus complète et plus précise.....	111
II. 1. Vers la compréhension globale des mécanismes moléculaires impliqués lors de l'embryogenèse précoce du tournesol grâce à l'utilisation des microarrays.....	111
II. 2. Amélioration de la découverte de gènes clés du développement embryonnaire végétal.	112
II. 3. Construction d'un outil à la disposition de la communauté scientifique.....	113

PARTIE 3 - PERSPECTIVES : VALIDATION POST-GENOMIQUE

Chapitre I – Nécessité de développer des outils bioinformatiques adaptés à la gestion de large volume de données et à la découverte de gènes candidats pertinents.....115

I. Développement de programmes informatiques d'analyse plus performants.....115

II. Nécessité d'une mutualisation des données disponibles et d'une uniformisation de l'annotation fonctionnelle des séquences.....115

II.1. Archivage des données d'expression génique au sein de bases de données publiques.....115

II.2. Vers une uniformisation de l'annotation fonctionnelle des séquences.....116

Chapitre II – Perspectives concernant l'étude des mécanismes moléculaires contrôlant l'embryogenèse précoce chez le tournesol.....118

I. Vers la découverte de nouveaux gènes et mécanismes clés : Elargissement du nombre de gènes et de conditions testées.118

II. Analyse moléculaire et fonctionnelle des gènes candidats identifiés : Exemple du gène MAGO NASHI.....119

II.1. Présentation de MAGO NASHI : principales caractéristiques moléculaires.....119

II.1.1. MAGO NASHI : une protéine hautement conservée au sein des différentes espèces animales et végétales.119

II.1.2. MAGO NASHI : une protéine ubiquitaire.....120

II.2. Fonctions de MAGO NASHI chez les animaux.....120

II.2.1. MAGO NASHI : une protéine essentielle au développement de l'oocyte chez la Drosophile.....120

II.2.1.1. MAGO NASHI est impliqué dans la formation des axes antéro-postérieur et dorso-ventral de l'oocyte de Drosophile.120

II.2.1.2. MAGO NASHI permet l'ancrage de déterminants maternels au pôle postérieur de l'oocyte et l'assemblage du pôle germinatif chez la Drosophile.....121

II.2.2. Rôles de MAGO NASHI chez d'autres espèces animales.....122

II.3. Mécanismes d'action de MAGO NASHI chez les animaux.....122

II.3.1. MAGO NASHI : un constituant d'un complexe ribonucléoprotéique (mRNP).....122

II.3.2. Régulation post-transcriptionnelle de l'activité de MAGO NASHI.....123

II.4. Très peu de données sont disponibles concernant les fonctions des homologues végétaux de MAGO NASHI.....	124
II.5. Vers la caractérisation de l'homologue de MAGO NASHI du tournesol : Démarches expérimentales entreprises et résultats préliminaires obtenus.....	124
<i>II.5.1. Vers la connaissance de l'organisation génomique du gène MAGO NASHI chez le tournesol : recherche du nombre de copies du gène MAGO NASHI chez le tournesol.....</i>	<i>124</i>
<i>II.5.1.1. Criblage des banques d'ADNc et recherche d'EST in silico.</i>	<i>124</i>
<i>II.5.1.2. Des expériences de Southern Blots montrent qu'il existe plus d'une copie du gène.....</i>	<i>125</i>
<i>II.5.1.3. Criblage des banques BAC pour accéder à la séquence génomique des différentes copies du gène.....</i>	<i>125</i>
<i>II.5.1.4. Modèle putatif de l'organisation génomique de MAGO NASHI chez le tournesol.....</i>	<i>127</i>
II.5.2. Vers la compréhension de la régulation transcriptionnelle de MAGO NASHI chez le tournesol..	127
<i>II.5.2.1. Obtention de profils d'expression spécifiques pour chacune des copies du gène.....</i>	<i>127</i>
<i>II.5.2.2. Caractérisation des éléments régulateurs des promoteurs des différentes copies du gène.....</i>	<i>128</i>
II.5.3. Démarches entreprises pour tenter d'élucider la fonction et le mécanisme d'action de MAGO NASHI chez les végétaux.....	128
<i>II.5.3.1. Recherche de mutants végétaux du gène MAGO NASHI.....</i>	<i>128</i>
<i>II.5.3.2. Recherche de l'homologue du tournesol de TSUNAGI-Y14, le principal partenaire de MAGO NASHI.....</i>	<i>129</i>
<i>II.5.3.3. Perspectives pour tenter d'élucider le rôle de MAGO NASHI dans l'embryogenèse végétale.....</i>	<i>130</i>

Chapitre III - Extension du modèle.....132

I. Mise en relation des données d'expression avec les connaissances génétiques.....132

II. Confrontation 'Transcriptome-Protéome-Métabolome' pour mettre en évidence les mécanismes de régulation transcriptionnelle et post-transcriptionnelle.....132

III. Transposition du modèle aux autres plantes dicotylédones.....133

PARTIE IV - MATERIELS ET METHODES

Chapitre I – Matériel Biologique.....134

I. Matériel Végétal: Helianthus annuus L. Ha300b.....134

II. Souches bactériennes.....	134
Chapitre II – Méthodes.	135
I. Méthodes générales de biologie moléculaire.....	135
I.1. Extraction d'ARN.....	135
I.2. Préparation de bactéries compétentes <i>E. Coli</i> DH5α et électroporation.....	135
I.3. RT-PCR quantitative.....	136
II. Construction et analyses des banques d'ADNc.....	136
II.1. Construction des banques d'ADNc de référence: HaGlbR, HaHeaR, HaCotR.....	136
<i>II.1.1. Obtention des ADNc à cloner.....</i>	<i>136</i>
<i>II.1.2. Clonage.....</i>	<i>137</i>
II.2. Construction de la banque SSH: HaHeaS.....	137
<i>II.2.1. Obtention des ADNc à cloner.....</i>	<i>137</i>
<i>II.2.2. Hybridations entre ADNc driver et tester</i>	<i>138</i>
<i>II.2.3. Amplification des ADNc préférentiellement exprimés dans la condition d'intérêt.....</i>	<i>139</i>
<i>II.2.4. Clonage.....</i>	<i>139</i>
II.3. Normalisation des banques redondantes.....	140
<i>II.3.1. Préparation des macroarrays à partir des banques d'ADNc.....</i>	<i>140</i>
<i>II.3.2. Hybridation des macroarrays avec des amorces overgos spécifiques des séquences redondantes.</i>	<i>141</i>
<i>II.3.2.1. Marquage radioactif des sondes overgos.....</i>	<i>141</i>
<i>II.3.2.2. Hybridation des macroarrays.</i>	<i>141</i>
<i>II.3.2.3. Lavages des membranes.</i>	<i>141</i>
<i>II.3.3. Détection des clones redondants.....</i>	<i>142</i>
II.4. Séquençage des clones.	142
<i>II.4.1. Extraction de l'ADN plasmidique.</i>	<i>142</i>
<i>II.4.2. Réactions de séquences.</i>	<i>142</i>
<i>II.4.3. Purification des réactions de séquences.</i>	<i>143</i>
<i>II.4.4. Séquençage.</i>	<i>143</i>
II.5. Analyses bioinformatiques.	143
III. Construction et analyse des microarrays.....	144
III.1. Construction des microarrays.	144
<i>III.1.1. Sélection et réarrangement des clones bactériens.....</i>	<i>144</i>
<i>III.1.2. Obtention des produits PCR et vérification.....</i>	<i>144</i>
<i>III.1.2.1. Conditions PCR</i>	<i>144</i>

<i>III.1.2.2. Vérification des produits PCR.....</i>	144
<i>III.1.2.3. Evapoconcentration des produits PCR et réarrangements.</i>	145
III.1.3. Spotting des microarrays.	145
III.1.4. Traitement de fixation des produits PCR sur le microarrays.....	145
III.2. Hybridation des microarrays.....	145
<i>III.2.1. Préparation des sondes complexes d'ADNc marquées.....</i>	145
<i>III.2.2. Préparation des sondes oligonucléotides.....</i>	146
<i>III.2.3. Conditions d'hybridation des microarrays.....</i>	146
<i>III.2.3.1. Description du dispositif expérimental.</i>	146
<i>III.2.3.2. Conditions d'hybridation des microarrays.</i>	146
<i>III.2.3.3. Déshybridation des microarrays.</i>	147
<i>III.2.4. Analyse des données d'hybridation des microarrays.</i>	147
<i>III.2.4.1.Obtention et traitement des images d'hybridation.</i>	147
<i>III.2.4.2. Analyse statistique de l'expression des gènes.</i>	147
IV. Etude de l'organisation génomique du gène MAGO NASHI.....	148
IV.1. Southern Blots.....	148
<i>IV.1.1. Extraction d'ADN génomique.....</i>	148
<i>IV.1.2. Digestion de l'ADN par des enzymes de restriction.....</i>	148
<i>IV.1.3. Transfert de l'ADN génomique digéré sur membrane de nylon.....</i>	149
<i>IV.1.4. Hybridation des membranes.....</i>	149
<i>IV.1.4.1. Obtention et marquage radioactif de la sonde.....</i>	149
<i>IV.1.4.2. Conditions d'hybridation.</i>	149
<i>IV.1.4.3. Lavage des membranes.</i>	149
IV.2. Sous-clonage et séquençage de clones BAC.	150
REFERENCES BIBLIOGRAPHIQUES.....	151

ANNEXES

RESUME

Chez les végétaux supérieurs, l'embryogenèse est une phase clé du développement au cours de laquelle l'embryon établit les principales structures qui formeront la future plante et synthétise et accumule des réserves définissant le rendement et la qualité nutritionnelle des graines. La compréhension des événements moléculaires et physiologiques menant à la formation de la graine est donc d'un intérêt agronomique majeur. Le tournesol (*Helianthus annuus L.*), qui présente une inflorescence portant des embryons de taille relativement grande et dont le développement est synchronisé et un génome maintenant bien caractérisé, constitue un bon système expérimental pour l'étude de l'embryogenèse zygotique chez les dicotylédones.

Afin de caractériser le transcriptome au cours de l'embryogenèse précoce chez le tournesol, des banques d'ADNc de référence ont été construites à partir d'embryons au stade globulaire, cœur et cotylédonaire qui constituent des stades clé du développement. Une banque d'ADNc de type SSH (Suppressive Subtractive Hybridization) comparant les transcrits d'embryons aux stades cœur et cotylédonaire et enrichie en gènes préférentiellement exprimés au stade cœur a également été réalisée. Au total, 23 800 clones ont été générés dont 4 118 ont été séquencés et 1 690 séquences EST publiées (numéro d'accension : AJ827751-AJ829440). Au cours d'analyses *in silico*, les séquences obtenues au sein de notre laboratoire à partir des banques de référence ont été couplées à d'autres données EST issues de banques d'ADNc produites à partir d'ovules non fécondés (HadevR7) ou d'embryons cotylédonaires à des stades plus tardifs (HaDevR5 et HaDevR6) réalisées dans le cadre du programme Génoplante. L'ensemble des séquences analysées représente un total de 7 106 EST et un unigène de 3064 séquences (Ben et al., 2005). La caractérisation fonctionnelle des séquences uniques de chacune des banques d'ADNc a permis d'avoir une vue globale des fonctions cellulaires impliquées à chaque stade de développement et de leur importance relative. Les profils fonctionnels, très similaires pour chacun des stades de développement étudiés, ont montré que les embryons zygotiques sont engagés dans des processus de différenciation et présentent une activité cellulaire très élevée. Des analyses de northern virtuels ont révélé des gènes différenciellement exprimés entre les différentes conditions. Des séquences montrant de fortes similarités avec des gènes jouant des rôles clés dans l'embryogenèse animale ou végétale ont également pu être identifiées comme, par exemple, le gène MAGO NASHI participant à l'élaboration des axes embryonnaires chez la *Drosophile* et dont la caractérisation moléculaire et fonctionnelle chez le tournesol a été entreprise au laboratoire.

Par la suite, ces banques d'ADNc, associées à d'autres banques d'ADNc préparées à partir de tissus d'origines très diverses, ont été utilisées pour construire un microarray sur membrane nylon comptant 8 185 séquences uniques. Ce microarray a été hybridé avec des ADNc issus d'embryons dans les trois stades de développement d'intérêt (globulaire, cœur et cotylédonaire) ainsi que des ADNc préparés à partir d'ovules non fécondés et de feuilles, ces derniers échantillons représentant respectivement un témoin de tissus reproducteurs non fécondés et un témoin de tissus somatiques. L'analyse statistique des résultats de ces expériences, réalisée en pratiquant deux anovas interconnectées, a révélé 744 gènes différenciellement exprimés entre aux moins deux conditions. Une analyse plus fine, effectuée après avoir défini 4 contrastes particulièrement intéressants (embryons globulaires vs embryons cœur, embryons précoces (globulaire et cœur) vs embryons cotylédonaires, embryons (globulaire, cœur et cotylédonaire) vs ovules non fécondés et tissus reproductifs (embryons et ovules) vs feuille), a mis en évidence de 106 gènes différenciellement exprimés entre les embryons et les ovules jusqu'à 437 entre les feuilles et l'ensemble des tissus reproducteurs. L'étude du profil d'expression de ces gènes nous a permis de dégager des hypothèses sur les mécanismes induits au cours des phases précoces de l'embryogenèse chez les plantes impliquant notamment une régulation transcriptionnelle et post-transcriptionnelle fine et spécifique à chacun des stades de développement et diverses voies de transduction du signal. La biogenèse de plastes fonctionnels ainsi que la sénescence semble également jouer un rôle essentiel dans le déroulement normal de l'embryogenèse.

Mots clés : *Helianthus annuus L.*, embryogenèse, EST, banques d'ADNc, microarray, analyse d'expression de gènes.

PARTIE I

ETUDE BIBLIOGRAPHIQUE



Figure 1: Le croissant fertile, berceau de l'agriculture.

(d'après fr.wikipedia.org/wiki/Croissant_fertile)

: croissant fertile

Damascus : oasis de Damas

Chapitre I - L'embryogenèse chez les végétaux supérieurs.

I. L'embryogenèse : une étape clé du développement des plantes.

I.1. Une phase essentielle du développement.

Chez les végétaux supérieurs, l'embryogenèse est l'étape du développement située entre la fécondation de l'oeuf et les étapes de dormance et de germination de la graine aboutissant à l'obtention d'une nouvelle plante. Comme pour les animaux, ce processus joue un rôle fondamental dans le cycle de reproduction des plantes à fleurs. A l'issue de ce processus, l'embryon mature contenu dans la graine représente l'image réduite de la plantule qui se développera après germination. Ainsi, au cours de l'embryogenèse, les potentialités du futur individu sont définies : l'embryon acquiert un patron morphologique correct conditionnant le plan d'organisation de la plante adulte, avec la mise en place d'organes fonctionnels et de tissus différenciés, et accumule des réserves essentielles au bon déroulement de la germination et au développement de la future plantule. De la bonne exécution du programme de l'embryogenèse dépendra donc la vigueur de la future plante.

I.2. La graine : un organe végétal d'un intérêt agronomique capital.

Il y a environ 10 000 ans, les hommes, ayant probablement observé qu'une graine tombée sur le sol générerait une nouvelle plante, prirent l'initiative de récolter des graines et de les semer afin de renouveler ce qu'ils consommaient. De chasseurs-cueilleurs puisant leur subsistance dans la nature, ils deviennent des producteurs. Débute alors l'une des aventures les plus marquantes de l'histoire de l'humanité: l'agriculture qui naît au Moyen-Orient dans la région du Croissant Fertile, les premiers signes de domestication des céréales ayant été repérés chez les Khiamiens de l'oasis de Damas (Fig.1). Elle apparaît ensuite presque simultanément en Chine, au Sahara et dans la Cordillère des Andes. Avec l'apparition de l'agriculture, un processus de sélection artificielle des plantes cultivées, probablement initialement inconscient, s'est enclenché permettant la modification des caractères génétiques et ainsi l'augmentation de la productivité. Avec la domestication des végétaux, les graines, bien que depuis toujours utilisées dans l'alimentation humaine, deviennent alors un centre d'intérêt majeur pour l'homme, grâce à leur rôle d'organe de multiplication.

De nos jours, les graines constituent toujours le composant principal de l'alimentation mondiale et couvrent à elles seules 55% des besoins énergétiques journaliers de l'homme. Leur richesse provient de leur contenu en amidon, protéines et acides gras, substances emmagasinées au cours du développement embryonnaire dans le but de nourrir la plantule au cours de la germination. Le rendement et les qualités germinatives et nutritionnelles des graines, qui sont les principales cibles des sélectionneurs pour les plantes cultivées,

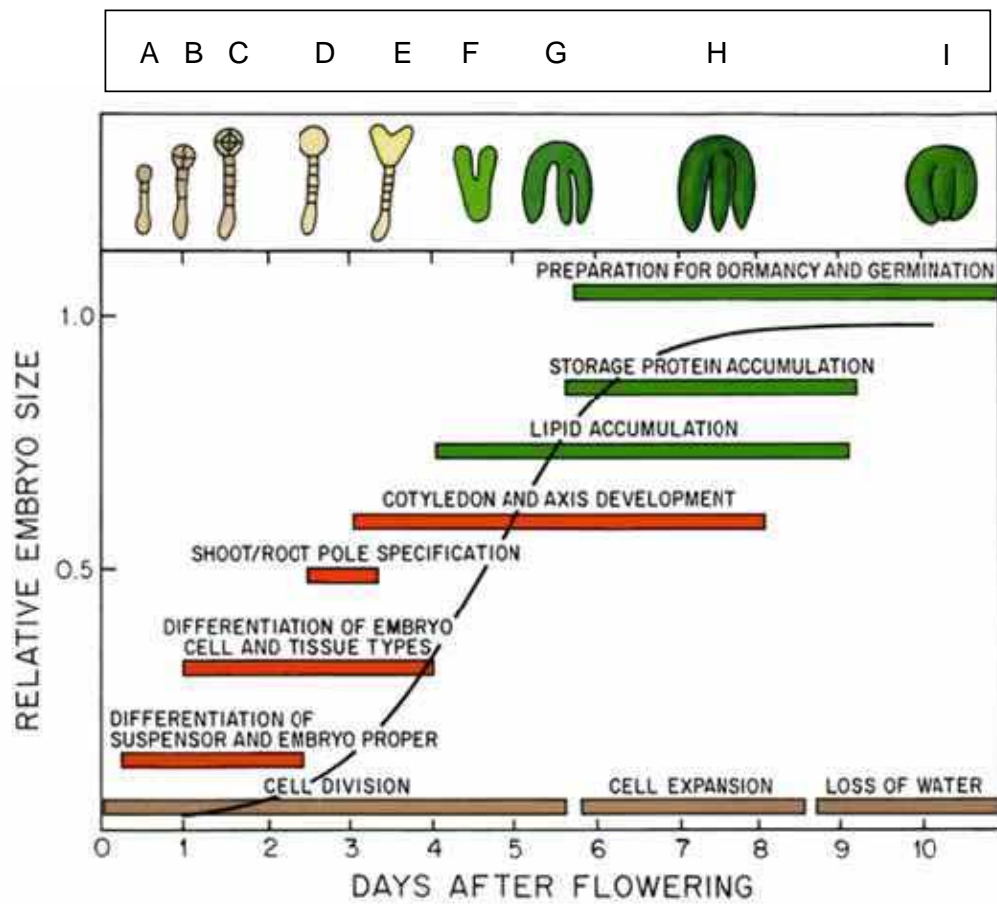


Figure 2: Différentes phases du développement embryonnaire d'*Arabidopsis thaliana*

(d'après <http://www.mcdb.ucla.eduResearch/Goldberg/research/interests-index.htm>)

- A = embryon une cellule
- B = Stade octant
- C = Stade dermatogène
- D = Stade globulaire
- E = Stade cordiforme
- F = Stade torpille
- G-I = Stade cotylédonaire

dépendent largement du développement de l'embryon. L'embryogenèse végétale est donc d'un intérêt agronomique tout particulier et d'une importance capitale pour l'économie.

Outre leur utilisation pour l'alimentation humaine, les graines sont aussi employées pour la nutrition animale ou l'industrie (lipochimique, pharmaceutique, cosmétique...).

I.3. La séquence de développement de la graine comporte trois phases clés.

Du fait de son importance biologique et de son intérêt économique, l'embryogenèse végétale a été largement étudiée. Des études anatomiques et physiologiques détaillées ont été menées. A l'aide de critères morphologiques et physiologiques, l'embryogenèse des végétaux supérieurs a pu être divisée en trois phases qui se déroulent de manière successive lors du développement normal de la graine (Fig.2):

- La **phase d'embryogenèse précoce**, depuis le zygote jusqu'au stade coeur, correspond à une étape de *morphogenèse* et consiste en l'établissement du plan d'organisation de la plantule qui résulte de la superposition d'un axe apico-basal correspondant à la mise en place des différents organes et d'un axe radial reflétant la formation des différents tissus.

- La **phase de maturation** de l'embryon qui s'étend jusqu'au stade cotylédonaire est caractérisée par l'*accumulation de substances de réserves* dans les cotylédons, l'hypocotyle et/ou l'albumen. Ces substances seront utilisées par la jeune plantule lors de la germination. Cette étape se traduit également par la croissance des différents organes de l'embryon.

- Enfin, la **phase de dessiccation** de la graine et de l'embryon, au cours de laquelle la teneur en eau des tissus chute de 90%, prépare l'embryon à la phase de dissémination. Cette déshydratation, qui met en jeu le métabolisme de l'acide abscissique (ABA), s'accompagne d'un fort ralentissement de l'activité physiologique de la graine qui entre en dormance (état de repos végétatif). Elle présente ainsi une capacité de résistance exceptionnelle aux facteurs physiques et chimiques de l'environnement et peut se maintenir plusieurs années dans cet état. La graine, pourvue des réserves qui lui permettront de reprendre sa croissance en présence d'eau et d'oxygène, germe lorsque les conditions extérieures sont favorables.

Dans la suite de ce mémoire, nous nous intéresserons plus particulièrement aux phases de morphogenèse et de maturation de l'embryogenèse.

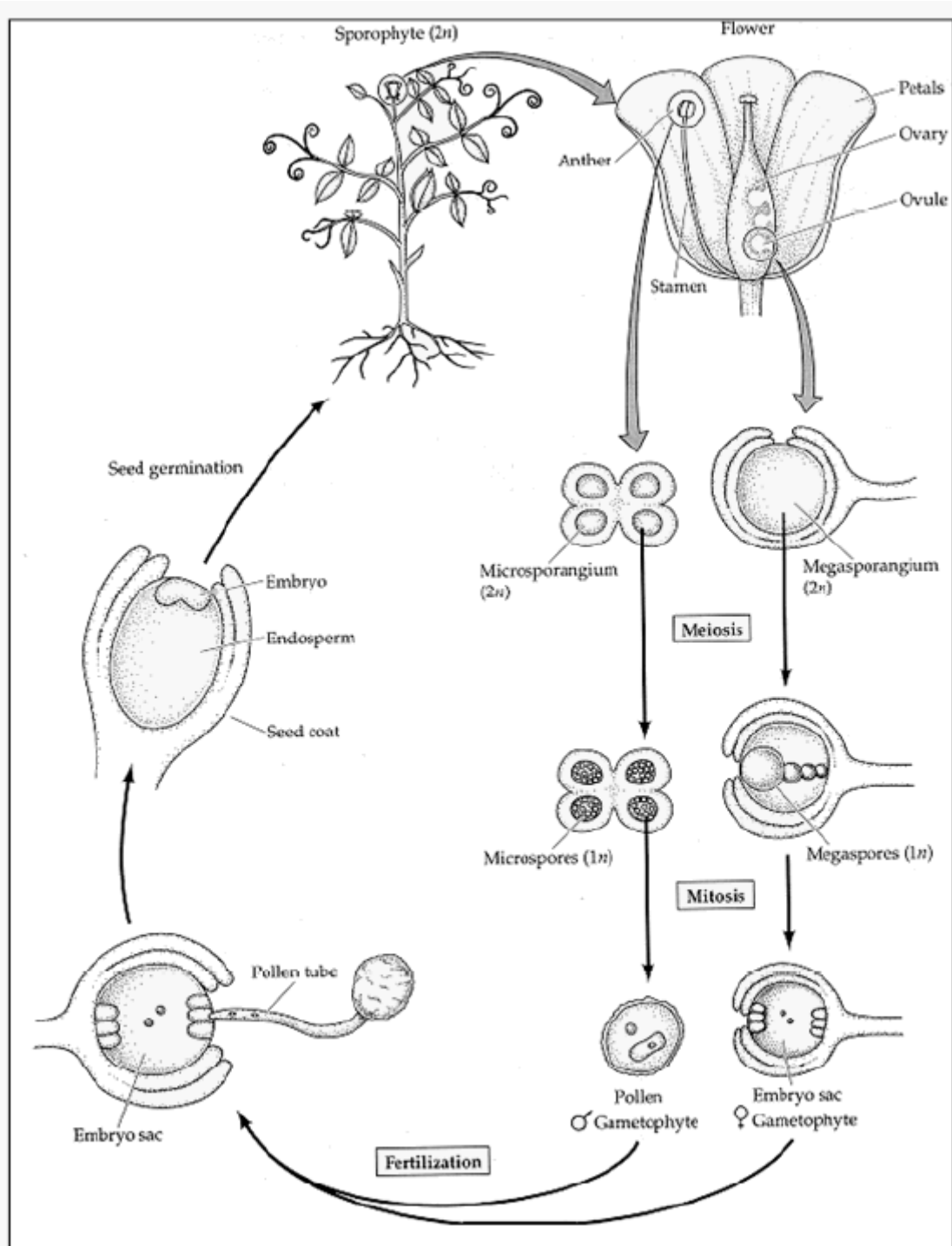


Figure 3: Cycle de développement des Angiospermes (plantes à fleur).

(d'après <http://zygote.swarthmore.edu/phyto1.html>)

II. Description histologique et morphologique de l'embryogenèse des Dicotylédones.

La formation des spores et des gamètes ainsi que le développement de l'embryon sont aujourd'hui bien caractérisés au niveau morphologique grâce notamment à des analyses histologiques et microscopiques. La capselle (*Capsella bursa-pastoris*, *Brassicaceae*) a longtemps servi de modèle du développement embryonnaire chez les Dicotylédones (Schulz and Jensen, 1968). Son étude a notamment permis de caractériser les plans de divisions du zygote ainsi que les stades embryonnaires successifs, en fonction du nombre de cellules dérivées de la cellule apicale (2 cellules, quadrant, octant) ou de la forme de l'embryon (globulaire, triangulaire, coeur, torpille, cotylédonaire...). Les progrès réalisés en génétique et en biologie moléculaire font maintenant d'*Arabidopsis thaliana* et du maïs (*Zea mays*) des modèles de choix pour l'étude de l'embryogenèse zygotique chez les dicotylédones et les monocotylédones, respectivement. Ainsi, les informations fournies dans ce paragraphe sur l'embryogenèse des dicotylédones concernent principalement le développement embryonnaire d'*Arabidopsis thaliana* qui reste actuellement le mieux décrit et qui implique une séquence de divisions relativement invariante et bien conservée chez la plupart des plantes dicotylédones.

Le cycle biologique des plantes à fleurs alterne entre deux phases : une phase gamétophytique haploïde et une phase sporophytique diploïde (Fig.3). La phase gamétophytique correspond à la production des gamètes: à la suite de la méiose, les spores forment le grain de pollen et le sac embryonnaire. La phase sporophytique débute avec la formation du zygote issu de la fusion des gamètes, comprend les étapes d'embryogenèse et de germination de la graine, se poursuit avec le développement de la plante mature et l'établissement des organes végétatifs et reproducteurs et s'étend jusqu'à la production des spores dans les organes reproducteurs.

Dans ce paragraphe, la période comprise entre la germination de la graine et la formation de la fleur de la phase sporophytique ne sera pas présentée.

II.1. La fécondation: étape d'initiation de l'embryogenèse.

II.1.1. Les acteurs de la fécondation.

Chez les Angiospermes, la reproduction sexuée se déroule au sein de la fleur (Fig.3) qui réunit :

- les pièces fertiles mâles constituées par les étamines composées du filet et de l'anthère,
- les pièces fertiles femelles constituées par le pistil composé de carpelles contenant l'ovaire à l'intérieur duquel sont enfermés les ovules,
- les pièces stériles constituées des sépales et des pétales qui ont un rôle protecteur et, dans le cas de pollinisation entomophile, un rôle attracteur pour les insectes.

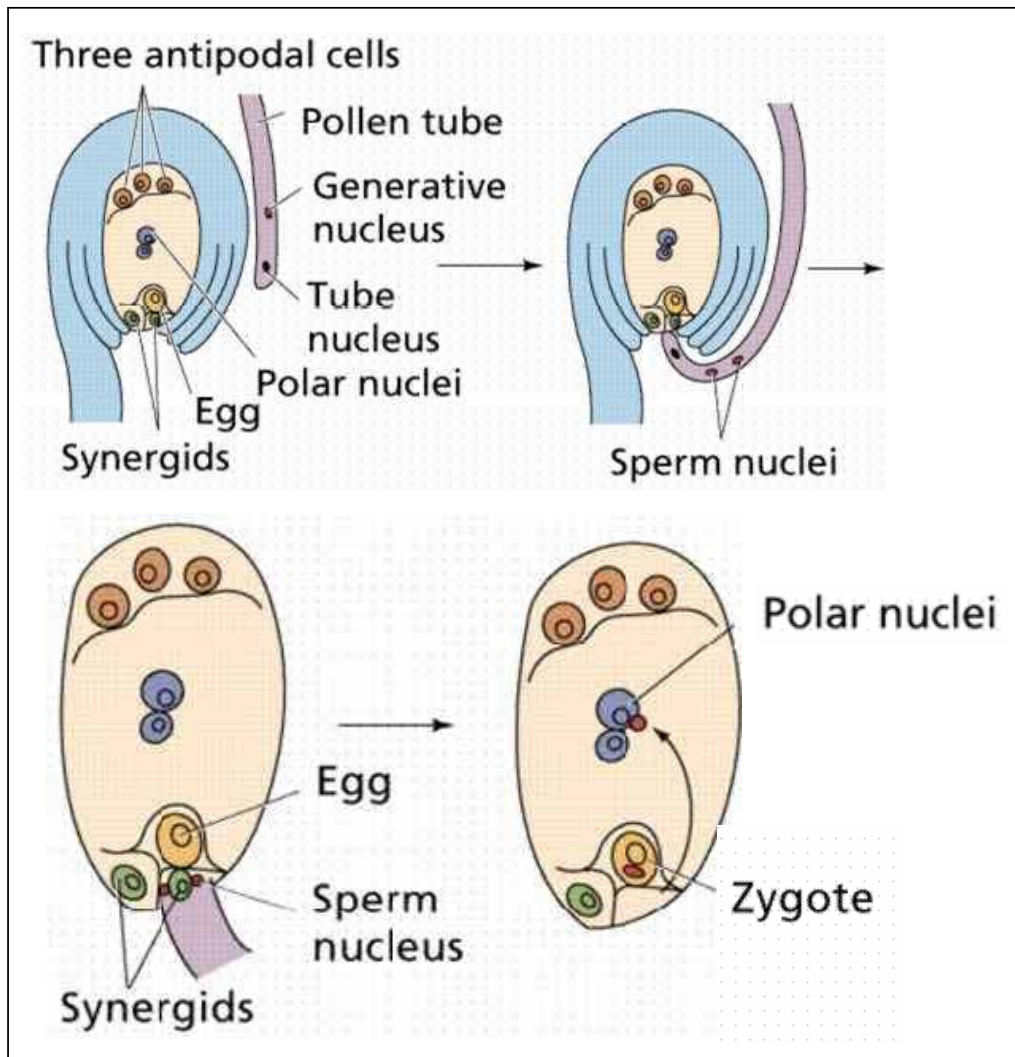


Figure 4: Déroulement de la double fécondation.

(d'après <http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookflowersII.html>)

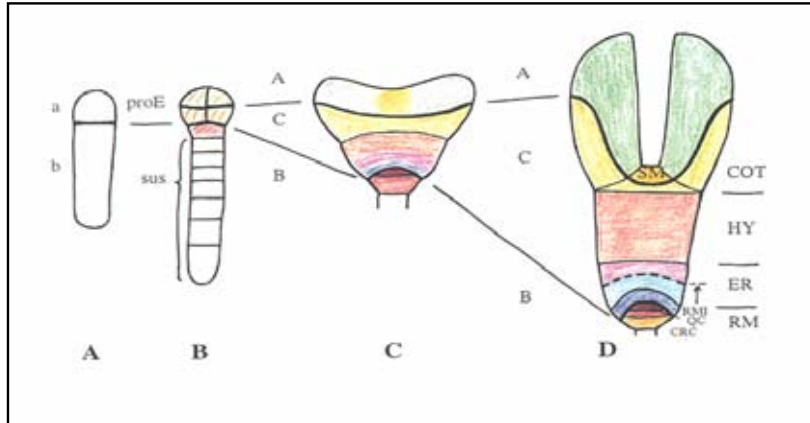
Les gamétophytes mâles ou grains de pollen sont produits au sein des anthères alors que le gamétophyte femelle ou sac embryonnaire est formé dans le pistil (Fig.3). Les sporocytes, ou cellules sporogènes, contenus dans les sacs polliniques des anthères et au sein du nucelle de l'ovule subissent une méiose aboutissant à la formation de 4 spores haploïdes appelées tétraspores ou microspores pour l'appareil reproducteur mâle et mégaspores pour l'appareil reproducteur femelle (Yang et Sundaresan, 2000). Au sein des organes reproducteurs mâles, chaque tétraspoire subit une maturation qui la transforme en un grain de pollen : chacune génère par mitose une cellule générative renfermant deux noyaux reproducteurs (gamètes mâles) intervenant dans la fécondation et une cellule végétative accumulant des réserves (amidon, lipides et protéines) qui serviront ultérieurement à construire le tube pollinique.

En revanche, chez la plupart des Dicotylédones et chez quelques Monocotylédones, suite à la méiose femelle, trois des quatre mégaspores avortent alors que l'unique spore survivante évolue pour produire le gamétophyte femelle. A maturité, le sac embryonnaire contient 8 noyaux issus de trois mitoses successives de la mégaspore et sept cellules localisées de manière polarisée : du côté du micropyle, se trouve une triade comprenant une grosse cellule centrale, l'oosphère (gamète femelle), et deux cellules latérales, les synergides; du côté du hile, sont situées trois antipodes et au centre du sac, la cellule centrale contenant deux noyaux polaires. Le sac embryonnaire est entouré de deux tissus sporophytiques, les intégruments internes et externes, qui pourraient jouer un rôle dans le développement de l'embryon et de l'albumen.

II.1.2. La double fécondation : une originalité des Angiospermes.

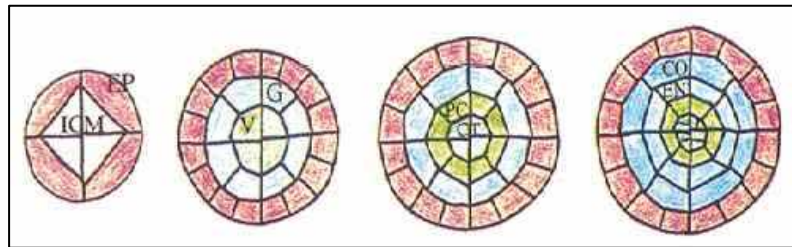
La double fécondation a été décrite dès 1898 par le russe Navashin (Navashin, 1898). A la suite de la libération des grains de pollen après déhiscence de l'anthère et à leur dépôt sur le stigmate de la fleur (pollinisation) par le biais d'agents les plus divers (principalement le vent ou les insectes), on assiste à la germination du grain de pollen. Après hydratation, la cellule végétative développe le tube pollinique qui traverse le style, atteint l'ovule dans la cavité ovarienne puis pénètre dans le sac embryonnaire par le micropyle (Fig.4). La cellule germinative déverse ensuite les deux noyaux reproducteurs haploïdes dans l'une des cellules synergides. Se produit alors la double fécondation au cours de laquelle l'un des noyaux spermiques fusionne avec le noyau haploïde de l'oosphère formant ainsi le zygote diploïde et le second noyau mâle s'unit aux deux noyaux polaires haploïdes de la cellule centrale pour former l'albumen, un tissu triploïde qui a pour rôle de fournir des nutriments à l'embryon et/ou à la plantule. L'hypothèse d'une programmation des cellules spermiques pour féconder la cellule œuf ou l'endosperme, s'appuyant notamment sur l'existence d'un dimorphisme des cellules spermiques chez plusieurs espèces (Saito et al., 2000), reste encore à vérifier.

Après la double fécondation, le pistil évolue en fruit tandis que le zygote, qui suit le programme de développement de l'embryogenèse, se transforme en graine.



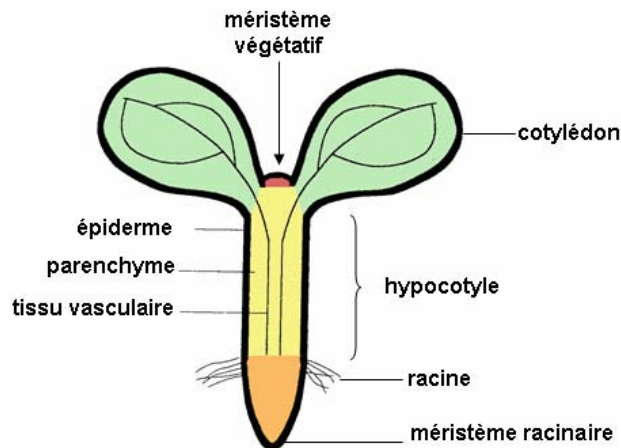
A: Mise en place de l'axe apico-basal (d'après Jürgens, 1995).

proE: proembryo, sus: suspensor, RMI: Root Meristem Initials, QC: Quiescent Center, CRC: Central Root Cap, COT: Cotyledons, HY: hypocotyl, ER: Embryonic Root, RM: Root Meristem, SM: Shoot Meristem



B: Formation de l'axe radial (d'après Jürgens, 1995).

EP: epidermis, ICM: Inner Cell Mass, G: Ground Tissue, V: Vasculated Primordium, PC: Pericycle, CT: Conductive Tissue, Co: Cortex, EN: Endodermis



C: Représentation schématique du plan d'organisation de la plantule selon les axes apico-basal et radial.

Figure 5: Formation des axes apico-basal et radial au cours de l'embryogenèse.

II.2. Mise en place des axes apico-basal et radial de l'embryon.

Le patron morphologique embryonnaire s'organise autour de deux axes, l'axe apico-basal et l'axe radial (Fig.5), qui sont mis en place au cours de l'embryogenèse précoce (Jürgens, 2001). L'axe apico-basal est composé de différentes structures, véritables ébauches des organes de la future plante, qui sont, du haut vers le bas, le méristème apical, les cotylédons, l'hypocotyle, la radicule et le méristème racinaire. L'axe radial correspond à l'édification des différents types de tissus, organisés en couches concentriques autour de l'axe apico-basal, avec de la périphérie vers le centre, l'épiderme, le cortex, l'endoderme, le péricycle et les tissus vasculaires.

II.2.1. La première division asymétrique traduit la polarisation du zygote.

Au cours du cycle biologique des organismes multicellulaires tels que les végétaux supérieurs, une cellule unique, le zygote, est à l'origine d'un individu constitué de nombreux types cellulaires différenciés et spécialisés. Les divisions asymétriques qui génèrent deux cellules dont les contenus cytoplasmiques sont différents et qui résultent d'une polarisation de la cellule mère peuvent constituer l'un des mécanismes menant à de telles différenciations cellulaires.

II.2.1.1. Déroulement de la première division du zygote.

- *Arabidopsis thaliana*: l'exemple le mieux décrit.

Chez *Arabidopsis thaliana*, à la suite de la fécondation, le zygote unicellulaire s'allonge selon l'axe apico-basal. Cette étape d'élongation du zygote semble être un événement très important de l'embryogenèse. En effet, l'analyse phénotypique des mutants *yoda* a montré qu'un allongement trop important ou trop faible mène à des perturbations du développement de l'embryon ou du suspenseur (Lukowitz et al., 2004). Ensuite, le réseau de microtubules corticaux du zygote, jusqu'alors disposé en un maillage inorganisé, s'oriente en faisceaux parallèles perpendiculairement à cet axe. Le noyau, à l'origine situé en position centrale dans la cellule zygotique, migre vers la partie apicale alors qu'une large vacuole se développe dans la région basale. En pré-prophase, les microtubules se resserrent pour former un anneau autour du noyau. Cet anneau, appelé bande pré-prophasique, indique la position du futur plan de division du zygote. Le réarrangement des constituants cytoplasmiques reflète une polarisation du zygote qui aboutit à une division cellulaire asymétrique.

Le zygote se sépare transversalement en deux cellules de taille et de devenir très différents (West and Harada, 1993):

- une petite cellule apicale riche en cytoplasme qui sera à l'origine de la majeure partie de l'embryon proprement dit,

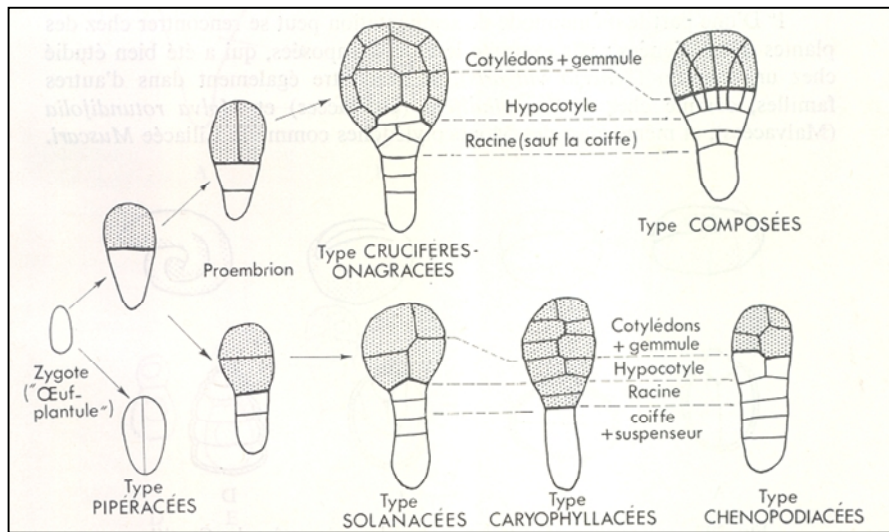


Figure 6: Principaux types d'organisation dans les premiers stades de l'embryogenèse chez les Dicotylédones (d'après Champagnat et al., 1969).

- et une grande cellule basale très vacuolisée qui formera principalement un suspenseur filamenteux de 6 à 9 cellules, qui jouerait le rôle de 'cordon ombilical' acheminant les nutriments et les facteurs de croissance puisés dans les tissus maternels et nécessaires au développement normal de l'embryon (Yeung et Meinke, 1993). Au cours du développement embryonnaire, le suspenseur dégénère à l'exception de la cellule apicale (hypophyse) qui génèrera une partie de la racine de l'embryon.

- **Autres types de première division zygotique.**

En comparaison du modèle de division zygotique d' *Arabidopsis thaliana* décrit ci-dessus, de nombreuses variations sont observées chez d'autres espèces végétales (Kaplan and Cooke, 1997). Chez les dicotylédones, six types de segmentation de l'embryon, associés à un changement de l'origine des différentes structures de l'axe apico-basal, ont pu être répertoriés (Fig.6). Contrairement à l'ensemble des autres dicotylédones dont la première division est transversale, le type Pipéracées est caractérisé par une première division verticale. Ainsi, la polarisation de l'embryon au cours de la première division asymétrique, qui constitue une étape de l'embryogenèse essentielle au bon déroulement du programme de développement du zygote, ne serait donc pas tant déterminée par l'orientation du plan de division ou par la différence de taille des cellules filles mais plutôt par une répartition asymétrique de constituants cytoplasmiques tels que des ARNm ou des protéines.

II.2.1.2. Hypothèses sur l'origine de la polarité du zygote.

Plusieurs hypothèses concernant l'origine de la polarité du zygote ont été émises. De manière générale, l'asymétrie et les différents potentiels développementaux peuvent être déterminés:

- de façon extrinsèque: les deux cellules issues de la division sont identiques mais subissent des stimuli environnementaux distincts entraînant différents devenir.
- de façon intrinsèque par le biais d'une distribution différentielle de certains déterminants de différenciation cellulaire.

Cependant, dans les deux cas, le processus de différenciation cellulaire requiert l'existence d'une polarité dans l'environnement ou au sein même de la cellule.

- **Une origine maternelle de la polarité du zygote ?**

Chez les Angiospermes, le sac embryonnaire et l'oosphère présentent une organisation polarisée suivant l'axe micropyle-chalaze (Fig.7) (Reiser and Fisher, 1993). En effet, au niveau du sac embryonnaire, les synergides et l'oosphère sont situées du côté du micropyle et les antipodes du côté chalazique. De même, l'oosphère non fécondée est relativement riche en cytoplasme du côté chalazique et est constituée d'une grande vacuole du côté micropylaire. L'embryon est lui-même orienté de manière polarisée et invariante au sein des tissus maternels environnants : l'extrémité apicale vers la chalaze et l'extrémité basale vers le micropyle. Le fait que la position de l'axe apico-basal de l'embryon soit calquée sur l'orientation de l'axe chalaze-micropyle du

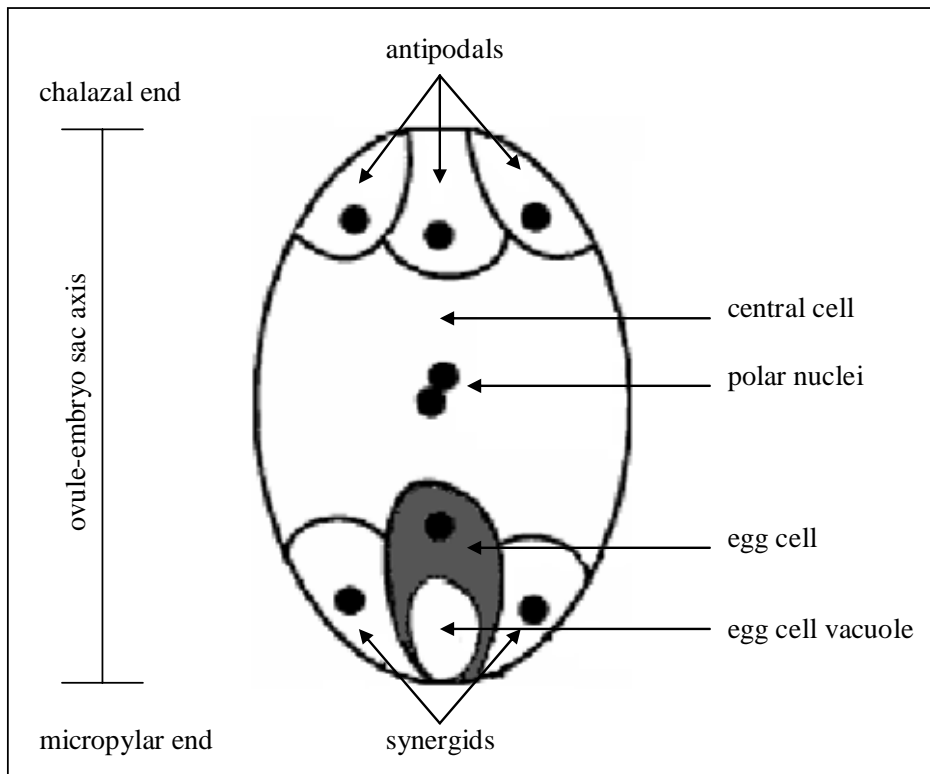


Figure 7: Structure polarisée du sac embryonnaire et de l'oosphère
(d'après Reiser and Fisher, 1993).

gaméophyte femelle peut laisser supposer que des facteurs positionnels issus de la phase haploïde du cycle pourraient servir de guide à l'asymétrie de l'embryon. Ainsi, la polarité du zygote pourrait être d'origine maternelle. Les déterminants de cette polarisation précoce au sein des tissus maternels n'ont pas encore été identifiés. Le gène SHORT INTEGUMENT1/DICER-LIKE1 (SIN1/DCL1) reste encore aujourd'hui l'un des seuls composants maternels décrit comme ayant une influence sur l'asymétrie embryonnaire (Ray et al., 1996). Le produit de ce gène, localisé dans le noyau, interviendrait, comme son homologue animal DICER, dans la production de micro-ARN potentiellement impliqués dans des mécanismes d'inhibition génique ou gene silencing (Papp et al., 2003). Ces molécules générées au niveau des tissus maternels pourraient agir comme signaux directs ou par le biais de mécanismes indirects pour réguler le développement embryonnaire.

- **La fécondation, un évènement accentuant la polarité du zygote ?**

Chez de nombreuses espèces végétales, le contenu cellulaire du zygote est réorganisé à la suite de la fécondation: le réticulum endoplasmique, les plastes et les mitochondries sont redistribués, ce qui accentue l'organisation polarisée préexistante de l'oosphère.

La polarité du zygote pourrait également résulter de l'expression de gènes induits suite à la fécondation. Ainsi, des phénomènes épigénétiques conduisant au silencing d'une grande partie du génome paternel durant les stades précoces de l'embryogenèse pourraient mener à la détermination exclusivement maternelle de la polarité du zygote (Vielle-Calzada et al., 2000) (cf PartI Ch.IV.1.1. p. 16).

- **Un gradient hormonal à l'origine de la polarité du zygote ?**

La polarité pourrait être induite par la présence d'un gradient hormonal (notamment d'un gradient d'auxine). Au cours de l'embryogenèse, deux gradients auxiniques, l'un intrinsèque à l'embryon et l'autre externe d'origine maternelle, pourraient être responsables de la transition entre la symétrie radiale et la symétrie bilatérale (Zimmermann, 1993). Alors que le gradient intrinsèque à l'embryon apparaîtrait au cours de l'embryogenèse, le gradient externe serait présent dès le début du développement embryonnaire. Le phénotype du mutant *gnom*, caractérisé par des défauts le long de l'axe apico-basal plus ou moins sévères liés à une altération du transport auxinique et visibles dès la première division zygotique, conforte l'hypothèse d'un rôle de l'auxine dans l'acquisition de la polarité embryonnaire (Steinmann et al., 1999).

- **Origine endogène au zygote de la polarité embryonnaire ?**

Chez les végétaux supérieurs, plusieurs cas de développement embryonnaire en l'absence de tissus maternels tels que le développement *in vitro* de zygotes issus de la fusion artificielle de gamètes chez le maïs (Dumas and Mogensen, 1993) ou d'embryons somatiques formés à partir de différents explants (protoplastes, microspores, explants tissulaires, cellules) (Zimmermann, 1993) ont été rapportés suggérant que des facteurs maternels ne sont pas obligatoirement requis pour l'établissement de l'asymétrie de l'embryon. La polarité

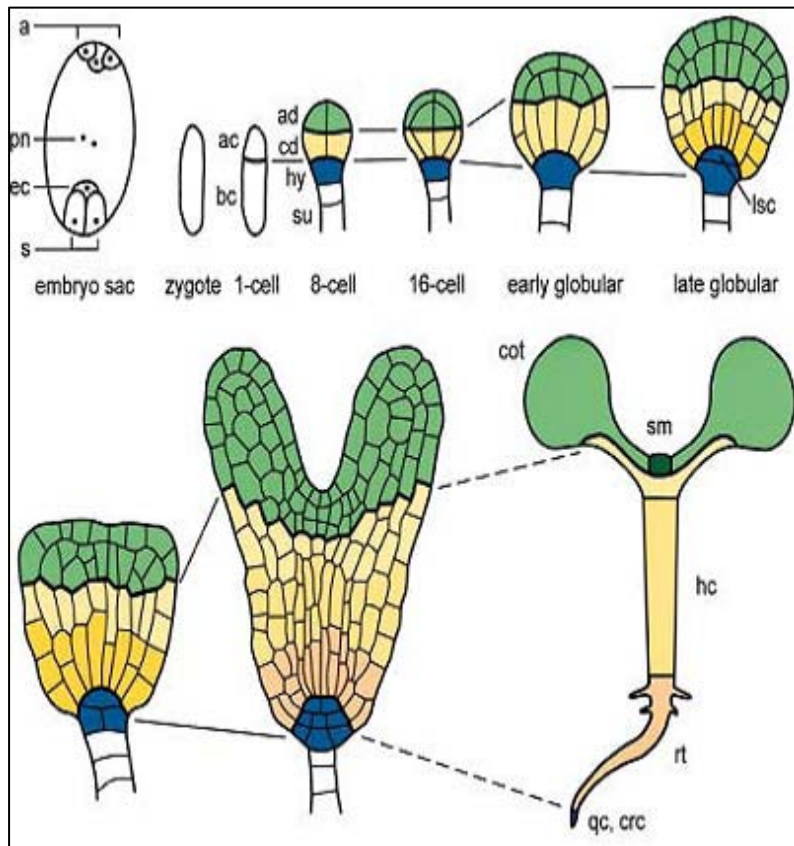


Figure 8. Développement de l'axe apico-basal de l'embryon d'*Arabidopsis thaliana*.
(D'après Laux et al., 2004)

a, antipodes; ac, apical daughter cell; ad, apical embryo domain; bc, basal daughter cell; cd, central embryo domain; cot, cotyledons; crc, central root cap; ec, egg cell; hc, hypocotyl; hy, hypophysis; lsc, lens-shaped cell; pn, polar nuclei; qc, quiescent center; rt, root; s, synergids; sm, shoot meristem; su, suspensor.

embryonnaire pourrait alors être induite par un gradient externe de substances sécrétées par les cellules ou présentes dans le milieu de culture (Van Engelen and de Vries, 1992; De Jong et al., 1993; Zimmermann, 1993). Toutefois, ces composés pourraient aussi mimer un effet maternel.

- **Bilan**

Bien que plusieurs exemples montrent que la polarité embryonnaire peut s'établir en absence de toute influence maternelle, celle-ci ne peut pas être totalement exclue. L'acquisition de l'asymétrie pourrait même dépendre de deux mécanismes superposés, l'un d'origine maternelle et l'autre endogène au zygote.

II.2.2. De l'embryon deux cellules jusqu'à l'embryon globulaire : mise en place des axes apico-basal et radial.

Chez *Arabidopsis thaliana*, la petite cellule apicale issue de la première division asymétrique du zygote suit un patron de divisions orientées très conservé (Fig.8) (West and Harada, 1993). La première série de divisions, comprenant deux premières segmentations longitudinales suivies d'une troisième transversale, aboutit à la formation du stade octant (8 cellules) caractérisé par l'organisation des cellules en deux rangées séparées par la frontière O' : la rangée supérieure, qui est destinée à former les cotylédons et l'apex végétatif apical, est à l'origine du domaine apical de l'embryon alors que la rangée inférieure, qui génère l'hypocotyle et les cellules initiales de la racine, formera le domaine central. La partie basale de l'embryon qui correspond au reste de la radicule est représentée par l'hypophyse (cellule apicale du suspenseur).

La vague suivante de divisions, orientée de manière péricline (parallèle à la surface de l'embryon) et menant au stade 16 cellules ou dermatogène, initie l'établissement de l'axe radial en faisant apparaître le premier tissu de l'embryon, le protoderme, qui est le précurseur de l'épiderme. Chaque cellule donne en se divisant une cellule externe à l'origine du protoderme et une cellule interne qui produit, par divisions périclines, les cellules précurseurs du tissu vasculaire (ou procambium) au centre entourées du tissu parenchymateux qui constitue un territoire méristématique et qui subit des divisions anticlines (perpendiculaires à la surface de l'embryon).

L'embryon est alors au stade globulaire : les trois couches tissulaires primordiales sont différenciées (protoderme, tissu parenchymateux méristématique et procambium) et organisées spatialement en trois couches concentriques conduisant à la mise en place de la symétrie radiale de l'embryon.

Parallèlement, le suspenseur, issu d'une série de divisions transverses de la cellule basale, est entièrement formé dès le stade globulaire précoce et compte alors entre 7 à 9 cellules vacuolisées.

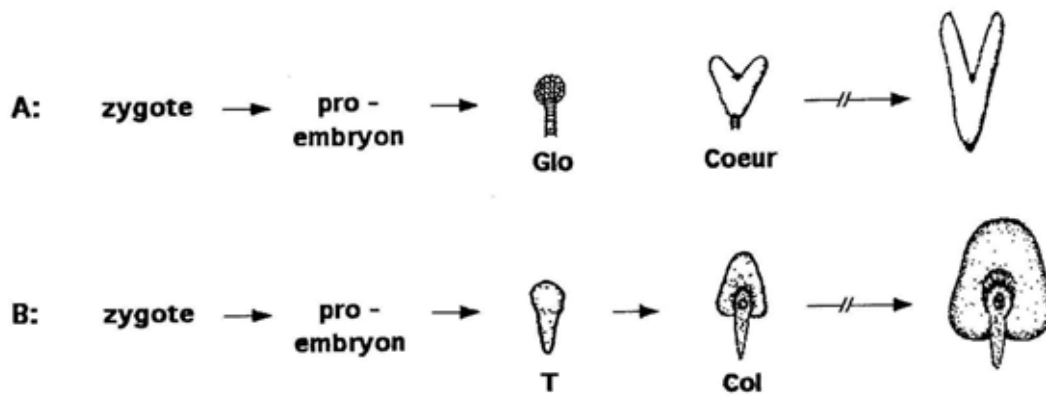


Figure 9: Les stades caractéristiques de l'embryogénèse des plantes dicotylédones (A) et Monocotylédones (B).

(d'après Breton, 1993)

Glo: embryon globulaire

T: embryon en phase de transition

Col: embryon coléoptilaire

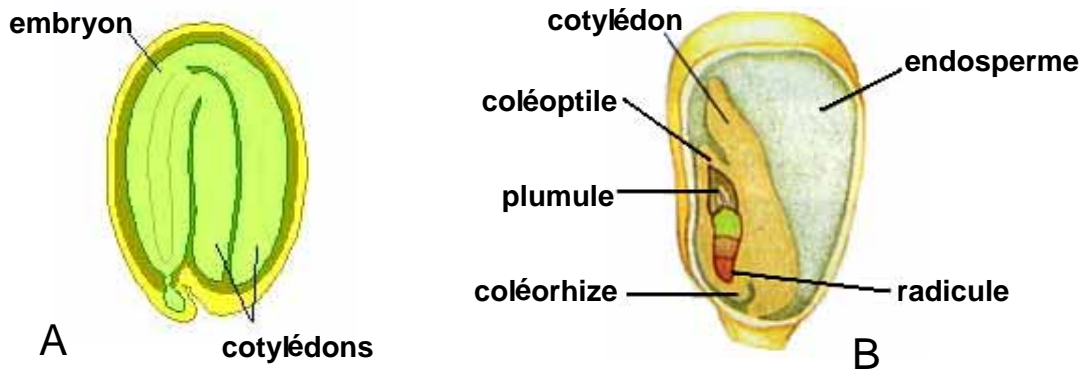


Figure 10: Structure d'une graine exalbuminée (*Arabidopsis thaliana*, A) et d'une graine Albuminée (maïs, B)

II.2.3. Transition globulaire-coeur : acquisition de la symétrie bilatérale et organisation définitive de la plante.

La transition entre les stades globulaire et coeur se traduit par le passage d'une symétrie radiale à une symétrie bilatérale, établie suite à l'apparition de 2 structures embryonnaires majeures : l'axe délimité par les méristèmes végétatif et racinaire et les cotylédons qui se développent de part et d'autre du méristème apical. Les méristèmes végétatif et racinaire, formés au cours du développement embryonnaire, sont ultérieurement responsables de la croissance et de l'architecture post-embryonnaire de la plante.

A partir du stade coeur, atteint 3 à 4 jours après la fécondation, le plan d'organisation et les principaux tissus de l'embryon mature sont établis. Ainsi, l'embryon possède les 3 tissus (protoderme, précurseurs du parenchyme et procambium) et les 4 structures (cotylédons, hypocotyle, méristèmes végétatif et racinaire) qui composent la plantule. Toutefois, l'embryon a encore relativement peu grossi.

II.2.4. Stade cotylédonnaire : croissance et accumulation des réserves.

Le stade cotylédonnaire, ultime étape du développement embryonnaire, constitue une phase active de division et élongation cellulaire au cours de laquelle les cotylédons croissent rapidement, l'axe embryonnaire s'allonge et les tissus parenchymateux et vasculaires envahissent peu à peu ces structures. Les cellules du cortex et de l'endoderme se différencient à partir des cellules méristématiques du parenchyme et vont accumuler des protéines et des lipides de réserves. Le protoderme externe se différencie en un épiderme.

II.3. Autres modèles de développements embryonnaires.

II.3.1. Comparaison avec l'embryogenèse zygotique chez les Monocotylédones.

Malgré des différences morphologiques importantes, les plans d'organisation et la séquence des événements aboutissant à l'embryon mature sont globalement conservés pour l'ensemble des plantes, qu'elles soient monocotylédones ou dicotylédones. L'embryogenèse de ces deux classes de plantes est même parfaitement comparable jusqu'au stade 8 cellules. Ainsi, chez le maïs qui constitue la plante modèle pour l'étude de l'embryogenèse des monocotylédones (Dumas et Mogensen, 1993), le développement embryonnaire suit, dans l'ensemble, le même schéma que chez *Arabidopsis thaliana* (Fig.9). Toutefois, quelques différences existent.

- Le développement du maïs s'effectue selon un plan de division différent de celui d'*Arabidopsis*, l'embryon et le suspenseur résultant tous deux d'une série de divisions asymétriques.
- Alors que le patron de divisions cellulaires d'*Arabidopsis* est très conservé, il peut varier d'un individu à l'autre chez le maïs.

- Chez le maïs, l'axe établi par les méristèmes caulinaire et racinaire de l'embryon est oblique par rapport à l'axe embryon-suspenseur alors que ces deux axes sont confondus chez *Arabidopsis*.

- Dans le cas du maïs, le méristème apical commence à produire des feuilles durant le développement embryonnaire, l'entrée en dormance de la graine ayant lieu lorsque la plantule est constituée de 6 primordia foliaires. Chez *Arabidopsis thaliana*, les premières feuilles n'apparaissent qu'après la germination de la graine.

- Enfin, chez le maïs, les réserves sont accumulées au sein de l'albumen, qui constitue la plus grande partie du grain à maturité, d'où la dénomination de grain albuminé (Fig.10). En revanche, la graine d'*Arabidopsis* où les réserves sont contenues au sein même de l'embryon est une graine exalbuminée, tout comme pour le tournesol.

II.3.2. Principales différences entre embryogenèse végétale et animale.

Bien que les végétaux présentent de nombreux avantages pour l'étude de l'embryogenèse (facilité de contrôle de la fécondation, descendants obtenus en très grand nombre, etc...), l'embryogenèse animale reste encore mieux connue. Plusieurs caractéristiques différencient l'embryogenèse des végétaux supérieurs de celles des animaux (Meinke, 1991) :

- Une particularité des végétaux supérieurs est l'existence d'une double fécondation qui conduit à la formation de l'embryon d'une part et de l'albumen d'autre part. Une seule fécondation a lieu chez les animaux.

- Dans le cas des végétaux, polarité structurale et différenciation des tissus au sein de l'embryon se mettent en place en l'absence de tout mouvement cellulaire.

- Chez les végétaux supérieurs, les cellules reproductrices dérivent du méristème végétatif alors que les animaux ont une lignée germinale distincte séparée des cellules somatiques très tôt lors de l'embryogenèse.

- La structure de l'embryon végétal, qui ne comprend qu'un nombre limité de types cellulaires et d'organes, est relativement simple comparée à celle des animaux.

- La différenciation des organes végétatifs et floraux se produit de manière continue dans les méristèmes, ce qui conduit à l'obtention d'une plante adulte très différente de l'embryon ou de la plantule. Chez les animaux, les décisions concernant le développement sont prises lors de l'embryogenèse précoce.

III. Méthodes d'étude des mécanismes moléculaires impliqués au cours de l'embryogenèse végétale.

Les premières données bibliographiques disponibles sur le programme génétique induit au cours du développement embryonnaire des plantes concernaient les gènes exprimés tardivement durant l'embryogenèse zygotique, comme les gènes codant pour les protéines de réserves ou ceux induits par l'acide abscissique lors de la déshydratation des graines (Galau et al., 1986 ; Skriver and Mundy, 1990). Du fait de la petite taille des embryons dans les phases précoces de leur développement et de leur localisation à l'intérieur des tissus de la graine, l'étude de l'expression des gènes durant l'embryogenèse précoce est difficile. Ainsi, pendant longtemps, la seule alternative pour étudier les phases d'initiation de l'embryogenèse au niveau moléculaire consistait à analyser l'induction d'embryons somatiques (Zimmermann, 1993). Par la suite, des approches de mutagenèse menées chez *Arabidopsis* et le maïs ont permis de caractériser des mutants létaux affectant l'embryogenèse précoce (Jürgens et al, 1991 ; Meinke, 1985 ; Clark and Sheridan, 1991). La caractérisation des gènes mutés a déjà permis de déterminer la nature et la fonction de plusieurs gènes clé de l'embryogenèse précoce. Toutefois, les informations ainsi obtenues restent encore insuffisantes et parcellaires et ne permettent pas d'élucider l'ensemble des mécanismes moléculaires et physiologiques menant à la formation de la graine. Plus récemment, des approches moléculaires directes menées sur de jeunes embryons zygotiques dans le but de cloner des gènes exprimés durant cette phase de l'embryogenèse ont été entreprises chez plusieurs espèces végétales. Toutefois, nous ne disposons actuellement que de très peu d'informations concernant les programmes génétiques induits au cours du développement embryonnaire précoce.

III.1. L'analyse par mutagenèse, une source d'informations abondantes mais insuffisantes.

La grande majorité des connaissances actuelles concernant le programme génétique induit au cours de l'embryogenèse des végétaux supérieurs provient de l'étude de mutants. Des mutants du développement embryonnaire ont été identifiés chez un grand nombre d'espèces végétales (*Arabidopsis*, maïs, orge, carotte, riz, pois) ayant subi différents types de mutagenèse (chimique, insertionnelle ou aux rayons X).

Chez le maïs, deux classes de mutants ont été identifiés : les mutants *dek* (defective-kernel : grain altéré) (Neuffer and Sheridan, 1980 ; Scanlon et al., 1994) altérés simultanément au niveau de l'embryon et de l'albumen et les mutants *emb* (embryonic defective) (Clark and Sheridan, 1991) pour lesquels seul l'embryon est affecté. Parmi ces derniers, certains mutants semblent altérés au niveau de la mise en place du patron embryonnaire, d'autres au niveau de la croissance cellulaire ou du métabolisme (Clark and Sheridan, 1991), seuls le clonage et la caractérisation moléculaire des gènes sauvages pouvant déterminer leur fonction.

Chez *Arabidopsis*, plus de 2000 mutants affectant différents évènements au cours de l'embryogenèse ont été isolés par mutagenèse chimique EMS (éthyl méthane sulfonate) (Mayer et al., 1991 ; Meinke, 1985), par insertion d'ADN-T (Devic et al., 1996, Errampalli et al., 1991, McElver et al., 2001) ou par transposon tagging (Altmann et al., 1995 ; Goldberg et al., 1994). Deux stratégies complémentaires, basées sur le criblage phénotypique de siliques immatures (Meinke, 1985) et de plantules en germination (Jürgens et al., 1991), ont été employées afin d'isoler des mutants altérés au niveau de la phase d'embryogenèse précoce. En fonction des données issues de ces deux approches de mutagenèse, différentes estimations du nombre de gènes essentiels au développement de la graine ont été effectuées. Ainsi, Meinke évalue à environ 500 le nombre de gènes EMB chez *Arabidopsis* (Meinke, 1991), alors que Jürgens estime que 4000 gènes seraient essentiels au développement normal de l'embryon parmi lesquels environ 40 gènes seraient responsables de la mise en place du patron morphologique (Jürgens et al., 1991).

III.1.1. Criblage de siliques immatures.

La stratégie choisie par Meinke (1985) consiste à sélectionner les siliques immatures dont 25% des graines sont arrêtées dans leur développement (taux compatible avec la présence d'une seule mutation récessive) après autofécondation des plantes mutagénéisées. Les graines qui n'achèvent pas complètement l'embryogenèse correspondent à des mutants embryo-défectueux de type *emb* et se distinguent des graines sauvages et des ovules non fécondés par leur taille et /ou leur couleur selon la nature de l'altération. Cependant, le grand nombre de mutants *emb* ainsi identifiés laisse penser que les gènes altérés sont majoritairement impliqués dans des fonctions cellulaires de base (fonctions de ménage) et que seulement peu d'entre eux sont réellement impliqués dans la formation du plan d'organisation de l'embryon. Afin de discriminer ces deux types de gènes, des expériences de sauvetage d'embryons ont été menées. En effet, si un embryon mutant *emb* est capable de se développer normalement sur un milieu enrichi en éléments nutritifs, cela signifie que le gène sauvage n'est pas spécifiquement requis pour l'aboutissement des stades tardifs du développement mais est plutôt impliqué dans une voie de biosynthèse (vitamines, acides aminés...). L'étude des altérations de gènes de ménage entraînant des mutations *emb* fournit toutefois d'importantes informations sur les fonctions métaboliques essentielles au développement de l'embryon.

III.1.2. Criblage de plantules en germination.

La stratégie adoptée par Jürgens et al. (1991), adaptée de l'approche génétique appliquée à la *Drosophile* et basée sur le criblage phénotypique des plantules en germination, permet d'isoler plus spécifiquement les gènes impliqués dans l'établissement du plan d'organisation de l'embryon. En effet, cette approche repose sur le postulat que la mutation de gènes responsables de la formation du patron n'interfère pas avec le déroulement de l'embryogenèse mais aboutit à la formation de plantules anormales alors que l'inactivation de gènes codant pour des fonctions cellulaires de base aboutit à la mort de l'embryon. Les mutations altérant le

plan d'organisation de l'embryon agissent à 3 niveaux différents du patron morphologique (Mayer et al., 1991) : l'axe apico-basal, l'axe radial et la forme globale. Cependant, malgré le grand nombre de mutants isolés, peu de gènes potentiellement impliqués dans l'établissement du plan d'organisation ont été clonés jusqu'à présent.

Si le crible défini par Meinke, qui mène majoritairement à l'identification de gènes de ménage, peut sembler peu stringent, le crible adopté par Jürgens, destiné à mettre en évidence les gènes spécifiquement impliqués dans la formation du patron embryonnaire, peut, au contraire, empêcher la détection de gènes jouant un rôle clé au cours de l'embryogenèse. De plus, cette méthode de sélection, s'appuyant sur un parallèle effectué entre l'embryogenèse végétale et animale et interprétant les données moléculaires obtenues en fonction du modèle précédemment décrit dans le règne animal, peut être soumise à controverse.

III.1.3. Limites de l'approche par mutagenèse.

Bien que l'analyse de mutants ait déjà permis la caractérisation de nombreux gènes jouant des rôles clés dans le développement embryonnaire des végétaux, cette approche ne permet pas l'accès à l'ensemble des gènes potentiellement intéressants du fait de l'existence de redondance de séquences ou de fonctions au sein du génome. De plus, les mutations à un stade aussi critique du développement sont souvent létales.

III.2. L'analyse par les méthodes de la génomique fonctionnelle, une approche prometteuse.

Etant donné les limites de l'approche par mutagenèse pour l'étude de l'embryogenèse précoce, des méthodes de transcriptomique, permettant de visualiser l'expression différentielle de gènes de manière spatiale, temporelle ou conditionnelle, à plus ou moins grande échelle, ont été utilisées afin de caractériser des gènes clés du développement embryonnaire inaccessibles par criblage de mutants.

III.2.1. Les techniques de transcriptomique à faible débit.

Depuis une dizaine d'années, plusieurs techniques de criblage différentiel ont permis de caractériser des gènes spécifiquement exprimés dans l'embryon ou jouant un rôle essentiel au cours de l'embryogenèse précoce. Ainsi, chez *Arabidopsis thaliana*, les approches par Differential Display ont révélé les gènes ATS1 et ATS3 (*Arabidopsis thaliana seed 1 et 3*) embryo-spécifiques (Nuccio and Thomas, 1999) ou encore le gène AGL15 (Heck et al., 1995), codant pour une protéine de type MADS box qui régulerait l'expression d'une enzyme du métabolisme des gibbérellines (Wang et al., 2004). Le gène PEL1 codant pour une protéine à doigt de zinc a été identifié chez *Arabidopsis thaliana* par soustraction virtuelle, une méthode d'enrichissement en cDNA différentiels par hybridation sur filtres haute densité (Li and Thomas, 1998).

Enfin, le gène ATML1 (*Arabidopsis thaliana* meristem L1 layer) a été isolé chez *Arabidopsis thaliana* grâce à des séquences ovules-spécifiques de l'orchidée *Phalaenopsis* (Nadeau et al., 1996).

Toutefois, ces techniques, qui ne permettent généralement que l'étude d'un nombre réduit de gènes, ont un rendement limité.

III.2.2. Les approches EST et microarrays ou l'étude globale de l'expression du génome.

