



HAL
open science

Vers une meilleure compréhension du rôle des gènes LEC par analyse fonctionnelle de variants

Camille Salaün

► **To cite this version:**

Camille Salaün. Vers une meilleure compréhension du rôle des gènes LEC par analyse fonctionnelle de variants. Biologie moléculaire. Université Paris-Saclay, 2023. Français. NNT : 2023UPASB005 . tel-04368891

HAL Id: tel-04368891

<https://theses.hal.science/tel-04368891v1>

Submitted on 2 Jan 2024

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

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

Vers une meilleure compréhension du rôle des gènes *LEC* par analyse fonctionnelle de variants

*Towards a better understanding of LEC genes
role(s) by functional analysis of variants*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°567 : Sciences du végétal : du gène à l'écosystème (SEVE)

Spécialité de doctorat : Biologie

Graduate School : BioSphERA - Biologie, Société, Ecologie &
Environnement, Ressources, Agriculture & Alimentation

Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche : Institut Jean-Pierre Bourgin
(Université Paris-Saclay, INRAE, AgroParisTech)
sous la direction de **Bertrand DUBREUCQ**, directeur de recherche

Thèse soutenue à Versailles, le 03 février 2023, par

Camille SALAÜN

Composition du Jury

Membres du jury avec voix délibérative

Jean-Denis FAURE

Professeur, AgroParisTech
Université Paris-Saclay

Président

Julia BUITINK

Directrice de recherche, INRAE
Université d'Angers

Rapporteur & examinatrice

Vanessa VERNOUD

Chargée de recherche, HDR, INRAE
Université Bourgogne Franche-Comté

Rapporteur & examinatrice

Julie BOUDET

Maîtresse de conférences
Université Clermont-Auvergne

Examinatrice

Benoit LANDREIN

Chargé de recherche, CNRS
Ecole Normale Supérieure de Lyon

Examineur

Thomas WIDIEZ

Directeur de recherche, INRAE
Ecole Normale Supérieure de Lyon

Examineur

Titre : Vers une meilleure compréhension du rôle des gènes LEC par analyse fonctionnelle de variants

Mots clés : développement, graine, embryogenèse, contrôle transcriptionnel, facteurs pionniers, LEC2

Résumé : Les plantes peuvent se reproduire de manière sexuée ou asexuée. Ces deux types de reproduction conduisent à la formation d'un embryon, zygotique ou somatique, à l'origine de la future plante. Ces processus développementaux sont régis par de fortes régulations de l'expression des gènes, notamment par des facteurs de transcription (FT). LEAFY COTYLEDON 2 (LEC2) est un FT connu pour être un régulateur maître de l'embryogenèse et du développement de la graine. Avec ABCISSIC ACID INSENSITIVE 3 (ABI3), FUSCA 3 (FUS3) et LEAFY COTYLEDON 1 (LEC1), LEC2 forme le groupe des LAFL et contrôle divers aspects du développement de la graine, allant de la formation de l'embryon à l'accumulation des composés de réserve, ou encore l'acquisition de la tolérance à la dessiccation. Au-delà de leur rôle dans la graine, les LAFL sont également essentiels à l'initiation et au maintien de la capacité embryogène des cellules végétales.

Mon travail de thèse vise à mieux connaître la structure, la fonction et le mode d'action de LEC2, qui restent encore mal compris. Différentes approches de mutagenèse ont été mises en place, de manière à générer de nouveaux variants de LEC2, présentant une activité différentielle, que ce soit dans les processus d'embryogenèse somatique ou au cours de la maturation de la graine. Les résultats obtenus ont montré la présence de domaines d'activation de la transcription (DA) dans les extrémités non structurées de la protéine et jusqu'alors non identifiés. Des mutations réalisées dans ces DA suggèrent qu'ils ne sont pourtant pas indispensables à l'activité de LEC2 *in planta*, bien qu'ils soient fortement conservés chez les plantes dicotylédones. Des analyses associant des données de RNA-seq, de CHIP-seq, d'ATAC-seq, ainsi que des données fonctionnelles, semblent mettre en évidence une double fonction de LEC2 jusqu'ici inconnue, comme activateur transcriptionnel d'une part, mais aussi comme facteur pionnier, capable de se lier à la chromatine et de modifier l'accessibilité de ses cibles sans en activer la transcription. Ces résultats permettent d'éclairer le rôle des gènes *LEC1* et *LEC2* dans les processus allant de la différenciation cellulaire à l'accumulation des réserves.

Title : Towards a better understanding of LEC genes roles by functional analysis of variants.

Keywords : development, seed, embryogenesis, transcriptional regulation, pioneer factors, LEC2

Abstract : Plants can reproduce sexually or asexually. Both types of reproduction lead to the formation of an embryo, either zygotic or somatic, which give rise to the future plant. These developmental processes are controlled by a strong regulation of gene expression, notably by transcription factors (TFs). LEAFY COTYLEDON 2 (LEC2) is a TF known to be a master regulator of embryogenesis and seed development. Together with ABCISSIC ACID INSENSITIVE 3 (ABI3), FUSCA 3 (FUS3) and LEAFY COTYLEDON 1 (LEC1), LEC2 forms the LAFL group of transcription factors, and controls various aspects of seed development, ranging from embryo formation to accumulation of reserve compounds, or acquisition of desiccation tolerance. Beyond their role in the seed, LAFLs are also essential for the initiation and maintenance of the embryogenic capacity of plant cells.

My thesis work aims at better understanding the structure, function, and mode of action of LEC2, which are still poorly understood. Different mutagenesis approaches have been implemented to generate new variants of LEC2 with a differential activity during somatic embryogenesis or seed maturation. The results obtained showed the presence of transcription activation domains (ADs) within the unstructured ends of the protein and never identified so far. Interestingly, mutations in these ADs suggest that they are, however, not essential for LEC2 activity *in planta*, despite their high conservation in dicotyledonous plants. The combination of RNA-seq, CHIP-seq, and ATAC-seq data, together with functional analysis seem to reveal a double function of LEC2, previously unknown, as a transcriptional activator on the one hand, but also as a pioneer factor in the other hand, able to bind to chromatin and modify the accessibility of its targets without activating their transcription. These results provide insight into the role of *LEC1* and *LEC2* in processes ranging from cell differentiation to reserve accumulation.

Remerciements

J'avais pourtant bien dit que je ne voulais pas faire de thèse ! Et pourtant, s'il y a une chose que je ne regrette pas, c'est bien d'avoir tenté cette aventure, qui a été formatrice à tant de niveaux. J'avais aussi dit que si je faisais une thèse, ce serait avec Bertrand et personne d'autre, car j'ai vu dès mon stage de M1, puis de M2, que je ne pouvais pas rêver meilleur directeur de thèse (sans vouloir faire de zèle, ce n'est pas mon genre !...). Merci Bertrand de m'avoir encadrée avec autant de patience, il en fallait avec moi, d'avoir su rester toujours positif face à mes doutes et mes angoisses, de m'encourager et de m'aider à prendre confiance en moi. Je sais que je n'ai pas été facile tous les jours, la preuve en est de ce cri de détresse : « Martine ! Elle va finir par me tuer cette doctorante ! ». Je ne compte pas tout ce que j'ai appris pendant ces quatre années en cumulé, que j'ai passées sous ta supervision. Merci infiniment.

Martine, merci pour les nombreux conseils que tu m'as donnés, l'aide que tu m'as apportée, que ce soit pour des manipes en tout genre ou pour des relectures de thèse, de revue, de lettres de motivation et même de mails ! Merci surtout de rigoler à mes blagues, c'est toujours un vrai bonheur d'entendre ton rire éclater dans le labo ! Johanne, merci d'avoir été une super maîtresse de stage en M1, de m'avoir donné envie de revenir en M2 et de poursuivre en thèse. Merci pour tes encouragements pendant ces années, ta douceur et ta gentillesse ! Quelle joie de mettre toutes les trois notre « Grand Chef » en boîte !

Merci à Loïc pour son accueil dans l'équipe, la passion et l'énergie dont il fait preuve sont très motivantes. Je remercie tous les membres de l'équipe SEEDDEV, en particulier Alexandra, qui m'a appris à dompter les levures, Delphine avec qui j'ai eu la joie de partager le bureau pendant un temps et qui m'a appris à faire des western-blots et Massi qui m'a apporté une aide considérable pour l'analyse de RNA-seq.

Je remercie nos voisins de labo, les membres de l'équipe PHYGERM, en particulier le Professeur RAJJOU, de m'avoir poussée à faire une thèse et de m'encourager avec des verres de rhum, Anne et Abdel avec qui il est toujours agréable de discuter dans les couloirs, et les filles de DYSCOL, Carine, Isabelle et Christelle, pour leur gentillesse, leur sourire et leur aide pour la biochimie. Merci aussi aux habitants du 1^{er} étage pour leur accueil, ainsi que Nico, Patrick, Jean-Christophe et tous nos amis du sous-sol, avec qui nous passons toujours de très bons moments lorsqu'ils remontent à la surface pour boire un café au soleil ou une bière après le travail.

J'ai vite compris pendant ma courte expérience que l'on n'arrive pas à faire grand-chose tout seul et que l'expertise de chacun, de l'institut ou d'ailleurs, est d'une aide précieuse. En cela, je dois énormément à Lionel, dont le support m'a été indispensable pour la réalisation d'un bon nombre des manipes et qui, même une fois parti aux États-Unis, a continué d'assurer un service après-vente et a répondu à nos appels au secours. Je remercie Fabien et Anouchka qui ont travaillé avec nous sur la mise en place de la

mutagenèse CRISPR, Jasmine de m'avoir aidée à traiter mes données *ADpred*, Jérôme VERDIER qui nous a sauvé la vie en nous rendant lisibles les données d'ATAC-seq. Merci aussi à François ROUDIER et à Daniel BOUYER de m'avoir appris à faire du ChIP-seq ainsi que toutes les personnes qui m'ont gentiment accueillie pendant une semaine au RDP à Lyon. Je remercie tous ceux qui m'ont aidée pour des manipes, qui m'ont prodigué des conseils tout au long de ma thèse.

Je remercie le LabEx SPS qui a financé mon contrat doctoral, l'IJPB et toutes les personnes qui contribuent à nous offrir un super cadre de travail : les serristes, les plateformes, le secrétariat, les référents de thèse... Merci à l'école doctorale Sciences du Végétal pour son accompagnement et à l'association Doc' en Herbe dont j'ai eu la chance d'être présidente quelques temps, ainsi que mes collègues du bureau, Thibault, Sami, Garry et Antoine D.

Je remercie les membres de mon comité de thèse, Hervé ETIENNE, Fabien NOGUÉ et Jérôme VERDIER de m'avoir accompagnée. Je remercie également Julia BUITINK, Vanessa VERNOUD, Julie BOUDET, Benoit LANDREIN, Thomas WIDIEZ et Jean-Denis FAURE d'avoir accepté de faire partie de mon jury de thèse et d'évaluer mon travail.

Enfin, ces années de thèse n'auraient pas été si belles sans mes copains non permanents Sami, Justine, Houssein et Romane. Ça a commencé avec une bière et un camembert partagés dans la cafète, puis les apéros covid et houmous à partir de 18h (couvre-feu oblige) chez Justine, les bars Versaillais, et j'en passe. L'amitié que nous avons su créer m'est très précieuse. Merci aussi à Léa, la petite dernière, Jean, Damien, ma chère Nathalie (et Victor, toujours prêt à rincer les doctorants en perdition).

Antoine, merci d'être présent chaque jour, de me supporter, m'encourager, m'écouter, me consoler, me rassurer, de discuter mes résultats, même si ça me fait râler. On a vécu ta rédaction, on a vécu la mienne, maintenant, on va pouvoir s'amuser ! J'ai hâte de poursuivre cette aventure à tes côtés.

Je remercie mon amie de toujours, Camille D. pour son soutien sans faille, Jeanne qui mène le même parcours que le mien, mais en découpant des cerveaux de souris. Nos séances télétravail/rédaction de thèse étaient très fructueuses. Merci à Vincent, mon éternel binôme de M1, cette personne formidable à qui je peux demander de faire des stats à 22h30 pour le lendemain matin (« j'espère que je serai dans tes remerciements de thèse ! » – voilà !). Merci à mes chers copains Rémi, Marie, Léonard et à la TFB. Merci aussi à mes vieux copains Normands de la rue des Moulins pour leur gentillesse et leur folie.

Je remercie mes parents adorés, de toujours m'épauler et m'encourager dans ce que j'entreprends, mes grandes sœurs et mes neveux que j'aime tant !

Pour terminer, je tiens à remercier LEC2, mon compagnon de route depuis le M1, qui m'en a fait voir de toutes les couleurs, mais que je n'ai pas épargné non plus, à le couper en morceaux et à le cloner dans tous les sens. LEC2, protéine de m..., qui va sacrément me manquer !

Liste des abréviations

AA : acide aminé

ABA : acide abscissique

ABI3 : ABSCISIC ACID INSENSITIVE 3

ADN : acide désoxyribonucléique

AGL15, AGL62 : AGAMOUS-LIKE 15, 62

AID : ACTIVATION-INDUCED CYTIDINE DEAMINASE

AP2-ERF : APETALA 2/ETHYLENE RESPONSE FACTOR

ARN : acide ribonucléique

At : *Arabidopsis thaliana*

BBM : BABY BOOM

Bn : *Brassica napus*

CME : *Cold Motif Element*

CRISPR : *Clustered Regularly Interspaced Short Palindromic Repeats*

DA : domaine d'activation de la transcription

dCas9 : *catalytically dead Cas9*

EFS : EARLY FLOWERING IN SHORT DAYS

EMF : EMBRYONIC FLOWER

ES : embryogenèse somatique

ET : élément transposable

EZ : embryogenèse zygotique

FIE : FERTILIZATION INDEPENDENT ENDOSPERM

FIS : FERTILIZATION INDEPENDENT SEED

FLC : FLOWERING LOCUS C

FRI : FRIGIDA

FT : facteur de transcription

FUS3 : FUSCA 3

GA : acide gibbérellique / gibbérellines

GFP : GREEN FLUORESCENT PROTEIN

HAT : histone acétyltransférase

KO : *knock-out*

L1L : LEC1-LIKE

LAF1 : LEC1, ABI3, FUS3, LEC2
LEC1, LEC2 : LEAFY COTYLEDON 1, 2
MAC : méristème apical caulinaire
MAR : méristème apical racinaire
MCP : MS2 COAT PROTEIN
MEA : MEDEA
miRNA : ARN non codant
MS1 : MULTICOPY SUPPRESSOR OF IRA1
nCas9 : *nickase Cas9*
NDR : *Nucleosome-Depleted Region*
NF-Y : NUCLEAR TRANSCRIPTION FACTOR Y
NFR : *Nucleosome-Free Region*
NLS : *Nuclear Localization Signal*
OLE1 : OLEOSIN 1
p : promoteur
PAM : *Protospacer Adjacent Motif*
pb : paire de bases
PcG : groupe polycomb
PIC : *Pre-Initiation Complex*
PKL : PICKLE
PLT2 : PLETHORA 2
PRC1, PRC2 : POLYCOMB REPRESSIVE COMPLEX 1, 2
PRE : *Polycomb Responsive Element*
RdDM : RNA-directed DNA-methylation
RLE : *cis-Repressive LEC2 Element*
TrxG : groupe trithorax
TSS : *Transcription Starting Site*
VAL1, VAL2 : VIVIPAROUS1/ABI3-LIKE 1, 2
VRN : VERNALIZATION
WOX : WUSCHEL RELATED HOMEODOMAIN
WUS : WUSCHEL

Nomenclature

Pour plus de précision dans le vocabulaire, j'ai parfois choisi d'utiliser les termes anglais pour définir les principes, inscrits alors en italique dans le texte.

pLEC2 : promoteur du gène

LEC2 : gène

LEC2 : protéine

lec2 : gène muté

lec2 : protéine mutée

[lec2] : phénotype du mutant

Δ LEC2₁₋₈₅ : délétion des acides aminés 1 à 85 de la protéine LEC2

LEC2 ATG-M104 : fragment de la protéine LEC2 allant de l'ATG jusqu'à la méthionine en position 104, substituée en codon STOP.

S50 : acide aminé sérine situé en position 50 de la protéine

Y₂₆₅A : substitution d'une tyrosine en alanine en position 265 de la protéine

:: : fusion entre un promoteur et un gène

: : fusion traductionnelle entre deux gènes

Table des matières

INTRODUCTION	1
1. REPRODUCTION DES PLANTES	1
1.1. Reproduction zygotique.....	1
1.2. Reproduction clonale.....	5
2. CONTROLE DE L'EXPRESSION DES GENES	6
2.1. Les régulations chromatinienne : contrôle de l'accessibilité des régions régulatrices de l'expression des gènes.....	7
2.2. Les promoteurs, lieu de fixation de la machinerie basale de transcription	9
2.3. Les facteurs de transcription, régulation ciblée de l'expression	10
2.4. Outils pour l'étude des interactions protéines/ADN/chromatine	14
3. REGULATION DU DEVELOPPEMENT ET DU REMPLISSAGE DE LA GRAINE.....	15
3.1. Contrôle épigénétique du développement de la graine.....	15
3.2. Contrôle hormonal du développement de la graine	16
3.3. Contrôle transcriptionnel du développement de la graine.....	18
4. REVUE : « CONTROLE GENETIQUE ET MOLECULAIRE DE L'EMBRYOGENESE SOMATIQUE »	24
5. OBJECTIFS DE LA THESE.....	41
CHAPITRE 1 : APPROCHE SANS A PRIORI	42
1. INTRODUCTION.....	42
2. RESULTATS.....	43
2.1. Génération et validation des surexprimeurs de LEC2.....	43
2.2. Choix d'un outil de mutagenèse <i>in planta</i>	46
2.3. Mise en place de l'approche « CRISPR-X ».....	47
2.4. Approche « EvolvR ».....	50
3. DISCUSSION ET DERNIERE APPROCHE ENVISAGEE	51
4. QUEL PROMOTEUR POUR SUREPRIMER <i>LEC2</i> ?	53
4.1. Phénotypes surexprimeurs de <i>AtLEC2</i> sous contrôle de <i>pAtUbq10</i>	54
4.2. Le phénotype surexporteur est plus fort chez le mutant <i>lec2-1</i>	55
4.1. L'utilisation de <i>pAtUbq10</i> vaut-elle mieux que celle de <i>pCaMV35S</i> ?.....	55
CHAPITRE 2 : APPROCHE AVEC A PRIORI	57
1. INTRODUCTION.....	57
1.1. Données préliminaires : une partie de l'extrémité N-terminale est nécessaire à l'activité transcriptionnelle de LEC2 dans la mousse.....	57

1.2.	Affinement de la zone de régulation en N-ter de LEC2	58
2.	MISE EN EVIDENCE DU SIGNAL DE LOCALISATION NUCLEAIRE (NLS) DE LEC2.....	59
2.1.	Un NLS est prédit en partie centrale de AtLEC2	59
2.2.	Validation expérimentale de la prédiction du NLS de AtLEC2	60
2.3.	Localisation de variants de LEC2 <i>in planta</i>	60
2.4.	Conclusion et discussion	61
3.	MISE EN EVIDENCE DES DOMAINES D'ACTIVATION DE LA TRANSCRIPTION DE LEC2	62
3.1.	Des domaines d'activation de la transcription sont prédits en N et C-ter de LEC2.....	62
3.2.	Validation expérimentale de la prédiction des DA de AtLEC2.....	63
3.3.	L'absence de DA réduit l'activité transcriptionnelle de AtLEC2 dans la mousse	64
3.4.	Conclusion et discussion	65
4.	FONCTIONNALITE DE LEC2 SANS DOMAINE D'ACTIVATION <i>IN PLANTA</i>	66
4.1.	LEC2 sans domaine d'activation restaure le phénotype sauvage <i>in planta</i>	67
4.2.	Effet des délétions sur le contenu en réserves des graines	68
4.3.	Conclusion et discussion	70
5.	CAS PARTICULIER : LE VARIANT LEC2 Y ₂₅₇ A Y ₂₆₅ A	71
6.	DISCUSSION GENERALE SUR L'APPROCHE A <i>PRIORI</i> REALISEE SUR LEC2	73
	CHAPITRE 3 : DOUBLE FONCTION DE LEC2	76
1.	INTRODUCTION.....	76
2.	LA PERTE D'UN DOMAINE D'ACTIVATION NE SEMBLE PAS IMPACTER LE TRANSCRIPTOME DES GRAINES EN DEVELOPPEMENT	76
3.	ACTION TRANSCRIPTIONNELLE DE LEC2 SUR SES CIBLES	77
3.1.	LEC2 se lie à des cibles et active leur transcription	78
3.2.	LEC2 se lie à des cibles sans activer leur transcription.....	78
4.	ACTION TRANSCRIPTIONNELLE DE LEC2 ET D'OUVERTURE DE LA CHROMATINE SUR SES CIBLES	79
4.1.	Certains <i>loci</i> sont différentiellement accessibles chez le mutant <i>lec2</i>	79
4.2.	Analyse croisée des données de RNA-seq, CHIP-seq et ATAC-seq.....	79
5.	CONCLUSION ET DISCUSSION	80
	DISCUSSION GENERALE.....	83
1.	LA POSITION DU NLS DE LEC2 N'A PAS PU ETRE CONFIRMEE.....	83
2.	LA REALISATION D'UN CRIBLE SUR L'APPARITION D'EMBRYONS SOMATIQUES EST-ELLE PERTINENTE ?	84
3.	EMBRYOGENESE SOMATIQUE OU ZYGOTIQUE ? LEC2, RESPONSABLE D'UN « ETAT EMBRYONNAIRE »	86
4.	LES DOMAINES D'ACTIVATION DU FACTEUR DE TRANSCRIPTION LEC2.....	87
4.1.	Pourquoi avoir un FT avec des DA dispensables ?.....	87

4.2. Conservation de LEC2 à travers les espèces	89
PERSPECTIVES.....	91
MATERIEL ET METHODES.....	93
1. MATERIEL.....	93
1.1. Matériel végétal.....	93
1.2. Souches bactériennes	93
1.3. Souche de levure.....	93
1.4. Vecteurs plasmidiques.....	94
2. METHODES.....	94
2.1. Culture et transformation des plantes.....	94
2.2. Biologie moléculaire	98
2.3. Biochimie	110
2.4. Cytologie et imagerie.....	116
2.5. Analyses bio-informatiques.....	117
REFERENCES BIBLIOGRAPHIQUES	120
ANNEXES	132
Annexe 1 : Séquences des ARN guides utilisés pour la mutagenèse CRISPR-X.....	133
Annexe 2 : Amorces pour le génotypage des plantes.....	134
Annexe 3 : Autres amorces utilisées pour les clonages ou la vérification des clonages et des plantes transformées.....	134
Annexe 4 : Liste des variants et fragments de LEC2 clonés	137
Annexe 5 : Séquences des 60 orthologues de AtLEC2 utilisés pour la prédiction et la conservation des domaines d'activation de la transcription	144
Annexe 6 : Exemples de traces d'ATAC-seq et ChIP-seq obtenues pour plusieurs gènes et représentatives des profils pouvant être observés	151
Annexe 7 : Milieux et solutions utilisés pour la culture et la transformation des mousses.....	159

Introduction

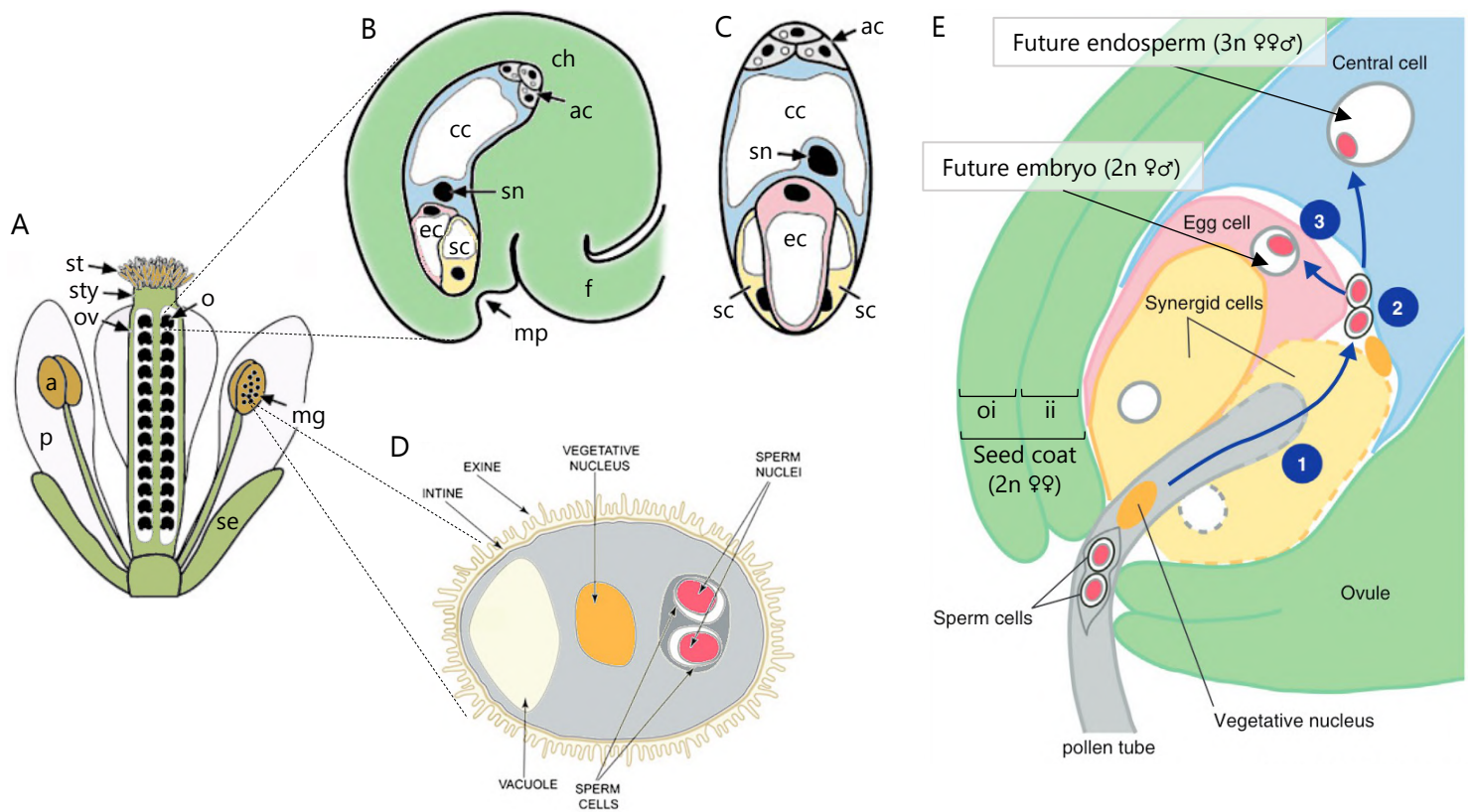


Figure Intro-1: Représentation schématique des gamétophytes mâle et femelle et de la double fécondation chez *Arabidopsis thaliana*. A) Fleur. B) Ovule. C) Gamétophyte femelle ou sac embryonnaire. D) Gamétophyte mâle ou grain de pollen. E) Le tube pollinique pénètre dans le sac embryonnaire par le micropyle et délivre les deux cellules spermatiques dans une des deux synergides (1), les cellules spermatiques restent immobiles (2) avant de fusionner avec les noyaux de la cellule végétative centrale et de la cellule œuf (3).

Abréviations: a, anthère ; ac, antipode ; cc, cellule végétative centrale ; ch, chalaze ; ec, cellule œuf ; f, funicule ; ii, tégument interne ; mg, gamétophyte mâle ; mp, micropyle ; o, ovule; oi, tégument externe ; ov, ovaire ; p, pétale ; se, sépale ; sn, noyaux de la cellule centrale ; st, stigmate ; sty, style.

Modifié d'après Drews and Yadegari, 2002 (A-C), Grobei et al., 2009 (D) et Hamamura et al., 2012 (E).

Introduction

1. Reproduction des plantes

Les plantes sont capables de se reproduire de manière sexuée ou asexuée. La reproduction sexuée, ou zygotique, assure le maintien de la variabilité génétique et la dispersion des individus grâce au pollen et aux graines, tandis que la reproduction asexuée permet la propagation clonale d'individus génétiquement identiques à la plante mère. Ces deux types de reproduction peuvent apporter différents avantages sélectifs mais présentent également de grands intérêts industriels et agronomiques. En effet, les graines représentent une part importante du rendement agricole et constituent un élément essentiel de l'alimentation humaine et animale, tandis que la reproduction clonale est un levier important pour la multiplication de certaines espèces d'intérêt agronomique et forestier.