L'essor récent des méthodes d'analyse de l'expression génique par séquençage d'EST à haut débit ou grâce à l'utilisation de la technologie microarray a permis d'envisager une étude exhaustive et simultanée de l'expression de la totalité du génome au cours du développement embryonnaire. Toutefois, du fait des difficultés rencontrées pour l'extraction d'ARN à partir de jeunes embryons, des données génomiques concernant les stades précoces du développement embryonnaire végétal ont longtemps manqué, les premières données EST obtenues à partir de graines immatures d'*Arabidopsis thaliana* ayant été publiées en 2000 (White et al., 2000). Bien que ces 10 500 séquences s'appliquent à des stades encore trop tardifs de l'embryogenèse (de 5 à 10 jours après floraison), leur analyse *in silico* a tout de même permis d'identifier un grand nombre de nouveaux gènes (24%), potentiellement spécifiques de la graine et de caractériser le profil fonctionnel des gènes exprimés dans la graine d'*Arabidopsis*. Par la suite, cette banque d'ADNc a servi à la construction de puces à ADN utilisées pour une étude comparative des transcriptomes de graines immatures, de feuilles et de racines (Girke et al., 2000) qui a révélé des gènes surexprimés dans la graine parmi lesquels certains codent pour des facteurs de régulation potentiels tels que des facteurs de transcription, des protéines kinases ou des phosphatases. Ces microarrays ont également été employés pour étudier la cinétique d'expression des gènes impliqués dans l'accumulation des réserves protéiques et lipidiques au sein de la graine et ont notamment permis d'élaborer le premier modèle de régulation transcriptionnelle menant à l'allocation des ressources en carbone parmi les différents types de réserves durant le développement embryonnaire (Ruuska et al., 2002). Des travaux plus récents visant à caractériser les programmes transcriptionnels induits au cours de stades reproductifs plus précoces ont été menés à partir de boutons floraux avant pollinisation, de fleurs pollinisées et de siliques récoltées deux jours après la fécondation et à l'aide de microarrays comptant plus de 22 000 gènes (Hennig et al., 2004). Cependant, toutes ces études, conduites à partir d'un mélange de tissus reproductifs (tissus embryonnaires, albumen et enveloppes de la graine) et non d'embryons isolés à un stade précis du développement, ne permettent pas de décrire précisément les gènes exprimés de manière spécifique au sein des tissus embryonnaires. Plus récemment, une étude microarray, réalisée à partir d'embryons d'*Arabidopsis thaliana* microdisséqués au laser, a permis une comparaison fine des transcriptomes des régions apicales et basales d'embryons aux stades globulaire et cœur (Casson et al., 2005). Jusqu'à 65% du génome seraient exprimés dans l'embryon en formation et les régions basales et apicales au stade cœur seraient caractérisées par une plus grande complexité de l'expression des gènes par rapport au stade globulaire, ceci étant probablement lié à la complexification de l'activité cellulaire au cours de la transition entre ces deux stades critiques de développement.

Des études transcriptomiques ont également été menées chez plusieurs autres plantes mais les connaissances restent encore incomplètes du fait de l'exploitation insuffisante de ces approches pour l'étude de l'embryogenèse végétale. Chez le maïs (*Zea mays*), la première banque d'ADNc fût construite en 1995 à partir d'embryons au stade de transition (Breton et al., 1995) et plus récemment, une approche utilisant la technologie de microarray a permis de disséquer les différentes voies métaboliques impliquées au cours du développement embryonnaire (entre 5 et 40 jours après la fécondation) en s'appuyant sur les profils d'expression de 1 500 gènes (Lee et al., 2002). Plusieurs projets ayant pour but de comparer les programmes transcriptionnels induits dans les tissus maternels et au sein de l'embryon formé à la suite de la fécondation ont été développés comme chez l'orge (*Hordeum vulgare* L.) (Sreenivasulu et al., 2004) ou encore chez le riz (*Oryza sativa* L.) (Lan et al., 2004). D'autres travaux, menés chez le maïs (Grimanelli et al., 2005) et chez le blé (*Triticum aestivum* L.) (Sprunck et al., 2005), visent à caractériser plus particulièrement la transition entre les programmes génétiques maternel et zygotique. Une étude de l'expression génique des facteurs de transcription impliqués au cours de la formation de la graine a débuté chez *Arabidopsis thaliana* (Folter et al., 2004) et chez le riz (Ye et al., 2004). Enfin, des études comparatives du transcriptome exprimé durant l'embryogenèse chez *Arabidopsis thaliana* et chez des espèces présentant d'importantes différences métaboliques avec cette plante modèle, telles que le sésame (*Sesamum indicum* L.) ou encore le ricin (*Ricinus communis* L.) dont les graines qualifiées de 'non vertes' n'effectuent pas la photosynthèse, ont révélé des dissemblances notables entre les proportions relatives des principales fonctions cellulaires impliquées au cours du développement de la graine chez ces différentes espèces (van de Loo et al., 1995 ; Suh et al., 2003).

En permettant l'analyse simultanée de l'expression d'un très grand nombre de gènes, ces études de transcriptomique à haut débit ont fourni une première image globale des principaux processus cellulaires impliqués à différentes étapes de l'embryogenèse végétale et chez différentes espèces. En révélant des gènes jouant potentiellement un rôle clé au cours de cette phase critique du développement et en apportant des informations sur les principales fonctions cellulaires et voies de régulation engagées, elles constituent la première étape vers la compréhension des processus moléculaires et physiologiques contrôlant le développement embryonnaire chez les végétaux.

IV. Modèle génétique et fonctionnel de l'établissement du patron embryonnaire chez les dicotylédones.

IV.1. L'élaboration du plan d'organisation de l'embryon végétal implique une régulation génétique très fine.

IV.1.1. Inégalité des sexes lors de l'embryogenèse précoce : rôle prédominant du génome maternel.

Jusqu'à trois à quatre jours après la fécondation (stade 32 à 64 cellules), l'activité génétique au sein de l'embryon semble essentiellement dépendre de la transcription des allèles d'origine maternelle, les allèles paternels étant transitoirement réprimés. Cette inhibition du génome paternel s'effectue au niveau transcriptionnel du fait de mécanismes de régulation épigénétiques tels que la méthylation de l'ADN, l'acétylation des histones ou le remodelage de la chromatine, entraînant des phénomènes d'empreinte génétique (ou genomic imprinting). Ainsi, l'imprinting est une forme de régulation épigénétique où l'expression de l'allèle d'un gène dépend du sexe du parent dont il provient. Des analyses génétiques menées chez *Arabidopsis thaliana* ont déjà permis d'identifier 20 gènes soumis à un tel contrôle et impliqués dans diverses fonctions biologiques parmi lesquels le gène GNOM (Vieille-Calzada, 2000). Toutefois, s'il est probable qu'une grande majorité des allèles paternels subissent une telle répression, certains transcrits paternels exprimés précocement au cours de l'embryogenèse tels que le gène PROLIFERA ont également été décrits (Kohler and Grossniklaus, 2005 ; Scholten et al., 2002 ; Springer et al., 2000 ; Weijers et al., 2001). Ainsi, l'empreinte génétique ne serait pas un mécanisme général touchant la totalité du génome au cours de l'embryogenèse précoce mais ne s'appliquerait qu'à des gènes spécifiques.

Les gènes appartenant à la classe FIS (Fertilization Independent Seed formation), comprenant les gènes FIS1/MEA, FIS2 et FIS3/FIE dont la mutation entraîne un développement autonome de l'endosperme en absence de fécondation, ne semblent subir ce type de régulation épigénétique qu'au niveau de l'endosperme, les deux allèles étant exprimés précocement au sein de l'embryon (Kinoshita et al., 1999). Ceci indiquerait que les phénomènes d'imprinting affecteraient plus particulièrement l'endosperme. Ces gènes qui codent pour des répresseurs transcriptionnels du groupe Polycomb (PcG) initialement décrit chez la drosophile formeraient des complexes protéiques inhibant des gènes impliqués dans le développement de l'endosperme. Récemment, grâce à des analyses génétiques menées chez *Arabidopsis thaliana*, 70 mutations entraînant un arrêt du développement embryonnaire à des stades précoces (du zygote unicellulaire au stade cœur) et affectant des gènes à effet maternel aux fonctions biologiques diverses ont été identifiées (Pagnussat et al., 2005). Parmi ces gènes, on note de nombreux facteurs de transcription appartenant à différentes familles (MYB, WRKY et TCP) ainsi que des signaux de transduction tels que des protéines kinases ou la phospholipase D. Le mécanisme d'action de ces gènes sur le développement embryonnaire, par un vrai effet

maternel du gamétophyte femelle (c'est-à-dire grâce à l'accumulation de facteurs maternels) ou par une latence dans l'activation du génome d'origine paternelle, reste encore à déterminer.

IV.1.2. Gènes impliqués dans la mise en place de l'axe apico-basal.

L'axe apico-basal, qui constitue l'axe principal de croissance des racines et des pousses végétatives après la germination, est établi le plus précocement, les premiers signes de polarité selon cet axe étant observables dès la première division du zygote.

IV.1.2.1. Suspenseur vs embryon : acquisition de devenir cellulaires distincts dès la première division asymétrique.

Des études d'expression génique ont montré que des programmes spécifiques de transcription étaient initiés très précocement lors du développement embryonnaire. Dès la première division asymétrique, les deux cellules filles donnant respectivement l'embryon et le suspenseur ont un devenir différent et expriment des programmes génétiques distincts. En effet, alors que l'oosphère et le zygote expriment un mélange des ARNm codant pour les facteurs de transcription WUSCHEL HOMEODOMAIN 2 et 8 (WOX2 et WOX8), ces transcrits sont séparés au cours de la première division zygotique, WOX2 étant désormais exprimé spécifiquement dans les cellules issues de la petite cellule apicale alors que l'expression de WOX8 est caractéristique du développement précoce de la cellule basale de l'embryon (Haecker et al, 2004). De même, à la suite de la première division asymétrique, seule la cellule apicale exprime le gène *ATML1* (*Arabidopsis thaliana* Meristem Layer 1) (Lu et al., 1996 ; Sessions et al., 1999), codant pour une protéine à homéobox alors que l'expression de *PIN7* (*PINFORMED7*) codant pour l'un des membres de la famille PIN intervenant dans les processus d'efflux auxinique est réduite à la cellule basale (Gälweiler et al., 1998 ; Friml, 2003). Ainsi, la première division asymétrique du zygote, en ségrégant de manière différentielle les transcrits de certains gènes, crée deux cellules d'identité différente à l'origine de la polarité apico-basale.

Cependant, étant donné que la cellule la plus apicale du suspenseur, l'hypophyse, participe à la formation du méristème racinaire de l'embryon, il semblerait que la cellule basale exprime, au moins en partie, le programme embryogène. Cette hypothèse semble être confirmée par l'observation de certains mutants comme *suspensor*, *raspberry 1* et *2* ou *twin* dont le suspenseur commence une série de divisions anormales prenant des caractéristiques embryonnaires après l'arrêt du développement de l'embryon. Ainsi, chez les mutants *suspensor* (Schwartz, 1994), des corps protéiques et lipidiques et des grains d'amidon s'amassent au niveau du suspenseur, alors que ceux-ci se limitent normalement à l'embryon. Chez les mutants *raspberry 1* et *2* (Yadegari, 1994) le gène *AtLTP*, codant pour une protéine de transport lipidique homologue à EP2 de la carotte et exprimée normalement seulement dans le protoderme de l'embryon, est transcrit à la fois dans

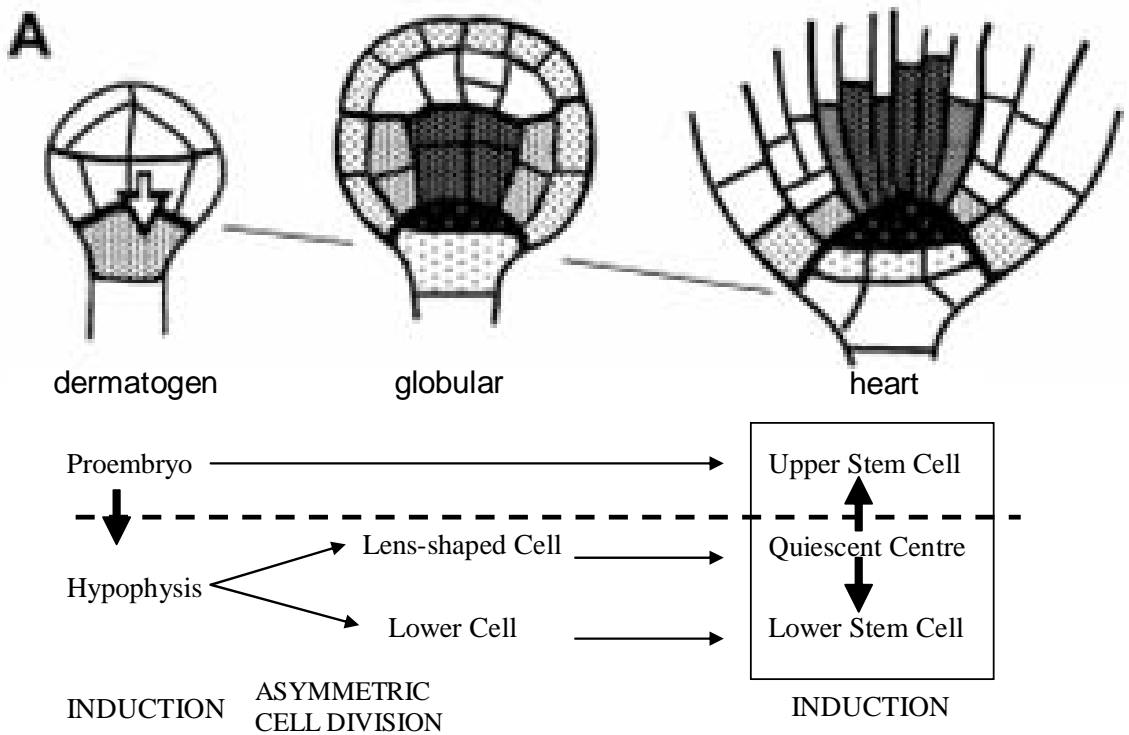


Figure 11: Modèle de formation du méristème racinaire. (d'après Jürgens, 2003)

Un signal provenant du proembryon induit le devenir cellulaire de l'hypophyse qui se divise alors de façon asymétrique. La cellule fille apicale est à l'origine du centre quiescent (en foncé) qui confère l'identité de cellule souche aux cellules adjacentes (en pointillé).

l'embryon et le suspenseur. Enfin, le suspenseur des mutants *twinn* se redifférencie pour donner un embryon secondaire suite ou non à l'arrêt du premier (Zhang and Sommerville, 1997).

Le devenir cellulaire du suspenseur serait également dépendant du gène YODA codant pour une MAPKK KINASE (Lukowitz et al., 2004) et du gène LEC1 exprimé seulement lors de l'embryogenèse et codant pour une sous-unité du facteur de transcription se liant à la boîte CCAAT indispensable au maintien du devenir cellulaire suspenseur (Lotan et al., 1998).

Ainsi, suspenseur et proembryon interagiraient : l'embryon signalerait au suspenseur de conserver son état différencié ; lors de l'arrêt du développement de l'embryon, comme pour les mutants *suspensor* et *raspberry*, cette signalisation est stoppée et le suspenseur adopte le programme génétique embryonnaire. Alternativement, un équilibre de facteurs de croissance pourrait maintenir simultanément les devenirs développementaux de l'embryon et du suspenseur (Goldberg et al., 1994).

IV.1.2.2. Mise en place du méristème racinaire.

Le méristème racinaire (Fig.11), constitué d'un ensemble de cellules souches organisées en files, se forme à partir de 2 types cellulaires distincts : l'étage supérieur des cellules souches provient des cellules basales du proembryon, alors que l'étage inférieur dérive de l'hypophyse (cellule apicale du suspenseur) dont le développement serait induit par un signal provenant du proembryon probablement de nature auxinique (Jürgens, 2001, 2003). L'hypophyse subit une division asymétrique générant une petite cellule apicale qui sera à l'origine du centre du méristème racinaire, dit centre quiescent, et une grande cellule basale qui formera la columelle dont la fonction est de protéger le méristème et de percevoir les stimuli gravitropiques. Le centre quiescent, qui est composé de 4 cellules non cyclantes encadrées par deux files de cellules souches, a pour rôle de maintenir l'état indifférencié des cellules souches adjacentes par le biais d'interactions locales médiées par un signal de nature inconnue. Les cellules filles issues de l'étage supérieur de cellules souches contribuent à prolonger les files cellulaires de la racine.

La spécification correcte du groupe de cellules issues de l'hypophyse est indispensable à l'établissement du méristème racinaire comme l'indique l'absence d'apex racinaire dans les mutants du 'groupe hypophysaire' tels que le mutant *hobbit* (Scheres et al., 1995). Le gène HOBBIT, qui code pour une protéine semblable à la sous-unité CDC27 du complexe promoteur de l'anaphase, pourrait être impliqué dans le processus de dégradation protéique (Blilou et al., 2002). Les altérations dues à la mutation du gène HOBBIT pourraient être dues au fait que l'hypophyse ne réponde pas correctement aux signaux potentiellement émis par le proembryon (Jürgens, 2001).

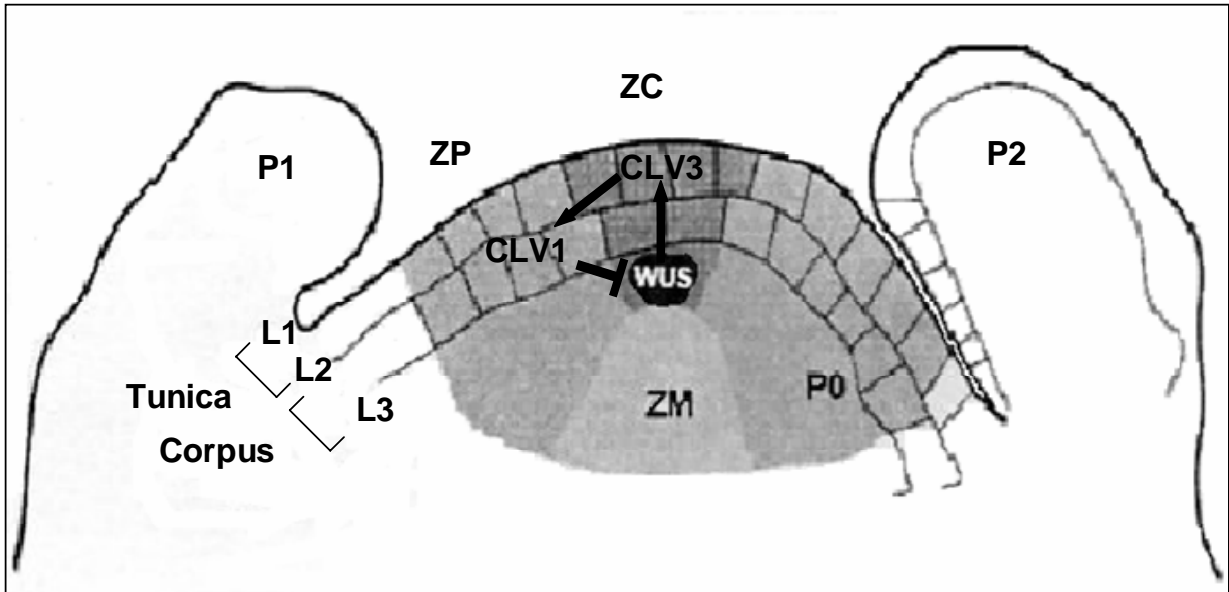


Figure 12: Modèle de fonctionnement du méristème apical.

ZC: Zone centrale
 ZM: Zone médullaire
 ZP: Zone périphérique
 P0, P1, P2: primordia foliaires.

→ Activation ⊣ Répression

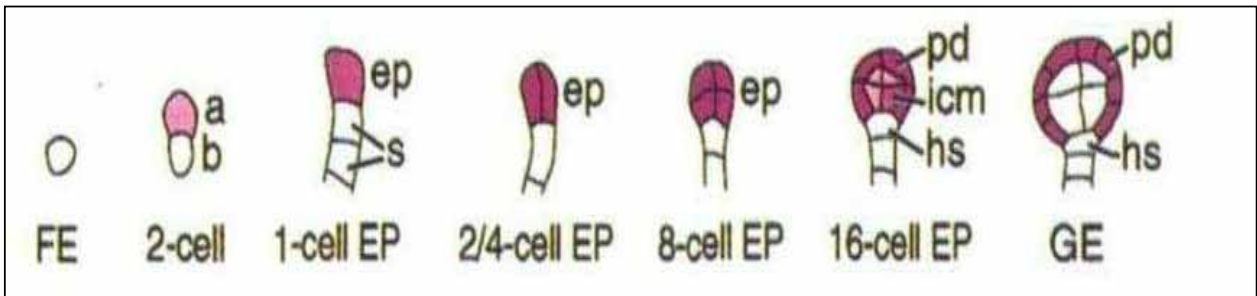


Figure 13: Représentation schématique de la distribution des ARNm d'ATML1 et de PDF2.
 (adaptée de Lu et al., 1996)

a, apical cell; b, basal cell; ep, embryo proper; FE, fertilized egg; GE, globular-stage embryo; hs, hypophysis; icm, inner cell mass; pd, protoderm; s, suspensor.

IV.1.2.3. Mise en place du méristème apical.

A un stade plus avancé du développement, à l'opposé du méristème racinaire, un ensemble de cellules se différencie pour donner le méristème apical caulinaire, lieu de production des organes aériens post-embryonnaires.

Ce méristème s'organise en 3 zones (Fig. 12):

- la zone centrale composée en son sommet de cellules à division lente et à sa base d'un centre organisateur, véritable réservoir de cellules souches,
- la zone périphérique entourant la zone centrale et constituant le site d'initiation des primordia au niveau duquel les cellules filles provenant du centre organisateur se divisent rapidement,
- la zone médullaire constituée de cellules à division rapide qui contribuent aux tissus internes de la tige.

La formation du méristème apical est soumise à un système de régulation complexe, où le centre organisateur qui synthétise des signaux contrôlant la taille et le maintien du méristème, joue un rôle central (Fig. 12). Ainsi, l'expression du gène *WUSCHEL* (*WUS*) au sein de ces cellules permet de garder les cellules souches sus-jacentes dans un état indifférencié (Mayer et al., 1998). En réponse, ces dernières expriment *CLAVATA3* (*CLV3*), qui diffuse vers les cellules avoisinantes jouant le rôle de ligand interagissant avec des récepteurs. *CLAVATA1* (*CLV1*), un récepteur à activité kinase, constituerait l'un de ces récepteurs putatifs qui, en induisant l'inhibition de *WUS* par rétro-contrôle négatif, limite l'expansion du centre organisateur (Fig. 12) (Carles and Fletcher, 2003).

IV.1.3. Gènes intervenant dans la mise en place de l'axe radial.

Le patron radial de l'embryon est composé de 3 tissus de base : le protoderme (précurseur de l'épiderme), le parenchyme et le tissu vasculaire. La délétion d'un tissu primaire pouvant être une cause de mortalité et les mutants affectés dans l'élaboration des tissus étant difficiles à identifier, seulement un très faible nombre de mutants présentant des défauts dans le patron radial a pu être isolé.

IV.1.3.1. Gènes impliqués dans la spécification et le maintien du protoderme.

Au cours de la mise en place de l'axe radial, des programmes transcriptionnels distincts au sein des différentes couches tissulaires périclinales sont identifiables dès l'apparition des deux premières couches de cellules au stade dermatogène (16 cellules).

Les gènes *ATML1* et *PDF2* (Protodermal Factor 2), qui sont spécifiquement exprimés au niveau du protoderme dès le stade 16 cellules (Fig. 13), sont indispensables à la spécification et au maintien de l'identité de la couche externe dans la région apicale de l'embryon dont l'intégrité semble essentielle pour l'organisation des cellules sous-jacentes (Abe et al., 2003). Les transcrits de ces deux gènes sont répartis de

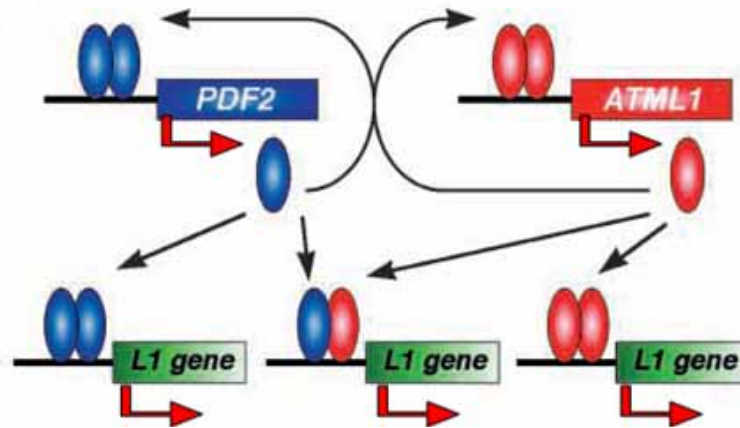


Figure 14: Schéma des fonctions régulatrices de PDF2 (bleu) et ATML1 (rouge) pour l'expression de gènes spécifiques de la couche L1. (adapté de Abe et al., 2003)

Les flèches rouges traduisent l'activation transcriptionnelle médiée par l'interaction de PDF2 et /ou ATML1 avec l'élément régulateur L1-box du promoteur des gènes. Le schéma évoque la formation potentielle d'homo et d'hétéro-dimères ainsi que la possibilité d'autorégulation positive de ces deux gènes.

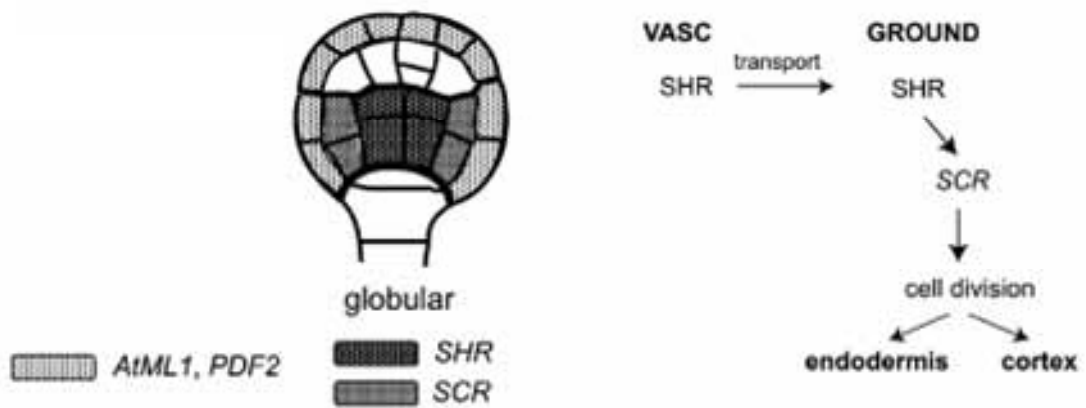


Figure 15: Modèle génétique de la mise en place des tissus internes (adaptée de Jürgens, 2003).

Dans la région centrale de l'embryon globulaire, les cellules internes sont à l'origine des cellules provasculaires (VASC), qui expriment SHR, et des cellules du tissu parenchymateux (GROUND), qui expriment SCR. Le facteur de transcription SHR migre dans le tissu parenchymateux où il active SCR menant à des divisions asymétriques formant l'endoderme et le cortex.

manière homogène au sein de l'embryon jusqu'au stade octant ; étant donné la présence sur leur promoteur de sites de liaison aux protéines à homéodomaines de type WUSCHEL, leur confinement au sein des cellules protodermiques pourrait être lié à une inhibition de leur expression au sein des cellules internes au stade dermatogène par ces facteurs de transcription (Ingram, 2004 ; Mayer et al., 1998). De plus, en se liant à des éléments régulateurs de type L1-box des promoteurs de certains gènes, les protéines ATML1 et PDF2 contrôleraient l'expression d'autres transcrits présents plus tardivement au sein du protoderme tels que PDF1 (Protodermal Factor 1), LTP1 (Lipid Transfer Protein 1) ou encore ETTIN (ETT) (Fig. 14) (Ingram, 2004). Les gènes KEULE et KNOLLE d'*Arabidopsis thaliana* constituent deux autres gènes dont la mutation perturbe la différenciation tissulaire lors de l'établissement du protoderme. Les produits de ces deux gènes qui sont respectivement une protéine SEC1 (Assaad et al., 2001) et une SYNTAXINE (Lukowitz et al., 1996), interagissent entre eux et sont indispensables à la séquestration, au transport et à la fusion des vésicules golgiennes lors de la cytokinèse. Au sein de l'axe radial de l'embryon, une cytokinèse complète serait nécessaire à la ségrégation physique des déterminants définissant le devenir des cellules internes et protodermiques (Laux and Jürgens, 1997).

IV.1.3.2. Gènes impliqués dans la différenciation des tissus internes.

Chez les mutants *scarecrow* (*scr*), *short root* (*shr*) et *pinocchio* (*pic*), la division asymétrique du tissu parenchymateux générant le cortex et l'endoderme n'a pas lieu (Scheres et al., 1995). Ces mutants ne présentent donc qu'une seule couche de cellules exprimant, selon les cas, uniquement des caractéristiques du cortex comme pour *shr* ou de l'endoderme pour *pic* ou les propriétés des deux tissus comme chez *scr* où les deux devenirs cellulaires ont été établis mais ne se séparent pas car les cellules du parenchyme ne se divisent pas.

Contrairement au gène SHR qui semble réellement impliqué dans les processus de spécification du devenir cellulaire au sein de l'axe radial de l'embryon, les défauts observés chez le mutant *scr* seraient uniquement dus à la réduction du nombre de cellules, ces altérations pouvant être éliminées par une mutation *fass* à l'origine d'une augmentation du nombre de couches cellulaires (Scheres et al., 1995). SCR et SHR codent tous deux pour des facteurs de transcription de la famille GRAS (GIBBERELLIN-INSENSITIVE, REPRESSOR of gal-3, SCARECROW) (Di Laurenzio et al., 1996). SHR, initialement exprimé au sein des cellules provasculaires, migre dans les cellules adjacentes du tissu parenchymateux où il induit SCR qui, suite à la division asymétrique de ce tissu, restera confiné dans l'endoderme (Fig. 15) (Nakajima et al., 2001).

IV.1.4. Gènes intervenant dans l'établissement de la symétrie bilatérale : Mise en place des organes latéraux.

Au sein du méristème, les cellules filles générées par les divisions perpétuelles sont intégrées dans les primordia d'organes situés en périphérie. Alors que les gènes STM, WUS et CLV contrôlent le maintien de la zone méristématique, un autre pool de gènes régule la distribution des cellules au niveau des organes latéraux. Plusieurs processus interdépendants interviennent successivement dans la production des organes : recrutement des cellules dans les primordia, croissance et séparation des organes et spécification de leur identité.

Différents gènes sont impliqués dans le recensement d'un groupe de cellules du méristème pour former un primordium. SHOOT MERISTEMLESS (STM) d'*Arabidopsis thaliana* (KNOTTED1 (KNT1) chez le maïs) est un facteur de transcription appartenant à la famille de gènes KNOX (KNOTTED1-like homeobox) qui, en inhibant le facteur de transcription de type MYB ASYMETRIC LEAVES 1 (AS1) d'*Arabidopsis thaliana* (ou ROUGH SHEATH (RS2) chez le maïs), empêche la formation d'organes latéraux en maintenant les cellules dans un état méristématique (Byrne et al., 2000). L'inactivation du gène STM semble donc indispensable à la formation des primordia foliaires. Cependant, les mécanismes de régulation menant à l'inhibition de ce gène n'ont pas encore été élucidés (Jürgens, 2003).

Un autre réseau de gènes incluant CUP-SHAPED COTYLEDON 1, 2 et 3 (CUC1, 2 et 3), codant pour des facteurs de transcription de type NAC, permet de définir les limites des primordia et de mener à la formation d'organes distincts (Fig. 16). Chez les dicotylédones, l'action de ces gènes, exprimés au niveau du point d'ancrage des futurs cotylédons de part et d'autre du méristème apical et inhibant la croissance au centre du domaine apical, permet de passer de la symétrie radiale de l'embryon globulaire à la symétrie bilatérale de l'embryon cœur. Ils interagissent également avec les gènes STM, PIN1 et MP selon des voies de régulation fines et complexes: PIN1 et MP, qui sont impliqués dans la voie de signalisation auxinique, influencent le patron d'expression spatial de CUC1, en inactivant ce gène au niveau des cotylédons, alors qu'ils activent CUC2 à la frontière méristème-cotylédons (Aida et al., 2002). En revanche, CUC1 et CUC2 sont nécessaires pour l'expression de STM qui entraîne la formation du méristème apical. Au cours des phases tardives d'embryogenèse, STM régulerait à son tour le profil d'expression des gènes CUC.

IV.2. La différenciation cellulaire au sein de l'embryon végétal nécessite l'échange d'indices de position.

Contrairement à l'embryogenèse animale, le développement embryonnaire chez les végétaux n'implique pas de processus de migration cellulaire mais est dépendant de la position des cellules. La spécification du devenir cellulaire au cours de l'embryogenèse végétale nécessite donc un échange intercellulaire d'indices de position pouvant impliquer différents procédés et agents de transduction du signal. La différenciation cellulaire peut, par exemple, être liée à l'héritage de différents déterminants développementaux par les

cellules filles à la suite d'une division asymétrique. Différents processus développementaux peuvent également se déclencher au sein de cellules identiques grâce à la perception d'informations positionnelles distinctes médiées par des molécules signal agissant à courte et/ou à longue distance telles que les hormones, les composants pariétaux ou des particules diffusibles.

IV.2.1. La paroi cellulaire : un composant actif de la transduction du signal au sein de l'embryon.

IV.2.1.1. Propriétés chimiques de la paroi végétale et régulation du développement embryonnaire.

Chez les végétaux, au cours des phases précoces du développement embryonnaire, certaines substances dérivées de la paroi cellulaire semblent constituer des molécules signal impliquées dans les communications intercellulaires. Des expériences menées sur des embryons somatiques ont montré que la dégradation de protéines pariétales de type arabinogalactanes (AGP) par des chitinases pourrait générer des particules solubles et diffusibles constituant d'excellents messagers potentiels (Vroemen et al., 1999). D'autres études d'embryogenèse somatique ont révélé que des protéines pariétales liées à la pathogénicité (protéines de type PR) telles qu'une osmotine ou encore une β -1-3-glucanase favorisaient l'induction de l'embryogenèse somatique (Helleboid et al., 2000). Les pectines, qui constituent avec l'hémicellulose le principal complexe polysaccharidique de la paroi cellulaire, pourraient également être impliquées dans la communication intercellulaire au sein de l'embryon grâce à leur rôle dans l'adhésion cellulaire (Malinowski and Filipecki, 2002). Le clivage des chaînes de xyloglucanes par les endotransglycosylases contribuerait à augmenter la plasticité et l'élasticité de la paroi cellulaire orientant ainsi la croissance cellulaire et pourrait également être à l'origine de composés oligosaccharidiques intervenant dans la transduction du signal. L'activité de ces enzymes a été observée lors de l'induction d'embryons somatiques ainsi que dans les stades précoces du développement d'embryons à partir de cultures cellulaires chez le concombre (Malinowski and Filipecki, 2002). Enfin, les peroxidases qui réduisent la plasticité de la paroi végétale et qui inhibent la croissance cellulaire pourraient également jouer un rôle au cours de la morphogenèse de l'embryon (Malinowski and Filipecki, 2002).

IV.2.1.2. Les plasmodesmes, douanes régulant les échanges moléculaires entre cellules.

Les cellules végétales sont interconnectées par des plasmodesmes, canaux cytoplasmiques traversant la paroi cellulaire. Les plasmodesmes qui créent une continuité symplastique entre les cellules adjacentes constituent des voies de communication intercellulaire (échanges de nutriments ou de facteurs développementaux...) facilitant la coordination des processus de croissance et de développement au sein d'un organisme. Ces structures membranaires particulières, dont l'ouverture peut être régulée en fonction du stade de développement, des tissus et des conditions physiologiques, seraient impliquées dans la régulation

d'évènements morphogénétiques complexes en autorisant le transport sélectif de déterminants moléculaires spécifiques (Crawford and Zambryski, 2001 ; Gisel et al., 1999 ; Han et al., 2000 ; Lucas et al., 1995 ; Sessions et al., 2000). L'existence de tels canaux au sein des embryons végétaux laisse penser qu'ils pourraient également jouer un rôle dans le contrôle du développement embryonnaire. Ainsi, il a été proposé qu'étant donné la petite taille de l'embryon, des gradients stables de concentration de morphogènes diffusables pourraient s'établir par le biais des plasmodesmes, notamment pendant les stades précoces de l'embryogenèse (Jürgens et al., 1991). Plus récemment, des travaux menés sur des embryons d'*Arabidopsis thaliana* à l'aide d'un traceur fluorescent ont mis en évidence une réduction de la taille limite d'exclusion des plasmodesmes (taille maximale des macromolécules pouvant diffuser librement entre les cellules) au stade torpille précoce de l'embryogenèse (Kim et al., 2002). Chez *Arabidopsis*, cette diminution de la dilatation des plasmodesmes correspond à l'initiation du développement autotrophique de l'embryon. Au cours de cette étude, deux mutants embryo-défectueux, *ise1* et *ise2* (*increased size exclusion limit of plasmodesmata*), caractérisés par un développement embryonnaire lent et des plasmodesmes restant dilatés au stade torpille, ont été identifiés. La caractérisation fonctionnelle des gènes affectés par ces mutations devrait fournir des indices sur le fonctionnement de ces canaux au cours de l'embryogenèse.

IV.2.2. Les hormones végétales, médiateurs indispensables à la formation d'un patron embryonnaire correct.

IV.2.2.1. L'auxine, molécule coordinatrice de l'établissement du patron embryonnaire.

L'identification de mutants embryo-défectueux affectés dans le transport ou la voie de réponse à l'auxine tels que les mutants *gnom*, *monopteros*, *pin-formed1* ou encore *bodenlos*, indique que l'auxine pourrait jouer un rôle capital dans la coordination de l'organisation de l'embryon en fournissant une information de position (Weijers and Jürgens, 2005). Cette molécule signal serait impliquée dans la mise en place de différentes régions de l'embryon à différents stades du développement. Elle serait en particulier liée à la spécification de la cellule apicale issue de la première division zygotique, à la formation du méristème racinaire à partir de l'hypophyse et enfin à l'établissement de la symétrie bilatérale. Les décisions développementales de l'embryogenèse précoce coïncident avec l'accumulation dynamique de l'hormone à des endroits spécifiques de l'embryon, liée à un transport localisé le long de l'axe apico-basal grâce à des protéines d'influx (dont AUX1 représente un bon candidat) et d'efflux de la famille PIN.

- **Un transport actif d'auxine accompagne la différenciation cellulaire au sein de l'embryon.**

Un premier pic d'auxine se crée dans la cellule apicale issue de la première division asymétrique du zygote grâce à un transport en provenance de la cellule basale et médié par PIN7 exprimé dans la cellule basale au niveau de la membrane apicale (Fig. 17) (Friml et al., 2003).

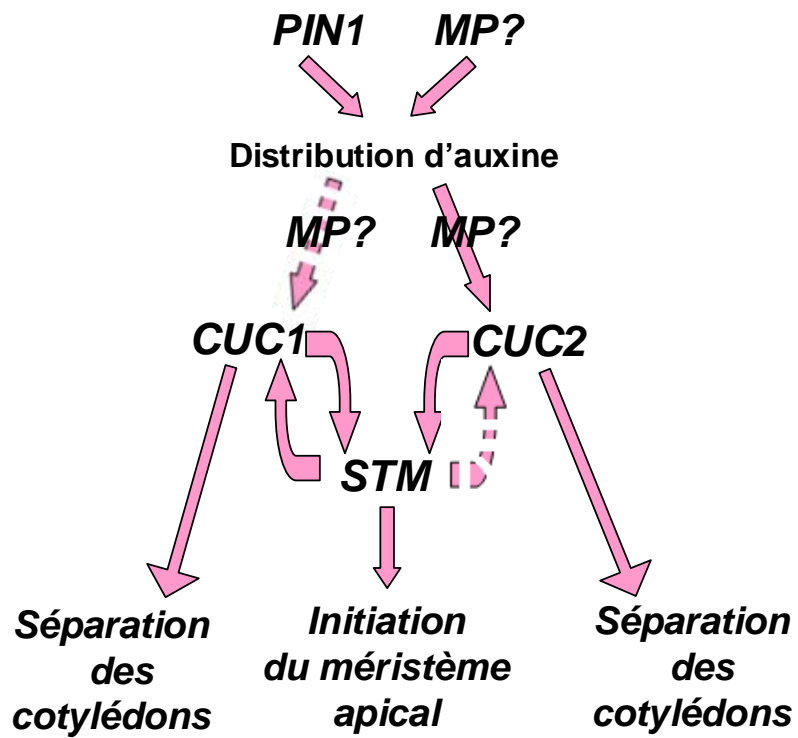


Figure 16: Modèle de mise en place de la symétrie bilatérale (d'après Aida et al., 2002).

Les flèches en pointillés indiquent un effet sur le patron d'expression alors que les flèches pleines indiquent une activation.

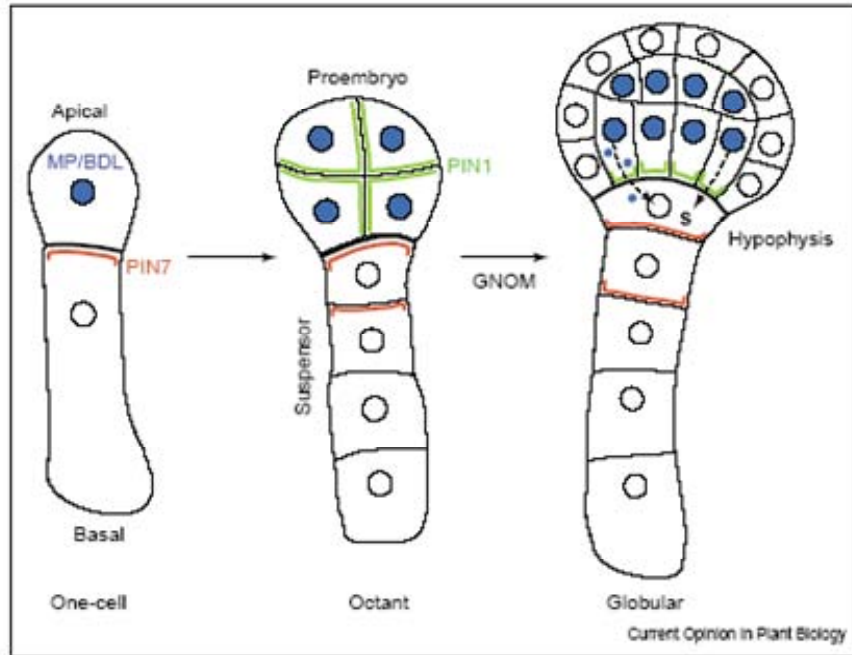


Figure 17 : Action de l'auxine au cours de la formation des axes embryonnaires.
(d'après Weijers et Jürgens, 2005)

Au stade globulaire, du fait d'une redistribution dynamique des protéines PIN1, PIN4 et PIN7, le flux s'inverse et devient basipète générant une accumulation d'auxine au niveau de l'hypophyse (Steinmann et al., 1999; Friml et al., 2002, 2003). Les protéines PIN1, initialement localisées de manière non polarisée, sont alors recrutées sur la membrane basale des cellules du proembryon via un processus d'exocytose médié par le régulateur endosomal du bourgeonnement vésiculaire GNOM. Parallèlement, la polarité de PIN7 s'inverse, ce transporteur faisant désormais face à la membrane basale des cellules du suspenseur. Au sein du méristème racinaire, le pic d'hormone est lié à l'état de différenciation, son déplacement s'accompagnant de modifications du devenir des cellules en position distale.

Au cours de la transition globulaire-coeur, suite à une réorientation apicale de PIN1, le flux d'auxine est dirigé vers l'extrémité des futurs cotylédons (et éliminé de la zone située entre les deux cotylédons) où l'hormone, en stimulant la croissance cellulaire, joue un rôle dans l'induction du développement des primordia latéraux et la détermination de leur position radiale et contribue donc à l'établissement de la symétrie bilatérale (Sabatini et al., 1999; Benkova et al., 2003; Reinhardt et al., 2003).

Ainsi, les auxines pourraient agir à la fois comme des messagers intercellulaires impliqués dans les processus d'organisation en tant que signaux de position comme dans le cas du méristème racinaire ou par signalisation à distance régulant la morphogenèse comme par exemple au niveau des cotylédons.

- **Quels sont les gènes impliqués dans la réponse à l'auxine au cours de l'embryogenèse ?**

MONOPTEROS (MP) et BODENLOS (BDL), protéines indispensables au flux correct et à la localisation coordonnée de l'auxine, semblent jouer un rôle central dans la réponse à l'auxine au cours de l'embryogenèse. Ainsi, MP, qui code pour le facteur de transcription ARF5, est capable de se lier à des éléments de promoteurs de gènes inductibles par l'auxine et est impliqué très précocément dans la mise en place du schéma d'organisation de l'embryon et des travées vasculaires (Hardtke and Berleth, 1998; Ulmasov et al., 1999). BDL code pour la protéine IAA12, un inhibiteur potentiel de MP dont la dégradation suite à un signal auxinique intracellulaire autorise l'activation par MP de gènes cibles tels que des membres de la famille de protéines à homéodomains du type HD-Zip requis pour la formation du tissu vasculaire (Baima et al., 2001; Hamann et al., 2002). MP et BDL sont co-exprimés au cours de l'embryogenèse: leurs ARNm, initialement accumulés de manière homogène au sein de la cellule apicale du proembryon et de ses dérivés, sont graduellement restreints au niveau du futur tissu vasculaire (Fig. 17). Ces gènes sont également exprimés dans la cellule apicale de l'hypophyse où ils participeraient au contrôle de la différenciation cellulaire. Les gènes régulés par l'auxine et à l'origine de devenirs cellulaires particuliers dans l'embryon pourraient donc être des gènes activés par MP de façon BDL-dépendante. Au sein de l'embryon, la réponse à l'auxine MP/BDL-dépendante pourrait s'effectuer soit directement par migration des protéines le long de l'axe principal du proembryon induisant la spécification successive des différentes régions ou grâce à l'activation d'un messager secondaire mobile agissant au niveau de l'hypophyse. Les membres de la famille

des peptides de type CLAVATA 3 (CLV3) pourraient constituer de bons candidats potentiels pour cette cascade de signalisation. En effet, CLV3, impliqué dans la communication intercellulaire, permet un maintien des cellules souches au niveau du méristème apical (Haeker and Laux, 2001). Or, un processus similaire pourrait également intervenir au niveau du méristème racinaire (Fiers et al., 2004).

Le gène AUXIN RESISTANT 6 (AXR6) (Hobbie et al., 2000) dont la mutation entraîne une réponse réduite à l'auxine et un phénotype similaire à celui observé chez les mutants *mp* et *bdl* (pas de racine, un seul cotylédon, défauts dans les tissus vasculaires) pourrait agir de concert avec ces deux gènes dans la voie de la réponse à l'auxine au cours du développement embryonnaire. Ce gène codant pour une sous-unité de la ligase de l'ubiquitine SKP1/CULLIN/F-BOX pourrait être impliqué dans la dégradation de protéines IAA en réponse à la signalisation auxinique (Hellmann et al., 2003). AXR6 (et *asf1*) montre que l'auxine est également indispensable pour des divisions cellulaires correctes et le devenir cellulaire du suspenseur (Souter and Lindsey, 2000).

PINOID, une sérine-thréonine kinase s'exprimant dès le stade globulaire, assurerait également une régulation négative de la signalisation par l'auxine (Christensen et al., 2000).

IV.2.2.2. Les stérols : régulateurs de la division et de l'expansion cellulaire lors de l'embryogenèse précoce?

Les stérols sont impliqués dans diverses fonctions de la biologie des organismes eucaryotes. Constituants membranaires (Bloch, 1983), ils agissent également comme des molécules signal notamment pour la régulation du développement (Edwards and Ericsson, 1999). Chez les végétaux, les phytostérols les plus abondants sont le sitostérol et le campestérol. Le campestérol constitue le précurseur des brassinostéroïdes qui sont des phytohormones jouant un rôle dans la croissance post-embryonnaire en régulant la division et l'élongation cellulaire (Clouse and Sasse, 1998). Le clonage du gène FACKEL (FK) codant pour une stérol C-14 reductase et la caractérisation des mutants embryo-défectueux *fk* constituent le premier lien établi entre les stérols et le développement embryonnaire. Ce gène, exprimé dès le stade globulaire, contrôle l'orientation du plan de division cellulaire (Jang et al., 2000; Schrick et al., 2000). Par la suite, HYDRA1 (HYD1) et CEPHALOPOD (CPH/SMT1), deux autres gènes codant pour des enzymes de la voie de biosynthèse des stérols (une stérol C-8,7 isomérase et une C-24 sterol methyl transférase, respectivement) et jouant un rôle au cours de l'embryogenèse, ont été isolés suite à des criblages génétiques visant à identifier des phénotypes embryo-défectueux ou des plantules de type *fk* (racine et hypocotyle courts et cotylédons malformés). Les défauts engendrés par l'altération de ces gènes sont dus à des divisions et des expansions cellulaires anormales au cours de la transition entre les stades globulaire et coeur. Alors que les gènes FK et HYD1 agissent de manière séquentielle et interdépendante, CPH agit indépendamment de ces gènes. Ainsi, les enzymes FK et HYD1, prédites comme étant des protéines membranaires du réticulum endoplasmique, pourraient interagir directement entre elles alors que la protéine CPH, qui serait un composant principal des particules lipidiques intracellulaires, séparée spatialement des deux autres protéines agirait seule. De plus,

des dosages par GC-MS du contenu en différents phytostérols de mutants *cph* ont montré que le taux de campestérol, qui est le précurseur des brassinostéroïdes, reste inchangé. Les défauts dans l'embryon ne seraient donc pas dus à l'absence de brassinostéroïdes mais plutôt à l'absence d'autres stérols essentiels ou à la présence de stéroïdes anormaux.

IV.2.2.3. Le rôle des cytokinines au cours de l'embryogenèse précoce reste encore à définir.

A ce jour, encore très peu de données sont disponibles en ce qui concerne le rôle des cytokinines au cours de l'embryogenèse.

L'identification du mutant *Amp-1*, présentant un taux élevé de cytokinines et un phénotype complexe affectant notamment le nombre de cotylédons, a montré l'importance de l'équilibre local cytokinines/auxine au stade globulaire tardif pour la détermination du nombre et de l'orientation des cotylédons (Chaudhury et al, 1993).

Les cytokinines pourraient également être impliquées dans l'organisation du patron radial. La mutation du gène WOODEN LEG (WOL)/CYTOKININ RECEPTOR 1(CRE1) codant pour un transducteur de signal 'two component hybrid' activé par des cytokinines extracellulaires, entraîne l'omission d'un cycle de divisions cellulaires au sein du primordium vasculaire et l'absence de phloème, toutes les files de cellules se différenciant en protoxylème (Mahonen et al., 2000; Hwang and Sheen, 2001).

En analogie avec les voies de signalisation à deux composants bactériennes, WOL/CRE1, après avoir perçu des signaux médiés par les cytokinines à la surface des cellules, pourrait transmettre le signal au noyau pour réguler l'expression de gènes contrôlant les divisions cellulaires au sein du primordium vasculaire. Toutefois, WOL/CRE1 ne serait pas directement impliqué dans la spécification du devenir cellulaire mais plutôt dans la régulation du nombre de couches cellulaires. Ainsi, le manque de phloème dans les mutants *wol* serait dû au fait que le xylème étant spécifié antérieurement, sa formation consommerait la totalité des cellules disponibles.

IV.3. Les récentes analyses transcriptomiques offrent un premier aperçu des principales fonctions cellulaires et voies de régulation impliquées au cours de l'embryogenèse végétale.

IV.3.1. Un large éventail de fonctions cellulaires indispensables au développement embryonnaire.

Récemment, plusieurs projets visant à caractériser le programme génétique induit au cours de différentes phases de l'embryogenèse végétale, par le biais des méthodes d'analyse globale d'expression génique, ont été entreprises chez diverses espèces (Girke et al., 2000 ; Hennig et al., 2004 ; Sprunck et al., 2005 ; Sreenivasulu et al., 2004 ; White et al., 2000). Ces études, bien qu'elles aient généralement été menées à

partir d'un mélange de tissus reproductifs et rarement à partir d'embryons isolés, ont fourni d'importantes données concernant les voies métaboliques et de régulation impliquées dans différents stades et processus de la formation des graines. Elles ont notamment permis d'effectuer une première caractérisation fonctionnelle globale de l'activité cellulaire d'embryons à différents stades de développement révélant qu'une large gamme de fonctions cellulaires était indispensable au développement embryonnaire. Ces analyses transcriptomiques, menées chez diverses espèces végétales, permettent également une comparaison globale du programme transcriptionnel induit dans différents types de graines (dicotylédones/monocotylédones, graines vertes/non vertes). Les profils fonctionnels des gènes exprimés dans les graines révélés par toutes ces études sont caractérisés par une richesse en gènes impliqués dans des fonctions cellulaires de base, les gènes associés au métabolisme primaire et à la synthèse de protéines, d'ADN ou d'ARN constituant systématiquement les principales catégories fonctionnelles. Ces profils fonctionnels illustrent le fort niveau d'activité cellulaire des tissus embryonnaires. En revanche, les facteurs de transcription et les composants de voies de signalisation ne représentent qu'une faible proportion des gènes exprimés. Toutefois, au cours d'expériences de microarrays visant à décrire les patrons d'expression des facteurs de transcription impliqués au cours de la formation de la graine chez *Arabidopsis thaliana* (de Folter et al., 2004), 70 facteurs de transcription surexprimés dans les phases d'embryogenèse précoce (entre 4 à 8 jours après fécondation) ont pu être mis en évidence parmi lesquels des membres des familles MYB et MADS, APETALA (AP2)/ERF (facteurs de réponse à l'éthylène) et bHLH (basic helix-loop-helix). Ces résultats sont en accord avec ceux obtenus lors d'une étude visant à caractériser le transcriptome induit pendant les stades précoces du développement reproductif chez *Arabidopsis thaliana* au cours de laquelle les familles de facteurs de transcription YABBY, MADS box et MYB sont apparues sur-exprimées (Hennig et al., 2004) mais aussi avec les données obtenues à partir d'embryons immatures de riz où les familles de facteurs de transcription les plus représentées sont AP2/ERF, MYB et les facteurs régulateurs de croissance (GRF) (Ye et al., 2004).

IV.3.2. Comparaison 'tissus maternels vs. embryonnaires' et identification des particularités du programme embryonnaire.

Une comparaison du transcriptome induit dans les tissus maternels et au sein de l'embryon formé suite à la fécondation a été réalisée chez l'orge (*Hordeum vulgare* L.) (Sreenivasulu et al., 2004) ou encore chez le riz (*Oriza sativa* L.) (Lan et al., 2004). Chez l'orge, les transcriptomes des deux types de tissus, qui sont très ressemblants durant l'embryogenèse précoce, divergent progressivement, les gènes induits au sein des tissus embryonnaires étant alors notamment caractérisés par une plus grande proportion du métabolisme des carbohydrates (incluant par exemple la synthèse d'amidon). De plus, les programmes génétiques des deux tissus, et plus particulièrement celui de l'embryon, subissent des bouleversements très importants durant la transition entre la phase de pré-stockage des réserves caractérisée par une division cellulaire très dynamique (de 0 à 5 jours après floraison) et la phase de dépôt des réserves (à partir du 9^{ème} jour après floraison) avec

une très grande augmentation de l'expression des gènes liés à la photosynthèse et à la production d'énergie, l'accumulation d'ATP semblant être indispensable pour la mise en place des réserves. Chez le riz, comme chez *Arabidopsis thaliana*, on assiste dans les jours qui suivent la fécondation à une activation des gènes associés à la synthèse protéique et au métabolisme des carbohydrates et des acides aminés. Des gènes liés au stress sont aussi surexprimés lors de la pollinisation et de la fécondation, suggérant que les voies de signalisation impliquées au cours de ces deux processus biologiques pourraient être très proches de celles provoquées par les réactions de défense ou de réponse au stress.

IV.3.3. Des programmes génétiques spécifiques à chaque type de graines : exemple des graines vertes et non vertes.

Une étude comparative du transcriptome exprimé dans les graines vertes d'*Arabidopsis thaliana* et dans les graines 'non vertes' du sésame (*Sesamum indicum* L.) a révélé d'importantes différences dans les profils fonctionnels de ces deux types de graines (Suh et al., 2003). Si les classes fonctionnelles correspondant au métabolisme et à l'organisation cellulaire constituent les groupes les plus représentés chez *Arabidopsis thaliana* et le sésame, la catégorie concernant la synthèse protéique est plus importante chez le sésame. De plus, en comparaison avec les transcrits les plus fréquents dans les graines d'*Arabidopsis thaliana*, la quasi absence des protéines de liaison aux chlorophylles a et b dans la graine de sésame et la relative abondance d'EST codant pour des protéines de transfert lipidique, une protéine majeure du latex, une metallothionéine ou encore des allergènes laissent supposer l'implication de processus cellulaires particuliers chez le sésame caractéristiques du développement des graines 'non vertes'. Cette hypothèse semble être confirmée par le fait que le malate pourrait être la principale source de pouvoir réducteur et de carbone pour la biosynthèse des réserves lipidiques dans les graines non photosynthétiques (Eastmond et al., 1997), comme le montre la forte expression de la malate déshydrogénase cytosolique (cMDH) et la faible expression de la pyruvate kinase plastidique alors que les principaux substrats pour l'anabolisme des lipides dans l'embryon d'*Arabidopsis thaliana* sont des intermédiaires de la glycolyse.

IV.4. Elaboration du patron embryonnaire chez les dicotylédones : organisation modulaire de l'embryon ou croissance différentielle ?

Chez *Arabidopsis thaliana*, certaines mutations affectant des gènes impliqués dans l'élaboration de l'axe apico-basal se traduisent par la délétion complète d'une ou plusieurs régions embryonnaires sans pour autant affecter la formation des autres parties de l'embryon. Cette observation suggérerait une organisation modulaire de l'embryon où les différents domaines se développeraient de manière autonome (Goldberg et al., 1994). Selon ce modèle, le plan d'organisation apico-basal de l'embryon est formé de 3 régions différentes : la région apicale formée des cotylédons, de l'apex végétatif et de la partie supérieure de l'hypocotyle, la région centrale constituée par la majorité de l'hypocotyle et la région basale composée de la

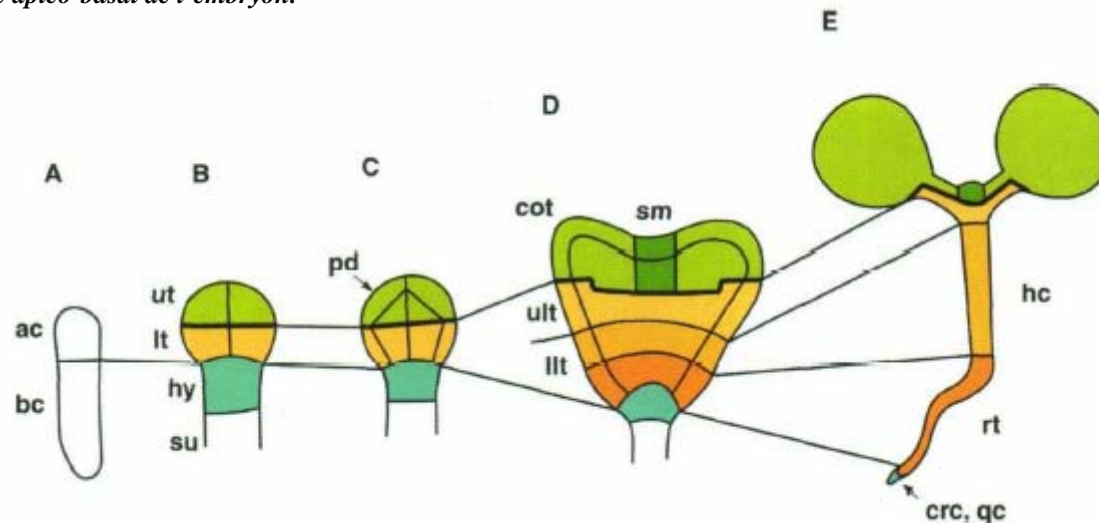
MUTANT	PROTEINE	ROLE	PHENOTYPE	PROFIL EXPRESSION	REFERENCES
Délétion centrale Fackel (FK)	C-14 stérol réductase	Division et croissance cellulaires.	Plantules avec hypocotyle très raccourci. Les cotylédons sont directement attachés à la radicule.	Absent dans l'albumen. Expression homogène dans l'embryon dès les stades précoces puis confinement au méristème caulinaire et à la pointe racinaire.	Schrack, 2000 Mayer, 1991
Délétion apicale Gurke (GK)	ACCase 1		Plantules sans cotylédons ni de méristème apical.		Mayer, 1991 Baud, 2004
Délétions apicale et basale Gnom/Emb30 (GN)	ARF-GEF associé à la membrane	Trafic vésiculaire	Elongation selon l'axe apico-basal réduite et première division plutôt symétrique. Défauts le long de l'axe apico-basal.	Expression ubiquitaire	Mayer, 1991 Shevell, 1994 Steinmann, 1999
Délétion basale Monopteros (MP)	Facteur de transcription	Axialisation des cellules en réponse à l'auxine	Plantule sans structures dérivées du domaine basal de l'embryon. Réduction générale du tissu vasculaire, formé par des lignes non continues.	Présent dans toutes les cellules subépidermales au stade globulaire précoce puis confinement progressif au sein du tissu vasculaire. Expression faible détectée dans les principaux tissus.	Mayer, 1991 Hardtke, 1998

Tableau 1 : Description de mutants affectés dans l'établissement de l'axe apico-basal de l'embryon.

Figure 18 : Etablissement des différents domaines chez *Arabidopsis thaliana*, (d'après Laux 1997).

A : Embryon une cellule
B : Octant
C : Stade dermatogène
D : Stade coeur
E : Plantule

ac : cellule apicale ; bc : cellule basale ; ut : domaine supérieur
lt : domaine central ; hy : hypophyse ; su : suspenseur ; pd : protoderme
cot : cotylédons ; sm : méristème caulinaire ; ult : domaine central supérieur
llt : domaine central inférieur ; hc : hypocotyle ; rt : radicule ;
qc : centre quiescent ; crc : coiffe racinaire



partie inférieure de l'hypocotyle et de l'apex racinaire (Fig.18). Une description détaillée des gènes dont la mutation provoque de tels défauts morphologiques est disponible dans le tableau 1. Chez ces mutants, les défauts histologiques conduisant à la suppression d'une région peuvent être retracés jusqu'à des stades très précoces du développement du proembryon impliquant qu'une carte du devenir cellulaire serait élaborée très tôt au cours du développement. Ainsi, la limite entre les zones basale et centrale apparaît dès la première division asymétrique du zygote ; en revanche, la frontière entre les zones apicale et centrale, constituée par la limite O', s'établit lors de la première série de divisions transverses dans l'embryon menant au stade octant. Chacun des gènes impliqués dans l'élaboration d'une région particulière de l'axe apico-basal définirait un patron de division spécifique à ce territoire, des défauts de divisions cellulaires au sein d'un territoire n'affectant pas la formation des territoires adjacents. Toutefois, des interactions entre les cellules des différents domaines permettraient une coordination du développement embryonnaire global. Ainsi, comme indiqué par le phénotype du mutant *gurke* présentant une altération de la zone apicale, les cotylédons seraient initiés dans la zone apicale qui par la suite recruterait la zone centrale dans leur formation. De plus, dans chaque région embryonnaire, des facteurs de transcription spécifiques, interagissant avec différents éléments régulateurs des promoteurs, seraient exprimés, la combinaison de ces facteurs et de ces éléments générant le patron transcriptionnel global au sein de l'embryon.

De plus, la caractérisation phénotypique de mutants embryo-défectueux a révélé que si la différenciation tissulaire pouvait se dérouler indépendamment de la morphogénèse (par exemple, chez le mutant *gnom* ou *fackel*), en revanche, la morphogénèse, ne peut avoir lieu sans qu'il y ait différenciation cellulaire (Mayer et al., 1991 ; Goldberg et al., 1994). Ainsi, alors que la perte d'un domaine embryonnaire n'altère pas l'élaboration des différentes couches de tissus dans les autres régions, le bon établissement de l'axe radial semble indispensable à la construction correcte de l'axe apico-basal, les cellules au sein de l'axe radial interagissant probablement entre elles le long de l'axe principal de l'embryon. Les mutants *keule* et *knolle* d'*Arabidopsis thaliana*, présentant simultanément des défauts de morphologie et une altération des processus de la différenciation tissulaire lors de l'établissement du protoderme, illustrent parfaitement ce phénomène.

Le modèle d'organisation modulaire de l'embryon végétal a été proposé par Jürgens et dérive du modèle de développement de la Drosophile qui se base sur l'existence de différents domaines embryonnaires. Kaplan et Cooke (1997) ont proposé un modèle différent où les mutations de type *monopteros*, *gurke*, *gnom* ou *fackel* ne résulteraient pas de la délétion d'un domaine embryonnaire mais d'un changement de croissance différentielle. Ces gènes seraient, en fait, impliqués dans des processus cellulaires fondamentaux dont l'altération empêcherait le bon déroulement du développement embryonnaire entraînant de graves défauts morphologiques pouvant aller jusqu'à la délétion de régions entières et menant fréquemment à l'arrêt du développement embryonnaire. Ce nouveau modèle expliquerait pourquoi aucun des gènes clonés responsables des phénotypes mutants n'est spécifique à l'embryon.

Chapitre II - Le tournesol : un modèle d'étude complémentaire de l'embryogenèse zygotique chez les Dicotylédones.

I. Le tournesol : une plante d'intérêt agronomique.

I.1. Le tournesol : une plante ayant subi les effets de la domestication.

L'étude de l'embryogenèse, qui constitue une étape clé du développement des végétaux au cours de laquelle les potentialités de la future plante sont définies, a pour but principal l'amélioration des caractéristiques agronomiques des plantes cultivées grâce aux méthodes biotechnologiques récemment développées. Avec l'apparition de l'agriculture et la domestication des plantes, des modifications de plusieurs caractères génétiques tels que la réduction de la dispersion des graines, l'amélioration du rendement et de la qualité nutritionnelle des graines, ou encore la synchronisation de la floraison et de la maturation des graines, ont progressivement eu lieu. De ce fait, l'hypothèse de l'existence de différences notables entre les mécanismes moléculaires et physiologiques induits lors du développement embryonnaire des plantes cultivées et non cultivées ne peut être rejetée. Ainsi, les plantes oléoprotéagineuses telles que le tournesol pourraient disposer de voies métaboliques et de mécanismes de régulation du développement embryonnaire particuliers. *Arabidopsis thaliana*, qui constitue la plante modèle pour l'étude de l'embryogenèse chez les dicotylédones, est une mauvaise herbe ne présentant aucun intérêt agronomique et n'ayant donc jamais subi de pression de sélection artificielle. Dès lors, on peut penser que ce système expérimental ne permet pas d'avoir accès aux spécificités des plantes domestiquées. De même, avant d'entreprendre des programmes d'amélioration génétique des semences, les connaissances obtenues chez *Arabidopsis thaliana* devront être préalablement validées chez les espèces cultivées d'intérêt.

Une autre différence métabolique majeure existe entre les graines d'*Arabidopsis thaliana* et les graines de tournesol : alors que les premières sont chlorophylliennes et décrites comme des 'graines vertes' (green seeds), les graines de tournesol, tout comme les graines de sésame ou de ricin, sont dépourvues de tout pigment chlorophyllien et sont donc qualifiées de 'graines non vertes' (ou 'non green seeds'). Une étude comparant les gènes exprimés dans les graines de sésame et d'*Arabidopsis thaliana* semblent indiquer que ces caractéristiques cacheraient des adaptations métaboliques différentes notamment au cours de la phase de maturation de la graine pendant l'accumulation des réserves en huile (Suh et al., 2003).

I.2. Historique de la culture du tournesol.

Le tournesol, *Helianthus annuus* L., est une plante originaire d'Amérique du Nord : la plus ancienne région de culture du tournesol se situe en Arizona et au Nouveau Mexique. Il y a environ 3000 ans avant notre ère, les Amérindiens utilisaient déjà ses graines et son huile à des fins alimentaires, médicinales, tinctoriales et

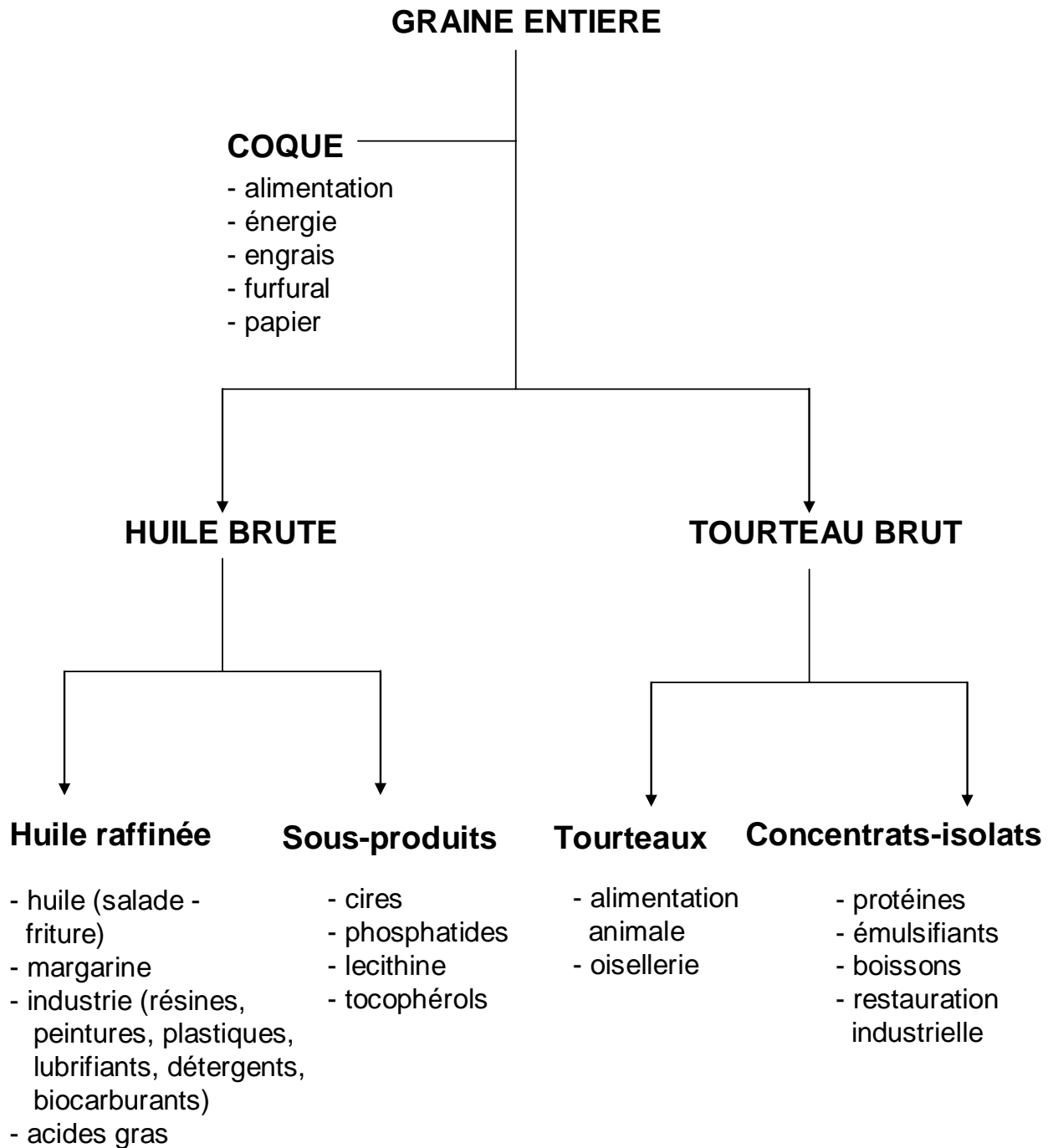


Figure 19: Valorisation du tournesol.

cérémoniales. Au XVIème siècle, le tournesol fût introduit en Europe par les navigateurs espagnols de retour des Amériques, où il fût d'abord utilisé comme plante ornementale puis, au XVIIIème siècle, comme plante oléagineuse à usage alimentaire suite à la sélection par les russes de variétés à gros capitules et à grosses graines ou à forte teneur en huile. Toutefois, la culture du tournesol n'a connu de véritable essor qu'aux alentours des années 60 avec l'exploitation de la vigueur hybride et surtout grâce à la découverte de la stérilité mâle cytoplasmique par Leclercq (1966) ainsi que celle des gènes de restauration de la fertilité. En France, le développement de cette culture, aujourd'hui largement étendue dans le Sud-Ouest, le Centre et en Côte d'Or, serait plus particulièrement lié à la mise au point de nouvelles variétés (par croisement et sélection), notamment d'hybrides résistants au mildiou, ainsi qu'à un évènement historique, la décolonisation, qui mit fin au quasi monopole de l'huile d'arachide.

I.3. Quelques généralités sur le tournesol.

Le tournesol cultivé (*Helianthus annuus* L.) est une espèce diploïde ($2n=34$ chromosomes) appartenant à la famille des Composées et à la sous-famille des Tubiliformes. Le genre *Helianthus* L. regroupe plus de 60 espèces végétales parmi lesquelles de nombreuses plantes herbacées, annuelles ou pérennes et comporte deux espèces cultivées majeures, le tournesol (*Helianthus annuus* L.), cultivé principalement pour ses graines riches en huile et le topinambour (*Helianthus tuberosus* L.), produit pour la valeur alimentaire de ses tubercules.

Le nom botanique *Helianthus* (du grec *hélios* : 'soleil' et *anthos* : 'fleur') provient du mouvement héliotropique du capitule, *annuus* faisant allusion à sa durée de vie annuelle.

I.4. Intérêt économique et production.

I.4.1. Le tournesol, une culture aux débouchés très diversifiés.

La culture de tournesol présente de nombreux débouchés que ce soit dans l'alimentation humaine, animale ou dans la valorisation des sous-produits des industries huilières (Fig.19).

I.4.1.1. Production d'huile alimentaire.

Le tournesol est principalement destiné à la production d'huile alimentaire, qu'elle soit d'assaisonnement, de cuisson ou utilisée pour la fabrication de margarine. Il constitue, avec le colza et l'olivier, l'une des trois sources principales d'huile alimentaire en Europe. L'huile de tournesol présente d'excellentes vertus pour la santé humaine du fait notamment de sa grande richesse en acides gras insaturés (de 85 à 91%) tels que l'acide oléique (23%) jouant un rôle essentiel dans la prévention des affections cardio-vasculaires ou encore l'acide linoléique (65%) qui est un acide gras essentiel (non synthétisé par le métabolisme et impérativement apporté par l'alimentation) connu pour réduire le taux de cholestérol sanguin et pour favoriser le processus

de coagulation du sang. L'huile de tournesol est également riche en divers produits insaponifiables tels que les stérols ou les tocophérols (vitamine E) qui possèdent une activité antioxydante et limitent son rancissement. Ainsi, l'huile de tournesol, qui est dotée de grandes qualités nutritionnelles, est aujourd'hui consommée dans de nombreux pays tant pour ses propriétés culinaires que pour ses caractéristiques organoleptiques et diététiques.

1.4.1.2. Utilisation des protéines en alimentation humaine.

De par leurs propriétés technologiques (solubilité dans l'eau, capacité d'absorption des matières grasses, aptitude au fouettage et à l'émulsification) ou nutritionnelles (absence de goût spécifique, bon équilibre en minéraux et en vitamines hydrosolubles du groupe B, faible taux en sucre), les protéines issues des graines de tournesol possèdent de nombreuses applications en alimentation humaine.

1.4.1.3. Utilisation des tourteaux en alimentation animale.

Les tourteaux, qui constituent le produit secondaire de l'extraction de l'huile à partir des graines, contiennent jusqu'à 35% de protéines, parmi lesquelles des protéines riches en acides aminés soufrés notamment la méthionine très recherchée dans l'alimentation du bétail. Ces résidus, qui ne présentent aucun facteur anti-nutritionnel, constituent donc un bon aliment de base pour les animaux d'élevage polygastriques. Cependant, les tourteaux de tournesol sont caractérisés par un apport énergétique faible, un contenu en fibres trop élevé et une carence en lysine, acide aminé indispensable à l'alimentation animale. Ces inconvénients peuvent toutefois être compensés en pratiquant un bon décorticage ainsi qu'une supplémentation en lysine. Depuis 1978, la consommation de tourteaux progresse régulièrement avec une augmentation moyenne de 25% par an.

Les graines peuvent également être directement utilisées dans les oisellereries.

1.4.1.4. Valorisation non alimentaire.

La valorisation non alimentaire du tournesol représente un atout non négligeable de cette culture. L'industrie chimique utilise l'huile de tournesol extraite à chaud pour la fabrication de nombreux produits tels que des lubrifiants, des plastiques, des peintures, des résines, des détergents, des biocarburants, des émulsifiants... Les autres sous-produits du tournesol peuvent également être valorisés : les coques des graines peuvent être brûlées à des fins énergétiques ou utilisées pour la fabrication de papier, la tige riche en chlorure de potassium (environ 36%) peut être employée pour la préparation d'engrais, les ligules peuvent entrer dans la composition d'une teinture jaune naturelle. Le tournesol comporte aussi des applications médicinales : les graines et les feuilles possèdent par exemple des propriétés efficaces contre les affections pulmonaires. Récemment, un programme de recherche européen visant à valoriser les phytostérols présents dans les

exudats de l'industrie huilière et possédant des propriétés anticancéreuses et de prévention des maladies cardiovasculaires a été développé.

1.4.2. Importance économique.

Avec le soja, le colza, le palmier à huile et l'arachide, le tournesol est l'une des plantes oléagineuses les plus importantes dans le monde. Sa culture représente de l'ordre de 20 millions d'hectares pour une production de 26.1 millions de tonnes en 2003. Les principaux pays producteurs sont les pays de l'ex-URSS (avec une production supérieure à 4 millions de tonnes en Ukraine et en Russie), l'Argentine (3.75 millions de tonnes) et l'Union Européenne (plus de 3 millions de tonnes). Au sein de l'Union Européenne, la France, avec une surface cultivée stable de 700 000 hectares et une production de 1.4 millions de tonnes, est le premier pays producteur, suivie de la Hongrie et de l'Espagne. Alors que le rendement moyen mondial est d'environ 13q/ha, le rendement moyen en France, qui oscille entre 20 et 25 q/ha, est de loin le plus élevé.

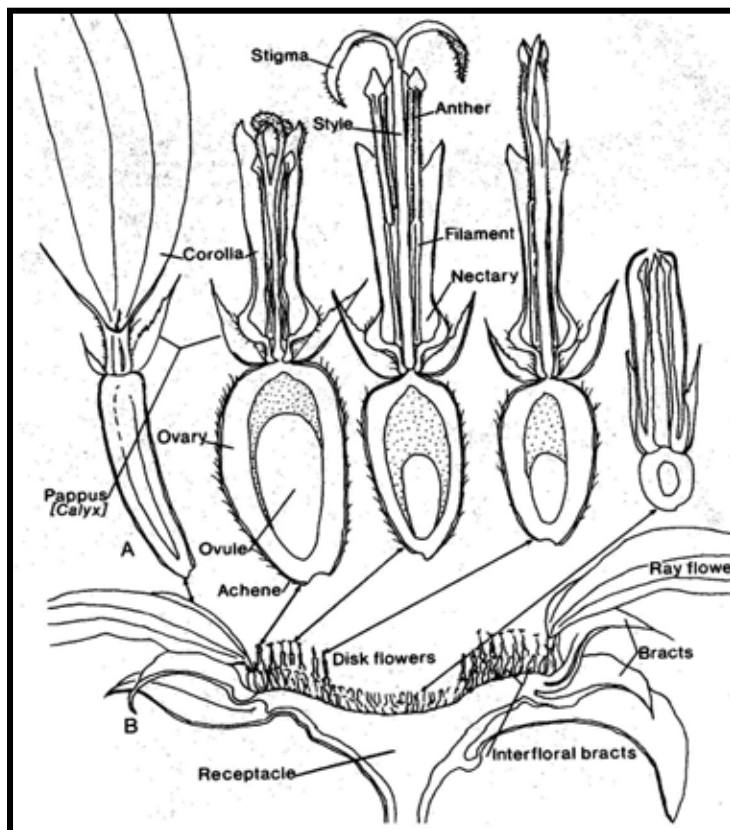
Le tournesol, ayant permis de produire environ 8 millions de tonnes d'huile alimentaire en 2002 (soit plus de 8.5% de la production mondiale d'huile végétale), se classe au 4^{ème} rang des ressources oléoprotéagineuses après le soja, l'huile de palme et le colza représentant respectivement 25.7, 25, et 12.4 millions de tonnes d'huile produite. Actuellement, le tournesol représente environ 50% des graines oléagineuses triturées en France et 18% pour l'ensemble de l'Union Européenne où il constitue la deuxième culture oléoprotéagineuse après le colza. La production française d'huile de tournesol qui s'élevait à 46 000 tonnes en 2002 représente quasiment 50% de la production européenne et plus de 5% de la production mondiale.

En France comme en Europe, on assiste, ces dernières années, à un recul de l'ordre de 20% du tournesol industriel au profit des variétés cultivées à des fins alimentaires. Toutefois, 75% des surfaces en tournesol industriel restent encore aujourd'hui implantés dans l'hexagone.

Afin de faire face à la concurrence des autres plantes oléagineuses, les producteurs et semenciers accordent de plus en plus d'importance à l'amélioration du rendement et de la qualité des graines de tournesol, les principaux critères pris en compte pour la sélection variétale étant l'augmentation de la productivité (en grains et en huile), la qualité des graines (taux protéique et lipidique, nature des acides gras, aptitude au décorticage), mais aussi la tolérance (aux parasites, aux contraintes climatiques), la résistance à la verse et la précocité.



A. Photographie d'un capitule de tournesol



B. Représentation schématique d'un capitule de tournesol et Identification des différents types de fleurs.

Figure 20: Le capitule, inflorescence du tournesol.

II. Une plante présentant de nombreux atouts pour l'étude de l'embryogenèse.

II.1. Un appareil reproducteur à l'anatomie avantageuse pour l'étude de l'embryogenèse.

Les organes reproducteurs du tournesol possèdent plusieurs caractéristiques morphologiques intéressantes conférant à cette plante de nombreux avantages pour l'étude des phases précoces de l'embryogenèse.

Tout d'abord, l'inflorescence du tournesol, qui est un capitule (Fig.20), porte un grand nombre de fleurs ce qui permet de récolter une quantité importante d'embryons à partir d'un même individu. Sur le capitule, les fleurs sont de deux types :

- au centre, de 50 à 3000 fleurs tubulées hermaphrodites (ou fleurons) assurent la reproduction.
- à la périphérie, disposées sur un seul rang, se trouvent des fleurs ligulées stériles de couleur jaune ayant pour rôle principal d'attirer les insectes vecteurs de la pollinisation (pollinisation entomophile).

Au sein d'un même capitule, la floraison des fleurons est décalée, ce qui permet d'avoir accès simultanément à des embryons à différents stades de développement. Elle se déroule pendant 10 à 15 jours de la périphérie vers le centre, 3 à 4 rangs de fleurons s'ouvrant chaque jour. Le tournesol est une plante prothandre: les anthères libèrent le pollen alors que l'appareil reproducteur femelle n'est pas encore mature ; puis, le style croît et le stigmate bifide s'ouvre et devient réceptif au pollen pendant quelques jours. La fécondation a généralement lieu le deuxième jour après quoi les stigmates flétrissent rapidement.

Bien que le tournesol soit une plante de préférence allogame, l'autofécondation reste possible.

Les graines de tournesol, qui mesurent environ 1.5cm, sont beaucoup plus grandes comparées aux graines d'*Arabidopsis thaliana* présentant un diamètre de 500µm. La récolte et le criblage des différents stades de développement, qui sont réalisés par microdissection sous une loupe binoculaire, sont donc moins laborieux. Enfin, la graine du tournesol ou akène (fruit sec ne s'ouvrant pas à maturité) est exalbuminée : les réserves en protéines (représentant de 15 à 21%) et les réserves en huile (de 20 à 60%) contenues dans les oléosomes, qui sont des vésicules intracellulaires, sont accumulées au sein même de l'embryon. L'étude du développement embryonnaire du tournesol est donc d'un intérêt tout particulier.

II.2. L'embryogenèse somatique du tournesol : un modèle d'étude des phases d'initiation de l'embryogenèse.

Chez le tournesol, des embryons somatiques peuvent être obtenus à partir de protoplastes d'hypocotyle (cellules végétales dépourvues de paroi) cultivés *in vitro* (Fig. 21). Dans ce modèle d'embryogenèse somatique, un protoplaste en culture en milieu solide (enrobé dans une matrice d'agarose) se divise de manière asymétrique et forme un embryoïde, structure proche d'un embryon zygotique dont le développement s'arrête prématurément (dès le stade cœur). En revanche, un protoplaste cultivé en milieu

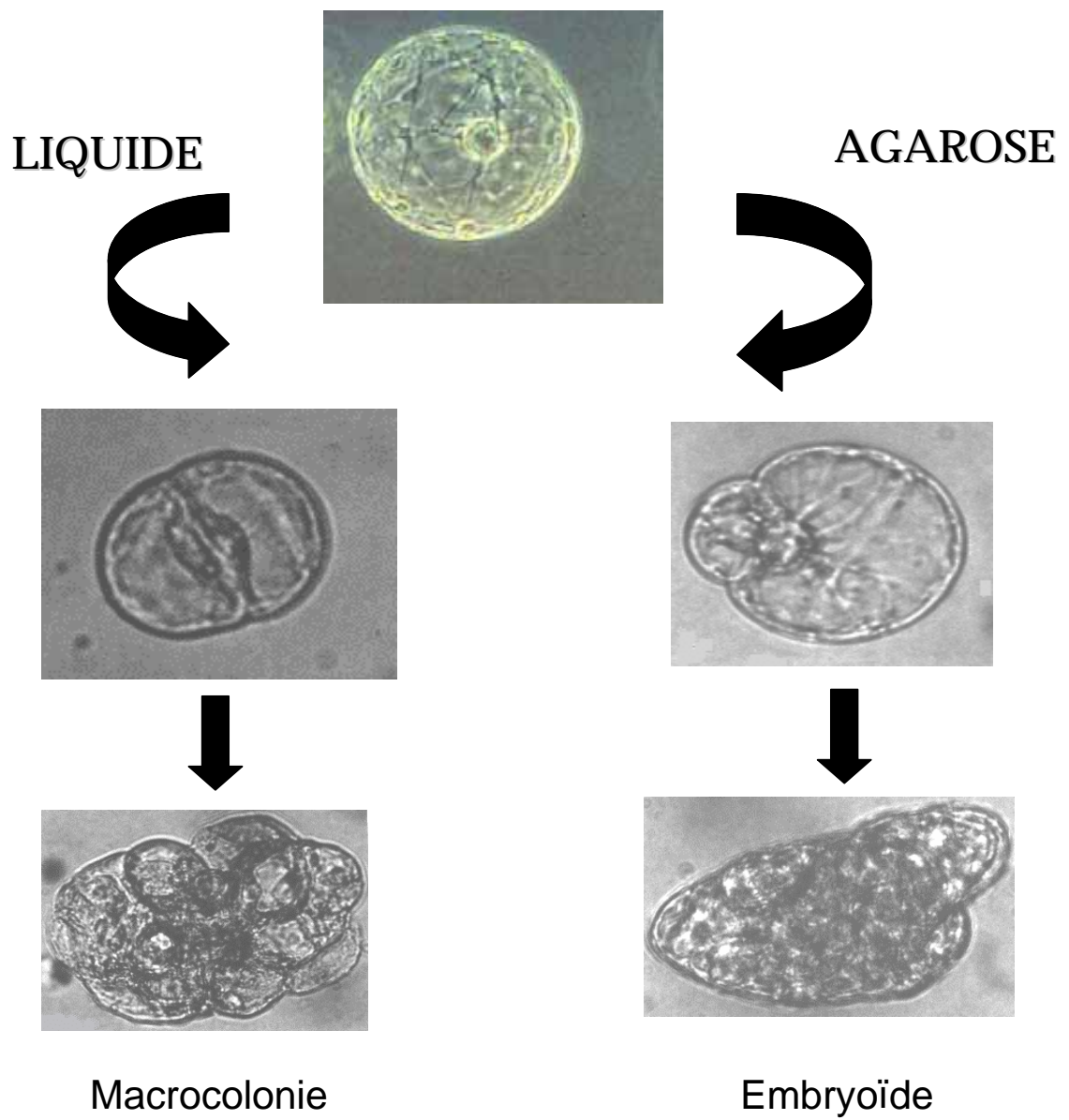


Figure 21: Le modèle somatique : un système pour l'étude de la première division asymétrique.

liquide (sans agarose) se divise symétriquement générant alors une macrocolonie, véritable amas de cellules somatiques (Chanabé et al., 1989). Ce système d'embryogenèse somatique a été proposé comme modèle d'étude de l'acquisition de la polarité embryonnaire et de l'induction de la première division cellulaire asymétrique (Alibert et al., 1994).

II.3. Des outils disponibles en constant développement.

Au cours de ces dernières années, une large gamme d'outils génétiques, génomiques et informatiques a été développée afin de faciliter l'étude de cette plante d'intérêt agronomique encore mal connue. Des cartes génétiques couvrant la quasi-totalité du génome du tournesol (représentant environ 300 Mb) sont actuellement disponibles (Flores Berrios et al., 2000 ; Tang et al., 2002 ; Al Chaarani et al., 2004). Deux banques BACs ont également été décrites (Caboche et Boucly, 2000 ; Gentzbittel et al., 2002). Parallèlement, des programmes de séquençage d'EST à grande échelle ont été entrepris permettant la création de plusieurs banques d'ADNc, obtenues à partir de divers tissus et conditions de culture, et générant plus de 66 000 entrées dans Genbank. Une collection de 2000 mutants d'irradiation est actuellement criblée à la recherche de plantes montrant des altérations du développement embryonnaire selon les critères phénotypiques définis par Meinke (Meinke, 1985). Un site Web nommé Iccare (<http://genopole.toulouse.inra.fr/bioinfo/Iccare/>, Muller et al., 2004) permettant l'étude de la synténie entre le génome du tournesol et d'*Arabidopsis thaliana* et aidant également à la désignation de sondes spécifiques a été réalisé.

III. Etat actuel des connaissances sur le développement embryonnaire chez le tournesol.

III.1. Embryogenèse somatique et analyse du transcriptome lors de l'acquisition de la polarité embryonnaire.

Afin d'élucider les mécanismes moléculaires et physiologiques impliqués dans l'établissement de la polarité embryonnaire chez le tournesol, une analyse du transcriptome induit dans les phases initiales de l'embryogenèse a été menée dans notre laboratoire en s'appuyant sur le modèle d'embryogenèse somatique *in vitro*. Quatre banques d'ADNc ont été construites à partir de protoplastes cultivés en conditions embryogènes (milieu solide) et non embryogènes (milieu liquide) : une banque de référence pour chaque condition et deux banques SSH réciproques. Pour chacune des banques, environ 1 000 EST ont été séquencées parmi lesquelles 2479 séquences, représentant 1502 gènes potentiellement différents, ont été soumises. Suite à une recherche d'homologie par BLASTX avec l'ensemble des protéines répertoriées dans

les banques de données, les principales fonctions cellulaires impliquées dans le développement des protoplastes dans les deux conditions de culture ont été mises évidence. Dans les deux cas, les gènes codant pour des protéines du métabolisme protéique et du métabolisme primaire sont les plus représentés, ce profil fonctionnel reflétant un niveau élevé d'activité cellulaire des protoplastes cultivés *in vitro*. La différence majeure entre les protoplastes en conditions embryogènes et non embryogènes consiste en une proportion très importante des gènes codant pour des protéines associées au stress dans le cas des protoplastes cultivés en milieu solide. Si ce résultat peut être lié aux conditions de culture difficiles que doivent subir les protoplastes inclus dans l'agarose, il peut aussi dévoiler une réelle intervention des phénomènes de stress dans l'initiation du développement embryonnaire et l'acquisition de la polarité embryonnaire. Des gènes potentiellement différenciellement exprimés entre les deux conditions de culture ont également été identifiés grâce à des analyses de northern digital parmi lesquelles une protéine ribosomiale et une protéine de fonction inconnue qui seraient surexprimées en situation embryogène. (Tamborindéguy et al., 2004a).

Ces banques d'ADNc, couplées à la banque d'ADNc réalisée à partir d'embryons zygotiques au stade cotylédonaire produite dans le cadre de ma thèse, ont ensuite été utilisées pour construire un microarray sur membrane nylon comptant 2368 séquences différentes représentant 1911 gènes uniques. L'hybridation de ce microarray avec des populations d'ADNc issues de protoplastes cultivés en conditions embryogènes et non embryogènes ainsi que des protoplastes fraîchement isolés et fraîchement inclus dans l'agarose a révélé 369 gènes différenciellement exprimés entre au moins deux conditions. L'analyse des patrons d'expression de ces gènes a permis d'élaborer un modèle hypothétique du processus d'acquisition de la polarité embryonnaire dans lequel le calcium, l'auxine et la sécrétion de composants cellulaires joueraient des rôles essentiels (Tamborindéguy, soumis).

III.2. Description de QTLs impliqués dans les processus du développement embryonnaire du tournesol.

Le développement rapide et récent de marqueurs moléculaires a contribué à l'obtention de cartes génétiques saturées ayant permis la cartographie de QTLs. Ainsi, plusieurs QTLs liés à l'embryogenèse somatique ou contrôlant d'importantes caractéristiques agronomiques des graines de tournesol telles que le poids des graines, la composition en huile des graines ou encore la capacité à germer ont pu être caractérisés (Al-Chaarani et al., 2004, 2005 ; Bert et al., 2003, Flores Berrios et al., 2000).

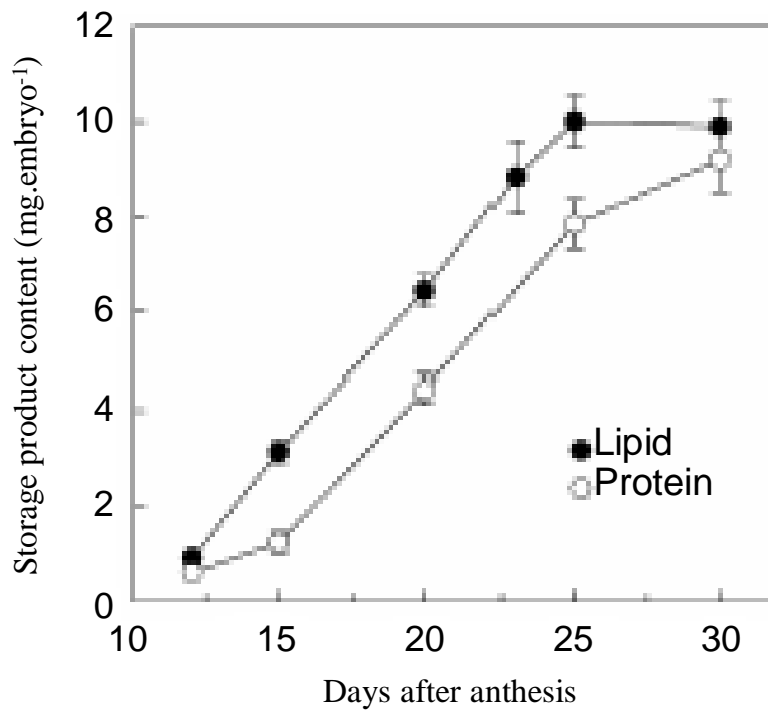


Figure 22: Cinétique d'accumulation des réserves lipidiques et protéiques au sein de la graine de Tournesol (d'après Pleite, 2005)

III.3. Contrôle moléculaire de la mise en place des réserves au sein de l’embryon de tournesol.

La graine de tournesol étant exalbuminée, les réserves sont accumulées au sein même de l’embryon. Ces réserves, qui sont principalement de nature lipidique et protéique, confèrent à cette culture une grande valeur agronomique. Cependant, très peu de données sont disponibles concernant le contrôle génétique de la mise en place des réserves au cours de l’embryogenèse chez le tournesol.

III.3.1. Cinétique de l’accumulation des réserves au sein de l’embryon.

L’accumulation des réserves au sein de l’embryon a lieu au cours de la phase de maturation de l’embryogenèse. Chez le tournesol, le stockage des réserves en huile est linéaire entre le 12^{ème} et le 25^{ème} jour après l’anthèse puis il cesse alors que le dépôt de protéines s’effectue de manière linéaire entre le 15^{ème} et le 25^{ème} jour après l’anthèse et se poursuit jusqu’au 31^{ème} jour (Fig. 23) (Pleite et al., 2005). La quantité d’amidon emmagasinée dans l’embryon de tournesol, qui représente moins de 1.6% des contenus en huiles et en protéines cumulés, est infime ; ce composé pourrait en fait être recyclé dans l’embryon comme dans les autres graines oléagineuses telles que le colza (*B. napus*) (da Silva et al., 1997).

III.3.2. Modèle de synthèse des acides gras au sein de l’embryon de tournesol.

Les acides gras sont synthétisés au sein des plastes de l’embryon qui, pour le tournesol comme pour le sésame ou encore le ricin, sont hétérotrophes c’est-à-dire qu’ils ne sont pas capables de produire de chlorophylle et de réaliser la photosynthèse. Le malate et le pyruvate constituent les métabolites les plus susceptibles d’être utilisés pour l’incorporation de carbone lors de la synthèse des acides gras, le malate représentant la principale source (Pleite et al., 2005). Le pouvoir réducteur nécessaire à la production des réserves en huile peut être produit grâce au métabolisme du malate par l’enzyme NADP-malique et/ou du glucose-6-phosphate par la voie oxydative des pentoses phosphate. Le rôle prépondérant du malate dans la synthèse des réserves en huile chez les plantes aux graines ‘non vertes’ a déjà été rapporté chez le ricin (Eastmond et al., 1997).

Les études moléculaires visant à caractériser la régulation de la synthèse des réserves lipidiques dans la graine ont surtout porté sur l’analyse de la réaction de désaturation de l’acide oléique en acide linoléique, constituant les principaux acides gras des réserves en triacylglycérols (TAG) de graines oléagineuses (Garcia-Diaz et al., 2002 ; Martinez-Rivas et al., 2000). Cette réaction est catalysée par l’enzyme oleoyl phosphatidylcholine desaturase microsomale (ou FAD2) dont le niveau d’activité est contrôlé par la température et la disponibilité en oxygène.

Plus récemment, les gènes codant pour des enzymes clés de la biosynthèse des acides gras chez le tournesol, qui sont respectivement la stearyl-acyl carrier protein (ACP) desaturase et les acyl-ACP thioesterase HaFatA1 et HaFatB dont la spécificité de substrat détermine les acides gras disponibles pour la synthèse des lipides membranaires et de réserves dans la graine, ont été clonés (Serrano-Vega et al., 2003, 2005). Toutefois, aucune donnée concernant le profil de ces gènes au cours de l'embryogenèse ne sont disponibles.

III.3.3. Les hélianthinines, principales protéines de réserves du tournesol.

Les hélianthinines, qui sont des globulines de type 11S, représentent les principales protéines de réserves du tournesol et sont exprimées exclusivement dans la graine en formation (Bogue et al., 1990). Ces protéines sont codées par une petite famille multigénique incluant un minimum de deux sous-familles divergentes. Deux membres de cette famille, HaG3-A et HaG3D, ont été isolés et caractérisés. Les protéines précurseurs prédites à partir des séquences des gènes HaG3 ont une taille de 493 acides aminés, sont riches en glutamine et autres acides aminés riches en azote et contiennent la séquence NGVEETICS hautement conservée parmi les protéines de réserves 11S et impliquée dans les processus protéolytique de ces polypeptides (Vonder Haar et al., 1988). Des éléments régulateurs du promoteur ont été identifiés parmi lesquels trois motifs de liaison aux protéines nucléaires de la graine ainsi que des éléments régulés par l'acide abscissique (Jordano et al., 1989 ;Nunberg et al., 1994).

La séquence complète du gène HaG5 codant pour une albumine 2S représentant un autre type de protéine de réserve a également été décrite (Allen et al., 1987). Le précurseur protéique prédit, qui est riche en résidus glutamine (24%), a une taille de 295 acides aminés et subit d'importantes modifications post-transcriptionnelles. HaG5 constitue l'un des membres d'une petite famille multigénique comprenant au moins deux gènes divergents.

Chapitre III – Intérêts et objectifs du projet de thèse.

Les connaissances actuelles sur l'embryogenèse végétale sont encore trop parcellaires.

Jusqu'à très récemment, la grande majorité des connaissances sur les mécanismes moléculaires régulant le développement embryonnaire des végétaux provenait de l'étude de mutants menée sur des plantes modèles telles qu'*Arabidopsis thaliana* ou le maïs. Bien que l'analyse par mutagenèse, qui a déjà permis le clonage et la caractérisation d'un grand nombre de gènes jouant des rôles clés dans cette phase biologique, soit très prolifique, elle revêt quelques inconvénients et connaît des limites. Du fait de l'existence de redondances génétiques et fonctionnelles n'autorisant pas la visualisation phénotypique des effets des mutations et du fort taux de létalité entraîné par des mutations à des stades aussi critiques de développement, cette approche ne permet pas d'identifier tous les gènes essentiels au développement embryonnaire et d'élucider l'ensemble des processus nécessaires à la formation de la graine. De nos jours, le développement de techniques permettant l'analyse simultanée de l'ensemble du génome à partir d'échantillon végétal réduit a permis d'intensifier les recherches portant sur cette étape du développement des végétaux qui revêt un intérêt biologique et agronomique évident et d'accélérer les découvertes concernant les gènes impliqués dans la formation de l'embryon, celles-ci restant toutefois encore insuffisantes et trop parcellaires.

Les connaissances sur l'embryogenèse du tournesol sont quasi-inexistantes.

Bien qu'étant une plante d'intérêt agronomique, le tournesol est encore peu étudié et mal connu. Jusqu'à ces dernières années, les seules données disponibles quant à la régulation génétique du développement de la graine de tournesol concernaient la description moléculaire de gènes impliqués dans l'accumulation des réserves en huile et exprimés tardivement au cours de l'embryogenèse ou portaient sur la caractérisation de régions du génome régulant des caractères quantitatifs liés à l'embryogenèse (taille des graines, accumulation des réserves en huile, germination...). Plus récemment, des travaux visant à analyser le transcriptome lors de l'acquisition de la polarité embryonnaire et s'appuyant sur un modèle expérimental d'embryogenèse somatique *in vitro* ont été menés au sein de notre laboratoire et ont permis d'identifier des processus moléculaires pouvant jouer un rôle clé dans l'étape d'initiation du développement embryonnaire. De larges programmes de séquençage d'EST ont également été entrepris par la communauté scientifique générant une grande quantité de ressources concernant les gènes du tournesol exprimés dans différentes conditions et différents tissus. Pour autant, du fait des difficultés inhérentes à l'étude des jeunes embryons, des données portant sur les phases précoces de morphogenèse de l'embryogenèse ont toujours manqué.

Présentation du projet de thèse.

L'objectif de mon travail de thèse est d'étudier, par les méthodes de la génomique fonctionnelle, le transcriptome induit au cours de l'embryogenèse précoce chez le tournesol et plus particulièrement aux stades globulaire, cœur et cotylédonaire qui constituent des étapes clés du développement embryonnaire. En

effet, ceux-ci correspondent respectivement au début de la mise en place de la symétrie radiale (et notamment à la formation du protoderme), à l'établissement de la symétrie bilatérale et au début de la phase de croissance et de maturation de l'embryon. De plus, de nombreux mutants embryon-défectueux d'*Arabidopsis thaliana* sont arrêtés au stade globulaire tardif. Ceci impliquerait que la transition du stade globulaire au stade cœur représenterait un processus de différenciation complexe et critique au cours duquel un grand nombre de gènes serait impliqué.

Afin de caractériser les gènes induits lors de ces phases précoces de l'embryogenèse, des banques d'ADNc de référence ont été construites pour chacun des stades d'intérêt (HaGlbR, HaHeaR et HaCotR pour les stades globulaire, cœur et cotylédonaire précoce, respectivement) ainsi qu'une banque de type SSH comparant les transcrits d'embryons aux stades cœur et cotylédonaire et enrichie en gènes préférentiellement exprimés au stade cœur. L'obtention et l'analyse des banques d'ADNc à partir des embryons immatures de tournesol à des stades précoces de développement, qui constituent les outils essentiels à mon étude et une ressource unique pour la communauté scientifique, sont présentées dans le premier chapitre de résultats de cette thèse.

Par la suite, ces banques, en association avec d'autres banques d'ADNc de tournesol réalisées au sein de notre laboratoire et pour le programme Génoplante, ont été utilisées pour la fabrication d'une puce d'ADNc sur membrane nylon comptant environ 10 000 séquences. L'analyse transcriptomique, comparant les gènes exprimés dans les trois stades de développement embryonnaire d'intérêt ainsi que dans les feuilles et les ovules non fécondés qui constituent respectivement des témoins physiologiques externes de tissus somatiques et reproducteurs, a permis d'identifier des gènes différenciellement exprimés entre les différentes conditions étudiées, d'observer des groupes d'expression et de mettre en évidence les processus moléculaires et physiologiques successivement engagés dans l'établissement du patron morphologique de l'embryon. La construction des microarrays et leur analyse sont décrites dans le second chapitre de résultats.

PARTIE II

RESULTATS ET DISCUSSION

Chapitre I – Approche EST : Création de ressources et analyses d’expression génique *in silico*.

L’approche EST, qui consiste à séquencer à grande échelle des clones ADNc issus d’une condition particulière, permet d’accéder directement aux gènes exprimés dans un tissu particulier, à un moment spécifique ou dans une condition donnée. L’annotation de ces séquences sur la base de leur homologie avec des gènes déjà décrits permet d’obtenir des indices quant à leur fonction putative et quant aux mécanismes moléculaires de régulation et aux principales voies métaboliques mis en jeu dans des conditions physiologiques particulières. De plus, cette stratégie facilite la découverte de nouveaux gènes et procure de précieuses informations concernant le génome d’organismes dont le séquençage complet n’est pas encore envisagé. Cette méthode relativement rapide et applicable à tous les organismes et à toutes les situations est apparue très séduisante et a été adoptée massivement générant, à ce jour, plus de 27 500 000 séquences EST obtenues à partir de plusieurs centaines d’organismes différents et répertoriées dans les banques de données publiques.

Chez les végétaux, cette approche génomique a été utilisée chez diverses plantes (Channelière et al., 2002 ; Crookshanks et al., 2001 ; Fernandes et al., 2002 ; Hofte et al., 1993; Journet et al., 2002) pour l’étude de différents processus biologiques tels que le développement, la réponse à différents stress ou la symbiose (Gyorgyey et al., 2000 ; Hu et al., 2003 ; Lim et al., 1996 ; Zhang et al., 2001).

Afin de caractériser les gènes induits au cours des phases successives de l’embryogenèse précoce du tournesol, nous avons construit des banques d’ADNc de référence à partir d’embryons à des stades clés du développement (globulaire, cœur et cotylédonaire) ainsi qu’une banque de type SSH enrichie en gènes exprimés au stade cœur après soustraction avec les transcrits d’embryons cotylédonaires. Après séquençage d’un nombre représentatif de clones pour chacune de ces banques et analyse bioinformatique des EST générées, une première ébauche des profils fonctionnels caractéristiques de chacun des stades de développement a été obtenue.

I. Choix méthodologiques.

I.1. Choix du matériel végétal: une stratégie « stades de développement » préférée à « jours après pollinisation ».

Chez le tournesol, les embryons ont un développement décalé dans le temps en fonction de leur position sur le capitule. Une récolte massive à une date précise après la floraison entraînerait la présence dans l’échantillon d’embryons à différents stades de développement ce qui empêcherait l’identification précise

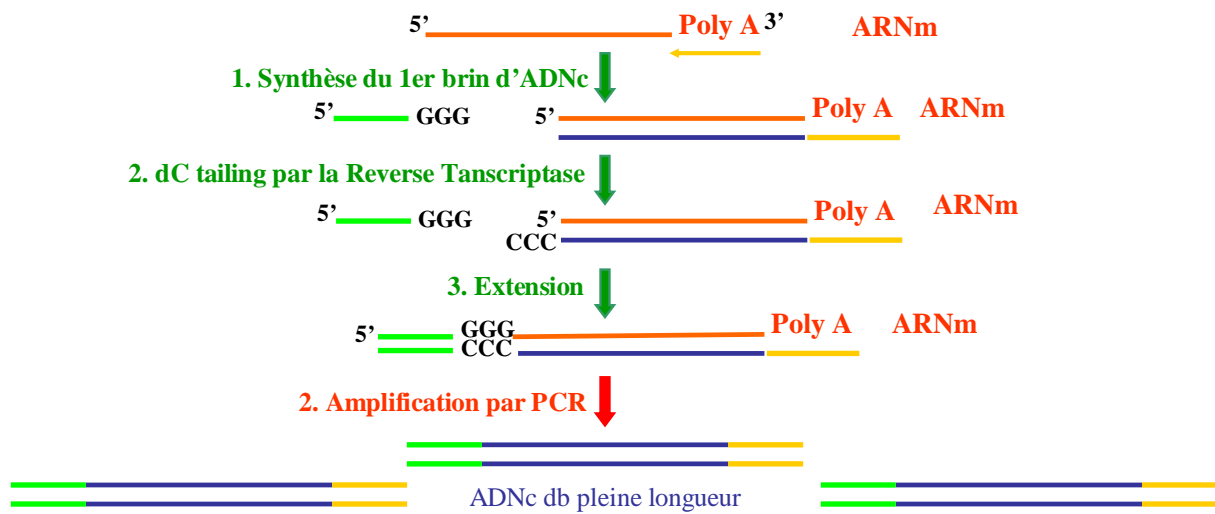


Figure 23. Principe de l'amplification des ADNc par la technologie SMART

La technologie SMART est une méthode d'amplification des ADNc par PCR permettant d'obtenir des ADNc pleine longueur à partir de faibles quantités d'ARNm.

1. Les ARNm sont rétrotranscrits à partir d'un oligo(dT) modifié contenant une séquence qui permettra ultérieurement l'amplification par PCR à partir d'une amorce spécifique de cette séquence.

2. Lorsque la réaction atteint l'extrémité 5', l'enzyme ajoute des cytosines du côté 3' de l'ADNc grâce à son activité terminal transférase.

3. Cette fraction de cytosine s'apparie avec la séquence oligo(G) d'un oligonucléotide créant ainsi une matrice étendue. L'enzyme change alors de brin et poursuit la réplication jusqu'à l'extrémité de l'oligonucléotide générant au final un ADNc pleine longueur contenant à chaque extrémité des séquences pouvant servir d'amorces pour une amplification par PCR.

4. Le second brin d'ADNc est synthétisé par PCR. L'amplification permet d'augmenter la quantité d'ADNc obtenue. De plus, si la réaction est arrêtée en phase exponentielle, les proportions relatives des différents ADNc de l'échantillon sont conservées.

des fonctions cellulaires successivement engagées au cours de l'embryogenèse précoce. Or, chaque stade correspond à un processus particulier de la formation de la graine, le stade globulaire coïncidant avec l'établissement des axes embryonnaires, le stade cœur à la mise en place de la symétrie bilatérale et le stade cotylédonaire à l'accumulation de réserves. Pour le tournesol, la récolte en masse des embryons en fonction du nombre de jours après la pollinisation n'est donc valable que pour des stades tardifs du développement. Les banques d'ADNc réalisées au cours de ce projet de thèse s'appliquant aux phases précoces de l'embryogenèse ont été construites à partir d'embryons à des stades précis du développement (globulaire, cœur et cotylédonaire). Préalablement à l'extraction d'ARNm, les embryons ont été microdisséqués sous une loupe binoculaire et triés en fonction de leur taille et de leur morphologie.

I.2. La technologie SMART est utilisée pour pallier aux faibles quantités de matériel végétal.

L'application des méthodes de génomique fonctionnelle pour l'analyse de l'embryogenèse précoce a longtemps été limitée du fait des difficultés rencontrées pour extraire le matériel végétal en quantités suffisantes. De nos jours, le développement de différentes techniques permettant l'amplification des gènes transcrits en amplifiant les ADNc par PCR a rendu plus accessibles les approches de transcriptomique pour de telles études réalisées à partir d'une très faible quantité de cellules ou de tissus difficiles à prélever. Ainsi, des banques d'ADNc ont pu être construites à partir de seulement 10 oocytes de souris (Revel et al., 1995) ou de 10 cellules de myélome (Belyavsky et al., 1989). Toutefois, si ces méthodes d'amplification permettent le clonage des gènes exprimés, les proportions initiales des différents ARNm présents dans le tissu ne sont pas toujours respectées. En effet, un biais est introduit du fait des différences de taille des transcrits, les ADNc les plus courts étant favorisés au cours de la réaction de PCR. Pour contrecarrer cet artefact, plusieurs solutions ont été proposées comme, par exemple, un fractionnement des ADNc selon leur taille suivi d'une amplification différentielle (Wellenreuther et al., 2004) ou tout simplement le rallongement de la durée de la phase d'extension lors de la PCR associé à l'arrêt de la réaction en phase exponentielle comme dans le cas de la technologie SMART. Ainsi, la SMART, dont le principe est décrit dans la figure 23 ci-contre, est une méthode d'amplification des ADNc par PCR favorisant l'obtention des ADNc pleine longueur à partir d'une faible quantité d'ARN de départ (Zhu et al., 2001). De plus, dans ce cas, le dosage des différents transcrits présents initialement dans le tissu n'étant pas modifié, la comparaison des niveaux d'expression des gènes exprimés est donc possible. Cette technique, qui a déjà permis de construire des banques d'ADNc à partir de pétales de rose (Channeliere et al., 2002) ou de protoplastes de tournesol cultivés en conditions embryogènes (Tamborindéguy et al, 2004a), a été utilisée, durant la thèse, pour la réalisation des trois banques d'ADNc de référence et de la banque SSH mais également pour l'obtention des sondes complexes utilisées pour l'hybridation des microarrays.

I.3. Les banques d'ADNc de référence : un échantillonnage de l'ensemble des gènes exprimés.

Une banque d'ADNc standard a été produite pour chacun des stades de développement embryonnaire étudiés (HaGlbR, HaHeaR et HaCotR pour les stades globulaire, cœur, cotylédonaire, respectivement) générant un total de 17 664 clones. Environ 1 000 clones de chacune de ces banques ont été séquencés afin d'avoir accès à un éventail assez large des gènes exprimés à chaque stade de développement et au final, 1 478 EST ont été publiées (numéros d'accèsion : AJ827751-AJ829228).

Théoriquement, ce type de banques d'ADNc permet de cloner l'ensemble des gènes exprimés dans une condition donnée et fournit, grâce au séquençage aléatoire d'un échantillon représentatif de clones, une vue globale de l'ensemble des fonctions biologiques impliquées et de leur importance relative. Les banques d'ADNc de référence permettent donc de décrire globalement l'activité cellulaire associée aux tissus testés et d'accéder aux principales voies de régulation induites.

Toutefois, ces banques ne sont généralement pas normalisées : la proportion de clones représentant un transcrite particulier correspond à son taux réel d'expression dans la cellule. Les gènes les plus exprimés sont donc les plus séquencés entraînant un fort taux de redondance dans les données générées et une difficulté d'accès aux transcrits les plus rares tels que les facteurs de transcription ou les signaux de transduction qui jouent pourtant un rôle essentiel dans la régulation du programme génétique induit dans une condition particulière. Plusieurs solutions ont été envisagées pour remédier à ce problème, la première, triviale mais peu économique, étant un séquençage à grande échelle. Afin de diminuer la redondance des clones les plus fréquents, une normalisation 'artificielle' des banques peut être effectuée par hybridation des clones bactériens déposés sur filtre avec une sonde marquée spécifique des séquences les plus représentées. Cette technique, si elle s'applique uniquement à un nombre très réduit de gènes, ne modifie pas le dosage initial des autres transcrits et n'empêche donc pas la comparaison des fréquences de chaque gène afin de déduire leurs niveaux d'expression relatifs (Ohlrogge and Benning, 2000). Une telle approche a été menée au cours d'analyse d'EST issues de graines d'*Arabidopsis thaliana* dans le but d'éliminer les inserts présentant un taux de redondance très élevé tels que ceux codant pour des protéines de réserves (White et al., 2000). Nous-mêmes y avons eu recours pour les banques construites à partir des embryons aux stades globulaire et cœur (HaGlbR et HaHeaR) pour lesquelles deux séquences montrant des similarités avec une protéine prédite et un transporteur membranaire étaient très abondantes.

Une autre possibilité pour favoriser la découverte de nouveaux gènes est d'augmenter la diversité des conditions étudiées (tissus, stades de développement, conditions de cultures, stress biotiques ou abiotiques...) en construisant, par exemple, des banques d'ADNc à partir d'un mélange d'ARNm. Cette stratégie a été adoptée chez diverses espèces et a notamment permis de générer environ 44 000 séquences EST chez le tournesol. Cependant, cette méthode ne permet pas l'analyse comparative du niveau d'expression des gènes entre les différents tissus. Afin de s'affranchir de cette contrainte, un système

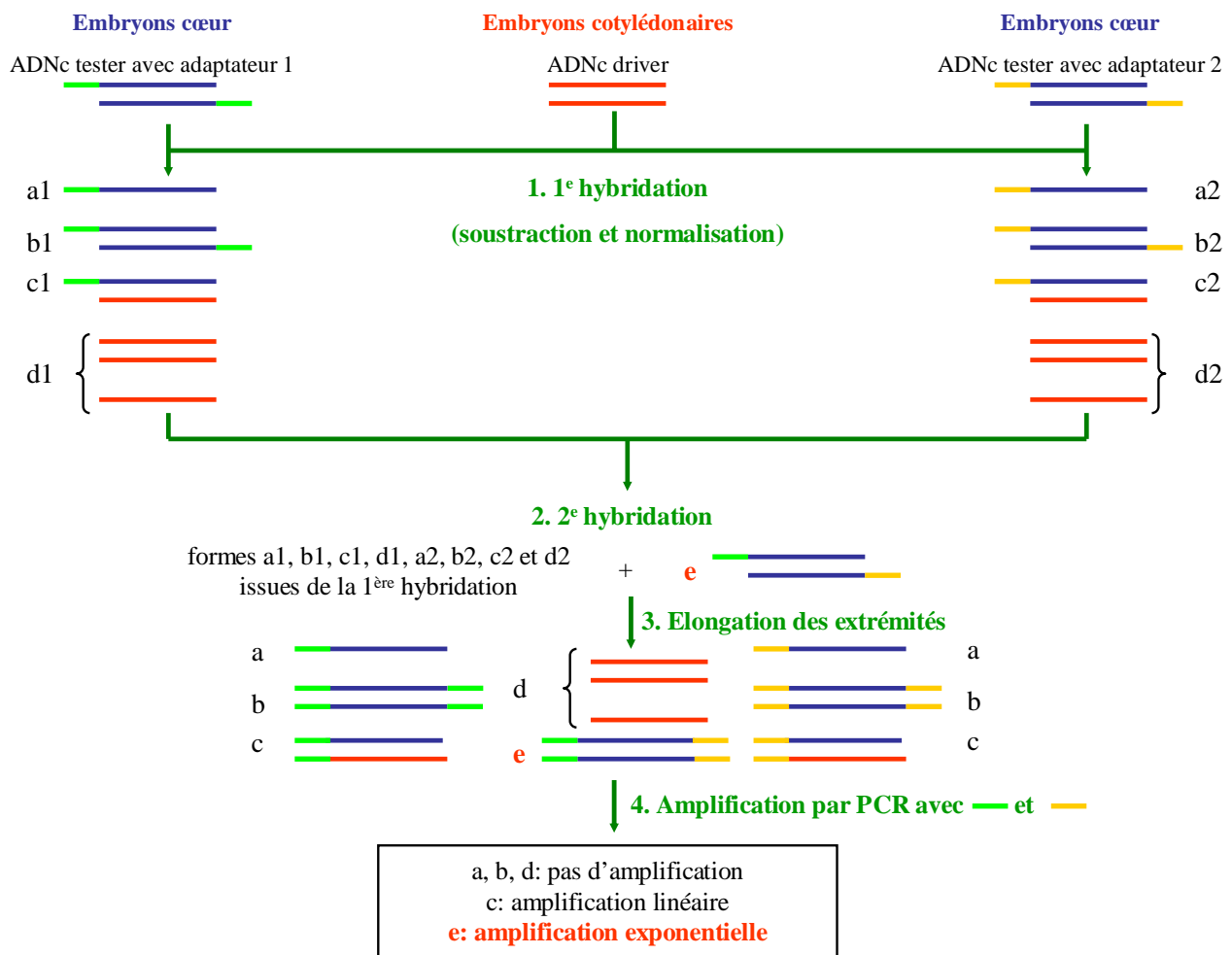


Figure 24. Principe de la technique SSH.

Après extraction et rétro-transcription des ARNm issus des 2 conditions à comparer, l'ADNc d'intérêt (tester) est séparé en 2 fractions égales auxquelles sont ajoutés deux adaptateurs différents. Deux hybridations sont alors réalisées.

1. Pour la 1^{ère} hybridation, les ADNc à soustraire (driver) sont ajoutés en excès à chaque fraction d'ADNc tester; après dénaturation et renaturation, plusieurs possibilités se présentent: les brins complémentaires des gènes présents dans les deux conditions (tester et driver) peuvent s'hybrider ensemble (formes c1 et c2), reformer les homoduplex initiaux (formes b1, b2, d1 et d2) ou rester en simplex (formes a1, a2, d1 et d2). Selon l'origine du brin d'ADNc, chacune des formes possède ou non des adaptateurs aux extrémités. Les simplex de type 'a' sont particulièrement intéressants: ils sont enrichis en séquences différenciellement exprimées (les gènes communs aux deux conditions créant des hétéroduplex de type 'c') et ils sont présents à la même concentration même pour des gènes plus ou moins exprimés.

2. Au cours de la 2^{ème} hybridation, les 2 fractions de tester sont mélangées sans dénaturation préalable. Un nouveau forme d'homoduplex (forme e), portant un adaptateur différent à chaque bout, peut se former.

3. et 4. Après élongation des extrémités de chaque complexe, une PCR est effectuée à l'aide d'amorces spécifiques des adaptateurs. A l'issue de cette PCR, seuls les hétéroduplex 'e', correspondant aux séquences préférentiellement exprimées dans la conditions d'intérêt sont amplifiés exponentiellement. La forme 'b', représentant les gènes fortement exprimés, ne peut pas être amplifiée du fait de la présence de régions répétées dans les adaptateurs entraînant la formation de structures de type pan.

d'identification du tissu d'origine des séquences s'appuyant sur l'ajout de 6 nucléotides dans l'oligo dT utilisé pour la synthèse du premier brin d'ADNc a récemment été élaboré (Qiu et al., 2003).

I.4. Les banques SSH : enrichissement en gènes faiblement et préférentiellement exprimés pour un stade de développement.

Les banques SSH (Suppressive Subtractive Hybridization) sont des banques soustraites et normalisées. Le principe des banques SSH s'appuyant sur deux hybridations successives est schématisé dans la figure 24 ci-contre. Celles-ci sont construites à partir de deux tissus différents dont les populations en ARNm sont soustraites, les ADNc clonés représentant les gènes préférentiellement exprimés dans l'une des deux conditions. Ce type de banques permet donc d'identifier les voies métaboliques et circuits de régulation spécifiquement impliqués dans le tissu d'intérêt en éliminant les fonctions cellulaires de base communes aux deux conditions testées. De plus, théoriquement, du fait de la normalisation, une seule copie de chaque transcrite, qu'il soit fortement exprimé ou pas, est clonée. Ainsi, contrairement aux banques de référence, l'accès aux ARNm les plus rares est facilité mais la comparaison entre la fréquence des EST dans différents tissus n'est plus possible.

Plusieurs études visant à étudier l'expression génique différentielle dans des mutants ou des plantes sauvages (Han et al., 2004) ou à caractériser les gènes spécifiquement induits lors de la floraison (Kim et al., 1999), de la sénescence (Hinderhofer and Zentgraf, 2001), du remplissage de graines (Jang et al., 2003) ou encore en réponse à des stress biotiques (Verica et al., 2004; Xiong et al., 2001) ou abiotiques (Watt, 2003) ont employé cette technique et ont confirmé la valeur des ressources ainsi créées pour la découverte de gènes, avec, dans tous les cas, entre 20 et 50% des EST ne présentant pas de similarités avec des séquences connues.

Lors de la caractérisation du transcriptome dans une condition particulière, la création d'une banque de type SSH associée à la construction d'une banque de référence améliore la probabilité d'identifier l'ensemble des gènes exprimés tout en autorisant l'analyse du profil d'expression du tissu.

Au cours de ce travail de thèse, une banque de type SSH comparant les transcrits d'embryons au stade cœur et cotylédonaire et enrichie en gènes préférentiellement exprimés au stade cœur a été produite afin d'isoler les gènes spécifiquement impliqués dans l'établissement du patron morphologique de l'embryon et dans l'acquisition de la symétrie bilatérale par rapport aux gènes de croissance ou de maturation embryonnaire et d'identifier ainsi plus précisément les processus engagés au cours des étapes précoces de morphogénèse embryonnaire.

II. Approche fonctionnelle de l'embryogenèse précoce du tournesol par analyse bioinformatique d'EST.

La construction et l'analyse des banques embryogènes de référence et SSH, qui ont permis la création de ressources et d'outils génomiques essentiels à notre étude, constituent le premier pas vers la caractérisation du transcriptome induit lors de l'embryogenèse précoce du tournesol.

L'analyse des banques de référence construites au cours de cette thèse associées aux données d'autres banques reproductives réalisées pour le programme Génoplante a fait l'objet d'une publication.

Les résultats des analyses de la banque SSH sont présentés dans une seconde partie.

II.1. Article: Comparative analysis of early embryonic sunflower cDNA libraries.

Afin de caractériser les fonctions cellulaires successivement engagées au cours du développement embryonnaire zygotique du tournesol, des analyses *in silico* ont été effectuées à partir des séquences issues des banques d'ADNc standard embryogènes produites au sein de notre laboratoire (HaGlbR, HaHeaR et HaCotR) ainsi que des données EST issues de banques d'ADNc obtenues à partir d'ovules non fécondés (HadevR7) ou d'embryons cotylédonaire à des stades plus tardifs (HaDevR5 et HaDevR6) réalisées dans le cadre du programme Génoplante. L'ensemble des séquences analysées représente un total de 7 106 EST et un unigène (nombre de gènes potentiellement différents) de 3064 séquences (Ben et al., 2005).

Résumé des principaux résultats obtenus

Une caractérisation fonctionnelle des séquences uniques de chacune des banques d'ADNc a été effectuée par recherche d'homologies avec des protéines de fonctions connues ou prédites répertoriées dans les banques de données publiques (programme BLASTX) et a permis d'avoir une vue globale des fonctions cellulaires impliquées à chaque stade de développement et de leur importance relative. Les profils fonctionnels sont apparus très similaires pour chacun des stades de développement étudiés, les catégories fonctionnelles les plus représentées correspondant systématiquement aux gènes codant pour des protéines potentiellement impliquées dans le métabolisme protéique et le métabolisme primaire suivis des gènes intervenant dans les processus transcriptionnels ou dans les mécanismes de réparation et de dégradation de l'ADN. Ce profil fonctionnel montre que, tout au long de leur développement, les embryons zygotiques présentent une activité cellulaire très élevée et que, bien plus que de simples amas de cellules en division, ils constituent des structures largement engagées dans des processus de différenciation. Cependant, pour un grand nombre des séquences analysées (49.7% de l'unigène global), aucune fonction potentielle n'a pu être attribuée. Ces EST, qui peuvent représenter des gènes jouant un rôle clé au cours de l'embryogenèse mais dont la fonction n'a pas encore été élucidée ou encore des séquences spécifiques du tournesol, constituent une ressource d'un très grand intérêt.

Bien que les patrons fonctionnels soient très proches pour tous les stades de développement étudiés, des séquences potentiellement spécifiques à chacun de ces stades ont tout de même pu être mises en évidence et des analyses de northern virtuels basées sur la statistique *R* décrite par Stekel (2000) ont permis de révéler des gènes différentiellement exprimés entre les différentes conditions. Enfin, des séquences montrant de fortes similarités avec des gènes jouant des rôles clés dans l'embryogenèse animale ou végétale tels que les gènes MAGO NASHI, ARGONAUTE, PINHEAD, SCARECROW ou LEC1, ont également été identifiées.

Comparative analysis of early embryonic sunflower cDNA libraries

Cécile Ben¹, Tarek Hewezi¹, Marie Françoise Jardinaud¹, Frédérique Bena²,
Nathalie Ladouce¹, Sébastien Moretti¹, Cecilia Tamborindeguy¹, Thierry Liboz¹,
Michel Petitprez¹, and Laurent Gentzbittel^{1,*}

¹*Institut National Polytechnique de Toulouse - Ecole Nationale Supérieure de Toulouse, Laboratoire de Agronomie Biotechnologie et Amélioration des Plantes Pôle de Biotechnologie Végétale, IFR40, 18 Chemin de Borde Rouge, Auzeville, 31326 Castanet Tolosan, France. (*author for correspondence: e-mail gentz@ensat.fr);*

²*Institut de Biologie Moléculaire des Plantes, 12, rue du général Zimmer, 67084 Strasbourg cedex, France*

Received 21 October 2004; accepted in revised form 12 December 2004

Key words: cDNA libraries, embryogenesis, expressed sequence tags (ESTs), *Helianthus annuus* L., seed development

Abstract

To gain information concerning cell functions and activities during sunflower embryogenesis, an expressed sequence tag (EST) approach was used to analyse gene expression in the early stages of sunflower embryos development. Confocal microscopy observations of whole-mounted embryos allowed us to identify precisely the major steps of the zygotic embryonic development. A time-course analysis was then employed to collect the embryonic material. Three cDNA libraries were constructed from microdissected embryos, and three other cDNA libraries were created using a classical day after pollination schedule. A total of 7106 ESTs were produced and assembled. The total number of putative different genes represents about 43.1% (3064 tentative contigs and singlets) of the analysed sequences. The unigenes that showed similarity to proteins with known or predicted functions (50.3%) were classified into 15 different functional categories. The functional profiles were found to be quite similar for all studied embryo stages but statistical analysis revealed that successive and coordinate sets of genes are expressed at each embryonic stage. The analysis allowed us to identify abundant and differentially expressed genes at the early stages of embryos development as well as some putatively interesting genes, showing strong similarities with genes playing key roles in plant and animal embryogenesis. The data presented in this study not only provide a first global overview of the genes expression profile during sunflower embryogenesis but also represent an original and valuable tool for developmental genomics studies on exalbuminous dicots.

Introduction

Grain quality and yields, which are the main objectives for crop plant breeders, depend largely on embryo structure and development. Indeed, in higher plants, embryogenesis is a key step of development during which the embryo acquires a correct morphological pattern and accumulates storage reserves. Dicotyledonous embryogenesis is remarkably conserved among exalbuminous plants (plants without endosperm in seed).

Arabidopsis thaliana embryos development, which involved a relatively invariant sequence of cellular divisions, still remains the best characterized (West and Harada, 1993; Goldberg *et al.*, 1994; Jurgens, 1995, 2001). In *Arabidopsis*, embryogenesis begins with a double fertilization during which the two sperm cells of the pollen grain merge with the egg cell and the central cell to produce the zygote and the endosperm, respectively. The zygote then divides asymmetrically giving rise to two cells of different size and

fate: the small apical cell forms the major part of the embryo proper whereas the big basal cell develop into the suspensor, a transient embryonic structure which possibly functions as an ‘umbilical cord’ conveying nutrients, hormones and growth factors necessary for a normal embryo development. The suspensor senesces during early embryogenesis (heart-shaped stage), except for the uppermost cell which belongs to the embryonic root. Subsequently, the embryo develops according to two different axes, the apical-basal and the radial axes, to establish different organs and tissues. Although, many morphological and histological processes of Angiosperms (flowering plants) embryogenesis are now well known, molecular and physiological events leading to the seed formation are still far to be completely understood. Indeed, the study of early stages of embryogenesis often remains difficult as the embryo, constituted of few very small cells, is embedded in several layers of maternal tissues. To circumvent such difficulties, researches are generally undergone on model plants such as *Fucus* algae (Belanger and Quatrano, 2000), *A. thaliana* mutants (Jurgens *et al.*, 1997; Souter and Lindsey, 2000; Chaudhury *et al.*, 2001; Jurgens, 2001), or somatic embryos (Zimmerman, 1993). However, the analysis of these different systems still gives rise to incomplete and insufficient information.

Large-scale single-pass sequencing of cDNA prepared from tissues with specific metabolism activity has provided an efficient and rapid mean towards isolation of new genes (van de Loo *et al.*, 1995; Cahoon *et al.*, 1999). In addition to providing an efficient tool for gene cloning, EST data set can also provide information on gene expression in different plant tissues by comparing EST frequencies in different cDNA libraries (White *et al.*, 2000; Fernandes *et al.*, 2002; Ogiwara *et al.*, 2003; Ronning *et al.*, 2003; Suh *et al.*, 2003).

The advantage of *in silico* analysis of EST collection is that a high number of gene sequences can be analysed without limitation and the analysis is relatively rapid (Fedorova *et al.*, 2002). Also, developmentally regulated genes can be identified by *in silico* analysis of EST data set derived from the same tissue at different time points.

Recently, sunflower (*Helianthus annuus* L.), which presents an inflorescence harbouring

numerous and relatively big embryos whose development is synchronized, has been proposed as a putative complementary model for the study of embryogenesis in exalbuminous dicotyledonous (Tamborindeguy *et al.*, 2004). Moreover, for this oleaginous culture with important agronomic outlets (food agro-industry, lipo-chemistry, animal feeding), the study of embryogenesis is of a particular interest, because oils and proteins storage reserves are directly accumulated into the embryo proper. In the past few years, several tools have been developed, making easier the study of this poorly investigated plant of agricultural interest and allowing a good knowledge of its large genome: genetic linkage maps now cover almost the entire sunflower genome (Tang *et al.*, 2002; Al Chaarani *et al.*, 2004), two BAC libraries are available (Caboche and Boucly, 2000; Gentzmittel *et al.*, 2002).

In sunflower, large scale ESTs sequencing programs have also been achieved, giving rise to 63 112 entries in Genbank. This significant EST data set was generated from cDNA libraries prepared from a wide range of biologically and agronomically relevant tissues. Among this EST collection, the largest group of sequences (approximately 44 000) originated from sequencing of mixed population of cDNAs from 11 different tissues. However, construction and analysis of cDNA libraries of developing sunflower embryos remain lacking. The study presented here represents a medium-scale ESTs sequencing project aiming to gain information on cell functions and activities during the early stages of sunflower embryos development. In order to analyse the transcriptional program induced during the embryo patterning, reference standard cDNA libraries from sunflower embryos at different stages of development were produced. Several thousands of ESTs from each library have been sequenced and annotated. After evaluating the value and the reliability of this original ESTs data set, we analysed the genes expression patterns during the different developmental stages to estimate the similarity and the evolution of the induced genes. The functional characterization of the cDNA sequences allowed us to present a global picture of the major cellular functions involved in each stage of embryo development as well as the relative importance of these functions during embryos development.

Results

Sunflower embryo development is very representative of dicotyledonous exalbuminous embryogenesis

Embryogenesis is known to be remarkably conserved among exalbuminous dicots. Confocal microscopy observations of sunflower embryos at different stages of development confirmed that only minor differences exist between the development of sunflower and *A. thaliana* embryos. As illustrated in Figure 1A, embryogenesis begins in sunflower with the double fertilization event

during which pollen nuclei fertilize the oosphere as well as the central cell to form the zygote and the endosperm, respectively. Then, the zygote undergoes an invariant sequence of cell divisions leading to successive well-characterized stages of embryo development. The zygote first divides asymmetrically to give rise to a big basal cell which will form the suspensor and a small apical cell. The last one develops into the embryo proper (Figure 1B). After two longitudinal and one transversal divisions of the apical cell, the embryo reaches the octant stage (Figure 1C). At this stage, the embryo is separated into two layers of four cells. The uppermost layer gives rise to the most apical

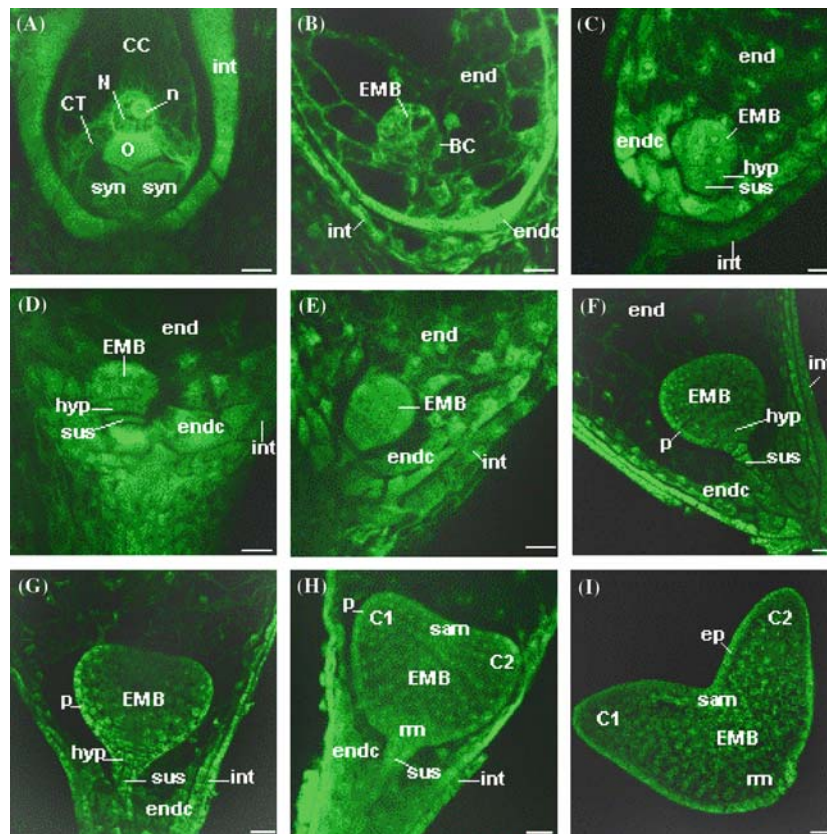


Figure 1. Sunflower embryo development. Confocal observations in ovules at different stages of development. During the double fertilization, pollen nuclei fertilize the oosphere as well as the central cell of the ovule (A) to form the zygote and the endosperm, respectively. Then, the zygote undergoes a first asymmetrical division giving rise to a big basal cell and a small apical cell which will develop in the embryo proper (embryo at 4-cell stage) (B). At the octant stage (C), the embryo is separated into two cells layers which contribute to different embryo structures. Radial pattern formation begins at the globular stage of the embryo (D–F) with the apparition of the protoderm. Finally, the embryo acquires a bilateral symmetry and successively goes through globular-heart stages transition (G), heart-shaped (H) and cotyledonary stages (I). BC: basal cell, C1 and C2: cotyledons, CC: central cell, CT: central cell cytoplasmic threads, EMB: embryo proper, end: endosperm, endc: external endosperm cell layer, ep: epiderm, hyp: hypophyse, int: integument, N: central cell nucleus, n: central cell nucleoli, O: oosphere, p: protoderm, mm: root meristem, sam: shoot apical meristem, sus: suspensor, syn: synergide. Bars: 15 μm (A, B, C), 20 μm (D, E), 25 μm (F, G, H), 50 μm (I).

structures of the embryo (the shoot apical meristem and a large part of the cotyledons). The lower layer forms the basal region of the cotyledons as well as the hypocotyl, the embryonic root and part of the root meristem. At the globular stage of the pro-embryo (Figure 1D–F), radial pattern formation begins and coincides with the apparition of periclinal divisions giving rise to a peripheral cellular layer, the protoderm, which is the epiderm precursor. Later, asymmetric cell divisions will produce the ground meristem (cortex and endoderm precursors) and the procambium (vascular tissue precursors) (not shown). Finally, cellular differentiation and elongation phases will successively lead to embryo heart-shaped stage (Figure 1G and H), corresponding to the establishment of the bilateral symmetry and to cotyledonary stage (Figure 1I). This microscopic observation has let us to identify precisely three different stages of embryos development that were used to construct the libraries named HaGlbR for the globular stage, HaHeaR for the heart stage and HaCotR for the cotyledonary stage, respectively.

Complexity and redundancy vary greatly among embryonic cDNA libraries

The cDNA libraries sampled in this study represent major stages in the development and physiology of sunflower embryos. For the earliest embryos stages of development, embryos were microdissected and screened depending on their size and morphology. According to these criteria, embryos were classified into three different categories and cDNA libraries were constructed accordingly: the globular and transition globular-heart shaped embryos (HaGlbR library), the heart

shaped embryos (HaHeaR library) and the early cotyledonary embryos (HaCotR library). For later stages, two libraries, named HaDevR5 and HaDevR6 were prepared from embryos collected from capitulae at 4 and 7 dap, corresponding to mid- and late cotyledonary embryos, respectively. The HaDevR7 library was constructed from unfertilized ovules.

Each library contains between 2304 arranged clones for HaDevR5 to 6528 clones for HaHeaR (Table 1). A total of 12 158 single-pass sequencing reactions were performed from the 5' end. After quality assessment, using PHRED for trimming of low-quality sequences and removing sequences of less than 100 bp excluding from poly-A tail, a total of 7106 good quality sequences was generated and further used for computational analysis. The good quality sequences rates vary from 16.6% for HaGlbR to 87.8% for HaDevR5. The average length of these ESTs was 394 bases with a range of 194 bases for HaGlbR to 531 bases for HaDevR6. The rate of contaminations by chloroplastic and mitochondrial sequences was determined using the BLASTN program on *Arabidopsis* sequences and was found to be very low to unexisting.

In order to generate the unigene (set of unique sequences) of each library, all ESTs were assembled by PHRAP program, thereby generating consensus sequences. The percentage of ESTs that was clustered into contigs ranges from 30.5% for HadevR7 to 77.2% for HaDevR6 (Table 2). This data reflects the redundancy rate within each library. In this manner, HaDevR6 appears to be a redundant library. The HaGlbR and HaHeaR libraries are enriched in the same two sequences that show homologies with an integral membrane transporter protein and a hypothetical 14.3 kDa

Table 1. cDNA libraries description and EST collection statistics.

Libraries	Developmental stage	No of clones	No of sequenced ESTs	No of analysed ESTs	ESTs average length (bp)
HaDevR7	Unfertilized ovary	2496	2496	827	413
HaGlbR	Globular/Globular-heart transition embryos	6144	1130	187	194
HaHeaR	Heart-shaped embryos	6528	1332	320	313
HaCotR	Early cotyledonary embryos	4992	1152	971	416
HaDevR5	Mid cotyledonary embryos (4 dap)	2304	2304	2023	493
HaDevR6	Late cotyledonary embryos (7 dap)	3744	3744	2778	531
Total or average	–	26208	12158	7106	394

Table 2. Evaluation of the cDNA libraries complexity and redundancy.

Developmental stage	HaDevR7 Unfertilized ovules	HaGlbR Globular embryos	HaHeaR Heart-shaped embryos	HaCotR Early cotyledonary embryos	HaDevR5 Mid cotyledonary embryos (4 dap)	HaDevR6 Late cotyledonary embryos (7 dap)
No of analyzed ESTs	827	187	320	971	2023	2778
No of assembled ESTs (% of the library ESTs) ^a	252 (30.5)	137 (73.3)	174 (54.4)	414 (42.6)	1194 (59)	2146 (77.2)
No of contigs	95	22	43	148	358	514
Average contig length (bp)	542	277	470	538	655	697
Maximum No of ESTs per contig	12	60	25	15	20	47
No of unigene ^b (% of the library ESTs) ^c	670 (81)	72 (38.5)	189 (59.1)	705 (72.6)	1187 (58.6)	1146 (41.2)

^aPercentage calculated by dividing the number of assembled ESTs of each library by the total number of analysed ESTs of each library.

^bThe number of unigene corresponds to the sum of the contigs and singletons for each library.

^cPercentage calculated by dividing the number of unigene for each library by the total number of analysed ESTs of each library.

protein (AJ827951 and AJ828208). In contrast, HaDevR7 library presents a very low redundancy level and so seems to be a very informative and useful tool for gene discovery.

Cluster size varies significantly from moderately abundant contigs (12–15 maximum ESTs per contig) to large clusters containing up to 60 ESTs for HaGlbR. Generated contigs show an average length varying from 277 bp for HaGlbR to 697 bp for HaDevR6.

The complexity of the libraries is evaluated by the size of the unigene and its proportion relative to the number of initial ESTs. The HaGlbR is clearly the less informative library, with a unigene corresponding to 38.5% of the total number of ESTs in this library. The HaCotR and HaDevR7 libraries are among the most complex libraries with a unigene size ranging from 72.6% to 81% of the ESTs.

Each library contributes significantly to the global unigene

The 7106 ESTs derived from all studied libraries were grouped together and assembled using PHRAP program. Three thousand sixty four unique sequences representing 43.1% of the analysed sequences were identified (Table 3). The contribution of each library to this global unigene was evaluated. HaDevR5 sequences contribute to about 22.9% of the total number of unigene contrary to HaGlbR sequences that represent only

1.4%. This great gap between the contributions of the different libraries in the global unigene is due to the high differences in the number of ESTs analysed for each library. In fact, if we consider the proportion of library-specific contigs expressed relatively to the number of ESTs of that library, the HaGlbR library contributes almost equally to the HaDevR6 library. The HaCotR and HaDevR7 are found to be the most contributing libraries, a result in accordance with their high complexity.

Interestingly, this analysis allows us to identify 2367 clusters (contigs and singletons) specific to a single library. These results indicate clearly that the information derived from the different libraries is not redundant but rather than complementary. They also point out that embryos are rapidly evolving structures, stressing the need for a more detailed time-course analysis by increasing the number of different libraries.

Functional characterization of expressed genes in sunflower embryos: the zygotic embryos are highly differentiating structures

In order to attribute a putative function to the clustered ESTs, protein homologies were searched in public databases (SwissProt and TrEMBL) using BLASTX program and *E*-values below 10^{-5} . Thereby, a putative function was assigned to 50.3% of contigs and singletons. According to the BLASTX results, clusters were distributed into 15

Table 3. Evaluation of the contribution of each library to global unigene.

	HaDevR7	HaGlbR	HaHeaR	HaCotR	HaDevR5	HaDevR6	Total
No of analysed ESTs	827	187	320	971	2023	2778	7106
No of contigs specific to single library ^a	38	9	23	59	93	214	436
No of singlets	394	33	75	366	609	454	1931
(% of the library ESTs) ^b	(47.6)	(17.6)	(23.4)	(37.7)	(30.1)	(16.3)	—
No of unigene specific to single library ^c	432	42	98	425	702	668	2367
(% of the library ESTs) ^d	(52.2)	(22.5)	(30.6)	(43.8)	(34.7)	(24)	—
(Contribution to global unigene (%)) ^e	(14.1)	(1.4)	(3.2)	(13.9)	(22.9)	(21.8)	—
No of unspecific contigs ^f							697
Global unigene ^g							3064
(% of total ESTs) ^h							(43.1)

^aContigs specific to a library are contigs only containing ESTs from a single library.

^bPercentage calculated by dividing the number of singlets by the total number of analysed ESTs in each library.

^cThe number of unigene specific to a single library corresponds to the sum of the contigs and singletons specific to this library.

^dPercentage calculated by dividing the unigene specific to a single library by the total number of analysed ESTs of this library.

^ePercentage calculated by dividing the unigene specific to a single library by the global unigene.

^fUnspecific contigs are contigs containing ESTs from different libraries.

^gThe global number of unigene corresponds to the sum of the unigenes specific to each library and unspecific contigs.

^hPercentage calculated by dividing the global number of unigene by the total number of analysed ESTs.

groups based on the function classification proposed by KEGG catalog : <http://www.genome.jp/kegg/> (Table 4). The functional classification was performed for each library to try to determinate precisely and specifically the major functions implicated at each stage of development. In this

case, the rate of consensus sequences showing no similarities with known proteins varies from 18.1% for HaDevR6 up to 53.3% for HaGlbR (Table 4). The low percentage of HaGlbR unigenes having a putative function could reflect the presence of potentially specific genes expressed in

Table 4. Functional characterization of genes expressed during embryo development in sunflower.