1.1. Reproduction zygotique

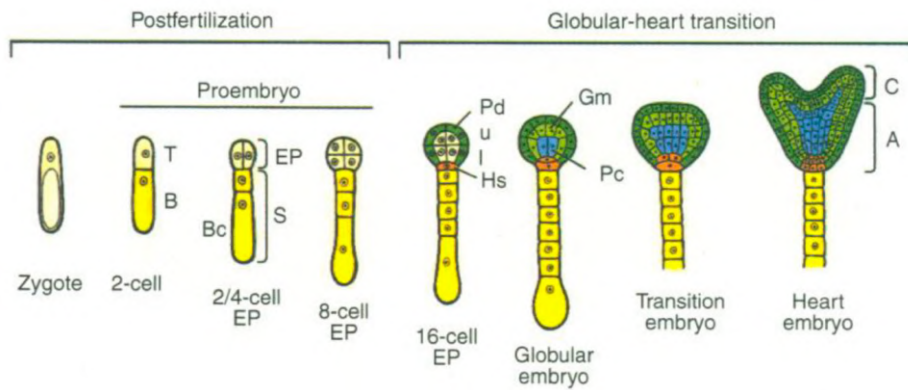
Les graines permettent la propagation des espèces et jouent un rôle très important dans l'adaptation des plantes aux environnements terrestres (Kenrick and Crane, 1997; Han et al., 2017). Les plantes à graines, ou Spermaphytes, sont divisées en deux sous-groupes : les Gymnospermes, ou plantes à ovules nus, qui représentent environ 1 000 espèces végétales vivantes, et dont font partie les pins et les ginkgos par exemple, et les Angiospermes, ou plantes à fleurs, qui représentent la grande majorité (plus de 95%) des 400 000 espèces végétales actuelles connues.

1.1.1. Fécondation et formation de la graine

1.1.1.1. La double-fécondation

Chez les Angiospermes, la double fécondation de la cellule œuf par deux cellules spermatiques donne naissance à un embryon diploïde et à un albumen triploïde se développant dans une enveloppe, formant une graine (Figure Intro-1) (Drews and Yadegari, 2002). Le gamétophyte femelle, ou sac embryonnaire, qui est contenu dans l'ovule, est formé d'un tissu nourricier, le nucelle et de deux enveloppes protectrices : les téguments interne et externe. La cellule mère ($2n$) à l'origine du gamétophyte femelle subit une division méiotique et donne ainsi 4 tétra-spores (n) dont 3 dégénèrent et une, la mégaspore, subit une succession de 3 mitoses sans cytokinèse pour donner le sac embryonnaire (Drews and Koltunow, 2011). Il s'agit ainsi d'une structure polarisée contenant 8 noyaux haploïdes répartis au sein de 7 cellules. Du côté du pôle antérieur, 2 synergides entourent

A



B

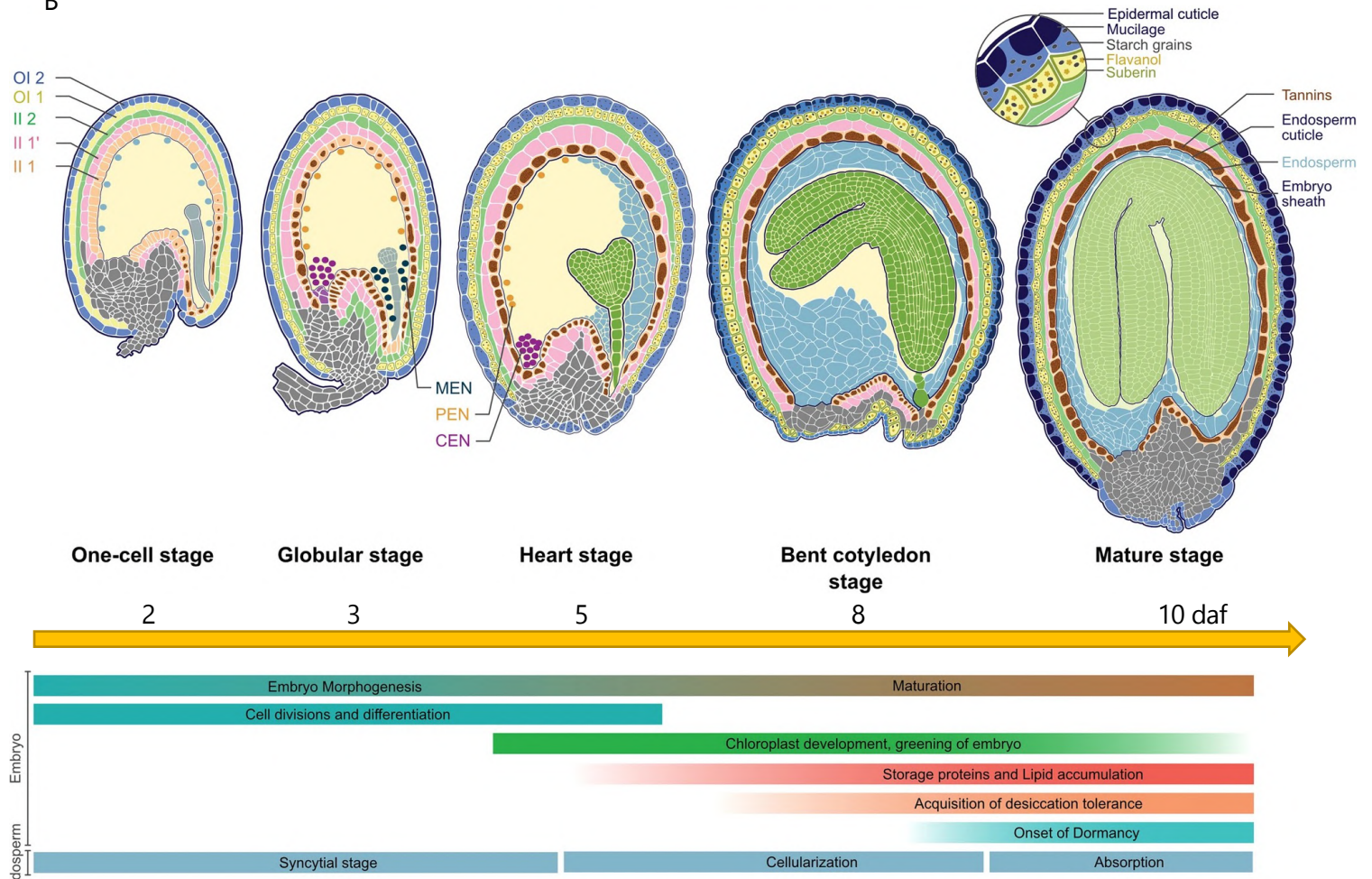


Figure Intro-2 : Représentation schématique du développement de la graine chez *Arabidopsis thaliana*.

A) Embryogenèse précoce : le zygote subit une division asymétrique donnant une cellule apicale et une cellule basale, respectivement à l'origine de l'embryon et du suspenseur. B) Développement de la graine montrant les trois compartiments : l'enveloppe, l'albumen et l'embryon. Les différentes couches de l'enveloppe formées à partir des téguments sont indiquées. L'albumen est représenté en bleu.

Abréviations : A, axe ; B, cellule basale du zygote ; Bc, cellule basale du suspenseur ; C, cotylédons ; CEN, albumen chalazal ; daf, jours après fécondation ; EP, proembryon ; Gm, méristème basal ; Hs, hypophyse ; I, partie supérieure (futurs parties aériennes) ; II 1/1'/2, couches internes des téguments ; MEN, albumen micropylaire ; OI 1/2, couches externes des téguments ; PEN, albumen périphérique ; PC, procambium ; Pd, protoderme ; S, suspenseur ; T, cellule apicale du zygote ; u, partie inférieure (futurs hypocotyle et racine).

Modifié d'après Goldberg et al., 1994 (A) et Verma et al., 2022 (B).

l'oosphère, tandis que le pôle postérieur est formé de 3 antipodes formant la chalaze. Enfin, la cellule végétative centrale contient deux noyaux (Figure Intro-1B-C) (Drews and Yadegari, 2002; Sprunck and Groß-Hardt, 2011). Le grain de pollen mature, quant à lui, est formé d'une cellule végétative et de deux cellules spermatiques, ou gamètes mâles, provenant de la division d'une cellule reproductrice ou génératrice (Figure Intro-1D) (Grobei et al., 2009; Twell, 2011). Au moment de la fécondation, un grain de pollen mature va germer sur le pistil de la fleur et le tube pollinique s'allonger jusqu'à l'entrée du micropyle et libérer les deux cellules spermatiques. L'une d'elles fécondera l'oosphère pour donner l'embryon, ou zygote principal, et l'autre fécondera les deux noyaux de la cellule végétative, donnant l'albumen, ou zygote accessoire (Figure Intro-1E) (Kessler and Grossniklaus, 2011; Hamamura et al., 2012). Ainsi, la graine est une structure bien particulière constituée de trois éléments d'origine et de génotype différents : l'embryon ($2n$) et l'albumen ($3n$) d'origine maternelle et paternelle (issus de la fécondation) et l'enveloppe, formée par les téguments ($2n$) d'origine maternelle (Ingouff et al., 2007). Il est à noter que, bien que les téguments soient d'origine maternelle (ne proviennent pas du processus de fécondation), leur développement est induit par la fécondation et accompagne la croissance de l'embryon et de l'albumen (Ingram, 2010; Figueiredo and Köhler, 2014).

À la suite de la double fécondation, la graine va donc pouvoir se développer suivant deux phases principales : l'embryogenèse, qui suit la fécondation, avec la formation de l'embryon et de l'albumen puis, la maturation, pendant laquelle la graine se différencie et accumule ses réserves (Goldberg et al., 1994). Chez *Arabidopsis*, ces deux phases durent entre 6-8 jours et environ 2 semaines, respectivement.

1.1.1.2. Formation de l'albumen

L'albumen, qui est donc formé à partir de la fusion des noyaux de la cellule centrale, binucléée et d'une cellule spermatique, va se développer suivant deux grandes phases. Chez la plupart des espèces, il se développe selon une succession de divisions mitotiques sans cytokinèse, c'est-à-dire sans formation de parois cellulaires (Costa et al., 2004). Le sac embryonnaire est alors composé de noyaux espacés régulièrement, qui vont entourer l'embryon jusqu'au stade globulaire (4 jours après fécondation) : on parle d'albumen syncytial. Puis, l'albumen se cellularise au fur et à mesure du développement des cotylédons de l'embryon (Figures Intro-2B et 4) (Hehenberger et al., 2012). Cette cellularisation se fait de manière concentrique en partant du micropyle jusqu'à la chalaze. Enfin, dans le cas des graines exalbuminées, dans lesquelles l'albumen tient très peu d'espace, il se réduit progressivement pour ne former qu'une couche cellulaire en laissant la place à l'embryon qui forme la majeure partie de la graine (Figure Intro-3A). L'albumen apporte à l'embryon des

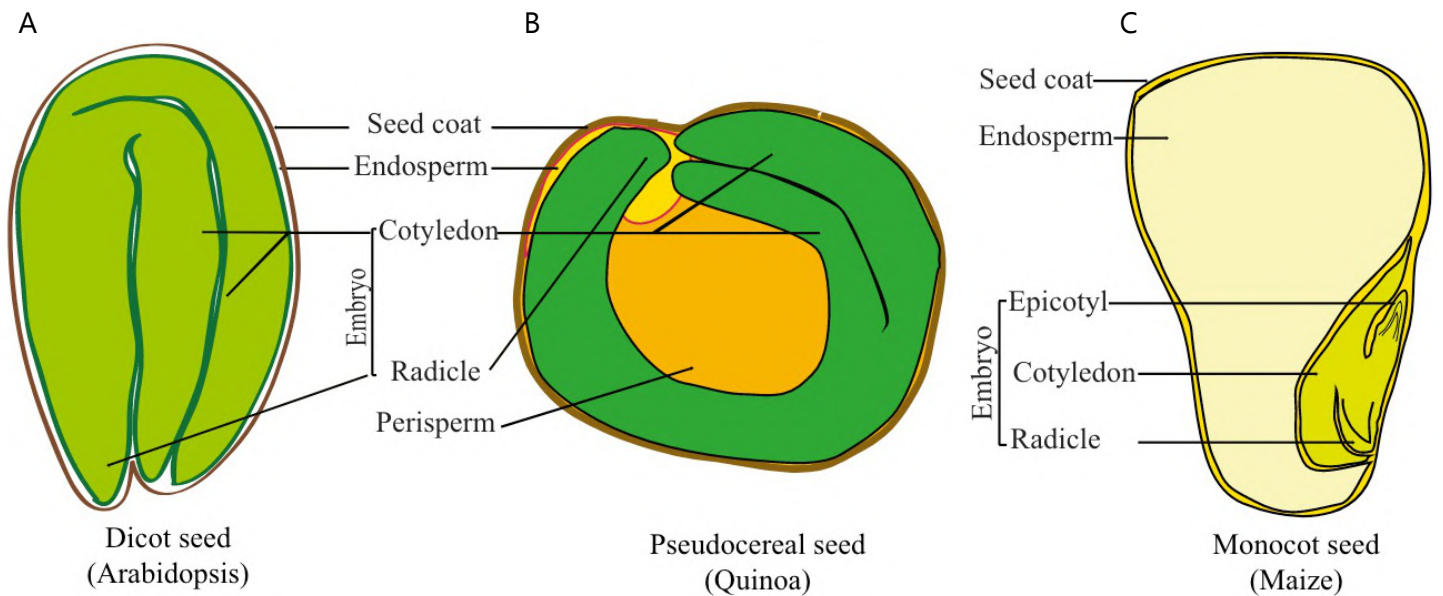


Figure Intro-3 : Représentation schématique de différents types de graines matures. A) Chez les graines exalbuminées, comme *Arabidopsis thaliana*, les deux cotylédons sont les principaux organes de stockage des réserves et tiennent la plus grande place dans la graine. L'albumen est réduit à une seule couche cellulaire (l'aleurone) et entoure l'embryon. B) Chez les graines à périsperme, comme le quinoa, la plus grande partie des réserves est accumulée dans le périsperme, un tissu d'origine maternelle provenant du nucelle. C) Chez les graines albuminées, comme le maïs, la graine mature contient un petit embryon entouré d'un albumen persistant, dans lequel sont stockés les composés de réserve. La couche à aleurones est la plus externe de l'albumen.

D'après Fatihi et al., 2016.

nutriments nécessaires à sa croissance provenant de la plante mère et joue également un rôle essentiel dans la signalisation avec l'embryon et les téguments pour coordonner la croissance de la graine (Baud et al., 2005; Berger et al., 2006; Ingram, 2010; Doll and Ingram, 2022). Chez les espèces albuminées, comme les céréales ou les Solénacées, l'albumen est persistant et constitue une part importante du volume de la graine. Il sert alors également de tissu nourricier lors de la germination et apporte l'énergie et les nutriments nécessaires à la plantule avant la mise en place de la photosynthèse (Figure Intro-3) (Yan et al., 2014).

1.1.1.3. Formation de l'embryon

L'embryogenèse zygotique permet la mise en place de l'architecture de la future plante, avec l'ébauche des futurs organes et des méristèmes apicaux caulinaire et racinaire, reliés par un axe embryonnaire. L'embryon se développe d'après un patron bien défini par une succession de divisions cellulaires coordonnées permettant la mise en place d'une polarité apico-basale et d'un axe de symétrie radiale (Zhang and Laux, 2011). Après la fécondation, le zygote s'allonge puis se divise en deux cellules asymétriques dont l'une, basale, est allongée, contrairement à la deuxième, apicale, qui est plus petite (Figure Intro-2A). La cellule apicale va alors suivre une série de divisions orientées, d'abord longitudinales, puis transverses, jusqu'à former une structure globulaire, entourée d'un protoderme. La partie supérieure de cette sphère sera à l'origine des futures parties aériennes de la plante, tandis que la partie inférieure donnera l'hypocotyle et la racine. La cellule basale, quant à elle, va uniquement subir des divisions transversales et former le suspenseur, qui relie l'embryon aux tissus maternels via le micropyle. La cellule la plus haute du suspenseur, l'hypophyse, produira, en se divisant, le futur méristème apical racinaire et le centre quiescent (4 jours après fécondation). Au stade cœur, l'embryon acquiert une symétrie bilatérale avec la croissance des cotylédons (5-6 jours après fécondation). Les ébauches des futurs organes sont visibles, avec l'hypocotyle, la radicule, l'épiderme et le cortex, ainsi que des tissus provasculaires. L'embryon poursuit sa croissance, les cotylédons s'allongent, le méristème apical caulinaire se forme au stade torpille (7 jours après fécondation), puis les cotylédons se retournent (8 jours après fécondation). L'albumen se réduit alors progressivement pour laisser place à l'embryon mature, qui occupe tout l'espace disponible dans la graine (10 jours après fécondation) (Baud et al., 2002).

1.1.2. Maturation de la graine

Comme décrit ci-dessus, après 10 jours de développement, l'embryon a atteint sa taille mature (fin des divisions et phase de différenciation uniquement) et la graine commence à accumuler des composés de réserve, pendant la phase dite de maturation précoce (Figures Intro-2B et 4) (Baud et

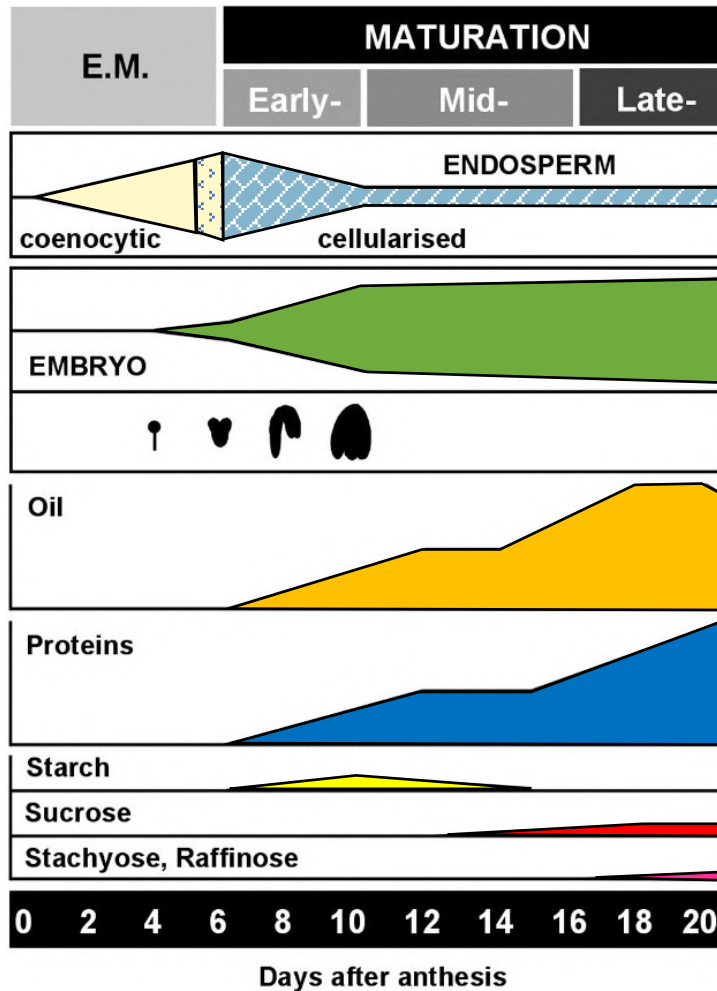


Figure Intro-4 : Représentation schématique du développement de la graine chez *Arabidopsis thaliana*. Le volume relatif de l'albumen et de l'embryon dans la graine est représenté au cours du développement. L'évolution de la teneur en huile, protéines, amidon et sucres au cours de la maturation est également indiquée.

Abréviations : E.M., morphogenèse embryonnaire

Modifié d'après Baud et al., 2008.

al., 2008). Jusqu'alors, les nutriments nécessaires au développement de la graine étaient essentiellement contenus dans l'albumen (parfois dans la chalaze et les tissus maternels comme le nucelle chez certaines espèces) qui ne représente plus, à ce stade, chez *Arabidopsis thaliana*, qu'une couche cellulaire. Chez cette espèce oléoprotéagineuse, la graine synthétise principalement des composés de stockage azotés (protéines) et carbonés (lipides), stockés dans des corps protéiques et lipidiques respectivement, des structures intracellulaires contenues principalement dans les cotylédons. Avant même le début de la phase de maturation, tandis que l'embryogenèse est toujours en cours, la graine accumule transitoirement de l'amidon, stocké dans les téguments externes et dans l'embryon, principalement au niveau de la radicule (Focks and Benning, 1998; Baud et al., 2002). La graine synthétise également des hexoses au cours de la maturation, qui sont dégradés par la suite. En fin de maturation, du sucrose et des oligosaccharides sont produits par la graine et qui persistent jusque dans la graine mature. Tout au long de la maturation, la graine synthétise des acides gras, dont l'accumulation suit le même profil que l'augmentation de la masse de la graine et dont la majeure partie (93% de la graine sèche) est formée par des triacylglycérols (TAG) (Baud et al., 2002). Enfin, les graines synthétisent des protéines de réserve dont la majorité, chez *Arabidopsis thaliana*, sont des albumines 2S (ou arabines) et les globulines 12S (ou cruciférines) (Heath et al., 1986). De la même manière que pour les acides gras, la quantité de protéines accumulées dans la graine augmente tout au long du développement (Baud et al., 2002). La dernière étape du développement de la graine est la maturation tardive, ou dessiccation. La graine mature perd son eau, qui représente alors, en fonction de l'espèce, 5 à 20% de la graine sèche et entre en quiescence métabolique, *i.e.* un arrêt du développement, en attendant que les conditions soient favorables à la germination. La survie de la graine à l'état sec est possible grâce au cytoplasme qui passe d'un état liquide à vitreux. La viscosité induit une faible mobilité moléculaire et assure la stabilité biochimique des graines. Elle est ainsi corrélée à la viabilité et la longévité de la graine sèche (Buitink and Leprince, 2008). Les graines dites orthodoxes sont capables de survivre de nombreuses années dans des conditions de stress comme de fortes variations de températures, grâce notamment à leur état sec, qui assure leur longévité (Sano et al., 2016). Cependant, certaines graines sont dites récalcitrantes, elles ont alors un phénotype vivipare et peuvent germer directement sur la plante mère après la maturation sans passer par un état sec et de quiescence métabolique. C'est le cas par exemple du palétuvier ou de certaines espèces comme le cacaotier ou le caféier (Farnsworth, 2000).

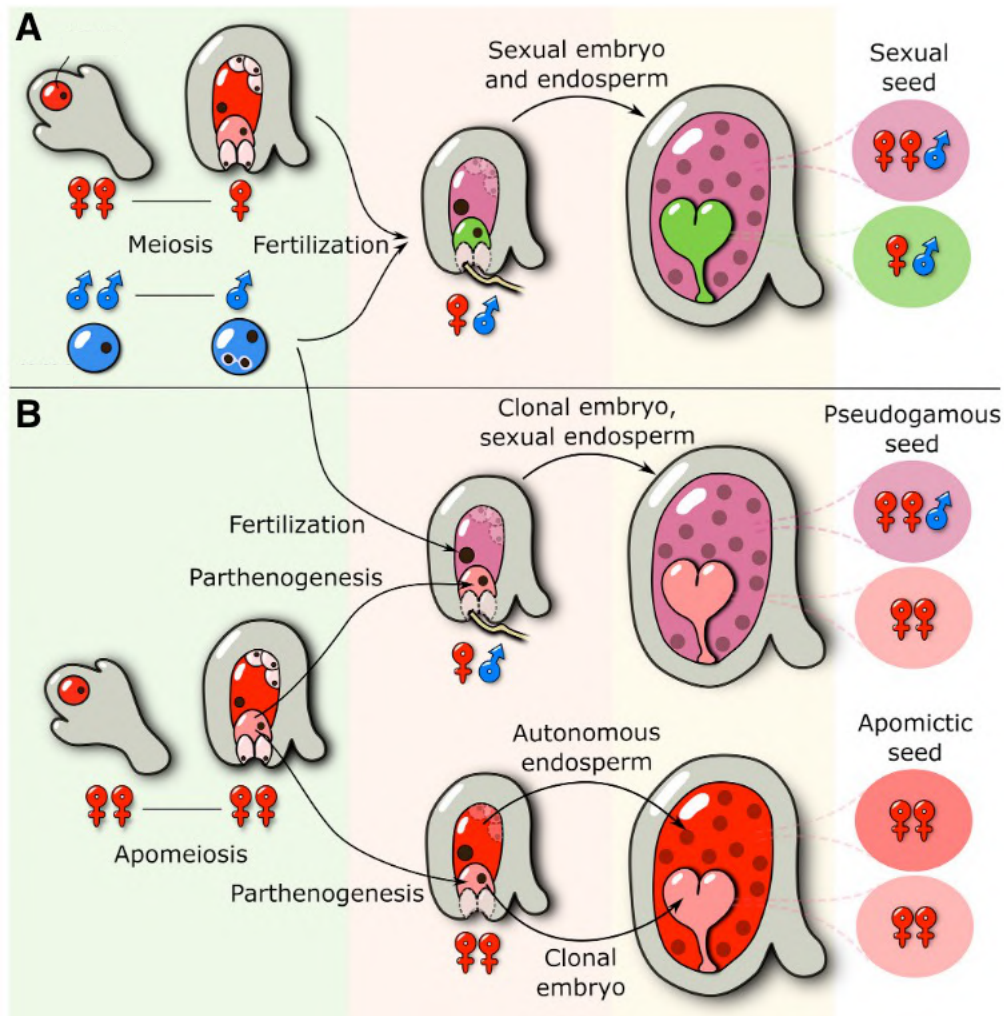


Figure Intro-5 : Initiation de la formation de la graine : comparaison entre la reproduction sexuée et asexuée (apomixie). A) La double fécondation des noyaux haploïdes de la cellule œuf et de la cellule végétative issus d'une méiose conduisent à la formation d'un embryon $2n$ et d'un albumen $3n$. B) Les cellules de la mégaspore ayant subi une apoméiose sont diploïdes : la cellule œuf ($2n$) donne un embryon clonal ($2n$) par parthénogénèse. Les graines dites pseudogames nécessitent la fécondation de la cellule végétative pour former l'albumen, tandis que les graines apomictiques peuvent former un albumen autonome sans fécondation.

Modifié d'après Figueiredo and Köhler, 2018.

1.2. Reproduction clonale

Si l'abondance de plantes à graines suggère que la reproduction zygotique apporte un avantage sélectif certain, les plantes n'ont cependant pas perdu leur capacité à se reproduire de manière asexuée. La formation d'une graine enveloppant un embryon n'est pas spécifique de la reproduction sexuée. En effet, certaines espèces d'Angiospermes, dites apomictiques, sont capables de former une graine et un embryon sans fécondation. C'est le cas du pissenlit, du mil, de l'aubépine ou encore de la ronce, de quelques espèces de pommiers, ou de certains agrumes (Hand and Koltunow, 2014). Dans cette situation, la mégaspore ne subit pas de méiose et est donc diploïde et non pas haploïde, on parle alors d'apoméiose ou améiose (Drews and Koltunow, 2011; Hand and Koltunow, 2014). Le gamétophyte femelle contient dans ce cas une cellule œuf diploïde qui permet la formation d'un embryon sans fécondation (parthénogenèse). La plupart des graines apomictiques sont pseudogames, c'est-à-dire que la formation de l'albumen nécessite tout de même la fécondation de la cellule centrale par une cellule spermatique, mais certaines plantes peuvent initier la formation d'un albumen autonome (Figure Intro-5) (Figueiredo and Köhler, 2018).