Functions classes	Unigene	HaDevR7	HaGlbR	HaHeaR	HaCotR	HaDevR5	HaDevR6
Lipids and fatty acids metabolism	1.6	1.3	2.7	1.1	1.2	2.2	1.7
Carbohydrates metabolism	3.7	3.4	1.3	3.2	5.9	3.6	2.7
Nucleotide metabolism	0.8	0.6	0.0	0.0	0.7	0.4	1.0
Energy metabolism	2.7	1.9	1.3	4.9	2.7	2.5	3.9
Amino acids metabolism	2.1	1.3	0.0	2.7	4.0	2.1	1.4
Transcription	3.5	3.3	0.0	2.2	4.8	2.1	3.1
Membrane transport	2.3	2.4	1.3	3.2	3.0	1.5	2.4
Signal transduction	3.0	2.3	1.3	3.2	2.9	3.6	2.8
Protein sorting and degradation	4.6	5.0	0.0	3.2	4.2	5.6	5.5
Structure (cytoskeleton, membrane, cell wall...)	1.8	1.1	2.7	1.1	2.9	1.8	1.5
Cell division, growth and differentiation	1.2	1.1	0.0	1.1	1.4	1.6	1.1
DNA structure, repair and degradation	3.8	6.7	1.3	4.3	3.8	6.5	4.9
Stress	2.0	2.4	1.3	1.6	2.2	2.2	1.6
Other	6.1	5.0	1.3	5.4	5.4	5.9	7.1
Translation	11.1	11.0	22.7	17.3	14.3	19.2	21.5
Unknown function	19.1	15.2	9.3	10.3	18.8	17.3	19.7
No hit	30.6	35.7	53.3	35.1	21.9	21.9	18.1

Clusters of the global unigene and of the unigenes specific to each library have been classified according to their function using the KEGG catalog.

The data are presented as the percentage of unigene for each functional category.

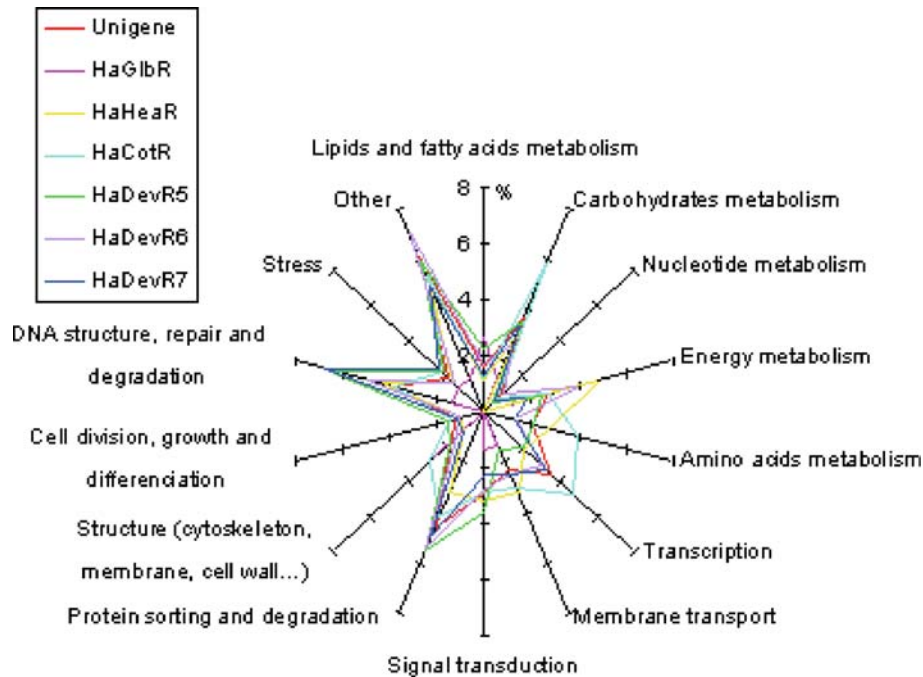


Figure 2. Similarity of the functional profiles between the different cDNA libraries. No hit, unknown function and translation classes were omitted from the radar plot. The coordinates on each axis are the percentage of unigene for each functional category.

this physiological stage, which is very difficult to investigate and probably less extensively sampled in other plant EST programs.

The functional profiles are quite similar for all studied embryo stages (Figure 2). Indeed, in each case, the largest groups, with the exception of the previously reported 'no hits' and 'unknown proteins' categories, consists of genes putatively encoding proteins associated with protein synthesis and processing (from 16.1% of HaDevR7 unigene up to 27% of HadevR6 unigene) and primary metabolism (between 3.6% for HaGlbR and 14.4% for HaCotR). In most cases, genes involved in transcription process and regulation and in DNA repair and degradation mechanisms constitute significant classes too. Remarkably, in all studied libraries, the group representing genes related to stress response is very low (below 2.5%), confirming the good conditions of plant material dissection and libraries construction.

A global view of the results over all libraries indicates that more than 26% of the unigene were found to belong to two major functional categories related to protein synthesis and primary metabolic pathways (Figure 2). Clusters thought to be involved either in transcription or DNA modification

each consist of more than 3% of the total number of the identified unigene. This functional profile reflects the high level of cellular activity that characterizes differentiating tissues. Out of the 3064 unique sequences, 30.6% of clusters do not match with any sequences indexed in public databases and could represent new genes (Table 4). Likewise, 19.1% of the global unigene show similarities with proteins of unknown function. The abundance of clusters with no homology or unknown functions (49.7% of the global unigene) may reflect the presence of some interesting but yet unstudied genes having a key role in sunflower embryo development.

Digital northern identifies putative differentially expressed genes among the embryonic libraries

The *R* statistic described by Stekel *et al.* (2000) was used to identify the contigs showing the most highly differences in EST abundance among the libraries. We limited our analyses to contigs with an *R* statistic >6 (15 in total), which represents moderate to highly expressed genes (from 5 to 88 ESTs per contig) (Table 5). This *R*-value is thought to provide a 91.5% 'true positive' rate. A

Table 5. Differentially expressed clusters in different stages of sunflower embryos development ($R > 6$).

Cluster Identification ^a	HaDevR7	HaGlbR	HaHeaR	HaCotR	HaDevR5	HaDevR6	total ^b	R-value	Annotation	E-value
	Unfertilized ovules	Globular embryos	Heart-shaped embryos	Early cotyledonary embryos	Mid cotyledonary embryos-4 dap	Late cotyledonary embryos -7 dap				
Contig 1140 (AJ828208)	2 (0.24%)	62 (33.16%)	22 (6.88%)	0 (0%)	1 (0.05%)	1 (0.04%)	88	102.14	Hypothetical 14.3 kDa protein	2.00E -07
Contig 1137 (AJ828160)	0 (0%)	23 (12.30%)	24 (7.50%)	0 (0%)	0 (0%)	0 (0%)	47	55.39	Hypothetical protein	2.00E -13
Contig 1118 (AJ827993)	0 (0%)	1 (0.53%)	14 (4.38%)	7 (0.72%)	0 (0%)	0 (0%)	22	19.32	Cell wall-associated hydrolase	5.00E -28
Contig 1033 (AJ828217)	0 (0%)	0 (0%)	9 (2.81%)	0 (0%)	0 (0%)	0 (0%)	9	12.29	No hit	-
Contig 1119 (AJ828349)	1 (0.12%)	3 (1.60%)	10 (3.13%)	5 (0.51%)	0 (0%)	3 (0.11%)	22	11.78	Apocytchrome <i>b</i>	1.00E -46
Contig 1138 (CD852621)	1 (0.12%)	0 (0%)	0 (0%)	0 (0%)	12 (0.59%)	47 (1.69%)	60	10.45	Tuber agglutinin	2.00E -18
Contig 1091 (AJ827761)	0 (0%)	7 (3.74%)	2 (0.63%)	0 (0%)	0 (0%)	5 (0.18%)	14	9.79	Cytochrome P450 like_TBP	2.00E -37
Contig 1110 (CD851159)	1 (0.12%)	0 (0%)	1 (0.31%)	14 (1.44%)	3 (0.15%)	0 (0%)	19	9.55	Elongation factor $I \alpha$	0.0
Contig 1136 (CD852534)	0 (0%)	2 (1.07%)	0 (0%)	0 (0%)	3 (0.15%)	30 (1.08%)	35	8.61	No hit	-
Contig 1106 (CD855158)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	18 (0.65%)	18	6.85	60S ribosomal protein L38-like protein	2.00E -30
Contig 888 (AJ828041)	0 (0%)	0 (0%)	5 (1.56%)	0 (0%)	0 (0%)	0 (0%)	5	6.83	Putative aminopeptidase	2.00E -25
Contig 988 (AJ828207)	1 (0.12%)	2 (1.07%)	4 (1.25%)	0 (0%)	0 (0%)	0 (0%)	7	6.71	Putative senescence- associated protein	4.00E -91
Contig 1104 (CD853176)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	17 (0.61%)	17	6.47	60S ribosomal protein L10a-1	3.00E -94
Contig 1057 (CD852123)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	11 (0.54%)	0 (0%)	11	6.20	No hit	-
Contig 1124 (CD853276)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (0.15%)	22 (0.79%)	25	6.08	60S ribosomal protein L10	2.00E -95
No. analysed ESTs	827	187	320	971	2023	2778	7106			

For each contig, the number of ESTs identified in each library is presented and is followed by the percentage, calculated by dividing this number by the total number of analysed ESTs in each library.

^aContig name and accession number of a representative EST.

^bTotal number of ESTs in each contig.

large fraction of the genes appearing to be differentially expressed belongs to the protein synthesis functional class and seems to be over-expressed during late embryogenesis stages. For example, three ribosomal proteins (all over-represented in the HaDevR6 library) and one translation elongation fraction (over-abundant in the HaCotR library) have been identified. Among the putative differentially expressed proteins, proteins involved in energetic and primary metabolism such as apocytochrome *b*, cytochrome P450 like-TBP and aminopeptidase were found to be highly represented in HaHeaR and HaGlbR libraries respectively. This results reflect the vast increase in general metabolism occurring in these highly differentiating and developing tissues. Also, a protein having an important role in cell growth such as cell wall-associated hydrolase was found to be very abundant in heart-shaped embryos. Other genes defined as differentially expressed between the different stages of embryo development are involved in more scattered cell functions. For example, a putative senescence-related protein appears to be highly expressed during very early stages of embryo development. Likewise, a protein showing significant homology with an agglutinin is over-represented in the HaDevR6 library. Three of the differentially expressed genes, encoding proteins similar to hypothetical proteins and tuber agglutinin, also constitute highly abundant transcripts with contigs containing between 47 and 88 ESTs (Table 5).

Lastly, several contigs showing significant level of differential expression do not present homology with any already identified proteins or match with

hypothetical proteins of unknown function. Some of these genes may encode proteins that are specific to sunflower embryogenesis.

To validate digital northern analysis, six clones identified as over expressed in early stages of embryo development were subjected to real-time RT-PCR analysis. Quantitative RT-PCR confirmed that the level of expression of all the selected genes is higher in the early embryos stages (globular and heart stages) than in the cotyledonary stages (Table 6). Not unexpectedly, the ratios of expression levels determined by real-time RT-PCR is largely higher than those calculated from digital northern. This is undoubtedly due to the high sensitivity of real-time RT-PCR to detect the transcripts abundance compared to the *in silico* EST analysis which is not an absolute measure of gene expression.

Clustering of the cDNA libraries according to gene expression patterns mimics the successive embryo developmental events

To establish the relationship between the different cDNA libraries, they were associated via their common patterns of gene expression. Since HaDevR7 library was prepared from unfertilized ovules, a distinct tissue type from embryos, and showed a very different pattern of genes expression, it was excluded from this analysis. Using the method described by Ewing *et al.* (1999), we have built a dendrogram based on the calculation of pairwise distances to cDNA libraries. The cluster analysis presented in Figure 3 elucidates that the libraries HaGlbR and HaHeaR on one hand and

Table 6. Fold changes of mRNA level of some differentially expressed clone using real-time RT-PCR.

Cluster identification	Fold change					
	Unfertilized ovules	Globular embryos	Heart-shaped embryos	Early cotyledonary embryos	Mid cotyledonary embryos-4 dap	Late cotyledonary embryos-7 dap
Contig 1140	93	8778	5531	3	1	1
Contig 1137	39	165	347	3	1	2
Contig 1033	137	128	2091	16	1	1
Contig 1091	43	729	667	2	1	2
Contig 888	14	7	183	2	1	1
Contig 988	59	103	358	2	1	3

After normalization of the expression level of the target genes using β -Actin as an internal control, the lowest expression level of each clones in the six tissues under study was used to determine the fold change of transcript abundance.

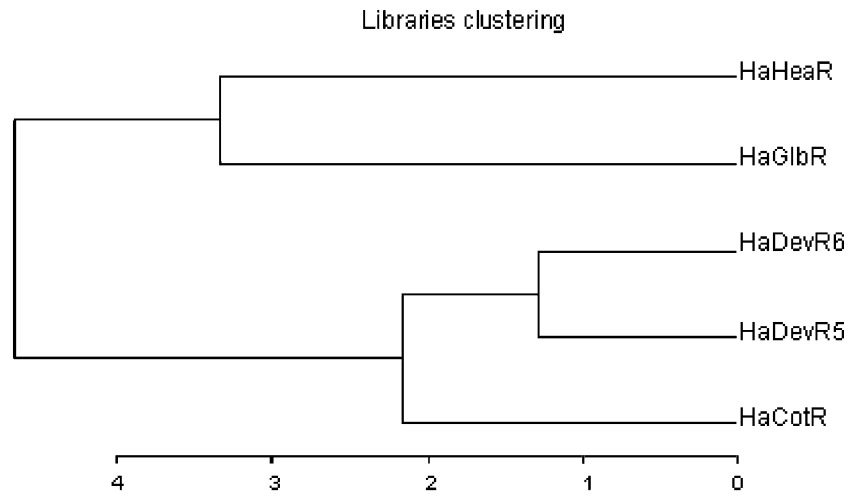


Figure 3. Dendrogram showing embryonic cDNA libraries similarities. Hierarchical clustering based on the average linkage and a Euclidean distance computed from the coordinates of the projection of the libraries on the four factorial axes of the FCA.

HaCotR, HaDevR5 and HaDevR6 on the other hand are closely linked and could express the most similar subsets of genes.

A more direct analysis, using the absolute frequencies of contigs per libraries to compute a factorial correspondence analysis, was performed (Figure 4). Remarkably, the biological sequence of embryo development, theoretically represented by

the HaGlbR, HaHeaR and HaCotR libraries in order, is perfectly respected along the first axis, confirming the reliability of the embryo screening method used in the first approach. HaDevR5 and HaDevR6 libraries appear to be closer to HaCotR library than to the libraries prepared from embryos at earlier stages of development. Thus, among all the studied libraries, only HaGlbR and

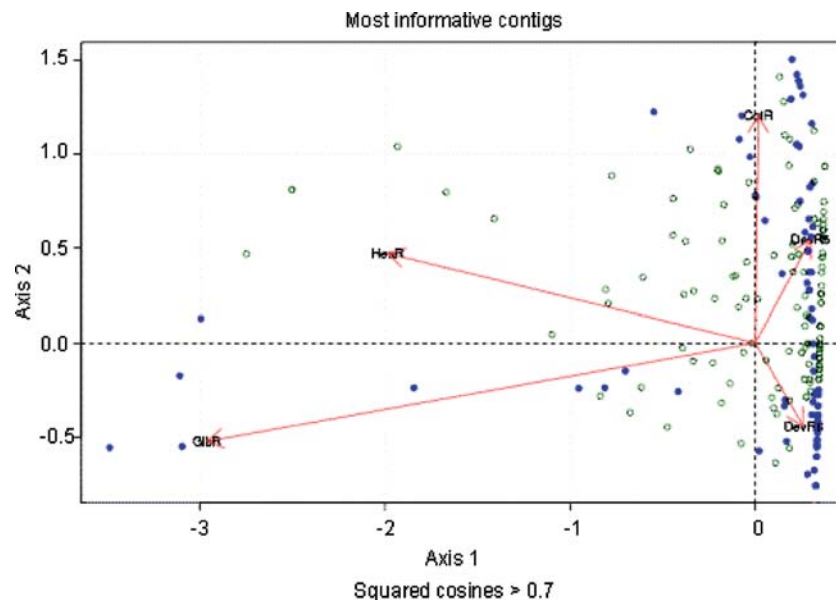


Figure 4. Factorial correspondence analysis of the libraries and contigs. Graphical representation of the projection of the contigs on the two main axes. The plan represents 67% of the information. Libraries are represented as vectors and contigs with points. Dotted circles represent contigs with a good projection on the plan and open circles represent contigs with less quality on the plan.

HaHeaR should contain genes induced at very early stages of embryogenesis.

Discussion

In the present work, which aims at gaining insight into the transcriptome induced during sunflower embryogenesis, we report the generation of more than 7100 EST sequences from sunflower embryos at diverse stages of development. As revealed by the results of sequences clustering, this original ESTs collection shows a low redundancy rate and is proved to be highly complex. Functional characterization of this data set reveals very similar profiles for each stage of embryo development. However, the types of genes expressed during the successive stages of development are different as illustrated by factorial correspondence analysis of the libraries and contigs (Figure 4). This analysis allowed us to identify numerous specific and differentially expressed genes, although the investigated tissues are closely related.

Libraries clustering analysis was performed to assess the similarities between the cDNA libraries. First, it points out that the biological sequence of embryo development is clearly respected between the HaGlbR, HaHeaR and HaCotR libraries. Second, HaDevR5 (4 dap) and HaDevR6 (7 dap) both appear to be closer to HaCotR library than to the libraries prepared from embryos at earlier stages of development. These findings are clearly exemplified in Figures 3 and 4. However, we found that the contribution of each of these three libraries to the global unigene is rather similar. This demonstrates that the development phases spanned by the HaDevR5 and HaDevR6 libraries are not redundant with the early cotyledonary stage.

Numerous embryo-defective mutants are known to be arrested at the transition between globular to heart-shaped stages. These developmental stages appear to constitute essential steps during which numerous substantial genes could be expressed. Therefore, the HaGlbR and HaHeaR cDNA libraries represent a rare resource to improve the discovery of genes involved at these crucial steps of development. Further sequencing effort of these libraries could be of interest.

HaGlbR and HaHeaR libraries have been experimentally normalized to increase gene discovery by reducing the frequency of highly

expressed genes. For these two cDNA libraries, two sequences with homologies to an integral membrane transporter protein and a hypothetical 14.3 kDa protein (AJ827951 and AJ828208) were found to exhibit high level of redundancy. These two genes might play important roles in these critical steps of development. In fact, some integral membrane proteins essential for plant embryogenesis have already been reported (Schrack *et al.*, 2000; Sparkes *et al.*, 2003). We also postulate that functional analysis of the hypothetical 14.3 kDa protein will be of great interest to clarify its role in sunflower embryo development. This protein has already been reported as over-expressed in sunflower protoplasts cultured in embryogenic conditions (Tamborindéguy *et al.*, 2004). cDNA libraries normalization does not necessarily limit the usefulness of the EST data set for comparison of gene expression levels. For example, in a study conducted on *Arabidopsis* seeds genomics, the removing of only the abundant storage-protein genes from the cDNA library seems not to alter the relative ESTs proportions of the vast majority of the other mRNAs (Ohlrogge and Benning, 2000; White *et al.*, 2000). Similarly, the normalization of HaGlbR and HaHeaR libraries, for reducing the frequency of two redundant sequences, could not prevent us to perform such digital northern.

Interestingly, the functional profile of the global unigene expressed in sunflower embryos is quite similar to the patterns of expression described for *Arabidopsis* developing seeds (White *et al.*, 2000). In *Arabidopsis* seeds, protein and primary metabolisms were also found to constitute the most abundant classes representing 7.1% and 17.4% of the unigene, respectively. However, the difference between these percentages reported by White *et al.* (2000) in *Arabidopsis* and those given in our study could be due to that sunflower contains non-green developing seed, and thus metabolic pathways involved in biosynthesis and accumulation of storage lipids and proteins could be different from those in green developing seeds. Other noticeable differences concern proteins thought to be involved in cell division and growth as well as cell structure which were found to be significantly more abundant in *Arabidopsis* seed. Our results indicate that differentiation-related functions are strongly activated in the sunflower developing embryos rather than cell

multiplication-related functions. This may be due to differences in growth rate between *Arabidopsis* and sunflower embryos.

Among genes appearing to be differentially expressed between the different developmental embryos stages, we found genes encoding ribosomal proteins. These results are consistent with the well-documented roles of ribosomal proteins in plants and animals cell division, growth and development (Barakat *et al.*, 2001). Also, they confirm previously reported data suggesting that ribosomal genes appear to be differentially regulated in different tissue types (Karsi *et al.*, 2002). Interestingly, a putative senescence-related protein appears to be highly expressed during the early stages of embryo development. This could be related to suspensor degeneration or to the formation of the future vascular bundles.

This analysis also permitted us to identify some putatively interesting genes, showing similarities with genes playing key roles in plant and animal embryogenesis. These clusters are listed in Table 7 with a brief description of their functions in embryo development. For example, *MAGO NASHI*-like sequences were identified with an *E*-value of 10^{-75} in the libraries from embryos collected on capitulae at 4 dap (HaDevR5) and 7 dap (HaDevR6). The *MAGO NASHI* gene was first described in *Drosophila melanogaster* with an assigned role in axis formation during oogenesis

(Boswell *et al.*, 1991). The *MAGO NASHI* protein is highly conserved among animals and plants suggesting that the roles of this protein could be conserved among the two kingdoms. The involvement of this protein in polarity acquisition in sunflower embryos is under investigation. A sunflower homologue of the *PINHEAD (PNH)* *A. thaliana* gene (CD852635, $E = 2 \times 10^{-83}$) was also found in the HaDevR5 library. The *PNH* gene is required for the shoot apical meristem formation during embryogenesis in *A. thaliana* (Newman *et al.*, 2002). Genetic analyses has shown that *PNH* acts concomitantly with *ARGONAUTE (AGO)*, a gene involved in the establishment of bilateral embryo symmetry (Lynn *et al.*, 1999). Interestingly, a homolog of *AGO* is present in the HaCotR library (AJ828503, $E = 6 \times 10^{-60}$). cDNA clones with high sequence similarity to *Arabidopsis* transcription factor leafy cotyledon 1 (*LEC1*), an essential regulator required for normal development during the early and late phase of embryogenesis (Kwong *et al.*, 2003; Lee *et al.*, 2003), and to *SCARECROW* a transcription factor involved in the asymmetric cell division within the ground tissue of the *Arabidopsis* root meristem (Heidstra *et al.*, 2004) have also been identified.

Most of genes playing key roles in plant embryogenesis have been identified using mutagenesis approach. Analysis of ESTs from developing seeds have already been established for

Table 7. Clusters showing similarities with plant and animal genes involved in embryogenesis.

Cluster identification ^a	Annotation	<i>E</i> -value	Role in embryogenesis	References
Contig 587 (CD852416)	Mago nashi-like protein	4.00E – 75	Axis formation during <i>D. melanogaster</i> oogenesis	Boswell <i>et al.</i> (1991)
Contig 811 (CD851235)	LEC1-like protein	8.00E – 53	Maintain of suspensor cell fate, specification of cotyledon identity in the early morphogenesis phase, initiation and/or maintain of the maturation phase and inhibition of precocious germination.	Kwong <i>et al.</i> (2003) Lee <i>et al.</i> (2003)
HaCotR007C02 (AJ828503)	Argonaute	6.00E – 60	Establishment of bilateral embryo symmetry in <i>A. thaliana</i>	Lynn <i>et al.</i> (1999)
DH0ALL30ZF02ZZM1 (CD852635)	PINHEAD protein (ZWILLE protein)	2.00E – 83	Regulation of axis determinacy in <i>A. thaliana</i>	Newman <i>et al.</i> (2002)
DH0AFF28ZF12ZZM1 (CD856397)	Putative SCARECROW gene regulator	6.00E – 06	Radial patterning of ground tissue in root in <i>A. thaliana</i>	Heidstra <i>et al.</i> (2004)

^aCluster name and accession number of a representative EST.

several dicotyledonous species (van de Loo *et al.*, 1995; White *et al.*, 2000; Suh *et al.*, 2003). To our knowledge, cDNA libraries generated from specific stages of embryo development (globular, heart-shaped and cotyledonary) have not yet been reported. Although the *Helianthus annuus* public EST data set available on the NCBI dbEST already consists of more than 63 100 entries, the 7106 sequences described herein were obtained from well defined stages of embryos. This set of EST represents unique resource, even among other dicotyledonous sequencing data and could be very fruitful to improve our knowledge about the crucial steps of plants embryogenesis development. This study will be completed by a microarray analysis of about 10 000 unique sequences obtained from a wide range of sunflower tissues. This should yield much more precise results about the gene expression profiles during this complex and essential developmental process. The functional validation through ectopic or antisenses expression is difficult to achieve in sunflower. It could be also postulated that major modifications of the key genes for embryogenesis will led to lethal phenotypes. We postulate that the creation of a collection of TILLING (Targeting Induced Local Lesions IN Genomes) mutants could be of interest to circumvent the major drawbacks to develop a reverse-genetic approach in sunflower.

Materials and methods

Tissues staining and confocal microscopy

Embryo sacs containing embryos at different developmental stages were sampled and fixed in a Picroformol Bouin acid solution for 4 h. Then, they have been rinsed for 30 min in two 50% ethanol bathes and stored at 4 °C in a 70% ethanol and 2% eosin solution. Confocal microscopy was performed using a Leica microscope with the 488 nm excitation line of an argon laser and a long pass emission filter at 520 nm.

cDNA libraries construction

For HaGlbR, HaHeaR and HaCotR cDNA libraries, the sunflower Ha300b genotype was used because of the big size of its capitulum and its high

yield in seeds. These cDNA libraries were created from immature embryos at different developmental stages and generated by auto-fertilization. Embryos were sampled by dissection of the seeds under binoculars, classified according to their developmental stages and stored indefinitely at -20 °C in RNA Later (Ambion). Then, total RNA was isolated according to Chomczynski and Sacchi protocol (1987).

For HadevR5, HaDevR6 and HaDevR7 cDNA libraries, the sunflower PSC8 genotype was used. Total RNA was extracted from unfertilized ovules or immature embryos collected on capitulae at different days after pollination (4 and 7 dap).

All cDNA libraries were built using the SMART cDNA Library Construction Kit (Clontech) according to the manufacturer's instructions. After ligation and amplification in the phagemid λ Triplex2, cDNAs were cloned into the pTriplex2 vector into the *E. coli* BM 25-8 strain by *in vivo* mass excision.

HaGlbR and HaHeaR libraries which originally showed high level of redundancy were experimentally normalized. To achieve this, the entire set of bacterial clones was arrayed in double offset on a macroarray filter using the Biogrid (BioRobotics) and hybridized with overgo probes specific to the two highly frequent sequences and designed on the GSC website (<http://www.genome.wustl.edu/tools>). Overgo probes hybridizations were realised according to Cai *et al.* protocol (1998).

Expressed sequence tag sequencing

Sequencing was carried on 1130 cDNAs clones from HaGlbR, 1332 cDNAs clones from HaHeaR, 1152 cDNAs clones from HaCotR, 2304 cDNAs clones from HaDevR5, 3744 cDNAs clones from HaDevR6 and 2496 cDNAs clones from HaDevR7 library (Table 1). For the sequencing of HaGlbR1, HaHeaR1 and HaCotR1 cDNA libraries, plasmid DNA were mini-prepared using the Montage Plasmid Miniprep₉₆ kit (Millipore) following manufacturer's instructions. cDNA clones sequencing reactions were performed from the 5' end, using BigDye Terminator v2.0, v3.0 and v3.1 (Applied Biosystems) technology according to manufacturer's instructions and purified with the Sequencing Reaction Clean Up kit (Millipore). Sequencing was carried

out on an ABI PRISM 3700 (Applied Biosystems). The HadevR5, HaDevR6 and HaDevR7 cDNA libraries were sequenced from the 5' end at the National Sequencing Center – Génoscope, Evry, France.

In silico sequence data analysis

Base calling and trimming of low-quality sequences was done using PHRED. Cloning vector was masked owing to cross-match program. Sequences that were shorter than 100 bp (excluding from poly-A tail) were removed for further analysis. Clustering, allowing contigs production, was realized with the PHRAP program. Homology searches for function assignment were performed on the Swissprot database for *E*-values below 10^{-5} . For contigs and singletons for which no matches were found, TrEMBL comparisons were made. Mitochondrial and plastid sequences were identified using the BLASTN program ($E < 10^{-20}$) on *A. thaliana* sequences. Contigs and singletons were classified in seventeen different functional groups using the kyoto encyclopedia of genes and genomes (KEGG) catalog (<http://www.genome.ad.jp/kegg/>). The method proposed by Stekel *et al.*, (2000) for comparing gene expression profiles from multiple cDNA libraries was used to assess the degree of difference in expression between ESTs from the six cDNA libraries.

Real-time RT-PCR

The expression levels of six differentially expressed ESTs were tested to confirm the results obtained from digital northern analysis. cDNA were produced using the SMART PCR cDNA Synthesis Kit (Clontech) according to the manufacturer's instructions. Gene specific primers were designed using the Primer Express software, Version 2.0 (Applied Biosystems). The reaction was performed in 20 μ l containing 10 μ l 2X Syber Green Mastermix (Applied Biosystems), 300 nM of each primer and 10 ng of cDNA. The PCR reactions were run in ABI PRISM 7900HT Sequence Detection System (Applied Biosystems) using the following program: 50 °C for 2 min, 95 °C for 10 min and 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Following PCR amplification, the reactions were subjected to temperature ramp to create the dissociation curve, measured as changes

in fluorescence measurements as a function of temperature, by which the non-specific products can be detected. The dissociation program was 95 °C for 15 s, 60 °C for 15 s followed by 20 min of slow ramp from 60 to 95 °C. Three replicates of each reaction were performed and β -Actin, a putative constitutively expressed gene, was used as internal control for normalization. The relative changes in gene expression were quantified using $2^{-\Delta\Delta}$ CT method as described by Livak and Schmittgen (2001).

Statistical analysis

The factorial correspondence analysis (FCA) was based on the contingency table of the number of occurrence of each contigs in each library. Contigs with an occurrence less than five in a library were removed from the data set. Hierarchical clustering was based on the average linkage and a Euclidean distance computed from the coordinates of the projection of the libraries on the four factorial axes of the FCA. Computations were done using R 1.9.0, on a PC running Linux (kernel 2.4–21) and Suse 9.0.

Accession numbers

HaGlbR: AJ827751–AJ827937; HaHeaR: AJ827938–AJ828257; HaCotR: AJ828258–AJ829228; HaDevR5: CD850803–CD852825; HaDevR6: CD852826–CD855603; HaDevR7: CD 855604–CD856478.

Acknowledgments

This work was partially supported by funds from the Génoplante GOP-HG01 project and from the Toulouse Genomics Center (French Network of Genomics Centers). We thank our colleagues in the sunflower Génoplante project for valuable discussions and suggestions and Cécile Donnadiou Tonon for advice and access to robotic equipments for libraries construction. Cécile Ben and Cecilia Tamborindeguy are supported by doctoral fellowships of the French MENRT; Tarek Hwezi, Frédérique Bena, Nathalie Ladouce and Sébastien Moretti were supported by grants from the Génoplante GOP-HG01 project.

References

- Al-Chaarani, G., Gentzbittel, L., Huang, X.Q. and Sarrafi, A. 2004. Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* DOI 10.1007/s00122-004-170-1 (In press)
- 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(2): 398–415.
- Belanger, K. and Quatrano, R. 2000. Polarity: the role of localized secretion. *Curr. Opin. Plant Biol.* 3: 67–72.
- Boswell, R.E., Prout, M.E. and Steichen, J.C. 1991. Mutations in a newly identified *Drosophila melanogaster* gene, mago nashi, disrupt germ cell formation and result in the formation of mirror-image symmetrical double abdomen embryos. *Development* 113: 373–384.
- Caboche, M. and Boucly, M. 2000. The Genoplante programme, a mobilizing programme in plant genomics. *Comptes Rend. Acad. Agric. France* 86: 159–173.
- Cahoon, E.B., Carlson, T.J., Ripp, K.G., Schweiger, B.J., Cook, G.A., Hall, S.E. and Kinney, A.J. 1999. Biosynthetic origin of conjugated double bonds: production of fatty acid components of high-value drying oils in transgenic soybean embryos. *Proc. Natl. Acad. Sci. USA* 96(22): 12935–12940.
- Cai, W.W., Reneker, J., Chow, C.W., Vaishnav, M. and Bradley, A. 1998. An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. *Genomics* 54(3): 387–397.
- Chaudhury, A., Koltunow, A., Payne, T., Luo, M., Tucker, M., Dennis, E. and Peacock, W. 2001. Control of early seed development. *Annu. Rev. Cell Dev. Biol.* 17: 677–699.
- Chomczynski, P. and Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162: 156–159.
- Ewing, R.M., Kahla, A.B., Poirot, O., Lopez, F., Audic, S. and Claverie, J.M. 1999. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res.* 9(10): 950–959.
- Fedorova, M., van de Mortel, J., Matsumoto, P.A., Cho, J., Town, C.D., VandenBosch, K.A., Gantt, J.S. and Vance, C.P. 2002. Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiol.* 130(2): 519–537.
- Fernandes, J., Brendel, V., Gai, X., Lal, S., Chandler, V.L., Elumalai, R.P., Galbraith, D.W., Pierson, E.A. and Walbot, V. 2002. Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and microarray hybridization. *Plant Physiol.* 128(3): 896–910.
- Gentzbittel, L., Abbott, A., Galaud, J., Georgi, L., Fabre, F., Liboz, T. and Alibert, G. 2002. A bacterial artificial chromosome (BAC) library for sunflower, and identification of clones containing genes for putative transmembrane receptors. *Mol. Genet. Genomics* 266: 979–987.
- Goldberg, R., de Paiva, G. and Yadegari, R. 1994. Plant embryogenesis: zygote to seed. *Science* 266: 605–614.
- Heidstra, R., Welch, D. and Scheres, B. 2004. Mosaic analyses using marked activation and deletion clones dissect *Arabidopsis* SCARECROW action in asymmetric cell division. *Genes Dev.* 18(16): 1964–1969.
- Jurgens, G. 1995. Axis formation in plant embryogenesis: cues and clues. *Cell* 81: 467–470.
- Jurgens, G. 2001. Apical-basal pattern formation in *Arabidopsis* embryogenesis. *EMBO J* 20: 3609–3616.
- Jurgens, G., Grebe, M. and Steinmann, T. 1997. Establishment of cell polarity during early plant development. *Curr. Opin. Cell. Biol.* 9: 849–852.
- Karsi, A., Patterson, A., Feng, J. and Liu, Z. 2002. Translational machinery of channel catfish: I. A transcriptomic approach to the analysis of 32 40S ribosomal protein genes and their expression. *Gene* 291: 177–186.
- Kwong, R.W., Bui, A.Q., Lee, H., Kwong, L.W., Fischer, R.L., Goldberg, R.B. and Harada, J.J. 2003. LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell.* 15(1): 5–18.
- Lee, H., Fischer, R.L., Goldberg, R.B. and Harada, J.J. 2003. *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc. Natl. Acad. Sci. USA* 100(4): 2152–2156.
- 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. *Meth. Meth* 25(4): 402–408.
- Lynn, K., Fernandez, A., Mitsuhiro, A., 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.
- Meinke, D.W. 1985. Embryo-lethal mutants of *Arabidopsis thaliana*: analysis of mutants with a wide range of lethal phases. *Theor. Appl. Genet.* 69: 543–552.
- Newman, K.L., Fernandez, A.G. and Barton, M.K. 2002. Regulation of axis determinacy by the *Arabidopsis* PINHEAD gene. *Plant Cell.* 14(12): 3029–3042.
- Ogihara, Y., Mochida, K., Nemoto, Y., Murai, K., Yamazaki, Y., Shin-I, T. and Kohara, Y. 2003. Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. *Plant J.* 33(6): 1001–1011.
- Ohlrogge, J. and Benning, C. 2000. Unraveling plant metabolism by EST analysis. *Curr. Opin. Plant Biol.* 3(3): 224–228 (Review).
- Ronning, C.M., Stegalkina, S.S., Ascenzi, R.A., Bougri, O., Hart, A.L., Utterbach, T.R., Vanaken, S.E., Riedmuller, S.B., White, J.A., Cho, J., Pertea, G.M., Lee, Y., Karamycheva, S., Sultana, R., Tsai, J., Quackenbush, J., Griffiths, H.M., Restrepo, S., Smart, C.D., Fry, W.E., Van Der Hoeven, R., Tanksley, S., Zhang, P., Jin, H., Yamamoto, M.L., Baker, B.J. and Buell, C.R. 2003. Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol.* 131(2): 419–429.
- Schrick, K., Mayer, U., Horrichs, A., Kuhnt, C., Bellini, C., Dangl, J., Schmidt, J. and Jurgens, G. 2000. FACKEL is a sterol C-14 reductase required for organized cell division and expansion in *Arabidopsis* embryogenesis. *Genes Dev.* 14(12): 1471–1484.
- Souter, M. and Lindsey, K. 2000. Polarity and signalling in plant embryogenesis. *J. Exp. Bot.* 51: 971–983.
- Sparkes, I.A., Brandizzi, F., Slocombe, S.P., El-Shami, M., Hawes, C. and Baker, A. 2003. An *Arabidopsis* pex10 null mutant is embryo lethal, implicating peroxisomes in an essential role during plant embryogenesis. *Plant Physiol.* 133(4): 1809–1819.
- Stekel, D., Git, Y. and Falciani, F. 2000. The comparison of gene expression from multiple cDNA libraries. *Genome Res.* 10: 2055–2061.

- Suh, M.C., Kim, M.J., Hur, C.G., Bae, J.M., Park, Y.I., Chung, C.H., Kang, C.W. and Ohlrogge, J.B. 2003. Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. *Plant Mol. Biol.* 52(6): 1107–1123.
- Tamborindéguy, C., Ben, C., Liboz, T. and Gentzbittel, L. 2004. Sequence evaluation of four specific cDNA libraries for developmental genomics of sunflower. *Mol. Genet. Genomics* 271(3): 367–375.
- Tang, S., Yu, J.K., Slabaugh, B., Shintani, K. and Knapp, J. 2002. Simple sequence repeat map of the Sunflower genome. *Theor. Appl. Genet.* 105: 1124–1136.
- van de Loo, F.J., Turner, S. and Somerville, C. 1995. Expressed Sequence Tags from Developing Castor Seeds. *Plant Physiol.* 108(3): 1141–1150.
- West, M. and Harada, J.J. 1993. Embryogenesis in higher plants: an overview. *Plant Cell* 5: 1361–1369.
- White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., de Ilarduya, O.M., Jaworski, J.G., Ohlrogge, J.B. and Benning, C. 2000. A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiol.* 124(4): 1582–1594.
- Zimmerman, J.L. 1993. Somatic embryogenesis: a model for early development in higher plants. *Plant Cell* 5: 1411–1423.

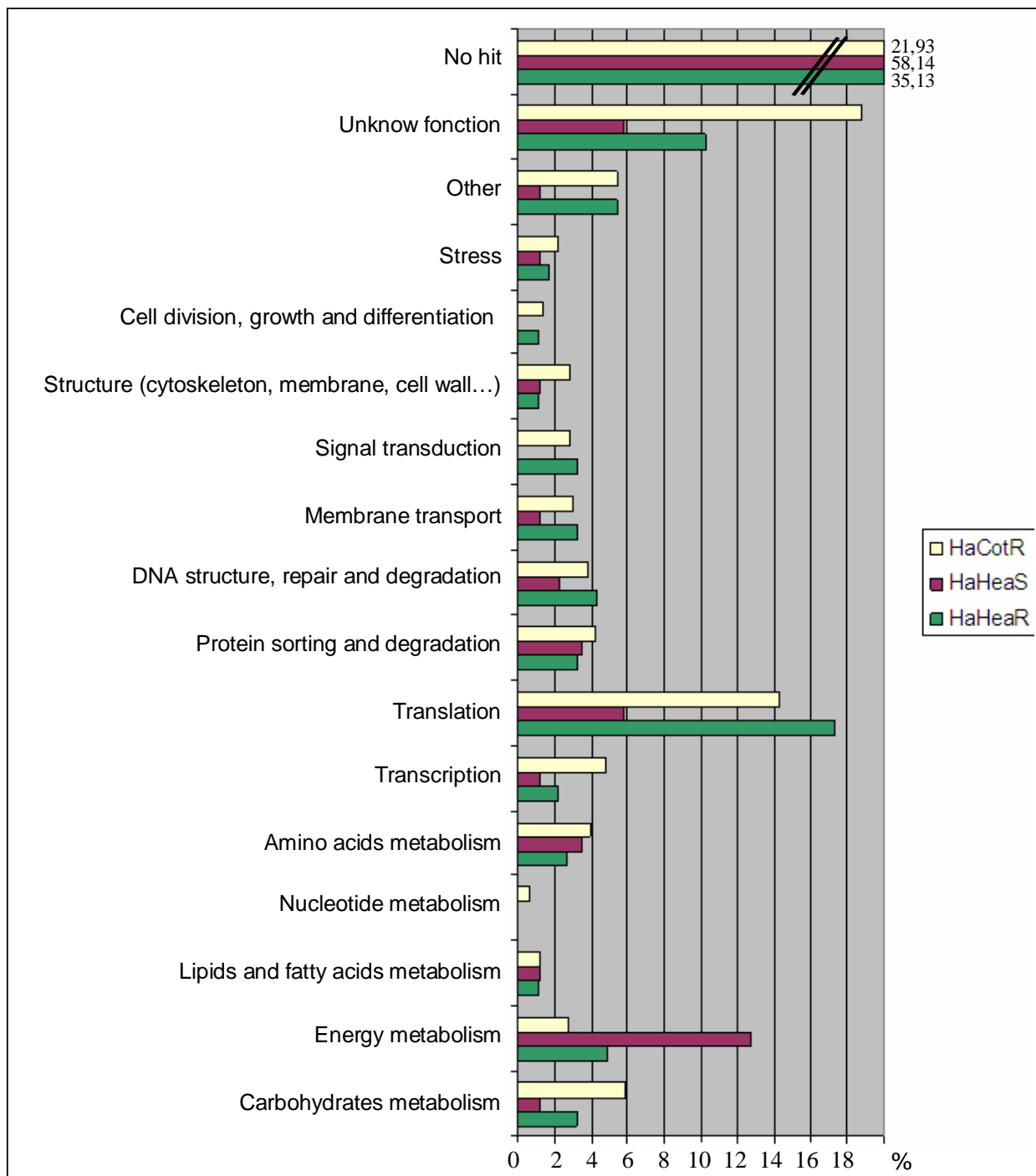


Figure 25: Classification fonctionnelle de l'unigène des banques HaHeaR, HaHeaS et HaCotR.

Pour chacune des banques d'ADNc, les contigs et singletons obtenus après assemblage des EST par le programme Phrap ont été classés selon leur fonction putative en utilisant les critères du KEGG. Les données sont représentées comme le pourcentage de l'unigène pour chaque catégorie fonctionnelle.

II.2. Analyse de la banque SSH enrichie en transcrits spécifiques d'embryons au stade cœur.

Parallèlement à la construction et à l'analyse des banques d'ADNc standards supposées contenir un échantillonnage de l'ensemble des gènes exprimés dans la condition étudiée, une banque d'ADNc de type SSH comparant les transcrits d'embryons aux stades cœur et cotylédonaire et enrichie en gènes préférentiellement exprimés au stade cœur a été réalisée. Le but de cette démarche était de caractériser le transcriptome spécifiquement induit au cours des étapes d'élaboration du plan d'organisation embryonnaire et d'acquisition de la symétrie bilatérale afin d'identifier des voies potentielles de régulation transcriptionnelle du processus complexe de morphogenèse embryonnaire.

Cette banque compte 6144 clones dont 504 ont été séquencés. Au final, 212 EST, d'une longueur moyenne de 264 bp, ont été publiées (numéro d'accèsion : AJ829229-AJ829440) représentant 135 séquences potentiellement uniques. Selon les résultats de l'assemblage par le programme Phrap, l'unigène est réparti en 129 singletons et 8 contigs regroupant entre 2 et 64 EST. Le taux de redondance relativement élevé de cette banque, qui met en évidence un dysfonctionnement de l'étape de normalisation lors de la construction de la banque, est en fait dû à l'abondance d'une seule et même séquence représentant 30% du total des EST soumises. L'annotation fonctionnelle de ce contig révèle une similarité avec la protéine *ycf3* du photosystème I illustrant l'importance de la biogenèse de plastides fonctionnels au stade cœur de l'embryogenèse végétale comme indiqué précédemment par l'analyse de mutants menées sur la graine d'*Arabidopsis thaliana* (Apuya et al., 2002 ; Despres et al., 2001 ; Uwer et al., 1998). Malgré la fréquence élevée de cette séquence, le pourcentage de séquences redondantes (assemblées dans un contig regroupant au moins deux séquences) de cette banque, qui est de 39.1%, reste inférieur à celui des banques de référence HaHeaR et HaCotR (54.4% et 42.6%, respectivement) révélant un enrichissement en séquences uniques de la banque SSH comparée aux banques d'ADNc standards.

Comme pour les banques de référence, la caractérisation fonctionnelle de l'unigène de cette banque, par recherche d'homologies avec le programme BLASTX dans les bases de données publiques, a mis en évidence l'importance relative des gènes impliqués dans la synthèse protéique (Fig.25). En revanche, en comparaison des résultats obtenus pour les banques de référence, on assiste à une diminution des gènes associés au métabolisme primaire de base, gènes qui correspondent le plus souvent à des fonctions de ménage. Ceci laisserait supposer une bonne efficacité de la soustraction des deux populations de transcrits pendant la création de la banque. En contrepartie, nous observons une augmentation des gènes impliqués dans le métabolisme énergétique suggérant que le stade cœur correspondrait à une phase d'activité cellulaire intense. Ces données semblent en accord avec des observations faites chez *Arabidopsis thaliana* où le stade cœur correspond à l'initiation du développement autotrophique des embryons. En effet, alors que la croissance des jeunes embryons dépend de l'apport de nutriments et de facteurs de croissance issus des tissus maternels *via* le suspenseur, cette structure embryonnaire éphémère dégénère au stade cœur du développement embryonnaire.

Théoriquement, la technique SSH permet d'identifier des gènes exprimés faiblement. Or, comparée aux banques de référence, l'analyse de la banque SSH n'a pas révélé un pourcentage plus élevé de séquences prétendues rares telles que les facteurs de transcription, les signaux de transduction ou encore les facteurs de croissance et de différenciation. Ceci pourrait témoigner d'une profondeur de séquençage insuffisante de cette banque qui mériterait peut être d'être plus exploitée.

Un pourcentage très élevé d'EST et supérieur à celui des banques de référence ne montre aucune homologie avec des séquences répertoriées confirmant la puissance de cette technique pour la découverte de nouveaux gènes.

III. Difficultés rencontrées pour l'obtention des données EST.

La construction de banque d'ADNc et le séquençage haut débit d'EST constituent une approche expérimentale très lourde et très technique exigeant un certain savoir-faire. Plusieurs phases du protocole telles que l'extraction d'ARNm en quantité et qualité suffisantes ou les conditions de culture bactérienne représentent des étapes limitantes pour la qualité de l'information obtenue au final. Au cours de la création de ces ressources, plusieurs difficultés ont été rencontrées.

III.1. Un fort taux de redondance des banques HaGlbR, HaHeaR et HaHeaS.

Malgré la normalisation, la banque SSH présente un fort taux de redondance avec la présence en grande quantité d'une séquence similaire à la protéine ycf 3 du photosystème I. De même, les banques construites à partir des embryons aux stades globulaire et cœur (HaGlbR et HaHeaR) contiennent deux séquences très redondantes montrant des similarités avec une protéine prédite et un transporteur membranaire. Afin de favoriser la découverte de nouveaux gènes, un pré-criblage de ces banques par hybridation des clones déposés sur filtre avec des sondes overgos spécifiques de chacune des séquences montrant un fort taux de redondance a été effectué avant d'achever le séquençage. La technique utilisant des sondes overgos pour identifier les clones homologues à ces séquences a été choisie pour sa facilité de mise en œuvre, sa très grande spécificité et la qualité des résultats d'hybridation obtenus, très propres, très lisibles et faciles à interpréter. Malgré l'élimination des séquences les plus abondantes des banques d'ADNc de référence, l'analyse des niveaux d'expression génique reste toujours possible.

III.2. Existence de plusieurs facteurs expérimentaux limitant le rendement du séquençage.

Un effet ‘plaque de culture bactérienne’ sur la qualité du séquençage des EST a été mis en évidence révélant l’importance des conditions de culture des clones bactériens dans les plaques mères de culture. Ainsi, la banque HaCotR, qui présente un rendement de séquençage très élevé (de l’ordre de 85%), a été ordonnée dans des plaques de culture à 96 puits contenant 200µl de milieu nutritif. En revanche, les banques HaGlbR et HaHeaR dont le taux de retour de séquences est beaucoup moins satisfaisant (de l’ordre de 40%) ont été directement repiquées dans des plaques à 384 puits dans un volume de milieu de 80µl. On peut supposer qu’un trop faible volume de culture ne permet pas une bonne homogénéisation de la culture par agitation ni une aération suffisante menant probablement à la sécrétion par les cellules stressées de toxines inhibant la réaction de séquence.

La nature de la souche bactérienne et du vecteur de clonage utilisés pour la construction des banques influe également sur le rendement de l’extraction de l’ADN plasmidique et sur les conditions à pratiquer au cours de la réaction de séquençage (T°C d’appariement différente pour chaque combinaison d’amorces spécifique de chaque plasmide). Pour ces travaux de thèse, la souche bactérienne d’*E.Coli* BM25.8 associée au vecteur de clonage pTriplex2 a été utilisée pour la construction des trois banques d’ADNc de référence alors que pour la banque SSH, les ADNc ont été clonés dans le plasmide PCRII introduit dans la souche bactérienne *E. Coli* DH5α. Pour l’extraction d’ADN plasmidique, de meilleurs rendements ont été obtenus avec la souche bactérienne d’*E.Coli* BM25.8 associée au vecteur de clonage pTriplex2. Les conditions de réaction de séquence spécifiques à chacun de ces vecteurs de clonage sont détaillées dans la partie Matériels et Méthodes.

Conclusion

Les banques d’ADNc réalisées au cours de cette thèse à partir de très jeunes embryons de tournesol constituent une ressource inédite pour l’étude de l’embryogenèse précoce du tournesol mais également de l’ensemble des espèces dicotylédones exalbuminées. Malgré des taux de redondance plus ou moins élevés selon les banques, l’annotation des EST a révélé la présence d’un grand nombre de nouvelles séquences ainsi qu’une très grande diversité fonctionnelle (métabolismes de base, structure cellulaire, signalisation, régulation...) attestant d’une bonne qualité des banques d’ADNc. L’analyse bioinformatique des EST a également permis d’obtenir un premier aperçu des programmes génétiques successivement induits au cours de stades critiques du développement embryonnaire. Toutefois, il est à noter que l’absence d’une séquence dans une banque, qui peut tout simplement être due au fait que l’ADNc n’a pas été cloné ou séquencé, n’implique pas obligatoirement que ce gène n’est pas exprimé dans la condition étudiée. De même, une

séquence présente exclusivement dans une condition n'est pas forcément différentiellement exprimée entre les stades de développements testés. Ainsi, si l'approche EST constitue la première étape vers la caractérisation du transcriptome lors de l'embryogenèse précoce du tournesol, les résultats obtenus doivent être confirmés et complétés par une analyse du niveau d'expression génique par le biais de la technique microarray appliquée à un beaucoup plus grand nombre de gènes.

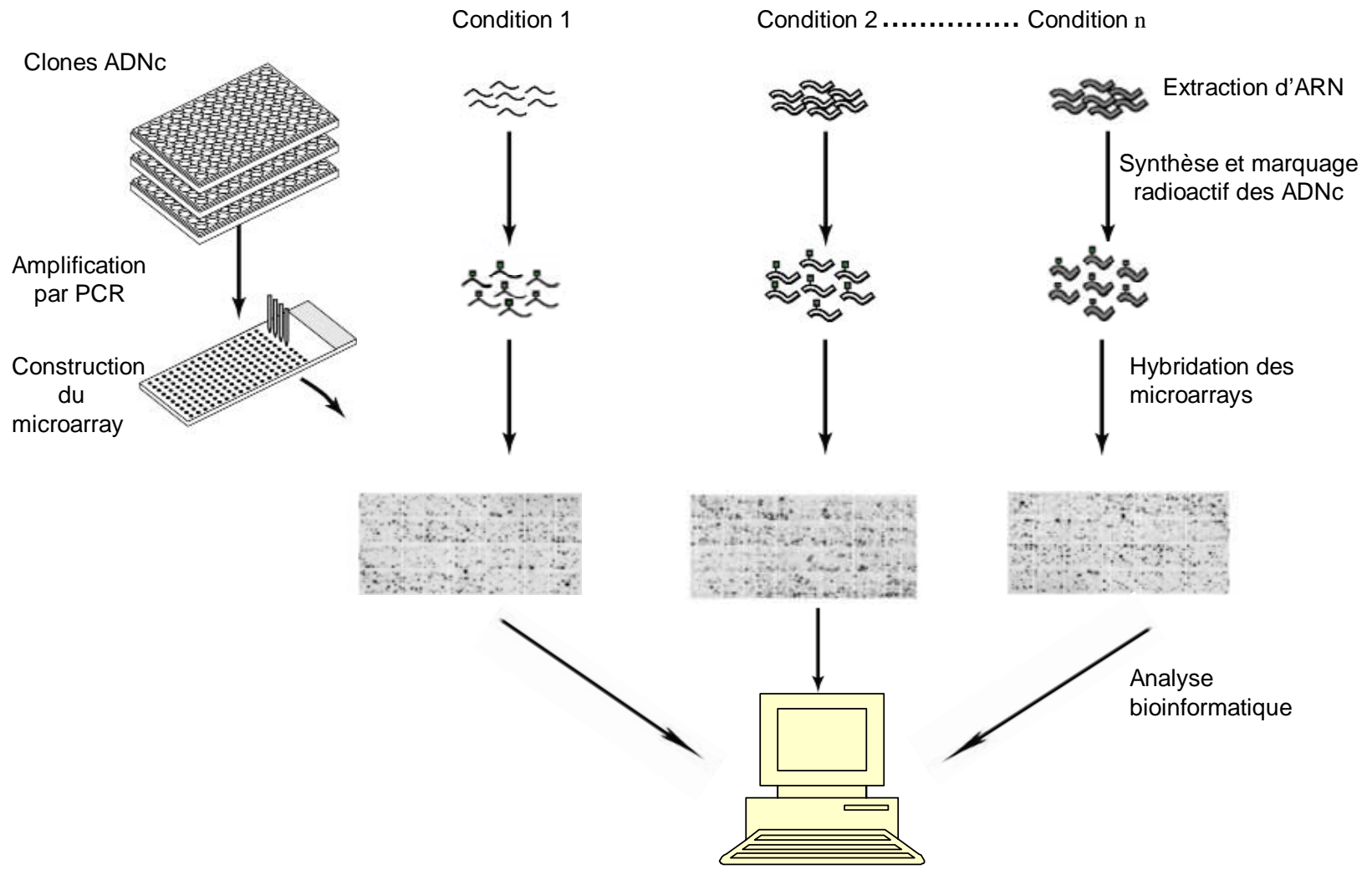


Figure 26: Schéma représentant la construction et l'utilisation des microarrays.

Chapitre II – Approche ‘Microarrays’ : Analyse simultanée de l’expression de 8 185 gènes lors de l’embryogenèse précoce du tournesol.

Depuis plusieurs années, de nombreuses approches d’analyse des patrons d’expression génique ont été développées afin de caractériser les gènes et plus généralement les programmes génétiques induits dans les cellules dans des conditions particulières. Parmi ces méthodes, on trouve les techniques de criblage différentiel telles que le Differential Display (Liang and Pardee, 1992) ou le cDNA-AFLP (Bachem et al., 1996, Tamborindéguy et al., 2004b), l’analyse de collection d’EST suite à la construction de banques soustractives (Kuang et al., 1998) ou par technique SAGE (Velculescu et al., 1995) ou encore l’utilisation de microarrays (Schena et al., 1995) qui, comparée aux techniques précédentes, revêt l’immense avantage de permettre l’étude simultanée de plusieurs milliers de gènes différents et la comparaison de diverses conditions. Du fait de ses atouts non négligeables, cette technique a été largement utilisée durant ces dernières années chez diverses espèces végétales dans le but d’élucider les mécanismes moléculaires contrôlant divers processus biologiques tels que le remplissage du grain (Ruuska et al., 2002), la sénescence (Lin and Wu, 2004) ou encore les interactions plantes-pathogènes (Wan et al., 2002).

Le principe de la technique de microarray (Fig.26) consiste à extraire les ARNm de chaque condition d’intérêt et à marquer radioactivement ou par fluorescence les ADNc obtenus à partir de ces ARN constituant la sonde complexe. Ces échantillons d’ADNc marqués sont ensuite hybridés à des molécules d’ADN représentatives de différents gènes qui ont été fixées sur un support solide (membrane nylon ou lame de verre). L’ADN fixé au support se trouvant en excès par rapport à la sonde complexe et l’hybridation étant arrêtée en phase linéaire, l’intensité du signal radioactif ou fluorescent mesurée suite à l’hybridation reflète la quantité des différents transcrits au sein de la population d’ARNm de départ. Après conversion des signaux d’intensité en valeurs numériques, une analyse statistique des données d’hybridation est réalisée *in silico* afin d’identifier les gènes différentiellement exprimés entre les conditions testées.

Afin de compléter les informations préliminaires concernant les gènes exprimés au cours des phases précoces de l’embryogenèse obtenues par l’analyse *in silico* des données EST, une approche de transcriptomique ‘haut débit’ utilisant la technique de microarrays a donc été entreprise.

I. Choix méthodologiques et mise en œuvre technique.

I.1. Construction d'un microarray ADNc sur membrane nylon.

I.1.1. Raisons du choix de la technique employée.

Aujourd'hui, trois types différents de puces à ADN distingués selon la nature des molécules déposées et du support utilisé, la densité des spots et les méthodes de fabrication et d'hybridation, sont disponibles : les macroarrays, les microarrays sur lame de verre ou membrane de nylon et les puces à oligos.

Les **macroarrays**, constitués de produits PCR déposés sur membrane de nylon voire même directement de clones bactériens renfermant les séquences d'intérêt, ont été utilisés à plusieurs reprises pour étudier l'expression de gènes chez les animaux (Nguyen et al., 1995 ; Pietu et al., 1996) et chez les végétaux (Fedorova et al., 2002 ; Sreenivasulu et al., 2002). Toutefois, étant donnée la faible densité du dépôt (inférieure à 50 spots/cm²), seul un nombre restreint de séquences peut être intégrées sur ce type de puce rendant ce procédé inadéquat pour l'analyse de l'expression d'un grand nombre de gènes. De plus, la taille de ces filtres à ADN est relativement importante ce qui entraîne des complications techniques lors de l'hybridation avec un besoin d'une grande quantité de sondes complexes, ou lors de l'analyse des données du fait d'un risque d'hétérogénéité du signal rendant la quantification difficile.

Les **microarrays**, caractérisés par une densité de dépôt beaucoup plus élevée (jusqu'à 20 000 spots/cm²) autorisant le dépôt de plusieurs milliers de gènes sur une puce de taille réduite, sont beaucoup mieux adaptés à l'analyse simultanée et à grande échelle de l'expression de grandes quantités de gènes différents. La technique microarray a été développée sur deux supports de nature différente : les *lames de verre* ou les *membranes de nylon*, chacun de ces supports revêtant des avantages et des inconvénients. Alors que les microarrays sur membranes nylon sont le plus souvent hybridés avec des sondes complexes marquées radioactivement, les puces sur lames de verre sont généralement hybridées à l'aide de sondes froides fluorescentes, deux sondes pouvant être hybridées pour une même hybridation diminuant ainsi le nombre d'hybridations nécessaires à l'analyse. Cependant, alors que les microarrays sur lame de verre sont pour l'instant à usage unique, les filtres sur membrane nylon peuvent être réutilisés plusieurs fois (jusqu'à 5 hybridations successives réalisées à partir d'une seule et même membrane) réduisant ainsi le coût non négligeable de ces analyses. De plus, la capacité de réhybridation des microarrays sur membrane de nylon permet d'élaborer un plan d'expérience autorisant la comparaison directe des différentes conditions étudiées entre elles alors que les analyses utilisant les lames de verre implique la comparaison des conditions d'intérêt avec un tissu témoin. Enfin, contrairement aux microarrays sur lames de verre très exigeants en quantité de sondes complexes, seulement une faible quantité d'ARN est nécessaire pour l'hybridation des microarrays de nylon hybridés avec des sondes marquées radioactivement (Bertucci et al., 1999).

Enfin, outre une densité inégalable (250 000 spots/cm²), les **puces à oligos**, où la cible est constituée par des oligonucléotides courts (de 20 à 80 mers), présentent de nombreux avantages pour ce type d'étude. En effet, dans ce cas, les molécules déposées sur lame de verre sont synthétisées *in vitro* : le design de ces séquences oligonucléotidiques spécifiques d'un gène qui se fait *in silico* ne nécessite que la connaissance de la séquence d'intérêt et n'exige donc pas la possession physique du clone bactérien. De plus, elle laisse à l'utilisateur un large champ d'action, celui-ci pouvant désigner les oligos au sein de domaines très conservés afin d'étudier le comportement de familles multigéniques entières ou au contraire au niveau de régions hautement variables pour discriminer chacun des membres d'une même famille et analyser précisément les profils d'expression de différents isoformes. Toutefois, souvent utilisée dans l'optique de créer un microarray représentatif de l'ensemble des gènes d'un organisme, cette approche requiert une connaissance minimale du génome de l'espèce étudiée et a donc été majoritairement employée pour des espèces modèles (Harmer et al., 2000 ; Hwang et al., 2005) même si son utilisation commence à se généraliser (Arpat et al., 2004). De plus, le coût encore très élevé de cette technique, qui exige l'emploi de méthodes et d'équipements particuliers, freine son expansion. Parmi les puces à oligos, les puces de type Affimetrix Genechips, construites par synthèse *in situ* des oligos directement sur la lame de verre par photolithographie, sont également une solution technique très intéressante mais restreinte à un très petit nombre d'espèces.

Etant donné le nombre relativement élevé des séquences constituant l'unigène sélectionné pour notre étude (supérieur à 8100 séquences potentiellement uniques) et la faible quantité de matériel végétal disponible pour la production des sondes complexes, la technique utilisant des microarrays de capacité intermédiaire entre les macroarrays et les puces à oligos nous est apparue la mieux adaptée. De plus, les puces sur membrane de nylon permettant plusieurs hybridations successives à partir d'un même filtre et la mise au point d'un plan expérimental efficace nous ont semblé plus attrayantes que les microarrays sur lame de verre.

La stratégie adoptée au sein du laboratoire a donc consisté à produire des microarrays à partir de produits PCR spottés sur membrane de nylon hybridés à l'aide de sondes marquées radioactivement.

1.1.2. Construction du microarrays.

1.1.2.1. Construction d'une puce générale plutôt qu'une puce thématique.

Le microarray à 10K construit durant cette thèse a été produit à partir d'un unigène total de 8 185 séquences uniques obtenu par assemblage grâce au programme Phrap d'une collection d'EST comptant plus de 25 840 séquences issues de banques d'ADNc de tournesol réalisées au sein de notre laboratoire et pour le programme Génoplante. Ces banques d'ADNc standards ou SSH ont été préparées à partir de tissus d'origines très diverses : tissus embryonnaires, ovules non fécondés, protoplastes cultivés en conditions embryogènes ou non embryogènes mais également feuille, apex, tige, plantes ayant subi des stress dus à des

pathogènes... Les banques d'ADNc produites à partir d'embryons à différents stades de développement au cours de cette thèse (HaGlbR, HaHeaR, HaCotR et HaHeaS) ont contribué à un enrichissement non négligeable de l'unigène déposé sur le microarray. En effet, l'assemblage des 1 690 EST publiques issues de ces banques avec l'ensemble des EST provenant des autres banques d'ADNc a généré 383 séquences uniques pouvant être exclusivement attribuées aux 4 banques d'ADNc embryogènes réalisées au laboratoire, ce qui représente 4.7% de l'unigène total présent sur la puce et 22.7% des séquences publiques de ces banques.

En augmentant l'unigène testé sur le microarray, la construction d'une puce générale créée à partir de banques d'ADNc d'origines diverses plutôt que d'une puce thématique ne comportant uniquement des clones issus de banques d'ADNc construites à partir de tissus embryonnaires, devraient permettre de compléter le réseau des fonctions cellulaires impliquées dans les phases précoces du développement embryonnaire et optimiser nos chances d'identifier des gènes jouant un rôle clé dans ce processus biologique en révélant des gènes non clonés dans les banques d'ADNc embryogènes.

1.1.2.2. Sélection de l'EST la plus représentative pour chaque contig.

Afin de sélectionner l'EST la plus représentative pour chaque contig, l'EST la plus longue et située en région 3' du cluster est choisie. Ces clones ADNc, susceptibles d'être situés dans la région 3' non traduite très variable, sont les plus spécifiques de chaque contig. Pour les contigs les plus longs, lorsque les EST représentant les extrémités 3' et 5' ne se superposent pas, un clone à chaque extrémité est gardé afin de révéler toute variabilité entre les signaux d'hybridation pour les EST situées dans les deux régions opposées.

1.1.2.3. Choix des témoins.

Afin d'évaluer la qualité des puces à ADNc et des résultats d'hybridation obtenus, différents témoins ont été ajoutés. Ainsi, pour valider l'intégrité des sondes complexes utilisées pour l'hybridation, des gènes potentiellement ubiquitaires tels que la calmoduline, l'ubiquitine ou la β -tubuline ont été intégrés au microarray. De même, afin de vérifier l'absence de toute contamination et pour évaluer le bruit de fond, différents témoins négatifs tels que de l'eau, des séquences spécifiques au règne animal (desmine et séquences de porc) ou encore des gènes d'origine bactérienne tels que les gènes NPTII et GUS ont également été introduits.

Au final, 10 368 produits PCR, comptant 9 404 clones d'ADNc de tournesol et des témoins positifs et négatifs, ont été déposés en double dépôt sur le microarray.

I.2. Elaboration d'un plan d'expérience performant pour faciliter l'analyse des données.

I.2.1. Sélection des échantillons à tester : Stades embryonnaires d'intérêt et témoins.

Afin de caractériser les gènes successivement induits au cours de l'embryogenèse précoce chez le tournesol, le microarray à 10K a été hybridé avec des sondes complexes préparées à partir d'embryons aux trois stades de développement d'intérêt (globulaire, cœur et cotylédonaire). De plus, afin de valider la spécificité des profils d'expression obtenus pour les tissus embryonnaires, deux témoins externes ont également été employés :

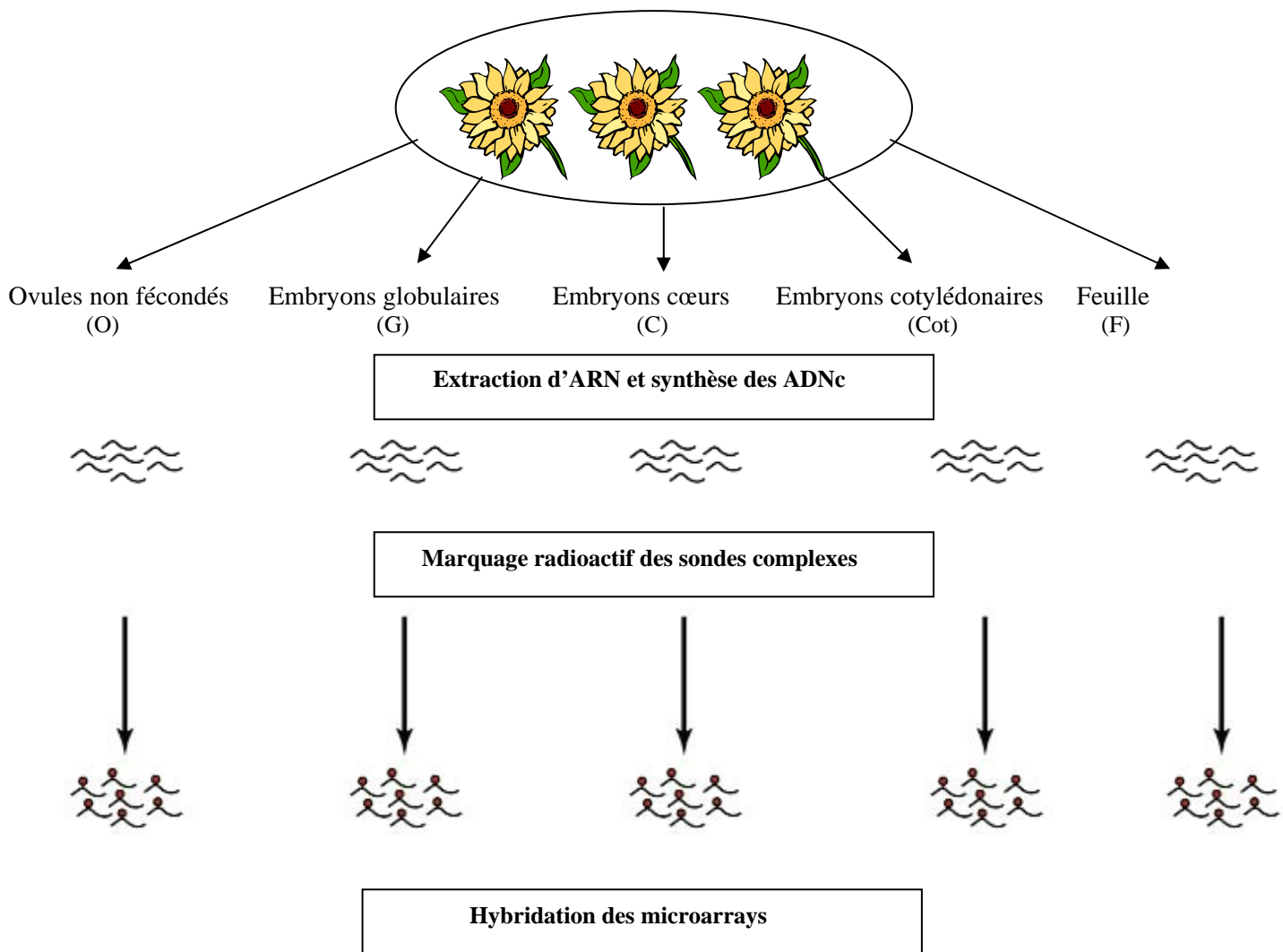
- des ovules non fécondés, témoignant de l'activité cellulaire caractéristique de tissus reproductifs non engagés dans un programme embryogène,
- et la feuille, représentant un témoin somatique photosynthétique.

Etant donné le manque de connaissances actuel concernant les gènes induits au cours des phases précoces de l'embryogenèse végétale, l'introduction comme témoin externe dans notre plan d'expérience de la feuille, qui constitue un tissu dont le profil transcriptionnel est bien connu et caractérisé par une forte expression des gènes liés à la production d'énergie par la photosynthèse, permettra de valider l'ensemble des résultats d'expression génique obtenus à l'issue de notre analyse statistique.

I.2.2. Importance des répétitions des mesures.

L'obtention de plusieurs mesures pour une même condition permet d'augmenter la confiance que l'on peut accorder aux conclusions déduites suite à l'analyse des résultats. Les répétitions peuvent être techniques ou biologiques. Les répétitions techniques consistent à utiliser, pour plusieurs hybridations, le même échantillon d'ADNc issu d'une seule et même extraction d'ARNm et permettent donc d'évaluer la variabilité liée à la manipulation (rendement du marquage de la sonde, conditions d'hybridation et de lavages des membranes...). En revanche, les répétitions biologiques, qui sont effectuées à partir de différentes extractions d'ARNm issues d'individus distincts, permettent de prendre en compte la variabilité existante au sein de la population d'individus. Toutefois, afin de limiter le nombre d'hybridations, plusieurs sondes complexes issues de différentes extraction d'ARNm pour une même condition peuvent être mélangées préalablement au marquage radioactif plutôt que de réaliser plusieurs hybridations biologiques. En règle générale, il est conseillé d'effectuer 3 à 4 répétitions pour chaque mesure et les répétitions biologiques sont préférables aux répétitions techniques (Churchill, 2002).

Dans notre cas, les difficultés rencontrées pour récolter le matériel végétal en quantité suffisante ne nous ont pas permis de réaliser de véritables répétitions biologiques. Toutefois, pour chaque stade de développement étudié, des embryons ont été prélevés sur différents individus afin de tenir compte de la variabilité introduite



	Membrane 1	Membrane 2	Membrane 3	Membrane 4	Membrane 5
Répétition 1	G	C	Cot	O	F
Répétition 2	C	G	O	F	Cot
Répétition 3	Cot	F	G	C	O
Répétition 4	O	Cot	F	G	C

Facteur d'efficacité $E = 0,94$ (par rapport au plan en blocs complet)

Figure 27 : Plan d'expérience utilisé pour l'hybridation des microarrays.

par l'échantillonnage des tissus. En revanche, pour chaque condition étudiée, 4 répétitions techniques ont été réalisées.

1.2.3. Carré Latin Incomplet : Intérêt et puissance du dispositif expérimental utilisé.

La capacité de réhybridation des microarrays sur membrane nylon autorise l'élaboration d'un plan d'expérience qui, lorsqu'il est adapté au problème biologique posé, permet d'augmenter la sensibilité de l'analyse en fixant des seuils de signification rendant les conclusions déduites des résultats analysés plus fiables.

Le plan d'expérience mis en œuvre pour cette étude, schématisé dans la figure 27 ci-contre, correspond à un Carré Latin Incomplet (ou Youden Square) dans lequel 5 blocs (ou membranes) ont chacun été hybridés avec 4 traitements différents parmi les 5 conditions étudiées. Chaque traitement a été hybridé 4 fois sur 4 puces différentes (soit 4 répétitions techniques). Ce dispositif expérimental comptant un total de 20 hybridations différentes à une efficacité de 0.94 comparé au plan en blocs complet correspondant impliquant 25 hybridations.

Parmi la large gamme de dispositifs expérimentaux disponibles, nous avons choisi un plan en blocs incomplets équilibrés du type Carré de Youden car, comme tout plan en blocs, il permet d'estimer la variabilité due à la réalisation des expériences mais, comme les Carrés Latins, outre l'hétérogénéité entre les différents blocs (ici, les membranes), il permet également de contrôler une seconde source de variation orthogonale aux traitements et aux blocs (ici, le cycle d'hybridation), et ceci avec un nombre réduit de répétitions. Or, dans ce type d'expériences, la variabilité introduite au cours des différentes répétitions techniques, qui peut notamment être due à des différences du rendement lors du marquage radioactif de la sonde, de l'efficacité d'hybridation et de lavage des membranes ou encore à des variations du temps d'exposition des membranes avant l'obtention des images de l'hybridation, est susceptible de constituer la source principale de variation.

I.3. L'analyse statistique des données est basée sur deux ANOVA interconnectées.

Le but ultime de l'analyse des données d'hybridation des microarrays est de proposer des gènes candidats susceptibles de jouer un rôle clé dans le processus analysé et pour lesquels il serait intéressant de mener une étude plus approfondie. Ce procédé d'analyse implique trois phases successives essentielles pour garantir la validité des conclusions émises:

- l'évaluation de la qualité des puces à ADN afin de s'assurer de la fiabilité des signaux obtenus,
- la normalisation des données permettant d'enlever la part de variation apportée par l'expérimentation et ne reflétant pas réellement une modification du niveau d'expression des gènes due à un changement de l'état physiologique du tissu étudié,

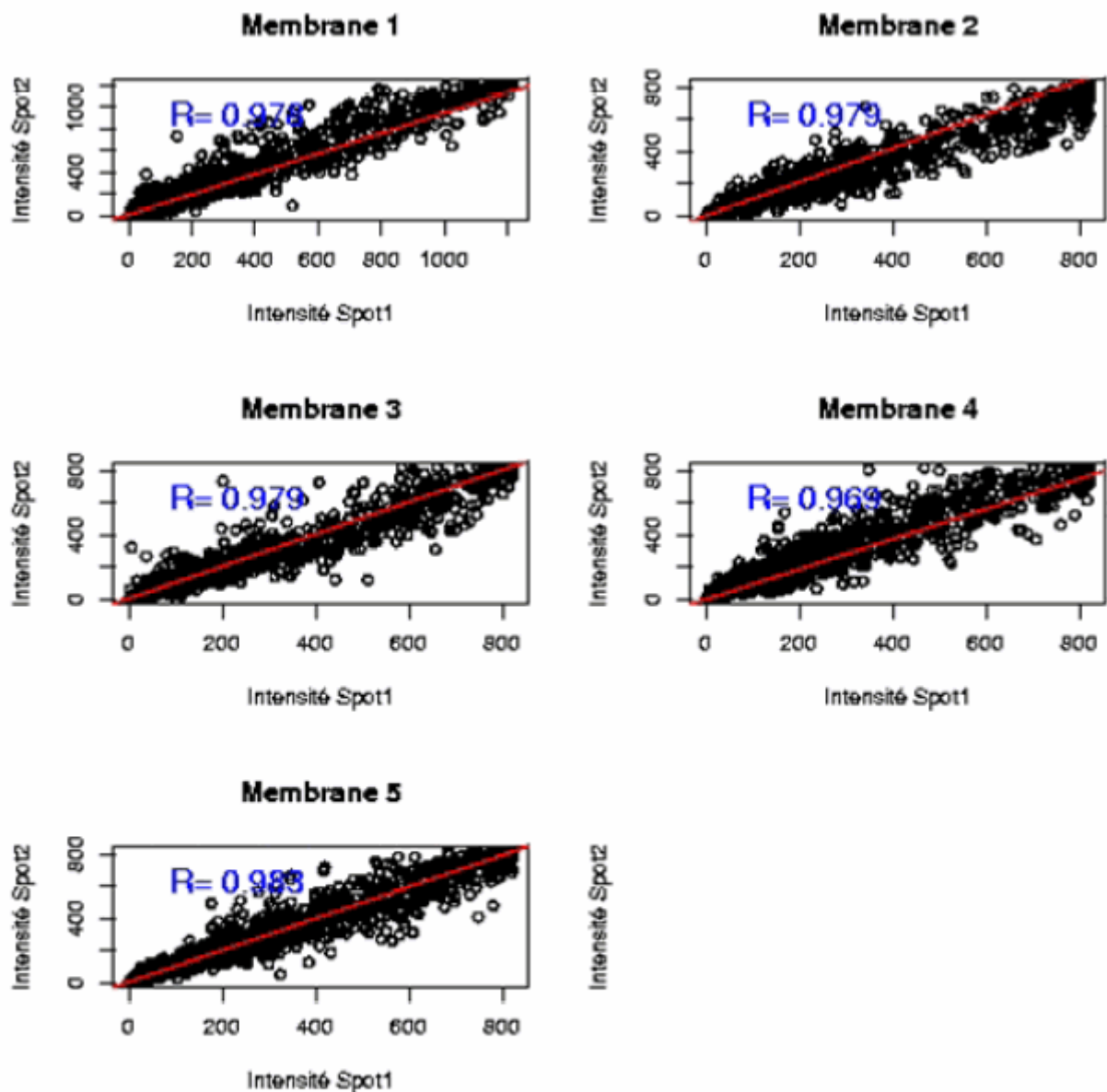


Figure 28: Analyse de la qualité des microarrays grâce à l'hybridation oligo-vecteurs.

Graphique montrant la corrélation entre les signaux d'hybridation (valeurs brutes) des deux spots de chaque clone pour l'hybridation vecteur de chaque membrane. La valeur R correspond au coefficient de corrélation.

- l'identification des gènes différentiellement exprimés et la caractérisation de leurs profils d'expression en fonction des différentes conditions testées.

1.3.1. Contrôle de la qualité des microarrays et de la fiabilité des résultats des hybridations.

Préalablement à l'exploitation des données d'hybridation des microarrays pour analyser le transcriptome dans les différentes conditions étudiées, il est indispensable de vérifier la qualité des résultats obtenus. Ce contrôle s'effectue à deux niveaux :

- tout d'abord, une hybridation oligos-vecteurs permet de valider la qualité générale du microarray produit, -ensuite, la comparaison des données des différentes répétitions techniques permet d'évaluer la reproductibilité des résultats.

1.3.1.1. Contrôle de la qualité des microarrays grâce à l' hybridation oligos-vecteurs.

L'hybridation oligos-vecteurs consiste à hybrider le microarray avec un mélange stoechiométrique d'amorces oligomères marquées radioactivement et spécifiques des séquences bordant le site de clonage des plasmides contenant les ADNc déposés sur la puce (M13R et Triplex3). Cette hybridation, réalisée pour les 5 membranes utilisées pour nos expériences, permet de vérifier l'architecture globale des microarrays (forme et densité des spots), l'homogénéité de la quantité déposée de produits PCR entre les différentes membranes et entre les différentes aiguilles de l'outil du robot ainsi que le rendement de PCR entre les différents types de vecteurs de clonage utilisés pour la construction des banques d'ADNc.

Les images des microarrays obtenues suite à l'hybridation oligos-vecteurs ont mis en évidence une augmentation très nette de la concentration des produits PCR déposés sur les champs situés sur les colonnes externes des membranes (8 champs au total) due à un dysfonctionnement du système d'humidification du robot au cours de la production des filtres. La concentration trop élevée de ces spots, entraînant une saturation du signal radioactif et ne permettant pas l'individualisation des spots, nous a contraint de retirer ces valeurs pour notre analyse statistique. Ainsi, afin de garantir la fiabilité des résultats, les analyses statistiques ont été réalisées à partir de 378 560 mesures de qualité élevée représentant 87.5% des 432 640 valeurs initiales.

Pour chaque membrane, la corrélation entre les signaux obtenus après l'hybridation oligos-vecteurs pour les deux spots de chacun des clones est très proche de 1 (Fig. 28).

1.3.1.2. Evaluation de la reproductibilité entre hybridations.

Une deuxième étape d'évaluation de la qualité des résultats consiste à contrôler la reproductibilité expérimentale entre les répliques techniques en comparant les différentes hybridations réalisées à partir d'un même tissu. Les hybridations d'un tissu particulier sur les différentes membranes n'étant pas réalisées

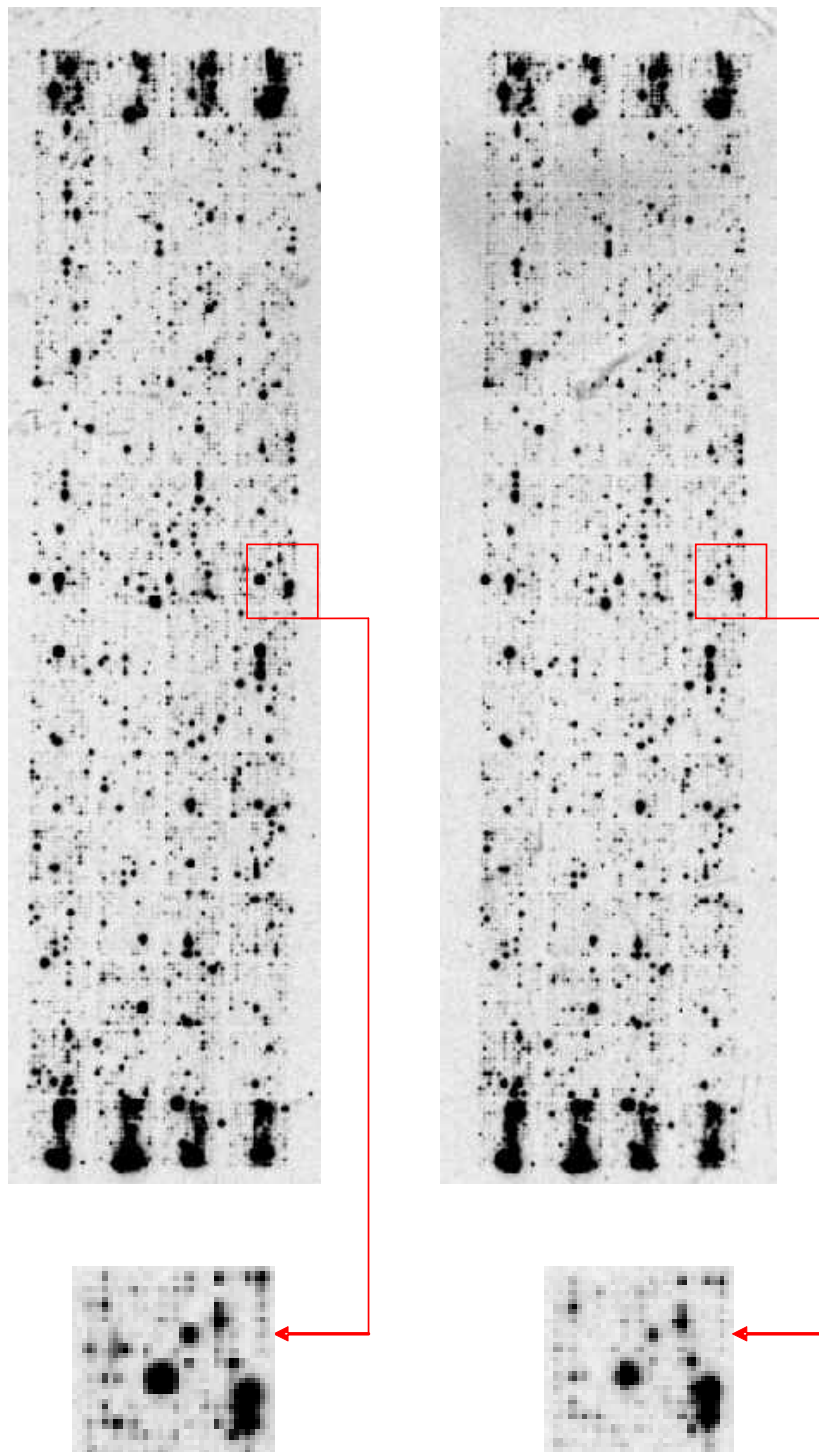


Figure 29: Deux membranes différentes hybridées avec l'ADNc d'embryons au stade cœur.

simultanément, une variation dans les données peut être introduite du fait de différences dans le rendement du marquage des sondes complexes, de différences dans les conditions d'hybridation et/ou de lavage des membranes ou dans le temps d'exposition avant révélation au phosphoimager. La figure 29, présentant deux membranes hybridées avec un même traitement, met en évidence la très grande similarité des deux profils d'hybridation. La figure 30 qui représente la matrice de dispersion pour les différentes répétitions techniques effectuées à partir des ovules non fécondés révèle une excellente corrélation entre les 4 répétitions effectuées sur des membranes indépendantes. Des données similaires ont été obtenues avec les autres traitements, les coefficients de corrélation pouvant atteindre 0.985. Ces résultats témoignent d'une forte reproductibilité des signaux d'intensité obtenus pour les différentes répétitions techniques.

1.3.2. La normalisation des données : Identification et retrait de la variabilité expérimentale.

Malgré toutes les précautions prises pour limiter les sources de variation dues à la réalisation technique des expériences, la variabilité expérimentale existe et a plusieurs origines : elle peut, par exemple, résulter de l'échantillonnage des tissus ou de la procédure de production des sondes complexes, être due à des variations incontrôlables des conditions d'hybridation ou être directement liée à la construction des microarrays (différentes quantités de produits PCR déposées selon les aiguilles...). Ainsi, afin de pouvoir comparer le niveau d'expression des gènes obtenus suite à plusieurs expériences successives, une normalisation préalable des données est indispensable. Plusieurs méthodes permettant d'évaluer la variabilité expérimentale ont été développées (Leung and Cavalieri, 2003 ; Quackenbush, 2001).

1.3.2.1. Normalisation à l'aide de gènes de ménage.

Les gènes de ménage (ou housekeeping genes) sont des gènes dont l'expression est supposée constante dans tous les types cellulaires et conditions physiologiques. La comparaison des signaux d'intensité obtenus pour ces gènes entre les différentes hybridations permet de corriger la variation expérimentale. Toutefois, cette approche requiert un grand nombre de gènes de ce type, présentant un niveau d'expression constant, d'intensité moyenne à élevée, parmi une large gamme de conditions. Or, de tels gènes sont difficiles à sélectionner.

1.3.2.2. Normalisation à l'aide de témoins externes (spikes).

Cette approche utilise des gènes témoins exogènes (provenant d'un autre organisme) déposés sur le microarray et dont différentes concentrations connues d'ADNc sont ajoutées à la sonde complexe préalablement à l'hybridation. La normalisation des données s'effectue en comparant les signaux obtenus pour ces gènes pour les différentes hybridations. Cependant, cette technique souffre d'imprécisions : l'ajout des ADNc dans la sonde complexe à une concentration précise est très difficile, des infimes variations de la

Matrice de dispersion des répétitions techniques OVULE

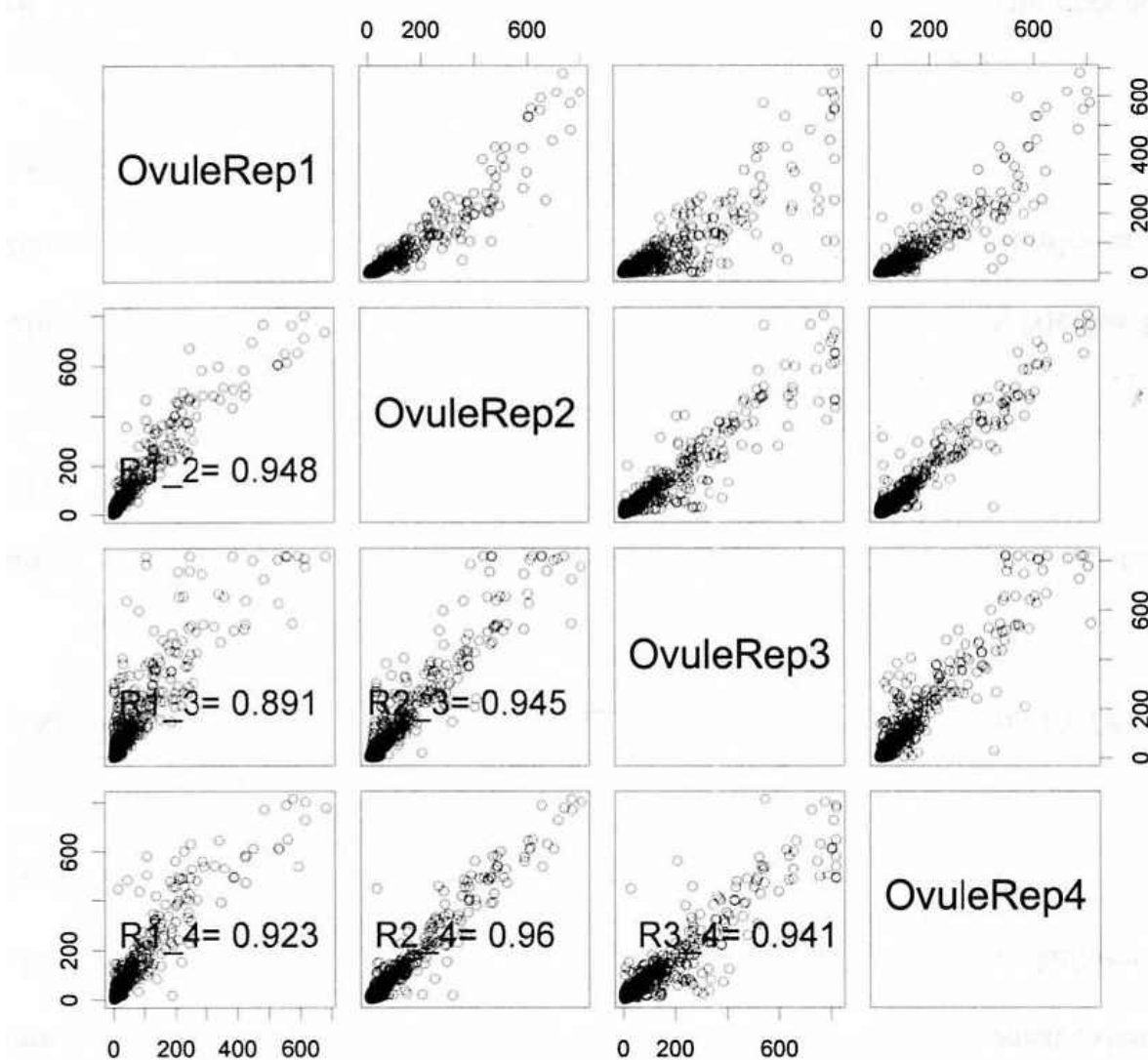


Figure 30: Evaluation de la reproductibilité des hybridations.

Exemple des 4 répétitions réalisées à partir des ovules non fécondés. La valeur R représente le coefficient de corrélation.

quantité entraînant des erreurs systématiques. De plus, elle ne tient pas compte de toutes les sources de variation expérimentale comme notamment la quantité de sonde complexe utilisée.

1.3.2.3. Normalisation globale à l'aide de tous les gènes présents sur la puce.

La normalisation globale tenant compte de tous les gènes présents sur la puce s'appuie sur l'hypothèse que la majorité des gènes présents sur la puce ne présentent pas de variation du niveau d'expression entre les conditions testées. Cette normalisation peut être réalisée par analyse de variance (ANOVA), qui est une technique facile à mettre en œuvre, ou par le calcul du Z score, où le signal obtenu pour chaque hybridation est centré et réduit. L'ANOVA qui, contrairement au Z score, permet d'évaluer l'ensemble des différents effets, est une approche plus fiable.

Toutefois, cette approche de normalisation globale peut ne pas être appropriée dans le cas d'analyses effectuées à partir de données issues d'une puce thématique, où les gènes testés sur le microarray sont directement liés au phénomène biologique étudié, pour lesquelles la sous ou sur-expression globale des gènes pourraient brouiller la variabilité existante.

1.3.2.4. Normalisation par régression et normalisation non linéaire.

Dans le cas d'échantillons très proches, la majorité des gènes devraient être exprimés à des niveaux similaires. La normalisation par régression consiste à ajuster les valeurs telles que la pente de la représentation graphique par scatterplot de l'intensité des gènes dans les deux conditions soit égale à 1. Cependant, l'intensité des spots n'est souvent pas linéaire, elle peut dépendre notamment de la localisation des gènes sur le microarray. Dans ces cas, des méthodes de régression locale non linéaire telles que la technique LOWESS (LOcally WEighted Scatterplot Smoothing) sont plus appropriées.

1.3.3. Identification des gènes différentiellement exprimés.

Plusieurs méthodes d'identification des gènes différentiellement exprimés plus ou moins sensibles et fiables sont aujourd'hui disponibles (Leung and Cavalieri, 2003, Murphy, 2002).

1.3.3.1. Définition d'un fold change, une méthode peu spécifique et peu sensible.

Au cours des premières analyses microarrays réalisées, un gène était déclaré différentiellement exprimé entre deux conditions si le ratio de l'intensité mesurée dans les deux conditions était supérieur à un certain seuil fixé arbitrairement (habituellement de 2). Cette approche, qui ne s'appuie sur aucun test statistique pour valider la signification de la différence d'expression, n'est pas très rigoureuse, une variation d'un facteur 2 n'ayant pas la même signification selon le niveau d'expression. Basée sur un seuil n'ayant aucune

légitimité biologique, elle est à l'origine de faux positifs, notamment pour les gènes faiblement exprimés, et de faux négatifs, en particulier pour les gènes fortement exprimés.

1.3.3.2. Visualisation de ratios inhabituels, une méthode très utilisée mais biaisée.

Cette méthode considère un gène comme différentiellement exprimé si le ratio des valeurs pour ce gène dans les deux conditions testées s'éloigne du ratio des moyennes des deux conditions d'un certain facteur fixé arbitrairement (généralement de ± 2 écarts types). Cependant, comme la technique se basant sur le fold change, cette approche génère un fort taux de faux positifs.

1.3.3.3. Analyse de variance.

Lorsque les hybridations des microarrays obéissent à un plan d'expérience pré-établi incluant des répétitions des mesures, l'identification des gènes différentiellement exprimés peut se faire par une analyse de variance, cette approche ayant l'avantage de sélectionner les gènes selon la signification de la variation du niveau d'expression entre les deux conditions et non selon la différence entre les intensités absolues du signal obtenues.

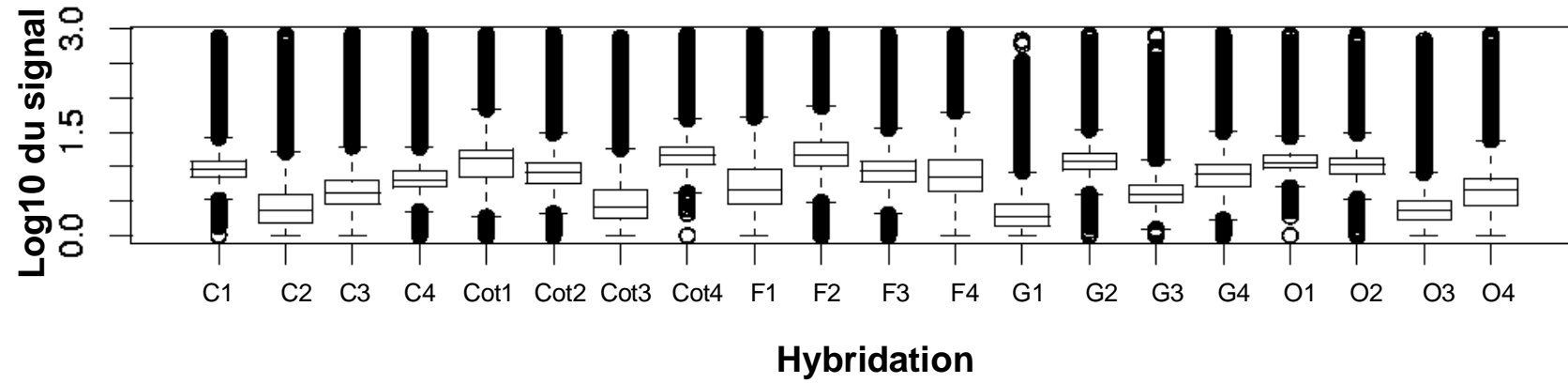
1.3.4. Méthode d'analyse statistique choisie pour notre étude.

Au vu des informations rapportées ci-dessus sur les différentes méthodes d'analyse microarrays disponibles, nous avons choisi d'utiliser pour l'analyse statistique de nos résultats la méthode décrite par Wolfinger (Wolfinger et al., 2001) qui s'appuie sur deux ANOVA interconnectées : l'ANOVA de normalisation, portant sur l'ensemble des données, suivie d'une ANOVA individuelle pour chacun des gènes réalisée sur les résidus de normalisation. Le fait d'effectuer une ANOVA pour chaque gène permet de satisfaire les postulats essentiels à ce type d'analyse qui sont l'indépendance des données et l'homogénéité des variances.

La première analyse de variance, réalisée à partir des log₁₀ des signaux mesurés, permet d'éliminer la part de variation imputée à l'expérimentation (différence due à la richesse de chaque tissu, procédure de marquage des sondes, efficacité d'hybridation, durée des lavages et de l'exposition). La transformation en log₁₀ des valeurs mesurées permet de diminuer l'effet d'échelle entre les signaux les plus faibles et les plus élevés et de faciliter l'étape de normalisation des données.

La technique de normalisation utilisée est une technique globale basée sur l'expression de tous les gènes présents sur la puce : contrairement aux méthodes de normalisation s'appuyant sur des gènes de ménage ou sur des témoins exogènes localisés à des endroits précis du microarray, elle permet de tenir compte des effets locaux ou dépendant du niveau d'intensité.

(A) rough expression levels



(B) experiment-wise normalized expression levels

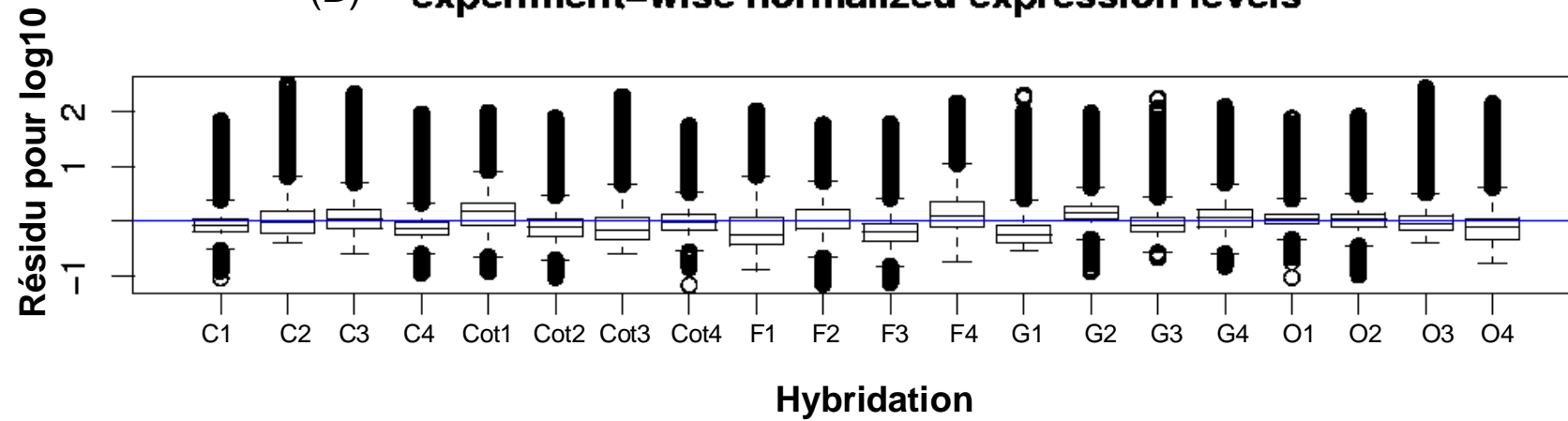


Figure 31: Répartition des signaux pour chaque hybridation avant (A) et après normalisation (B).

Dans notre cas, chaque signal peut être décomposé mathématiquement selon le modèle suivant :

$$y_{ghijk} = \mu + T_h + M_i + R_j + r_{ghijk}$$

où :

y_{ghijk} est l'intensité du gène g dans le traitement h , dans la membrane i pour la répétition technique j et pour le spot k ,

μ correspond à la moyenne globale,

T_h , M_i et R_j représentent respectivement les effets principaux du traitement, de la membrane et de la répétition technique,

et r_{ghijk} est un résidu normalisé par rapport aux effets principaux testés.

La figure 31 qui représente, pour chacune des 20 hybridations, les boîtes à moustaches des données brutes avant normalisation et des résidus obtenus après normalisation, montre que, suite à la normalisation, la moyenne des signaux est proche de zéro, ce qui prouve que la méthode choisie pour la normalisation est efficace.

Les résidus, qui correspondent aux valeurs normalisées pour les facteurs considérés, représentent le niveau d'expression de chaque gène dans chaque condition et constituent le jeu de données pour la seconde ANOVA effectuée pour chacun des gènes.

Selon ce second modèle d'ANOVA, les résidus de la normalisation peuvent être décomposés en :

$$r_{ghijk} = G_g + (GT)_{gh} + e_{ghijk}$$

où:

r_{ghijk} est la valeur normalisée du gène g dans le traitement h , dans la membrane i pour la répétition technique j et pour le spot k ,

G_g représente l'effet principal du $g^{\text{ième}}$ gène,

$(GT)_{gh}$ constitue l'effet d'interaction entre le gène g et le traitement h ,

e_{ghijk} correspond à l'erreur stochastique.

Grâce à l'estimation de l'interaction entre un gène et les traitements, cette ANOVA permet d'identifier les gènes dont la variation du niveau d'expression peut réellement être attribuée à la nature du tissu. Un seuil de signification (ou p-value) est défini selon la correction de Bonferroni : tous les gènes dont la p-value calculée est supérieure au seuil fixé sont déclarés différentiels. La correction de Bonferroni peut sembler trop stringente et limiter la puissance d'identification de gènes significativement différentiels. Toutefois, l'un des buts de l'analyse statistique des données microarrays étant de révéler des candidats potentiels pour des études plus poussées, il nous est donc apparu essentiel de réduire au maximum le risque de sélectionner des faux positifs (gènes déclarés différentiellement exprimés alors que non différentiels) tout en ayant conscience du risque d'omettre certains gènes réellement différentiels.

1.3.5. Classification des gènes différentiels en fonction des profils d'expression.

Si l'analyse statistique du niveau d'expression des gènes entre différents traitements suffit à identifier des gènes potentiellement intéressants en révélant leur sur ou sous-expression dans chacune des conditions étudiées, l'analyse des gènes selon leur profil d'expression pour l'ensemble des tissus testés fournit une information sur la fonction cellulaire dans laquelle ils sont potentiellement impliqués et permet donc d'évaluer la pertinence de ces gènes candidats. Les analyses par clustering, qui sont des techniques de classification visant à regrouper les gènes présentant des comportements similaires, permettent, lorsqu'elles sont couplées à une représentation graphique des résultats d'assemblage, de visualiser très clairement les profils d'expression des gènes dans le temps, dans l'espace ou dans différents tissus. Ces techniques s'appuient sur le postulat que des gènes contrôlant un processus biologique particulier doivent être co-régulés et doivent donc montrer des profils d'expression identiques. Dans la plupart des cas, ces méthodes ne nécessitent pas une connaissance préalable du processus sous-jacent (méthodes d'assemblage non supervisées).

Plusieurs techniques de clustering répondant à différents algorithmes et appliquant différentes méthodes de calcul de distance métrique qui permettent d'évaluer le niveau de similarité entre les profils d'expression des gènes aboutissent à différentes associations des gènes entre eux et ainsi à différentes représentations (Butte, 2002 ; Leung and Cavalieri, 2003 ; Quackenbush, 2001). Il convient donc d'élire l'algorithme et la méthode de calcul les plus adéquats pour classer nos données.

1.3.5.1. Méthodes de mesures de similarité.

Il existe diverses méthodes de calcul de distances permettant d'évaluer le degré de similarité entre les gènes et les conditions testés au cours d'analyses microarrays, les plus répandues étant la distance euclidienne et la corrélation de Pearson.

- **Calcul de la distance euclidienne.**

La distance euclidienne est une généralisation du théorème de Pythagore à un espace à n dimensions où chaque gène est considéré comme un point, chaque axe constitue un échantillon biologique et les coordonnées sur chaque axe correspondent aux niveaux d'expression du gène pour l'échantillon considéré. Cette méthode est sensible aux niveaux absolus d'expression. L'inconvénient principal de cette méthode de calcul est qu'elle ne permet pas de révéler les gènes corrélés négativement.

- **Coefficient de corrélation de Pearson.**

Le coefficient de corrélation de Pearson est calculé à partir des distances de chaque point à la droite de régression linéaire. Cette technique de calcul est sensible à l'allure générale du profil d'expression et suppose que les gènes interagissent selon un modèle linéaire.

I.3.5.2. Méthodes de clustering.

Seules les méthodes de clustering non supervisées, qui ne nécessitent aucune connaissance biologique préalable concernant les gènes censés être liés fonctionnellement et qui s'appliquent donc à notre cas de figure, seront décrites dans ce paragraphe. Celles-ci sont de deux sortes hiérarchiques ou non hiérarchiques.

- **Clustering hiérarchique.**

Le clustering hiérarchique est une approche agglomérative au cours de laquelle les profils d'expression de chacun des gènes sont regroupés selon leur similitude de manière itérative jusqu'à obtenir un arbre unique représentant l'ensemble des gènes testés. Cette technique de classification, caractérisée par sa simplicité et la clarté de la visualisation des résultats sous forme de dendrogramme qu'elle fournit, compte parmi les plus utilisées. Toutefois, elle présente quelques limites. Un mauvais adressage d'un gène au début du processus ne peut pas être corrigé ultérieurement. De même, avec l'accroissement d'un cluster, il se peut que le vecteur d'expression représentant le cluster ne soit plus réellement représentatif d'aucun des gènes qu'il contient. De plus, cette méthode de clustering ne tient pas compte des associations négatives.

- **Méthodes de clustering non hiérarchique.**

Les méthodes de clustering non hiérarchiques telles que le clustering par k-means ou par self-organising map (SOM) constituent des approches divisives formant un réseau dans lequel les gènes sont associés à des groupes d'expression prédéfinis, par partitions successives, en fonction de la similarité de leur profil d'expression avec le profil d'expression de référence défini pour chaque groupe. La technique de détermination des vecteurs d'expression de référence diffère entre le clustering par SOM et par k-means. Si les avantages de cette procédure sont l'absence de hiérarchie et la délimitation nette des groupes d'expression, un inconvénient évident est la nécessité de déterminer au préalable le nombre de partitions, l'utilisation d'autres techniques telles que l'ACP pouvant aider à l'estimation du nombre de clusters représentant au mieux les données. Cependant, tout comme pour les dendrogrammes, les associations négatives sont difficiles à mettre en évidence par ces méthodes d'assemblage.

I.4. Validation des gènes différentiels par RT-PCR quantitative, une méthode rapide et très sensible.

Diverses techniques peuvent être utilisées pour valider les résultats des expériences de microarrays parmi lesquelles l'hybridation *in situ*, le northern blot ou encore la RT-PCR quantitative. Du fait de la difficulté rencontrée pour extraire le matériel biologique en grande quantité, nous avons choisi d'utiliser la RT-PCR quantitative qui a l'avantage d'être très sensible, spécifique et rapide. Cette technique repose sur la détection de fluorescence émise par le produit PCR néoformé et mesurée en phase exponentielle de la réaction de

PCR. Parmi les deux chimies disponibles pour réaliser la RT-PCR en temps réel (sondes TaqMan ou Sybr Green), nous avons opté pour le Sybr Green, qui est un produit intercalant se liant au niveau du petit sillon de l'ADN double brin. Ainsi, plus la quantité d'amplicons double brin produits augmente, plus le signal fluorescent est élevé. Bien que le Sybr Green, qui se lie à tout fragment d'ADN double brin soit moins spécifique que les sondes TaqMan, qui constituent des primers désignés spécifiquement pour une séquence pré-définie, ce procédé reste le plus abordable économiquement et la non-spécificité de la réaction peut être évaluée.

Deux méthodes d'analyse des résultats de RT-PCR quantitative peuvent être utilisées :

- la quantification absolue qui détermine le nombre absolu de copies du transcrit d'intérêt en comparant son signal PCR à une courbe standard,
- et la quantification relative qui évalue le changement du niveau d'expression du gène d'intérêt par rapport à une référence.

Le changement relatif du niveau d'expression des gènes peut être quantifié selon la méthode du $2^{-\Delta\Delta CT}$ (Livak and Schmittgen, 2001) basée sur la détermination d'un seuil C_T (Threshold Cycle) correspondant au nombre de cycles de PCR pour lequel le signal d'amplification est supérieur au bruit de fond alors que la courbe d'amplification est toujours en phase exponentielle. La comparaison entre le C_T du gène d'intérêt par rapport au C_T du gène témoin (gène dont l'expression est constitutive parmi l'ensemble des conditions testées, dans notre cas, il s'agit de l'actine) permet d'évaluer le niveau d'expression du gène dans les différentes conditions.

II. Vers l'élaboration d'un modèle fonctionnel de l'embryogenèse précoce chez le tournesol.

Article - Towards a functional analysis of early embryogenesis in exalbuminous dicotyledonous: Comparison of transcriptional profiles of developing sunflower embryos using a 10K cDNA microarray.

Towards a functional analysis of early embryogenesis in exalbuminous dicotyledonous: Comparison of transcriptional profiles of developing sunflower embryos using a 10K cDNA microarray.

Keywords: *Helianthus annuus* L., embryogenesis, global gene expression, cDNA microarrays, SMART technology.

Corresponding author: Laurent Gentzbittel*

Laboratoire de Biotechnologie et Amélioration des Plantes, Ecole Nationale Supérieure Agronomique de Toulouse, Avenue de l'Agrobiopole, BP 107, Auzeville Tolosane, 31326 Castanet Tolosan, France

* Correspondance : Tel : +33-562-193596
 Fax: +33-562-193587
 E-mail: gentz@ensat.fr

Abstract

To identify key genes involved in critical phases of sunflower seed development, we compared gene expression patterns of embryos at globular, heart-shaped and cotyledonary stages, using a microarray approach. Leaves and unfertilized ovules, corresponding respectively to a photosynthetic somatic tissue and to a non-embryonic reproductive tissue, were employed as external controls. Statistical analysis, performed using two-step ANOVA models, revealed that 744 cDNA clones can be considered as differentially expressed between at least two of the five physiological conditions under study.

Functional characterization of the differentially expressed genes pointed out major changes in metabolic activity levels and cellular processes involved during embryo development. For all stages of embryo development investigated, a large fraction of differentially expressed genes appears to be involved in transcriptional and translational processes. Heart-shaped and cotyledonary stages are also characterized by an important cell-to-cell communication as well as high metabolic and energetic activities. The major changes in expression patterns of genes implicated in energetic activities raise the question of the transition to autotrophic development as early as heart stage. The expression of a large number of genes involved in lipids and amino acids metabolism is likely to indicate that storage reserves accumulation is occurring as soon as the cotyledonary stage. Genes showing strong similarities with genes playing key roles in plant or animal embryogenesis such as KEULE or Ras- related protein RAB11, were also identified among genes highly expressed at diverse embryo developmental stages. Surprisingly, among the reproductive tissues under study, cotyledonary embryos and unfertilized ovules appear to induce the most similar transcriptional programs.

The gene expression profiles derived from this study can provide additional insights into physiological processes that are active during the early stage of embryo development in non-green developing-seeds.

Introduction

Seeds constitute the major nutritive resource for humanity. In exalbuminous plants (plants without endosperm) such as sunflower, pea or soybean, protein and lipid reserves are directly accumulated into the embryo proper. Thus, nutritional value of the seed entirely depends on correct embryo development. Embryogenesis also represents a key step of plant life cycle during which embryo develops the main structures of the future individual according to a well-defined morphological pattern. In dicotyledonous species, the embryo, generated after the double fertilization, a characteristic process of Angiosperms (flowering plants), follows a relatively invariant sequence of divisions going through successive well-identified stages corresponding to crucial events of the development (Jürgens, 2001, 2003; Laux et al., 2004). During the globular stage, the two main axes of development are established: the apico-basal axis, along which functional embryonic organs (including shoot apical meristem, cotyledons, hypocotyl, radicle and root meristem) will later get organized and the radial axis, around which different tissues (epidermis, cortex, endoderm, pericycle and vascular tissues) will set up in a concentric manner. Then, the embryo undergoes cellular differentiation and acquires a bilateral symmetry to reach the heart-shaped stage. At this stage, the embryo exhibits its final body plan: all the embryonic tissues and structures are completely formed. The ultimate stage of the embryogenesis is the cotyledonary stage which corresponds to an active phase of cell division and elongation and to the accumulation of reserves essential for the subsequent germination and the normal development of the future seedling.

In the past decades, plant embryogenesis has been the object of intense histological, biochemical, genetic and molecular studies, notably in the *Arabidopsis thaliana* plant model. Many genes involved in diverse processes of embryo development have already been characterised through a wide range of techniques such as differential screening approaches (Heck et al., 1995; Li and Thomas, 1998), promoter /enhancer trapping (Topping et al., 1994) or powerful mutational screens (Mayer et al., 1991; Meinke and Sussex, 1979). Recently, very sensitive techniques adapted to the study of transcriptional activities of scarce biological tissues and allowing genome-wide gene expression analysis, such as *in silico* expressed sequence tags (EST) analysis, SAGE (Serial Analysis of Gene Expression) or microarrays technology, have been developed and gave the opportunity to obtain global gene expression patterns during seed development in *Arabidopsis thaliana* and a few other well-characterized plant species (Breton et al., 1995; Girke et al., 2000; Lee et al., 2002; Sreenivasulu et al. 2004; Zhu et al., 2003). The first EST data analysis from *Arabidopsis thaliana* developing seeds was performed on more than 10 500 sequences obtained from seeds harvested between 5 and 10 after flowering and allowed to identify numerous new genes, potentially seed-specific, and provided the first functional profiles for exalbuminous dicotyledonous seed characterized by very active primary metabolism and nucleic acids and proteins synthesis pathways and low proportion of transcription factors and proteins involved in signal transduction (White et al., 2000). These EST were further used to construct a

cDNA microarray notably employed to describe transcriptional kinetics of genes involved in proteins and lipids reserves accumulation in seed, giving rise to the first transcriptional regulatory model controlling the carbon resources allocation among the different types of reserves during embryogenesis (Ruuska et al., 2002). Recently, microarray analysis studying earlier stages of *Arabidopsis thaliana* reproductive development (Hennig et al., 2004) confirmed that transcriptional modifications during this phase of plant development primarily involved primary metabolism as well as transcription and identified YABBY, MADS box and MYB transcription factors families as playing putatively key roles in reproduction process. Gene expression studies in *Arabidopsis* and in a small number of well studied plants have thus significantly improved our knowledge about the major cellular functions involved during seed formation in exalbuminous dicotyledonous and revealed some putative key genes of plant embryogenesis.

Compared to other dicotyledonous plants, sunflower (*Helianthus annuus* L.), which is one of the main oleaginous culture, presents numerous advantages for the study of early stages of zygotic embryogenesis (Ben et al., 2005). First, its capitulum harbours a large quantity of relatively big embryos with synchronised development, making easier the identification and the isolation of embryos at diverse stages of development and allowing the simultaneous analysis of the effect of a stress over the range of all embryo development stages. Several genetic tools providing a good knowledge of its large genome are now available such as genetic linkage maps covering almost the entire genome (Tang et al., 2002; Al Chaarani et al., 2004), BAC libraries (Gentzittel et al., 2002) and a large EST collection. A somatic embryogenesis system was also proposed as a model to study embryonic polarity acquisition (Tamborindéguy et al., 2004). Lastly, compared to the dicotyledonous plant model *Arabidopsis thaliana* which contains ‘green seeds’, sunflower, such as other crop species of major interest (e.g. cotton), holds ‘non green seeds’, whose formation might involved particular cellular processes, and has undergone domestication steps, whose consequences cannot be evaluated in *Arabidopsis thaliana* which is a weed.

In a first step towards the understanding of the mechanisms involved in early sunflower embryogenesis, an *in silico* analysis was performed on more than 7 100 EST from cDNA libraries prepared from embryos at globular, heart-shaped and cotyledonary stages as well as unfertilized ovules (Ben et al., 2005). This study, which gave a first global overview of the cell functions and activities involved successively in sunflower embryo development, revealed that differentiation-related functions such as protein metabolic pathways and transcription were strongly activated in all stages rather than cell multiplication-related functions. It also pointed out several putatively interesting sequences such as potential differentially expressed genes among the embryonic stages as revealed by digital northern or sequences showing similarities with genes playing key roles in animal or plant embryogenesis.

In order to characterize further the molecular mechanisms during the early stages of sunflower embryogenesis, a nylon microarray containing 8 185 unique sequences, obtained from a wide range of sunflower tissues, was constructed and used to investigate the gene expression patterns of the three main

stages of embryo development. cDNA from unfertilized ovules and leaves were employed respectively as external controls for photosynthetic somatic tissue and non-embryonic reproductive tissue. Two-way analysis of variance was used to identify the differentially expressed genes. Assignment of putative functions to the differentially expressed genes allowed us to reveal the major cellular functions occurring during the early stages of sunflower embryogenesis.

Results

Construction of a cDNA microarray representing 8 185 sunflower unique sequences

A total of 25 842 EST, generated from cDNA libraries prepared from different sunflower tissues such as leaves, apex, shoot as well as unfertilized ovules, embryos at diverse stages of development, protoplasts cultivated in embryogenic and non embryogenic conditions or plants subjected to various pathogen stress conditions were clustered into 8 185 contigs. For 1 219 contigs, two non-overlapped EST at both 5' and 3' ends were selected. For the remaining 6 966 contigs, only one EST at the 3' end was selected from each. Thus, a total of 9 404 sequences corresponding to 8 185 potentially unique sequences were selected. In order to perform quality assessment of the microarray experiments, different types of controls were employed: on one hand, potentially constitutive genes such as calmoduline, ubiquitine and β -tubuline were integrated as positive controls to check the integrity of the cDNA complex probe used for hybridization; on the other hand, negative controls including water, animal specific sequences such as desmine or EST from pig as well as bacterial specific genes such as NPTII or GUS were designed to test for contaminations and to define the background level. Finally, the 9404 sunflower cDNAs clones as well as the positive and negative controls were spotted onto nylon membranes and resulted in a total of 10,368 data points.

Rehybridization capacities of the nylon microarrays allow for adopting powerful experimental design

Contrary to experiments carried on glass-slide microarrays that cannot be re-used, we could benefit of the rehybridization abilities of the microarrays spotted on nylon membrane to set up a randomized balanced bloc design. This design allows estimating the variability due to experimental procedures and provides results based on formal statistical significance thresholds, making the statistical analysis easier and more reliable. Our experimental design is equivalent to an incomplete Latin Square (or Youden Square) in which each of the five tissues under investigation (globular, heart-shaped and cotyledonary embryos, unfertilized ovules and leaves) are challenged four times on four different nylon microarrays (slides). A total of five slides was used for the experiment, each slide being hybridized four times with four different tissues (Table 1). For each tissue, cDNA were synthesized from a single one RNA extraction performed on embryos or ovules harvested on various sunflower capitulae of the same genotype. Thus, even if the experimental design

involves technical repetitions, the biological variability between different individuals is taken into account during the plant samples collection.

Quality assay of microarray and hybridizations reveal good experimental reproducibility

To ensure the experimental reproducibility and the validity of our data, we first hybridized each slide with radio-labeled oligonucleotides to estimate the quality of the nylon microarrays. This hybridization, performed with a stoichiometric mix of primers specific to sequences flanking the cloning site of the spotted cDNA insert clones (M13R and Triplex3), allowed us to check the global architecture of the microarray (shape and density of the spots) as well as the homogeneity of the quantity of the spotted PCR products between the different slides and spotting tool pins. A second quality assessment step consists in evaluating the experimental reproducibility between repeats by comparing the different hybridizations realized from the same tissue. As hybridizations of a particular tissue on the different slides are not realized simultaneously, variability in the raw data could thus be introduced due to differences in the yield of complex probes radio-labeling, in hybridization or washing conditions, or in exposure time.

After quality trimming, we decided to keep for further statistical analysis 378, 560 high quality measured signals out of the 432, 640 initial intensity values (87.5%) to guarantee the reliability of our results. The figure 2, representing the pairwise scatterplots obtained from the selected signals, reveals excellent correlations between the four repeats on independent slides for each tissue with R-squared ranging from 0.693 to 0.985. This observation reflects the high reproducibility of measurements of gene expression levels between replicates.

Identification of differentially expressed genes using analysis of variance method

To identify differentially expressed genes, we applied a statistical method using two interconnected analysis of variance, respectively named the ‘normalization’ and the ‘gene’ models (Wolfinger et al., 2001). The first ANOVA model eliminates, from the measured signal, the undesirable variation due to systematic experimental effects which does not reflect a modification of the gene expression due to difference between the developmental stages. The normalization model, which is realized on log₁₀-transformed values for all the 378, 560 signal intensity measures (including clones and controls), revealed significant effects of the tissue factor, membrane factor and hybridization factor, illustrating the necessity of the normalization step in the analysis of microarray data (Table 2).

Residuals of this model, which represent the normalized expression values, were used as input data for the gene model which allows us to identify genes for which the level of expression varies depending on the stage of development. Each clone was spotted in duplicate onto the array and four hybridizations were realized for every sample, thus yielding 8 measurements for each combination of one clone and tissue. A clone was considered expressed in a tissue if, among all the hybridization measures, at least one spot corresponding to this clone showed normalized expression values greater than the maximum normalized expression value of the negative controls. Following this rule, 14.2% of the clones were identified as being

expressed in at least one of the five tested conditions. The Bonferroni correction, used to deal with multiple testing issues (9 464 tests), generates a very conservative experiment-wise significance value of $5.42 \cdot 10^{-6}$, resulting in a very low false-positive acceptance rate. According to this p-value threshold, 744 clones corresponding to 7.9% of the analysed clones were found to be differentially expressed among at least two studied conditions. To determine more precisely the origin of variation, the 4 degrees of freedom (df) of the ‘Gene-by-Tissue’ interaction factor were further split into four linear contrasts aiming at testing the difference between ‘Leaves’ versus ‘Reproductive Tissues’ (including embryos and unfertilized ovules), ‘Unfertilized Ovules’ versus ‘Embryos’ (every embryonic stages confounded), ‘Globular Embryos’ versus ‘Heart-shaped Embryos’ and ‘Early Embryos’ (i.e. globular and heart-shaped embryos) versus ‘Cotyledonary Embryos’. For every contrast under study, the results concerning differentially expressed genes are summarized in Table 3 and 4 and illustrated as Volcano plots of significance against magnitude of effects in Figure 2. As expected, the major part of the differential expression of the analysed clones can be assigned to the variability existing between the genetic programs of leaves versus reproductive tissues (58.8% of the set of differential clones) whereas 15.6% of the global differential genes show differences between unfertilized ovules and embryos and 21.4% and 25.1% of the genes are differentially expressed when comparing globular *versus* heart-shaped embryos and early embryos *versus* cotyledonary embryos.

Cluster analysis reveals the temporally regulation of transcriptional activity during sunflower embryogenesis

Clustering methods, which aims at assembling genes showing similar expression patterns, lead to the identification of genes that are co-regulated or implied in a common biological process. These techniques could also help to assign a putative role to sequences with unknown functions, based on the notion of ‘guilt by association’.

A hierarchical cluster analysis, using the average linkage method and the Pearson’s coefficient correlation, was performed on the normalized expression values of the 744 differentially expressed genes. Five major groups of genes have been identified (Fig.3). This graphical representation offers a very clear visualization of the dynamics of gene expression during sunflower embryogenesis and reveals a continuous wave of genes induction and repression with, on one hand, genes highly expressed during very early stages of development and slightly expressed later (cluster 2), or in contrast, genes showing very low expression levels during early phases of embryo development and further switched on during middle (cluster 3) or late stages of development (cluster 5). The putative functions of the sequences assembled in each cluster, performed with BLASTX program to search for homologies with known sequences, allowed us to assign the main cellular functions associated with each expression profile. Genes over-expressed at very early stages of embryo development and assembled in cluster 2 are mostly related to DNA structure, transcription, translation or signal transduction. In contrast, genes with high transcription activity during the later stage of embryogenesis (clusters 3 and 5) were found to encode products with predicted function related to

translation, metabolic and energetic pathways and intercellular communication, besides high rates of genes involved in transcriptional and translational processes.

Cluster analysis also associated the different tissues based on the similarity of their gene expression profiles (Fig.3). As expected, leaf, which was introduced in our analysis as a somatic external control and which constitutes a very specialized tissue mainly dedicated to photosynthesis, seems to show a very different genetic program from the other four reproductive tissues.

Among the embryonic tissues, globular embryos appeared to be the most divergent whereas heart-shaped and cotyledonary embryos are more closely linked. Surprisingly, cotyledonary embryos and unfertilized ovules seem to express the most similar subsets of genes pointing out the probable resemblance of the functional cellular status of these two types of tissues.

Energy metabolism is highly over-expressed in leaves compared to reproductive tissues

Among the five main groups of genes revealed by clustering analysis, the largest cluster gathers genes over-expressed in leaves compared with reproductive tissues. As expected, the functional annotation of these sequences revealed the dominance of genes involved in energy and photosynthesis pathways such as genes coding for ATP synthase, components from the oxydo-reduction chain as well as RUBISCO, proteins involved in photosystems I and II assembly, or chlorophyll a/b binding proteins, all these sequences representing more than 28% of the genes with known function present in the cluster. 20.2% of the genes with known function assembled in this cluster shows homologies with genes implied in several metabolism pathways including carbohydrates, lipids or amino acids synthesis. Another substantial set of sequences present in this cluster, representing 15.2% of the genes with an associated function, corresponds to genes involved in protein synthesis and degradation pathway. The over-expression of these particular categories of genes illustrates the high energetic activity of leaf tissue. This expected result validates the genes expression profiles obtained from the statistical analysis of our hybridization microarray data.

Embryogenic process implies major modifications between maternal and filial embryonic genetic programs

116 cDNA clones, representing 1.24% of the sequences spotted onto the microarray, were found to be differentially expressed between unfertilized ovules and embryonic tissues. Only 9 genes related to primary metabolism pathways showed a significant variation in expression level, thus revealing that basal cellular activities are similar in these tissues. However, despite the apparent slight conservation of global metabolic pathways, some crucial differences exist between the two tissues as demonstrated by the differential expression of genes involved, for example, in gene regulation. Indeed, 15 genes potentially involved in translation (ribosomal proteins or initiation factors) are over-expressed in embryos whereas this category is inexistent among the genes highly expressed in ovules. In turn, genes involved in proteins maturation and degradation process represent 9.1% of the total number of genes over-expressed in ovules but only 2.8% of

the sequences over-expressed in embryos. Some genes monitoring gene expression at the transcriptional level are also differentially expressed between embryonic and non-embryonic tissues as revealed by the over-expression of a CCAAT-binding transcription factor in ovules and the large proportion in embryonic tissue of over-expressed DNA-binding proteins of unknown function as well as H2A, H2B and H3 histones which are thought to control gene transcription activity by differentially modeling the chromatin. Among the genes over-expressed in embryos, a large fraction (more than 30%) corresponds to genes participating in nucleic acids and protein processes whereas these functional categories only account for 13.6% of the genes over-expressed in unfertilized ovules. Also, among the 44 genes over-expressed in unfertilized ovules, 9.1% are related to energy metabolism whereas this pathway is inexistent among the genes over-expressed in embryos showing that unfertilized ovules are highly active tissue. Lastly, compared to embryonic tissues, unfertilized ovules also show numerous highly expressed proteins related to senescence process.

Transcriptional regulation plays a central role in globular embryos.

Among the 8 185 putative unique sequences under study, only 4 genes were revealed as being over-expressed in globular embryos compared to heart-shaped embryos. This result could demonstrate that the large majority of biological pathways induced in the globular embryo carries on during the subsequent stage of embryo development. The functional characterization of these 4 genes provides interesting clues about the regulatory networks controlling the genetic program induced during embryo patterning revealing a crucial role for transcriptional regulation during early events of plant embryogenesis. Indeed, apart from one sequence showing no homology with any known or predicted protein, the three other genes, encoding a histone H4, a G-box binding transcription factor and a putative small ribonucleoprotein, are all associated with transcription process. According to hierarchical clustering results, we found that the clone showing no homology with known sequences was assembled with genes encoding proteins involved in protein metabolism. Applying the concept of ‘guilty by association’, we could hypothesis that this clone is involved in translation or protein maturation processes.

Recovery of a high cellular activity at the heart-shaped embryo stage

Functional classification of genes over-expressed in heart-shaped embryo compared to globular embryo reveals an important increase in cellular activity at this stage of development with the switch on of a wide range of cellular functions related to primary and energetic metabolism pathways, gene expression regulation or signal transduction. At the exception of ‘no hits’ and ‘unknown proteins’ which represent together up to 48.7% of the genes over-expressed in heart-shaped embryos, the largest functional category consists in genes associated with energy production and in particular genes encoding enzymes involved in glycolysis or Krebs cycle as well as proteins necessary for functional plastids biogenesis. Primary metabolism corresponds to 8.45% of the genes highly expressed in heart-shaped embryos and is characterised by a preponderance of lipids and amino acids synthesis processes. Genes involved in protein synthesis and degradation pathways are also present in high proportion, representing 7.15% of the over-

expressed genes. Another significant category, representing 7.1% of the genes, contains genes with a more surprising cellular function related to senescence process. Among the sequences over-expressed in heart-shaped embryo, we also identified a wide variety of genes involved in signal transduction such as diverse kinases, a putative RAS-related protein RAB11C or a 14-3-3-like protein as well as a few genes engaged in transcription among which 4 transcription factors with a NFkB (nuclear factor of kappa light) and one member from the bHLH family. Genes related to stress, mostly monooxygenases, are also present.

Cotyledonary embryo: maintenance of a high cellular activity associated with growth and reserves accumulation.

Compared to earlier stages of embryo development, cotyledonary embryos are characterized by a dramatic increase in protein metabolism pathways, which account for more than 21% of the over-expressed genes. Concerning the genes involved in protein folding and degradation identified as over-expressed in cotyledonary embryos, we found a majority of heat shock protein whereas none of these proteins appeared to be highly expressed at earlier stages of embryo development.

A large part of the variation in gene expression between cotyledonary and early embryos can be attributed to three significant functional classes, representing respectively 10.6%, 7.9% and 7.3% of the over-expressed genes in cotyledonary embryo, and corresponding to genes associated with primary metabolism, energy production including glycolysis and photosynthesis pathways and genes involved in DNA structure containing exclusively genes coding for histones H2B and H3.

Among the genes weakly expressed at early embryo stages and switched on at cotyledonary stage, we found genes involved in cell division and growth such as genes coding for the cytokinesis-related SEC1 protein, a putative auxin-independent growth protein or a prohibitin 1 like- protein as well as genes coding for proteins participating in cell wall synthesis such as cellulose synthase. Although these clones do not represent a large part of the genes preferentially expressed in cotyledonary embryo (only 2.6%), their presence reflects the high dividing and growing status of embryos at this stage of development. Also, genes coding for membrane and cell wall related proteins such as arabinogalactan protein (AGP), known for their putative role as signal molecule in somatic embryogenesis (Vroemen et al., 1999), appeared to be highly induced at the cotyledonary stage whereas such sequences were inexistent among the clones over-expressed at earlier stages of development. Among the sequences strongly expressed in the later embryonic stage under study, we could identify genes involved in signal transduction such as GTP-binding protein, kinases or hydroxyproline-rich glycoprotein. Few genes engaged in transcription process among which 4 transcription factors with one CONSTANS B box zinc finger, one BTF3 and one G-box B12D were preferentially expressed in cotyledonary embryos. Genes related to antioxidant functions, all encoding peroxidases except one polygalacturonase, are also present. Among the sequences over-expressed in cotyledonary embryo, we still find genes related to senescence process. Sequences annotated as “no hits or unknown”, corresponding to 23.8% of the clones over-expressed in cotyledonary embryos, represent the largest functional category.

Differential expressed genes identified by microarray analysis are confirmed by real time RT-PCR.

To validate microarray results, 9 clones identified as over-expressed at the different embryonic stages of interest were subjected to real-time RT-PCR analysis. The expression profiles obtained from quantitative RT-PCR were similar to those generated by microarray analysis (table 5).

Discussion

Microarray analysis is an efficient method to identify marker genes of early plant embryogenesis

In the present work, which aims at gaining information about the transcriptome induced during sunflower early embryogenesis as a complementary model to *Arabidopsis thaliana*, we analysed the expression profile of 8 185 putative different genes by comparing crucial embryo stages of development using the microarray technology. To our knowledge, such a study analyzing the metabolic pathways and the molecular regulatory mechanisms involved at precise stages of dicotyledonous early embryo development as well as their dynamics during the complete embryonic process has not yet been reported. This was allowed by the unique properties of sunflower inflorescence facilitating the harvesting of numerous embryos at diverse stages of development.

Several lines of evidence, such as the excellent correlations between the experimental repeats, testify of the high level of reproducibility and reliability of our microarray results. To perform statistical analysis, we used a method relying on analysis of variance which, contrary to common fold change-based analysis, takes into account all sources of variation likely to influence gene expression and depends on p-value significance thresholds to identify differentially expressed genes. To identify the differentially expressed genes, the stringent p-value of $5.42 \cdot 10^{-6}$ was used ensuring us to yield a very low false positive rate. Compared to other selection criteria employed to identify differentially expressed genes, ANOVA based microarrays analysis are more powerful as illustrated by volcano plot representations of significance against magnitude which revealed that many clones with low variance of the gene expression can be highly significant whereas some genes presenting high variance (more than the 2-fold change value usually arbitrarily defined as threshold) can sometimes constitute artifacts. The good concordance between the results derived from microarray analysis and quantitative RT-PCR experiments also confirms the high specificity of probe hybridization.

The validity of the gene expression profiles obtained in our study is reinforced by the nature of the genes declared to be over-expressed in leaf tissue. Considering the current lack of knowledge available concerning transcriptional activity during early phases of plant embryogenesis, we needed to validate the results

obtained from our statistical analysis by testing a tissue with a well-characterized gene expression profile. In this aim, we introduce the very specialized photosynthetic leaf tissue as an external control in our experimental design.

As expected, cluster analysis performed on differentially expressed clones, generated a large group of genes over-expressed in leaf tissue compared to reproductive tissues which contained a high proportion of sequences involved in energy and photosynthesis as well as primary metabolism pathways (carbohydrates, lipids and amino acids synthesis). These clustering results, in accordance with well-known physiological evidences, confirm the reliability of the gene expression profiles obtained during this study.

Among the 9 404 different clones tested in these microarrays experiments, only 744 were differentially expressed between at least two of the five conditions under study. This result demonstrates the very strong resemblance between the genetic programs induced during the different phases of embryonic process as well as between maternal unfertilized tissue and filial embryonic tissues. This notion of strong similarity between transcriptional profiles of reproductive tissues at diverse stages of development was previously demonstrated by *in silico* analysis of embryonic cDNA libraries (Ben et al., 2005).

Clustering analysis was performed to assess the resemblance between the five studied conditions based on transcriptional profiles similarities. Among reproductive tissues, globular embryos appeared to express the most divergent subsets of genes whereas genetic programs induced in heart-shaped and cotyledonary embryos seem to be more closely related. A possible explanation for this result could rely on the induction of very few but specific sets of genes in globular embryo such as genes necessary for polarity acquisition and for the establishment of the two main embryonic axes. At the heart-shaped stage, although the embryo still carries on patterning events leading to bilateral symmetry, it could also become nutritiously independent from surrounding maternal tissues, as the suspensor senesces, thus undergoing a capital and definitive transition towards an autotrophic development. This shift towards autonomous embryo development may be related to the activation of energetic and metabolic synthesis pathways that should be maintained until complete seed development. Surprisingly, unfertilized ovules and cotyledonary embryos present the most similar gene expression patterns. At cotyledonary stage, the embryo, which has achieved its final body plan organization, starts an active phase of growth and cell elongation and accumulates storage reserves. These cell activities mainly involved primary basal metabolism pathways like those primarily probably implied in unfertilized ovules.

As illustrated by the gene expression profiles generated by the global hierarchical clustering analysis (Fig.3), transcriptional activity of early plant embryogenesis is characterized by successive waves of induction of different types of genes. Functional characterization of the genes assembled in the different clusters revealed that embryogenesis initiation especially requires active signal transduction and transcriptional and translational regulations. While the embryo development progresses, induced cellular functions diversify:

later stages of embryo development are also characterized by high metabolic and energetic activity as well as an important cell-to-cell communication.

This evolution of embryo cellular activity during sunflower embryogenesis was confirmed by functional annotation of the genes declared as differentially expressed between the different contrasts defined for ANOVA analysis. Indeed, the functional characterization of the four genes over-expressed in globular embryos pointed out that, during embryogenesis initiation, gene expression is submitted to a specific and precise transcriptional regulation essential for the establishment of a correct genetic program suitable for the subsequent embryo patterning events. However, the lack of data concerning genes over-expressed in globular embryos prevented us to decipher the global biological mechanisms specifically involved during this crucial stage of development corresponding to the establishment of the two embryonic main axes. Strikingly, all the clones declared to be over-expressed in globular embryos come from cDNA libraries prepared from embryonic tissues at early stages of development. Among cDNA clones declared as over-expressed in heart-shaped embryos, genes associated to energetic and metabolic pathways are present in high proportion with preponderance of lipids and amino acids synthesis processes for primary metabolism pathways, probably linked to the initiation of lipid reserves accumulation and to an active protein synthesis. During cotyledonary stage, a high cellular activity is maintained and various cellular functions are induced from primary and energetic metabolism pathways to gene expression regulation and cellular division and growth. This ultimate stage of embryo development is also characterized by a high activation of translation process, probably associated with the onset of protein reserves accumulation.

The putative function of about 300 of the differentially expressed genes could not be assigned and annotated as no hits or unknown. The specific expression patterns of some of these genes suggest their potential importance in embryos development. These genes may be of great importance in future functional genomics study of sunflower embryogenesis.

Sunflower embryos are highly differentiating structures.

Careful examination of the differentially expressed genes at the 3 different stages of embryo development revealed that 5 cDNA clones related to cell cycle functions can be considered as differentially expressed. Among these clones, we find a gene encoding a PREG-like protein which is a member of the cyclin family, a putative AUXIN-INDEPENDENT GROWTH protein, a gene coding for the DIMINUTO/DWARF1 integral membrane protein, implicated in steroid synthesis and controlling cell elongation (Klarhe et al., 1998), a PROHIBITIN 1 corresponding to a highly conserved protein, inhibiting cell proliferation and acting as an important regulator in cell survival (McClung et al., 1995), as well as the well-characterized KEULE gene, coding for a SEC1 protein known to be involved in vesicular fusion during cytokinesis and previously described as playing a central role in radial axis formation during plant embryogenesis (Assaad et al., 2001). Surprisingly, cell division and growth processes appeared to be especially active during late stages of embryo development. In fact, during early phase of embryogenesis, as the embryo establishes the different

organs and structures, its size does not really increase. In turn, at the cotyledonary stage, while the embryo body plan is completely organised, the embryo grows and accumulates reserves. All these genes involved in cell proliferation and elongation might constitute key elements of embryo patterning control. However, their restricted number, compared to the large amount of differentially expressed genes involved in transcription, protein synthesis and degradation process or signal transduction, suggest that much more than simple proliferating and growing cellular masses, sunflower embryos seem to be engaged in very complex and active differentiation events regulated as much at transcriptional level as at post-transcriptional and post-translational levels.

Diverse transcriptional regulatory mechanisms are involved during early plant embryogenesis.

We have already mentioned that different types of histones are over-expressed in specific stages of embryo development, the histone H4 being over-expressed during the early stages of embryo development (globular and heart-shaped) whereas histones H2A, H2B and H3 are particularly strongly expressed in cotyledonary embryos. These specific expression patterns could be related to previously described results reporting that, in mealybug (*Planococcus citri*), histone H4 was shown to play a role in genomic imprinting phenomenon. Depending on its acetylation level, it regulates the activity of maternal and paternal genomes during early events following the fertilization (Ferraro et al., 2001).

Transcriptional regulation of the genetic programs induced during sunflower embryo differentiation is also illustrated by the differential expression of transcription factors from various families in each stage of embryogenesis. For example, a CCAAT-binding transcription factor was identified as over-expressed in unfertilized ovules compared to embryos. In heart-shaped embryos, a member of the bHLH transcription factor family, already reported as constituting an important family of transcription regulators for *Arabidopsis thaliana* early embryogenesis (de Folter et al., 2004) as well as a NFkappaB, constituting an essential survival factor mediating cellular response to a variety of extracellular stress stimuli and implicated in the control of cell proliferation and apoptosis in animals (Rayet et al., 1999), were found to be highly expressed. A G-box binding transcription factor appeared to be strongly expressed in globular embryos and further in cotyledonary embryos. Promoter G-box elements were already shown to confer seed-specific expression for a gene encoding a barley aldose reductase correlated with desiccation tolerance acquisition (Roncarati et al., 1995) as well as for a gene coding for a *Brassica napus* oleosin (Keddie et al., 1994). Thus, the G-Box binding transcription factor identified in our study as over-expressed in globular and cotyledonary embryos could represent an essential regulator leading to seed-specific activation of genes playing key roles in plant embryogenesis. Detailed functional analyses should be of great interest to clarify its role in sunflower embryo development. Among transcription factors over-expressed in cotyledonary embryos, are a member from the seed-specific VP1/ABI3 regulatory factor family, required for the abscisic acid induction of maturation-specific genes including Lea (late embryogenesis abundant protein) genes leading to acquisition of desiccation tolerance and arrest in embryo growth (Hoecker et al., 1995) and a putative CONSTANS B-box zinc finger protein, a clock-regulated transcription factor implicated in photoperiod-regulated responses,

controlling the expression of flowering as well as tuberization inductive genes (Martinez-Garcia et al., 2002; Yanovsky and Kay, 2003).

Embryonic genetic program is controlled through different post-transcriptional regulation mechanisms.

During embryo development, different post-transcriptional regulation mechanisms appear to be required to control the genetic programs induced at each stage.

One of the genes highly expressed in ovules encodes a polyubiquitin. Ubiquitination was already found to be an important post-translational regulatory process of the unfertilized egg cell in wheat (Sprunck et al., 2005) where it could participate to the establishment and maintenance of the functional status of the unfertilized ovule, maybe by preventing the onset of DNA replication and other cell-cycle related events before the fertilization occurs.

Among the two genes implicated in protein degradation pathway and over-expressed in embryos compared to unfertilized ovules, one shows similarities with a trypsin proteinase inhibitor. Proteinase inhibitors have been shown to be developmentally expressed in seeds and reserve organs and are known to be involved in reserve control (De Leo, 1998). Such proteins, which have already been reported as preferentially expressed in sunflower protoplasts cultivated in embryogenic conditions by analysis of EST frequency in cDNA libraries (Tamborindeguy et al., 2004) and microarray experiments (Tamborindeguy et al., submitted), could constitute essential elements of the post-translational regulatory network involved in embryonic process.

Among the genes over-expressed in globular embryos, we found a putative ribonucleoprotein. Ribonucleoproteins constitute members of complexes implicated in various biological processes regulating gene expression such as mRNA cellular localisation, maturation, degradation and translation. Such complexes were shown to play critical roles during animal embryogenesis (Stebbins-Boaz and Richter, 1997). However, nothing is known about their putative role and involvement during plant embryogenesis.

Strikingly, half of the genes involved in protein maturation pathways over-expressed in cotyledonary embryos corresponds to heat-shock proteins whereas these genes are inexistent among the sequences highly expressed at earlier stages of embryo development. Among these differentially expressed chaperone proteins, Hsp90 and Hsp70 are the most abundant. These two proteins, known to be involved in the folding and assembly of proteins playing roles in signal transduction, cell regulation and transcriptional activation, operate as part of a multichaperone complex in which Hsp70 provides substrate for Hsp90-mediated folding (Zhang et al., 2002).

Different types of translation factors seem to be induced at each phase of embryo development: For example, two translation initiation factors are over-expressed during early embryo stages, whereas exclusively translation elongation factors appeared to be over-expressed in cotyledonary embryos. These results indicate clearly that each tissue employed different types of translation initiation factors to achieve

protein synthesis. Notably, the expression profile of the translation elongation factor EF1 α obtained by microarray analysis, with a high expression in the ultimate stage of embryogenesis, is in accordance with the results previously obtained from embryonic cDNA libraries analysis (Ben et al., 2005).

Post-transcriptional regulation of the genetic programs specifically involved at each development stage seems also to rely on the nature of the ribosomal proteins expressed in each tissue. The differential expression of ribosomal proteins under different conditions has already been described (Karsi et al., 2002) as well as their role in plant and animal cell division growth and development (Barakat et al., 2001). Moreover, a 60S ribosomal L10 protein, over-expressed at the cotyledonary stage by our microarray experiments, shows exactly the same profile according to digital northern performed on EST collection from cDNA libraries (Ben et al., 2005), suggesting the implication of this kind of post-transcriptional gene expression regulation during embryogenesis.

Active signal transduction and intercellular communication are necessary for correct embryo patterning.