Bien que les plantes à graines représentent la majeure partie du règne végétal, d'un point de vue évolutif, l'embryon est apparu avant la graine, ce qui pourrait expliquer pourquoi de nombreux types de cellules végétales, autres que les cellules zygotiques, sont capables d'induire de l'embryogenèse (Radoeva and Weijers, 2014). Cette propriété spécifique aux cellules végétales est mise à contribution dans les processus d'embryogenèse somatique (ES). Cette méthode, utilisée pour la propagation clonale, repose sur la totipotence des cellules végétales, *i.e.*, leur capacité à se différencier et se redifférencier en n'importe quel type cellulaire (voir Introduction – partie 4. [Revue : « Contrôle génétique et moléculaire de l'embryogenèse somatique »](#)). L'embryogenèse somatique permet, à partir d'un seul explant, de produire plusieurs individus génétiquement identiques constituant des clones, sous forme d'embryons somatiques, susceptibles de se développer pour donner une plante mature. L'ES est utilisée comme outil de propagation pour certaines espèces d'intérêt comme le caféier, le pin ou encore le cacaotier dont la reproduction sexuée est impossible ou inefficace et dont les programmes de sélection agronomiques reposent sur la multiplication de clones élites (Loyola-Vargas and Ochoa-Alejo, 2016). L'embryogenèse somatique (ou "adventive") est un moyen de reproduction asexuée que l'on trouve chez certaines espèces, ou un processus artificiel qui implique l'utilisation de différentes phytohormones et des conditions de culture *in vitro* particulières. Un exemple naturel bien décrit est celui de *Kalanchoë daigremontiana* qui forme de véritables embryons à la périphérie de ses feuilles, qui se développent en plantules somatiques (Garcês et al., 2007). D'une façon générale, la totipotence de nombreuses

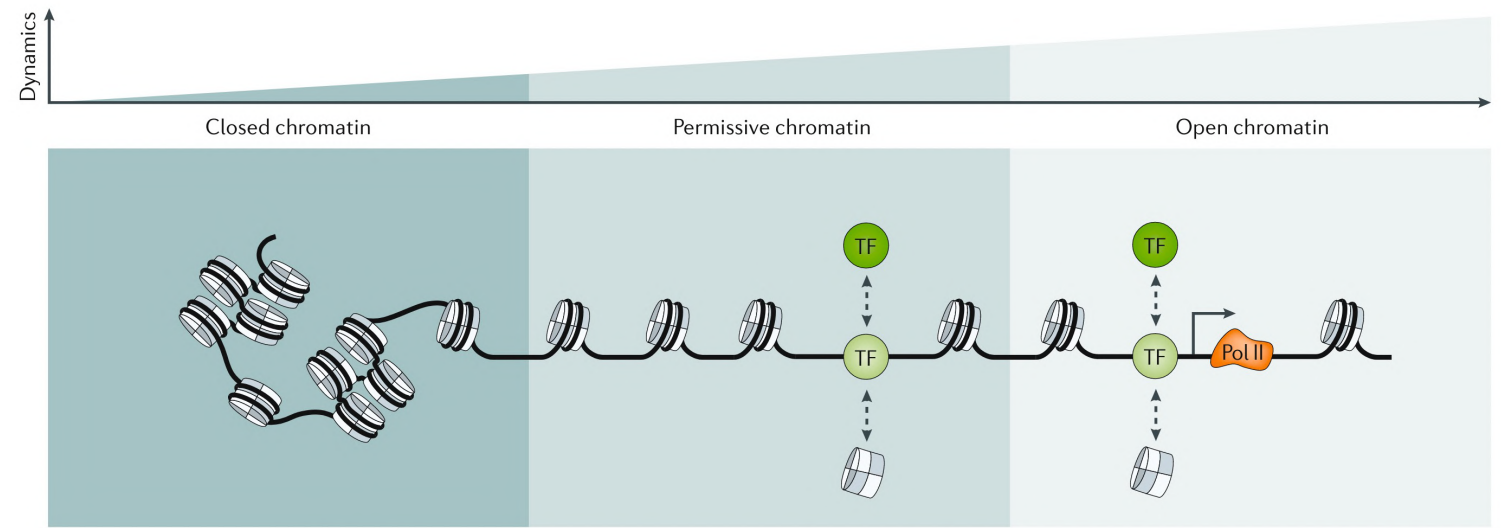


Figure Intro-6 : La distribution des nucléosomes et la dynamique de la chromatine influe sur l'accessibilité de l'ADN pour la fixation des protéines régulatrices comme les facteurs de transcription et la machinerie transcriptionnelle. Lorsque la chromatine est fermée, l'ADN est enroulé autour des nucléosomes et inaccessible aux protéines régulatrices ainsi qu'aux remodeleurs chromatininiens. La chromatine ouverte permet aux facteurs de transcription de se fixer sur leur cible et de rendre l'ADN encore plus accessible afin d'initier la transcription.

Abréviations : Pol II, ARN polymérase II ; TF, facteur de transcription.

D'après Klemm et al., 2019.

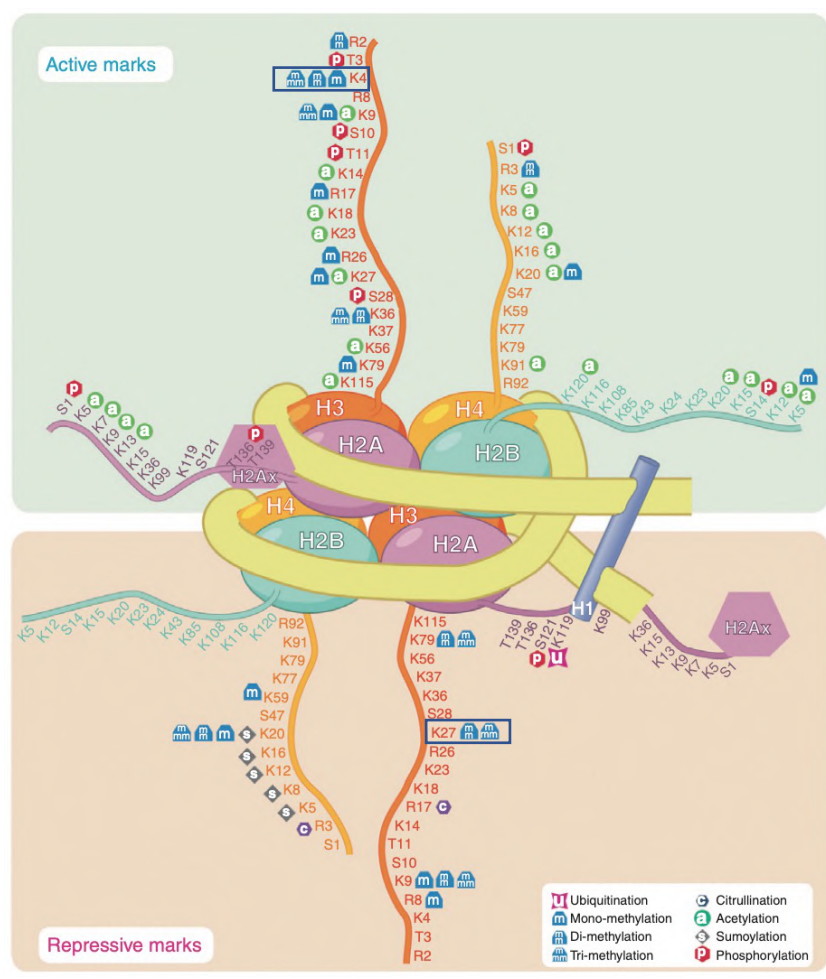


Figure Intro-7 : Représentation schématique d'un nucléosome. L'ADN est entouré autour d'un octamère formé de quatre dimères d'histones H2A, H2B, H3 et H4. L'histone H1, la protéine de liaison, est liée à l'ADN entre les nucléosomes. Les différents acides aminés formant les queues des histones sont représentés ainsi que les types de modifications post-traductionnelles spécifiques de chaque résidu. Les marques activatrices et répressives sont représentées respectivement dans la partie haute et la partie basse de la figure. Acides aminés : lysine (K), arginine (R), serine (S) et thréonine (T).

D'après Sawan and Herceg, 2010.

cellules végétales permet la propagation d'explants par bouturage en régénérant une plante complète. Certaines espèces comme le fraisier sont également capables de se reproduire de façon asexuée grâce à des stolons qui peuvent s'enraciner et donner une nouvelle plante (Chen et al., 2019).

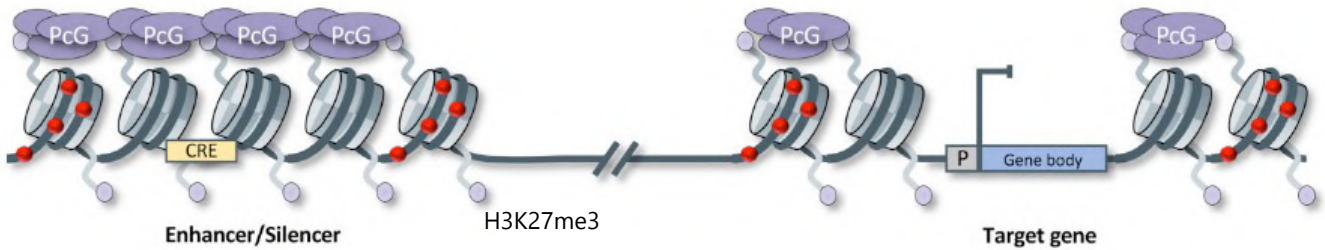
Ces différents moyens de propagation sexuée ou asexuée sont autant d'outils biotechnologiques qui peuvent être mis à profit pour la multiplication d'espèces d'intérêt agronomique. Comme tout processus développemental, ils sont soumis à de très fortes régulations génétiques, épigénétiques ou hormonales et la compréhension de ces régulations est un levier technologique important pour la production de ressources végétales. En effet, dans le cas de la reproduction sexuée, chaque étape du développement de la graine allant de la fécondation jusqu'à la germination est critique pour obtenir une production de graines en quantité et de qualité, en termes de remplissage (qualité et quantité de protéines, de lipides et de sucres) ou de vigueur germinative (homogénéité, rapidité de la germination et bon développement des plantules). Pour les espèces dont la reproduction sexuée est impossible ou inefficace, la reproduction asexuée, quelle que soit la méthode utilisée, est nécessaire à la propagation, mais peut être difficile à mettre en œuvre. Par exemple, certaines espèces sont récalcitrantes aux processus d'embryogenèse somatique, d'autres ne sont pas capables de former des graines apomictiques. L'identification des facteurs pouvant bloquer ces mécanismes est nécessaire si l'on souhaite améliorer la propagation asexuée des espèces végétales.

2. Contrôle de l'expression des gènes

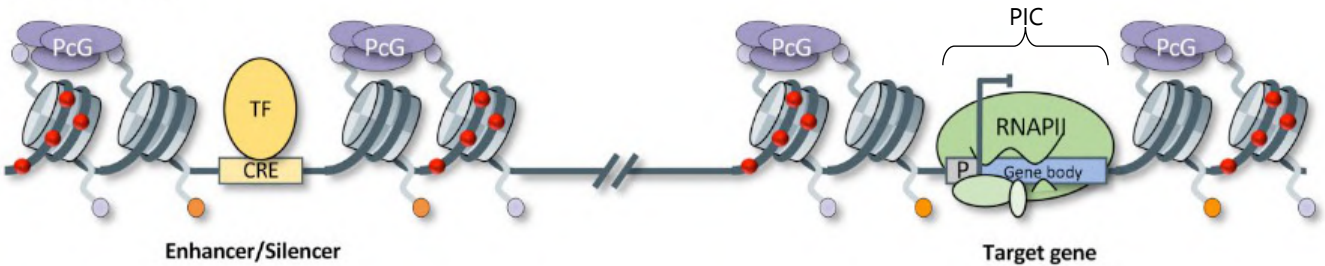
La régulation de l'expression de gènes est indispensable au bon fonctionnement de la cellule, à sa différenciation et la formation de tissus ou à l'adaptation du métabolisme et du développement de la plante aux fluctuations de l'environnement. Afin de contrôler finement l'expression des gènes, les eucaryotes utilisent de nombreux mécanismes (Carels, 2015) : la structure de la chromatine (Kouzarides, 2007) et la méthylation de l'ADN (Chan et al., 2005), l'initiation de la transcription, l'épissage alternatif (Szakonyi and Duque, 2018) ou encore la stabilité des ARN messagers (Sorenson et al., 2018). Par la suite, l'accumulation des protéines est régulée par différents moyens comme le contrôle de la traduction, les modifications post-traductionnelles, le trafic intracellulaire et la dégradation des protéines (Mazzucotelli et al., 2008). Je me concentrerai ici sur les régulations épigénétiques et transcriptionnelles de l'expression des gènes.

Le profil spatio-temporel d'expression des gènes au cours du développement est finement régulé par des éléments *cis*-régulateurs de l'ADN, comprenant des promoteurs, des *enhancers* ou des *silencers* et leurs interactions avec des facteurs *trans*-régulateurs comme les facteurs de

A Repressed CRMs



B Poised CRMs



C Active enhancer



D Active silencer

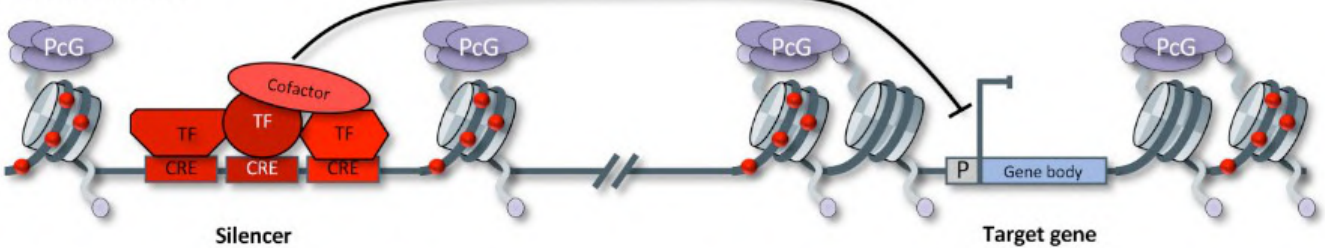


Figure Intro-8 : Représentation schématique de la dynamique de la chromatine et de l'activité des motifs *cis*-régulateurs pour la régulation de la transcription des gènes. A) Le motif *cis*-régulateur est réprimé, le gène est transcriptionnellement inactif : les éléments *cis*-régulateurs sont bloqués par les nucléosomes, la méthylation de l'ADN et les groupes polycomb maintiennent des marques répressives H3K27me3. B) L'élément *cis*-régulateur est accessible pour la liaison par les facteurs de transcription, de même que le promoteur, qui est accessible par la machinerie basale de la transcription, mais les histones environnantes sont faiblement acétylées et des marques répressives H3K27me3 sont toujours maintenues par les groupes Polycomb, ce qui entrave l'élongation par l'ARN polymérase II, le gène n'est donc pas transcrit. C) L'*enhancer* est actif et accessible pour la liaison par de nombreux facteurs de transcription et cofacteurs, les histones sont acétylées et l'ADN déméthylé, la machinerie basale est fixée sur le promoteur et la transcription est active. D) Le *silencer* est actif, les facteurs de transcription et les cofacteurs se lient aux éléments *cis*-régulateurs et recrutent les groupes Polycomb pour catalyser l'induction de marques répressives H3K27me3 et le *silencing* du gène.

Abréviations : CRE, élément *cis*-régulateur ; CRM, motif *cis*-régulateur ; P, promoteur ; PcG, complexe Polycomb ; PIC, complexe de pré-initiation ; RNAPII, ARN polymérase II ; TF, facteur de transcription.

Légende : cercles violets, marques répressives (H3K27me3) ; cercles oranges, marques activatrices (acétylation des histones) ; cercles rouges : méthylation de l'ADN.

D'après Schmitz et al., 2022.

transcription spécifiques, les remodeleurs de la chromatine ou encore la machinerie transcriptionnelle avec l'ARN polymérase II (Hajheidari and Huang, 2022).

2.1. Les régulations chromatinienne : contrôle de l'accessibilité des régions régulatrices de l'expression des gènes

Le premier levier de contrôle de l'expression d'un gène est la structure de la chromatine qui l'entoure et qui peut être plus ou moins compactée, rendant l'ADN accessible, ou non, à la machinerie transcriptionnelle (Figure Intro-6). L'hétérochromatine définit un état dans lequel l'ADN est condensé et inaccessible, empêchant l'expression des gènes, à l'inverse de l'euchromatine, dans laquelle l'ADN est décondensé, accessible et où la transcription est possible (Klemm et al., 2019). L'ADN est rendu accessible ou non, d'une part, par la méthylation des cytosines, qui semble pouvoir protéger le génome contre l'insertion des rétrotransposons et moduler l'expression de familles de gènes répétés, pour revue (Chan et al., 2005) ; et, d'autre part, via l'organisation des nucléosomes, des sous-unités répétées de chromatine pour revue (Olins and Olins, 2003). Chez les plantes, la méthylation est ciblée sur des séquences CG, CNG (où N peut être n'importe quel nucléotide) et CHH (où H peut être un A, C ou T) et est induite par des ADN-méthyltransférases. Lorsque la méthylation est exercée sur un promoteur, elle a pour effet d'inhiber la transcription mais, au niveau de séquences codantes, elle n'affecte généralement pas l'expression du gène. Les nucléosomes, quant à eux, sont des octamères formés par quatre paires d'histones : H2A, H2B, H3 et H4. L'ADN (environ 147 pb, en fonction des espèces) est enroulé autour d'un octamère d'histones. La dynamique des nucléosomes, c'est-à-dire leur positionnement, leur assemblage et désassemblage est un premier facteur influant l'accessibilité de l'ADN à la machinerie transcriptionnelle. Les gènes exprimés constitutivement sont généralement situés dans des régions sans nucléosomes (NFR, *Nucleosome-Free Region*), tandis que les gènes dont l'expression est contrôlée, par exemple, seulement en conditions de stress, sont présents dans des régions dans lesquelles la présence de nucléosomes est régulée (NDR, *Nucleosome-Depleted Region*) (Lai and Pugh, 2017). Les nucléosomes sont également soumis à des régulations post-traductionnelles sur la queue polypeptidique N-terminale des histones, induisant des variations dans leur séquence (Figure Intro-7). Les différents variants d'histones peuvent influencer positivement ou négativement la compaction de la chromatine, et, ainsi, l'accessibilité de l'ADN (Figure Intro-8) (Lai and Pugh, 2017). De manière générale, l'acétylation des lysines par les histones acétyltransférases (HAT) permet la décondensation de l'ADN autour des nucléosomes, et induit donc un état de l'ADN plus accessible à la machinerie transcriptionnelle. La méthylation des lysines par les histones méthyltransférases

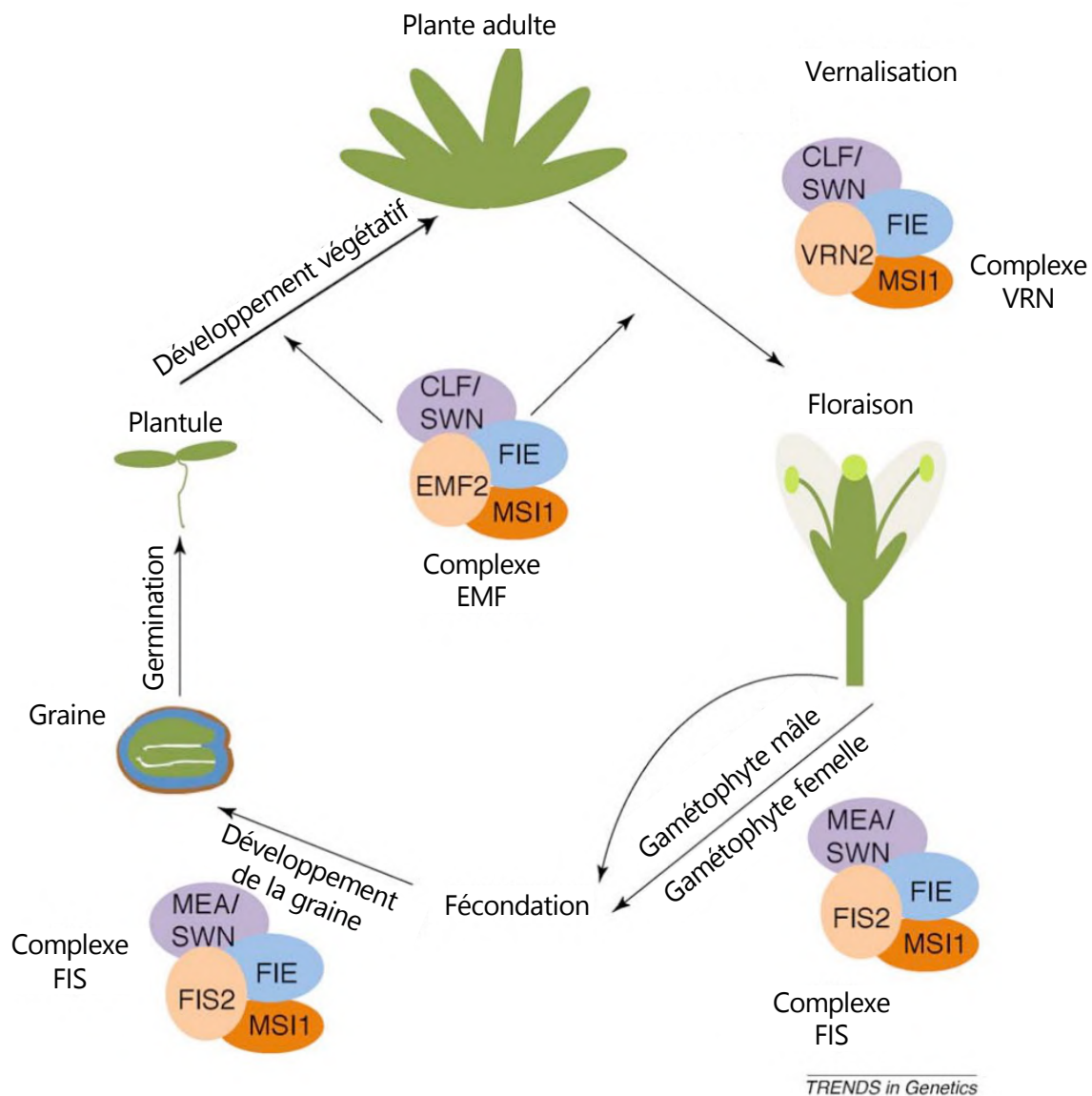


Figure Intro-9 : Le Complexe Répressif Polycomb 2 (PRC2) régule divers mécanismes tout au long développement de la plante et peut prendre trois formes différentes en fonction des protéines qui le forment.

- Le complexe EMF, formé par CLF/SWN, EMF2, FIE et MSI1 promeut le développement végétatif de la plante, retarde la reproduction et maintient les cellules dans un état différencié.
- Le complexe VRN, formé par CLF/SWN, VRN2, FIE et MSI1 induit le *silencing* au locus FLC et permet ainsi la floraison après le processus de vernalisation.
- Le complexe FIS, formé par MEA/SWN, FIE, FIS2 et MSI1 empêche la formation et le développement de la graine en absence de fécondation et est nécessaire au bon développement de la graine.

Modifié d'après Hennig and Derkacheva, 2009.

peut avoir différents effets. Par exemple, la triméthylation de la lysine 27 de l'histone H3 (H3K27me3) est une marque chromatinienne répressive, tandis que la triméthylation de la lysine 4 de l'histone H3 (H3K4me3) est une marque de décondensation de l'ADN (Sawan and Herceg, 2010; Schmitz et al., 2022). Les groupes Polycomb (PcG) et Trithorax (TrxG), sont des complexes protéiques conservés chez les eucaryotes et jouant des rôles antagonistes dans la régulation épigénétique. Ce sont des acteurs majeurs de l'induction de marques épigénétiques stables, héréditaires à travers les divisions cellulaires, mais également dynamiques, en réponse à des signaux développementaux et environnementaux, voir (Mozgova and Hennig, 2015) pour revue. Le Complexe Répressif Polycomb 2 (PRC2) a une activité histone méthyltransférase et est capable d'induire des marques H3K27me3, tandis que PRC1 induit l'ubiquitination de l'histone H2A (H2Aub) (Figure Intro-8). Le dogme traditionnel propose que la sous-unité Pc de PRC1 se lie aux marques H3K27me3 déposées par PRC2 pour ubiquitiner H2A (Wang et al., 2004), mais chez les plantes, H2Aub pourrait être indépendante de H3K27me3. Il semblerait, en effet, que PRC1 et PRC2 soient recrutés indépendamment et établissent des sites de liaison l'un pour l'autre induisant ainsi de manière complémentaire, une plus forte répression au niveau de leurs cibles (Yang et al., 2013; Mozgova and Hennig, 2015). Chez *Arabidopsis thaliana*, trois différents complexes PRC2 ont été identifiés, les protéines formant le complexe pouvant différer en fonction du stade de développement, tout au long du cycle de vie de la plante. Dans les trois types de complexe PRC2, l'activité méthyltransférase est conférée par CURLY LEAF (CLF), SWINGER (SWN) ou MEDEA (MEA). La liaison aux marques H3K27me3 est assurée par FERTILIZATION-INDEPENDANT ENDOSPERM (FIE) et la liaison aux nucléosomes, par MULTICOPY SUPPRESSOR OF IRA 1 (MSI1). Enfin, la spécificité de chaque complexe peut être assurée par trois différentes protéines : FERTILIZATION INDEPENDANT SEED 2 (FIS2), EMBRYONIC FLOWER 2 (EMF2) et VERNALIZATION 2 (VRN2), dont le rôle est d'apporter la stabilité du complexe. C'est ainsi que le complexe PRC2-FIS est impliqué dans le développement de la graine et sa répression en absence de fécondation, tandis que PRC2-EMF promeut le développement végétatif, retarde la reproduction et maintient les cellules dans un état différencié et que PRC2-VRN réprime l'expression de *FLC* (*FLOWERING LOCUS C*) pour induire la floraison après vernalisation (Figure Intro-9) (Hennig and Derkacheva, 2009). Certaines protéines sont présentes dans les différents complexes PRC2 (*e.g.* FIE et MEA), et la spécificité d'action est apportée par les membres FIS2, EMF2 et VRN2, dont la fonction n'est pas redondante. Par exemple, EMF2 et VRN2 ne peuvent pas se substituer à FIS2 dans la répression de la formation de l'albumen (Roszak and Köhler, 2011). Le complexe PRC1 peut également prendre plusieurs formes et réguler différents aspects du développement en lien avec les sous-unités de PRC2. Les sous-unités des

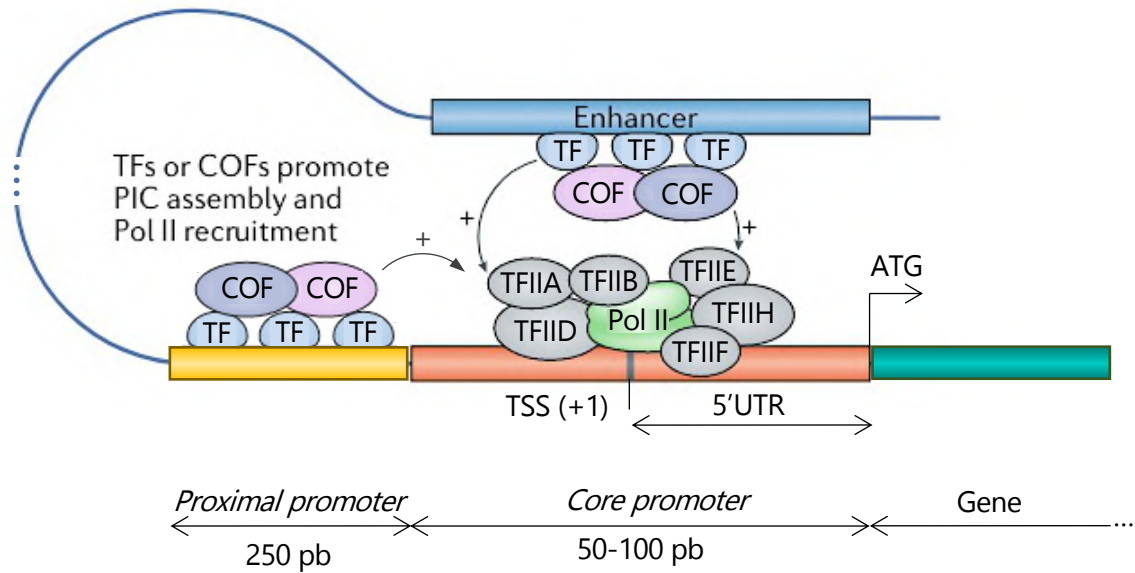


Figure Intro-10 : Représentation schématique de la structure d'un promoteur eucaryote au moment de la formation du complexe de pré-initiation (PIC). Le *core promoter* contient des boîtes de fixation reconnues par la machinerie basale d'initiation de la transcription, *i.e.* les facteurs de transcription généraux en complexe avec l'ARN polymérase II. Le *core promoter* est suffisant pour une activation directe de la transcription mais avec une activité très faible qui peut être supprimée par la chromatine ou activée par les *enhancers*. Le promoteur proximal et les *enhancers* contiennent les sites de fixation des facteurs de transcription spécifiques qui permettent d'augmenter le niveau de transcription indépendamment de la distance qui les relie au promoteur ou de leur orientation. Les facteurs de transcription se lient grâce à leur domaine de liaison à l'ADN et régulent le niveau de transcription en recrutant la machinerie basale de transcription ou des cofacteurs de transcription grâce à leur domaine d'activation de la transcription. Les cofacteurs n'ont pas de domaine de liaison et sont recrutés par les facteurs de transcription spécifiques pour réguler la transcription.

Abréviations : COF, cofacteur de transcription ; PolII, ARN polymérase II ; TF, facteur de transcription ; TSS, site d'initiation de la transcription ; TFIIA/B/D/E/F/H, facteurs de transcription généraux.

Modifié d'après Haberle and Stark 2018.

Complexes Répressifs Polycomb (PRC) ont une interaction très stable avec l'ADN pendant la réplication, ce qui permet de maintenir la répression au cours des divisions cellulaires et assure l'héritabilité de l'état chromatinien, bien que les marques H3K27me3 elles-mêmes ne soient pas maintenues (Francis et al., 2009; Petruk et al., 2012). Les PRC peuvent interagir avec des facteurs de transcription à travers l'une de leurs sous-unités afin de cibler spécifiquement un locus tandis que la compétition de différents facteurs de transcription à un même site peut moduler leur activité (Mozgova and Hennig, 2015; Tao et al., 2019). Les complexes TrxG quant à eux, sont des antagonistes de la répression induite par les PRC grâce à leur activité méthyltransférase sur H3K4me3 ou acétyltransférase sur H3K27ac (Schuettengruber et al., 2011). Neuf états possibles de la chromatine peuvent ainsi être définis, en fonction de la séquence d'ADN, de la méthylation des GC, de l'acétylation ou la méthylation des histones, ou des types de variants d'histones et sont représentatifs de l'accessibilité de l'ADN et de l'expression des gènes (Sequeira-Mendes et al., 2014).

2.2. Les promoteurs, lieu de fixation de la machinerie basale de transcription

2.2.1. Structure

La régulation de la transcription et l'initiation de la transcription des gènes sont assurées, entre autres, via leur séquence promotrice. Un promoteur est constitué d'une partie centrale, appelée *core promoter* et d'une partie proximale (Figure Intro-10).