The highly differentiating status of the embryos is also demonstrated by the diversity and the specificity of the signal transduction pathways involved at each stage of development. Thus, heart-shaped embryos highly expressed a 14-3-3 protein, a Ras- related protein RAB11C whereas cotyledonary embryos over-expressed a GTP binding protein and a hydroxyproline-rich glycoprotein as well as calmodulin. The members of the 14-3-3 isoforms family have been shown to be developmentally regulated during animal embryogenesis, where they take part in cell differentiation processes (Fu et al., 2000). Recently, results derived from experiments carried on somatic and microspore embryogenesis models could suggest that 14-3-3 proteins might also play critical roles during plant embryogenesis (Maraschin et al., 2003; Rienties et al., 2005). Rab11 GTPase, which is implicated in the regulation of various vesicle transport processes, was shown to be essential for oskar mRNA localization during *Drosophila* oogenesis, a crucial event leading to the correct elaboration of the future embryonic body plan (Dollar G et al., 2002). Casein kinase I and a calcium-dependent protein kinase (CDPK) are highly expressed in heart-shaped embryos and two other kinases, among which a putative serine/threonine kinase, are over-expressed in cotyledonary embryo. Calcium-dependent protein kinase was shown to be temporally associated to oil bodies during seed maturation in sandalwood (*Santalum album* L.) and active CDPK protein was previously detected in sunflower and various other plant species suggesting that CDPK and Ca²⁺ may have a regulatory role during oil accumulation and oil body biogenesis (Anil et al., 2003). In mammals, isoforms of casein kinase I phosphorylate many different substrates among which are key regulators notably controlling cell differentiation and proliferation (Knippschild et al., 2005). It is may be important to mention that genes encoding Ras-like GTP-binding protein, calmodulin or calcium-dependent protein kinase, were also found to be highly expressed in the developing non-green seeds of sesame (*Sesamum indicum*) compared to *Arabidopsis thaliana* green seeds according to EST data analysis

(Suh et al., 2003) suggesting that these genes could participate in signal transduction pathways specifically or preferentially involved during non-green seeds formation.

Transition to embryo autotrophic development: a critical phase of plant embryogenesis.

The heart-shaped stage of embryo development is characterized by a substantial increase in energy synthesis processes which is further maintained during the cotyledonary stage. This high energetic activity could be linked to the initiation of autotrophic embryo development phase. Indeed, while growth of early embryos depends upon the suspensor connection to nutritive maternal tissues, this embryonic structure senesces during heart-shaped embryo stage and the embryo should activate proper metabolic pathways to be able to develop in an autonomous way. The active energetic status for these two embryonic stages is illustrated by the over-expression of genes encoding ATP synthases as well as genes coding for various enzymes involved in glycolysis as well as Krebs cycle such as aconitate hydratase, pyruvate kinase or lactate dehydrogenase which were found to be highly expressed in heart-shaped embryo or phosphoglycerate mutase, triose phosphate isomerase, enolase, aldolase or 3-phosphoglycerate kinase found to be over-expressed during cotyledonary stage. In addition to the production of energy and reducing power, these energetic pathways provide pyruvate and acetyl-coA which represent two components at the crossroads of many metabolic pathways. Consequently, these two processes play a central role in cellular metabolism. Genes involved in plastids biogenesis such as genes encoding proteins involved in photosystem I (PSI) assembly as well as RUBISCO, cytochromes and chlorophyll a/b binding protein, are also induced in heart-shaped embryos. In mature chloroplasts, the thylakoids membrane contains two types of photosystems, PSI and PSII, differing by their light absorption properties. Our microarray results showed that the two types of photosystems were active in leaves whereas only PSI was induced in embryos. Plastids are known to shelter fatty acids biosynthesis. So, the presence of functional plastids in sunflower seed could be essential for correct embryo development.

Senescence, a crucial regulatory process of sunflower embryo patterning.

Several genes encoding senescence-related proteins show high level of expression at diverse stages of embryo development. Three of these genes were identified as highly expressed in cotyledonary embryos and senescence-related genes represent up to 12.1% and 13.9% of the clones with known function over-expressed in unfertilized ovules and in heart-shaped embryos. These observations are in agreement with the results obtained from digital northern realised on EST generated from sunflower embryonic cDNA libraries which revealed the over-expression of one putative senescence-associated protein in heart-shaped embryos (Ben et al., 2005). Moreover, among the genes revealed as over-expressed in at heart-shaped embryo stage, we can also find several enzymes participating in protein degradation process such as cysteine proteases, a clp-like proteinase or ubiquitin, previously described as acting in senescence or programmed cell death (Ingvarsdén and Veierskov, 2001; Nakabayashi et al., 1999; Xu and Chye, 1999). The implication of

senescence process and programmed cell-death during plant gametogenesis and embryogenesis has already been reported (Wu and Cheun, 2000, Bozhkov et al., 2005). In unfertilized ovules, activation of this cell death mechanism could be related to gametogenesis events and more particularly to the degeneration of the three macrospores which do not develop to form the embryo sac (Wu and Cheun, 2000). During embryogenesis, senescence could be linked to various process such as morphogenesis (Lam, 2004), suspensor senescence (Schwartz et al., 1997) or vascular tissues formation (Fukuda, 1996). Gene expression analysis such as *in situ* hybridization may be useful in this context to identify more precisely the tissue-specific expression patterns of the senescence encoding genes. This may shed light into the biological or morphological process in which these genes are involved.

Owing to this global transcriptomic study, several hypotheses concerning key biological processes involved in sunflower embryo development could be proposed. First, our results revealed that embryos are under an active differentiating process rather than exclusively proliferating events. They also revealed the high complexity and subtlety of the gene expression regulation networks, occurring at transcriptional or post-transcriptional levels.

Our results also pointed out that embryo cells actively communicate through a wide range of signal transduction pathways. Lastly, three rather surprising processes emerged as being absolutely essential for plant embryogenesis: senescence, active energy production and the biogenesis of functional plastids. Although this study allowed us to highlight some of the major biological pathways involved in early phases of sunflower embryogenesis, some links between those processes are still lacking. Lastly, as morphological and histological events are very similar among dicotyledonous plants, this study could provide clues for the characterization of molecular events occurring during seed formation in other exalbuminous dicots species.

Material and methods

Plant material

Helianthus annuus L. genotype Ha300b, selected for its high yield in seeds, was grown under standard conditions in greenhouse (14 h photoperiod and 23°C/18°C light dark conditions). Unfertilized ovules and immature embryos, generated by hand auto-pollination, were sampled by microdissection of the seeds using binoculars, classified according to their developmental stages (unfertilized ovules, globular, heart-shaped and cotyledonary embryos) and stored indefinitely at -20°C in RNA Later (Ambion, Austin, TX, USA). Leaf samples were collected from B6 stage plants.

Microarray construction

Expressed sequence tags from different sunflower cDNA libraries were clustered using the Phrap assembly program to identify unigenes. 9 404 clones representative of 8 185 unigenes were rearranged into 96-well

microplates containing 200 µl of culture medium (LB Freezing) with 100 mg/L Ampicillin and incubated at 37°C or 31°C (depending on the bacterial strain used for cloning vector) in a shaking incubator (200rpm) for 24 hours.

5 µl of bacterial suspension was transferred into thermowell 96-well microplates containing 95 µl of MilliQ water. The plates were then heated for 10 min at 95°C followed by centrifugation at 1200 g for 3 min to pellet cellular debris. 8 µl of the supernatant of the rough bacterial lysate were used to amplify the cDNA insert using primers specific for sequences flanking the vector cloning site : for cDNA cloned into PCRII, pSport1 and pGMT vectors, the universal M13 primers (M13 forward 5'GTTTTCCCAGTCACGACGTTG 3' and M13 reverse 5' TGAGCGGATAACAATTTTCACACAG 3') were used whereas for inserts in pTriplex2 forward and reverse oligonucleotides were forward 5' TCCGAGATCTGCACGAGC 3' (Triplex5) and reverse 5' TAATACGACTCACTATAGGG 3' (Triplex3), respectively. PCR reaction was performed in 100 µl final volume with 10 µl 10X buffer, 2.5 mM of MgCl₂, 0.2 µM of each primer, 200 µM of each dNTP and 3 units of Taq DNA Polymerase (Quiagene, Courtaboeuf, France). DNA samples were denatured at 95°C for 2 min followed by 40 cycles of amplification at 95°C for 30 sec, 52°C for 30 sec and 72°C for 2 min followed by a final extension step for 7 min at 72°C. PCR fragments were visualized on 1% agarose gel to ensure adequate PCR amplification. PCR products showing double bands were removed or replaced by another cDNA clone belonging to the same contig. The PCR samples were then evapoconcentrated and resuspended into 30 µl of MilliQ water to obtain a concentration of approximately 300-500 ng/µl. Finally, the concentrated PCR products were arrayed in duplicates onto Hybond N+ nylon membrane (Amersham, Orsay, France) using the *MicroGrid II* (Biorobotics LTD, Cambridge, U.K.) with 64 microarraying pins. A 13x13 gridding pattern was applied. Relative humidity was maintained to 70 % and temperature at 21°C throughout the arraying procedure. The spotting of the microarrays was performed at the "Centre de Ressources-Génotypage, Séquencage" of Toulouse, France.

After printing, PCR products were fixed on the nylon membrane by placing the microarrays facing up onto Whatman paper moistened with denaturation solution (1.5 M NaCl and 0.5 M NaOH) followed by neutralization solution (1.5 M NaCl and 1M Tris-HCl, pH 7.4) for 20 min each. The treated filters were then baked at 80°C for 2 hours followed by UV crosslinking with UV Stratalinker 1800.

RNA isolation and cDNA probes preparation

Total RNA was isolated from leaves, immature embryos and unfertilized ovules tissues according to Chomczynski and Sacchi protocol (1987). To control for biological variation between individuals, plant material samples from different individuals were pooled prior to RNA extraction, for each tissue under study. cDNA were produced from DNase-treated RNA using the SMART PCR cDNA synthesis kit (Clontech, Palo Alto, CA, USA) according to manufacturer's instructions. Then, for each condition, complex probe was prepared from 100 ng of cDNA labeled with [α -³³P] dCTP using Megaprime DNA Labeling System (Amersham, Orsay, France) following the manufacturer's specifications. Before use for hybridization, the radiolabeled probes were purified from unincorporated labeled nucleotides by passing

through Probe Quant G-50 Micro Columns (Amersham, Orsay, France), denatured at 100°C for 10 min and cooled on ice.

Preparation of oligo probes

500 ng of each reverse primers used to amplify the EST (M13 reverse and Triplex3) were end-labeled with 3 µL of [γ -³³P]dATP (30 µCi), using 200 units of polynucleotide kinase (Gibco, Invitrogen, Cergy Pontoise, France). The reaction was incubated for 60 min at 37°C and then heated for 10 min at 95°C to end the reaction. Radiolabeled oligo probes mix was purified through Probe Quant G-50 Micro Columns (Amersham, Orsay, France).

Microarray hybridization and image processing

An incomplete Latin square (balanced incomplete blocks design) was chosen for the experimental design of our microarray analysis. Five microarrays were respectively hybridized with four of the five tissues under study, every pair of treatments appearing an equal number of times between the different filters (3 times in our case). This experimental design has the advantage to allow the control of two supplementary sources of variation (microarrays and technical replication) in addition to the treatment effect.

For hybridization with complex probes, membranes were first pre-hybridized for 24 h at 65°C in a buffer containing 0.5% SDS, 5X SSC, 0.1% each of Ficoll 400, Polyvinylpyrrolidone and Albumine Bovine Fraction and 100 µg.ml⁻¹ of Salmon sperm DNA (Sigma, Saint Quentin Fallavier, France). Then, pre-hybridized microarrays were incubated for 24 h at 65°C in the same buffer containing the radiolabeled complex probe. Membranes were washed twice in 0.1% SDS, 0.1X SSC at 65°C for 15 min each and exposed for 3 to 6 nights to Fuji film imaging plate as a radioactive energy sensor. Radioactive intensity of the spots was captured using Bio-Imaging Analyser BAS-5000 (Fujifilm) with density gradation of 16 bit/Pixel at a resolution of 25 µm/Pixel. Finally, spots assignment and intensities quantification were realised using ArrayGauge V.1.3 (Fujifilm). After subtracting background from each spot intensity, numeric values were used for statistical analysis.

To test the quality of the spotting on the nylon membranes, each microarray was also hybridized with labeled oligonucleotides. The hybridization buffer used for this checking experimental phase is the same as the one used for complex probes hybridization. After the pre-hybridization of the membranes for 12h at 42°C, labeled oligonucleotides probe was added and membranes were incubated for 24h at 42°C. Filters were then washed for 10 min at room temperature in 2X SSC, 0.1% SDS followed by 5 min at 42°C. Membranes were exposed to Fuji film imaging plate for 1 night.

Between each round of hybridization, filters were dehybridized by washing twice in boiling SDS 0.5%, SSC 0.1X for 15 min each and scanned after one night exposure to confirm the dehybridization.

Microarray data analysis

Statistical analyses were performed as described by Wolfinger (Wolfinger et al., 2001), using the R 2.0.0 software. Two interconnected ANOVA models were performed, the “normalization” model and the “gene”

model. The normalization model, which uses log₁₀ of signal intensity, accounts for experiment-wide systematic effects. The residuals of this model are considered as normalized values and are used as input data for the gene models. Genes showing a least one signal intensity above the maximum of the negative controls normalized values were retained as expressed genes. The gene models are fit separately to the normalized data from each gene. A test of heterogeneous variance for the normalized expression levels among tissues was done using the Levene test for each gene-model. For genes showing a homogenous variance and a significant tissue effect, the “Gene-by-Tissue” sum-of-squares were further split into four linear contrasts that were used as *post hoc* test to identify the origin of variability. We tested for the difference between ‘leaves’ versus ‘mean of reproductive tissues’ (including unfertilized ovules and embryos), between ‘mean of the embryos’ (all embryo stages confounded) versus ‘unfertilized ovules’, between ‘globular embryos’ versus ‘heart-shaped embryos’ and between ‘mean of early embryo stages’ (i.e. globular and heart-shaped) versus ‘cotyledonary embryos’. The Bonferroni method was used to conservatively reduce errors due to multiple testings. Volcano plots of significance against magnitude of effects were drawn for each couple of comparisons. Clustering of normalized expression levels of genes that showed a significant overall tissue effect using ANOVA method was done using classical hierarchical clustering with average linkage and Pearson’s coefficient correlation distance calculation technique.

Real-time PCR

The expression levels of 9 differentially expressed EST were analyzed by Q-RT-PCR to confirm the results obtained from microarray analysis. Gene specific primers were designed using the Primer Express software version 2.0 (Applied Biosystems, Courtaboeuf, France). The reaction was performed in a final volume of 10 µl composed by 5µl 2X Syber Green Mastermix (Applied Biosystems), 300 nM of each primer and 6 ng of cDNA. The PCR reactions were run in ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). PCR program was: 50°C for 2 min, 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. Following the PCR amplification, the reactions were then subjected to a dissociation program of 95°C for 15 s, 60°C for 15 s followed by 20 min of slow ramp temperature from 60°C to 95°C. Three replicates of each reaction were performed. β-Actin, a putative constitutively expressed gene, was used as an internal control for normalization. Quantification of relative changes in gene expression was performed using the $2^{-\Delta\Delta CT}$ method as described by Livak (Livak and Schmittgen, 2001).

Acknowledgements

The authors wish to thank Cecilia Tamborindegy, Nathalie Ladouce and Frédérique Bena for the construction of some cDNA libraries used to product the cDNA microarray and Patrick Bermudes for his technical support for sunflower culture. We also thank our colleagues from the “Resource Genomics Center” of Toulouse for advice and access to microarray equipments as well as our colleagues from the “Plateforme Bioinformatique” for providing efficient tools for sequences annotation. Cécile Ben is supported by

fellowship of the French MENRT. This work was partially supported by grants from Génoplante GOP-HG01 programme and from the Toulouse Genomics Center (French Network of Genomics Centers).

REFERENCES

Al-Chaarani GR, Gentzbittel L, Huang XQ, Sarrafi A. (2004) Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (*Helianthus annuus* L.). *Theor Appl Genet.* 109(7):1353-60.

Anil VS, Harmon AC, Rao KS. (2003) Temporal association of Ca(2+)-dependent protein kinase with oil bodies during seed development in *Santalum album* L.: its biochemical characterization and significance. *Plant Cell Physiol.* 44(4):367-76.

Assaad FF, Huet Y, Mayer U, Jurgens G. (2001) The cytokinesis gene KEULE encodes a Sec1 protein that binds the syntaxin KNOLLE. *J Cell Biol.* 152(3):531-43.

Barakat A, Szick-Miranda K, Chang IF, Guyot R, Blanc G, Cooke R, Delseny M, Bailey-Serres J. (2001) The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiol.* 127(2):398-415.

Ben C, Hewezi T, Jardinaud MF, Bena F, Ladouce N, Moretti S, Tamborindéguy C, Liboz T, Petitprez M, Gentzbittel L. (2005) Comparative analysis of early embryonic sunflower cDNA libraries. *Plant Mol Biol.* 57(2):255-70.

Bozhkov PV, Filonova LH, Suarez MF. (2005) Programmed cell death in plant embryogenesis. *Curr Top Dev Biol.* 67:135-79.

Breton C, Chaboud A, Matthys-Rochon E, Bates EE, Cock JM, Fromm H, Dumas C. (1995) PCR-generated cDNA library of transition-stage maize embryos: cloning and expression of calmodulin genes during early embryogenesis. *Plant Mol Biol.* 27(1):105-13.

Chomczynski P, Sacchi N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem.* 162(1):156-9.

- de Folter S, Busscher J, Colombo L, Losa A, Angenent GC. (2004) Transcript profiling of transcription factor genes during silique development in *Arabidopsis*. *Plant Mol Biol.* 56(3):351-66.
- De Leo F, Bonade-Bottino MA, Ceci LR, Gallerani R, Jouanin L. (1998). Opposite effects on spodoptera littoralis larvae of high expression level of a trypsin proteinase inhibitor in transgenic plants *Plant Physiol.* 118(3):997-1004.
- Dollar G, Struckhoff E, Michaud J, Cohen RS. (2002) Rab11 polarization of the *Drosophila* oocyte: a novel link between membrane trafficking, microtubule organization, and oskar mRNA localization and translation. *Development.* 129(2):517-26.
- Ferraro M, Buglia GL, Romano F. (2001) Involvement of histone H4 acetylation in the epigenetic inheritance of different activity states of maternally and paternally derived genomes in the mealybug *Planococcus citri*. *Chromosoma.* 110(2):93-101.
- Fu H, Subramanian RR, Masters SC. (2000). 14-3-3 proteins: structure, function and regulation. *Annual Review of Pharmacology and Toxicology* 40, 617–647.
- Fukuda H. (1996) Xylogenesis: Initiation, Pogression, and cell death. *Annu Rev Plant Physiol Plant Mol Biol.* 47:299-325.
- Gentzbittel L, Abbott A, Galaud JP, Georgi L, Fabre F, Liboz T, Alibert G. (2002) A bacterial artificial chromosome (BAC) library for sunflower, and identification of clones containing genes for putative transmembrane receptors. *Mol Genet Genomics.* 266(6):979-87.
- Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J. (2000) Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol.* 124(4):1570-81.
- Heck GR, Perry SE, Nichols KW, Fernandez DE. (1995) AGL15, a MADS domain protein expressed in developing embryos. *Plant Cell.* 7(8):1271-82.
- Hennig L, Gruissem W, Grossniklaus U, Kohler C. (2004) Transcriptional programs of early reproductive stages in *Arabidopsis*. *Plant Physiol.* 135(3):1765-75.
- Hoecker U, Vasil IK, McCarty DR. (1995) Integrated control of seed maturation and germination programs by activator and repressor functions of Viviparous-1 of maize. *Gene Dev.*9(20):2459-69.

- Ingvarsdén C, Veierskov B. (2001) Ubiquitin- and proteasome-dependent proteolysis in plants. *Physiol Plant*. 112(4):451-459.
- Jurgens G. (2001) Apical-basal pattern formation in *Arabidopsis* embryogenesis. *EMBO J*. 20(14):3609-16.
- Jurgens G. (2003) Growing up green: cellular basis of plant development. *Mech Dev*. 120(11):1395-406.
- Karsi A, Patterson A, Feng J, Liu Z. (2002) Translational machinery of channel catfish: I. A transcriptomic approach to the analysis of 32 40S ribosomal protein genes and their expression. *Gene*. 291(1-2):177-86.
- Keddie JS, Tsiantis M, Piffanelli P, Cella R, Hatzopoulos P, Murphy DJ. (1994) A seed-specific *Brassica napus* oleosin promoter interacts with a G-box-specific protein and may be bi-directional. *Plant Mol Biol*. 24(2):327-40.
- Klahre U, Noguchi T, Fujioka S, Takatsuto S, Yokota T, Nomura T, Yoshida S, Chua NH. (1998) The *Arabidopsis* DIMINUTO/DWARF1 gene encodes a protein involved in steroid synthesis. *Plant Cell*. 10(10):1677-90.
- Knippschild U, Gocht A, Wolff S, Huber N, Lohler J, Stoter M. (2005) The casein kinase 1 family: participation in multiple cellular processes in eukaryotes. *Cell Signal*. 17(6):675-89.
- Lam E. (2004) Controlled cell death, plant survival and development. *Nat Rev Mol Cell Biol*. 5(4):305-15.
- Laux T, Wurschum T, Breuninger H. (2004) Genetic regulation of embryonic pattern formation. *Plant Cell*. 16 Suppl:S190-202.
- Lee JM, Williams ME, Tingey SV, Rafalski JA. (2002) DNA array profiling of gene expression changes during maize embryo development. *Funct Integr Genomics*. 2(1-2):13-27.
- Li Z, Thomas TL. (1998). PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in *Arabidopsis*. *Plant Cell*. 10(3):383-98.
- Livak KJ, Schmittgen TD. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC(T)} Method. *Methods*. 25(4):402-8.
- Maraschin Sde F, Lamers GE, de Pater BS, Spalink HP, Wang M. (2003) 14-3-3 isoforms and pattern formation during barley microspore embryogenesis. *J Exp Bot*. 54(384):1033-43.

Martinez-Garcia JF, Virgos-Soler A, Prat S. (2002) Control of photoperiod-regulated tuberization in potato by the *Arabidopsis* flowering-time gene CONSTANS. Proc Natl Acad Sci U S A. 99(23):15211-6.

Mayer U, Ruiz RAT, Berleth T, Misera S, Jurgens G . (1991). Mutations affecting body organization in the *Arabidopsis* embryo. Nature. 353:402-407.

McClung JK, Jupe ER, Liu XT, Dell'Orco RT. (1995) Prohibitin: potential role in senescence, development, and tumor suppression. Exp Gerontol. 30(2):99-124.

Meinke DW, Sussex IM. (1979) Embryo-lethal mutants of *Arabidopsis thaliana*. A model system for genetic analysis of plant embryo development. Dev Biol. 72(1):50-61.

Nakabayashi K, Ito M, Kiyosue T, Shinozaki K, Watanabe A. (1999) Identification of clp genes expressed in senescing *Arabidopsis* leaves. Plant Cell Physiol. 40(5):504-14.

Rayet B, Gelinac C. (1999) Aberrant rel/nfkb genes and activity in human cancer. Oncogene. 18(49):6938-47.

Rienties IM, Vink J, Borst JW, Russinova E, de Vries SC. (2005) The *Arabidopsis* SERK1 protein interacts with the AAA-ATPase AtCDC48, the 14-3-3 protein GF14lambda and the PP2C phosphatase KAPP. Planta. 221(3):394-405.

Roncarati R, Salamini F, Bartels D. (1995) An aldose reductase homologous gene from barley: regulation and function. Plant J. 7(5):809-22.

Ruuska SA, Girke T, Benning C, Ohlrogge JB. (2002) Contrapuntal networks of gene expression during *Arabidopsis* seed filling. Plant Cell. 14(6):1191-206.

Schwartz B. W., Vernon D. M., and Meinke D. W. (1997) Development of the suspensor: Differentiation, communication and programmed cell death during plant embryogenesis. In “ Cellular and Molecular Biology of Plant Seed Development” (B. A. Larkins and I. K. Vasil, Eds.), vol 2, pp. 53-72. Kluwer Academic Press, Dordrecht, The Netherlands.

Sprunck S, Baumann U, Edwards K, Langridge P, Dresselhaus T. (2005) The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.). Plant J. 41(5):660-72.

- Sreenivasulu N, Altschmied L, Radchuk V, Gubatz S, Wobus U, Weschke W. (2004) Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains. *Plant J.* 37(4):539-53.
- Stebbins-Boaz B, Richter JD. (1997) Translational control during early development. *Crit Rev Eukaryot Gene Expr.* 7(1-2):73-94.
- Suh MC, Kim MJ, Hur CG, Bae JM, Park YI, Chung CH, Kang CW, Ohlrogge JB. (2003) Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. *Plant Mol Biol.* 52(6):1107-23.
- Tamborindeguy C, Ben C, Liboz T, Gentzbittel L. (2004) Sequence evaluation of four specific cDNA libraries for developmental genomics of sunflower. *Mol Genet Genomics.* 271(3):367-75.
- Tang S, Yu JK, Slabaugh B, Shintani K, Knapp J. (2002) Simple sequence repeat map of the sunflower genome. *Theor Appl Genet.* 105(8):1124-1136.
- Topping JF, Agyeman F, Henricot B, Lindsey K. (1994) Identification of molecular markers of embryogenesis in *Arabidopsis thaliana* by promoter trapping. *Plant J.* 5(6):895-903.
- Vroemen C, de Vries S, Quatrano R. (1999) Signalling in plant embryos during the establishment of the polar axis. *Semin Cell Dev Biol.* 10(2):157-64.
- White JA, Todd J, Newman T, Focks N, Girke T, de Ilarduya OM, Jaworski JG, Ohlrogge JB, Benning C. (2000) A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiol* 124(4):1582-94.
- Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS. (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol.* 8(6):625-37.
- Wu HM, Cheun AY. (2000) Programmed cell death in plant reproduction. *Plant Mol Biol.* 44(3):267-81.
- Xu FX, Chye ML. (1999) Expression of cysteine proteinase during developmental events associated with programmed cell death in brinjal. *Plant J.* 17(3):321-7.

Yanovsky MJ, Kay SA. (2003) Living by the calendar: how plants know when to flower. *Nat Rev Mol Cell Biol.* 4(4):265-75.

Zhang X, Beuron F, Freemont PS. (2002) Machinery of protein folding and unfolding. *Curr Opin Struct Biol.* 12(2):231-8.

Zhu T, Budworth P, Chen W, Provart N, Chang H, Guimil S, SuW, Estes B, Zou G, Wang X. (2003). Transcriptional control of nutrient partitioning during rice grain filling. *Plant Biotechnology Journal.* 1:59-70.

Figure 1. Reproducibility control between technical repeats.

Example from unfertilized ovules hybridizations.

Pairwise scatterplots were obtained from raw data.

R value corresponds to correlation coefficient.

Figure 2. Volcano plots of significance against magnitude of effects.

A. Linear contrast of mean normalized expression values in globular embryos *versus* mean normalized expression values in heart-shaped embryos.

B. Linear contrast of mean normalized expression values in early embryos *versus* mean normalized expression values in cotyledonary embryos.

C. Linear contrast of mean normalized expression values in embryos *versus* mean normalized expression values in unfertilized ovules.

D. Linear contrast of mean normalized expression values in reproductive tissues *versus* mean normalized expression values in leaves.

The horizontal line represents the Bonferroni correction p-value threshold.

Figure 3. Hierarchical clustering of the 744 differentially expressed clones, using average linkage and Pearson's coefficient correlation.

Genes are listed by lines and tissues in column.

L: Leaves, G: Globular embryo, H: heart-shaped embryo, Cot: Cotyledonary embryo, O: unfertilized ovule.

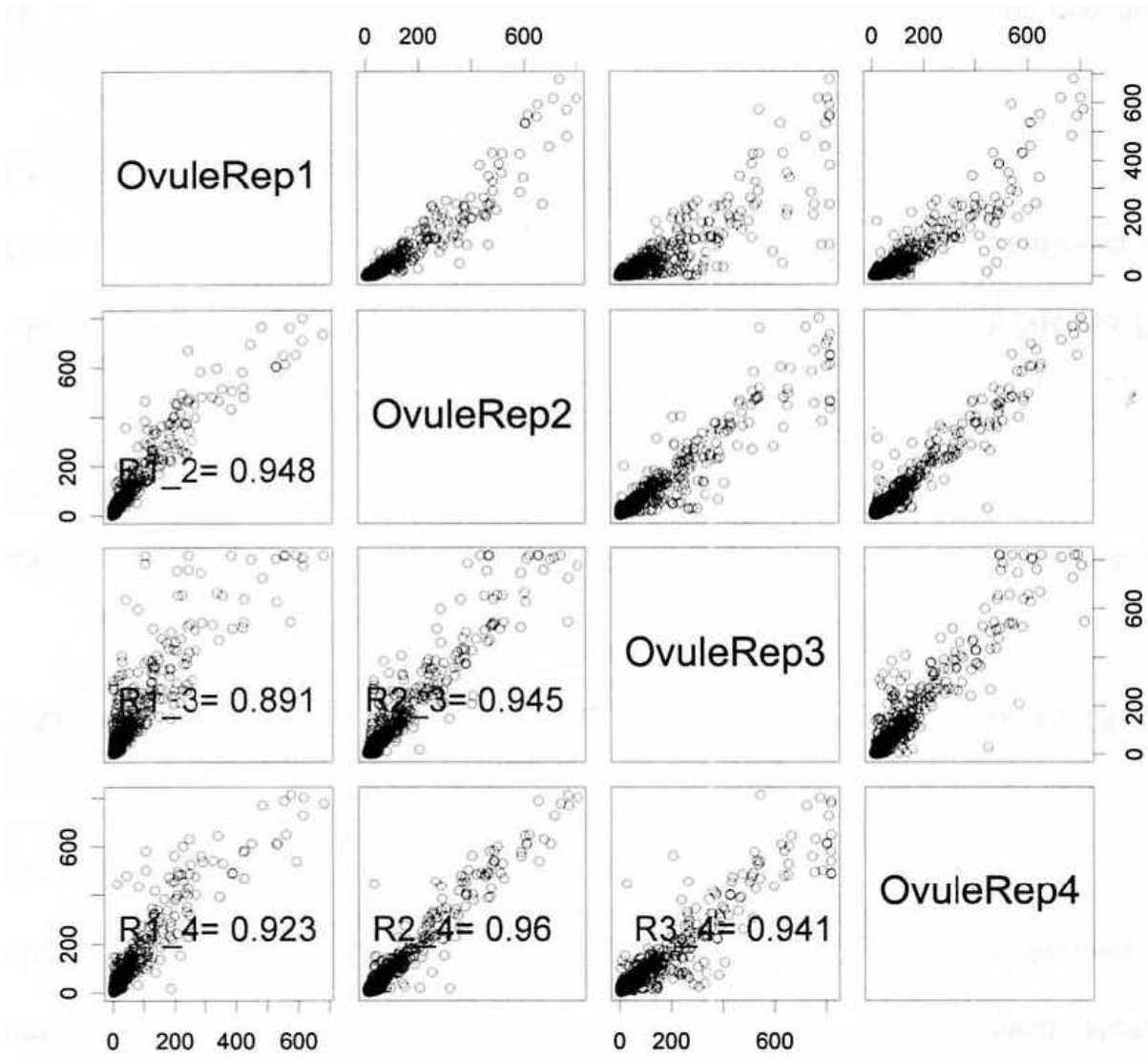
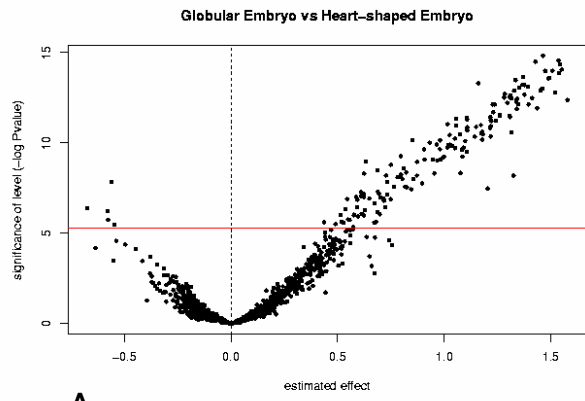
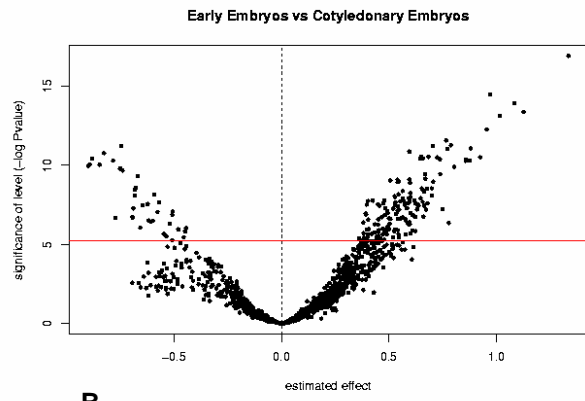


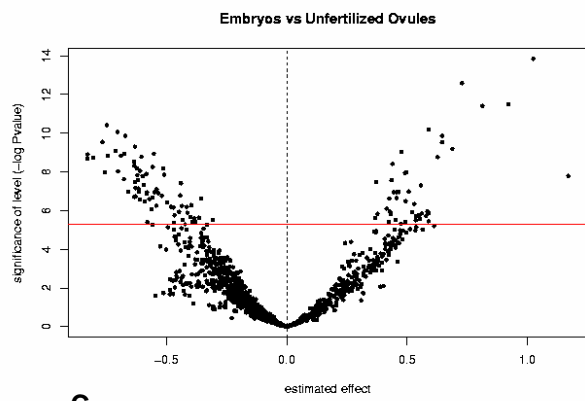
Figure 1.



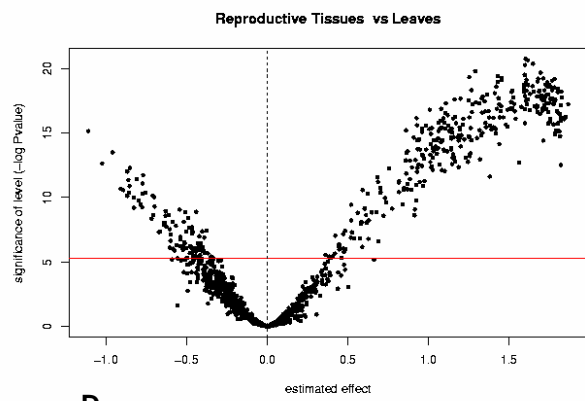
A



B



C



D

Figure 2.

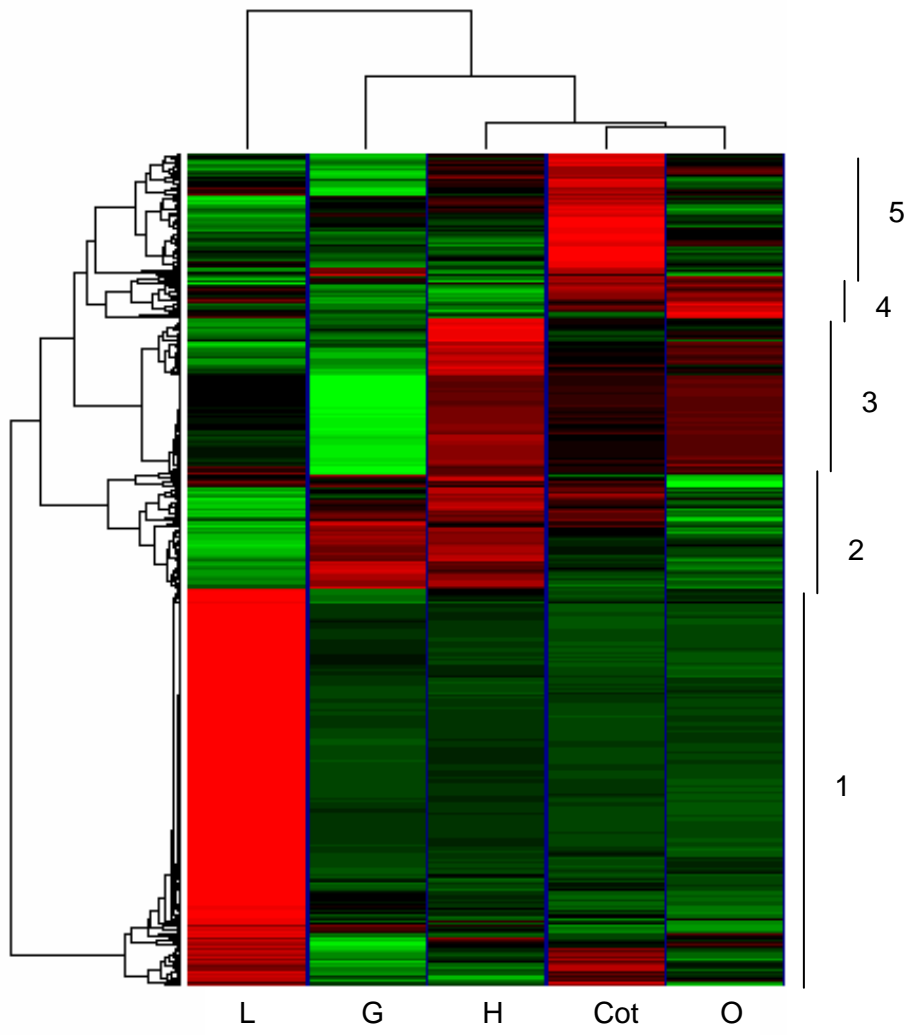


Figure 3.

	Slide 1	Slide 2	Slide 3	Slide 4	Slide 5
Repeat 1	G	H	Cot	O	L
Repeat 2	H	G	O	L	Cot
Repeat 3	Cot	L	G	H	O
Repeat 4	O	Cot	L	G	H

Table 1. Experimental design used for microarray hybridization.

An incomplete Latin Square design, using five different slides and four technical repeats, was used for our analysis.

Cot: Cotyledonary embryos, G: Globular embryos, H: Heart-shaped embryos, L: Leaves, O: Unfertilized ovules.

Source of variation	Df	Sum sq	Mean Square	Fvalue	Pr(>F)
Treatment	4	3718	930	7481.1	< 2.2e-16***
Membrane	4	1309	327	2632.7	< 2.2e-16***
Technical repeat	3	15974	5325	42851.9	< 2.2e-16***
Residuals	378548	47038	0.1243		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 0

Table 2. Analysis of variance table for the normalization model

Differential Clones (p value threshold = $5.42 \cdot 10^{-6}$)					Non Differential	Total of analysed clones
Leaves vs Reproductive Tissues	Unfertilized Ovules vs Embryos	Globular vs Heart-shaped Embryos	Early Embryos vs Cotyledonary	In at least 2 conditions among all		
437 (335/102)	116 (44/72)	159 (5/154)	187 (36/151)	744	8820	9464
FDR $1.17 \cdot 10^{-4}$	FDR $4.42 \cdot 10^{-4}$	FDR $3.23 \cdot 10^{-4}$	FDR $2.74 \cdot 10^{-4}$	FDR $6.89 \cdot 10^{-5}$		

Table 4: Summary of differential clones discovery using ANOVA.

In parenthesis: number of clones declared to be overexpressed in (condition 1 / condition 2).

Table 5. Fold Changes of mRNA level of selected differentially expressed clones obtained by quantitative RT-PCR and microarray analysis.

Up-regulated genes in heart- shaped embryos <i>versus</i> globular embryos			
Accession Number	Functional Annotation	Quantitative RT-PCR Fold Change	Microarray Fold Change
CD848057	PSI subunit	5.49	5.41
CX944786	bHLH transcription factor	342.41	5.22
CD857578	calcium-dependent protein kinase	4.08	16.41
AJ829314	unnamed protein product	8.33	6.36
CX944569	senescence related protein	2.1	12.52
Up-regulated genes in cotyledonary embryos <i>versus</i> early embryos			
Accession Number	Functional Annotation	Quantitative RT-PCR Fold Change	Microarray Fold Change
CD858195	Sec1 Keule cytokinesis-related protein	47.49	3.48
CD849586	EF1alpha	1.63	2.77
CX947490	glycinamide ribonucleotide synthetase	3.09	3.16
Up-regulated genes in unfertilized ovules <i>versus</i> embryos			
Accession Number	Functional Annotation	Quantitative RT-PCR Fold Change	Microarray Fold Change
-	Calmoduline	1.27	2.36

After normalization of the expression level of the target gene β -Actin as an internal control, the lowest expression level of each clone in the tested tissues was used to determine the fold change of transcript abundance.

Chapitre III – ‘Collection EST’ – ‘Analyse Microarrays’ : Atouts et contraintes des deux approches transcriptomiques utilisées.

Au cours d'études visant à caractériser le programme génétique induit lors d'un processus biologique particulier (par exemple au sein d'un tissu, à un stade de développement ou dans des conditions de culture spécifiques), la création d'une collection d'EST et l'analyse du profil d'expression de gènes par microarrays sont deux approches étroitement liées et véritablement complémentaires. La construction de banques d'ADNc à partir des tissus d'intérêt fournit les outils essentiels à la réalisation des microarrays, notamment dans le cas d'études menées sur des espèces pour lesquelles peu de séquences sont disponibles, et permet d'optimiser les chances de tester les gènes jouant un rôle clé dans le processus étudié. Le séquençage d'une fraction représentative des clones et l'analyse *in silico* des EST obtenues permet généralement de mettre en évidence les principales voies métaboliques et de régulation engagées. Les expériences de microarrays subséquentes, au cours desquelles un nombre beaucoup plus important de gènes peut être testé en introduisant, par exemple, sur la puce à ADN des EST issues d'autres tissus, permettent une étude plus globale et une quantification plus précise de l'activité transcriptionnelle induite au cours du phénomène biologique analysé.

Cependant, bien que ces techniques permettent d'appréhender le processus biologique étudié dans sa globalité en fournissant un premier aperçu de l'ensemble des fonctions cellulaires et voies de régulation impliquées, ces méthodes transcriptomiques connaissent également quelques limites.

I. La collection d'EST : une ressource précieuse mais ne donnant qu'un premier aperçu du profil fonctionnel.

I.1. Une avancée vers une meilleure connaissance du génome du tournesol.

A ce jour, plus de 66 000 EST de tournesol, issues de banques d'ADNc réalisées à partir de tissus d'origines très diverses, sont déjà disponibles dans les banques de données publiques. En fournissant des renseignements sur la séquence des gènes, cette collection d'EST, représentant environ 20 000 gènes potentiellement différents, procure d'importantes informations concernant le génome de cette plante dont le séquençage complet n'est pas envisagé. Des EST représentant différentes copies d'un même gène donnent par exemple des indices sur les familles multigéniques.

Suite à une recherche de similarités avec des protéines aux rôles biologiques connus, l'approche EST permet de décrire les fonctions des gènes différentiellement exprimés dans diverses conditions et d'imaginer ainsi les processus de régulation physiologique sous-jacents mis en œuvre par la plante, dans différents tissus, à

différents stades de développement ou en réponse à différentes conditions de culture. Ainsi, l'analyse d'EST *in silico* a été homologuée comme méthode permettant d'évaluer les profils d'expression des gènes dans différentes conditions (Ewing et al., 1999 ; Stekel et al., 2000).

De plus, certaines EST ne présentant aucune homologie avec des séquences déjà répertoriées peuvent constituer des gènes spécifiques au tournesol.

I.2. Une ressource précieuse pour l'étude de l'embryogenèse chez les plantes dicotylédones exalbuminées.

Les banques d'ADNc créées au cours de cette thèse s'appliquant aux stades très précoces du développement constituent une ressource unique tant pour le tournesol que pour l'ensemble des autres espèces végétales et peuvent notamment être utilisées pour l'étude de l'embryogenèse chez les autres plantes dicotylédones exalbuminées.

De plus, le pourcentage élevé de nouveaux gènes identifiés pour chacune des banques semble confirmer l'importance de créer des banques représentant des stades de développement très proches dans le temps afin de favoriser la découverte de nouveaux gènes. En effet, cette stratégie permet par exemple de mettre en évidence des gènes jouant un rôle clé au cours du développement mais dont l'expression est transitoire.

I.3. Les limites de l'approche EST.

Bien que l'approche EST ait largement convaincu par ces nombreux atouts et ait été rapidement adoptée chez un grand nombre d'organismes et pour l'étude d'une large gamme d'évènements physiologiques, cette méthode revêt toutefois quelques inconvénients qui peuvent en limiter l'utilisation.

- Tout d'abord, l'obtention et l'analyse de banques d'EST est un procédé relativement complexe exigeant un certain *savoir-faire technique* et beaucoup de soin, cette difficulté étant encore accentuée lorsque le matériel végétal de départ est difficile d'accès et disponible qu'en petites quantités comme dans le cas de l'embryogenèse végétale.

- De plus, étant donné le coût non négligeable du séquençage, le nombre d'EST générées est généralement limité n'offrant souvent qu'une *image restreinte* de l'expression génique dans la condition testée.

- La *redondance* des gènes fortement exprimés, bien qu'elle fournisse d'importants indices quant au profil d'expression génique dans le tissu d'origine, limite la découverte de gènes nouveaux. Celle-ci peut être en partie éliminée par la technique SSH qui permet de mieux cibler le séquençage grâce à la normalisation, le séquençage des gènes fortement exprimés étant ainsi réduit alors que la probabilité d'accéder aux gènes rares augmente. Toutefois, cette technique interdit toute analyse ultérieure du patron d'expression des gènes. Afin de remédier à ce problème pour les banques de référence les plus redondantes, nous avons adopté une autre stratégie consistant à réaliser un premier séquençage 'test' visant à rechercher les séquences

redondantes puis à procéder, si nécessaire, à un nettoyage de la banque par criblage des séquences les plus abondantes avant de compléter le séquençage.

- Un autre inconvénient de l'approche EST est l'éventualité de générer des *séquences de faible qualité* du fait du séquençage unique pour chaque EST pouvant entraîner un taux important d'erreur. Ainsi, les EST peuvent contenir des substitutions, des délétions ou des insertions comparées à la séquence d'ADN génomique d'origine, la région la plus fiable de l'EST étant située entre 100 et 300 bp. De même, la longueur moyenne des EST étant d'environ 400 bp, les EST ne représentent généralement que des fragments de gènes et non la séquence codante complète. Il existe également des risques de chimères correspondant à des EST où la région 5' montre des homologies avec une séquence et la région 3' avec une autre. De même, malgré un amorçage de la synthèse de l'ADNc au niveau de l'oligo dT au cours de la construction de la banque visant à sélectionner les ARNm, on peut constater dans de rares cas une contamination par des séquences mitochondriales, chloroplastiques ou bactériennes (vecteur de clonage).

- Une autre limitation de cette approche pour l'élucidation des voies de régulation transcriptionnelle impliquées dans une situation physiologique particulière vient du fait que les EST ne représentent que la partie codante des gènes et ne donnent *pas l'accès aux éléments régulateurs* des promoteurs qui sont uniquement disponibles dans l'ADN génomique.

- Enfin, les séquences de type '*no hit*' ne présentant aucune homologie avec des séquences connues, bien que très intéressantes pour la découverte de gènes potentiellement intéressants, ne permettent pas de déterminer les fonctions cellulaires principales mises en jeu au niveau d'un tissu. Or, ces séquences représentent un pourcentage très élevé pour chacune de nos banques (entre 21,9% pour HaCotR et 56,2% pour HaHeaS). Des analyses de clustering les associant grâce à leur profil d'expression avec des séquences de fonctions connues permettent de leur attribuer une fonction potentielle par extrapolation et de décider s'il est judicieux de les étudier plus en détail.

II. L'approche microarrays fournit une information plus complète et plus précise.

II. 1. Vers la compréhension globale des mécanismes moléculaires impliqués lors de l'embryogenèse précoce du tournesol grâce à l'utilisation des microarrays.

Les microarrays permettent une étude beaucoup plus exhaustive et plus précise du transcriptome que les approches EST. En effet, outre le fait d'analyser l'expression d'un grand nombre de gènes différents voire d'envisager l'étude de génomes complets, cette technologie permet une quantification directe et fiable du niveau d'expression des gènes comme le confirme la validation des résultats par RT-PCR quantitative.

De plus, l'information obtenue à l'issue d'expériences de microarrays est riche et touche trois niveaux différents de la description des mécanismes moléculaires impliqués dans le processus biologique étudié.

- Les analyses microarrays permettent tout d'abord de mettre en évidence les gènes différentiellement exprimés dans chacune des conditions investiguées. La caractérisation fonctionnelle des gènes surexprimés dans une condition permet de décrire les différentes fonctions cellulaires impliquées dans cette condition particulière. La mise en évidence de gènes spécifiques pour une condition permet également d'identifier des gènes marqueurs.

- Dans un second temps, la comparaison des patrons d'expression des gènes dans les différentes conditions et l'assemblage des séquences présentant des profils similaires au sein de groupes fonctionnels en utilisant les méthodes de clustering permettent de décrire les gènes co-exprimés et ainsi de visualiser les interactions et les mécanismes de co-régulation. Le clustering permet également d'attribuer une fonction putative aux gènes de fonction inconnue.

- Enfin, l'analyse microarray est une approche globale pour la caractérisation des mécanismes moléculaires impliqués dans un processus biologique permettant d'appréhender la question scientifique dans sa globalité en dégagant des hypothèses concernant le réseau de gènes et de protéines au niveau des systèmes biologiques.

II. 2. Amélioration de la découverte de gènes clés du développement embryonnaire végétal.

Bien que déjà très riche, l'unigène déposé sur le microarray construit pour cette étude ne contient pas la totalité de l'unigène disponible pour le tournesol dans les bases de données de séquences publiques. En effet, environ 20 000 gènes différents ont déjà été calculés soit plus de deux fois plus que l'unigène présent sur notre microarray. L'objectif serait donc d'élargir le nombre de gènes présents sur la puce à ADN à l'ensemble de l'unigène actuel afin d'obtenir une image complète des mécanismes moléculaires impliqués dans l'embryogenèse précoce et d'affiner les hypothèses de travail en élucidant les zones d'ombre qui persistent encore. Les origines diverses de ces séquences issues de banques d'ADNc réalisées à partir de conditions et de tissus différents devrait permettre d'enrichir notre outil en gènes impliqués dans de nouvelles voies métaboliques non présentes sur notre microarray et ainsi de révéler de nouvelles voies de régulation non dévoilées par notre étude. L'une des solutions serait de construire une puce à oligos représentant la totalité de l'unigène actuellement disponible. Cependant, ce procédé a un coût très élevé et nécessite des mises au point techniques importantes.

Toutefois, au cours de nos analyses, les gènes apparus comme différentiellement exprimés entre les différents stades embryonnaires provenaient plus particulièrement de banques d'ADNc produites à partir de tissus embryonnaires, notamment en ce qui concerne le stade globulaire. Ainsi, les autres banques d'ADNc de tournesol publiques ayant majoritairement été construites à partir de tissus somatiques classiques ne constituent donc peut être pas la ressource la plus appropriée pour identifier des gènes clés de

l'embryogenèse précoce. Une solution plus économique et certainement plus adaptée pour la découverte de gènes candidats pertinents serait donc de compléter l'unigène spotté sur notre microarray en ajoutant des clones non séquencés issus de banques d'ADNc préparées à partir d'embryons à des stades précoces et présentant un faible taux de redondance (forte probabilité d'isoler de nouveaux gènes).

II. 3. Construction d'un outil à la disposition de la communauté scientifique.

Le microarray que nous avons construit, qui contient 8 185 séquences uniques représentant une large gamme de fonctions cellulaires différentes, constitue un outil précieux à la disposition de la communauté scientifique. Son utilité et sa fiabilité ont été prouvées au cours de nos analyses visant à caractériser le transcriptome au cours de l'embryogenèse précoce chez le tournesol. Ce microarray a d'ores et déjà été employé pour caractériser les programmes génétiques induits chez le tournesol en réponse à des stress abiotiques (température et lumière) ainsi que pour identifier les gènes exprimés lors du remplissage du grain dans des conditions physiologiques contrastées (programme Génoplante GOP-HG01).

PARTIE III

PERSPECTIVES : VALIDATION POST-GENOMIQUE

L'étude réalisée au cours de cette thèse a permis d'obtenir un premier aperçu des principales fonctions cellulaires impliquées lors de l'embryogenèse précoce chez le tournesol et d'isoler des gènes candidats susceptibles de jouer un rôle clé dans ce processus biologique crucial. Cependant, la caractérisation du transcriptome induit au cours des phases précoces de la formation de la graine, grâce à l'analyse d'une collection d'EST obtenues à partir d'embryons zygotiques à différents stades de développement et à l'étude du profil d'expression de plus de 8000 gènes différents à l'aide de la technologie microarray, ne constitue que la première étape vers le décryptage des voies fonctionnelles et du réseau de régulation transcriptionnel engagés dans ce processus biologique. Si ces travaux ont d'ores et déjà permis de dégager un certain nombre d'hypothèses concernant la régulation moléculaire du développement embryonnaire précoce chez les plantes, celles-ci devront être validées chez le tournesol puis testées chez d'autres modèles végétaux afin d'évaluer le degré de conservation de ces mécanismes parmi différentes espèces végétales.

Par ailleurs, de larges zones d'ombre permettant de faire le lien entre les différents mécanismes et voies de régulation potentiellement impliqués subsistent et devront être élucidées afin de permettre une compréhension globale des événements moléculaires et physiologiques ayant lieu lors des phases précoces de l'embryogenèse végétale.

Chapitre I – Nécessité de développer des outils bioinformatiques adaptés à la gestion de large volume de données et à la découverte de gènes candidats pertinents.

I. Développement de programmes informatiques d'analyse plus performants.

Les analyses transcriptomiques utilisant la technologie microarray qui permet d'étudier le profil d'expression de plusieurs milliers de gènes simultanément sont à l'origine d'un très grand nombre de données. Nos expériences menées sur plus de 10 000 EST et visant à caractériser les gènes différemment exprimés au cours des phases précoces de l'embryogenèse chez le tournesol ont généré pas moins de 432 640 signaux d'intensité différents, ce nombre pouvant encore largement s'alourdir dans le cas d'expériences testant un plus grand nombre de conditions et obéissant à un dispositif expérimental plus complexe. Bien qu'il existe déjà un certain nombre d'outils permettant une automatisation des analyses (possibilité d'effectuer les analyses bioinformatiques en séries, programmes d'annotation automatique de séquences ...), les procédures d'analyses statistiques restent lourdes et exigent une certaine expertise bioinformatique. Etant donné le succès grandissant que rencontre ce type d'approche, il apparaît donc nécessaire de développer un certain nombre d'outils informatiques de calcul plus puissants, plus efficaces, et plus accessibles.

II. Nécessité d'une mutualisation des données disponibles et d'une uniformisation de l'annotation fonctionnelle des séquences.

II.1. Archivage des données d'expression génique au sein de bases de données publiques.

Un autre problème rencontré face à ces quantités importantes de données est la gestion et l'archivage des informations concernant les profils d'expression générées par ces analyses. Un grand nombre de bases de données indiquant les résultats d'expériences de microarrays menées sur différentes espèces végétales telles que GENEVESTIGATOR, BarleyExpress ou encore SGMD (Soybean Genomics and Microarray Database) sont déjà disponibles *via* internet (Alkharouf and Matthweus, 2004 ; Tang et al., 2005 ; Zimermann et al., 2004). Des banques de données plus généralistes cataloguant l'ensemble des données d'expression génique obtenues chez diverses espèces ont également été développées comme

par exemple les bases des données ArrayExpress et NCBI GEO (Barrett et al., 2005 ; Parkinson et al., 2005). Si de telles initiatives, offrant un accès libre et facile à l'ensemble des données de profils transcriptionnels afin de les confronter aux résultats de nos propres analyses et de les intégrer à nos réflexions, doivent impérativement être poursuivies, la comparaison entre différentes expériences de microarrays n'est pas toujours évidente. En effet, étant donné le large éventail de procédés techniques disponibles pour réaliser ces analyses (type de puce à ADN utilisé, dispositif expérimental mis en œuvre, chimie de marquage des sondes et procédure d'hybridation, méthode d'analyse statistique), une uniformisation des protocoles expérimentaux semble indispensable afin que les résultats obtenus soient comparables entre eux.

II.2. Vers une uniformisation de l'annotation fonctionnelle des séquences.

L'un des objectifs finaux des analyses microarrays est la sélection, parmi les gènes différentiellement exprimés, de gènes candidats susceptibles de jouer des rôles clés dans le processus investigué. Ce choix nécessite tout d'abord l'utilisation de méthodes d'analyse statistique adaptées pour les données de microarrays limitant notamment le taux de faux positifs. La sélection des gènes candidats les plus pertinents méritant une étude plus poussée exige également une recherche préalable de renseignements concernant la fonction putative du gène ou son profil d'expression dans différentes conditions ainsi qu'une connaissance au niveau moléculaire des homologues identifiés chez d'autres espèces ou des autres copies du gène dans le cas de famille multigénique. Il est également intéressant de disposer d'informations concernant les gènes co-régulés. Cependant, du fait de la dispersion des données existantes et de la multiplication des langages d'annotation fonctionnelle, cette étape reste extrêmement laborieuse. Ainsi, afin de faciliter les recherches, une homogénéisation du vocabulaire et des critères de classification fonctionnelle des séquences entraînant une meilleure lisibilité des données disponibles est indispensable. Des démarches dans ce sens ont déjà été entreprises avec, par exemple, la création de la banque de protéines non redondante PIR (Protein Information Resource) regroupant l'ensemble des informations concernant toutes les protéines décrites (Wu et al., 2002) ou encore l'élaboration par le Gene Ontology Consortium (GO) (<http://www.geneontology.org>) d'une nomenclature standardisée pour l'annotation de gènes indiquant la fonction moléculaire, le processus biologique et le composant cellulaire (Ashburner et al., 2000), cette nomenclature pouvant être notamment utilisée pour l'analyse automatisée des profils d'expression des gènes (Volinia et al., 2004). Un autre projet original de regroupement d'informations est le projet MedBlast (Tu et al., 2004) permettant d'accéder à toute la bibliographie évoquant une séquence particulière ou encore le projet développé chez *Arabidopsis thaliana* par Louis (Louis et al., 2002) associant des gènes en fonction de leur apparition commune dans la littérature selon l'hypothèse que des séquences citées conjointement sont susceptibles d'être engagées dans un même mécanisme. Ces initiatives doivent être poursuivies

dans l'espoir d'améliorer la sélection des gènes candidats et ainsi de mieux appréhender le problème biologique posé.

La difficulté pour l'annotation des EST est encore plus sensible pour les séquences ne montrant aucune homologie avec les séquences de fonction connue répertoriées dans les banques de données publiques (no hits). Ainsi, malgré le séquençage complet du génome de plantes modèles telles qu'*Arabidopsis thaliana*, on se heurte encore à l'existence de ce type de séquences. Malgré la possibilité de rechercher au sein de la séquence des motifs protéiques au rôle connu, les chances d'attribution d'une fonction à ces séquences sont limitées. Celles-ci, susceptibles d'être spécifiques du modèle végétal étudié, pourraient pourtant constituer d'excellents candidats.

Chapitre II – Perspectives concernant l'étude des mécanismes moléculaires contrôlant l'embryogenèse précoce chez le tournesol.

I. Vers la découverte de nouveaux gènes et mécanismes clés : Elargissement du nombre de gènes et de conditions testées.

Le séquençage complet du tournesol n'étant pas encore envisagé, l'espoir de décrypter un jour les profils d'expression de l'ensemble des gènes de cette plante dans différentes conditions est faible. Toutefois, comme nous l'avons déjà évoqué, nous disposons actuellement au laboratoire de plusieurs milliers de clones non séquencés issus de banques d'ADNc préparées à partir de tissus embryogènes et d'environ 20 000 séquences de tournesol potentiellement uniques répertoriées dans les banques de données publiques, ces ressources pouvant être utilisées pour enrichir l'unigène déposé sur le microarray construit au cours de cette étude, afin de tenter d'élucider les zones d'ombre et d'affiner les hypothèses de travail.

De nouvelles expériences de microarrays réalisées à partir d'embryons à d'autres stades de développement ou soumis à différents traitements peuvent également être envisagées. En effet, bien qu'à notre connaissance, le travail réalisé au cours de cette thèse constitue le premier projet d'analyse du transcriptome à des stades précis de l'embryogenèse précoce chez les dicotylédones, les stades de développement et la période de l'embryogenèse analysés peuvent encore être affinés et élargies, notamment en essayant de caractériser le programme génétique induit à des phases plus précoces du développement et constituant des étapes critiques de la formation du patron embryonnaire tels que le stade 16 cellules (ou dermatogène) correspondant à la mise en place du protoderme, le stade octant au cours duquel sont définis les domaines apicaux et centraux du proembryon ou même au stade deux cellules où s'établissent les devenir cellulaires de l'embryon et du suspenseur. De tels projets pourraient être réalisables en utilisant la technique de microdissection au laser (LCM, Laser Capture Microdissection) qui a déjà permis l'étude du transcriptome induit dans les domaines basal et apical d'embryon d'*Arabidopsis thaliana* au stade globulaire et cœur (Casson et al., 2005). Des analyses de l'expression génique induite dans des embryons prélevés à partir de plantes cultivées en conditions de stress (température, lumière, sécheresse, attaque de pathogènes...) pourraient également être menées afin de tenter d'isoler des gènes exprimés en réponse à des stress au cours de l'embryogenèse.

Enfin, le criblage de la collection de 2000 mutants d'irradiation produite au laboratoire à la recherche de plantes montrant des altérations du développement embryonnaire selon les critères phénotypiques définis par Meinke (Meinke, 1985) pourrait permettre d'identifier des gènes clés du développement embryonnaire.

II. Analyse moléculaire et fonctionnelle des gènes candidats identifiés : Exemple du gène MAGO NASHI.

La caractérisation moléculaire et fonctionnelle des gènes candidats chez le tournesol et la validation, sur notre modèle d'étude, des hypothèses dégagées suite à l'analyse de la collection d'EST et des résultats de microarrays, bien qu'absolument indispensables, restent difficiles du fait du caractère récalcitrant de cette plante à de nombreuses techniques de biologie moléculaire (difficulté pour la transformation génétique et pour la régénération *in vitro* à partir d'explants...) et du manque d'outils et de connaissances moléculaires et physiologiques concernant cette espèce végétale (faible quantité de séquences disponibles, peu de mutants...).

Toutefois, un certain nombre de démarches peut être entrepris dans le but de poursuivre les recherches visant à caractériser les gènes induits au cours de l'embryogenèse précoce chez le tournesol comme l'illustrent les travaux préliminaires concernant la description moléculaire et fonctionnelle du gène MAGO NASHI menés au sein de notre laboratoire.

Pourquoi avoir choisi d'étudier le gène MAGO NASHI de manière plus détaillée?

MAGO NASHI est l'un des gènes potentiellement intéressants mis en évidence parmi les séquences EST issues des banques d'ADNc réalisées durant cette thèse. Initialement décrit chez la Drosophile, le gène MAGO NASHI appartient au groupe de gènes postérieurs à effet maternel jouant un rôle essentiel dans le développement embryonnaire. De plus, cette protéine est hautement conservée parmi les organismes pluricellulaires animaux et végétaux avec plus de 60% des acides aminés conservés parmi les homologues identifiés. Ce haut degré de conservation entre animaux et végétaux peut laisser penser que les rôles de la protéine ont pu être conservés entre les deux règnes.

II.1. Présentation de MAGO NASHI : principales caractéristiques moléculaires.

II.1.1. MAGO NASHI : une protéine hautement conservée au sein des différentes espèces animales et végétales.

Les protéines prédites à partir des ADNc des différents homologues de MAGO NASHI déjà identifiés comptent environ 150 acides aminés et ont un poids moléculaire estimé compris entre 17 et 19 kDa (Annexe 1). Les régions les plus variables de la protéine se trouvent dans la partie N-terminale riche en acides aminés chargés ainsi que dans une région centrale d'environ 30 acides aminés. La région C-terminale est relativement hydrophobique. Plusieurs résidus Leucine, pouvant constituer des domaines leucine zipper impliqués dans la liaison à des protéines, sont conservés entre les différents

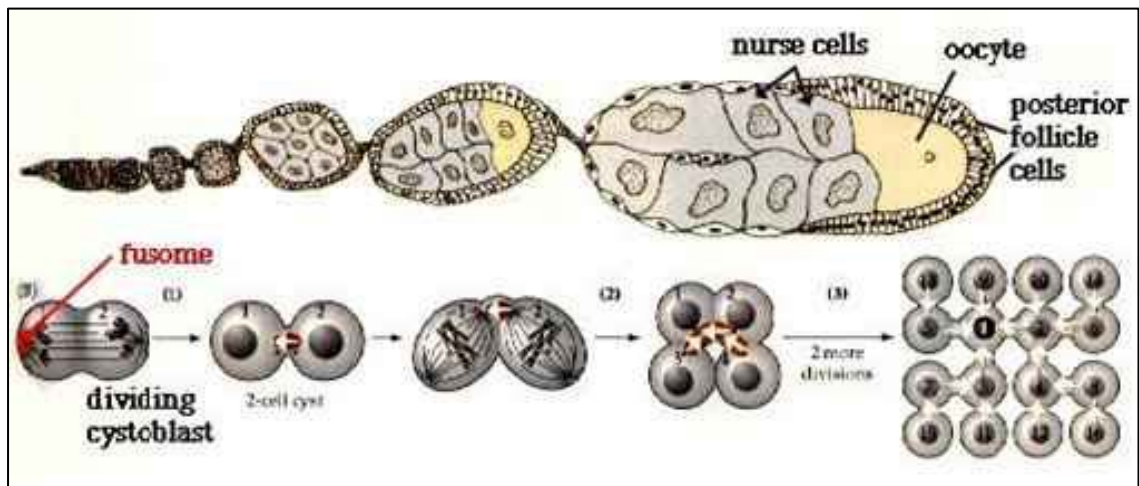


Figure 32 : Oogenèse chez la *Drosophile* (D'après Gilbert, 1997).

homologues. Fait surprenant, aucune homologie avec toute autre protéine ou domaine protéique de fonction biologique connue n'a pu être mise en évidence.

II.1.2. MAGO NASHI : une protéine ubiquitaire.

Que ce soit chez les animaux ou les végétaux, ARNm et protéine de MAGO NASHI sont ubiquitaires : ils sont présents dans tous les types de tissus et à tous les stades de développement. Ainsi, même si MAGO NASHI a tout d'abord été isolé à partir d'embryons de *Drosophile* et peut être identifié dans de nombreuses banques d'EST de tissus embryonnaires animaux et végétaux, des expériences de Northern Blots et d'immunolocalisation ont montré que l'expression de ce gène n'est pas limitée aux tissus embryonnaires et que les ARNm et la protéine de MAGO NASHI sont aussi synthétisés dans tous les tissus adultes analysés des différentes espèces étudiées.

Ainsi, en plus de son rôle probable dans l'oogenèse et le développement embryonnaire, le patron d'expression ubiquitaire de MAGO NASHI suggère que cette protéine pourrait avoir d'autres fonctions dans les tissus adultes. Chez la souris, par exemple, l'expression des ARNm de MAGO NASHI apparaît plus prononcée dans des tissus proliférant rapidement (thymus, moëlle osseuse, placenta). L'expression de MAGO NASHI pourrait donc être liée à la prolifération cellulaire selon un mécanisme encore inconnu (Zhao et al., 1998).

II.2. Fonctions de MAGO NASHI chez les animaux.

II.2.1. MAGO NASHI : une protéine essentielle au développement de l'oocyte chez la Drosophile.

La diversification cellulaire au cours du développement requiert souvent la localisation subcellulaire particulière d'indices moléculaires (ARN, protéines) qui servent à induire la différenciation de cellules génétiquement identiques en types cellulaires distincts. Chez la *Drosophile*, des asymétries générées pendant l'oogenèse fournissent l'information nécessaire pour l'établissement d'axes antéro-postérieur (AP) et dorso-ventral (DV) de l'oocyte et de l'embryon.

II.2.1.1. MAGO NASHI est impliqué dans la formation des axes antéro-postérieur et dorso-ventral de l'oocyte de Drosophile.

Chez la *Drosophile*, l'oogenèse est initiée lorsqu'une cellule souche de la lignée germinale se divise asymétriquement pour donner une cellule souche fille et un cystoblaste (Fig.32). Le cystoblaste subit 4 divisions mitotiques avec une cytokinèse incomplète pour former un cyste de 16 cellules dérivées de la lignée germinale et interconnectées par des ponts cytoplasmiques. Une couche de cellules somatiques

folliculaires entoure le cyste de 16 cellules pour former la chambre ovarienne. L'une des 2 cellules du cyste ayant 4 ponts cytoplasmiques se développe en oocyste et se place en position postérieure par rapport aux 15 autres cellules qui se différencient en cellules nurses. Cet arrangement de cellules de la lignée germinale génère l'asymétrie initiale qui mène à la polarisation de l'axe AP de l'oocyte. Cependant, de nombreuses étapes au cours de l'oogenèse sont nécessaires pour établir correctement les axes AP et DV de l'oocyte. Leur formation requiert notamment une série d'interactions inductives entre l'oocyte et les cellules somatiques folliculaires environnantes.

Au cours de l'oogenèse, après s'être placé en position postérieure par rapport aux cellules nurses dans la chambre ovarienne, l'oocyte exprime la protéine GURKEN (GRK), un membre de la famille TGF α (Transforming Growth Factor α), qui induit les cellules folliculaires adjacentes pour adopter un devenir de cellules postérieures. Ces cellules renvoient alors un signal non identifié à l'oocyte. Le gène MAGO NASHI serait impliqué dans les mécanismes de transduction de ce signal qui induit la réorganisation du cytosquelette de microtubules et ainsi la formation d'un gradient antéro-postérieur de microtubules au sein de l'oocyte définissant l'axe AP. Ces microtubules polarisés permettent alors la migration du noyau de l'oocyte et des ARNm GRK exprimés en périphérie du noyau d'une position postérieure vers une position antérieure dans l'oocyte. La protéine GRK alors synthétisée dans la région antéro-corticale de l'oocyte envoie un second signal vers les follicules environnants pour induire une position dorsale de ces cellules, établissant ainsi l'axe DV (Mickleman et al., 1997 ; Mohr et al., 2001 ; Newmark et al., 1997).

II.2.1.2. MAGO NASHI permet l'ancrage de déterminants maternels au pôle postérieur de l'oocyte et l'assemblage du pôle germinatif chez la Drosophile.

Le cytosquelette polarisé permet en outre de guider la localisation de déterminants abdominaux, les ARNm OSKAR (OSK) vers le pôle postérieur de l'oocyte appelé pôle germinatif. Une fois localisés, les ARNm OSK dirigent l'assemblage des composants du pôle germinatif et la localisation des déterminants de segmentation abdominale NANOS.

Cette réorganisation de déterminants cytoplasmiques mène à la différenciation cellulaire pendant l'embryogenèse.

Les fonctions de MAGO NASHI de transduction du signal 'cellules folliculaires postérieures-oocyte' avec la formation d'un réseau de microtubules polarisé et de localisation de déterminants cytoplasmiques au pôle postérieur semblent totalement indépendantes et diffèrent par leurs besoins d'activité du gène MAGO NASHI : le transport d'ARNm OSK est plus sensible à une réduction de la quantité de protéine MAGO NASHI alors que la transduction du signal requiert moins d'activité de la protéine (Mickleman et al., 1997).

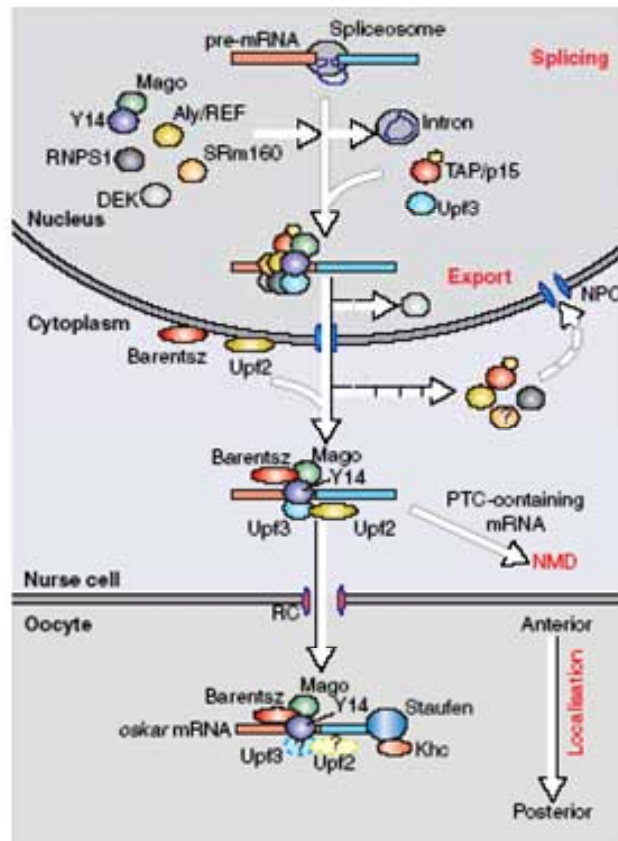


Figure 33: MAGO NASHI, un membre d'un complexe ribonucléoprotéique de jonction exon-exon: modèle d'assemblage et de fonctionnement.
(d'après Palacios, 2002).

Suite à l'épissage des pré-ARNm, les 6 composants du complexe de jonction exon-exon (Mago, Y14, RNPS1, DEK, SRM160 et Aly/REF) sont déposés en aval de la jonction exon-exon de l'ARNm. Au niveau du nucléoplasme, le complexe recrute Upf3 et le récepteur d'export hétérodimérique TAP/p15. Après translocation à travers les complexes de pores nucléaires (NPC), plusieurs composants du mRNP se dissocient et sont recyclés vers le noyau. En revanche, en l'absence de traduction, Y14, Mago et Upf3 semblent rester liés à l'ARNm. Upf2, un facteur de dégradation des ARNm non-sens, et Barentsz, une protéine requise pour la localisation des ARNm *oskar*, sont recrutés après la translocation dans le cytoplasme. Les ARNm contenant un codon de terminaison prématuré (PTC) sont dégradés alors que l'ARNm *oskar* est transporté du pôle antérieur vers le pôle postérieur de l'oocyte grâce à la liaison de composants de localisation spécifiques tels que Staufen et une kinésine chaîne lourde (Khc).