2.2.1.1. Le *core promoter* et la machinerie basale de transcription

Le *core promoteur* entoure le site d'initiation de la transcription (TSS, *Transcription Starting Site*) et comprend les 50-100 pb nécessaires à la reconnaissance et la liaison de la machinerie basale de transcription et d'initiation de la transcription, *i.e.* l'ARN polymérase II et les facteurs de transcription généraux (Haberle and Stark, 2018). Ces derniers sont au nombre de 6 et forment séquentiellement le complexe de pré-initiation de la transcription (PIC). TFIID est un complexe protéique comprenant la *TATA-Binding Protein* (TBP) ainsi que 13 facteurs associés à la TBP (TAF) et permet la reconnaissance du *core promoteur* (Louder et al., 2016). TFIIA et TFIIIB, assurent la liaison au *core promoteur*, TFIIIF forme un complexe avec l'ARN polymérase II, qui est recruté sur le *core promoteur*, enfin, TFIIIE puis TFIIH se lient au *core promoteur* (Sainsbury et al., 2015). Le *core promoteur* est lui-même caractérisé par la présence de plusieurs boîtes de fixation de la machinerie basale de transcription, dont la séquence et la position relative au TSS sont très conservées chez les eucaryotes. Quelques exemples de boîtes sont : la *TATA box*, reconnue par la *TATA-Binding Protein* (TBP) (Joshi, 1987; Yella and Bansal, 2017), *Inr* (*Initiator*), reconnue par TAF1 (Smale and Baltimore,

(a) AP2/ERF

(1) 136-146, (2) three-stranded antiparallel β -sheet packed with α -helix; binding to DNA major groove by β -sheet plane, (3) A/GCCGAC, AGCCGCC, (4) abiotic stress response, ethylene response, (5) 1gcc, 2gcc, 3gcc



(b) B3

(1) 118, (2) seven-stranded open β -barrel, (3) TGTCTC, CATGCA, (4) auxin response, abscisic acid response, (5) 1wid, 1yel



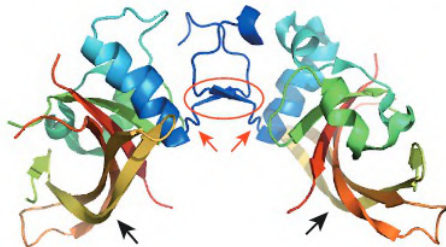
(c) WRKY

(1) 72, (2) four/five-stranded antiparallel β -sheet with a Zn-binding site; binding to DNA major groove by β -sheet rim (β -wedge), (3) TTGAC(C/T), (4) biotic and abiotic stress response, (5) 1wj2, 2ayd, 2lex



(d) NAC

(1) 94-106, (2) twisted six-stranded β -sheet packed with α -helices; dimer with rotational symmetry, (3) AN₅TCN₅ACACGCATGT, (4) biotic and abiotic stress response, (5) 1ut4, 1ut7, 3ulx, 4dul



(e) SBP

(1) 16-17, (2) two Zn-binding sites with long flexible basic loop, (3) CCGTAC(A/G), (4) flower development, Cu uptake, (5) 1ul4, 1ul5, 1wj0

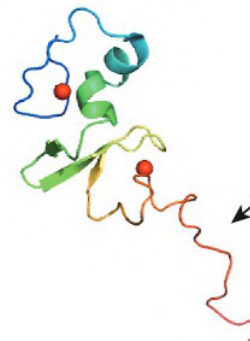


Figure Intro-11 : Structure des domaines de liaison à l'ADN des principales familles de facteurs de transcription spécifiques des plantes. (1) Nombre de facteurs prédits chez *Arabidopsis thaliana*, (2) description de la structure, (3) séquence reconnue, (4) fonctions majeures connues, (5) code associé au facteur de transcription dans la *Protein Data Bank (PDB)*, le code souligné étant celui utilisé pour la figure.

D'après Yamasaki et al., 2013.

1989; Zhu et al., 1995), *BRE* (*TFIIB Recognition Element*), fixée par le facteur de transcription général TFIIB (Deng and Roberts, 2005), *DPE* (*Downstream Promoter Element*), fixée par TAF6 (Burke et al., 1998), la *CCAAT box*, reconnue par les facteurs de transcription NF-Y (Dolfini et al., 2012) ou encore *TCT* (*Polypyrimidine Initiator*) (Bernard et al., 2010).

2.2.1.2. Le promoteur proximal et les facteurs de transcription spécifiques

Le *proximal promoter*, est situé en amont du *core promoter*. Il comprend les sites de liaison des facteurs de transcription spécifiques, qui vont permettre d'affiner le niveau de transcription du gène, ainsi que les *enhancers* (Heintzman and Ren, 2007). Les *enhancers* sont des régions d'ADN allant généralement d'une centaine à un millier de paires de bases, contenant des sites de fixation de facteurs de transcription et permettant de stimuler, à distance, l'initiation de la transcription médiée par les promoteurs (Kyrchanova and Georgiev, 2021). De manière générale, les gènes contrôlant des processus développementaux présentent des régulations transcriptionnelles complexes et sont contrôlés par plusieurs *enhancers* situés plus ou moins loin du promoteur en termes de nombre de paires de bases, mais qui peuvent se révéler très proches, compte tenu de la structure tertiaire de l'ADN. D'ailleurs, si les nucléosomes sont capables de rendre un promoteur plus ou moins accessible à la machinerie transcriptionnelle, il en va de même avec les *enhancers*, ce qui assure un contrôle supplémentaire de l'expression des gènes (Lai and Pugh, 2017). Afin de préparer le promoteur à l'initiation de la transcription, les facteurs de transcription spécifiques se lient aux séquences promotrices qu'ils reconnaissent au niveau du promoteur proximal ou sur les *enhancers*. Ils recrutent alors des coactivateurs, comme le complexe Mediator (Lewis and Reinberg, 2003; Plaschka et al., 2015), des remodeleurs de chromatine et des modificateurs d'histones, conduisant à une altération de la structure et du positionnement des nucléosomes. Ils réorganisent ainsi la structure de la chromatine sur le promoteur pour permettre la liaison et la stabilisation de la machinerie basale d'initiation de la transcription (ARN pol II et facteurs de transcription généraux), requise pour son bon positionnement et les interactions avec les autres protéines régulatrices, formant le complexe de pré-initialisation (PIC) (Heintzman and Ren, 2007; Haberle and Stark, 2018).

2.3. Les facteurs de transcription, régulation ciblée de l'expression

2.3.1. Les facteurs de transcription spécifiques

Chez les procaryotes et les eucaryotes, la régulation transcriptionnelle des gènes est un niveau de régulation exercé par les facteurs de transcription (FT). Ces protéines sont capables de reconnaître des séquences *cis*-régulatrices spécifiques situées sur les promoteurs et de recruter ou d'exclure la

machinerie transcriptionnelle afin d'activer ou de réprimer l'expression des gènes, pour revue (Slattery et al., 2014; Suter, 2020). Bien que les FT procaryotes reconnaissent et se lient très spécifiquement à leur cible, chez les eucaryotes, ils agissent souvent en complexe (Lai et al., 2018; Suter, 2020). Les facteurs de transcription possèdent généralement un domaine de liaison à l'ADN qui leur permet de reconnaître et de se fixer à un motif particulier situé sur ses cibles transcriptionnelles (Yamasaki et al., 2013; Hajheidari and Huang, 2022). Les activateurs transcriptionnels sont également pourvus d'un ou plusieurs domaines d'activation de la transcription, leur permettant le recrutement de complexes protéiques associés au remodelage de la chromatine ou l'initiation de la transcription (Erkina and Erkine, 2016). Ils peuvent aussi avoir un domaine d'interaction protéine-protéine, qui leur permet d'agir en complexe avec d'autres facteurs ayant ou non un domaine de liaison à l'ADN (Riechmann et al., 2000). La diversité de domaines de liaison à l'ADN définit les différentes familles de facteurs de transcription conservées de manière générale chez les eucaryotes, mais aussi spécifiquement chez les plantes (Figure Intro-11). Les facteurs de transcription peuvent également contenir un domaine de localisation nucléaire, nécessaire à leur dynamique de localisation et donc à leur fonction (Zhang et al., 2022). Ce domaine peut cependant être apporté par un partenaire protéique tiers, un autre facteur de transcription par exemple (Mitsuda and Ohme-Takagi, 2009). Les facteurs de transcription sont des régulateurs déterminants pour l'expression génique, en particulier pour les processus développementaux. Chez *Arabidopsis thaliana*, on estime que 5 à 10% des 25 500 gènes séquencés codent pour des facteurs de transcription, ce qui représente une plus forte proportion que chez certains animaux comme *Drosophila melanogaster* ou *Caenorhabditis elegans*, mais équivalente à celles des humains (Mitsuda and Ohme-Takagi, 2009). Ces données suggèrent que le contrôle transcriptionnel est très important chez les plantes, permettant probablement une plus grande plasticité de l'expression des gènes, afin de compenser leur immobilité face aux variations environnementales (Mitra et al., 2021).

2.3.2. Les régulateurs maîtres

Certains FT sont désignés comme étant des régulateurs maîtres car ils sont responsables de l'initiation et du maintien de l'identité cellulaire, par leur capacité à coordonner l'expression spatio-temporelle de certains gènes, transmise au travers des divisions (Oldfield et al., 2014). L'expression des régulateurs maîtres est généralement limitée dans le temps et dans l'espace à un ou plusieurs types cellulaires dont ils déterminent le devenir (Oestreich and Weinmann, 2012). Par exemple, chez les plantes, WUSCHEL (WUS) est un régulateur maître indispensable à la croissance, responsable du maintien de la niche de cellules souches dans le méristème apical caulinaire, de la différenciation des *primordia* d'organes et de la totipotence des cellules végétales (Jha et al., 2020). D'autres

régulateurs maîtres contrôlent des étapes clés du développement végétatif comme la floraison, à l'instar de SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), qui contrôle la transition entre la croissance végétative et reproductive (Kaufmann and Airoidi, 2018).

2.3.3. Les facteurs pionniers

Les facteurs de transcription peuvent également avoir un rôle de facteurs pionniers, c'est-à-dire qu'ils sont capables de se lier à l'ADN nucléosomal ou à la chromatine fermée, de l'ouvrir localement et de promouvoir le dépôt de marques épigénétiques permissives nécessaires à l'expression de gènes cibles, en permettant le recrutement d'autres facteurs de transcription qui peuvent également participer au maintien de l'ouverture des sites de liaisons à l'ADN pendant une longue période (Lai et al., 2018; Zaret, 2018; Iwafuchi-Doi, 2019; Karagianni et al., 2020). Les facteurs pionniers jouent un rôle essentiel dans les processus développementaux tels que la formation des organes, le développement des tissus ou encore la différenciation cellulaire. Si de nombreux facteurs pionniers ont été identifiés chez les animaux, ils sont moins connus chez les plantes, mais certains ont tout de même été étudiés, comme LEAFY (LFY), APETALA 1 (AP1), SEPALLATA 3 (SEP3) ou encore LEAFY COTYLEDON 1 (LEC1) (Lai et al., 2018). Les trois premiers interviennent dans l'établissement du méristème floral et dans la spécification de la fleur et maintiennent un état ouvert de la chromatine tout en contrebalançant l'effet répressur du complexe PRC2 par élimination des marques H3K27me3 (Pajoro et al., 2014; Sayou et al., 2016). Ils n'ont cependant pas d'homologue chez les animaux, contrairement à LEC1, qui est un facteur de transcription de type NF-YB et qui a été montré comme jouant un rôle dans la vernalisation et dans le développement des graines. LEC1, en effet, promeut le dépôt de marques H3K36me3 et contribue à éliminer les marques H3K27me3 induites par PRC2 dans l'embryon sur le gène *FLC* (voir ci-après, partie [2.3.3.1. Rôle de LEC1 \(NF-YB9\) comme facteur pionnier](#)) (Tao et al., 2017). Le groupe des facteurs de transcription NF-Y est très conservé et se lie, dans un complexe trimérique NF-YA/B/C, à la *CCAAT box*, présente dans environ 30% des promoteurs eucaryotes (Dolfini et al., 2012). Chez les plantes, on dénombre en général une dizaine d'homologues de chacune des sous-unités NF-YA, B ou C (Siefers et al., 2009; Petroni et al., 2013). Chez *Arabidopsis*, la fonction de LEC1 (NF-YB9) peut être partiellement restaurée par LEC1-LIKE (L1L/NF-YB6) et NF-YB10, indiquant une certaine redondance de fonction entre les différents NF-Y (Kwong et al., 2003). La capacité de fixation à l'ADN et d'activation de la transcription est apportée par NF-YA. NF-YB et C n'ont pas de domaine de liaison fort à l'ADN, ni de domaine d'activation de la transcription, mais s'assemblent pour former un hétérodimère (Romier et al., 2003). Il a été montré que le trimère NF-Y mime la forme d'une histone, qui lui permet

d'interagir avec la chromatine et de jouer son rôle de facteur pionnier (Nardini et al., 2013; Oldfield et al., 2014).

2.3.3.1. Rôle de LEC1 (NF-YB9) comme facteur pionnier

Le répresseur de la floraison *FLC* (*FLOWERING LOCUS C*) joue un rôle déterminant au cours du processus de vernalisation, *i.e.* lorsque les plantes ne peuvent fleurir et/ou se développer qu'après avoir été soumises à des températures froides. Avant la vernalisation, son expression est induite par des marques épigénétiques permissives de type H3K4me3 et H3K36me3 ainsi que par FRIGIDA (FRI), une protéine *scaffold*, permettant la formation d'un complexe protéique FRI-C, agissant comme un régulateur transcriptionnel spécifique de *FLC*, recrutant la machinerie basale de transcription et des modificateurs chromatinien, de façon à réprimer une floraison précoce (Choi et al., 2011). À l'inverse, lorsque les plantes sont traitées au froid, le complexe PRC2-VRN induit le dépôt de marques répressives H3K27me3 au locus *FLC* (Figure Intro-8). Les plantes sont alors vernalisées : le répresseur de floraison *FLC* ne s'exprime plus et les plantes peuvent fleurir. Néanmoins, les marques épigénétiques déposées au locus *FLC* sont remises à zéro dans l'embryon, ce qui permet l'expression de *FLC* à la génération suivante (Crevillén et al., 2014), on parle alors de *resetting*. C'est ainsi que LEC1, qui est exprimé spécifiquement pendant l'embryogénèse, est capable de réactiver *de novo* l'expression de *FLC* pendant l'embryogénèse précoce, permettant ainsi de ne pas transmettre à la future génération, l'état vernalisé acquis par le parent (Tao et al., 2017). Les NF-Y sont capables d'accéder aux éléments *cis*-régulateurs présents sur l'ADN, bien que leurs cibles soient rendues inaccessibles par les Polycomb. Ils forment une structure mimant l'ADN nucléosomal H2A-H2B, rendant la chromatine plus accessible et facilitant le déplacement des histones (Nardini et al., 2013; Oldfield et al., 2014). De cette même manière, LEC1 pourrait alors se lier au locus *FLC* dont l'expression est bloquée, déplacer les protéines Polycomb et faciliter le recrutement de EARLY FLOWERING IN SHORT DAYS (EFS), une méthyltransférase et ainsi le dépôt de marques actives H3K36me3 (Tao et al., 2017). La même équipe a montré que l'action de LEC1 comme facteur pionnier permet de recruter LEAFY COTYLEDON 2 (LEC2) et FUSCA 3 (FUS3) au niveau des boîtes *RY* situées au niveau des *CME* (*Cold Motif Element*) du locus *FLC* (Tao et al., 2019). LEC2 et FUS3 sont alors en compétition avec VIVIPAROUS 1/ABI3-LIKE (VAL) 1 et 2, également appelés HIGH-LEVEL EXPRESSION OF SUGAR-INDUCIBLE GENE 2 (HSI2) et HSI2-LIKE 1 (HSL1), des facteurs de transcription de type B3 (de la même famille que LEC2 et FUS3), dont le rôle est de maintenir les marques répressives déposées par les Polycomb. La liaison de LEC2 et FUS3 au locus *FLC* va alors déplacer les VAL et permettre le recrutement de FRI et d'autres protéines modificatrices de la chromatine comme EFS et ainsi conduire à l'ouverture de la chromatine permettant l'activation

Method	Input	TF source	Time	Relative cost	Advantages	Disadvantages
ChIP-seq	Cross-linked chromatin	Endogenous or <i>in vivo</i> expressed recombinant protein	4–5 days	Medium	Can be applied for a wide range of organisms	Low resolution of binding site maps; prone to false positive and false negative errors; low signal-to-noise ratio peaks
ChIP-exo	Cross-linked chromatin	Endogenous or <i>in vivo</i> expressed recombinant protein	4–5 days	High	High resolution of binding site maps	High technical complexity
ChIP-nexus	Cross-linked chromatin	Endogenous or <i>in vivo</i> expressed recombinant protein	4–5 days	High	High resolution of binding site maps	High technical complexity
CUT&RUN	Native chromatin	Endogenous or <i>in vivo</i> expressed recombinant protein	2 days	Medium	High resolution of binding site maps; high signal-to-noise ratio peaks	No published report for plant TFs
CUT&Tag	Native chromatin	Endogenous or <i>in vivo</i> expressed recombinant protein	2 days	Medium	High resolution of binding site maps; high signal-to-noise ratio peaks	No published report for plant TFs
DamID	Native chromatin	<i>In vivo</i> expressed recombinant protein	4–5 days	High	Identification of transient TF-DNA interactions	Low resolution of binding site maps
scCC	RNA	<i>In vivo</i> expressed recombinant protein	2 days	Medium	Simultaneous measure of transcriptome and TF binding profiles at the single-cell level	Low resolution of binding site maps; possible modification or silencing of target gene expression due to transposon integration
nextPBM	Randomized synthetic DNA	Endogenous or <i>in vivo</i> expressed recombinant protein	2–3 days	Medium	Captures the effect of TF–protein interactions and post-translational modifications on DNA binding specificity and affinity	Lack of endogenous genome sequence and chromatin context
PBM	Randomized synthetic DNA	<i>In vitro</i> or non-native cell expressed recombinant protein	2 days	Low	Identifies binding sequence motifs in a high-throughput manner	Lack of endogenous genome sequence and chromatin context
SELEX-seq	Randomized synthetic DNA	<i>In vitro</i> or non-native cell expressed recombinant protein	2 days	Low	Identifies binding sequence motifs in a high-throughput manner	Lack of endogenous genome sequence and chromatin context
DAP-seq	Genomic DNA	<i>In vitro</i> or non-native cell expressed recombinant protein	2 days	Low	High resolution of binding site maps in endogenous genome context; high signal-to-noise ratio peaks; can be easily performed in a high-throughput manner; can be used to dissect the direct and indirect binding sites and disentangle the cooperative action of TFs	Lack of chromatin context
Multi-DAP-seq	Genomic DNA	<i>In vitro</i> or non-native cell expressed recombinant protein	2 days	Low	Can be applied to non-model organisms; high resolution of binding site maps in endogenous genome context; high signal-to-noise ratio peaks; can be easily performed in a high-throughput manner	Lack of chromatin context; potential modification of DNA binding specificity and/or affinity due to incorporation of biotinylated lysine into the TF protein sequence

Figure Intro-12 : Outils pour l'étude des interactions entre les facteurs de transcription et l'ADN.

D'après Hajheidari and Huang, 2022.

transcriptionnelle. Ainsi, la liaison de *LEC2*, *FUS3* et des *VAL* sur les mêmes cibles permet une dynamique d'induction et de répression de l'expression de *FLC* tout au long du cycle de vie de la plante. Après la fécondation, l'expression de *LEC1* est activée, suivie par celle de *LEC2* et *FUS3*, conduisant à l'activation de *FLC* dans l'embryon. Puis, après la germination, *LEC2* et *FUS3* ne s'expriment plus, les *VAL* peuvent se lier au locus *FLC* et maintenir la répression médiée par les PcG.

2.4. Outils pour l'étude des interactions protéines/ADN/chromatine

L'expression des gènes est donc régulée par de nombreux mécanismes transcriptionnels, impliquant des régulations chromatinienne qui permettent l'accès de la machinerie transcriptionnelle aux promoteurs, ainsi que des régulations spécifiques par des facteurs de transcription. L'étude des interactions entre les protéines et la chromatine ou l'ADN, impliquées dans le contrôle de l'expression des gènes, est rendue possible grâce à l'émergence de nouvelles techniques de biochimie associées au séquençage à haut débit d'ADN ou d'ARNm. Elles permettent de mieux appréhender la manière dont l'expression des gènes est régulée. Cependant l'efficacité de ces différentes méthodes peut être altérée par le niveau d'expression des facteurs de transcription, qui est souvent très faible, ce qui peut affecter la fiabilité des résultats. Certaines d'entre elles sont difficiles à mettre en place, coûteuses ou encore peu développées chez les plantes (Figure Intro-12) (Hajheidari and Huang, 2022). Différentes méthodes, basées sur l'utilisation d'enzymes sensibles ou non à l'ADN (*e.g.* des DNases), bloquées ou non par les nucléosomes ou les facteurs de transcription, permettent également de déterminer l'accessibilité de la chromatine (Figure Intro-13) (Klemm et al., 2019).

Nous avons vu au cours des deux premières parties de cette introduction que les plantes sont capables de se reproduire de différentes manières, que ce soit par la formation d'embryons zygotiques à l'intérieur d'une graine, dans le cadre de la reproduction sexuée, ou bien par propagation végétative en réponse aux signaux environnementaux. L'expression des gènes contrôlant les divers aspects de la reproduction des plantes est soumise à de fortes régulations, comme celles décrites ci-dessus. Les prochaines parties ont pour but de décrire les principaux mécanismes moléculaires régulant le développement de la graine d'une part et l'induction de l'embryogenèse somatique d'autre part.

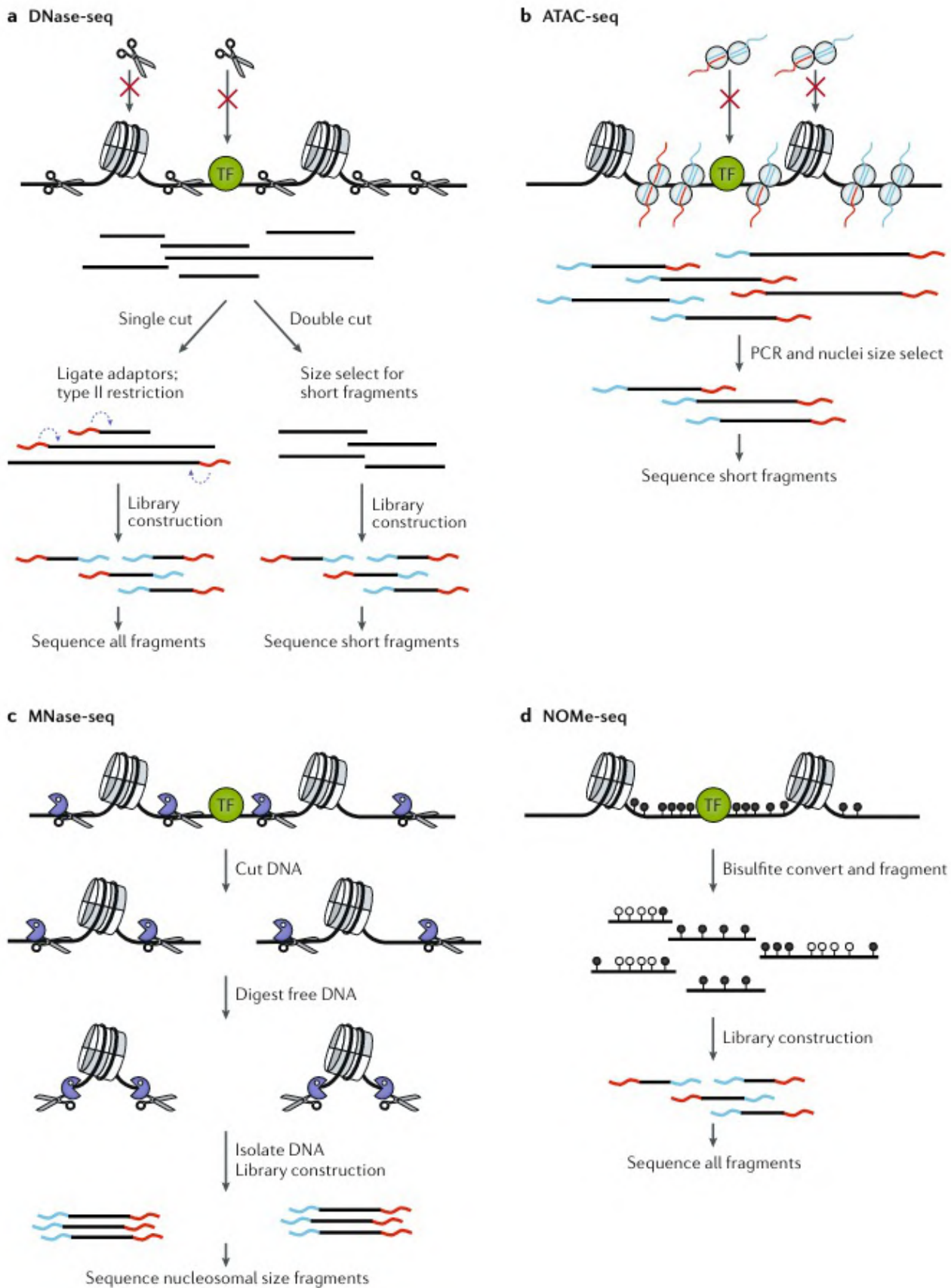


Figure Intro-13 : Principales méthodes pour la mesure de l'accessibilité de l'ADN. A) *DNase I hypersensitive site sequencing* (DNase-seq) : l'endonucléase DNase est utilisée pour cliver l'ADN accessible et non fixé par des protéines (facteurs de transcriptions ou nucléosomes). B) *Assay for transposase-accessible chromatin using sequencing* (ATAC-seq) : une transposase hyperactive (Tn5) est utilisée pour cliver et liguer simultanément des adaptateurs sur l'ADN accessible non lié par des facteurs de transcription ou des nucléosomes. C) *Micrococcal nuclease sequencing* (MNase-seq) : une endonucléase/exonucléase MNase est utilisée pour cliver et éliminer l'ADN accessible. D) *Nucleosome occupancy and methylome sequencing* (NOMe-seq) : une GpC méthyltransférase est utilisée pour méthyler l'ADN accessible, suivie d'une conversion au bisulfite des cytosines non méthylées en uracile.

Abréviations : TF, facteur de transcription.

D'après Klemm et al. 2019

3. Régulation du développement et du remplissage de la graine

3.1. Contrôle épigénétique du développement de la graine

Le développement de la graine et les transitions entre les différentes phases sont régulés par des modifications d'histones et le remodelage de la chromatine qui permettent l'expression de gènes spécifiques au bon moment, en réponse aux conditions environnementales, de telle manière que la reproduction sexuée et la propagation de la descendance se fassent de manière coordonnée, pour revue (Ding et al., 2022). Dans un premier temps, les PcG et notamment PRC2-FIS, sont au cœur de la régulation de la formation de la graine, avant la fécondation, en agissant à la fois sur la formation de l'albumen et sur celle de l'enveloppe. Comme décrit précédemment, le complexe PRC2-FIS est composé de quatre membres : FERTILIZATION INDEPENDANT SEED 2 (FIS2), MEDEA (MEA), FERTILIZATION INDEPENDANT ENDOSPERM (FIE) et MULTICOPY SUPPRESSOR OF IRA 1 (MSI1) (Figure Intro-9) (Hennig and Derkacheva, 2009). Avant fécondation, les PcG exprimés dans le gamétophyte femelle répriment la formation de l'enveloppe et le complexe PRC2-FIS réprime la réplication autonome de la cellule centrale en bloquant les divisions cellulaires (Roszak and Köhler, 2011; Figueiredo et al., 2015). Ainsi, la perte de fonction de PRC2-FIS dans les mutants *fis2* ou *mea* induit la réplication autonome de la cellule centrale, donc la formation d'un albumen asexué, mais pas la formation de l'enveloppe (Figure Intro-14). De l'autre côté, la perte de fonction des PcG sporophytiques dans le gamétophyte femelle promeut la formation de l'enveloppe. Ainsi, la perte de fonction de PRC2-FIS cumulée à celle des PcG sporophytiques induit la formation de graines autonomes avec un albumen et une enveloppe (Roszak and Köhler, 2011). En d'autres termes les réseaux génétiques et/ou hormonaux conduisant à la formation de graines apomictiques sont régulés par des marques épigénétiques répressives, notamment H3K27me3, et la perte de fonction des PcG, notamment PRC2-FIS, supprime cette répression et permet le développement de graines autonomes, *i.e.* sans fécondation. Après la fécondation, la répression de la formation de l'enveloppe est abolie par un signal hormonal lié à l'auxine, provenant de l'albumen, transporté vers les téguments et dépendant du facteur de transcription AGAMOUS-LIKE 62 (AGL62) (Figueiredo et al., 2016). La formation de l'albumen conduit également à la dégradation du nucelle, médiée par des signaux régulés par PRC2, AGL62, ainsi que le facteur de transcription TRANSPARENT TESTA 16 (TT16) (Xu et al., 2016b)

Au cours du développement, il a été montré que le complexe PRC2 n'est pas essentiel à l'établissement du plan d'organisation de l'embryon chez *Arabidopsis thaliana*, contrairement à ce qui est observé chez les animaux, bien que des mutations dans différentes histones

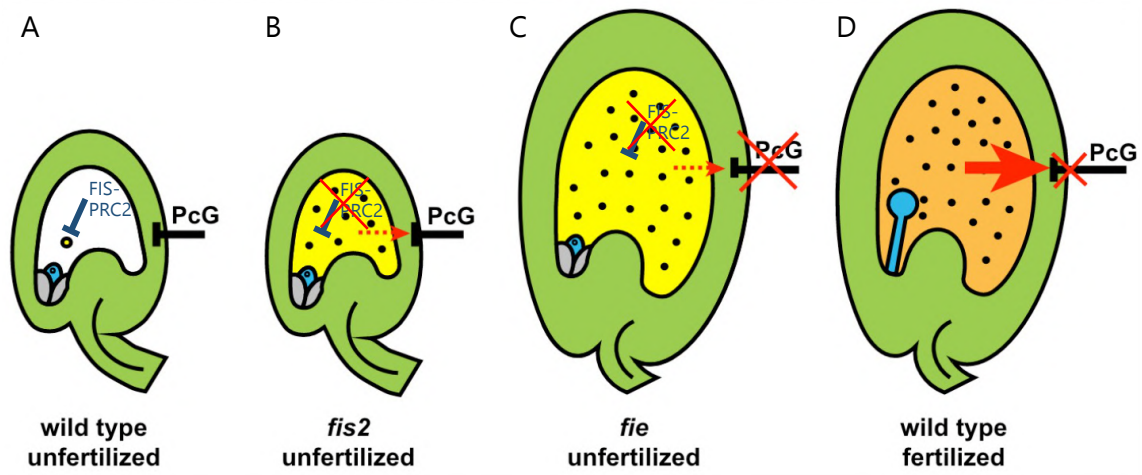


Figure Intro-14 : La formation de graines autonomes, *i.e.* sans fécondation est réprimée les PcG qui déposent des marques épigénétiques répressives. A) Dans les ovules sauvages, les PcG exprimés spécifiquement dans le gamétophyte femelle répriment le développement de l'enveloppe et PRC2-FIS réprime la réplication autonome de la cellule centrale. B) Chez les mutants *fis*, l'albumen se forme de manière autonome, mais la répression exercée par les PcG dans les téguments est maintenue, empêchant le développement de l'enveloppe. C) Chez les mutants *fie*, le blocage de la formation de l'enveloppe par les PcG est supprimé : l'albumen et l'enveloppe se forment de manière autonome. D) Lors du développement des graines sauvages après fécondation, un signal auxinique dépendant d'AGL62 est transmis par l'albumen sexué vers les téguments (représenté par la flèche rouge) et permet le retrait des marques répressives déposées par les PcG, conduisant à la formation de l'enveloppe. Dans les mutants *fis2* et *fie* le signal formé par l'albumen autonome est absent ou trop faible (représenté par la flèche rouge en pointillés) pour induire la formation de l'enveloppe.

Modifié d'après Roszak and Köhler, 2011.

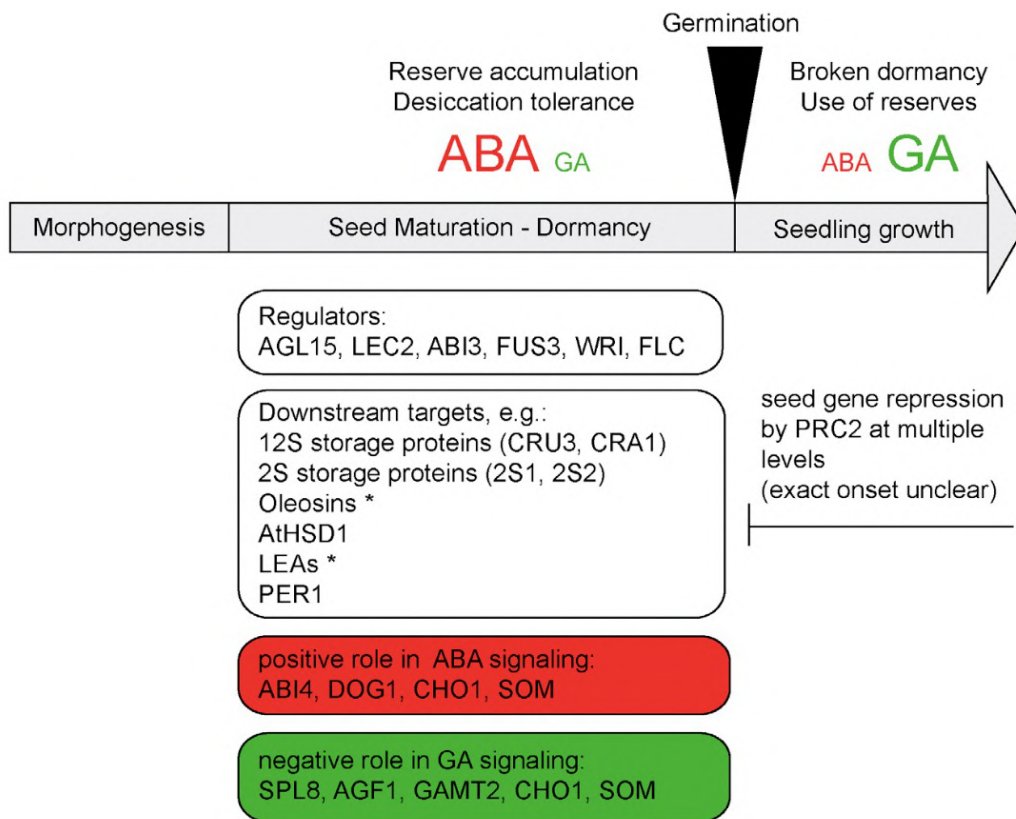


Figure Intro-15 : PRC2 réprime les gènes de maturation et de dormance dans les plantules. Les gènes représentés ici ont été identifiés comme des cibles de H3K27me3 et leur expression augmente significativement dans le mutant *fie*. Il s'agit de gènes régulateurs du développement, des intégrateurs de signaux environnementaux, des gènes impliqués dans l'accumulation de réserve et la tolérance à la dessiccation, ainsi que des acteurs des voies de signalisation de l'acide abscissique (ABA) et des gibbérellines (GA). La différence de taille des caractères pour ABA et GA représente les niveaux élevés et bas des hormones.

D'après Bouyer et al., 2011.

méthyltransférases conduisent à la formation d'embryons plus gros et de graines plus larges chez *Arabidopsis thaliana* (Liu et al., 2016; Cheng et al., 2018). Néanmoins, la fonction de PRC2 est nécessaire à la transition entre un état embryonnaire et un état végétatif (Bouyer et al., 2011). En effet, PRC2 contrôle le passage d'un état « graine », c'est-à-dire un état embryogène et d'accumulation de réserves, à la plantule en induisant le dépôt de marques H3K27me3 sur de nombreux gènes impliqués dans la maturation et la germination ainsi qu'au niveau de la signalisation hormonale (Figure Intro-15). La perte de fonction de PRC2 dans le mutant *fie* permet la formation d'un embryon et le développement précoce de la plantule de manière normale, mais après quelques jours de développement végétatif, la plantule est très affectée, avec l'apparition de calcs, d'embryons somatiques, la formation de poils racinaires au niveau des parties aériennes et de trichomes au niveau des poils racinaires (Bouyer et al., 2011). Ainsi, PRC2 assure, après la germination, la répression des gènes impliqués dans le développement de la graine afin d'assurer un développement végétatif normal.

Durant le développement de la graine, il a été observé que le niveau de méthylation dans le contexte CHH augmente de manière significative, mais cible plus particulièrement les éléments transposables (ET), probablement pour protéger la future plante de l'effet délétère qui peut être engendré par les ET (Bouyer et al., 2017; Lin et al., 2017). Des défauts de méthylation n'impactent pas le développement de la graine et, de manière générale, les gènes régulant les majeurs aspects du développement et de la maturation sont présents dans des régions génomiques non soumises à la méthylation, quel que soit le stade de développement de la graine (Lin et al., 2017). A la fin du développement, cependant, lors de la transition entre la phase de dormance et la germination, des modifications d'histones permettent l'activation de gènes spécifiques de la germination, par le dépôt de marques H3K4me3, tandis que les gènes de développement et de maturation de la graine sont réprimés par le remplacement de marques H3K4me3 par des marques H3K27me3 (Müller et al., 2012). De la même manière que lors de la répression du locus *FLC* en réponse au froid, des marques répressives sont déposées en réponse à la stratification des graines.

3.2. Contrôle hormonal du développement de la graine

3.2.1. L'auxine, régulateur de l'embryogenèse précoce

L'auxine joue un rôle majeur tout au long du cycle de vie de la plante, dans le contrôle des processus développementaux, comme l'initiation et la formation de nouveaux organes, les divisions, la croissance et la différenciation cellulaire, en réponse aux signaux environnementaux, voir (Casanova-Sáez and Voß, 2019) pour revue. La dynamique de la concentration d'auxine dans les

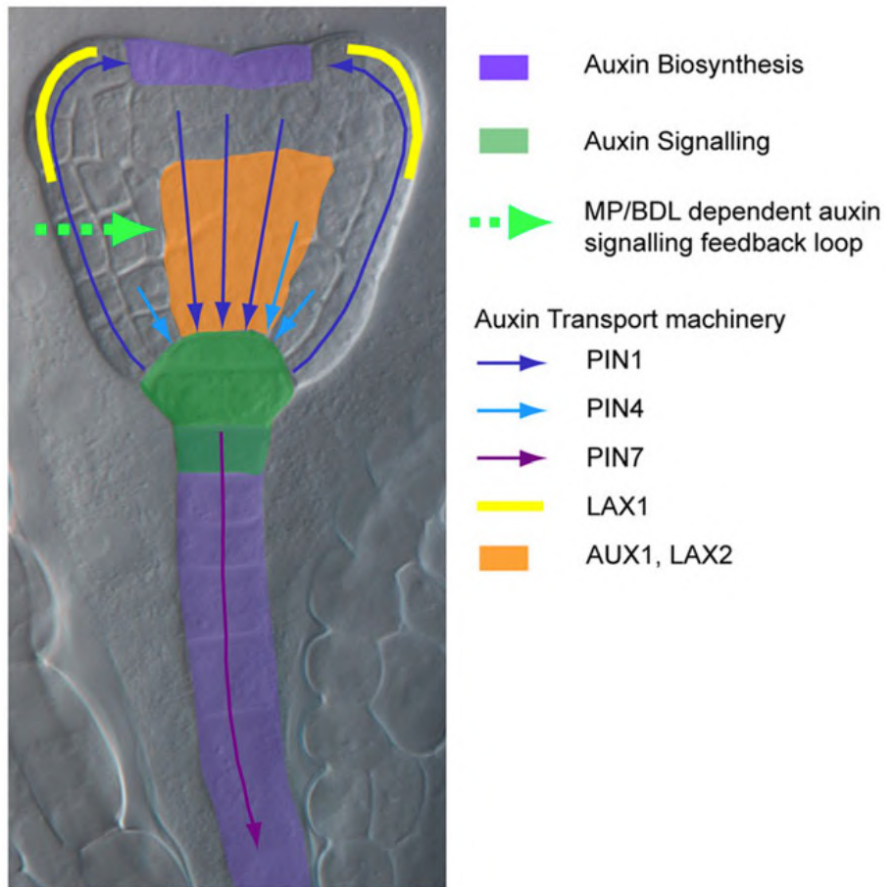


Figure Intro-16 : Modèle du rôle de l'auxine, sa concentration et son transport dans la mise en place de la polarité apicale-basale au cours de l'embryogenèse. La production d'auxine a lieu aux deux opposés : le méristème apical et le suspenseur. Le transport de l'auxine est médié par des transporteurs d'influx (AUX, LAX) et d'efflux (PIN). Une régulation transcriptionnelle des transporteurs d'auxine est médiée par MONOPTEROS (MP) et BODENLOS (BDL).

D'après Robert et al., 2015.

cellules, en particulier l'influx, l'efflux ainsi que la façon dont les gradients d'auxine sont établis et maintenus par un réseau de signalisation et un métabolisme très précis, sont au cœur du développement des plantes. Ainsi, le développement de la graine ne fait pas exception et est très finement régulé par l'auxine.

Dans un premier temps, l'auxine est nécessaire au développement de l'albumen en induisant la réplication de la cellule centrale après fécondation (Figueiredo et al., 2015). De plus, la production d'auxine dans l'albumen forme un signal dépendant du facteur de transcription AGL62 et nécessaire au développement de l'enveloppe (Figure Intro-14) (Figueiredo et al., 2016). Il a également été montré que la concentration d'auxine dans l'albumen est déterminante pour sa cellularisation, une trop forte concentration d'auxine la bloquant (Batista et al., 2019). Ainsi, l'auxine constitue un signal essentiel à l'initiation du développement de la graine avec, dans un premier temps, un effet sur l'albumen et l'enveloppe. Elle régule également la croissance de l'embryon, en particulier l'établissement de l'axe de polarité apico-basal lors des premières divisions cellulaires asymétriques du zygote. En effet, la biosynthèse, l'accumulation (minimums et maximums) et le transport polarisé de l'auxine par influx et efflux déterminent l'architecture de l'embryon, nécessitant l'expression des gènes de biosynthèse comme les *YUCCA* (*YUC*) mais aussi des transporteurs comme les *PIN-FORMED* (*PIN*) ou les *AUXIN1/LIKE-AUXIN* (*AUX/LAX*) (Robert et al., 2013). Au cours des premières divisions du zygote, on observe un transport polarisé de l'auxine de la cellule basale vers la cellule apicale, médié par le transporteur PIN7 (Zhang and Laux, 2011). Par la suite, le transport d'auxine se fait à la fois du suspenseur vers les cotylédons via PIN1 et LAX1, ainsi que du méristème apical caulinaire vers le méristème apical racinaire via PIN1, 4 et AUX1 et 2 (Figure Intro-16) (Robert et al., 2015; Figueiredo and Köhler, 2018).

3.2.2. L'acide abscissique et les gibbérellines, balance entre maturation et germination

Les gibbérellines jouent un rôle important au cours de l'embryogenèse tardive. Une altération de leur biosynthèse ou de la signalisation des gibbérellines conduit à un défaut de développement des cotylédons et de l'axe embryonnaire tandis qu'un excès conduit à des graines plus grosses (Hu et al., 2018).

Plus tard, le ratio entre l'acide abscissique (ABA) et les gibbérellines (GA) régule la maturation, la dormance et la germination des graines, pour revue, voir (Finkelstein et al., 2008). L'ABA est nécessaire pour la maturation, l'induction de la dormance et l'inhibition de la germination précoce (Figure Intro-15). Les graines déficientes en ABA ou dans la signalisation, comme les mutants *abi3*

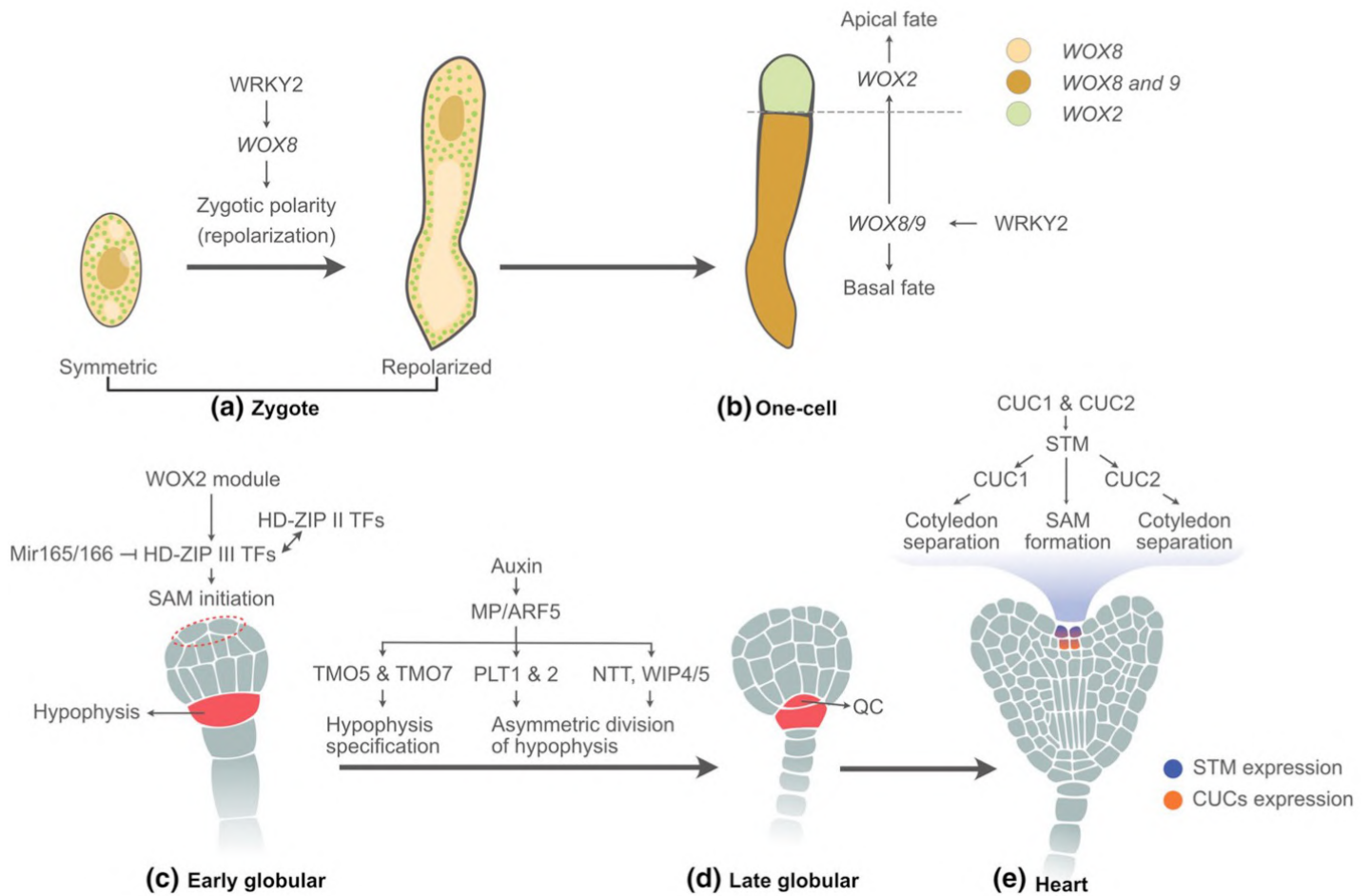


Figure Intro-17 : Facteurs de transcription impliqués dans la formation de l'embryon. A) Dans le zygote, WRKY2 active l'expression de WOX8 pour promouvoir la polarisation du noyau (marron) et des vacuoles (jaune) et la division asymétrique du zygote. B) WRKY2 et WOX8/9 régulent le développement du proembryon en activant l'expression de WOX2 dans la région apicale (vert). C) Au stade globulaire précoce, le module WOX2 (WOX1/2/3/5) est requis pour l'initiation du méristème apical caulinaire. Le module active l'expression des FT HD-ZIP III pour réprimer la différenciation des cellules souches méristématiques. L'expression des FT HD-ZIP III est contrôlée par les miRNA 165/166. Ces FT agissent de concert avec les FT HD-ZIP II pour réguler la formation de la région apicale pendant l'embryogenèse. L'hypophyse se spécifie *via* l'activation de TMO5/7 par MP/ARF5 en réponse à l'auxine. D) D'autres cibles d'ARF5 comme NTT, WIP4/5 et PLT sont requises pour la division asymétrique de l'hypophyse et la spécification du centre quiescent au stade globulaire tardif. E) Les CUC régulent la formation du domaine frontière entre les cotylédons et agissent de manière synergique avec STP pour contrôler la formation du méristème apical caulinaire.

Abréviations : QC, centre quiescent ; SAM, méristème apical caulinaire.

D'après Verma et al., 2022.

par exemple, présentent un défaut de dormance et continuent de se développer après l'embryogenèse *sensu stricto*, jusqu'à germer directement sur pied. L'ABA est synthétisé à la fois par l'embryon et l'albumen au cours du développement, avec un pic au moment de l'induction de la dormance et il confère à la graine sa capacité de tolérance à la dessiccation, permettant la survie à l'état sec pendant une longue période. Les GA, quant à elles, stimulent la germination en induisant l'accumulation d'enzymes hydrolytiques, en particulier dans les cellules de la couche à aleurones, capables de casser les barrières formées par l'albumen et l'enveloppe, de même que la mobilisation des réserves de la graine et en stimulant l'expansion de l'embryon. Les GA permettent également le passage d'un état embryonnaire à un état végétatif en modifiant l'accessibilité de certains *loci* par le remodeleur de la chromatine PICKLE (PKL) (Finkelstein et al., 2008). D'autres hormones comme les brassinostéroïdes et l'éthylène promeuvent la germination et lèvent la dormance.

3.3. Contrôle transcriptionnel du développement de la graine

Les différents aspects du développement de la graine, allant de l'embryogenèse à la germination en passant par la maturation sont largement régulés par de nombreux facteurs de transcription, pour revue (Verma et al., 2022). Les mécanismes les plus connus sont décrits dans ce chapitre.

3.3.1. Contrôle transcriptionnel au cours de l'embryogenèse

Chez *Arabidopsis thaliana*, lors des premières phases du développement de l'embryon, les facteurs de transcription de type WOX (WUSCHEL-RELATED HOMEODOMAIN) sont impliqués dans la mise en place, puis le maintien de la polarité apicale-basale lors des premières divisions cellulaires du zygote (Figure Intro-17) (Breuninger et al., 2008). Au stade globulaire, la formation de la structure apicale de l'embryon est également régulée par PHABULOSA (PHB), PHAVOLUTA (PHV) et REVOLUTA (REV), des facteurs de transcription de la famille des HOMEODOMAIN-LEUCINE ZIPPER de classe III et d'autres facteurs de classe II (HD-ZIP II et III), ainsi que des APETALA 2/ETHYLENE RESPONSE FACTOR (AP2-ERF), comme DORNRÖSCHEN (DRN) (Emery et al., 2003; Cole et al., 2009; Turchi et al., 2013). Plus tard, au stade cœur, la formation des domaines frontière (*i.e.* une file de petites cellules permettant d'individualiser les organes) au niveau des cotylédons, permet la mise en place de la symétrie bilatérale de l'embryon et est essentielle à la formation du méristème apical caulinaire. Elle est régulée par les facteurs de transcription CUP-SHAPED COTYLEDON (CUC), appartenant à la famille des NAC. Le FT SHOOT MERISTEMLESS (STM), de la famille des KNOX assure la formation du méristème apical caulinaire (MAC) (Aida et al., 1999). Les WOX sont également impliqués dans la balance hormonale entre l'auxine et les cytokinines pour la formation du MAC (Zhang et al., 2017). Les WOX sont nécessaires à l'expression de *CLAVATA 3 (CLV3)*, un

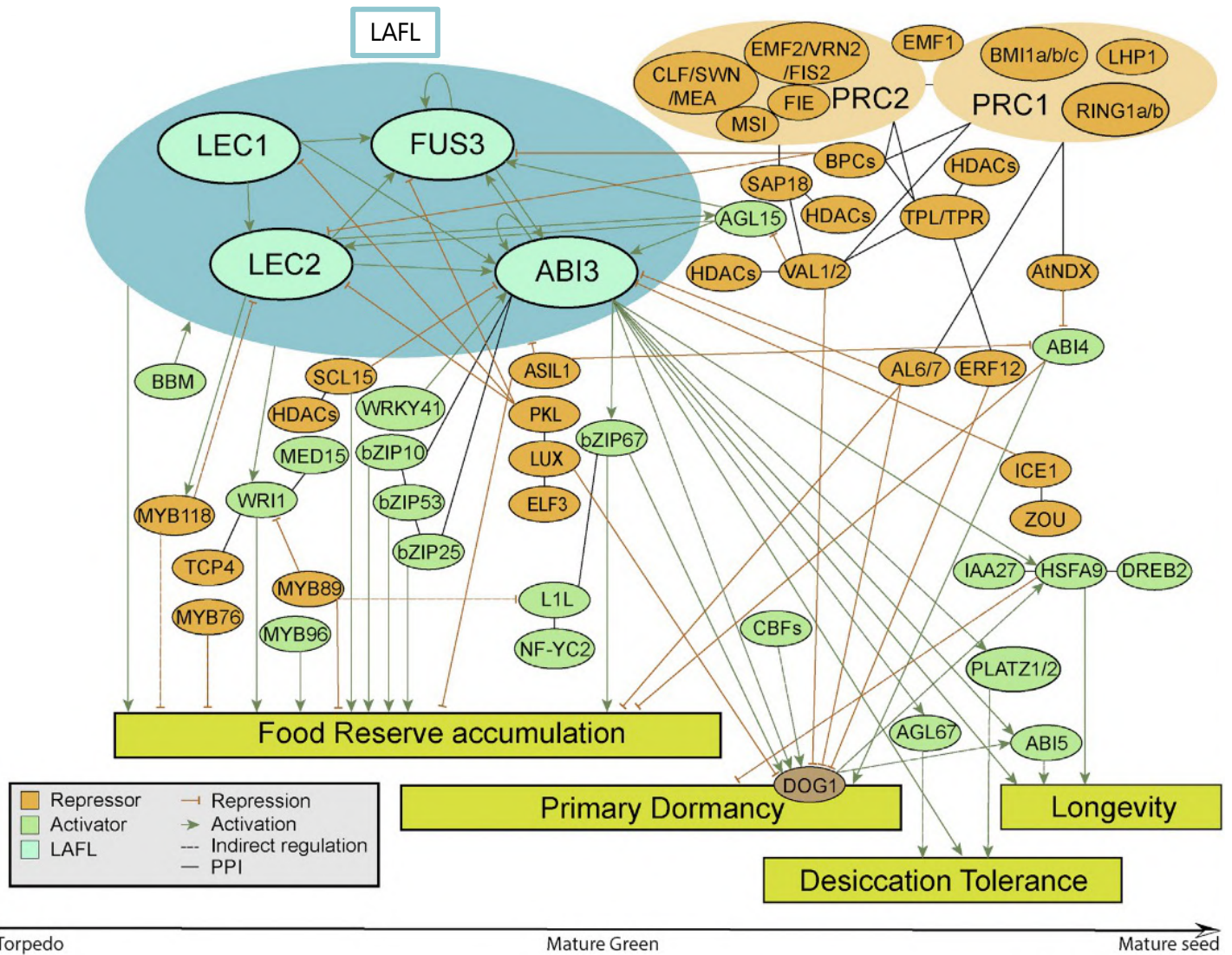


Figure Intro-18 : Les facteurs de transcription forment un réseau complexe pour réguler la maturation de la graine. Vue d'ensemble des événements orchestrant la maturation de la graine au travers des réseaux transcriptionnels et cascades de régulation. Les régulateurs maîtres LAFL agissent au sein d'un réseau central. Les régulations indirectes ont été définies sur la base d'études génétiques ou sur l'abondance de transcrits, en absence de données de liaison des facteurs de transcription sur les gènes régulés.

D'après Alizadeh et al., 2021.

peptide conduisant à la mise en place des cellules souches. Ces différentes protéines agissent de concert pour réguler finement l'architecture de l'embryon (Figure Intro-17). Il en va de même pour la formation de la partie racinaire et la mise en place de la symétrie radiale, qui est aussi régulée par de nombreux facteurs de transcription (Verma et al., 2022).

Le facteur de transcription BABY BOOM (BBM), membre de la famille des AP2-ERF, est lui-aussi un régulateur clé du développement de la graine. Il joue un rôle sur la formation de l'embryon et le développement de l'albumen (Chen et al., 2022). Il agit de manière redondante avec PLETHORA 2 (PLT2), également un FT AP2-ERF, pour réguler la formation de l'embryon, mais aussi la prolifération et la cellularisation de l'albumen. D'ailleurs, le double mutant *bbm plt2* est létal à l'état embryonnaire (Galinha et al., 2007). L'expression ectopique de *BBM* dans la cellule œuf du gamétophyte femelle conduit à la formation d'un embryon haploïde (parthénogenèse) (Chen et al., 2022) et sa surexpression induit la formation d'embryons somatiques, voir Introduction – partie 4. [Revue : « Contrôle génétique et moléculaire de l'embryogenèse somatique »](#). Enfin, BBM active la transcription des régulateurs maîtres de la maturation de la graine (Horstman et al., 2017).

3.3.2. Les LAFL, régulateurs maîtres du développement de la graine

Le processus de maturation de la graine est également soumis à un fort contrôle transcriptionnel (Figure Intro-18) (Baud et al., 2008; Alizadeh et al., 2021). LEAFY COTYLEDON 1 (LEC1), ABSCISIC ACID INSENSITIVE 3 (ABI3), FUSCA 3 (FUS3) et LEAFY COTYLEDON 2 (LEC2) sont tous les quatre des facteurs de transcription formant le groupe des LAFL et sont connus pour être des régulateurs maîtres du développement de la graine (Fatihi et al., 2016). LEC2, ABI3 et FUS3 font partie de la famille des FT à domaine B3, qui est spécifique des végétaux (Yamasaki et al., 2013; Carbonero et al., 2016; Sasnauskas et al., 2018). Le domaine B3 est composé d'environ 110 acides aminés fortement conservés (Figure Intro-11). Chez *Arabidopsis thaliana*, on dénombre 118 membres dans la famille des B3 (Swaminathan et al., 2008). Les AFL-B3 (*i.e.* ABI3, FUS3 et LEC2) reconnaissent et se lient spécifiquement aux boîtes RY. Ces dernières sont caractérisées par une séquence consensus « CATGCA » et sont particulièrement présentes dans les promoteurs de gènes spécifiques de la graine et impliqués dans les processus de maturation (Reidt et al., 2000; Baud et al., 2016). Comme décrit précédemment, LEC1, ou NF-YB9, est membre de la famille de facteurs de transcription NF-Y (Lee et al., 2003).