II.2.2. Rôles de MAGO NASHI chez d'autres espèces animales.

La caractérisation de MAGO NASHI chez d'autres espèces animales suggère que le rôle de MAGO NASHI dans la localisation des ARNm pourrait ne pas être limitée à la Drosophile. Chez *C. elegans*, MAGO NASHI est impliqué dans la détermination du sexe de lignées germinales hermaphrodites en stimulant l'oogenèse par le biais de l'inhibition d'un ou plusieurs gènes masculinisants, peut-être en mobilisant des protéines de liaisons à l'ARN (Li et al., 2000). La complémentation d'un mutant *mago nashi* de Drosophile par le gène MAGO NASHI de *C.elegans* démontre que la protéine MAGO NASHI de *C. elegans* constitue bien un homologue fonctionnel. Chez l'homme, la protéine MAGO NASHI interagit avec la protéine de liaison à l'ARN Y14 décrite comme essentielle dans la localisation d'ARN (Kataoka et al., 2001).

Des analyses du transcriptome induit lors de la spermatogenèse chez la truite ont identifié le gène MAGO NASHI comme un gène cible potentiel pour le contrôle de la puberté et de la fertilité. Une étude plus poussée visant à décrire le rôle et le mode d'action de ce gène au cours de la maturation du testis de la truite est en cours (Mazurais et al., 2005). MAGO NASHI pourrait également jouer un rôle au cours de l'embryogenèse chez le poisson zèbre (Zebrafish) où une analyse d'expression a révélé que le transcrit d'origine maternelle serait transporté, immédiatement après la fécondation, de l'hémisphère végétal vers le pôle animal probablement par le biais du réseau de microtubules. Le domaine d'expression est ensuite restreint progressivement au niveau du tube neural, des structures paraxiales et ventralement au niveau des canaux pronéphrétiques (Pozzoli et al. 2004). Toutefois, la fonction de MAGO NASHI au cours du développement embryonnaire du poisson zèbre reste encore à élucider.

II.3. Mécanismes d'action de MAGO NASHI chez les animaux.

II.3.1. MAGO NASHI : un constituant d'un complexe ribonucléoprotéique (mRNP).

Pour accomplir ses différentes fonctions, la protéine MAGO NASHI n'agit pas seule. Membre de complexes mRNP, elle interagit directement, avec d'autres partenaires protéiques tels que les protéines TAP et Y14 (codée par le gène RBM8 et homologue de TSUNAGI chez la Drosophile) mises en évidence chez l'homme (Kataoka et al., 2001).

MAGO NASHI serait une protéine navette entre le noyau et le cytoplasme de l'oocyte qui, par le biais du complexe mRNP, se lie de façon indirecte à certains ARNm épissés dans le noyau, reste accrochée pendant le transport et reste associée à l'ARNm nouvellement exporté dans le cytoplasme (Fig. 33). Ce modèle est en accord avec les résultats d'hybridation *in situ* montrant que MAGO NASHI est principalement situé dans le noyau mais aussi dans le cytoplasme de l'oocyte.

Si, contrairement aux protéines TAP et Y14, la protéine MAGO NASHI ne se lie pas directement aux ARNm épissés, sa liaison particulièrement forte avec ces 2 protéines suggère qu'elle doit avoir une grande influence sur l'interaction de ces facteurs avec les ARNm. MAGO NASHI pourrait servir d'adaptateur spécifique pour TAP et Y14, augmentant ainsi l'exportation d'ARNm épissés du noyau vers le cytoplasme.

Les complexes mRNP se lient aux ARNm épissés au niveau de régions situées 20 nucléotides en amont de jonctions exon-exon. Ces complexes pourraient servir d'indices déposés sur l'ARNm gardant la mémoire du site d'épissage et influençant l'exportation hors du noyau des ARNm en servant d'adaptateurs pour l'interaction des ARNm avec les microtubules du cytosquelette. Chez la Drosophile, le complexe MAGO NASHI/TSUNAGI est localisé au sein du cytoplasme du pôle postérieur de l'oocyte où il fonctionne tout d'abord pour interpréter le signal 'cellules folliculaires postérieures-oocyte'. Ce complexe serait ensuite impliqué dans le transport localisé de certains ARNm tels que les ARNm OSK puis, suite à la traduction de ces ARNm, il interviendrait dans l'assemblage et/ou le maintien de ces composants protéiques nécessaires à la détermination de la cellule germinale (Newmark and Boswell, 1994 ; Palacios, 2002).

Cependant, la conservation du complexe MAGO NASHI/Y14 chez les métazoaires et leur patron d'expression ubiquitaire suggèrent un rôle essentiel plus général de ces protéines. Comme composants du complexe jonction exon-exon, MAGO NASHI et Y14 sont susceptibles de s'associer à la plupart, sinon à tous les ARNm épissés et pourrait avoir une fonction plus générale dans le métabolisme post-épissage des ARNm tel que la dégradation cytoplasmique d'ARNm aberrants et le contrôle de la traduction (Le Hir et al., 2001 ; Palacios, 2002). Il a également été montré que les protéines MAGO NASHI et Y14 peuvent aussi assurer d'autres fonctions de manière indépendante.

II.3.2. Régulation post-transcriptionnelle de l'activité de MAGO NASHI.

Chez la Drosophile, il existe 2 types d'ARNm MAGO NASHI de tailles différentes (0.7kb et 1.1kb). Des expériences de RNA blots réalisées à partir d'ARN isolés à différents stades du cycle de vie de la Drosophile pour déterminer le patron d'expression dans le temps du gène ont montré que les 2 ARNm étaient détectés tout le long de la vie. Cependant, si les 2 ARNm sont détectés en même quantité au stade larvaire et adulte mâle et femelle, pendant les 4 premières heures de développement embryonnaire, les transcrits de 0.7kb sont plus abondants que les transcrits de 1.1kb. Ceci semble dû à des mécanismes de régulation post-transcriptionnelle utilisant les signaux de poly-adénylation, la région non traduite du transcrit 1.1kb contenant 3 sites AUUUA non présents chez les transcrits 0.7kb et impliqués dans l'instabilité des ARNm (Newmark and Boswell, 1994).

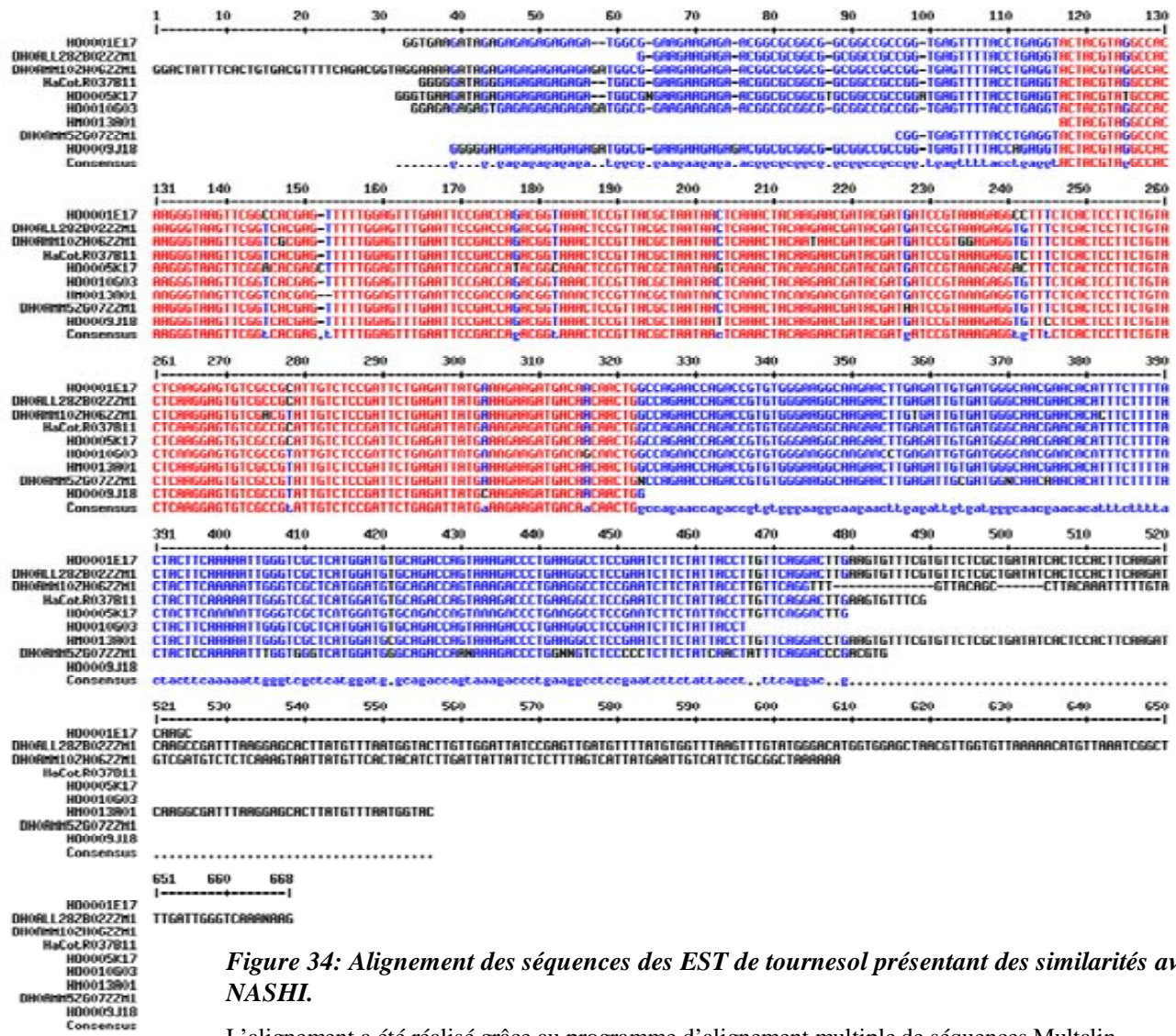


Figure 34: Alignement des séquences des EST de tournesol présentant des similarités avec MAGO NASHI.

L'alignement a été réalisé grâce au programme d'alignement multiple de séquences Multalin (<http://prodes.toulouse.inra.fr/multalin/multalin.html>)

II.4. Très peu de données sont disponibles concernant les fonctions des homologues végétaux de MAGO NASHI.

Très peu de données concernant la caractérisation des homologues végétaux de MAGO NASHI sont actuellement disponibles, la seule étude publiée portant sur l'homologue de ce gène chez le riz (Swidzinski et al., 2001). Au cours de ces travaux, des expériences de Southern blot ont montré qu'il existe probablement 2 copies du gène MAGO NASHI dans le génome du riz alors qu'il en existe une seule chez *Arabidopsis thaliana* située sur le chromosome 1. L'une des 2 copies du riz, dont la séquence a été décrite partiellement, possède 2 introns situés à la même position que ceux de l'homologue d'*Arabidopsis thaliana* mais à une position différente de celle de l'intron unique du gène de la Drosophile.

Des expériences de Northern blots et Western blots ont été réalisées pour décrire les profils d'expression du gène dans les feuilles, les racines et les graines. Un seul ARNm d'environ 0.9kb est présent en quantité semblable dans chacun des tissus. En revanche, la quantité en protéines est inférieure dans les racines par rapport aux feuilles et à la graine. Ceci peut être dû à une synthèse ou une stabilité réduites des protéines MAGO NASHI dans les racines.

Des expériences d'immunolocalisation avec des anti-corps dirigés contre la protéine MAGO NASHI du riz ont montré que celle-ci était principalement localisée dans le noyau. Toutefois, ces résultats doivent être considérés avec circonspection du fait d'un manque de spécificité des anti-corps utilisés.

Enfin, un homologue de RBM8/TSUNAGI montrant environ 60% d'identité en acides aminés sur une région de 80 acides aminés (le gène RBM8 humain codant pour une protéine prédite de 173 acides aminés) a été identifié dans la banque de données de séquences du riz.

Actuellement, les recherches en cours portent sur l'étude de la sur et sous expression de MAGO NASHI et de l'homologue du riz de RBM8 pour déterminer leur possible rôle dans la localisation des ARN dans les cellules végétales.

II.5. Vers la caractérisation de l'homologue de MAGO NASHI du tournesol : Démarches expérimentales entreprises et résultats préliminaires obtenus.

II.5.1. Vers la connaissance de l'organisation génomique du gène MAGO NASHI chez le tournesol : recherche du nombre de copies du gène MAGO NASHI chez le tournesol.

II.5.1.1. Criblage des banques d'ADNc et recherche d'EST in silico.

Après criblage des banques d'ADNc réalisées au laboratoire (embryons à différents stades de développement et protoplastes en conditions embryogènes ou non) à l'aide d'une sonde overgo

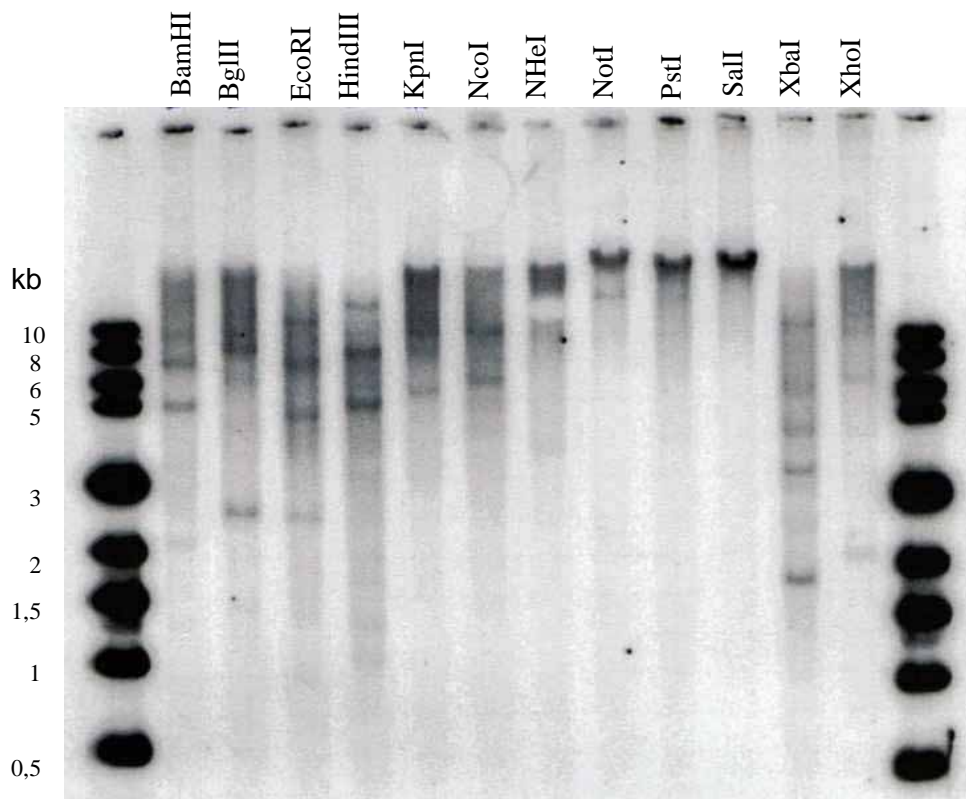


Figure 35: Expérience de Southern Blot réalisée à partir d'ADN génomique extrait du génotype PCS8 de Tournesol.

L'ADN a été digéré par 12 enzymes de restriction différentes et hybridé avec une sonde marquée radioactivement et préparée à partir de l'EST de tournesol HO001E17 homologue à MAGO NASHI.

désignée dans une région conservée du gène MAGO NASHI et recherche dans les bases de données de séquences de tournesol homologues au gène MAGO NASHI, 9 EST correspondant potentiellement à autant de copies différentes du gène ont pu être isolées (Fig. 34). Toutefois, cette estimation du nombre de copies du gène MAGO NASHI doit être pondérée car ces EST ont été obtenues à partir de 3 génotypes différents (PSC8 pour DH0ALL28ZB02ZZM1, DH0AMM5ZG07ZZM1 et DH0AMM10ZH06ZZM1; Emil pour HO0010G03, HM0013A01, HO0001E17, HO0005K17, HO0009J18 et Ha300b pour HaCotR037B11) et seule la séquence des 6 clones issus des banques produites au sein de notre laboratoire (HaCotR037B11, HO0010G03, HM0013A01, HO0001E17, HO0005K17, HO0009J18) a pu être vérifiée grâce à un deuxième séquençage. Cette validation a confirmé que les séquences de ces 6 clones étaient bien différentes.

Fait remarquable, l'ensemble de ces EST dérive de banques d'ADNc construites à partir de tissus embryogènes (embryons cotylédonaire, protoplastes cultivés en conditions embryogènes, embryons à 4 et 7 jours après fécondation). Cette observation conforte l'hypothèse d'un rôle clé du gène MAGO NASHI dans le développement embryonnaire du tournesol.

II.5.1.2. Des expériences de Southern Blots montrent qu'il existe plus d'une copie du gène.

Le profil d'hybridation obtenu grâce aux expériences de Southern Blots réalisées sur l'ADN génomique du génotype PSC8 qui comprend de 1 à 5 fragments de restriction selon l'enzyme utilisée pour la digestion de l'ADN confirmerait le fait que le génome du tournesol contiendrait plus d'une copie du gène MAGO NASHI (Fig. 35).

II.5.1.3. Criblage des banques BAC pour accéder à la séquence génomique des différentes copies du gène.

- **Criblage des banques BAC: 5 clones BAC positifs.**

Afin d'accéder à la séquence génomique des différentes copies du gène, deux banques BAC disponibles au laboratoire et construites respectivement à partir des génotypes Ha821 (Gentzittel et al., 2002) et PSC8 (non publiées) ont été criblées en utilisant une sonde overgo désignée au niveau d'une région conservée entre les différentes EST de tournesol présentant des similarités avec MAGO NASHI. Au total 5 clones BAC ont pu être isolés : il s'agit des clones 34G03 et 225K21 pour le génotype Ha821 et 13B07, 88L03 et 362P18 pour le génotype PSC8. Des expériences de Southern Blots menées sur ces clones BAC ont révélé que les clones 88L03 et 362P18 étaient recouvrants (Fig. 36). Par la suite, le sous-clonage et le séquençage ne seront donc entrepris que pour le clone 362P18.

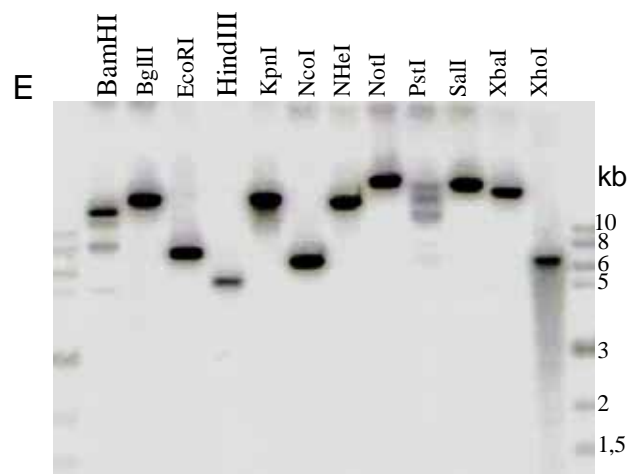
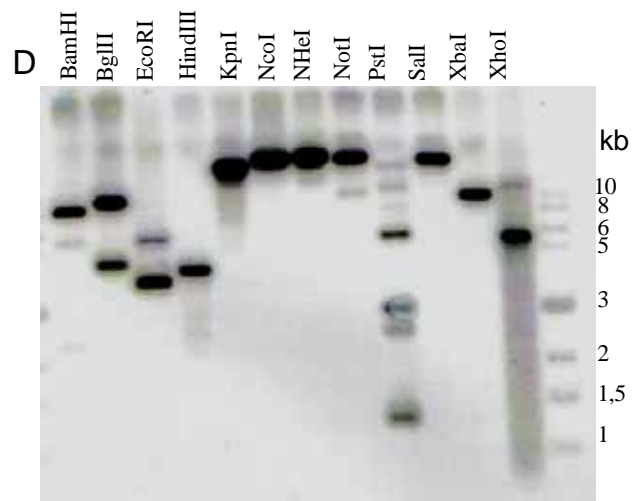
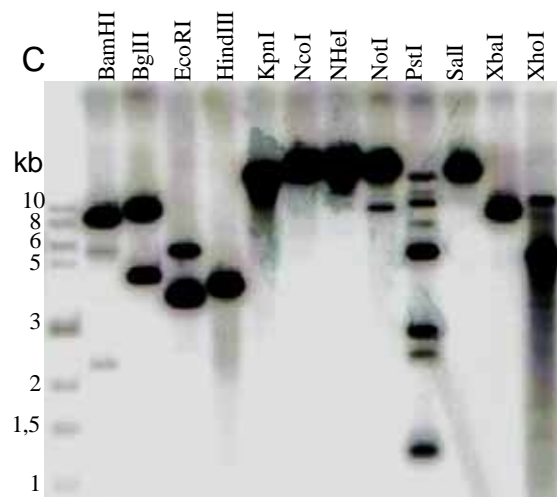
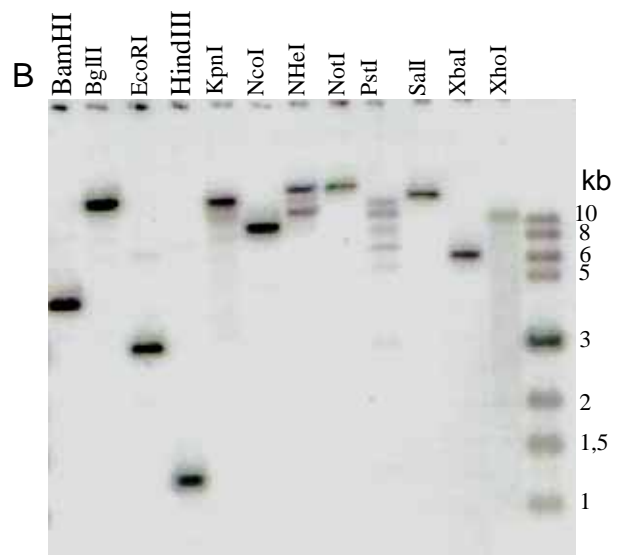
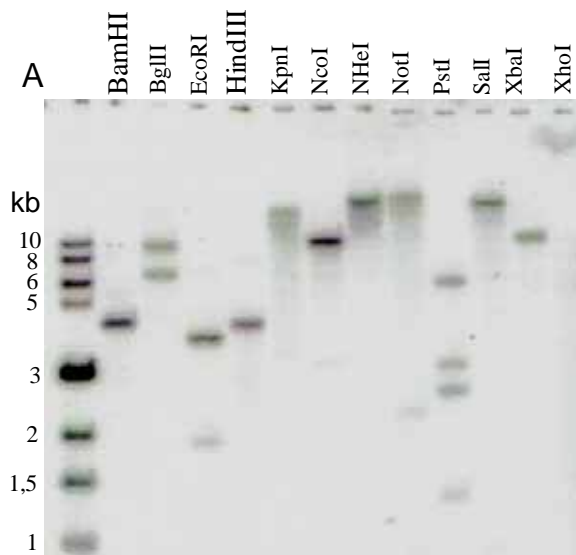


Figure 36: Southern Blots obtenus à partir de l'ADN extrait des clones BAC isolés après criblage des banques.

L'ADN a été digéré avec 12 enzymes de restriction différentes et hybridé avec une sonde marquée radioactivement préparée à partir de l'EST de tournesol HO001E17 homologue à MAGO NASHI.

- A: Clone BAC 34G03 (génotype Ha821)
- B: Clone BAC 225K21 (génotype Ha821)
- C: Clone BAC 362P18 (génotype PSC8)
- D: Clone BAC 88L03 (génotype PSC8)
- E: Clone BAC 13B07 (Génotype PSC8)

Les profils d'hybridation obtenus par Southern Blot des clones BAC sont très simples (présence d'1 ou 2 bandes quelle que soit l'enzyme de restriction). Ainsi, chaque clone BAC contiendrait soit une seule copie du gène MAGO NASHI soit plusieurs copies extrêmement proches les unes des autres.

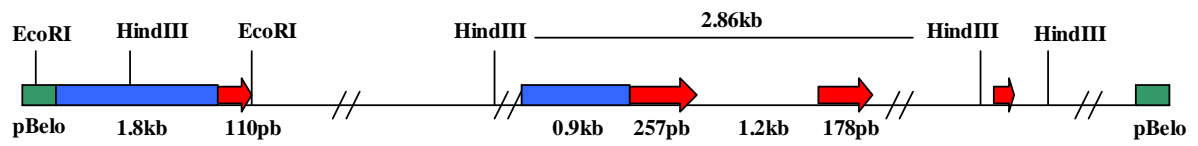
- **Sous-clonage et séquençage des clones BAC positifs.**

Afin de décrire l'organisation génomique des différentes copies de MAGO NASHI (nombre, taille et position des introns...) ainsi que leurs régions régulatrices (promoteurs et région 3' non traduite), le sous-clonage par digestion à l'aide des enzymes de restriction Hind III et EcoRI et le séquençage des clones BAC 34G03, 225K21, 13B07 et 362P18 a été entrepris. L'alignement des séquences partielles obtenues pour ces différents clones BAC est présenté en annexe 2. Les séquences générées à partir des clones BAC 34G03, 13B07 et 362P18 contiennent 2 introns ; concernant le clone 225K21, le sous-clonage ne nous a pour l'instant permis d'accéder qu'à une partie de la séquence du premier intron. Les deux introns identifiés dans la séquence des clones BAC sont situés à la même position que les deux introns des homologues de MAGO NASHI chez le riz (AN : OSM123) et *Arabidopsis thaliana* (AN : U89959), le premier intron étant localisé entre les deux premières régions très conservées de la séquence codante et le second intron au sein du domaine conservé carboxylique de la protéine. La taille des 3 exons identifiés est respectivement de 257 bp, 178 bp et 58 bp. Le premier intron qui a été complètement séquençé pour l'ensemble des clones BAC excepté pour le clone 225K21 a une taille de 1.3 kb pour le clone 362P18 et de 1.2 kb pour les clones 34G03 et 13B07. Concernant le second intron, seule la séquence complète du clone 362P18, représentant 0.8 kb, n'a pu être obtenue. La conservation de la position des introns au sein des homologues végétaux et la présence d'un seul intron au sein du gène MAGO NASHI de la drosophile indiquerait que les séquences introniques du gène se seraient établies indépendamment chez les plantes et chez les animaux.

Pour chacun des clones, une portion plus ou moins importante de la séquence 5' non traduite contenant le promoteur du gène a pu être obtenue allant de 2.7 kb pour le clone 13B07 à seulement 30 pb pour le clone 34G03 sous-cloné avec l'enzyme Hind III. Pour les clones 13B07 et 362P18, une partie de la séquence de la région 3' non traduite a également été générée représentant 180 bp pour 13B07 et environ 1kb pour 362P18.

Les résultats du sous-clonage du BAC 34G03 (génotype Ha821) par Hind III et EcoRI ont révélé deux séquences 5' non codantes très différentes, laissant supposer que ce clone BAC contiendrait deux copies du gène. De plus, selon l'alignement des séquences présenté en annexe 2, la séquence obtenue grâce au sous-clonage de ce clone BAC avec l'enzyme Hind III semble être identique à la séquence du clone BAC 13B07 (génotype PSC8). Les séquences des domaines 5' non traduits du clone BAC 362P18 (génotype PSC8) et du clone BAC 34G03 sous-cloné avec EcoRI sont identiques sur une région de 520 bp située directement en aval de la partie codante du gène puis divergent. La parfaite homologie d'une partie des séquences promotrices entre ces deux séquences suggérerait que ces deux

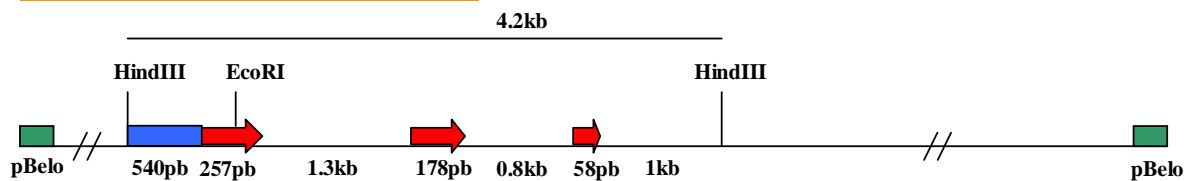
Clone BAC 34G03 (Ha821)



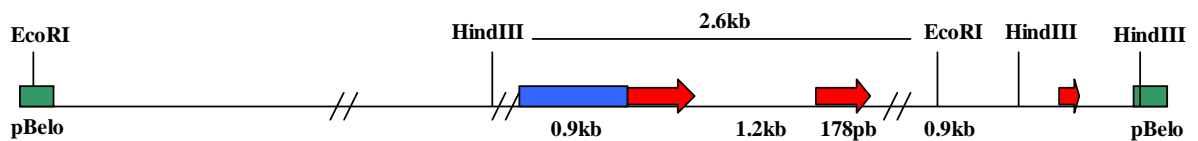
Clone BAC 225K21 (Ha821)



Clone BAC 362P18 (PSC8)



Clone BAC 13B07 (PSC8)



- : vecteur de clonage
- : région 5' non traduite
- : exons
- // : région non séquencée

Figure 37: Représentation schématique des données du sous-clonage des clones BAC

clones BAC pourraient représenter des copies identiques du gène MAGO NASHI. Toutefois, le sous-clonage du clone BAC 34G03 avec EcoRI n'ayant pas permis d'obtenir la séquence des introns, nous ne pouvons pas conclure quant à l'équivalence des copies portées par ces deux clones BAC.

II.5.1.4. Modèle putatif de l'organisation génomique de MAGO NASHI chez le tournesol.

Selon les données obtenues grâce aux démarches entreprises dans le but de décrire l'organisation génomique du gène MAGO NASHI chez le tournesol, nous pouvons supposer qu'il existe au moins 3 copies du gène de MAGO NASHI. Le criblage des banques BAC disponibles au laboratoire et le sous-clonage et le séquençage des clones BAC apparus positifs ont permis d'identifier 3 copies du gène pour le génotype Ha821, 1 copie étant située sur le clone 225K21 et 2 autres copies étant localisées sur le clone 34G03 (Fig. 37). Toutefois, nous ne disposons de la séquence génomique complète pour aucune de ces copies.

Concernant le génotype PSC8, uniquement 2 copies différentes, situées sur les clones 362P18 et 13B07, ont pu être mises en évidence, celles-ci correspondant potentiellement aux deux copies du gène situées sur le clone BAC 34G03. Concernant le clone BAC 362P18, le sous-clonage nous a permis d'accéder à la séquence complète du gène ainsi qu'à environ 1kb de chaque région 5' et 3' non codantes.

La comparaison des EST obtenues à partir de banques d'ADNc construites à partir du génotype PSC8 avec la séquence des clones BAC préparée à partir du même génotype indique que toutes les EST correspondraient à la copie de MAGO NASHI portée par le clone BAC 362P18 : l'EST DH0ALL28ZB02ZZM1 montre une homologie parfaite avec la région codante de la séquence de ce clone BAC et les rares différences entre les EST DH0AMM5ZG07ZZM1 et DH0AMM10ZH06ZZM1 et la séquence du clone BAC sembleraient dues à des erreurs de séquençage. Ainsi, il se pourrait que seule cette copie du gène soit exprimée au cours de l'embryogenèse.

II.5.2. Vers la compréhension de la régulation transcriptionnelle de MAGO NASHI chez le tournesol.

II.5.2.1. Obtention de profils d'expression spécifiques pour chacune des copies du gène.

L'une des premières démarches à entreprendre pour étudier la régulation transcriptionnelle de MAGO NASHI chez le tournesol est de caractériser le profil d'expression de chacune des copies identifiées. Une quantification du niveau d'expression au sein de différents tissus et à divers stades de développement peut, par exemple, être réalisée par RT-PCR quantitative qui constitue une méthode rapide et très sensible. Ces expériences, réalisées à partir d'amorces spécifiques de chacune des copies

du gène, permettraient de vérifier que seule la copie portée par le clone BAC 362P18 est transcrite dans les tissus embryonnaires et de déterminer les tissus au sein desquels les autres copies sont exprimés. De même, l'expression des différentes copies du gène pourrait être testée dans des organes issus de plantes cultivées dans des conditions particulières (stress salin, température, sécheresse...) afin de mettre en évidence une expression en réponse au stress.

Des expériences d'hybridation *in situ* pourraient également être réalisées afin d'observer le profil d'expression du gène au niveau tissulaire et cellulaire. La mise en évidence d'une ségrégation cellulaire particulière des ARNm de MAGO NASHI chez le tournesol suggérerait que le rôle de ce gène dans la localisation particulière d'ARNm spécifiques pourrait être conservé chez les végétaux. Afin d'étudier la répartition cellulaire des transcrits du gène MAGO NASHI, des expériences d'expression transitoire menées sur des protoplastes cultivés en conditions embryogènes ou non et transformés avec un plasmide contenant les régions codantes de MAGO NASHI et de la GFP (Green Fluorescent Protein) sous contrôle d'un promoteur constitutif tel que le promoteur 35S pourraient également être envisagées.

II.5.2.2. Caractérisation des éléments régulateurs des promoteurs des différentes copies du gène.

Lorsque les régions promotrices de l'ensemble des copies du gène MAGO NASHI du tournesol auront été séquencées, il serait intéressant de rechercher *in silico* la présence d'éléments régulateurs caractéristiques de chacune d'elles afin d'essayer d'identifier des structures déjà connues pour jouer un rôle au cours de l'embryogenèse.

II.5.3. Démarches entreprises pour tenter d'élucider la fonction et le mécanisme d'action de MAGO NASHI chez les végétaux.

II.5.3.1. Recherche de mutants végétaux du gène MAGO NASHI.

Une recherche de mutants du gène MAGO NASHI a été réalisée au sein des banques de mutants publiques d'*Arabidopsis thaliana* répertoriées sur le site internet TAIR (The *Arabidopsis* Information Resource, <http://www.arabidopsis.org/>). Cette recherche est restée infructueuse, aucun mutant altéré au niveau de la copie unique du gène MAGO NASHI d'*Arabidopsis thaliana* n'ayant pu être identifié. Chez la drosophile, la mutation du gène MAGO NASHI entraîne une stérilité de la descendance (Boswell et al., 1991). Si le même phénotype est observé après la mutation du gène MAGO NASHI d'*Arabidopsis thaliana*, il se peut que le mutant ait été éliminé au cours de la réalisation de la banque de mutants.

II.5.3.2. Recherche de l'homologue du tournesol de *TSUNAGI-Y14*, le principal partenaire de *MAGO NASHI*.

Chez les animaux, la protéine *MAGO NASHI* constitue l'un des éléments d'un complexe RNP au sein duquel elle interagit avec différents composants, *TSUNAGI/Y14* étant son principal partenaire. *TSUNAGI*, qui contient deux domaines RNP, est une protéine présentant de fortes similarités avec les protéines de liaison à l'ARN appartenant à la famille des ribonucléoprotéines (RNP). Comme *MAGO NASHI*, elle est caractérisée par un haut niveau de conservation parmi les différents homologues identifiés (60% d'identité entre l'homologue de la drosophile et de l'homme) et est exprimé à tous les stades de développement. Chez la Drosophile, il existe une seule copie du gène codant pour une protéine de 165 acides aminés ayant un poids moléculaire de 19 kDa (Mohr et al., 2001).

Une recherche des homologues végétaux putatifs du gène *TSUNAGI/Y14* a été menée par BLASTN au sein des EST végétales répertoriées dans les banques de séquences publiques. Aucune séquence EST présentant une homologie forte n'a pu être identifiée (e-value toujours inférieure à e-30). Cependant, dans le but de cribler les banques BAC et les banques d'ADNc disponibles au laboratoire à la recherche de l'homologue du tournesol de *TSUNAGI*, une sonde overgo a été désignée à partir d'une séquence EST de tournesol présentant des similarités avec *TSUNAGI* (AN : BU017638, e-value = 4e-29). Les amorces ont été désignées dans une région conservée entre cette séquence et tous les autres homologues de *TSUNAGI* déjà identifiés. L'une des difficultés rencontrées est que ces amorces présentent également une homologie inattendue avec le gène *Argonaute*. A la suite du criblage des banques d'ADNc, 3 clones positifs ont été révélés (HaCotR002E04, HaCotR017C06 et HaSemS4008H13) ; après séquençage, la caractérisation fonctionnelle de ces EST par recherche d'homologies dans les banques de données a montré que ces 3 clones présentaient des similarités avec la protéine *ARGONAUTE* et non avec *TSUNAGI*.

Le fait que l'on ne soit pas parvenu à identifier des clones ADNc correspondant au gène *TSUNAGI* parmi l'ensemble des EST végétales publiées ou par criblage des banques d'ADNc embryogènes produites au laboratoire laisse supposer que, chez les végétaux, ce gène pourrait subir une régulation transcriptionnelle extrêmement fine entraînant une expression très transitoire ou une activation dans des conditions physiologiques très particulières, rendant le clonage de ce gène difficile. Une autre hypothèse serait que les partenaires de *MAGO NASHI* sont différents au sein du règne animal et végétal. Au sein des collections d'EST générées au laboratoire à partir de tissus embryonnaires (embryons zygotiques à différents stades ou embryons somatiques), plusieurs protéines de type RNP, constituant des partenaires potentiels pour la protéine *MAGO NASHI*, ont pu être identifiées. Une étude plus poussée visant à évaluer leur capacité d'interaction avec *MAGO NASHI* serait donc intéressante. On peut également envisager de réaliser des expériences de double hybride pour tenter d'isoler l'ensemble des partenaires potentiels interagissant directement avec la protéine *MAGO NASHI*.

II.5.3.3. Perspectives pour tenter d'élucider le rôle de MAGO NASHI dans l'embryogenèse végétale.

- **Etude fonctionnelle par RNAi chez le tournesol.**

Le tournesol constitue une espèce végétale difficile à transformer génétiquement et le rendement de la transformation par *Agrobacterium tumefaciens* ne permet pas d'envisager une analyse systématique par *knockout* des gènes d'intérêt. Ainsi, chez cette plante, l'utilisation des techniques de génétique reverse par mutagenèse insertionnelle permettant une étude fonctionnelle plus poussée des gènes candidats est limitée. Récemment, une méthode permettant le silencing transitoire des gènes par RNAi a été mise au point au sein de notre laboratoire (Hewezi et al., 2004). Ainsi, l'analyse d'embryons issus de capitules infiltrés avec des short RNA spécifiques des différentes copies du gène de MAGO NASHI pourrait servir à évaluer le rôle respectif de chacune de ces copies au cours de l'embryogenèse précoce du tournesol.

- **Expériences de sur et sous-expression chez *Arabidopsis thaliana*.**

Arabidopsis thaliana, qui constitue l'une des plantes modèles pour l'étude des dicotylédones, dispose d'une gamme d'outils permettant d'effectuer des analyses fonctionnelles beaucoup plus large que le tournesol. Du fait de la facilité de transformation génétique de cette espèce végétale, on peut envisager de réaliser des expériences de sur ou sous-expression de MAGO NASHI afin de tenter d'élucider le rôle putatif de ce gène au cours de l'embryogenèse végétale. Cependant, il est à noter que, contrairement au tournesol qui contient au moins 3 copies du gène MAGO NASHI, *Arabidopsis thaliana* contient une seule copie et que le modèle fonctionnel de ce gène n'est peut être pas identique chez le tournesol et chez *Arabidopsis*.

- **Etude fonctionnelle par complémentation chez la Drosophile.**

Des études visant à déterminer si les différentes copies du gène MAGO NASHI du tournesol identifiées peuvent compléter fonctionnellement des mutants de la Drosophile dont le gène MAGO NASHI est altéré pourraient être menées en introduisant, par transgénèse, dans le génome des mutants un vecteur d'expression contenant l'ADNc de tournesol sous contrôle du promoteur du gène MAGO NASHI de la Drosophile comme cela a été décrit pour la caractérisation de l'homologue de MAGO NASHI de *C. Elegans* (Newmark et al., 1997).

Toute la démarche expérimentale entreprise pour l'étude détaillée du gène MAGO NASHI peut être généralisée à d'autres gènes potentiellement intéressants, tout en restant dans les limites du

raisonnable. Le TILLING (Targeting Induced Local Lesions IN Genomes) qui constitue une approche de mutagenèse classique suivie d'un criblage pour la détection de changements d'une base dans le gène d'intérêt (McCallum et al., 2000) pourrait être employé pour faciliter l'étude approfondie des gènes candidats. Mise au point chez *Arabidopsis thaliana* (Till et al., 2003), cette technique a déjà été utilisée avec succès pour mener des approches de génomique fonctionnelle chez des plantes cultivées difficiles à transformer génétiquement (Slade et Knauf, 2005).

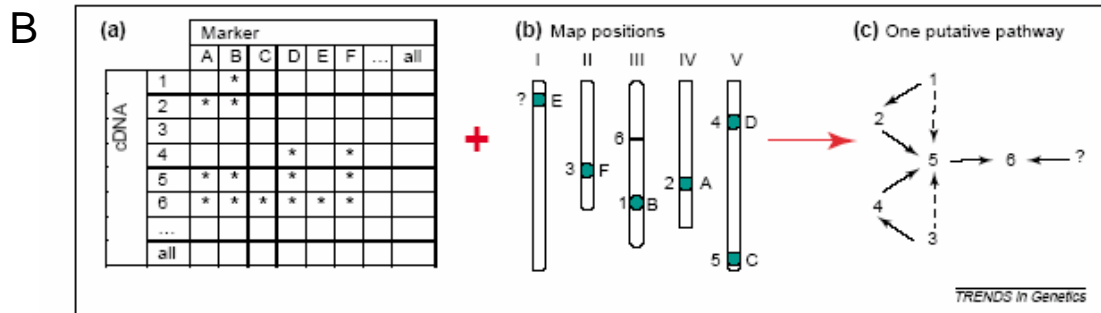
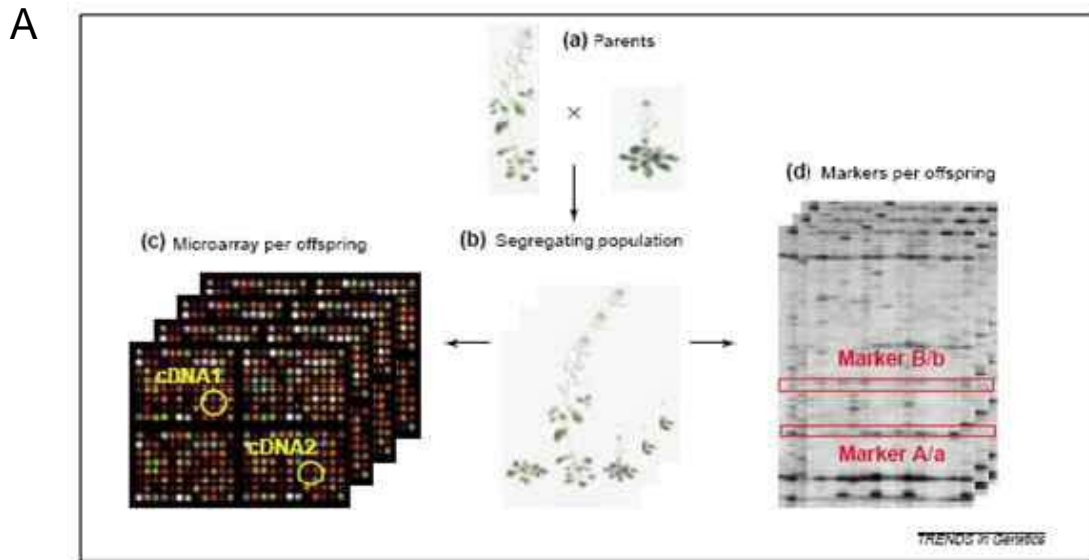


Figure 38: Principe de la génomique génétique. (d'après Jansen and Nap, 2001)

A. Combinaison d'études d'expression et d'analyses par marqueur moléculaire de populations de ségrégation.
 (a,b) Génération d'une population de ségrégation.
 (c,d) Pour analyser la population de ségrégation, chaque individu de la population est utilisé pour une analyse d'expression par microarray (c) et pour une analyse par marqueurs moléculaires (d).

B. Reconstitution de la voie de régulation.
 Les analyses QTL couplées aux données d'expression permettent de déduire les relations entre les ADNc testés.
 L'analyse QTL associe chaque ADNc du microarray avec les marqueurs d'une carte génétique.
 Ces données peuvent ensuite être utilisées pour étudier quels gènes influencent l'expression d'autres gènes et pour déduire dans quel ordre cette influence s'applique.

* = Différence significative entre les 2 allèles de chaque marqueur.

Chapitre III - Extension du modèle.

I. Mise en relation des données d'expression avec les connaissances génétiques.

Afin de compléter les connaissances apportées par nos travaux concernant le transcriptome induit lors de l'embryogenèse précoce chez le tournesol, des études de génomique génétique (genetical genomics) pourraient être menées. Ce type d'analyses, dont le principe est présenté dans la figure 38, associe les techniques d'expression génique et de cartographie génétique et consiste à détecter le loci génomique, appelé eQTL (expression quantitative trait loci), contrôlant les différences d'expression génique et susceptible de réguler la variation d'un caractère fonctionnel (Jansen and Nap, 2001). Bien que de telles études n'aient été menées jusqu'à aujourd'hui que sur des espèces modèles comme la souris, le maïs, l'homme ou encore la levure (Brem et al., 2002 ; Schadt et al., 2003), les outils nécessaires à leur réalisation, tels que des cartes génétiques denses, l'existence de populations de ségrégation ou encore un nombre important de séquences EST, sont actuellement disponibles pour le tournesol. Au cours de ce projet de thèse, plusieurs séquences présentant un profil d'expression intéressant mais ne montrant aucune homologie avec des gènes de fonction connue ont été identifiées. Les expériences de génomique génétique pourraient constituer une approche possible pour dégager des pistes concernant le rôle putatif de ces EST en les cartographiant et en recherchant un lien avec des QTL décrits précédemment comme influant sur les propriétés de la graine (poids, précocité, teneur en huile) (Al Chaarani et al., 2004, 2005) ou comme intervenant dans les processus d'embryogenèse somatique et de morphogenèse (Flores Berrios et al., 2000). En contrepartie, la cartographie d'EST auxquelles une fonction putative a pu être attribuée couplée à la recherche d'un lien avec des QTL donnerait des indices sur la voie métabolique dans laquelle les loci du QTL pourraient être impliqués.

II. Confrontation 'Transcriptome-Protéome-Métabolome' pour mettre en évidence les mécanismes de régulation transcriptionnelle et post-transcriptionnelle.

Les résultats d'analyses du transcriptome, basées sur l'analyse de collections d'EST ou d'expériences microarrays, sont le reflet de la population de gènes exprimés dans un tissu particulier à un moment donné. Elles sont donc le témoin des mécanismes de régulation transcriptionnelle contrôlant le processus biologique d'intérêt. Cependant, de nombreux phénomènes sont également soumis à un autre type de régulation de l'activité génique ayant lieu au niveau post-transcriptionnel. Ainsi, un ARNm présent dans une cellule n'est pas systématiquement traduit en protéine et la protéine produite

n'est pas obligatoirement active. La présence en quantité plus ou moins importante d'un transcrit codant pour une protéine n'est en fait qu'un indicateur de l'activité biologique et de la quantité de cette protéine dans la cellule. L'observation, parmi les clones d'ADNc ou au sein des gènes déclarés différentiels suite aux analyses de microarrays, d'une proportion élevée de séquences codant pour des protéines jouant un rôle potentiel dans la dégradation des ARNm ou dans le catabolisme protéique tels que l'ubiquitine ou les protéasomes confirme l'implication probable de ce type de contrôle au cours du développement embryonnaire précoce. Il serait donc intéressant à plus long terme de mener, en complément de cette analyse du transcriptome, des études de protéomique et de métabolomique permettant de mettre en évidence les phénomènes de régulation post-transcriptionnelle afin de mieux cibler les gènes et voies métaboliques essentiels au bon déroulement de l'embryogenèse.

III. Transposition du modèle aux autres plantes dicotylédones.

La collection d'EST générée au cours de cette thèse à partir d'embryons zygotiques de tournesol à des stades précoces du développement précisément identifiés constitue une ressource unique tant pour l'étude de l'embryogenèse chez cette plante d'intérêt agronomique majeur que pour l'ensemble des autres espèces dicotylédones exalbuminées.

Les gènes candidats et les mécanismes moléculaires potentiellement impliqués dans le contrôle de l'embryogenèse précoce du tournesol, identifiés suite à l'analyse des EST et des données de microarrays, pourraient servir de base pour l'étude de ce processus biologique chez d'autres modèles végétaux. Il serait également intéressant de comparer les différentes voies de régulation mises en œuvre chez différents types de plantes (domestiquées/non domestiquées, graines vertes/graines non vertes, exalbuminée/ albuminée, monocot/dicot) ou encore d'évaluer la similarité des événements moléculaires conduisant à l'établissement du patron morphologique avec d'autres dicotylédones dont le patron de développement est légèrement différent.

Dans l'éventualité d'une conservation de ces mécanismes parmi l'ensemble des plantes dicotylédones, le transfert de ces connaissances à des plantes modèles telle qu'*Arabidopsis thaliana* ou encore *Medicago truncatula* disposant d'un large arsenal d'outils génétiques et moléculaires permettrait de poursuivre cette étude plus aisément. Cette transposition permettrait d'appréhender ce processus dans sa globalité en effectuant, par exemple, des expériences de sur et sous expression (par mutagenèse, RNAi ou transformation génétique) de l'ensemble des gènes impliqués dans la voie de signalisation afin de caractériser la succession des événements menant au développement correct de l'embryon.

PARTIE IV
MATERIELS ET METHODES

Chapitre I – Matériel Biologique.

I. Matériel Végétal: *Helianthus annuus* L. Ha300b.

Pour réaliser ces travaux de thèse, le génotype Ha300b a été choisi pour son rendement élevé en graines. *Helianthus annuus* L. est cultivé en serre dans des conditions de culture standard (photopériode de 14h et température de 23°C à la lumière et 18°C à l'obscurité). Au moment de la floraison, les capitules sont autofécondés manuellement. Les embryons immatures sont récoltés par dissection des graines sous une loupe binoculaire puis triés selon leur stade de développement, en fonction de leur taille et de leur morphologie, et stockés dans du RNALater (Ambion, Austin, TX, USA) à -20°C.

II. Souches bactériennes.

Au cours de ces travaux, quatre souches bactériennes ont été utilisées :

- *E. Coli* DH5 α

Cette souche a été utilisée pour le clonage des ADNc de la banque SSH HaHeaS.

- *E. Coli* BM25-8 supE44, thi Δ (lac-proAB) [F' traD36, proAB+, lacIqZ Δ M15] λ imm434 (kanR)P1(camR)hsdR(rk12-mk12-)

Cette souche a été utilisée pour le clonage des ADNc des trois banques de référence HaG1bR, HaHeaR et HaCotR.

- *E. Coli* XL1-Blue endA1, gyrA96, hsdR17, lac-, recA1, relA1, supE44, thi-1 [F' lacIqZ Δ M51, proAB, Tn10]

Cette souche a été utilisée pour la construction des trois banques de référence.

- One Shot TOP10 Chemically Competent *E. Coli* F- mcrA Δ (mrr-hsdRMS-mcrBC) ϕ 80lacX74ercA1araD139 Δ (araleu) 7697 galU galK rpsL (Str^R) end A1 nupG

Cette souche a été utilisée lors du sous-clonage des clones BAC au cours de l'étude du gène MAGO NASHI.

Chapitre II – Méthodes.

I. Méthodes générales de biologie moléculaire.

I.1. Extraction d'ARN.

La méthode d'extraction d'ARN utilisée est adaptée du protocole décrit par Chomczynski et Sacchi (1987). Après élimination du RNA Later utilisé pour la conservation des échantillons végétaux, les embryons sont broyés puis incubés 10 minutes dans un bain-marie à 50°C dans 550µl de tampon d'extraction (500µl Tampon guanidium [59 mg guanidine thyocyanate, 734 mg de citrate de sodium, 74.4 mg EDTA, 500 mg sarcosyl dans 100 ml d'eau DEPC – pH ajusté à 7], 25µl β-mercaptoéthanol, 25µl SDS 10%) et 550 µl de phénol/chloroforme (1:1). Après une centrifugation de 10 minutes à 10000 rpm et à 4°C, la phase aqueuse est récupérée et subit deux lavages successifs au phénol/chloroforme et au chloroforme. Après une précipitation d'une nuit à -20°C dans de l'éthanol absolu et de l'acide acétique 1M, le culot est rincé à deux reprises dans 200µl de LiCl puis dans 200µl d'Acétate de Sodium 3M par centrifugations de 10 min. à 10 000 rpm à 4°C. A la suite d'une précipitation de 2 h. à -20°C dans de l'éthanol 70%, les ARN sont finalement repris dans 100µl d'eau DEPC.

Suit alors un traitement à la DNase réalisé à partir des 100µl d'ARN extraits avec 1µl de DNase RQ1 (Promega, Charbonnières, France) et 11µl de 10X Reaction Buffer (Promega, Charbonnières, France). Le mélange est incubé pendant 30 min. à 37°C puis la réaction est stoppée par ajout de 10µl de Solution Stop (Promega, Charbonnières, France) et par chauffage à 65°C durant 10 min.. On effectue ensuite un lavage au phénol/chloroforme et au chloroforme puis les ARN sont précipités pendant 2 h. à -20°C dans de l'éthanol absolu et de l'Acétate de Sodium 3M et le culot est repris dans 3µl d'eau DEPC.

I.2. Préparation de bactéries compétentes *E. Coli DH5α* et électroporation.

La production des bactéries compétentes *E. Coli DH5α* est réalisée selon le protocole décrit par Jacobs et al. (1990). Après avoir effectué une préculture en inoculant 50ml de LB Broth avec 100µl de bactéries *E. Coli DH5α* et en incubant 1 nuit à 37°C sous agitation (200rpm), 500ml de milieu 2XYT (16g Bacto tryptone, 10g Bacto yeast extract, 5g de NaCl pour un vol. final de 1L –pH ajusté à 7 avec NaOH 5N) sont inoculés avec 10 ml de préculture et incubés 3h à 37°C à 200 rpm. La culture bactérienne est ensuite refroidie 10 min. dans la glace puis centrifugée 15 min. à 4000g à 4°C. Après avoir éliminé le surnageant, le culot est resuspendu dans 500 ml d'eau UHQ stérile à 4°C. Après avoir renouvelé l'étape précédente, le culot bactérien est resuspendu dans 10 ml de glycérol 10% à 4°C et

centrifugé 15 min. à 2800 rpm à 4°C. Après avoir éliminé le surnageant, les bactéries sont aliquotées par 40µl et stockées à -80°C.

Les ADNc clonés sont introduits dans les bactéries *E. Coli* DH5α compétentes par électroporation avec un Gene Pulser II (Bio-Rad, Hercules, CA, USA) selon les instructions du fabricant.

I.3. RT-PCR quantitative.

La RT-PCR quantitative a été utilisée pour valider les résultats d'expression génique différentielle révélés par analyse des banques d'ADNc *in silico* et par les expériences de microarrays. Des amorces spécifiques à chaque EST ont été désignées à l'aide du logiciel Primer Express version 2.0 (Applied Biosystems, Forster City, CA, USA). Les réactions sont réalisées à l'aide d'un ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Forster City, CA, USA) à partir de 10µl de 2X Sybr Green Mastermix (Applied Biosystems, Forster City, CA, USA), 300 nM de chaque amorce et 10 ng d'ADNc selon le programme suivant : 50°C pendant 2 min., 95°C pendant 10 min., suivis de 40 cycles de 15 sec. à 95°C et 1 min. à 60°C. La réaction est ensuite soumise à un programme de dissociation de 15 sec. à 95°C, 15 sec. à 60°C puis une montée lente de 60°C à 95°C durant 20 min.. Pour chaque réaction, trois répétitions ont été réalisées. La quantification des changements du niveau d'expression des gènes est effectuée avec la méthode des $2^{-\Delta\Delta CT}$ décrite par Livak (Livak and Schmittgen, 2001), l'actine étant utilisée comme témoin constitutif pour la normalisation.

II. Construction et analyses des banques d'ADNc.

II.1. Construction des banques d'ADNc de référence: HaGlbR, HaHeaR, HaCotR.

Les trois banques de référence HaGlbR, HaHeaR et HaCotR, produites respectivement à partir d'embryons de tournesol immatures aux stades globulaire, cœur et cotylédonaire, ont été construites grâce au kit 'SMART™ cDNA library construction' (Clontech, Palo Alto, CA, USA) selon les instructions du fabricant.

II.1.1. Obtention des ADNc à cloner.

Pour chacun des échantillons, le premier brin d'ADNc est synthétisé en mélangeant 3µl d'ARN total à 1µl de SMART III Oligonucleotide (10 µM) et 1µl de CDS III/3' PCR Primer (10 µM). Le mélange est incubé pendant 2 min. à 72°C puis refroidi dans la glace pendant 2 min.. Après ajout de 2µl de 5X

First-Strand Buffer, 1µl de DTT (20mM), 1µl de dNTP (10mM) et 200 unités de MMLV PowerScript™ Reverse Transcriptase, le mélange est incubé 1h. à 42°C.

L'amplification des ADNc s'effectue simultanément à la synthèse du second brin : 2µl du premier brin d'ADNc sont mélangés à 2µl de 50X dNTP (10mM), 2µl de 5' PCR Primer (10µM), 2µl de CDS III/3' PCR Primer (10µM), 10µl de 10X Advantage 2 PCR Buffer et 2µl de 50X Advantage 2 Polymerase Mix dans un volume final de 100µl. Le programme PCR réalisé est le suivant : n cycles de 5 sec. à 95°C et 6 min. à 68°C. Afin que la proportion initiale de chaque transcrit présent dans l'échantillon soit conservée, la réaction d'amplification est arrêtée en phase exponentielle.

50µl d'ADNc sont ensuite digérés avec 2µl de protéinase K (20µg/µl) pendant 20 min. à 45°C. Après une phase de nettoyage au phénol/chloroforme/isoamyl alcool et de précipitation à l'éthanol 95% avec de l'acétate de sodium 3M, les ADNc sont resuspendus dans 80µl d'eau. Suit alors une étape de restriction par l'enzyme Sfi I (200 unités) pendant 2 h. à 50°C. Enfin, les ADNc sont fractionnés en fonction de leur taille à l'aide de la colonne CHROMASPIN-400. Les ADNc sélectionnés sont ensuite précipités à l'éthanol 95% avec de l'acétate de sodium 3M et resuspendus dans 7µl d'eau.

II.1.2. Clonage.

Le clonage s'effectue dans le phage λTriplex2 en incubant, durant une nuit à 16°C, 1 µl d'ADNc avec 1µl de vecteur (500 ng/µl), 0.5µl de 10X Ligation Buffer, 0.5µl d'ATP (10mM) et 200 unités de T4 DNA Ligase dans un volume final de 5µl. Suit alors l'étape d'encapsidation réalisée grâce au kit 'Gigapack III Plus Packaging Extract' (Stratagene, La Jolla, CA, USA) à partir de 4µl de ligation selon les recommandations du fabricant. Après titrage et amplification de la banque suite à la transformation de bactéries compétentes *E. coli* XL1-Blue, le phage λTriplex2 est converti en plasmide pTriplex2 grâce au système Cre-Lox lors d'une excision *in vivo* en masse dans les bactéries *E. coli* BM25-8.

II.2. Construction de la banque SSH: HaHeaS.

La banque HaHeaS est une banque soustraite créée à l'aide du kit 'PCR-Select™ cDNA Substraction'(Clontech, Palo Alto, CA, USA) à partir d'embryons aux stade cœur et cotylédonaire et enrichie en transcrits préférentiellement exprimés dans les embryons au stade cœur.

II.2.1. Obtention des ADNc à cloner.

Après l'extraction d'ARN, l'ADNc est synthétisé à l'aide du kit 'SMART™ PCR cDNA Synthesis' (Clontech, Palo Alto, CA, USA) selon le instructions du fabricant. Pour chacun des échantillons, le premier brin d'ADNc est synthétisé en mélangeant les 3µl d'ARN total à 1µl de SMART II A

Oligonucleotide (10 μ M) et 1 μ l de 3' SMART CDS Primer II A (10 μ M). Le mélange est incubé pendant 2 min. à 70°C puis 2 μ l de 5X First-Strand Buffer, 1 μ l de DTT (20mM), 1 μ l de 50X dNTP (10mM) et 200 unités de MMLV PowerScript™ Reverse Transcriptase sont ajoutés et le mélange est incubé 1h. à 42°C. Le premier brin d'ADNc est dilué d'un facteur 5 dans du TE et chauffé à 72°C pendant 7 min..

L'amplification des ADNc s'effectue simultanément à la synthèse du second brin : 10 μ l du premier brin d'ADNc sont mélangés à 2 μ l de 50X dNTP (10mM), 2 μ l de 5' PCR Primer II A (10 μ M), 10 μ l de 10X Advantage 2 PCR Buffer et 2 μ l de 50X Advantage 2 Polymerase Mix dans un volume final de 100 μ l. Le programme PCR réalisé est le suivant : 5 sec. à 95°C suivies de n cycles de 5 sec. à 95°C, 5 sec. à 65°C et 6 min. à 68°C. Afin que la proportion initiale de chaque transcrit présent dans l'échantillon soit conservée, la réaction d'amplification est arrêtée en phase exponentielle.

Les ADNc tester (cœur) et driver (cotylédonaire) sont digérés avec 15 unités d'enzyme de restriction RsaI pendant 1h.30min. à 37°C. Après l'arrêt de la digestion par ajout de 2.5 μ l de 20X EDTA/glycogène, les ADNc sont nettoyés au phénol/chloroforme/isoamyl alcool, précipités à l'éthanol 95% avec de l'acétate de sodium 3M et resuspendus dans 5.5 μ l d'eau.

L'ADNc tester est ensuite séparé en deux fractions identiques (de 2 μ l chacune) qui sont chacune liguée avec un adaptateur spécifique, 1 ou 2R, en incubant une nuit à 16°C chaque fraction d'ADNc avec 2 μ l de l'adaptateur spécifique (10 μ M chacun), 2 μ l de 5X Ligation Buffer et 400 unités de T4 DNA Ligase dans un volume final de 10 μ l. Ces adaptateurs serviront ultérieurement pour effectuer une amplification PCR des ADNc d'intérêt. La réaction est arrêtée en chauffant 5 min. à 72°C après ajout de 1 μ l de 20X EDTA/glycogène.

II.2.2. Hybridations entre ADNc driver et tester .

Deux hybridations successives des ADNc driver et tester visant à normaliser et à isoler les transcrits spécifiques de la population d'ADNc tester d'intérêt ont lieu. Au cours de la première hybridation, les 1.5 μ l d'ADNc driver sont ajoutés à 1.5 μ l de chaque fraction d'ADNc tester avec 1 μ l de 4X Hybridization Buffer. Les ADNc tester ayant été dilués par l'étape de ligation d'adaptateurs, les ADNc se retrouvent en excès dans le mélange réactionnel. Les échantillons sont alors chauffés pendant 1min.30sec. à 98°C afin de les dénaturer puis incubés 8h. à 68°C. A la suite de cette première hybridation, la population d'ADNc simple brin restante est normalisée et très enrichie en séquences différenciellement exprimées pour chacune des conditions tester et driver, les ADNc présents à la fois dans la population tester et driver formant des hybrides. Lors de la seconde hybridation, les deux échantillons d'ADNc issus de la première hybridation sont mélangés et 1 μ l d'ADNc driver fraîchement dénaturé est ajouté pour favoriser l'enrichissement en séquences différenciellement exprimées. Le mélange est alors incubé une nuit à 68°C sans phase de dénaturation préalable. Le

mélange, qui est dilué dans 200µl de Dilution Buffer et chauffé 7 min. supplémentaires à 68°C, contient alors de nouvelles molécules hybrides consistant en ADNc préférentiellement exprimés dans la condition tester d'intérêt contenant des adaptateurs différents à chaque extrémité.

II.2.3. Amplification des ADNc préférentiellement exprimés dans la condition d'intérêt.

Après avoir complété les brins manquants des adaptateurs par une brève incubation à 75°C afin de créer le site d'appariement pour le PCR Primer I, les ADNc différenciellement exprimés sont sélectivement amplifiés au cours de deux réactions PCR successives.

Au cours de la première amplification, 1µl d'ADNc sont mélangés à 0.5µl de 50X Advantage cDNA Polymerase mix, 1µl de PCR Primer I (10µM), 0.5µl de dNTP mix (10mM) et 2.5µl de 10X Advantage2 PCR Buffer dans un volume final de 25µl. Ce mélange réactionnel est soumis au programme de PCR suivant : 5 min. à 75°C et 25 sec. à 94°C suivi de 27 cycles de 10 sec. à 94°C, 30 sec. à 66°C et 1 min. 30 sec. à 72°C. Au cours de cette première amplification, seuls les ADNc double brin possédant 2 adaptateurs différents à chaque bout sont amplifiés de façon exponentielle.

Durant la seconde phase d'amplification, 1µl de produits PCR issus de la première réaction et dilués d'un facteur 10 est mélangé à 0.5µl de 50X Advantage cDNA Polymerase mix, 1µl de chaque Nested PCR Primer 1 et 2R (10µM chacun), 0.5µl de dNTP mix (10mM) et 2.5µl de 10X Advantage2 PCR Buffer dans un volume final de 25µl. Le programme de PCR effectué est le suivant : 12 cycles de 10 sec. à 94°C, 30 sec. à 68°C et 1 min. 30 sec. à 72°C. Cette phase d'amplification permet une réduction du bruit de fond et un enrichissement en séquences différenciellement exprimées.

II.2.4. Clonage.

Le clonage des ADNc s'effectue dans le plasmide PCRII à l'aide du kit 'TA Cloning kit Dual Promoter' (Invitrogen, Carlsbad, CA, USA) selon les conseils du fabricant. Ainsi, 3µl d'ADNc amplifiés sont mélangés à 2µl de PCRII vector, 1µl de 10X Ligation Buffer et 200 unités de T4 DNA Ligase dans un volume final de 10µl et incubés une nuit à 14°C.

Les plasmides sont ensuite introduits dans des bactéries compétentes *E. coli* DH5α par électroporation avec un Gene Pulser II (Bio-Rad, Hercules, CA, USA). Après une nuit de culture sur milieu sélectif (LB Agar contenant de l'ampicilline (100mg/L) et 40µl de X-Gal (40mg/ml), 6144 clones positifs ont été ordonnés dans des plaques de culture de 384 puits contenant 80µl de LB Freezing (6.270g K₂HPO₄, 1.796g KH₂PO₄, 0.499g citrate de sodium, 0.099g MgSO₄, 0.898g (NH₄)₂SO₄, 44 ml glycérol dans 1L de LB Broth) avec ampicilline (100mg/L) et incubés une nuit à 37°C sous agitation (200 rpm).

II.3. Normalisation des banques redondantes.

A la suite d'un séquençage préliminaire de 96 clones de chacune des banques visant à évaluer la redondance et la qualité de la banque (taille des inserts, contamination par de l'ADN d'origine bactérienne, mitochondriale ou chloroplastique), les banques HaGlbR, HaHeaR et HaHeaS ont présenté un taux de redondance relativement élevé pour quelques séquences correspondant à une séquence présentant des homologies avec une protéine *ycf3* impliquée dans l'assemblage du photosystème I pour la banque HaHeaS et deux séquences montrant des similarités avec une protéine prédite de 14.3 kDa et un transporteur membranaire pour les deux banques de référence HaGlbR et HaHeaR. Afin d'optimiser la découverte de gènes différents lors du séquençage de ces banques, un 'nettoyage' de ces banques pour ces séquences redondantes a été réalisé par criblage des clones bactériens déposés sur filtre nylon avec des sondes overgos spécifiques de chacune de ces trois EST. La technique de criblage par sonde overgos a été choisie pour sa spécificité très élevée, la très haute qualité des résultats obtenus ainsi que sa facilité de mise en œuvre.

Le même protocole a été suivi pour cribler les banques EST et BAC à la recherche de clones MAGO NASHI et TSUNAGI.

II.3.1. Préparation des macroarrays à partir des banques d'ADNc.

Chacune des banques d'ADNc a été spotted à l'aide du *BioGrid* (Biorobotics LTD, Cambridge, U.K.) en duplicat et selon un modèle de dépôt 4X4 sur des membranes nylon N+ Hybond (Amersham, Orsay, France) découpées au format 8X12cm et étalées sur du LB Agar. Après une nuit de culture à 37°C pour la banque HaHeaS clonée dans *E. coli* DH5 α ou 31°C pour les banques HaGlbR et HaHeaR clonées dans la souche bactérienne *E. coli* BM25-8, les membranes sont traitées afin de lyser les bactéries et de fixer les clones bactériens au support. Pour cela, les filtres sont déposés pendant 7 min. sur du papier Wathman imbibé d'une solution de (NaOH 0.5N, NaCl 1.5M) puis 7 min. sur du papier Wathman imbibé d'une solution de (Tris HCl 0.5N pH8, NaCl 1.5M). Après avoir laissé sécher pendant 1h minimum sur du papier peu absorbant, les membranes sont alors placées pendant 20 min. sur du papier Wathman imbibé d'une solution de (NaOH 0.4N) puis 7 min. sur du papier Wathman imbibé de solution de 5X SSPE (NaCl 0.9M, Sodium Phosphate 50mM, EDTA 5mM pH 7.7) avant de les laisser sécher pendant une nuit.

II.3.2. Hybridation des macroarrays avec des amorces overgos spécifiques des séquences redondantes.

Les amorces overgos des 3 séquences montrant un taux de redondance élevé ont été désignées sur le site web (<http://www.genome.wustl.edu/tools/>). L'hybridation des macroarrays avec les sondes overgos est réalisée selon le protocole décrit par Cai et al. (1998).

II.3.2.1. Marquage radioactif des sondes overgos.

Les sondes overgos sont tout d'abord marquées radioactivement avec du $\alpha^{32}\text{P}$ dATP et du $\alpha^{32}\text{P}$ dCTP : le mélange des deux oligos (10pmol/ μl chacun) correspondant à chaque sonde overgo est chauffé pendant 5 min. à 80°C puis 10 min. à 37 °C et enfin stocké dans la glace. Après ajout de 0.5 μl de BSA (2mg/ml), 2 μl d'OLB (-A, -C, -N₆) (OLB (-A, -C, -N₆) correspond au mélange de trois solutions A, B et C selon les proportions (1 :2.5 :1.5) où [solution O : 1.25M Tris-HCl pH8, 125 mM MgCl₂], [solution A : 1ml solution O, 18 μl 2-mercaptoéthanol, 5 μl 0.1MdTTP, 5 μl 0.1M dGTP], [solution B : 2M Hepes-NaOH pH6.6] et [solution C : 3mM Tris-HCl pH7.4, 0.2 mM EDTA]), 5 μCi de chaque élément radioactif et 2 unités d'enzyme de Klenow dans un volume final de 10 μl , le mélange est incubé pendant 1h à température ambiante. Après avoir éliminé les nucléotides non incorporés grâce à un passage sur colonne Sephadex G50 (Amersham, Orsay, France), les sondes sont dénaturées 10 min. à 95°C puis placées dans la glace.

II.3.2.2. Hybridation des macroarrays.

Les membranes sont préhybridées pendant 4 h. à 58°C dans 50 ml de solution d'hybridation préchauffée (BSA à 1%, EDTA 1mM, SDS à 7%, Sodium Phosphate à 0.5M) puis, après ajout de 10 μl de sonde overgos marquée radioactivement et dénaturée, les membranes sont hybridées pendant une nuit à 58°C.

II.3.2.3. Lavages des membranes.

Afin d'éliminer la sonde en excès, les membranes subissent deux rinçages de 30 min. à 58°C, le premier dans du SSC 2X (SSC 20X : 3M NaCl, 300mM Na citrate) et du SDS 0.1% et le second dans du SSC 1.5X et du SDS 0.1%.

II.3.3. Détection des clones redondants.

Les résultats d'hybridation sont révélés par scanning au PhosphoImager (FujiFilm) après une nuit d'exposition. Les clones apparaissant positifs à la suite de l'hybridation et correspondant donc aux séquences redondantes ne sont pas séquencés.

II.4. Séquençage des clones.

II.4.1. Extraction de l'ADN plasmidique.

L'ADN plasmidique contenant les clones à séquencer est extrait à l'aide du kit 'Montage Plasmid Miniprep96' (Millipore, Bedford, MA, USA) selon le protocole *Partial Lysate* proposé par le fabricant et reposant sur la lyse des clones bactériens et la purification de l'ADN plasmidique par filtration sous vide. Une pré-culture des clones bactériens dans des plaques de culture bactérienne de 96 puits contenant 200µl de LB2X (20g Tryptone, 10g Yeast Extract, 10g NaCl dans 1L d'eau – pH ajusté à 7 avec NaOH 10N) avec de l'ampicilline (100mg/L) pendant une nuit sous agitation (200 rpm) et à 31°C ou 37°C selon la souche bactérienne utilisée pour le clonage de la banque est tout d'abord effectuée. Puis, 10µl de la suspension bactérienne sont utilisés pour inoculer 1ml de LB2X avec ampicilline (100mg/L) contenu dans des plaques Deep Well de 96 puits mises à incuber une nuit à la température adéquate et à 320 rpm. Après une centrifugation de 7 min. à 1500g, le surnageant est éliminé et les culots bactériens sont resuspendus dans 150µl de la solution 1 de resuspension. A la suite de l'ajout de 150µl de la solution 2 de lyse, la plaque est agitée vigoureusement pendant 1 min. puis incubée à température ambiante pendant 2 min. supplémentaires avant l'ajout de 150µl de la solution 3 de neutralisation et une agitation vigoureuse de 2 min.. 200µl du lysat bactérien de chaque puit sont transférés dans la plaque *Clearing* placée sur un vacuum au dessus de la plaque *Plasmid*. Un vide de 8 inches de Hg est appliqué jusqu'à ce que la totalité du lysat se soit transférée dans la plaque *Plasmid* qui est alors soumise à un vide de 24 Inches de Hg jusqu'à ce que les puits soient vides. Après ajout de 200µl de solution 4 de rinçage dans chaque puit, l'étape de filtration à 24 Inches de Hg est renouvelée. Les plasmides sont ensuite resuspendus dans différents volumes de solution 5 d'injection selon la souche bactérienne utilisée pour le clonage des banques qui sont respectivement 25µl pour des plasmides introduits dans *E. coli* DH5α et 50µl pour BM 25-8.