Les LAFL forment une boucle d'autorégulation et sont capables d'induire directement leur propre expression (Santos Mendoza et al., 2005; To et al., 2006; Fatihi et al., 2016). Ils régulent l'expression de gènes impliqués dans les processus de maturation, comme les oléosines, les gènes de synthèse

Niveau relatif d'expression des LAFL, des VAL et PKL par rapport à EF1 α au cours du développement de la graine, données ePlant BAR Toronto

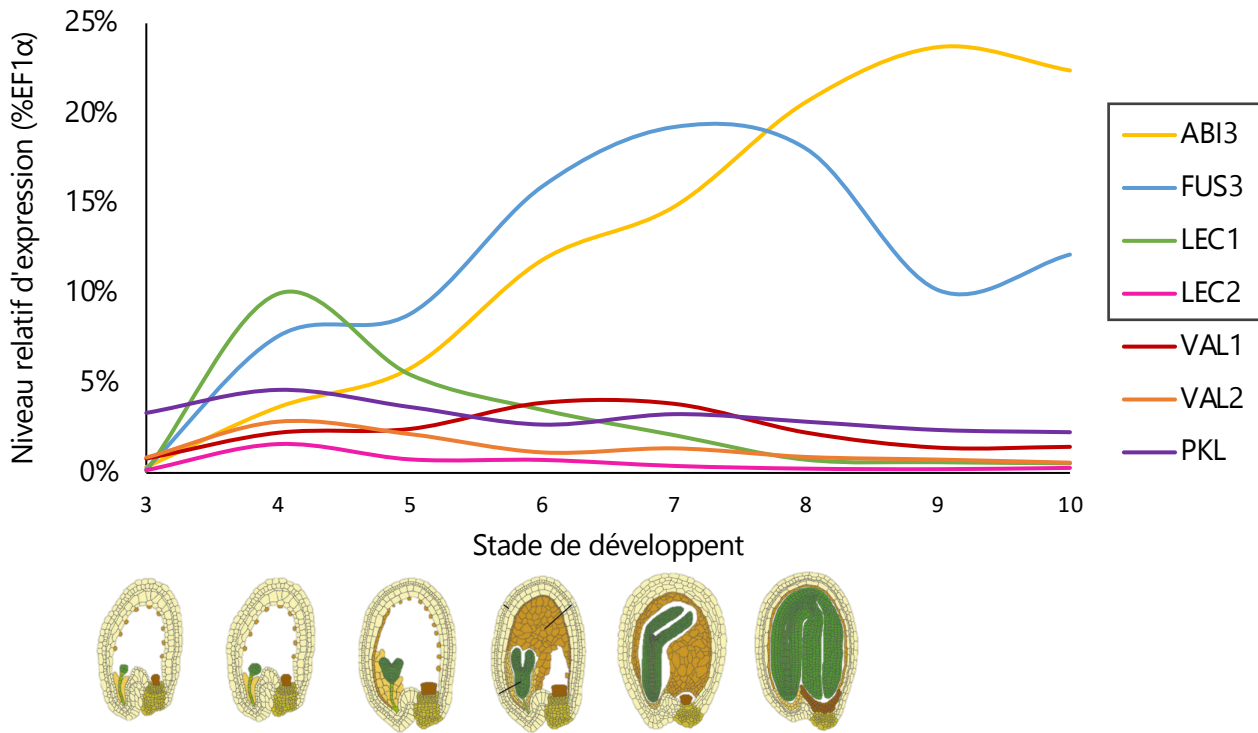


Figure Intro-19 : Niveau d'expression relatif des LAFL et de leurs régulateurs VAL1 et 2 et PKL au cours du développement de la graine. Les données d'expression sont issues de ePlant BAR Toronto (Waese et al., 2017) et sont exprimées en pourcentage d'expression relatif au gène constitutif EF1 α . Les schémas de graine en développement proviennent de Jo et al., 2019.

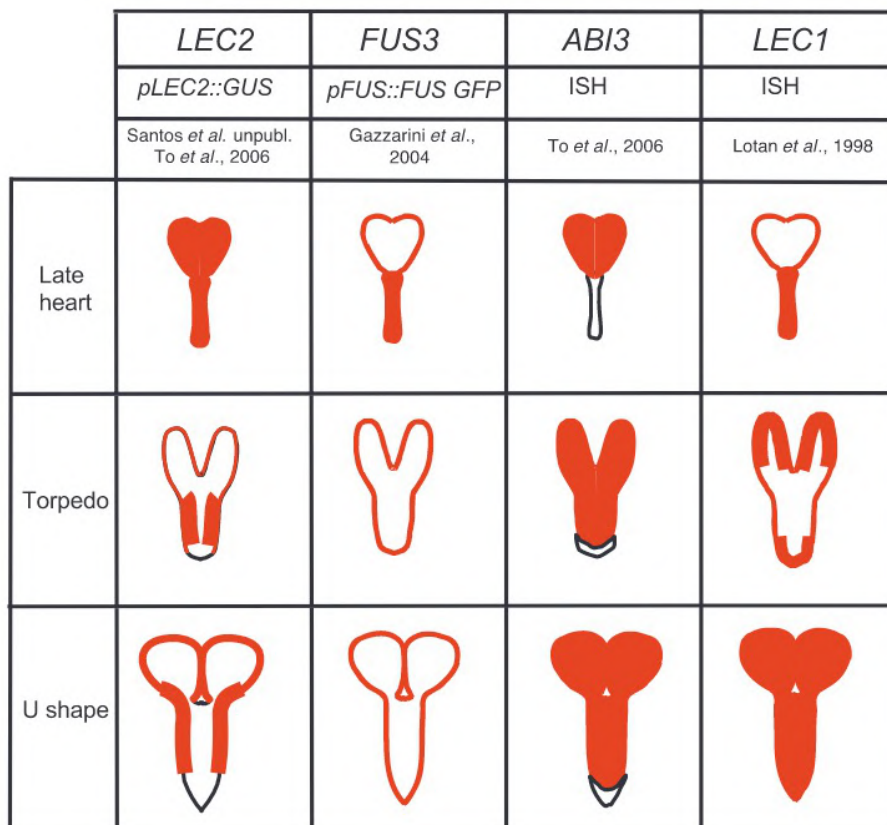


Figure Intro-20 : Patron d'expression des LAFL au cours du développement de la graine. Les domaines d'expression pour chaque facteur de transcription sont représentés en rouge. Les données d'expression sont issues de fusions avec un rapporteur GUS ou d'hybridation *in situ* (ISH).

des triacylglycérols, mais aussi des FT comme *WRINKLED 1 (WR1)*, qui code un régulateur de la biosynthèse d'huiles (Reidt et al., 2000; Braybrook et al., 2006; Baud et al., 2007; Wang et al., 2007; Santos-Mendoza et al., 2008; Stone et al., 2008; Boulard et al., 2017; Jo et al., 2019; Alizadeh et al., 2021). Les LAFL agissent ensemble et parfois avec d'autres facteurs comme les bZIP, pour former des complexes sur les promoteurs afin d'activer la transcription de gènes de biosynthèse des réserves, comme cela a été montré pour l'*OLEOSIN 1 (OLE1)* (Baud et al., 2016).

Les LAFL sont exprimés spécifiquement au cours du développement de la graine : *LEC1* et *LEC2* pendant les stades précoces de l'embryogenèse avec un pic au stade globulaire et *ABI3* et *FUS3* plus tardivement. *LEC2* est exprimé très faiblement (Figures Intro-19 et 20) (Santos-Mendoza et al., 2008; Fatihi et al., 2016; Waese et al., 2017). De manière générale, les mutants *laf1* présentent des phénotypes sévères avec des défauts de maturation et d'accumulation de réserves dans les graines et une intolérance à la dessiccation (Meinke et al., 1994; Baumlein et al., 1994; Santos-Mendoza et al., 2008; Roscoe et al., 2015). Lorsque les gènes « *lec* », i.e. *lec1*, *lec2* et *fus3* sont mutés, les cotylédons sont transformés en feuilles, avec l'apparition de trichomes et d'un système vasculaire complexe, démontrant que ces gènes sont requis pour le maintien de la spécificité d'identité embryonnaire des cotylédons et le devenir cellulaire au cours de l'embryogenèse précoce (Figure Intro-21) (Meinke et al., 1994). Les embryons peuvent également avoir des défauts au niveau du suspenseur alors anormalement développé ou au niveau des méristèmes apicaux caulinaire et racinaire. Les graines des mutants *lec1*, *lec2* et *fus3* présentent des taches pourpres à noires correspondant à une accumulation d'anthocyanes et de chlorophylle dans les cotylédons, marquant un stress lié à l'accumulation de sucres en réponse à la photosynthèse anormale dans les feuilles-cotylédons (Meinke et al., 1994). De façon plus spécifique, les mutants *laf1* ne présentent pas exactement les mêmes phénotypes (Barthole, 2013). Le mutant *abi3* n'a pas de défaut d'identité des cotylédons mais est insensible à l'acide abscissique, ce qui le rend particulièrement sensible à la dessiccation et engendre un défaut total de dormance avec un phénotype vivipare. Les embryons et graines du mutant *abi3* présentent une couleur verte persistante après la maturation, liée à l'accumulation de chlorophylle (Giraudat et al., 1992). Si le phénotype du mutant *lec2* est plus faible que celui observé pour *lec1*, les graines sont tout de même sensibles à la dessiccation et ne survivent pas très longtemps à l'état sec. Par exemple, pour maintenir les mutants *lec1* et *abi3* à l'état homozygote, il faut prélever les graines sur la plante mère avant la fin de la maturation et les semer aussitôt.

La surexpression des LAFL induit également des phénotypes forts, comme l'accumulation ectopique de triacylglycérols dans les feuilles dans le cas de *LEC2* (Figure Intro-22) (Mendoza et al., 2005) ou

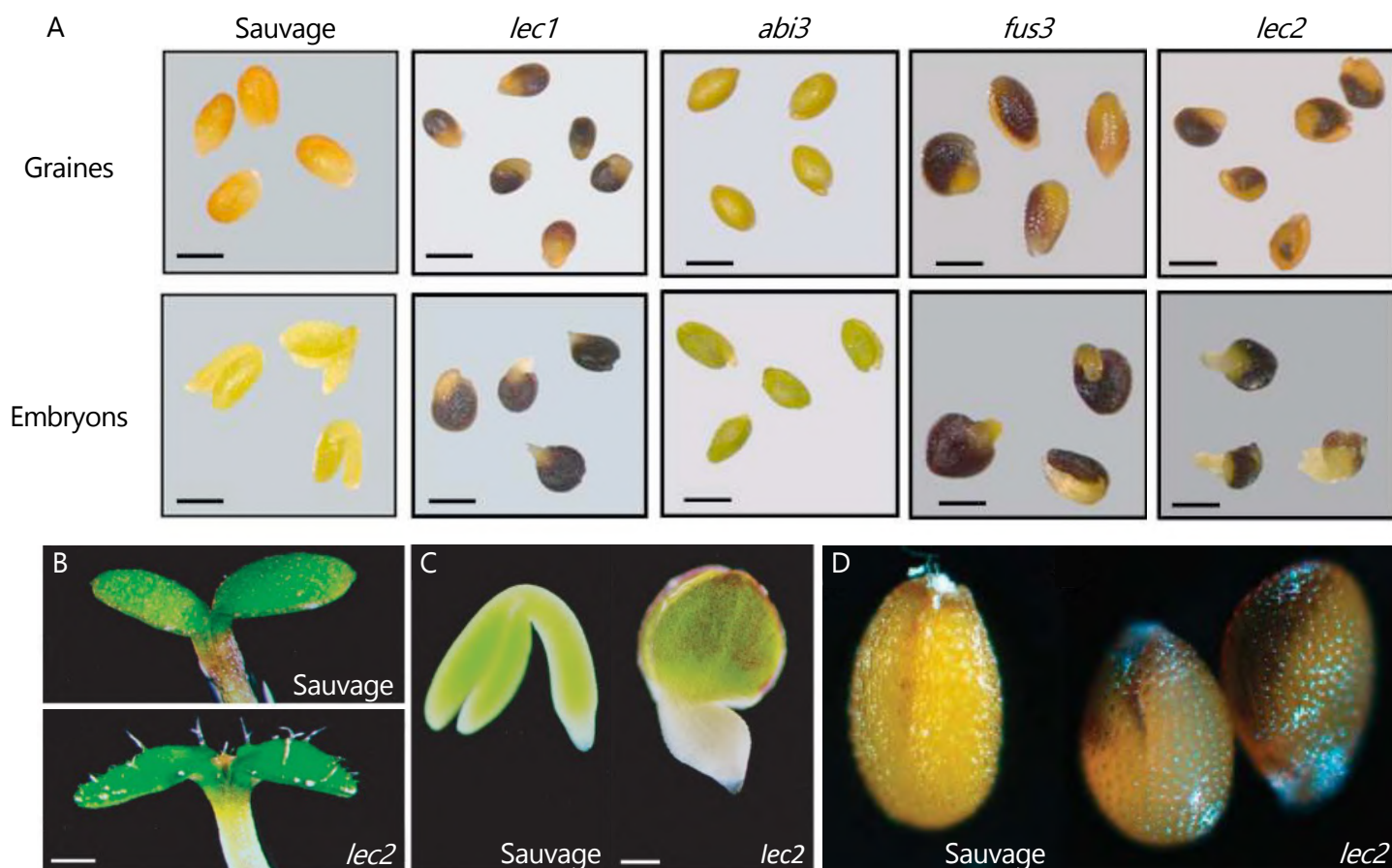


Figure Intro-21 : Phénotypes des mutants *lec1*, *abi3*, *fus3* et *lec2*. A) Phénotypes des graines et des embryons matures. Les mutants « *lec* » i.e. *lec1*, *fus3* et *lec2* sont caractérisés par une accumulation d'anthocyanes sur les cotylédons formant une tache marron visible sur la graine. Le mutant *abi3*, quant à lui, est caractérisé par une couleur verte persistante après la maturation de l'embryon et de la graine. B-D) Zoom sur le mutant *lec2*, qui comme les autres mutants *lec* a B) des cotylédons modifiés en feuilles avec la présence de trichomes et une accumulation d'anthocyanes visible au niveau C) des cotylédons et D) de la graine.

Barres d'échelle : A) 500 μm ; B) 300 μm ; C) 100 μm .

Modifié d'après Barthole, 2013 (A), Stone et al., 2001 (B-C), Santos Mendoza et al., 2005 (D).

de *LEC1* (Mu et al., 2008). De plus, la surexpression de *LEC2*, *LEC1* ou encore *FUS3* induit l'expression de gènes spécifiques de la maturation de la graine, comme les gènes de biosynthèse d'acides gras et de protéines de réserve, de tolérance à la dessiccation ou encore de dormance, ainsi que celle des autres *LAFL* (Mendoza et al., 2005; Wang et al., 2007; Mu et al., 2008). De manière intéressante, en fonction du stade de développement observé, *LEC1* n'induit pas l'expression des mêmes sets de gènes (Pelletier et al., 2017), montrant qu'il joue probablement plusieurs rôles dans la maturation de la graine.

Outre les effets sur l'accumulation de réserves et l'induction de la maturation, la surexpression des *LAFL*, en particulier *LEC1* et *LEC2* est suffisante pour induire la formation d'embryons somatiques de manière spontanée (Figure Intro-23) (Lotan et al., 1998; Stone et al., 2001). Cet aspect est développé dans la 4^{ème} partie de cette introduction – [Revue : « Contrôle génétique et moléculaire de l'embryogenèse somatique »](#). Ainsi, en plus de leur rôle dans la régulation de gènes associés aux processus de maturation, les *LAFL* agissent également sur des cibles liées à l'identité cellulaire, au maintien de la capacité embryogène ou encore à la signalisation hormonale, que ce soit pendant l'embryogenèse zygotique ou somatique (Braybrook and Harada, 2008; Stone et al., 2008; Jia et al., 2014; Pelletier et al., 2017). En conséquence, lorsque les *LAFL* sont mutés, les plantes perdent leur capacité embryogène, comme cela a été montré pour *lec2* (Gaj et al., 2005).

Les *LAFL* sont donc des régulateurs maîtres du développement de la graine jouant un rôle majeur au cours de la maturation, mais également lors de l'initiation et du maintien du caractère embryogène des cellules.

3.3.3. Régulations exercées sur les LAFL

Par le rôle clé des *LAFL* dans le développement de la graine et l'effet de leur surexpression capable d'induire l'embryogenèse, leur expression doit être finement régulée de manière spatio-temporelle. Les gènes de la famille des *LAFL* sont logiquement soumis à un fort contrôle transcriptionnel et post-transcriptionnel (Lepiniec et al., 2018). Ce chapitre décrit les principales régulations connues.

3.3.3.1. Régulations transcriptionnelles

Des analyses de relations structure-fonction réalisées avec le gène *LEC2* ont montré que les trois principaux éléments régulateurs présents sur son promoteur et requis pour sa régulation (positive comme négative) sont situés dans les 450 pb en amont du codon de démarrage de la traduction (Berger et al., 2011). Les éléments cis-régulateurs CARG reconnus par des facteurs de transcription de type MADS-box ainsi que les boîtes GAGA liées par des protéines BASIC PENTACYSTEINE (BCP), sont essentiels pour l'activation transcriptionnelle de *LEC2*. Or, si une partie de la séquence

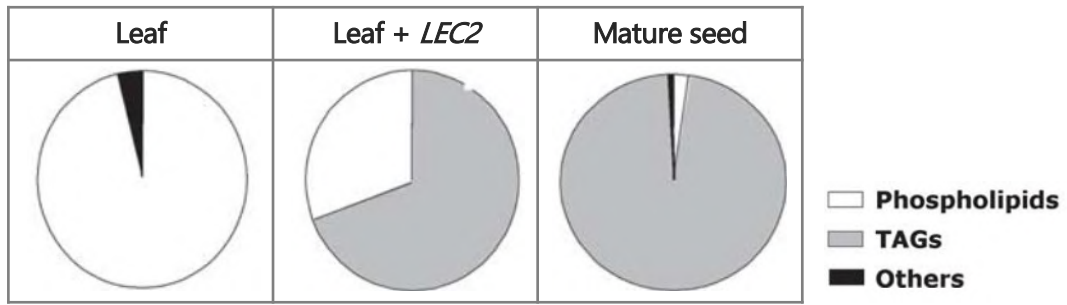


Figure Intro-22 : L'expression ectopique de *LEC2* dans les feuilles induit l'accumulation de lipides de réserves spécifiques des graines. Les feuilles sont normalement composées majoritairement de phospholipides et les graines de triacylglycérols. Les feuilles exprimant *LEC2* montrent une accumulation anormale de triacylglycérols.

Abréviations : TAGs, triacylglycérols.

Modifié d'après Mendoza et al., 2005.

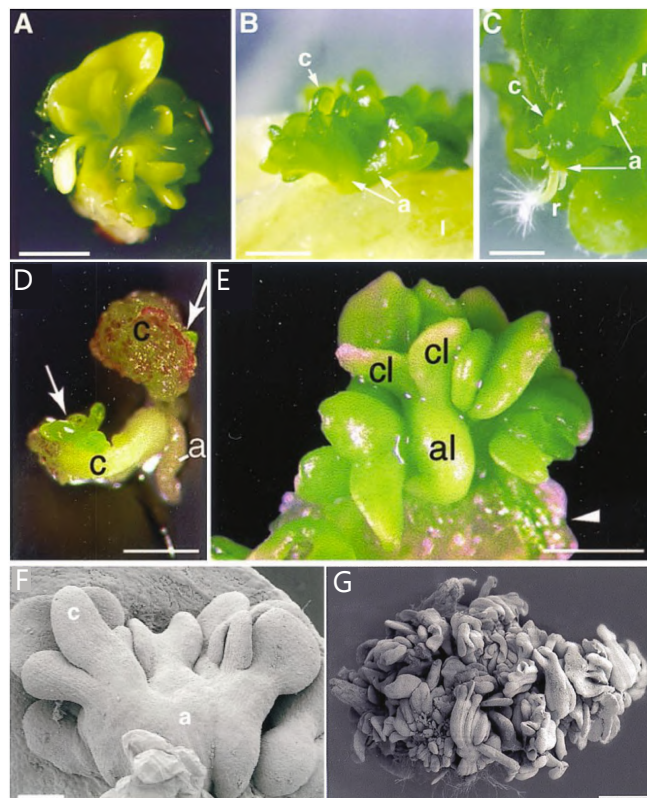


Figure Intro-23 : La surexpression de *LEC1* ou *LEC2* conduit à la formation spontanée d'embryons somatiques sur les plantules. Phénotypes surexprimeurs de A-C) *LEC1*, D-E) *LEC2*. Observation en microscopie électronique à balayage de plantules surexprimant F) *LEC1* et G) *LEC2*.

Barres d'échelle : A et C, 0,5 mm ; B, 0,1 mm ; D, E et G, 1 mm ; F, 0,05 mm.

Abréviations : a/al, axe ; c/cl, cotylédon ; r, racine.

Modifié d'après Lotan et al., 1998 (A-C, F) et Stone et al., 2001 (D-E, G)

promotrice est nécessaire à son expression, il a été montré qu'un élément cis-répresseur *RLE* (*cis-repressive LEC2 element*) est situé sur le promoteur entre 150 et 200 pb en amont du codon START de transcription et est nécessaire et suffisant pour la reconnaissance et le dépôt de marques répressives par les complexes PRC, conduisant à la répression transcriptionnelle du gène (Berger et al., 2011). *RLE* a été le premier *Polycomb Responsive Element* (PRE) putatif à être identifié chez les plantes. Il serait responsable de la répression de *LEC2* après le stade embryonnaire. Sans le *RLE*, l'expression de *LEC2* devient quasiment constitutive dans les parties végétatives. En plus des régulations exercées sur le *RLE* présent dans la séquence promotrice, d'autres acteurs épigénétiques induisent la répression ou l'activation de *LEC2* et des autres *LAFI*. En effet, les FT à domaine B3 VAL1/HSI2 et VAL2/HSL1, ainsi que le remodeleur de chromatine PKL, sont capables d'induire une répression chromatinienne de l'expression des *LAFI* en induisant le dépôt de marques répressives H3K27me3 pendant la germination (Suzuki et al., 2007; Jia et al., 2014). Lorsque ces gènes sont mutés, l'expression des *LAFI* est induite et l'efficacité de l'embryogenèse somatique des plantes est plus importante (Horstman et al., 2017). Comme vu précédemment, des modifications de la chromatine et de la méthylation de l'ADN interviennent lors du développement et de la germination de la graine (Bouyer et al., 2017). Même si les changements de méthylation observés dans des contextes non-CG ne semblent pas impacter le développement de la graine (Lin et al., 2017), il apparaît que les *LAFI* sont soumis à d'importants changements de leur état chromatinien, notamment lors de la germination (Müller et al., 2012). En effet, il a été montré que des marques épigénétiques déposées sur les *loci ABI3* ou *LEC2* par les complexes PRC1/PRC2 conduisent à l'inhibition de leur expression pendant cette phase. Dans les embryons en maturation, *ABI3* et *LEC2* sont actifs et associés à H3K4me3, alors qu'au cours de la germination, cette marque est remplacée par H3K27me3, qui est une marque de répression (Müller et al., 2012; Molitor et al., 2014). Le recrutement de PRC2 sur des (PRE) induirait le dépôt de marques répressives H3K27me3. Il a aussi été montré que ces facteurs de transcription sont régulés par des petits ARN non codants (miRNA) qui font partie intégrante de la boucle d'auto-régulation de ce groupe de FT (Lepiniec et al., 2018).

3.3.3.2. Régulations post-traductionnelles

Les modifications post-traductionnelles sont un moyen supplémentaire de réguler l'activité globale d'une protéine chez les organismes vivants et une façon de s'adapter aux changements environnementaux. Les modifications post-traductionnelles peuvent affecter l'activité spécifique d'une protéine, sa stabilité, sa localisation ou encore ses interactions avec d'autres protéines (Vu et al., 2018). Une protéine codée par un seul gène peut prendre différentes formes en fonction de l'épissage alternatif lors de la transcription/traduction ou de modifications post-traductionnelles

auxquelles elle est soumise, on parle alors d'isoformes protéiques, pouvant avoir des fonctions ou des activités distinctes (Smith et al., 2013). Si de nombreuses régulations transcriptionnelles ont été identifiées pour l'expression des *LAFL*, les régulations post-traductionnelles qui s'exercent sur ces régulateurs maîtres sont mal connues. Il a été montré que les protéines ABI3 et FUS3 ont un temps de demi-vie assez court : ABI3 est ubiquitinée et dégradée par le protéasome (Zhang et al., 2005) alors que FUS3 contient dans sa partie C-terminale un domaine PEST (séquence peptidique riche en prolines, acides glutamiques, serines et thréonines) servant de signal protéolytique nécessaire à sa dégradation par le protéasome (Lu et al., 2010; Rechsteiner and Rogers, 1996; Maruri-López et al., 2014). Des analyses de prédiction permettent de supposer que LEC2 est également ciblé par de l'ubiquitination et deux domaines PEST sont présents dans sa séquence (<http://emboss.bioinformatics.nl/cgi-bin/emboss/pepfind>). Il a été montré que chez le colza, l'UBIQUITIN-PROTEIN LIGASE 3 (BnUPL3) agit sur l'ubiquitination de LEC2 ainsi que sa dégradation par le protéasome (Miller et al., 2019). BnUPL3 est un orthologue d'AtUPL3, qui est décrit comme permettant une dégradation par le protéasome après ubiquitination de ses cibles via leur domaine de localisation nucléaire (Downes et al., 2003). Le mutant *upl3* chez *Arabidopsis* a un niveau d'accumulation de LEC2 plus important conduisant à une expression prolongée des gènes de la voie de biosynthèse des lipides au cours de la maturation, une plus forte quantité de lipides dans les graines, qui sont également plus grosses. Cela montre que l'activité de LEC2 comme régulateur de l'accumulation de réserves n'est pas uniquement liée à son expression ou son absence d'expression, mais aussi à son niveau d'accumulation.

Finalement, en plus des nombreuses régulations hormonales et transcriptionnelles exercées sur les *LAFL* pour contrôler leur expression, il apparaît que leur accumulation dans les cellules soit également contrôlée et des moyens sont mis en œuvre pour dégrader ces protéines et/ou réguler leur accumulation dans les cellules. Il semble en effet que le niveau d'expression et d'accumulation de ces régulateurs maîtres soit très important pour leur fonction, puisque lorsqu'ils sont trop exprimés, ils induisent des phénotypes forts, délétères pour la plante dont le développement végétatif est alors altéré. Il paraît donc important de comprendre exactement la structure, la fonction et les régulations des *LAFL*, si l'on veut améliorer la qualité des graines, par exemple, au niveau de la maturation et du remplissage chez les espèces d'intérêt agronomique.

4. Revue : « Contrôle génétique et moléculaire de l'embryogenèse somatique »

Comme évoqué précédemment, les plantes sont capables de se reproduire de manière sexuée et asexuée. Les mécanismes génétiques et moléculaires régulant ces différents types d'embryogenèse peuvent être communs. Les LAFL, en plus de leur rôle maître dans la régulation du remplissage de la graine jouent un rôle majeur dans la mise en place, l'initiation et le maintien du caractère embryogène des cellules végétales, bien qu'on ne sache pas par quel(s) mécanisme(s) ces facteurs agissent sur des processus différents. Nous avons rédigé un article de revue ayant pour but de donner un aperçu des mécanismes génétiques et moléculaires connus qui gouvernent les processus d'embryogenèse somatique et zygotique, avec un focus particulier sur le rôle des LAFL.

Salaün, C., Lepiniec, L., and Dubreucq, B. (2021). Genetic and Molecular Control of Somatic Embryogenesis. *Plants* 10: 1467.

Résumé : L'embryogenèse somatique est une méthode de reproduction asexuée qui peut avoir lieu naturellement chez différentes espèces végétales et qui est utilisée pour la propagation clonale, la transformation et la régénération de différentes espèces d'intérêt agronomique. L'embryogenèse somatique partage des similarités avec l'embryogenèse zygotique en termes de développement et de physiologie, car elle fait intervenir des acteurs communs du contrôle hormonal, transcriptionnel, développemental et épigénétique. Ici, nous donnons une vue d'ensemble des principaux réseaux de signalisation impliqués dans l'induction et la régulation de l'embryogenèse somatique en mettant l'accent sur les régulateurs maîtres du développement de la graine, les facteurs de transcription LEAFY COTYLEDON 1 et 2, ABSCISIC ACID INSENSITIVE 3 et FUSCA 3 dont le rôle précis durant l'embryogenèse zygotique et somatique reste à élucider.

Genetic and Molecular Control of Somatic Embryogenesis

Camille Salaün, Loïc Lepiniec  and Bertrand Dubreucq *

Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000 Versailles, France; camille.salaun@inrae.fr (C.S.); loic.lepiniec@inrae.fr (L.L.)

* Correspondence: bertrand.dubreucq@inrae.fr

Abstract: Somatic embryogenesis is a method of asexual reproduction that can occur naturally in various plant species and is widely used for clonal propagation, transformation and regeneration of different crops. Somatic embryogenesis shares some developmental and physiological similarities with zygotic embryogenesis as it involves common actors of hormonal, transcriptional, developmental and epigenetic controls. Here, we provide an overview of the main signaling pathways involved in the induction and regulation of somatic embryogenesis with a focus on the master regulators of seed development, LEAFY COTYLEDON 1 and 2, ABSCISIC ACID INSENSITIVE 3 and FUSCA 3 transcription factors whose precise role during both zygotic and somatic embryogenesis remains to be fully elucidated.

Keywords: embryogenesis; LAFL; transcriptional regulation; epigenetic control



Citation: Salaün, C.; Lepiniec, L.; Dubreucq, B. Genetic and Molecular Control of Somatic Embryogenesis. *Plants* **2021**, *10*, 1467. <https://doi.org/10.3390/plants10071467>

Academic Editors: Paloma Moncaleán and Itziar A. Montalbán

Received: 25 June 2021
Accepted: 13 July 2021
Published: 17 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Plants can propagate through sexual or asexual reproduction. In Angiosperms, zygotic or sexual reproduction is the most common and consists of the double fertilization of a female gametophyte (that contains the egg and the central cell) by the two sperm cells of a male gametophyte (pollen). Sexual reproduction leads to the formation of a seed that contains all the components required for seedling development: the differentiated zygotic embryo which will give rise to the future plant, storage compounds located in the cotyledons, endosperm and/or nucellus depending on the species, all surrounded by a protective tissue, the testa (formed by ovary integuments) [1]. Through meiosis and crosses, zygotic reproduction allows genetic diversity in the progeny and its propagation through seeds, that can be transported through wind, water (hydrochory) or animals (zoochory) [2]. However, plants are also able to reproduce in an asexual way, i.e., without meiosis and fertilization. For example, apomixis consists in the formation of a seed derived from a diploid cell of the ovule and can occur naturally in some species [3]. Vegetative reproduction is also widespread and used in horticulture, with different techniques such as layering, cuttings, and grafting.

This study focuses on somatic embryogenesis (SE), a complex process of clonal propagation by which plants can form embryos without meiosis and fertilization. This process involves the totipotency of plant cells, i.e., their ability for dedifferentiation and differentiation in new cellular types. The new plant derived from a somatic embryo is thus genetically identical to the mother plant. SE can occur spontaneously or in response to specific environmental conditions in some species such as *Kalanchoe* [4,5].

Understanding the mechanisms controlling somatic embryogenesis and its regulation is a key issue in plant biology, since clonal propagation through SE is widely used for various plants of agronomic interest, such as *Coffea* spp. [6], *Pinus* spp. [7], *Theobroma cacao* L. [8] and other species, as it facilitates the clonal multiplication of elite genotypes in a quite fast and efficient way [9]. Moreover, SE is also largely used for genetic transformation protocols of these agronomical species, and for other crops such as rice, soybean, maize, or wheat [10–12]. Because the efficiency of transformation is correlated with the embryogenic and regeneration capacity, comprehension of SE processes is of great importance.

However, somatic embryogenesis involves complex, or even not characterized, genetic inheritance [13]. Some species or specific genotypes can display a recalcitrance to SE and are unable to undergo it even under favorable conditions. Lastly, although somatic and zygotic embryogenesis (ZE) share some similar hormonal, transcriptional or epigenetic controls [14], they also display some specificities. In this review, the known regulatory mechanisms involved in SE and ZE are compared with a specific focus on the *LAFL* (*LEC1*, *ABI3*, *FUS3* and *LEC2*) gene regulatory network.

2. Onset of Somatic Embryogenesis

Somatic embryogenesis relies on the totipotency of plant cells, i.e., their capacity to dedifferentiate and differentiate in a new cell type [15,16]. Although some plant species such as *Kalanchoe daigremontiana* undergo spontaneous SE [4], this process is usually induced in vitro by a stress over the plant tissues. This stress, that is essential for SE induction, can take different forms (reviewed in [17]): a high level of plant hormones in the culture media is the most commonly used [18,19], but it also can be wounding, extreme pH, osmotic or heat shock, or treatments with different chemicals. The first stages of SE are characterized by the induction of the expression of numerous stress-related genes and especially those encoding transcription factors belonging to AP2/ERF, MYB, AUX/IAA, B3 or WUS/WOX families [19–21].

Somatic embryos can be induced directly from in vitro cultivated plant tissues (for example immature zygotic embryos) on a low-auxin medium. Alternatively, indirect SE can be induced by cultivating embryogenic tissues, such as callus, on an auxin-rich medium leading to the transition to somatic embryos by switching to a low-auxin medium [22,23]. The protocols used for industrial clonal reproduction or for the research on somatic embryos are usually based on indirect SE, since it allows the production of a large number of somatic embryos [22–24] and because direct SE can be ineffective for some species or genotypes, such as *Gossypium hirsutum* [25], *Capsicum chinense* Jacq., or *Cocos nucifera* L. [26].

3. Hormonal Control of Somatic Embryogenesis

Multiple events in the plant life cycle, such as development or responses to biotic and abiotic stresses, are controlled or regulated by hormones. Auxins are known to play essential roles in all the aspects related to plant development, from cellular division to elongation, through cell identity and organogenesis [27,28]. Auxin polar transport by the proteins PIN (efflux) and AUX (influx) leads to differential distributions, with gradients and maxima that are essential for plant development [29]. During zygotic embryo development, local auxin production, polar transport and gradient are mainly mediated by PIN proteins [30]. More specifically, in Arabidopsis, the expression of *PIN1* and *PIN7* genes in the apical part of the young embryo is at the basis of the acquisition of the apical-basal orientation [30,31], in concert with *AUX1*, *LAX1* and *LAX2*, playing a cooperative role [32].

Not surprisingly, auxin was thus shown to be a central player in the induction of SE as reviewed in [33]. Briefly, auxins and more specifically 2, 4-dichlorophenoxyacetic acid (2, 4-D), a synthetic auxin-like plant growth regulator, are largely used for inducing SE [19]. The treatment of plant tissues with 2, 4-D involves both a stress and the auxin response, biosynthesis and signaling pathway, as well as a transcriptional reprogramming and chromatin remodeling (as described below) [33]. Similar to zygotic embryogenesis, a polar transport and auxin gradient are necessary for the formation of a somatic embryo [34]. SE induction, as well as organogenesis and plant regeneration, depends on the addition of plant growth regulators (PGR) such as auxins and cytokinins [18]. Although many protocols adapted to different species and genotypes are available, a key element for embryo formation is the large variation in PGR concentration added to the culture media [24]. A majority of protocols are based on the use of auxins, alone or combined with cytokinins [24]. During SE induction, auxin response factor (*ARF*) genes display specific expressions being up- or down-regulated, suggesting that auxin signaling is central in the

process [35]. Moreover, *YUCCA* and *AUX/IAA* genes that are respectively involved in auxin biosynthesis and response, are transcriptionally regulated during SE, including by LAFL transcription factors [36,37].

Growing plant tissues *in vitro* with high concentrations of auxin (2, 4-D), triggers a general reprogramming of somatic cell transcriptomes and modulates the expression of many SE-associated transcription factors [33]. Moreover, when protoplasts are cultured with a medium containing high concentrations of 2, 4-D, the size of their nuclei is significantly increased, suggesting a reorganization of the chromatin [38]. Finally, hormonal stress induces a modification of chromatin state, leading to the activation of transcription factors such as *WUS*, *LEC* genes or *BBM* that are specific to embryogenic programs [39]. These results suggest that hormones, especially auxin, trigger a general reprogramming of gene expression through chromatin modifications and activation of specific transcription factors.

4. Transcriptional Control of Somatic Embryogenesis

The main way to regulate gene expression is the transcriptional initiation through transcription factors (TF). Transcriptional regulation performs an essential role in somatic embryogenesis [40]. In *Arabidopsis*, the ectopic expression of some transcription factors, such as *LEC* genes, *BBM*, *WUS/WOX* genes or *AGL15* can increase the efficiency of SE induction and lead to the formation of somatic embryos without adding hormones [23]. Mutations in these genes have a negative impact on the efficiency of SE induction [1,41,42]. Several of the TF are involved in hormonal signaling, while others control cellular differentiation and organogenesis. These transcription factors are described below.

4.1. LAFL (*LEC1*, *LEC2*, *ABI3*, *FUS3*)

ABSCISIC ACID INSENSITIVE 3 (*ABI3*), *FUSCA 3* (*FUS3*) and *LEAFY COTYLEDON 2* (*LEC2*) are transcription factors forming the AFL group. They belong to the plant-specific B3 family of transcription factors, characterized by a highly conserved B3 DNA binding domain of ~110 amino-acid residues first characterized in *ABI3/VP1* [43,44]. B3-containing factors can recognize specific target DNA with a 5'-GATC-3' core sequence, known as *RY* boxes and conserved within the promoters of storage genes [45,46]. Together with *LEAFY COTYLEDON 1* (*LEC1*), a NF-YB transcription factor [1] involved in CAAT box complexes [47], AFL form the LAFL group of master regulators of seed development [48–50]. They can genetically and/or physically interact altogether and form protein complexes to activate their target genes [46,51] and control various aspects of seed development, such as the accumulation of storage compounds or the acquisition of desiccation tolerance (Figure 1). The expression of LAFL genes is regulated by transcription factors, such as *BBM* [52], hormonal signaling (ABA, GA, IAA) [53] and chromatin modifications [52].

These genes are involved in somatic embryogenesis programs as well as in the initiation and maintaining of the embryogenic fate of plant cells [54] (Figure 1). Indeed, the ectopic expression of *LEC* genes (*LEC1*, *LEC2*, and *FUS3*) in *Arabidopsis* and in other species can induce the spontaneous formation of somatic embryos without addition of hormones in the culture medium [42,55–59]. The expression of *GhLEC1*, *GhLEC2* and *GhFUS3* is significantly higher in cotton (*Gossypium hirsutum*) genotypes permissive to callus differentiation, than in recalcitrant genotypes [60]. Consistent with this role, *lec1*, *lec2* and *fus3* simple and multiple mutations severely impair the ability to form somatic embryos in *Arabidopsis* [60].

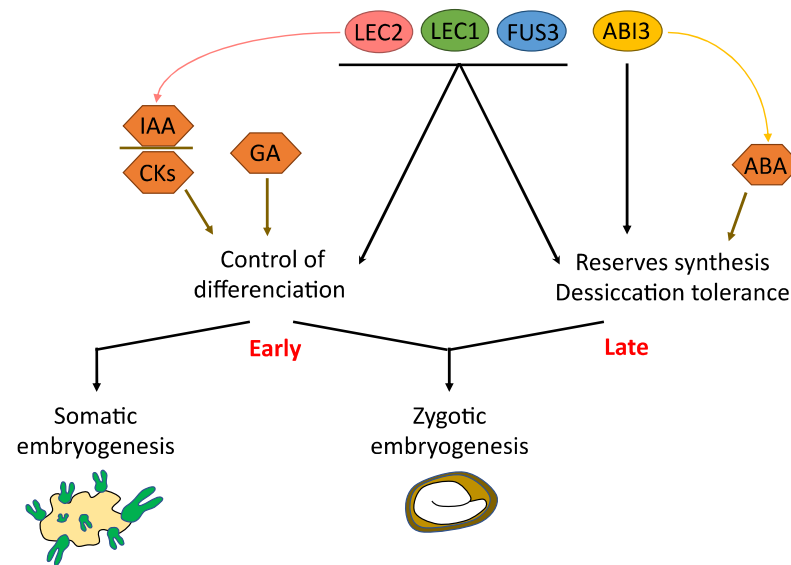


Figure 1. The roles of *LAFL* genes in somatic embryogenesis (SE) and zygotic embryogenesis (ZE). IAA: auxins, CKs: cytokinins, GA: gibberellins.

Several target genes of the LEC proteins can be directly involved in the regulation of somatic embryogenesis [33,61], such as *AGL15*, that is involved in hormonal signaling and controls the embryogenic induction [62]; *IAA30*, a main actor of auxin signaling and perception [63] or *LOB40*, a TF putatively involved in the formation of organ boundaries and gibberellins signaling [64]. LEC2 and *AGL15* control each other in a feedback regulatory loop [33]. LEC2 is also known to rapidly induce the expression of auxin-related genes, such as *IAA1*, *IAA17* and *ACS4*, and key enzymes involved in auxin biosynthesis such as the *YUC* genes (*YUC1*, *YUC2*, *YUC4*, *YUC10*) [37,65]. However, among these auxin genes, only *YUC4* is a direct target of LEC2. Then, it can be hypothesized that LEC2 (and LEC1) can induce SE processes through different mechanisms (Figure 2). The LEC proteins activate the maturation program, leading to the increase in various storage compounds and the decrease in water content. These physiological changes would trigger a stress inducing SE competency and/or promotion of auxin activity in competent cells [65]. In addition, some genes usually expressed during zygotic embryo development and known to bring an embryogenic competence, such as *AGL15* for example, are directly induced by LEC2 [65]. Another hypothesis can be a dual function of LEC2, able to act as a regulator of seed maturation by regulating specific targets in complex with specific partners, and as a regulator of embryogenesis (both ZE and SE), by regulating other targets, in cooperation with other partners (Figure 2).

ABI3 is involved in ABA-mediated developmental and metabolic processes, such as seed maturation, accumulation of seed proteins or the transition between embryo maturation and early seedling development [66]. *ABI3* is essential for the emergence of roots from callus after induction of organogenesis. The *abi3-6* mutant does not show any regeneration after passing from a hormone-rich to a hormone-free medium [67]. Nevertheless, the over-expression of *ABI3* does not induce somatic embryogenesis [68,69]. Because the expression of auxin-related genes is significantly weaker in the *abi3-6* mutant, *ABI3* can be involved in auxin distribution and/or homeostasis, with its role in ABA signal transduction and reserve accumulation, rather than in the initiation of SE as other LAFL.

Although very similar in their B3 sequences and binding capabilities [46], LAFL are thus involved to a different extent in somatic embryogenesis (Figures 1 and 2). This is consistent with their different DNA binding activities [46] and their specific expression patterns.

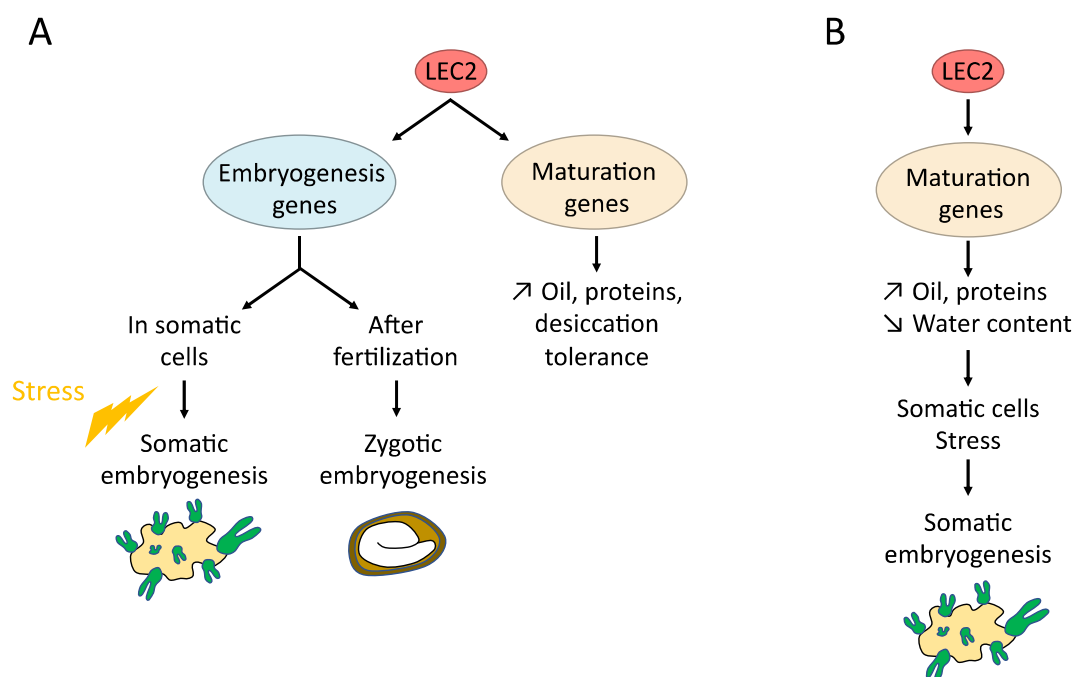


Figure 2. The possible roles of LEC2 in somatic embryogenesis. (A): LEC2 can have a dual function, one specific to embryogenesis, the other inducing maturation genes. The embryogenesis function may lead to SE or ZE depending on the type of cell it is expressed in, as well as possible stress and epigenetic modifications. (B): LEC2 induces stress through the accumulation of reserve specific compounds in the cells.

4.2. BABY BOOM (BBM)

BABY BOOM (BBM), also known as PLETHORA 4 (PLT4) is a transcription factor belonging to the AP2/ERF TF family characterized by a DNA binding domain of 70 amino-acid residues, that were first characterized in APETALA2 (AP2) and ethylene-responsive element binding proteins (EREBP) [70,71]. The AP2/ERF family is divided into two sub-families: TFs that count one AP2/ERF DNA binding domain and are generally involved in response to abiotic and environmental stresses [72], and those counting two DNA binding domains, known to regulate developmental processes. BBM contains two AP2/ERF domains and is a member of the AINTEGUMENTA-LIKE (AIL) subfamily, composed of eight members, all carrying specific functions in the division of meristematic and embryonic cells and involved in the regulation of SE and its induction when they are ectopically expressed [33]. *BBM* is expressed in the zygotic embryo and regulates cell identity and the growth of the root meristem [73]. The ectopic expression of *BBM* in *Arabidopsis thaliana* and *Brassica napus* leads to the spontaneous formation of somatic embryos and cotyledon-like structures over plantlets [70]. However, other phenotypes are also induced by ectopic expression of *BBM* in these species, such as callus proliferation, ectopic shoot formation, alterations in leaf morphology and better regeneration of explants during in vitro culture, without adding hormones in the medium. *BBM* can thus have a stimulating role in cell proliferation and morphogenesis during both SE and ZE. *BBM* directly binds promoter regions of *LAF1* genes through its ANT/AIL binding motif and *AGL15* via another motif or an intermediate protein [52]. *LAF1* genes are upregulated after *BBM* induction, and together with *AGL15*, are *BBM*-positive regulators during SE processes and are key components of the *BBM* signaling pathway, although these interactions may be indirect.

4.3. WUSCHEL (WUS) and WUSCHEL-RELATED HOMEODOMAIN (WOX) Genes

WUSCHEL (WUS) is a homeobox transcription factor that is described as a key regulator of meristematic stem cell fate [74]. It also performs an essential role in somatic embryogenesis, by promoting the vegetative-to-embryonic transition and the maintenance

of embryogenic cell identity [75]. WUS is involved in the regulation of the embryogenic cell (totipotency) and meristematic cell (pluripotency) fates [76]. In *Arabidopsis*, *wus* mutants can form somatic embryos, although seedlings display the same phenotype as seed-derived *wus* plantlets, i.e., an absence of functional embryonic shoot apical meristem (SAM), and the typical *wuschel* stop-and-go development described in [74] consisting, among others, in the formation of new shoot meristems instead of organs at the primordia [77]. The overexpression of *WUS* and gain of function mutations can increase somatic embryo production or induce the formation of somatic embryos from vegetative tissues without addition of external hormones in *Arabidopsis thaliana* [75]. In many other species, such as *Gossypium hirsutum* [78] or *Coffea canephora* [79], the overexpression of *WUS* is also able to significantly increase the ability of a plant to form somatic embryos. In *Gossypium hirsutum*, the ectopic expression of *AtWUS*, increases the efficiency of callus differentiation of recalcitrant genotypes [60]. *WUS* can also be used as a marker gene of dedifferentiation after SE induction in *Medicago truncatula* [80]. During SE, *WUS* is upregulated and transcriptionally induces *LEC1*, *LEC2*, and *AGL15* genes [76]. Thus, we can hypothesize that SE originates from a cellular reprogramming followed by a process similar to ZE, at least in the signaling cascade of some key genes. When some master genes of ZE are knockdown (including *LAFL*), SE is impaired, favoring such a hypothesis. Two fully independent pathways exist, yielding to similar developmental structures. This hypothesis can be tested by comparing SE efficiency of the double mutants (e.g., *wus/lec2*) or double ectopic expressions to the single mutations/ectopic expression.

The *WUSCHEL-RELATED HOMEODOMAIN* (*WOX*) genes share similar sequences with the *WUS* homeodomain and a specific *WUS* box located downstream of the homeodomain [81]. *WOX* genes are involved in the early embryonic patterning but also in different signaling pathways regulating several aspects of plant growth and induction of somatic embryogenesis [76]. Functions of the different known *WOX* genes in several species are reviewed in [76]. As an example, it was shown in *Medicago truncatula*, that *MtWOX9-1* overexpression improves SE efficiency, linked to an increase in *AGL15* and *AGL8* level of accumulation [82] and *WOX5* is significantly upregulated two days after SE induction and can serve, such as *WUS*, as a marker of dedifferentiation [80].

4.4. *AGAMOUS-LIKE 15 (AGL15)*

AGAMOUS-LIKE 15 (AGL15) is a transcription factor, belonging to the MADS domain protein family. These TFs are known to share a specific MADS DNA binding domain, containing a 55 to 60 amino-acid sequence conserved among eukaryotes [83]. *AGL* genes display diverse functions in plant development, more particularly in flower development, although *AGL15* is mainly detected in developing embryos [84,85]. *AGL15* can target and regulate the expression of numerous hormone-related genes, including gibberellic acid and ethylene metabolism but also auxin signaling via *GA2ox6* [86] and *IAA30* activation [62,86]. *AGL15* activates its own expression [87] and regulates *LEC2*, *ABI3* and *FUS3* in a positive feedback regulatory loop [36]. The ectopic expression of *AGL15* in *Arabidopsis thaliana* and in soybean increases SE capacity [88,89]. While the efficiency of SE plant tissue can decrease with time, overexpression of *AGL15* maintains SE capacity for years, showing the role of *AGL15* in the preservation of embryogenic capability [88]. The loss-of-function of *AGL15* reduces the frequency of somatic embryo development [89].

4.5. *WOUND INDUCED DEDIFFERENTIATION 1 (WIND1)*

WOUND INDUCED DIFFERENTIATION 1 (WIND1) is an AP2/ERF transcription factor with two AP2 domains involved in callus formation in response to wounding and performs a critical role in the acquisition of regeneration competency and regulating cytokinin signaling [90,91]. The over- and ectopic expression of *WIND1* leads to the callus formation on explants without induction through the addition of hormones in the culture medium or wounding [91]. *WIND1* is not directly involved in the induction of SE, but in the induction of callus formation and regeneration competency. However, the induction of

WIND1 expression followed by that of *LEC2* leads to the formation of callus within all the plant tissues that can be regenerated in a whole plant, showing that *WIND1* acts upstream from *LEC2* during SE and regeneration processes [92].

Through the example of several transcription factors, we have seen that transcriptional control performs a key role in the initiation and efficiency of somatic embryogenesis. Aside from their role in the transcriptional activation of their targets through the promoting regions, some eukaryotic TFs can activate their targets even if they are located in nucleosomal DNA that is less accessible. These specific TFs, called pioneer transcription factors, are usually submitted to strong (spatiotemporal) transcriptional or post-transcriptional controls and can act directly or indirectly by recruiting chromatin remodelers, thereby increasing local chromatin accessibility [93]. Pioneer factors are of primary importance in development since they allow the chromatin to be open and accessible for other factors to bind, and thus, are thought to enable cellular reprogramming [94]. This is the case for *LEC1*, that is a pioneer transcription factor capable of inducing chromatin modifications to different target genes: its overexpression leads to the spontaneous formation of somatic embryos on plantlets and its loss-of-function mutants are not able to do SE anymore [54,57,95–97].

5. Epigenetic Regulation of Somatic Embryogenesis

Although chromatin regulations are essential in developmental processes, relatively little data concerning the epigenetic regulation of somatic embryogenesis have been published (see [98] for review). DNA methylation profile and chromatin accessibility are constantly changing in response to cellular needs and environmental constraints [99]. The different abiotic stresses to which plants are subjected can either increase or reduce DNA methylation, depending on the type of stress (i.e., heavy metals, salt, temperature, culture density), the plant species or even the organ considered [100,101]. It should be noted that epigenetic changes induced can be transient or stable over a period of time, and can also be transmitted to the offspring [102]. Then, it can easily be hypothesized that epigenetic modifications correlate with SE processes [98]. In particular, the efficiency of cellular differentiation is linked to the methylation profile of DNA. The cellular reprogramming is often concomitant with significant changes in the chromatin status: not only DNA methylation but also histone modifications including methylation or acetylation [103,104]. Overall, DNA hypomethylation or histone acetylation are associated with transcriptional activation of regulatory genes that control development or hormone responses responsible for the totipotent state of the cells [39]. For example, in cotton, the level of CHH demethylation (one of the preferential DNA methylation contexts, with CG and CHG, where H can be an A, T or G) is a marker of somatic embryo differentiation [25]. By comparing SE recalcitrant *vs.* permissive genotypes, a hypermethylation of CHH is found, suggesting a repressive effect of methylation over SE-related genes expression, explaining why SE is less efficient for tissues with hypermethylated genomes [25]. Moreover, the DNA methylation level in callus decreases during embryogenesis and the important changes observed in the transcript levels during callus induction and somatic embryogenesis reflect the epigenetic reprogramming [39]. The use of deacetylation or demethylation inhibitors to promote the induction of somatic embryogenesis show that the chromatin profile of the tissue is a determining factor for SE induction, providing the evidence that chromatin regulations are directly involved in cell reprogramming and callus formation during SE [39,105–107].

POLYCOMB REPRESSIVE COMPLEXES (PRC) 1 and 2 are required for establishing and maintaining a stable epigenetic repression in response to developmental or environmental signals [108]. Briefly, PRC2 has a histone 3 lysine 27 trimethylation (H3K27me₃) activity and PRC1 can recognize H3K27me₃ and lead to chromatin compaction through histone H2A lysine ubiquitination (H2Aub) [49]. PRC2 is a main factor involved in cell identity to maintain a stable repression of embryogenesis program genes. PRC1 and PRC2 can repress the expression of several genes involved in cellular differentiation and callus development as well as somatic embryo formation including *WOX5*, *WOX8*, *AGL15*, *LEC1*, *LEC2*, *ABI3*, *FUS3*, *BBM*, *PIN1* or *PIN2* [109]. Consistent with these regulations, mutations

in PRC2 subunits can lead to the formation of callus or abnormal development of vegetative tissues that resemble SE [110–112].

PICKLE genes (*PKL* and *PIKLE RELATED 1* and *2* – *PKR1* and *2*) are members of the CHD3 family of chromatin ATPases remodelers [113] and *VAL1* and *2* (*VIVIPAROUS/ABI3-LIKE*) are transcriptional regulators containing DNA and chromatin binding domains [114]. Both *PKL* and *VAL* have a repressive effect on *LAFL* genes expression during zygotic embryo development [49,113–115]. In *pkl* and *val1* mutants, the chromatin-based repression of *LAFL* is less important and consequently, the SE capacity is increased [52].

Mutations in genes involved in chromatin modifications, such as the DNA METHYLTRANSFERASE 1 (*MET1*) [116], KRYPTONITE (*KYP*) – an H2K9 METHYLTRANSFERASE [117], JUMONJI 14 (*JMJ14*) – an H3K4 DEMETHYLASE [118] and the HISTONE ACETYLTRANSFERASE 1 (*HAC1*) [119] lead to an alteration of *WUS* expression and a poorer formation of seedlings de novo after regeneration. Moreover, the functional analysis of *met1* showed that DNA methylation and histone modification control de novo regeneration through *WUS* expression and auxin signaling [120].

Among chromatin-modification proteins involved in SE processes *AHL15*, an Arabidopsis AT-hook binding domain transcription factor that is important for chromatin opening, induces SE in absence of auxin treatment when it is overexpressed [121]. *AHL15* and its homologs are positively regulated during hormone-mediated induction of SE and are required for the induction through *BBM* overexpression. *AHL15* has a role in the level of heterochromatin in somatic cells: in loss-of-function mutants, the amount of heterochromatin is higher in in vitro culture and the capacity of cells to form somatic embryos after 2, 4-D induction is altered.

miRNAs are also important during plant development because they target regulatory genes including transcription factors and F-box proteins [122,123]. *DICER-LIKE 1* (*DCL1*), is a RNase III-like enzyme responsible for the biosynthesis of most plant miRNAs [124]. The *dcl1* mutant, which is affected in the production of miRNAs, is not capable of inducing somatic embryogenesis [125]. This suggests that some miRNAs are crucial for SE induction processes, including auxin-related miRNAs that target genes of auxin perception, biosynthesis and signaling, for example: miR165/166, miR167, miR164, miR390 and miR393 [126].