II.4.2. Réactions de séquences.

Les réactions de séquences sont réalisées à partir de 5µl d'ADN plasmidique mélangé à 1µl de BigDye Terminator v3.2 Cycle Sequencing Kit (Applied Biosystems, Forster City, CA, USA), 3µl de Buffer 2.5X et 1µl d'amorce (50µM). Les amorces utilisées diffèrent en fonction du plasmide employé pour

cloner la banque : il s'agit de Triplex5 (5' TCCGAGATCTGCACGAGC 3') pour le plasmide pTriplex2 et M13R (5' TGAGCGGATAACAATTTACACAG 3') pour PCR II. Les températures d'appariement caractéristiques de chaque amorce étant différentes (45°C pour Triplex5 et 42°C pour M13R), un programme PCR spécifique est appliqué pour chaque type de plasmide: 5 min à 94°C suivies de 40 cycles de 30 sec. à 94°C, 15 sec. à 45°C ou 42°C et 4 min. à 60°C.

II.4.3. Purification des réactions de séquences.

Les produits de réactions de séquences sont purifiés à l'aide du kit 'Montage Seq96 Sequencing Reaction Clean Up' (Millipore, Bedford, MA, USA). Après avoir ajouté 15µl de solution d'injection, les produits de réactions de séquences sont transférés dans la plaque Seq96 et soumis à un vide d'environ 25 Inches de Hg jusqu'à ce que les puits de la plaque Seq96 soient vides. Cette opération est renouvelée après ajout de 25µl de solution d'injection dans chacun des puits de la plaque. Les produits de réactions de séquences sont alors resuspendus dans 25µl de solution d'injection puis transférés dans une plaque d'injection adéquate pour le séquenceur.

II.4.4. Séquençage.

Le séquençage est réalisé à l'aide d'un séquenceur ABI Prism 3700 (Applied Biosystems, Forster City, CA, USA) selon les paramètres d'injection 2kV pendant 15 sec..

II.5. Analyses bioinformatiques.

Les séquences de mauvaise qualité sont filtrées avec le programme PHRED puis le vecteur est éliminé en utilisant le programme CROSS_MATCH. Les séquences ayant une taille inférieure à 100 paires de bases sans le polyA sont exclues pour la suite des analyses. Le contigage est réalisé à l'aide du programme PHRAP et les recherches d'homologies dans les bases de données sont effectuées avec le programme BLASTX sur la banque nr.

III. Construction et analyse des microarrays.

III.1. Construction des microarrays.

III.1.1. Sélection et réarrangement des clones bactériens.

Des EST provenant de différentes banques d'ADNc de tournesol ont été assemblées selon le programme Phrap pour identifier les séquences potentiellement différentes. 9 404 clones représentatifs de 8185 unigènes ont été réarrangés dans des plaques de 96 puits contenant 200µl de milieu de culture (LB Freezing) avec 100 mg.l⁻¹ d'ampicilline et incubés à 37°C ou 31°C (selon la souche bactérienne utilisée pour le clonage) dans un agitateur (200rpm) pendant 24 heures.

III.1.2. Obtention des produits PCR et vérification.

III.1.2.1. Conditions PCR.

5µl de suspension bactérienne sont transférés dans une plaque PCR 96 puits contenant 95µl d'eau UHQ. Les plaques sont ensuite chauffées pendant 10 min. à 95°C puis centrifugées à 1200 g pendant 3 min afin de déposer les débris cellulaires au fond des puits. 8µl de surnageant du lysat bactérien sont alors prélevés afin d'amplifier l'ADNc cloné en utilisant les amorces spécifiques des séquences flanquant le site de clonage du plasmide qui sont les primers M13 universels (M13 Forward 5' GTTTTCCAGTCACGACGTTG 3' and M13 Reverse 5' TGAGCGGATAACAATTTTCACACAG 3') pour les cDNA clonés dans les vecteurs PCRII, pSport1 and pGMT et les primers Triplex (Triplex5 5' TCCGAGATCTGCACGAGC 3' et Triplex3 5' TAATACGACTCACTATAGGG 3') pour les inserts clonés dans le plasmide pTriplex2. La réaction PCR est effectuée dans 100µl de volume final avec 10µl de 10X PCR buffer, 2.5 mM de MgCl₂, 0.2 µM de chaque primer, 200 µM de chaque dNTP et 3 unités de Taq DNA Polymerase (Quiagène, Courtaboeuf, France). Les échantillons d'ADN sont dénaturés pendant 2 min. à 95°C suivies de 40 cycles d'amplification de 30 sec. à 95°C, 30 sec. à 52°C et 2 min. à 72°C suivis d'une phase d'extension finale de 7 min. à 72°C.

III.1.2.2. Vérification des produits PCR.

Afin de s'assurer de la qualité des produits d'amplification obtenus (taille adéquate, quantité suffisante, absence de contaminations), les fragments sont visualisés sur un gel d'agarose à 1%. Les produits PCR présentant des bandes d'amplification multiples ainsi que les clones non amplifiés sont éliminés et remplacés par une autre EST appartenant au même contig.

III.1.2.3. Evapoconcentration des produits PCR et réarrangements.

Les produits PCR sont évaporés sous une hotte puis repris dans 30µl d'eau UHQ afin d'obtenir une concentration de 300 à 500 ng/µl. Les échantillons stockés initialement dans des plaques PCR de 96 puits sont transférés dans des plaques de 384 puits afin de réduire le temps de spotting des microarrays.

III.1.3. Spotting des microarrays.

Les produits PCR sont déposés en double dépôt sur une membrane de nylon Hybond N+ (Amersham, Orsay, France) à l'aide du *MicroGrid II* (Biorobotics LTD, Cambridge, U.K.) en utilisant un outil comptant 64 aiguilles selon un patron de dépôt 13x13. Au cours de la procédure de dépôt, une humidité relative de 70 % et une température de 21°C ont été maintenues.

III.1.4. Traitement de fixation des produits PCR sur le microarrays.

Suite au spotting des microarrays, les microarrays subissent des traitements permettant de fixer les produits PCR à la membrane de nylon. Les membranes de nylon sont tout d'abord placées pendant 20 min. sur du papier Whatman imbibé de solution de dénaturation (1.5 M NaCl et 0.5 M NaOH) puis 20 min. supplémentaires sur du papier Whatman imbibé de solution de neutralisation (1.5 M NaCl et 1M Tris-HCl, pH 7.4). Les filtres traités sont ensuite chauffés pendant 2h à 80°C puis crosslinker sous UV pendant 2 min avec un UV Stratalinker 1800.

III.2. Hybridation des microarrays.

III.2.1. Préparation des sondes complexes d'ADNc marquées.

A la suite de l'extraction d'ARN, les ADNc sont préparés à l'aide du kit 'SMART™ PCR cDNA Synthesis' (Clontech, Palo Alto, CA, USA) selon le même protocole que celui décrit pour la banque d'ADNc SSH. Ensuite, pour chacun des échantillons, 100 ng d'ADNc sont marqués à l'aide du kit 'Megaprime DNA Labeling System' (Amersham, Orsay, France). Pour cela, 100ng de sonde sont incubés pendant 5 min. à 100°C avec 5 µl de primers et de l'eau pour que le volume final de réaction à l'issue du marquage soit égal à 50µl. Puis, 10 µl de buffer 5X, 2 µl d'enzyme et 5 µl de α -³³P dCTP (50µCi) sont ajoutés et le mélange est incubé à 37°C pendant 30 min. Les sondes marquées sont purifiées par passage sur colonne Probe Quant G-50 Micro Columns (Amersham, Orsay, France) et dénaturées à 100°C pendant 10 min. puis refroidies dans la glace avant utilisation.

III.2.2. Préparation des sondes oligonucléotides.

500 ng de chaque amorce reverse des séquences flanquant le site de clonage des différents vecteurs utilisés pour les différentes banques (amorce M13R 5' TGAGCGGATAACAATTTACACAG 3' et Triplex3 5' TAATACGACTCACTATAGGG 3') sont marqués avec 3µl de [γ -33 P] dATP (30µCi) avec 200 unités de polynucléotide kinase en incubant le mélange 1h. à 37°C. Après arrêt de la réaction par un chauffage à 95°C pendant 10 min., les différents oligos marqués sont mélangés et purifiés par passage sur colonne Probe Quant G-50 Micro Columns (Amersham, Orsay, France).

III.2.3. Conditions d'hybridation des microarrays.

III.2.3.1. Description du dispositif expérimental.

Afin de faciliter l'analyse statistique de nos expériences de microarrays, un plan en blocs incomplets équilibrés de type carré latin incomplet a été élaboré. Cinq microarrays sont hybridés avec quatre des cinq tissus testés et 4 répétitions techniques sont réalisées pour chaque traitement. Outre l'effet du traitement, ce dispositif expérimental permet de contrôler deux autres sources de variabilité : la variation due à la membrane et la variation due à la répétition technique. Afin de contrôler la variation biologique existant entre différents individus, pour chaque traitement, les ARN sont extraits à partir d'embryons provenant de différents capitules.

III.2.3.2. Conditions d'hybridation des microarrays.

- **Hybridation avec les sondes complexes.**

Les membranes sont pré-hybridées pendant 24 h. à 65°C dans un tampon composé de SDS 0.5%, SSC 5X, 0.1% de Ficoll 400, 0.1% Polyvinylpyrrolidone, 0.1% d'Albumine Bovine Fraction et 100 µg/ml d'ADN de sperme de saumon (Sigma, Saint Quentin Fallavier, France) préalablement dénaturé pendant 10 min. à 95°C. Les microarrays sont ensuite hybridés dans le même tampon contenant la sonde marquée radioactivement pendant 24 h à 65°C puis rincés deux fois dans du SDS 0.1%, SSC 0.1X à 65°C pendant 15 min.

- **Hybridation avec la sonde oligo-vecteur.**

L'hybridation avec la sonde oligo-vecteur servant à vérifier la qualité des microarrays produits est réalisée avec le même tampon d'hybridation que pour les sondes complexes. Après une pré-hybridation des microarrays pendant 12h à 42°C, la sonde oligos est ajoutée et les membranes sont alors incubées pendant 24h à 42°C puis lavées dans du SSC 2X, SDS 0.1% pendant 10 min. à

température ambiante puis 5 min à 42°C. Les membranes sont exposées durant une nuit au contact d'écrans Fujifilm sensibles aux émissions radioactives puis révélées grâce au scanner Bio-Imaging Analyser BAS-5000 (Fujifilm).

III.2.3.3. Déshybridation des microarrays.

Entre chaque cycle d'hybridation, les microarrays sont déshybridés par deux lavages successifs de 15 min. dans du SDS 0.5%, SSC 0.1X porté à ébullition. L'absence de tout signal radioactif résiduel sur les membranes est vérifiée à l'aide du scanner Bio-Imaging Analyser BAS-5000 (Fujifilm) après une nuit d'exposition.

III.2.4. Analyse des données d'hybridation des microarrays.

III.2.4.1. Obtention et traitement des images d'hybridation.

A la suite de l'hybridation, les membranes sont exposées de 3 à 6 nuits au contact d'écrans Fujifilm capables de capter l'énergie radioactive. Après avoir scanné les écrans à l'aide du Bio-Imaging Analyser BAS-5000 (Fujifilm) selon une densité de 16 bit/Pixel et une résolution de 25 µm/Pixel, les images obtenues sont traitées grâce au logiciel ArrayGauge V.1.3 (Fujifilm) qui permet le repérage des différents spots et la quantification de leur intensité. Le bruit de fond global est ensuite soustrait à l'intensité de chaque spot.

III.2.4.2. Analyse statistique de l'expression des gènes.

Les analyses statistiques sont réalisées selon la méthode décrite par Wolfinger (Wolfinger et al., 2001) avec le logiciel R 2.0.0. Deux ANOVA interconnectées sont effectuées. La première ANOVA effectuée sur les log10 des valeurs permet d'éliminer la variabilité due à l'expérimentation. Les résidus de cette ANOVA constituent les valeurs normalisées utilisées pour réaliser la seconde ANOVA effectuée indépendamment pour chaque gène et permettant de déterminer l'effet de chaque traitement testé. Les gènes présentant au moins un spot dont le signal est supérieur au maximum de l'intensité des témoins négatifs sont considérés comme exprimés. Pour chaque gène, l'homogénéité de la variance des valeurs d'expression normalisées entre les différents stades de développement est vérifiée grâce au test de Levène et les effets des différents traitements sont mesurés. Pour les gènes présentant une variance homogène et un effet 'stade' significatif, la somme des carrés 'Gene-by-Stages of Development' est divisée en 4 contrastes linéaires utilisés comme test *post hoc* pour identifier l'origine de la variabilité. Les 4 contrastes définis comparent respectivement 'ovules non fécondés' *versus* 'embryons (tous stades confondus)', 'feuille' *versus* 'tissus reproductifs (incluant les

embryons et les ovules non fécondés), ‘embryons globulaires’ *versus* ‘embryons cœur’ et ‘embryons précoces’ (globulaires et cœur) *versus* ‘embryons cotylédonaire’. La correction de Bonferroni est utilisée pour réduire l’erreur introduite par les tests multiples.

Les gènes présentant des variations significatives de leur niveau d’expression entre les différents stades de développement étudiés sont classés selon leur profil d’expression grâce à une méthode de classification hiérarchique classique utilisant la technique du lien moyen et le coefficient de corrélation de Pearson pour le calcul de la distance entre les gènes et les conditions.

IV. Etude de l’organisation génomique du gène MAGO NASHI.

IV.1. Southern Blots.

IV.1.1. Extraction d’ADN génomique.

L’extraction d’ADN génomique est réalisée à partir de 10g de matériel végétal broyé dans l’azote liquide auxquels sont ajoutés 15 ml de tampon d’extraction (Tris HCl pH8 0.1M, EDTA 0.05M, NaCl 0.5M) et 2 ml de SDS 10%. Après une incubation de 10 min. à 65°C et ajout de 5 ml de KoAc 5M/3M, on effectue une centrifugation de 30 min. à 10 000g et à 4°C. Après récupération du surnageant et ajout de 0.5 vol. d’isopropanol, le mélange est incubé 20 min. dans la glace puis centrifugé pendant 20 min. à 4°C à 9 500 rpm. Le culot est repris dans 500µl de TE puis stocké à -20°C pendant une nuit. Après un lavage dans du phénol/chloroforme/alcool isoamylique (25 :24 :1), la phase aqueuse est récupérée et précipitée dans 2 vol. d’éthanol 100% et 0.1 vol. d’acétate de sodium 3M par centrifugation 15 min. à 12 000g et à 4°C. Le culot est alors repris dans 500µl de TE.

Les échantillons subissent ensuite un traitement à la RNase A : après avoir ajouté 5µl de RNase A à 10mg/ml, le mélange est incubé 20 min. à température ambiante puis lavé avec du phénol/chloroforme/alcool isoamylique (25 :24 :1) et précipité dans 2 vol. d’éthanol 100% et 0.1 vol. d’acétate de sodium 3M. Après un rinçage à l’éthanol 70%, le culot est repris dans 30µl de TE.

IV.1.2. Digestion de l’ADN par des enzymes de restriction.

Au cours des expériences de Southern Blot visant à caractériser l’organisation génomique du gène MAGO NASHI, l’ADN génomique a été digéré par différentes enzymes de restriction: EcoRI, BamHI, HindIII, BglII, KpnI, NcoI, NheI, NotI, PstI, SacI, XbaI et XhoI. 20µg d’ADNg sont incubés pendant une nuit à 37°C avec 200 unités d’enzymes, 10µl du tampon 10X adéquat, 1 µl de BSA 100X, dans un volume final de 100µl.

IV.1.3. Transfert de l'ADN génomique digéré sur membrane de nylon.

Le transfert de l'ADN sur membrane de nylon Hybond-N+ (Amersham, Orsay, France) est réalisé selon le protocole de transfert neutre. Après migration de l'ADN digéré par électrophorèse sur gel d'agarose 0.8%, le gel est placé dans une solution d'HCl 0.125M pendant 10 min. sous agitation (étape de dépurination) puis, après un bref rinçage à l'eau, le gel est incubé pendant 30 min. sous agitation dans la solution de dénaturation (87.66g NaCl et 20g NaOH dans 1L). Le gel est ensuite incubé pendant 30 min. dans la solution de neutralisation (87.66g NaCl, 60.5g Tris pH 7.5 dans un vol. de 1L) après avoir été rapidement rincé dans l'eau. Le transfert s'effectue par capillarité durant une nuit dans du 20X SSC. L'ADN est fixé à la membrane par UV crosslinking puis la membrane est rincée dans du 2X SSC avant d'être stockée à 4°C dans du saran.

IV.1.4. Hybridation des membranes.

IV.1.4.1. Obtention et marquage radioactif de la sonde.

La sonde spécifique de MAGO NASHI est obtenue par restriction du clone ADNc HO001E17 et découpage de la bande correspondant à l'insert après migration du plasmide digéré sur gel d'agarose. Après élution de l'ADN dans 300µl d'eau par passage sur filtre 0.2µm, l'ADN est précipité durant une nuit à -20°C dans 2.5 vol. d'éthanol absolu et 0.1 vol. d'acétate de sodium 3M. Après une centrifugation de 30 min. à 15 000g à 4°C et élimination du surnageant, le culot est rincé à l'éthanol 70% et repris dans 15 µl de TE.

100ng de sonde sont ensuite marqués radioactivement par incorporation de ³²P dCTP à l'aide du kit 'MegaPrime DNA labelling System' (Amersham, Orsay, France) selon les instructions du fabricant. Préalablement à l'hybridation, la sonde est purifiée par passage sur une colonne de type ProbeQuant G50 microcolumn (Amersham, Orsay, France) et dénaturée par chauffage 10 min. à 100°C.

IV.1.4.2. Conditions d'hybridation.

Les membranes sont pré-hybridées pendant 4h à 65°C dans un tampon contenant du SDS 7%, 0.5M phosphate B7.2 et 10 mM EDTA) puis hybridées dans le même tampon après ajout de la sonde radioactive pendant une nuit à 65°C.

IV.1.4.3. Lavage des membranes.

Les membranes subissent 3 rinçages pendant 30 min à 65°C, le premier dans une solution de 2X SSC, 0.1% SDS, le second dans une solution de 1X SSC, 0.1% SDS et enfin le dernier dans une solution de 0.1X SSC, 0.1% SDS. Les membranes sont ensuite mises au contact d'un écran sensible aux émissions

radioactives. Les résultats d'hybridation sont révélés par scanning au PhosphoImager (FujiFilm) après plusieurs nuits d'exposition.

IV.2. Sous-clonage et séquençage de clones BAC.

Pour réaliser le sous-clonage des clones BAC 34G03, 225K21, 362P18 et 13B07, identifiés par criblage des banques BAC avec des sondes overgos spécifiques du gène MAGO NASHI, l'ADN des BAC a été digéré avec les enzymes EcoRI et HindIII. Les produits de restriction de l'ADN des BAC sont ligués dans du pBlue Script, qui a également été digéré préalablement par les enzymes EcoRI et HindIII et déphosphorylé. Le plasmide est ensuite introduit dans des bactéries compétentes par transformation chimique grâce au kit 'One Shot TOP10 Competent cells' (Invitrogen, Carlsbad, CA, USA) selon les instructions du fabricant. 5µl de produit de ligation sont ajoutés à un échantillon de 50µl de bactéries One Shot et le mélange est mis dans la glace durant 30 min. puis incubé 30 sec. à 42°C et remis dans la glace avant d'ajouter en conditions stériles 250µl de milieu SOC pré-chauffé à 37°C. Les échantillons sont ensuite incubés pendant 1h. à 37°C à 225 rpm puis 20 à 200µl de chaque transformation sont étalés sur du LB Agar en conditions sélectives (100mg/L Ampicilline, 40µl de XGal (40mg/ml)) et incubés une nuit à 37°C. La totalité des colonies blanches susceptibles d'avoir intégré le plasmide est repiquée dans des plaques de culture bactérienne de 96 puits contenant 200µl de LB Freezing avec 100mg/L d'ampicilline et incubée une nuit à 37°C à 200 rpm. La présence d'un insert est vérifiée par PCR en utilisant les primers universels M13 reverse et forward selon le même protocole que celui utilisé pour construire le microarray. Les clones positifs sont séquencés en utilisant la même méthode que pour le séquençage des banques d'ADNc.

REFERENCES BIBLIOGRAPHIQUES

A

- Abe M, Katsumata H, Komeda Y, Takahashi T.** (2003). Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in *Arabidopsis*. *Development*. 130(4):635-43.
- Aharoni A, Vorst O.** (2001). DNA microarrays for functional plant genomics. *Plant Mol Biol*. 48:99-118.
- Aida M, Vernoux T, Furutani M, Traas J, Tasaka M.** (2002) Roles of PIN-FORMED1 and MONOPTEROS in pattern formation of the apical region of the *Arabidopsis* embryo. *Development*. 129(17):3965-74.
- Al-Chaarani GR, Gentzbittel L, Huang XQ, Sarrafi A.** (2004). Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (*Helianthus annuus* L.). *Theor Appl Genet*. 109(7):1353-60.
- Al-Chaarani GR, Gentzbittel L, Wedzony M, Sarrafi A.** (2005). Identification of QTLs for germination and seedling development in sunflower (*Helianthus annuus* L.). *Plant Science*. 169:221-27.
- Alibert G, Aslane-Chanabe C, Burrus M.** (1994). Sunflower tissue and cell cultures and their use in biotechnology. *Plant Physiol. Biochem*. 32, 31-44.
- Alkharouf NW, Matthews BF.** (2004). SGMD: the Soybean Genomics and Microarray Database. *Nucleic Acids Res*. 32(Database issue):D398-400.
- Allen RD, Cohen EA, Vonder Haar RA, Adams CA, Ma DP, Nessler CL, Thomas TL.** (1987). Sequence and expression of a gene encoding an albumin storage protein in sunflower. *Mol Gen Genet*. 210(2):211-8.
- Altmann T, Felix G, Jessop A, Kauschmann A, Uwer U, Pena-Cortes H, Willmitzer L.** (1995). Ac/Ds transposon mutagenesis in *Arabidopsis thaliana*: mutant spectrum and frequency of Ds insertion mutants. *Mol Gen Genet*. 247(5):646-52.
- Anil VS, Harmon AC, Rao KS.** (2003). Temporal association of Ca(2+)-dependent protein kinase with oil bodies during seed development in *Santalum album* L.: its biochemical characterization and significance. *Plant Cell Physiol*. 44(4):367-76.
- Apuya NR, Yadegari R, Fischer RL, Harada JJ, Goldberg RB.** (2002). RASPBERRY3 gene encodes a novel protein important for embryo development. *Plant Physiol*. 129(2):691-705.
- Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main D, Wood T, Leslie A, Wing RA, Wilkins TA.** (2004). Functional genomics of cell elongation in developing cotton fibers. *Plant Mol Biol*. 54(6):911-29.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G.** (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 25(1):25-9.
- Assaad FF, Huet Y, Mayer U, Jurgens G.** (2001). The cytokinesis gene KEULE encodes a Sec1 protein that binds the syntaxin KNOLLE. *J Cell Biol*. 152(3):531-43.

B

- Bachem CW, van der Hoeven RS, de Bruijn SM, Vreugdenhil D, Zabeau M, Visser RG.** (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(5):745-53.

- Baima S, Possenti M, Matteucci A, Wisman E, Altamura MM, Ruberti I, Morelli G.** (2001). The *arabidopsis* ATHB-8 HD-zip protein acts as a differentiation-promoting transcription factor of the vascular meristems. *Plant Physiol.* 126(2):643-55.
- Barakat A, Szick-Miranda K, Chang IF, Guyot R, Blanc G, Cooke R, Delseny M, Bailey-Serres J.** (2001). The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiol.* 127(2):398-415.
- Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R.** (2005). NCBI GEO: mining millions of expression profiles--database and tools. *Nucleic Acids Res.* 33(Database issue):D562-6
- Belanger K, Quatrano R.** (2000). Polarity: the role of localized secretion. *Curr Opin Plant Biol* 3: 67-72.
- Belyavsky A, Vinogradova T, Rajewsky K.** (1989) PCR-based cDNA library construction: general cDNA libraries at the level of a few cells. *Nucleic Acids Res.* 17(8):2919-32.
- Ben C, Hewezi T, Jardinaud MF, Bena F, Ladouce N, Moretti S, Tamborindeguy C, Liboz T, Petitprez M, Gentzbittel L.** (2005) Comparative analysis of early embryonic sunflower cDNA libraries. *Plant Mol Biol.* 57(2):255-70.
- Benkova E, Michniewicz M, Sauer M, Teichmann T, Seifertova D, Jurgens G, Friml J.** (2003). Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell.* 115(5):591-602.
- Bert PF, Jouan I, Tourvieille de Labrouhe D, Serre F, Philippon J, Nicolas P, Vear F.** (2003). Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.). 2. Characterisation of QTL involved in developmental and agronomic traits. *Theor Appl Genet.* 107(1):181-9.
- Bertucci F, Bernard K, Loriod B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K, Jordan BR.** (1999). Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum Mol Genet.* 8(9):1715-22.
- Blilou I, Frugier F, Folmer S, Serralbo O, Willemsen V, Wolkenfelt H, Eloy NB, Ferreira PC, Weisbeek P, Scheres B.** (2002). The *Arabidopsis* HOBBIT gene encodes a CDC27 homolog that links the plant cell cycle to progression of cell differentiation. *Genes Dev.* 16(19):2566-75.
- Bloch KE.** (1983).Sterol structure and membrane function. *CRC Crit Rev Biochem.* 14(1):47-92.
- Bogue MA, Vonder Haar RA, Nuccio ML, Griffing LR, Thomas TL.** (1990). Developmentally regulated expression of a sunflower 11S seed protein gene in transgenic tobacco. *Mol Gen Genet.* 222(1):49-57.
- Boswell RE, Prout ME, Steichen JC.** (1991). Mutations in a newly identified *Drosophila melanogaster* gene, mago nashi, disrupt germ cell formation and result in the formation of mirror-image symmetrical double abdomen embryos. *Development.* 113(1):373-84.
- Bozhkov PV, Filonova LH, Suarez MF.** (2005). Programmed cell death in plant embryogenesis. *Curr Top Dev Biol.* 67:135-79.
- Brem RB, Yvert G, Clinton R, Kruglyak L.** (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science.* 296(5568):752-5.
- Breton C, Chaboud A, Matthys-Rochon E, Bates EE, Cock JM, Fromm H, Dumas C.** (1995). PCR-generated cDNA library of transition-stage maize embryos: cloning and expression of calmodulin genes during early embryogenesis. *Plant Mol Biol.* 27(1):105-13.
- Broun P, Gettner S, Somerville C.** (1999). Genetic engineering of plant lipids. *Annu Rev Nutr.* 19:197-216.

Butte A. (2002). The use and analysis of microarray data. *Nat Rev Drug Discov.* 1(12):951-60.

Byrne ME, Barley R, Curtis M, Arroyo JM, Dunham M, Hudson A, Martienssen RA. (2000). Asymmetric leaves1 mediates leaf patterning and stem cell function in *Arabidopsis*. *Nature.* 408(6815):967-71.

C

Cahoon EB, Carlson TJ, Ripp KG, Schweiger BJ, Cook GA, Hall SE, Kinney AJ. (1999). Biosynthetic origin of conjugated double bonds: production of fatty acid components of high-value drying oils in transgenic soybean embryos. *Proc Natl Acad Sci U S A.* 96(22):12935-40.

Cai WW, Reneker J, Chow CW, Vaishnav M, Bradley A. (1998). An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. *Genomics* 54(3): 387-397.

Carles CC, Fletcher JC. (2003) Shoot apical meristem maintenance: the art of a dynamic balance. *Trends Plant Sci.* 8(8):394-401.

Casson S, Spencer M, Walker K, Lindsey K. (2005). Laser capture microdissection for the analysis of gene expression during embryogenesis of *Arabidopsis*. *Plant J.* 42(1):111-23.

Champagnat P, Ozenda P, Baillaud L. (1969). *Biologie végétale III. Croissance, Morphogénèse, Reproduction.* Ed Masson et Cie.

Chanabe C, Burrus M, Alibert G. (1989). Factors affecting the improvement of colony formation from sunflower protoplasts. *Plant Science* 64, 125-132.

Channeliere S, Riviere S, Scalliet G, Szecsi J, Jullien F, Dolle C, Vergne P, Dumas C, Bendahmane M, Hugueney P, Cock JM. (2002). Analysis of gene expression in rose petals using expressed sequence tags. *FEBS Lett.* 515(1-3):35-8.

Chaudhury AM, Letham S, Craig S, Dennis ES. (1993). amp1 – a mutant with high cytokinin levels and altered embryonic pattern, faster vegetative growth, constitutive photomorphogenesis and precocious flowering. *Plant J.* 4:907-916.

Chaudhury A, Koltunow A, Payne T, Luo M, Tucker M, Dennis E, Peacock W. (2001). Control of early seed development. *Annu Rev Cell Dev Biol* 17: 677-699.

Chomczynski P, Sacchi N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem.* 162(1):156-9.

Christensen SK, Dagenais N, Chory J, Weigel D. (2000). Regulation of auxin response by the protein kinase PINOID. *Cell.* 100(4):469-78.

Churchill GA. (2002). Fundamentals of experimental design for cDNA microarrays. *Nat Genet.* 32 Suppl:490-5.

Clark JK, Sheridan WF. (1991). Isolation and Characterization of 51 embryo-specific Mutations of Maize. *Plant Cell.* (9):935-951.

Clouse SD, Sasse JM. (1998). BRASSINOSTEROIDS: Essential Regulators of Plant Growth and Development. *Annu Rev Plant Physiol Plant Mol Biol.* 49:427-451.

Crawford KM, Zambryski PC. (2001). Non-targeted and targeted protein movement through plasmodesmata in leaves in different developmental and physiological states. *Plant Physiol.* 125(4):1802-12.

Crookshanks M, Emmersen J, Welinder KG, Nielsen KL. (2001). The potato tuber transcriptome: analysis of 6077 expressed sequence tags. *FEBS Lett.* 506(2):123-6.

D

da Silva PMFR, Eastmond PJ, Hil LM, Smith AM, Rawsthorne S. (1997). Starch metabolism in developing seeds of oilseed rape. *Planta.* 203:480-487.

De Folter S, Busscher J, Colombo L, Losa A, Angenent GC. (2004). Transcript profiling of transcription factor genes during silique development in *Arabidopsis*. *Plant Mol Biol.* 56(3):351-66.

De Jong AJ, Heidstra R, Spaik HP, Hartog MV, Meijer EA, Hendriks T, Schiavo FL, Terzi M, Bisseling T, Van Kammen A, De Vries SC. (1993). Rhizobium Lipooligosaccharides Rescue a Carrot Somatic Embryo Mutant. *Plant Cell.* 5(6):615-620.

De Leo F, Bonade-Bottino MA, Ceci LR, Gallerani R, Jouanin L. (1998). Opposite effects on *Spodoptera littoralis* larvae of high expression level of a trypsin proteinase inhibitor in transgenic plants *Plant Physiol.* 118(3):997-1004.

Despres B, Delseny M, Devic M. (2001). Partial complementation of embryo defective mutations: a general strategy to elucidate gene function. *Plant J.* 27(2):149-59.

Devic M, Albert S, Delseny M. (1996). Induction and expression of seed-specific promoters in *Arabidopsis* embryo-defective mutants. *Plant J.* 9(2):205-15.

Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN. (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell.* 86(3):423-33.

Dollar G, Struckhoff E, Michaud J, Cohen RS. (2002). Rab11 polarization of the *Drosophila* oocyte: a novel link between membrane trafficking, microtubule organization, and oskar mRNA localization and translation. *Development.* 129(2):517-26.

Dumas C, Mogensen HL. (1993). Gametes and Fertilization: Maize as a Model System for Experimental Embryogenesis in Flowering Plants. *Plant Cell.* 5(10):1337-1348.

E

Eastmond PJ, Dennis DT, Rawsthorne S. (1997). Evidence That a Malate/Inorganic Phosphate Exchange Translocator Imports Carbon across the Leucoplast Envelope for Fatty Acid Synthesis in Developing Castor Seed Endosperm. *Plant Physiol.* 114(3):851-856.

Edwards PA, Ericsson J. (1999). Sterols and isoprenoids: signaling molecules derived from the cholesterol biosynthetic pathway. *Annu Rev Biochem.* 68:157-85.

Errampalli D, Patton D, Castle L, Mickelson L, Hansen K, Schnall J, Feldmann K, Meinke D. (1991). Embryonic Lethals and T-DNA Insertional Mutagenesis in *Arabidopsis*. *Plant Cell.* 3(2):149-157.

Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S, Claverie JM. (1999). Large-scale statistical analyses of rice EST reveal correlated patterns of gene expression. *Genome Res.* 9(10): 950-959.

F

Fedorova M, Van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt JS, Vance CP. (2002). Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiol.* 130(2):519-37.

Fernandes J, Brendel V, Gai X, Lal S, Chandler VL, Elumalai RP, Galbraith DW, Pierson EA, Walbot V. (2002). Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. *Plant Physiol.* 128(3):896-910.

Ferraro M, Buglia GL, Romano F. (2001). Involvement of histone H4 acetylation in the epigenetic inheritance of different activity states of maternally and paternally derived genomes in the mealybug *Planococcus citri*. *Chromosoma.* 110(2):93-101.

Fiers M, Hause G, Boutilier K, Casamitjana-Martinez E, Weijers D, Offringa R, van der Geest L, van Lookeren Campagne M, Liu CM. (2004). Mis-expression of the CLV3/ESR-like gene CLE19 in *Arabidopsis* leads to a consumption of root meristem. *Gene.* 327(1):37-49.

Fletcher JC, Brand U, Running MP, Simon R, Meyerowitz EM. (1999). Signaling of cell fate decisions by CLAVATA3 in *Arabidopsis* shoot meristems. *Science.* 283(5409):1911-4.

Flores Berrios E, Gentzbittel L, Kayyal H, Alibert G, Sarrafi A. (2000). AFLP mapping of QTLs for in vitro organogenesis traits using recombinant inbred lines in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 101, 1299-1306.

Friml J, Benkova E, Blilou I, Wisniewska J, Hamann T, Ljung K, Woody S, Sandberg G, Scheres B, Jurgens G, Palme K. (2002). AtPIN4 mediates sink-driven auxin gradients and root patterning in *Arabidopsis*. *Cell.* 108(5):661-73.

Friml J, Vieten A, Sauer M, Weijers D, Schwarz H, Hamann T, Offringa R, Jurgens G. (2003). Efflux-dependent auxin gradients establish the apical-basal axis of *Arabidopsis*. *Nature.* 426(6963):147-53.

Fu H, Subramanian RR, Masters SC. (2000). 14-3-3 proteins: structure, function and regulation. *Annual Review of Pharmacology and Toxicology* 40, 617-647.

Fukuda H. (1996). Xylogenesis: Initiation, Progression, and cell death. *Annu Rev Plant Physiol Plant Mol Biol.* 47:299-325.

G

Galau G.A. et al. (1986). Abscisic acid induction of cloned cotton late embryogenesis abundant (LEA) messenger RNAs. *Plant Mol. Biol.* 7:155-170.

Galweiler L, Guan C, Muller A, Wisman E, Mendgen K, Yephremov A, Palme K. (1998). Regulation of polar auxin transport by AtPIN1 in *Arabidopsis* vascular tissue. *Science.* 282(5397):2226-30.

Garcia-Diaz MT, Martinez-Rivas JM, Mancha M. (2002). Temperature and oxygen regulation of oleate desaturation in developing sunflower (*Helianthus annuus*) seeds. *Physiol Plant.* 114(1):13-20.

Gentzbittel L, Abbott A, Galaud JP, Georgi L, Fabre F, Liboz T, Alibert G. (2002). A bacterial artificial chromosome (BAC) library for sunflower, and identification of clones containing genes for putative transmembrane receptors. *Mol Genet Genomics.* 266(6):979-87.

Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J. (2000). Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol.* 124(4):1570-81.

Gisel A, Barella S, Hempel FD, Zambryski PC. (1999). Temporal and spatial regulation of symplastic trafficking during development in *Arabidopsis thaliana* apices. *Development.* 126(9):1879-89.

Goldberg R, de Paiva G, Yadegari R. (1994). Plant embryogenesis: zygote to seed. *Science* 266: 605-614.

Grimanelli D, Perotti E, Ramirez J, Leblanc O. (2005). Timing of the Maternal-to-Zygotic Transition during Early Seed Development in Maize. *Plant Cell*. 17(4):1061-72.

Gyorgyey J, Vaubert D, Jimenez-Zurdo JI, Charon C, Troussard L, Kondorosi A, Kondorosi E. (2000). Analysis of *Medicago truncatula* nodule expressed sequence tags. *Mol Plant Microbe Interact*. 13(1):62-71.

H

Hachet O, Ephrussi A. (2001). *Drosophila* Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. *Curr Biol*. 11(21):1666-74.

Haecker A, Laux T. (2001). Cell-cell signaling in the shoot meristem. *Curr Opin Plant Biol*. 4(5):441-6.

Haecker A, Gross-Hardt R, Geiges B, Sarkar A, Breuninger H, Herrmann M, Laux T. (2004). Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. *Development*. 131(3):657-68.

Hamann T, Benkova E, Baurle I, Kientz M, Jurgens G. (2002). The *Arabidopsis* BODENLOS gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning. *Genes Dev*. 16(13):1610-5.

Han CU, Lee CH, Jang KS, Choi GJ, Lim HK, Kim JC, Ahn SN, Choi JE, Cha JS, Kim HT, Cho KY, Lee SW. (2004). Identification of rice genes induced in a rice blast-resistant mutant. *Mol Cells*. 17(3):462-8.

Han YZ, Huang BQ, Zee SY, Yuan M. (2000). Symplastic communication between the central cell and the egg apparatus cells in the embryo sac of *Torenia fournieri* Lind. before and during fertilization. *Planta*. 211(1):158-62.

Hardtke CS, Berleth T. (1998). The *Arabidopsis* gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J*. 17(5):1405-11.

Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA. (2000). Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science*. 290(5499):2110-3.

Heck GR, Perry SE, Nichols KW, Fernandez DE. (1995). AGL15, a MADS domain protein expressed in developing embryos. *Plant Cell*. 7(8):1271-82.

Heidstra R, Welch D, Scheres B. (2004). Mosaic analyses using marked activation and deletion clones dissect *Arabidopsis* SCARECROW action in asymmetric cell division. *Genes Dev*. 18(16):1964-9.

Helleboid S, Chapman A, Hendriks T, Inze D, Vasseur J, Hilbert JL. (2000). Cloning of beta-1,3-glucanases expressed during *Cichorium* somatic embryogenesis. *Plant Mol Biol*. 42(2):377-86.

Hellmann H, Hobbie L, Chapman A, Dharmasiri S, Dharmasiri N, del Pozo C, Reinhardt D, Estelle M. (2003). *Arabidopsis* AXR6 encodes CUL1 implicating SCF E3 ligases in auxin regulation of embryogenesis. *EMBO J*. 22(13):3314-25.

Hennig L, Gruissem W, Grossniklaus U, Kohler C. (2004). Transcriptional programs of early reproductive stages in *Arabidopsis*. *Plant Physiol*. 135(3):1765-75.

Hewezi T, Kallerhoff G, Alibert G. (2004). Local infiltration of high and low molecular weight RNA from silenced sunflower (*Helianthus annuus* L.) plants trigger post transcriptional gene silencing in non-silenced plants. *Plant biotechnology Journal* doi:10.1111/j.1467-7652.2004.00103.x.

Hinderhofer K, Zentgraf U. (2001). Identification of a transcription factor specifically expressed at the onset of leaf senescence. *Planta*. 213(3):469-73.

Hobbie L, McGovern M, Hurwitz LR, Pierro A, Liu NY, Bandyopadhyay A, Estelle M. (2000). The *axr6* mutants of *Arabidopsis thaliana* define a gene involved in auxin response and early development. *Development*. 127(1):23-32.

Hoecker U, Vasil IK, McCarty DR. (1995). Integrated control of seed maturation and germination programs by activator and repressor functions of *Viviparous-1* of maize. *Gene Dev*.9(20):2459-69.

Hofte H, Desprez T, Amselem J, Chiapello H, Rouze P, Caboche M, Moisan A, Jourjon MF, Charpentieu JL, Berthomieu P, et al. (1993). An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J*. 4(6):1051-61.

Hu W, Wang Y, Bowers C, Ma H. (2003). Isolation, sequence analysis, and expression studies of florally expressed cDNAs in *Arabidopsis*. *Plant Mol Biol*. 53(4):545-63.

Hwang I, Sheen J. (2001). Two-component circuitry in *Arabidopsis* cytokinin signal transduction. *Nature*. 413(6854):383-9.

Hwang YS, Bethke PC, Cheong YH, Chang HS, Zhu T, Jones RL. (2005). A gibberellin-regulated calcineurin B in rice localizes to the tonoplast and is implicated in vacuole function. *Plant Physiol*. 138(3):1347-58.

I

Ingram GC. (2004). Between the sheets: inter-cell-layer communication in plant development. *Philos Trans R Soc Lond B Biol Sci*. 359(1446):891-906.

Ingvarsdson C, Veierskov B. (2001). Ubiquitin- and proteasome-dependent proteolysis in plants. *Physiol Plant*. 112(4):451-459.

J

Jacobs M, Wnendt S, Stahl U. (1990). High-efficiency electro-transformation of *Escherichia coli* with DNA from ligation mixtures. *Nucleic Acids Res*. 18(6):1653.

Jang CS, Lee MS, Kim JY, Kim DS, Seo YW. (2003). Molecular characterization of a cDNA encoding putative calcium binding protein, *HvCaBP1*, induced during kernel development in barley (*Hordeum vulgare* L.). *Plant Cell Rep*. Aug;22(1):64-70.

Jang JC, Fujioka S, Tasaka M, Seto H, Takatsuto S, Ishii A, Aida M, Yoshida S, Sheen J. (2000). A critical role of sterols in embryonic patterning and meristem programming revealed by the *fackel* mutants of *Arabidopsis thaliana*. *Genes Dev*. 14(12):1485-97.

Jansen RC, Nap JP. (2001). Genetical genomics: the added value from segregation. *Trends Genet*. 17(7):388-91.

Jordano J, Almoguera C, Thomas TL. (1989). A sunflower helianthinin gene upstream sequence ensemble contains an enhancer and sites of nuclear protein interaction. *Plant Cell*. 1(9):855-66.

Journet EP, van Tuinen D, Gouzy J, Crespeau H, Carreau V, Farmer MJ, Niebel A, Schiex T, Jaillon O, Chatagnier O, Godiard L, Micheli F, Kahn D, Gianinazzi-Pearson V, Gamas P. (2002). Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. *Nucleic Acids Res*. 30(24):5579-92.

Jurgens G, Mayer U, Torres Ruiz R, Berleth T, Misera S. (1991). Genetic analysis of pattern formation in the *Arabidopsis* embryo. *Dev Suppl* 1: 27-38.

Jurgens G. (1995). Axis formation in plant embryogenesis: cues and clues. *Cell* 81: 467-470.

Jurgens G, Grebe M, Steinmann T. (1997). Establishment of cell polarity during early plant development. *Curr Opin Cell Biol* 9: 849-852.

Jurgens G. (2001). Apical-basal pattern formation in *Arabidopsis* embryogenesis. *EMBO J.* 20(14):3609-16.

Jurgens G. (2003). Growing up green: cellular basis of plant development. *Mech Dev.* 120(11):1395-406.

K

Kaplan DR, Cooke TJ. (1997). Fundamental Concepts in the Embryogenesis of Dicotyledons: A Morphological Interpretation of Embryo Mutants. *Plant Cell.* 9(11):1903-1919.

Karsi A, Patterson A, Feng J, Liu Z. (2002). Translational machinery of channel catfish: I. A transcriptomic approach to the analysis of 32 40S ribosomal protein genes and their expression. *Gene.* 291(1-2):177-86.

Kataoka N, Diem MD, Kim VN, Yong J, Dreyfuss G. (2001). Magoh, a human homolog of *Drosophila* mago nashi protein, is a component of the splicing-dependent exon-exon junction complex. *EMBO J.* 20(22):6424-33.

Keddie JS, Tsiantis M, Piffanelli P, Cella R, Hatzopoulos P, Murphy DJ. (1994). A seed-specific *Brassica napus* oleosin promoter interacts with a G-box-specific protein and may be bi-directional. *Plant Mol Biol.* 24(2):327-40.

Kim I, Hempel FD, Sha K, Pfluger J, Zambryski PC. (2002). Identification of a developmental transition in plasmodesmatal function during embryogenesis in *Arabidopsis thaliana*. *Development.* 129(5):1261-72.

Kim JY, Chung YS, Ok SH, Lee SG, Chung WI, Kim IY, Shin JS. (1999). Characterization of the full-length sequences of phospholipase A2 induced during flower development. *Biochim Biophys Acta.* 1489(2-3):389-92.

Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL. (1999). Imprinting of the MEDEA polycomb gene in the *Arabidopsis* endosperm. *Plant Cell.* 11(10):1945-52.

Klahre U, Noguchi T, Fujioka S, Takatsuto S, Yokota T, Nomura T, Yoshida S, Chua NH. (1998). The *Arabidopsis* DIMINUTO/DWARF1 gene encodes a protein involved in steroid synthesis. *Plant Cell.* 10(10):1677-90.

Knippschild U, Gocht A, Wolff S, Huber N, Lohler J, Stoter M. (2005). The casein kinase 1 family: participation in multiple cellular processes in eukaryotes. *Cell Signal.* 17(6):675-89.

Kohler C, Grossniklaus U. (2005). Seed development and genomic imprinting in plants. *Prog Mol Subcell Biol.* 38:237-62.

Kuang WW, Thompson DA, Hoch RV, Weigel RJ. (1998). Differential screening and suppression subtractive hybridization identified genes differentially expressed in an estrogen receptor-positive breast carcinoma cell line. *Nucleic Acids Res.* 26(4):1116-23.

Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ. (2003). LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell.* 15(1): 5-18.

L

- Lam E.** (2004). Controlled cell death, plant survival and development. *Nat Rev Mol Cell Biol.* 5(4):305-15.
- Lan L, Chen W, Lai Y, Suo J, Kong Z, Li C, Lu Y, Zhang Y, Zhao X, Zhang X, Zhang Y, Han B, Cheng J, Xue Y.** (2004). Monitoring of gene expression profiles and isolation of candidate genes involved in pollination and fertilization in rice (*Oryza sativa* L.) with a 10K cDNA microarray. *Plant Mol Biol.* 54(4):471-87.
- Laux T, Jurgens G.** (1997) Embryogenesis: A New Start in Life. *Plant Cell.* 9(7):989-1000.
- Laux T, Wurschum T, Breuninger H.** (2004). Genetic regulation of embryonic pattern formation. *Plant Cell.* 16 Suppl:S190-202.
- Le Hir H, Gatfield D, Braun IC, Forler D, Izaurralde E.** (2001). The protein Mago provides a link between splicing and mRNA localization. *EMBO Rep.* 2(12):1119-24.
- Lee H, Fischer RL, Goldberg RB, Harada JJ.** (2003). *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci U S A.* 100(4):2152-6.
- Lee JM, Williams ME, Tingey SV, Rafalski JA.** (2002). DNA array profiling of gene expression changes during maize embryo development. *Funct Integr Genomics.* 2(1-2):13-27.
- Leung YF, Cavalieri D.** (2003). Fundamentals of cDNA microarray data analysis. *Trends Genet.* 19(11):649-59.
- Li Z, Thomas TL.** (1998). PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in *Arabidopsis*. *Plant Cell.* 10(3):383-98.
- Li W, Boswell R, Wood WB.** (2000). mag-1, a homolog of *Drosophila* mago nashi, regulates hermaphrodite germ-line sex determination in *Caenorhabditis elegans*. *Dev Biol.* 218(2):172-82.
- Liang P, Pardee AB.** (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science.* 257(5072):967-71.
- Lim CO, Kim HY, Kim MG, Lee SI, Chung WS, Park SH, Hwang I, Cho MJ.** (1996). Expressed sequence tags of Chinese cabbage flower bud cDNA. *Plant Physiol.* 111(2):577-88.
- Lin JF, Wu SH.** (2004). Molecular events in senescing *Arabidopsis* leaves. *Plant J.* 39(4):612-28.
- Livak KJ, Schmittgen TD.** (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* 25(4):402-8.
- Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ.** (1998). *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell.* 93(7):1195-205.
- Louis A, Chiapello H, Fabry C, Ollivier E, Henaut A.** (2002). Deciphering *Arabidopsis thaliana* gene neighborhoods through bibliographic co-citations. *Comput Chem.* 26(5):511-9.
- Lu P, Porat R, Nadeau JA, O'Neill SD.** (1996). Identification of a meristem L1 layer-specific gene in *Arabidopsis* that is expressed during embryonic pattern formation and defines a new class of homeobox genes. *Plant Cell.* 8(12):2155-68.
- Lucas WJ, Bouche-Pillon S, Jackson DP, Nguyen L, Baker L, Ding B, Hake S.** (1995). Selective trafficking of KNOTTED1 homeodomain protein and its mRNA through plasmodesmata. *Science.* 270(5244):1980-3.

- Lukowitz W, Mayer U, Jurgens G.** (1996). Cytokinesis in the *Arabidopsis* embryo involves the syntaxin-related KNOLLE gene product. *Cell*. 84(1):61-71.
- Lukowitz W, Roeder A, Parmenter D, Somerville C.** (2004). A MAPKK kinase gene regulates extra-embryonic cell fate in *Arabidopsis*. *Cell*. 116(1):109-19.
- Lynn K, Fernandez A, Mitsuhiro A, Sedbrook J, Tasaka M, Masson P, Barton MK.** (1999). The PINHEAD/ZWILLE gene acts pleiotropically in *Arabidopsis* development and has overlapping functions with the ARGONAUTE1 gene. *Development*. 126: 469-481.

M

- Mahonen AP, Bonke M, Kauppinen L, Riikonen M, Benfey PN, Helariutta Y.** (2000). A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. *Genes Dev*. 14(23):2938-43.
- Malinowski R, Filipecki M.** (2002). The role of cell wall in plant embryogenesis. *Cell Mol Biol Lett*. 7(4):1137-51.
- Maraschin Sde F, Lamers GE, de Pater BS, Spaik HP, Wang M.** (2003). 14-3-3 isoforms and pattern formation during barley microspore embryogenesis. *J Exp Bot*. 54(384):1033-43.
- Martinez-Garcia JF, Virgos-Soler A, Prat S.** (2002). Control of photoperiod-regulated tuberization in potato by the *Arabidopsis* flowering-time gene CONSTANS. *Proc Natl Acad Sci U S A*. 99(23):15211-6.
- Martinez-Rivas JM, Garcia-Diaz MT, Mancha M.** (2000). Temperature and oxygen regulation of microsomal oleate desaturase (FAD2) from sunflower. *Biochem Soc Trans*. 28(6):890-2.
- Mayer KF, Schoof H, Haecker A, Lenhard M, Jurgens G, Laux T.** (1998). Role of WUSCHEL in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell*. 95(6):805-15.
- Mayer U, Ruiz RAT, Berleth T, Misera S, Jurgens G .** (1991). Mutations affecting body organization in the *Arabidopsis* embryo. *Nature*. 353:402-407.
- Mayer U, Buttner G, Jurgens G.** (1993). Apical-basal pattern formation in the *Arabidopsis* embryo: studies on the role of the *gnom* gene. *Development*. 117:143-54.
- Mazurais D, Montfort J, Delalande C, Gac FL.** (2005). Transcriptional analysis of testis maturation using trout cDNA macroarrays. *Gen Comp Endocrinol*. 142(1-2):143-54.
- McCallum CM, Comai L, Greene EA, Henikoff S.** (2000). Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol*. 123(2):439-42
- McClung JK, Jupe ER, Liu XT, Dell'Orco RT.** (1995). Prohibitin: potential role in senescence, development, and tumor suppression. *Exp Gerontol*. 30(2):99-124.
- McElver J, Tzafrir I, Aux G, Rogers R, Ashby C, Smith K, Thomas C, Schetter A, Zhou Q, Cushman MA, Tossberg J, Nickle T, Levin JZ, Law M, Meinke D, Patton D.** (2001). Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics*. 159(4):1751-63.
- Meinke DW, Sussex IM.** (1979). Embryo-lethal mutants of *Arabidopsis thaliana*. A model system for genetic analysis of plant embryo development. *Dev Biol*. 72(1):50-61.
- Meinke DW.** (1985). Embryo-lethal mutants of *Arabidopsis thaliana*: analysis of mutants with a wide range of lethal phases. *Theor. Appl. Genet*. 69: 543-552.
- Meinke DW.** (1991). Perspectives on Genetic Analysis of Plant Embryogenesis. *Plant Cell*. 3(9):857-866.

Micklem DR, Dasgupta R, Elliott H, Gergely F, Davidson C, Brand A, Gonzalez-Reyes A, St Johnston D. (1997). The mago nashi gene is required for the polarisation of the oocyte and the formation of perpendicular axes in *Drosophila*. *Curr Biol.* 7(7):468-78.

Mohr SE, Dillon ST, Boswell RE. (2001). The RNA-binding protein Tsunagi interacts with Mago Nashi to establish polarity and localize oskar mRNA during *Drosophila* oogenesis. *Genes Dev.* 15(21):2886-99.

Muller C, Denis M, Gentzbittel L, Faraut T. (2004). The Iccare web server: an attempt to merge sequence and mapping information for plant and animal species. *Nucleic Acids Res.* 32(Web Server issue):W429-34.

Murphy D. (2002). Gene expression studies using microarrays: principles, problems, and prospects. *Adv Physiol Educ.* 26(1-4):256-70.

N

Nadeau JA, Zhang XS, Li J, O'Neill SD. (1996). Ovule development: identification of stage-specific and tissue-specific cDNAs. *Plant Cell.* 8(2):213-39.

Nakabayashi K, Ito M, Kiyosue T, Shinozaki K, Watanabe A. (1999). Identification of clp genes expressed in senescing *Arabidopsis* leaves. *Plant Cell Physiol.* 40(5):504-14.

Nakajima K, Sena G, Nawy T, Benfey PN. (2001) Intercellular movement of the putative transcription factor SHR in root patterning. *Nature.* 413(6853):307-11.

Navashin S. (1898). Resultate einer Revision der Befruchtungsvorgänge bei *Lilium martagon* und *Fritillaria tenella*. *Bull Acad Imp Sci St Petersburg* 33, 39-47.

Neuffer MG, Sheridan WF. (1980). Defective kernel mutants of maize. Genetic and lethality studies. *Genetics.* 95:3029-3042.

Newman KL, Fernandez AG, Barton MK. (2002). Regulation of axis determinacy by the *Arabidopsis* PINHEAD gene. *Plant Cell.* 14(12):3029-3042.

Newmark PA, Boswell RE. (1994). The mago nashi locus encodes an essential product required for germ plasm assembly in *Drosophila*. *Development.* 120(5):1303-13.

Newmark PA, Mohr SE, Gong L, Boswell RE. (1997). Mago nashi mediates the posterior follicle cell-to-oocyte signal to organize axis formation in *Drosophila*. *Development.* 124(16):3197-207.

Nguyen C, Rocha D, Granjeaud S, Baldit M, Bernard K, Naquet P, Jordan BR. (1995) Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics.* 29(1):207-16.

Nuccio ML, Thomas TL. (1999). ATS1 and ATS3: two novel embryo-specific genes in *Arabidopsis thaliana*. *Plant Mol Biol.* 39(6):1153-63.

Nunberg AN, Li Z, Bogue MA, Vivekananda J, Reddy AS, Thomas TL. (1994). Developmental and hormonal regulation of sunflower helianthinin genes: proximal promoter sequences confer regionalized seed expression. *Plant Cell.* 6(4):473-86.

O

Ogihara Y, Mochida K, Nemoto Y, Murai K, Yamazaki Y, Shin-I T, Kohara Y. (2003). Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. *Plant J.* 33(6):1001-11.

Ohlrogge J, Benning C. (2000). Unraveling plant metabolism by EST analysis. *Curr Opin Plant Biol.* 3(3):224-228. Review.

P

Pagnussat GC, Yu HJ, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie LF, Ye D, Sundaresan V. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. *Development.* 132(3):603-14.

Palacios IM. (2002). RNA processing: splicing and the cytoplasmic localisation of mRNA. *Curr Biol.* 12(2):R50-2.

Papp I, Mette MF, Aufsatz W, Daxinger L, Schauer SE, Ray A, van der Winden J, Matzke M, Matzke AJ. (2003). Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. *Plant Physiol.* 132(3):1382-90.

Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Lara GG, Holloway E, Kapushesky M, Lilja P, Mukherjee G, Oezcimen A, Rayner T, Rocca-Serra P, Sharma A, Sansone S, Brazma A. (2005) ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 33(Database issue):D553-5.

Pietu G, Alibert O, Guichard V, Lamy B, Bois F, Leroy E, Mariage-Sampson R, Houlgatte R, Soularue P, Auffray C. (1996). Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res.* 6(6):492-503.

Pleite R, Pike MJ, Garces R, Martinez-Force E, Rawsthorne S. (2005). The sources of carbon and reducing power for fatty acid synthesis in the heterotrophic plastids of developing sunflower (*Helianthus annuus* L.) embryos. *J Exp Bot.* 56(415):1297-303.

Pozzoli O, Gilardelli CN, Sordino P, Doniselli S, Lamia CL, Cotelli F. (2004). Identification and expression pattern of mago nashi during zebrafish development. *Gene Expr Patterns.* 5(2):265-72.

Q

Quackenbush J. (2001). Computational analysis of microarray data. *Nat Rev Genet.* 2(6):418-27.

Qiu F, Guo L, Wen TJ, Liu F, Ashlock DA, Schnable PS. (2003). DNA sequence-based "bar codes" for tracking the origins of expressed sequence tags from a maize cDNA library constructed using multiple mRNA sources. *Plant Physiol.* 133(2):475-81.

R

Ray S, Golden T, Ray A. (1996). Maternal effects of the short integument mutation on embryo development in *Arabidopsis*. *Dev Biol.* 180(1):365-9.

Rayet B, Gelinas C. (1999). Aberrant rel/nfkb genes and activity in human cancer. *Oncogene.* 18(49):6938-47.

Reinhardt D, Pesce ER, Stieger P, Mandel T, Baltensperger K, Bennett M, Traas J, Friml J, Kuhlemeier C. (2003). Regulation of phyllotaxis by polar auxin transport. *Nature.* 426(6964):255-60.

Reiser L, Fischer RL. (1993). The Ovule and the Embryo Sac. *Plant Cell.* 5(10):1291-1301.

Revel F, Renard JP, Duranthon V. (1995). PCR-generated cDNA libraries from reduced numbers of mouse oocytes. *Zygote.* 3(3):241-50.

Rienties IM, Vink J, Borst JW, Russinova E, de Vries SC. (2005). The *Arabidopsis* SERK1 protein interacts with the AAA-ATPase AtCDC48, the 14-3-3 protein GF14lambda and the PP2C phosphatase KAPP. *Planta.* 221(3):394-405.

Roncarati R, Salamini F, Bartels D. (1995). An aldose reductase homologous gene from barley: regulation and function. *Plant J.* 7(5):809-22.

Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J, Perteau GM, Lee Y, Karamycheva S, Sultana R, Tsai J, Quackenbush J, Griffiths HM, Restrepo S, Smart CD, Fry WE, Van Der Hoeven R, Tanksley S, Zhang P, Jin H, Yamamoto ML, Baker BJ, Buell CR. (2003). Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol.* 131(2):419-429.

Ruuska SA, Girke T, Benning C, Ohlrogge JB. (2002). Contrapuntal networks of gene expression during *Arabidopsis* seed filling. *Plant Cell.* 14(6):1191-206.

S

Sabatini S, Beis D, Wolkenfelt H, Murfett J, Guilfoyle T, Malamy J, Benfey P, Leyser O, Bechtold N, Weisbeek P, Scheres B. (1999). An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell.* 99(5):463-72.

Scanlon MJ, Stinard PS, James MG, Myers AM, Robertson DS. (1994). Genetic analysis of 63 mutations affecting maize kernel development isolated from Mutator stocks. *Genetics.* 136(1):281-94.

Schadt EE, Monks SA, Drake TA, Lusk AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature.* 422(6929):297-302.

Schena M, Shalon D, Davis RW, Brown PO. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science.* 270(5235):467-70.

Scheres B, Di Laurenzio L, Willemsen V, Hauser M, Janmaat K, Weisbeek P, Benfey PN. (1995). Mutations affecting the radial organisation of the *Arabidopsis* root display specific defects throughout the embryonic axis. *Development.* 121:53-62.

Scholten S, Lorz H, Kranz E. (2002). Paternal mRNA and protein synthesis coincides with male chromatin decondensation in maize zygotes. *Plant J.* 32(2):221-31.

Schrack K, Mayer U, Horrichs A, Kuhnt C, Bellini C, Dangl J, Schmidt J, Jurgens, G. (2000). FACKEL is a sterol C-14 reductase required for organized cell division and expansion in *Arabidopsis* embryogenesis. *Genes Dev.* 14(12):1471-1484.

Schulz R, Jensen W. (1968). Capsella embryogenesis: the early embryo. *J Ultrastruct Res.* 22(5):376-92.

Schwartz BW, Yeung EC, Meinke DW. (1994). Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of *Arabidopsis*. *Development.* 120:3235-3245

Schwartz BW, Vernon DM, Meinke DW. (1997). Development of the suspensor: Differentiation, communication and programmed cell death during plant embryogenesis. In "Cellular and Molecular Biology of Plant Seed Development" (B. A. Larkins and I. K. Vasil, Eds.), vol 2, pp. 53-72. Kluwer Academic Press, Dordrecht, The Netherlands.

Serrano-Vega MJ, Venegas-Calero M, Garces R, Martinez-Force E. (2003). Cloning and expression of fatty acids biosynthesis key enzymes from sunflower (*Helianthus annuus* L.) in *Escherichia coli*. *J Chromatogr B Analyt Technol Biomed Life Sci.* 786(1-2):221-8.

Serrano-Vega MJ, Garces R, Martinez-Force E. (2005). Cloning, characterization and structural model of a FatA-type thioesterase from sunflower seeds (*Helianthus annuus* L.). *Planta.*

Sessions A, Weigel D, Yanofsky MF. (1999). The *Arabidopsis thaliana* MERISTEM LAYER 1 promoter specifies epidermal expression in meristems and young primordia. *Plant J.* (2):259-63.

- Sessions A, Yanofsky MF, Weigel D.** (2000). Cell-cell signaling and movement by the floral transcription factors LEAFY and APETALA1. *Science*. 289(5480):779-82.
- Shanklin J, Cahoon EB.** (1998). Desaturation and related modifications of fatty acids. *Annu Rev Plant Physiol Plant Mol Biol*. 49:611-641.
- Skriver K, Mundy J.** (1990) Gene expression in response to abscisic acid and osmotic stress. *Plant Cell*. 2(6):503-12.
- Slade AJ, Knauf VC.** (2005). TILLING moves beyond functional genomics into crop improvement. *Transgenic Res*. 14,109-15.
- Souter M, Lindsey K.** (2000). Polarity and signalling in plant embryogenesis. *J Exp Bot* 51: 971-983.
- Sparkes IA, Brandizzi F, Slocombe SP, El-Shami M, Hawes C, Baker A.** (2003). An *Arabidopsis* pex10 null mutant is embryo lethal, implicating peroxisomes in an essential role during plant embryogenesis. *Plant Physiol*. 133(4):1809-1819.
- Springer PS, Holding DR, Groover A, Yordan C, Martienssen RA.** (2000). The essential Mcm7 protein PROLIFERA is localized to the nucleus of dividing cells during the G(1) phase and is required maternally for early *Arabidopsis* development. *Development*. 127(9):1815-22.
- Sprunck S, Baumann U, Edwards K, Langridge P, Dresselhaus T.** (2005). The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.). *Plant J*. 41(5):660-72.
- Sreenivasulu N, Altschmied L, Panitz R, Hahnel U, Michalek W, Weschke W, Wobus U.** (2002). Identification of genes specifically expressed in maternal and filial tissues of barley caryopses: a cDNA array analysis. *Mol Genet Genomics*. 266(5):758-67.
- Sreenivasulu N, Altschmied L, Radchuk V, Gubatz S, Wobus U, Weschke W.** (2004). Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains. *Plant J*. 37(4):539-53.
- Stebbins-Boaz B, Richter JD.** (1997). Translational control during early development. *Crit Rev Eukaryot Gene Expr*. 7(1-2):73-94.
- Steinmann T, Geldner N, Grebe M, Mangold S, Jackson CL, Paris S, Galweiler L, Palme K, Jurgens G.** (1999). Coordinated polar localization of auxin efflux carrier PIN1 by GNOM ARF GEF. *Science*. 286(5438):316-8.
- Stekel D, Git Y, Falciani F.** (2000). The comparison of gene expression from multiple cDNA libraries. *Genome Res* 10: 2055-2061.
- Suh MC, Kim MJ, Hur CG, Bae JM, Park YI, Chung CH, Kang CW, Ohlrogge JB.** (2003). Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. *Plant Mol Biol*. 52(6):1107-23.
- Swidzinski JA, Zaplachinski ST, Chuong SD, Wong JF, Muench DG.** (2001). Molecular characterization and expression analysis of a highly conserved rice mago nashil homolog. *Genome*. 44(3):394-400.

T

- Tamborindeguy C, Ben C, Liboz T, Gentzbittel L.** (2004a). Sequence evaluation of four specific cDNA libraries for developmental genomics of sunflower. *Mol Genet Genomics*. 271(3):367-75.

Tamborindeguy C, Ben C, Jardinaud F, Gentzbittel L, Liboz T. (2004b). Mass cloning of differential and nondifferential transcript-derived fragments from cDNA-AFLP experiments in sunflower. *Plant Mol Bio Rep.* 22:367-75.

Tamborindeguy C, Ben C, Liboz T, Petitprez M, Gentzbittel L. (soumis). Towards a functional analysis of plant embryonic polarity acquisition: Comparison of transcriptional profiling of sunflower early somatic embryogenesis using cDNA microarrays.

Tang S, Yu JK, Slabaugh B, Shintani K, Knapp J. (2002). Simple sequence repeat map of the sunflower genome. *Theor Appl Genet.* 105(8):1124-1136.

Tang X, Shen L, Dickerson JA. (2005). BarleyExpress: a web-based submission tool for enriched microarray database annotations. *Bioinformatics.* 21(3):399-401.

Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Young K, Taylor NE, Henikoff JG, Comai L, Henikoff S. (2003). Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res.* 13(3):524-30.

Timmermans MC, Hudson A, Becraft PW, Nelson T. (1999). ROUGH SHEATH2: a Myb protein that represses knox homeobox genes in maize lateral organ primordia. *Science.* 284(5411):151-3.

Topping JF, Agyeman F, Henricot B, Lindsey K. (1994). Identification of molecular markers of embryogenesis in *Arabidopsis thaliana* by promoter trapping. *Plant J.* 5(6):895-903.

Tu Q, Tang H, Ding D. (2004). MedBlast: searching articles related to a biological sequence. *Bioinformatics.* 20(1):75-7.

U

Ulmasov T, Hagen G, Guilfoyle TJ. (1999). Activation and repression of transcription by auxin-response factors. *Proc Natl Acad Sci U S A.* 96(10):5844-9.

Uwer U, Willmitzer L, Altmann T. (1998). Inactivation of a glycyl-tRNA synthetase leads to an arrest in plant embryo development. *Plant Cell.* 10(8):1277-94.

V

Van de Loo FJ, Turner S, Somerville C. (1995). Expressed Sequence Tags from Developing Castor Seeds. *Plant Physiol* 108(3):1141-1150.

Van Engelen FA, De Vries SC. (1992). Extracellular proteins in plant embryogenesis. *Trends Genet.* 8(2):66-70.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. (1995). Serial analysis of gene expression. *Science.* 270(5235):484-7.

Verica JA, Maximova SN, Strem MD, Carlson JE, Bailey BA, Gultinan MJ. (2004). Isolation of EST from cacao (*Theobroma cacao* L.) leaves treated with inducers of the defense response. *Plant Cell Rep.* 23(6):404-13.

Vielle-Calzada JP, Baskar R, Grossniklaus U. (2000). Delayed activation of the paternal genome during seed development. *Nature.* 404(6773):91-4.

Volinia S, Evangelisti R, Francioso F, Arcelli D, Carella M, Gasparini P. (2004). GOAL: automated Gene Ontology analysis of expression profiles. *Nucleic Acids Res.* 32(Web Server issue):W492-9.

Vonder Haar RA, Allen RD, Cohen EA, Nessler CL, Thomas TL. (1988). Organization of the sunflower 11S storage protein gene family. *Gene.* 74(2):433-43.

Vroemen C, de Vries S, Quatrano R. (1999). Signalling in plant embryos during the establishment of the polar axis. *Semin Cell Dev Biol.* 10(2):157-64.

W

Wan J, Dunning FM, Bent AF. (2002). Probing plant-pathogen interactions and downstream defense signaling using DNA microarrays. *Funct Integr Genomics.* 2(6):259-73.

Wang H, Caruso LV, Downie AB, Perry SE. (2004). The embryo MADS domain protein AGAMOUS-Like 15 directly regulates expression of a gene encoding an enzyme involved in gibberellin metabolism. *Plant Cell.* 16(5):1206-19.

Watt DA. (2003). Aluminium-responsive genes in sugarcane: identification and analysis of expression under oxidative stress. *J Exp Bot.* 54(385):1163-74.

Weijers D, Geldner N, Offringa R, Jurgens G. (2001). Seed development: Early paternal gene activity in *Arabidopsis*. *Nature.* 414(6865):709-10.

Weijers D, Jurgens G. (2005) Auxin and embryo axis formation: the ends in sight? *Curr Opin Plant Biol.* 8(1):32-7.

Wellenreuther R, Schupp I, Poustka A, Wiemann S; The German cDNA Consortium. (2004). SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones. *BMC Genomics.* 5(1):36.

West M, Harada JJ. (1993). Embryogenesis in Higher Plants: An Overview. *Plant Cell.* 5(10):1361-1369.

White JA, Todd J, Newman T, Focks N, Girke T, de Ilarduya OM, Jaworski JG, Ohlrogge JB, Benning C. (2000). A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiol* 124(4):1582-94.

Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS. (2001). Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol.* 8(6):625-37.

Wu CH, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu ZZ, Ledley RS, Lewis KC, Mewes HW, Orcutt BC, Suzek BE, Tsugita A, Vinayaka CR, Yeh LS, Zhang J, Barker WC. (2002). The Protein Information Resource: an integrated public resource of functional annotation of proteins. *Nucleic Acids Res.* 30(1):35-7.

Wu HM, Cheun AY. (2000). Programmed cell death in plant reproduction. *Plant Mol Biol.* 44(3):267-81.

X

Xiong L, Lee MW, Qi M, Yang Y. (2001). Identification of defense-related rice genes by suppression subtractive hybridization and differential screening. *Mol Plant Microbe Interact.* 14(5):685-92.

Xu FX, Chye ML. (1999). Expression of cysteine proteinase during developmental events associated with programmed cell death in brinjal. *Plant J.* 17(3):321-7.

Y

Yadegari R, Paiva G, Laux T, Koltunow AM, Apuya N, Zimmerman JL, Fischer RL, Harada JJ, Goldberg RB. (1994). Cell Differentiation and Morphogenesis Are Uncoupled in *Arabidopsis* raspberry Embryos. *Plant Cell.* 6(12):1713-1729.

Yang WC, Sundaresan V. (2000). Genetics of gametophyte biogenesis in *Arabidopsis*. *Curr Opin Plant Biol.* (1):53-7.

Yanovsky MJ, Kay SA. (2003). Living by the calendar: how plants know when to flower. *Nat Rev Mol Cell Biol.* 4(4):265-75.

Ye R, Yao QH, Xu ZH, Xue HW. (2004). Development of an efficient method for the isolation of factors involved in gene transcription during rice embryo development. *Plant J.* 38(2):348-57.

Yeung EC, Meinke DW. (1993). Embryogenesis in Angiosperms: Development of the Suspensor. *Plant Cell.* 5(10):1371-1381.

Z

Zhang L, Ma XL, Zhang Q, Ma CL, Wang PP, Sun YF, Zhao YX, Zhang H. (2001). Expressed sequence tags from a NaCl-treated *Suaeda salsa* cDNA library. *Gene.* 267(2):193-200.

Zhang JZ, Somerville CR. (1997). Suspensor-derived polyembryony caused by altered expression of valyl-tRNA synthetase in the *twn2* mutant of *Arabidopsis*. *Proc Natl Acad Sci U S A.* 94(14):7349-55.

Zhang X, Beuron F, Freemont PS. (2002). Machinery of protein folding and unfolding. *Curr Opin Struct Biol.* 12(2):231-8.

Zhao XF, Colaizzo-Anas T, Nowak NJ, Shows TB, Elliott RW, Aplan PD. (1998). The mammalian homologue of *mago nashi* encodes a serum-inducible protein. *Genomics.* 47(2):319-22

Zhao XF, Nowak NJ, Shows TB, Aplan PD. (2000). MAGOH interacts with a novel RNA-binding protein. *Genomics.* 63(1):145-8.

Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques.* 30(4):892-7.

Zhu T, Budworth P, Chen W, Provart N, Chang H, Guimil S, SuW, Estes B, Zou G, Wang X. (2003). Transcriptional control of nutrient partitioning during rice grain filling. *Plant Biotechnology Journal.* 1:59-70.

Zimmerman JL. (1993). Somatic Embryogenesis: A Model for Early Development in Higher Plants. *Plant Cell.* 5(10):1411-1423.

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W. (2004). GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.* 136(1):2621-32.