In a recent study based on genome-wide analyses (ATAC-seq, ChIP-seq and RNA-seq), the authors observed a hierarchical organization of the transcriptional and chromatin mechanisms that regulate cellular reprogramming during plant somatic embryogenesis [127]. Their results suggest that the developmental stage of the explant used for tissue culture is at the top of the controlling hierarchy, as it determines the chromatin landscape. In fact, after germination, some specific chromatin permissive marks become inaccessible. This is the case for *AP2*, *B3* and *NF-Y* genes that encode the key TFs inducing somatic embryogenesis. After the developmental stage of the explant, the second level of this hierarchical organization is auxins that induce important changes in chromatin accessibility and, consequently in genetic expression of specific genes, especially *AP2* and *B3* transcription factors. Finally, the third level of the hierarchy is composed of the transcription factors themselves (*AP2*, *B3* and *NF-Y*), which are expressed thanks to their opened chromatin status, and that initiate or regulate embryo formation.

6. Further Prospects

The mechanisms underlying the control of somatic embryogenesis are multiple, involve complex gene regulatory networks, hormonal and epigenetic controls, and remain poorly understood. For instance, *SOMATIC EMBRYOGENESIS RECEPTOR KINASE* (*SERK*) genes are leucine-rich repeat receptor-like kinase (LRR-RKK) [128] first identified in carrot as markers of embryogenic cells and involved in the transition from the vegetative to embryogenic state of the cells, but also in response to environmental signals and in plant development [129]. In vitro, *SERK* expression can be induced by various stresses such as drought, wounding, or by altering hormone balance in the culture media [17].

From a gene regulatory network point of view, somatic embryogenesis shares similarities with zygotic embryogenesis, and some differences. The transcriptome of embryonic cells is similar to those of zygotic embryos at the octant stage, but not earlier [21,130], and many stress-related genes are also detected in somatic embryos [131]. This emphasizes the similarity of molecular controls involved but also some specificities. Moreover, the similarity may result from a convergence of different morphogenic pathways [21]. The production of embryogenic cells in *Arabidopsis* callus needs the repression of biochemical pathways and root meristem genes, while activating gene networks involved in shoot patterning and polarized cell growth rather than inducing specific zygotic embryo networks [21].

It can be hypothesized that SE is controlled by embryogenic genes that acquired different or more specialized functions (neo- or sub-functionalization) or specific regulations during evolution that remain under specific environmental or physiological conditions capable of inducing embryogenesis. Different master regulatory genes of zygotic embryogenesis, such as *LAFL*, can be involved in the cellular stress necessary for transcriptional reprogramming. Moreover, the AP2 family of transcription factors, which contains AINTEGUMENTA-LIKE (ANT), BABY BOOM (BBM) and PLETHORA (APB), is involved in the maintenance of sporophytic meristems in Angiosperms [132]. Nevertheless, in ferns these genes are involved in the establishment of a gametophytic meristem without fertilization, and the overexpression of *ANT* or the *Brassica napus* *BBM* gene significantly increases the formation of sporophytes without gamete (apogamy), whereas the knockdown of *ANT* reduces this ability [133]. This type of conservation of an ancestral function, also suggested for several transcription factors genes [132], can explain the ability of Angiosperms to produce somatic embryos.

The production of somatic embryos appears to be strongly linked to specific cellular events. For further comprehension of these mechanisms, a better understanding of the gene networks involved is necessary at the cellular level: characterization of single cell transcriptome, epigenome, and proteome are required to develop better knowledge about molecular events in place at the cellular level. Recent technologies of single cell analyses or cellular transcriptome atlas are thus highly important to better appreciate the early gene regulatory networks associated with somatic embryogenesis. These techniques applied to samples enriched in embryonic cells with the help of specific marker genes or specific morphological markers build the basis of a better knowledge on early controls of somatic embryogenesis. For example, the use of *proLEC1* or *proLEC2* fused to a fluorescent protein marker is considered as a valuable marker to detect early events of SE production as used in [21]. Similarly, to monitor early molecular changes during SE, using *proBBM* should certainly be considered [52,70].

LEC1 is acting as a pioneer TF during embryogenesis to control *FLC* expression [134]. *LEC2* forming some complexes with *LEC1* can be involved in similar regulations. This pioneer activity may induce the key genes involved in the reprogramming of a cell to form a somatic embryo (Figure 3). Nevertheless, TFs such as *LEC2*, *ABI3* or *FUS3*, which have several protein domains, can perform different roles (i.e., controlling early embryogenesis, reserve accumulation or desiccation tolerance), depending on their partners in different cellular environments (Figures 2 and 3). Alternatively, the different functions can be linked if the ectopic triggering of the maturation processes induces stress in vegetative tissues leading to the initiation of somatic embryogenesis.

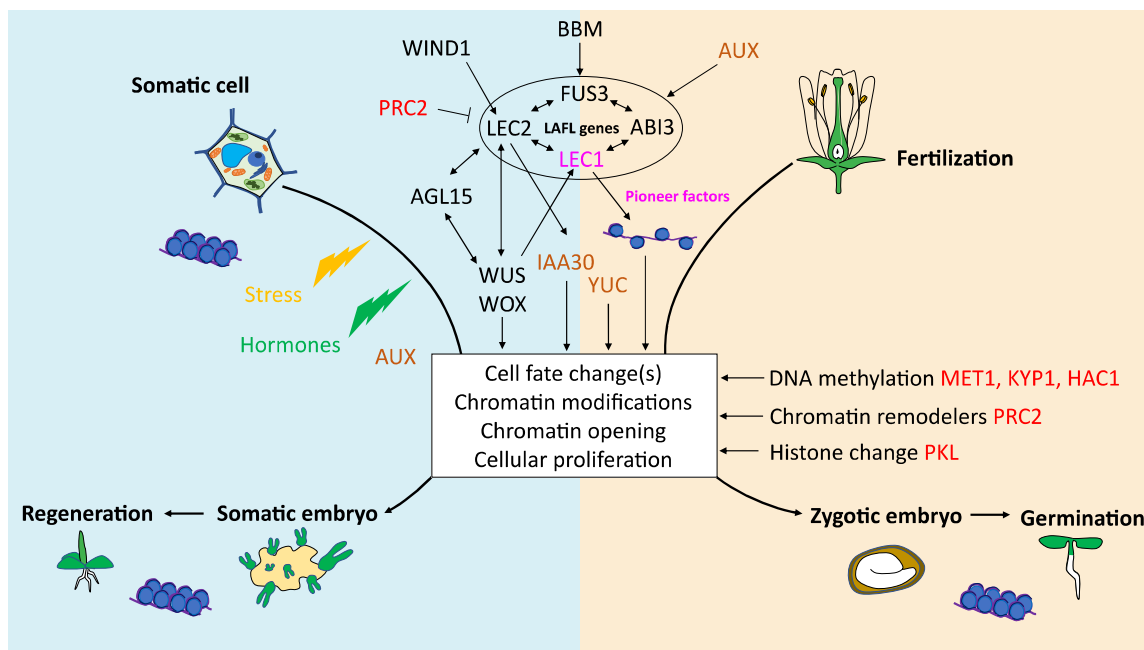


Figure 3. General overview of genes involved in somatic embryogenesis or zygotic embryogenesis in interaction with LEC1, ABI3, FUS3, and LEC2 (LAFL) transcription factors. Hormone-related genes are represented in brown, epigenetic regulators in red, transcription factors in black and pioneer factors in bright pink.

Author Contributions: Writing—review and editing, C.S., L.L., B.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work has benefited from a French State grant (Saclay Plant Sciences, reference ANR-17-EUR-0007, EUR SPS-GSR) managed by the French National Research Agency under an Investments for the Future program (reference ANR-11-IDEX-0003-02).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank M.M., S.K., H.W. and J.T. for critical reading of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

ABA	Abcisic Acid
ABI3	ABSCISIC ACID INSENSITIVE 3
AGL	AGAMOUS-LIKE
BBM	BABY BOOM
FUS3	FUSCA 3
GA	Gibberellic Acid
HAC1	HISTONE ACETYLTRANSFERASE 1
IAA30	INDOL-3-ACETIC ACID INDUCIBLE 30
KYP1	KRYPTONITE 1
LEC1	LEAFY COTYLEDON 1
LEC2	LEAFY COTYLEDON 2

MET1	METHYLTRANSFERASE 1
PGR	Plant Growth Regulators
PKL	PICKLE
PRC2	POLYCOMB RESPONSE COMPLEX 2
SE	Somatic Embryogenesis
WUS	WUSCHEL
ZE	Zygotic Embryogenesis

References

- Jo, L.; Pelletier, J.M.; Harada, J.J. Central role of the LEAFY COTYLEDON1 transcription factor in seed development. *J. Integr. Plant Biol.* **2019**, *61*, 12806. [[CrossRef](#)]
- Iluž, D. Zoochory: The dispersal of plants by animals. In *All Flesh Is Grass*; Dubinsky, Z., Seckbach, J., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2011; pp. 199–214, ISBN 978-90-481-9316-5.
- Figueiredo, D.D.; Köhler, C. Auxin: A molecular trigger of seed development. *Genes Dev.* **2018**, *32*, 479–490. [[CrossRef](#)] [[PubMed](#)]
- Garcês, H.M.P.; Champagne, C.E.M.; Townsley, B.T.; Park, S.; Malhó, R.; Pedroso, M.C.; Harada, J.J.; Sinha, N.R. Evolution of asexual reproduction in leaves of the genus *Kalanchoë*. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 15578–15583. [[CrossRef](#)] [[PubMed](#)]
- Garcês, H.M.P.; Koenig, D.; Townsley, B.T.; Kim, M.; Sinha, N.R. Truncation of LEAFY COTYLEDON1 protein is required for asexual reproduction in *Kalanchoë daigremontiana*. *Plant Physiol.* **2014**, *165*, 196–206. [[CrossRef](#)] [[PubMed](#)]
- Loyola-Vargas, V.M.; Avilez-Montalvo, J.R.; Avilés-Montalvo, R.N.; Márquez-López, R.E.; Galaz-Ávalos, R.M.; Mellado-Mojica, E. Somatic embryogenesis in *Coffea* spp. In *Somatic Embryogenesis: Fundamental Aspects and Applications*; Springer International Publishing: Cham, Switzerland, 2016; pp. 241–266, ISBN 9783319337050.
- Lelu-Walter, M.-A.; Klimaszewska, K.; Miguel, C.; Aronen, T.; Hargreaves, C.; Teyssier, C.; Trontin, J.-F. Somatic embryogenesis for more effective breeding and deployment of improved varieties in *Pinus* spp.: Bottlenecks and recent advances. In *Somatic Embryogenesis: Fundamental Aspects and Applications*; Loyola-Vargas, V.M., Ochoa-Alejo, N., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 319–365, ISBN 9783319337050.
- Alexandra Pila Quinga, L.; Heringer, A.S.; Pacheco de Freitas Fraga, H.; do Nascimento Vieira, L.; Silveira, V.; Steinmacher, D.A.; Guerra, M.P. Insights into the conversion potential of *Theobroma cacao* L. somatic embryos using quantitative proteomic analysis. *Sci. Hortic.* **2018**, *229*, 65–76. [[CrossRef](#)]
- Loyola-Vargas, V.M.; Ochoa-Alejo, N. Part II Somatic Embryogenesis of Basic Models and Industrial and Agronomical Crops. In *Somatic Embryogenesis: Fundamental Aspects and Applications*; Loyola-Vargas, V.M., Ochoa-Alejo, N., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 185–412, ISBN 9783319337050.
- Ochoa-Alejo, N. The uses of somatic embryogenesis for genetic transformation. In *Somatic Embryogenesis: Fundamental Aspects and Applications*; Loyola-Vargas, V.M., Ochoa-Alejo, N., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 415–434, ISBN 978-3-319-33705-0.
- Rasmussen, S.K.; Kumar Mall, T.; Duraiagaraja, S.; Agrawal, P.K.; Yadava, P.; Abhishek, A.; Singh, R.; Singh, I.; Kaul, T.; Pattanayak, A. Advances in maize transformation technologies and development of transgenic maize. *Front. Plant Sci.* **2017**, *7*, 1949. [[CrossRef](#)]
- Raza, G.; Singh, M.B.; Bhalla, P.L. Somatic embryogenesis and plant regeneration from commercial soybean cultivars. *Plants* **2020**, *9*, 38. [[CrossRef](#)]
- Etienne, H.; Guyot, R.; Beulé, T.; Breitler, J.-C.; Jaligot, E. Plant fidelity in somatic embryogenesis-regenerated plants. In *Somatic Embryogenesis: Fundamental Aspects and Applications*; Loyola-Vargas, V.M., Ochoa-Alejo, N., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 121–150, ISBN 9783319337050.
- Leljak-Levanić, D.; Mihaljević, S.; Bauer, N. Somatic and zygotic embryos share common developmental features at the onset of plant embryogenesis. *Acta Physiol. Plant.* **2015**, *37*, 1–14. [[CrossRef](#)]
- Fehér, A. Somatic embryogenesis—stress-induced remodeling of plant cell fate. *Biochim. Biophys. Acta Gene Regul. Mech.* **2015**, *1849*, 385–402. [[CrossRef](#)]
- Su, Y.H.; Tang, L.P.; Zhao, X.Y.; Zhang, X.S. Plant cell totipotency: Insights into cellular reprogramming. *J. Integr. Plant Biol.* **2020**, *63*, 228–243. [[CrossRef](#)]
- Nic-Can, G.I.; Avilez-Montalvo, J.R.; Aviles-Montalvo, R.N.; Márquez-López, R.E.; Mellado-Mojica, E.; Galaz-Ávalos, R.M.; Loyola-Vargas, V.M. The relationship between stress and somatic embryogenesis. In *Somatic Embryogenesis: Fundamental Aspects and Applications*; Springer International Publishing: Cham, Switzerland, 2016; pp. 151–170, ISBN 9783319337050.
- Skoog, F.; Miller, C.O. Chemical regulation of growth and organ formation in plant tissues cultured in vitro. *Symp. Soc. Exp. Biol.* **1957**, *11*, 118–130. [[PubMed](#)]
- Nic-Can, G.I.; Loyola-Vargas, V.M. The role of the auxins during somatic embryogenesis. In *Somatic Embryogenesis: Fundamental Aspects and Applications*; Loyola-Vargas, V.M., Ochoa-Alejo, N., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 171–182, ISBN 9783319337050.
- Nowak, K.; Gaj, M.D. Transcription factors in the regulation of somatic embryogenesis. In *Somatic Embryogenesis: Fundamental Aspects and Applications*; Springer International Publishing: Cham, Switzerland, 2016; pp. 53–79, ISBN 9783319337050.

21. Magnani, E.; Jiménez-Gómez, J.M.; Soubigou-Taconnat, L.; Lepiniec, L.; Fiume, E. Profiling the onset of somatic embryogenesis in *Arabidopsis*. *BMC Genom.* **2017**, *18*, 1–12. [[CrossRef](#)] [[PubMed](#)]
22. Ikeda-Iwai, M.; Satoh, S.; Kamada, H. Establishment of a reproducible tissue culture system for the induction of *Arabidopsis* somatic embryos. *J. Exp. Bot.* **2002**, *53*, 1575–1580. [[CrossRef](#)] [[PubMed](#)]
23. Tian, R.; Paul, P.; Joshi, S.; Perry, S.E. Genetic activity during early plant embryogenesis. *Biochem. J.* **2020**, *477*, 3743–3767. [[CrossRef](#)] [[PubMed](#)]
24. Gaj, M.D. Factors influencing somatic embryogenesis induction and plant regeneration with particular reference to *Arabidopsis thaliana* (L.) Heynh. *Plant Growth Regul.* **2004**, *43*, 27–47. [[CrossRef](#)]
25. Guo, H.; Fan, Y.; Guo, H.; Wu, J.; Yu, X.; Wei, J.; Lian, X.; Zhang, L.; Gou, Z.; Fan, Y.; et al. Somatic embryogenesis critical initiation stage-specific mCHH hypomethylation reveals epigenetic basis underlying embryogenic redifferentiation in cotton. *Plant Biotechnol. J.* **2020**, 1–3. [[CrossRef](#)]
26. Solís-Ramos, Y.L.; Andrade-Torres, A.; Sáenz Carbonell, L.A.; Oropeza Salín, C.M.; Castaño de la Ser, E. Somatic embryogenesis in recalcitrant plants. In *Embryogenesis*; Sato, K.-I., Ed.; InTech: Kyoto, Japan, 2012; pp. 597–618. [[CrossRef](#)]
27. Casanova-Sáez, R.; Voß, U. Auxin metabolism controls developmental decisions in land plants. *Trends Plant Sci.* **2019**, *24*, 741–754. [[CrossRef](#)] [[PubMed](#)]
28. Paque, S.; Weijers, D. Q&A: Auxin: The plant molecule that influences almost anything. *BMC Biol.* **2016**, *14*, 67. [[CrossRef](#)]
29. Petrášek, J.; Friml, J. Auxin transport routes in plant development. *Development* **2009**, *136*, 2675–2688. [[CrossRef](#)]
30. Robert, H.S.; Grones, P.; Stepanova, A.N.; Robles, L.M.; Lokerse, A.S.; Alonso, J.M.; Weijers, D.; Friml, J. Local auxin sources orient the apical-basal axis in *Arabidopsis* embryos. *Curr. Biol.* **2013**, *23*, 2506–2512. [[CrossRef](#)] [[PubMed](#)]
31. Xiong, F.; Liu, H.H.; Duan, C.Y.; Zhang, B.K.; Wei, G.; Zhang, Y.; Li, S. *Arabidopsis* JANUS regulates embryonic pattern formation through Pol II-Mediated transcription of *WOX2* and *PIN7*. *iScience* **2019**, *19*, 1179–1188. [[CrossRef](#)] [[PubMed](#)]
32. Robert, H.S.; Grunewald, W.; Sauer, M.; Cannoot, B.; Soriano, M.; Swarup, R.; Weijers, D.; Bennett, M.; Boutilier, K.; Friml, J. Plant embryogenesis requires AUX/LAX-mediated auxin influx. *Development* **2015**, *142*, 702–711. [[CrossRef](#)] [[PubMed](#)]
33. Wójcik, A.M.; Wójcikowska, B.; Gaj, M.D. Current perspectives on the auxin-mediated genetic network that controls the induction of somatic embryogenesis in plants. *Int. J. Mol. Sci.* **2020**, *21*, 1333. [[CrossRef](#)]
34. Márquez-López, R.E.; Pérez-Hernández, C.; Ku-González, Á.; Galaz-Ávalos, R.M.; Loyola-Vargas, V.M. Localization and transport of indole-3-acetic acid during somatic embryogenesis in *Coffea canephora*. *Protoplasma* **2018**, *255*, 695–708. [[CrossRef](#)] [[PubMed](#)]
35. Wójcikowska, B.; Gaj, M.D. Expression profiling of *AUXIN RESPONSE FACTOR* genes during somatic embryogenesis induction in *Arabidopsis*. *Plant Cell Rep.* **2017**, *36*, 843–858. [[CrossRef](#)]
36. Zheng, Y.; Ren, N.; Wang, H.; Stromberg, A.J.; Perry, S.E. Global identification of targets of the *Arabidopsis* MADS domain protein AGAMOUS-Like15. *Plant Cell* **2009**, *21*, 2563–2577. [[CrossRef](#)]
37. Wójcikowska, B.; Jaskóła, K.; Gašiorek, P.; Meus, M.; Nowak, K.; Gaj, M.D. LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of *Arabidopsis*, via YUCCA-mediated auxin biosynthesis. *Planta* **2013**, *238*, 425–440. [[CrossRef](#)]
38. Pasternak, T.; Miskolczi, P.; Ayaydin, F.; Mészáros, T.; Dudits, D.; Fehér, A. Exogenous auxin and cytokinin dependent activation of CDKs and cell division in leaf protoplast-derived cells of alfalfa. *Plant Growth Regul.* **2000**, *32*, 129–141. [[CrossRef](#)]
39. Pasternak, T.; Dudits, D. Epigenetic clues to better understanding of the asexual embryogenesis in *planta* and in vitro. *Front. Plant Sci.* **2019**, *10*, 1–5. [[CrossRef](#)] [[PubMed](#)]
40. Horstman, A.; Bemer, M.; Boutilier, K. A transcriptional view on somatic embryogenesis. *Regeneration* **2017**, *4*, 201–216. [[CrossRef](#)] [[PubMed](#)]
41. Santos Mendoza, M.; Dubreucq, B.; Miquel, M.; Caboche, M.; Lepiniec, L. LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in *Arabidopsis* leaves. *FEBS Lett.* **2005**, *579*, 4666–4670. [[CrossRef](#)] [[PubMed](#)]
42. Braybrook, S.A.; Harada, J.J. LECs go crazy in embryo development. *Trends Plant Sci.* **2008**, *13*, 624–630. [[CrossRef](#)] [[PubMed](#)]
43. Carbonero, P.; Iglesias-Fernández, R.; Vicente-Carbajosa, J. The AFL subfamily of B3 transcription factors: Evolution and function in angiosperm seeds. *J. Exp. Bot.* **2017**, *68*, 871–880. [[CrossRef](#)] [[PubMed](#)]
44. Sasnauskas, G.; Kauneckaitė, K.; Siksnys, V. Structural basis of DNA target recognition by the B3 domain of *Arabidopsis* epigenome reader VAL1. *Nucleic Acids Res.* **2018**, *46*, 4316–4324. [[CrossRef](#)] [[PubMed](#)]
45. Baumlein, H.; Villarroel, R.; Inz, D.; Wobus, U. Cis-analysis of a seed protein gene promoter: The conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *Plant J.* **1991**, *2*, 233–239.
46. Baud, S.; Kelemen, Z.; Thévenin, J.; Boulard, C.; Blanchet, S.; To, A.; Payre, M.; Berger, N.; Effroy-Cuzzi, D.; Franco-Zorrilla, J.M.; et al. Deciphering the molecular mechanisms underpinning the transcriptional control of gene expression by master transcriptional regulators in *Arabidopsis* seed. *Plant Physiol.* **2016**, *171*, 1099–1112. [[CrossRef](#)]
47. Lee, H.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2152–2156. [[CrossRef](#)]
48. Santos-Mendoza, M.; Dubreucq, B.; Bastien Baud, S.; Parcy, F.; Caboche, M.; Lepiniec, L. Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *Plant J.* **2008**, *54*, 608–620. [[CrossRef](#)] [[PubMed](#)]
49. Lepiniec, L.; Devic, M.; Roscoe, T.J.; Bouyer, D.; Zhou, D.X.; Boulard, C.; Baud, S.; Dubreucq, B. Molecular and epigenetic regulations and functions of the LAFL transcriptional regulators that control seed development. *Plant Reprod.* **2018**, *31*, 291–307. [[CrossRef](#)]

50. Kumar, V.; Jha, P.; Van Staden, J. LEAFY COTYLEDONs (LECs): Master regulators in plant embryo development. *Plant Cell. Tissue Organ Cult.* **2020**, *140*, 475–487. [[CrossRef](#)]
51. Boulard, C.; Fatihi, A.; Lepiniec, L.; Dubreucq, B. Regulation and evolution of the interaction of the seed B3 transcription factors with NF-Y subunits. *Biochim. Biophys. Acta Gene Regul. Mech.* **2017**, *1860*, 1069–1078. [[CrossRef](#)]
52. Horstman, A.; Li, M.; Heidmann, I.; Weemen, M.; Chen, B.; Muiño, J.M.; Angenent, G.C.; Boutilier, K.; Muino, J.M.; Angenent, G.C.; et al. The BABY BOOM transcription factor activates the LEC1-ABI3-FUS3-LEC2 network to induce somatic embryogenesis. *Plant Physiol.* **2017**, *175*, 848–857. [[CrossRef](#)]
53. Ledwoń, A.; Gaj, M.D. LEAFY COTYLEDON1, FUSCA3 expression and auxin treatment in relation to somatic embryogenesis induction in *Arabidopsis*. *Plant Growth Regul.* **2011**, *65*, 157–167. [[CrossRef](#)]
54. Gaj, M.D.; Zhang, S.; Harada, J.J.; Lemaux, P.G. Leafy cotyledon genes are essential for induction of somatic embryogenesis of *Arabidopsis*. *Planta* **2005**, *222*, 977–988. [[CrossRef](#)]
55. Stone, S.L.; Kwong, L.W.; Yee, K.M.; Pelletier, J.; Lepiniec, L.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11806–11811. [[CrossRef](#)]
56. Feeney, M.; Frigerio, L.; Cui, Y.; Menassa, R. Following vegetative to embryonic cellular changes in leaves of *Arabidopsis* overexpressing LEAFY COTYLEDON2. *Plant Physiol.* **2013**, *162*, 1881–1896. [[CrossRef](#)]
57. Lotan, T.; Ohto, M.A.; Matsudaira Yee, K.; West, M.A.L.; Lo, R.; Kwong, R.W.; Yamagishi, K.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* **1998**, *93*, 1195–1205. [[CrossRef](#)]
58. Liu, Z.; Ge, X.X.; Qiu, W.M.; Long, J.M.; Jia, H.H.; Yang, W.; Dutt, M.; Wu, X.M.; Guo, W.W. Overexpression of the CsFUS3 gene encoding a B3 transcription factor promotes somatic embryogenesis in Citrus. *Plant Sci.* **2018**, *277*, 121–131. [[CrossRef](#)] [[PubMed](#)]
59. Guo, F.; Liu, C.; Xia, H.; Bi, Y.; Zhao, C.; Zhao, S.; Hou, L.; Li, F.; Wang, X. Induced expression of AtLEC1 and AtLEC2 differentially promotes somatic embryogenesis in transgenic tobacco plants. *PLoS ONE* **2013**, *8*, e71714. [[CrossRef](#)]
60. Zheng, W.; Zhang, X.; Yang, Z.; Wu, J.; Li, F.; Duan, L.; Liu, C.; Lu, L.; Zhang, C.; Li, F. AtWUSCHEL promotes formation of the embryogenic callus in *Gossypium hirsutum*. *PLoS ONE* **2014**, *9*, e87502. [[CrossRef](#)] [[PubMed](#)]
61. Braybrook, S.A.; Stone, S.L.; Park, S.; Bui, A.Q.; Le, B.H.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3468–3473. [[CrossRef](#)]
62. Zheng, Q.; Zheng, Y.; Ji, H.; Burnie, W.; Perry, S.E. Gene regulation by the AGL15 transcription factor reveals hormone interactions in somatic embryogenesis. *Plant Physiol.* **2016**, *172*, 2374–2387. [[CrossRef](#)]
63. Sato, A.; Yamamoto, K.T. Overexpression of the non-canonical Aux/IAA genes causes auxin-related aberrant phenotypes in *Arabidopsis*. *Physiol. Plant.* **2008**, *133*, 397–405. [[CrossRef](#)]
64. Xu, C.; Luo, F.; Hochholdinger, F. LOB domain proteins: Beyond lateral organ boundaries. *Trends Plant Sci.* **2016**, *21*, 159–167. [[CrossRef](#)]
65. Stone, S.L.; Braybrook, S.A.; Paula, S.L.; Kwong, L.W.; Meuser, J.; Pelletier, J.; Hsieh, T.-F.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. *Arabidopsis* LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3151–3156. [[CrossRef](#)] [[PubMed](#)]
66. To, A.; Valon, C.; Savino, G.; Guillemot, J.; Devic, M.; Giraudat, J.; Parcy, F. A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. *Plant Cell Online* **2006**, *18*, 1642–1651. [[CrossRef](#)] [[PubMed](#)]
67. Sengupta, S.; Nag Chaudhuri, R. ABI3 plays a role in de-novo root regeneration from *Arabidopsis thaliana* callus cells. *Plant Signal. Behav.* **2020**, *15*, 1794147. [[CrossRef](#)] [[PubMed](#)]
68. Parcy, F.; Valon, C.; Raynal, M.; Gaubier-Comella, P.; Delseny, M.; Giraudat, J. Regulation of gene expression programs during *Arabidopsis* seed development: Roles of the ABI3 locus and of endogenous abscisic acid. *Plant Cell* **1994**, *6*, 1567–1582. [[CrossRef](#)]
69. Parcy, F.; Giraudat, J. Interactions between the ABI1 and the ectopically expressed ABI3 genes in controlling abscisic acid responses in *Arabidopsis* vegetative tissues. *Plant J.* **1997**, *11*, 693–702. [[CrossRef](#)]
70. Boutilier, K.; Offringa, R.; Sharma, V.K.; Kieft, H.; Ouellet, T.; Zhang, L.; Hattori, J.; Liu, C.-M.; van Lammeren, A.A.M.; Miki, B.L.A.; et al. Ectopic expression of BABY-BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* **2002**, *14*, 1737–1749. [[CrossRef](#)]
71. Riechmann, J.L.; Meyerowitz, E.M. The AP2/EREBP family of plant transcription factors. *Biol. Chem.* **1998**, *379*, 633–646. [[CrossRef](#)]
72. Mizoi, J.; Shinozaki, K.; Yamaguchi-Shinozaki, K. AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta Gene Regul. Mech.* **2012**, *1819*, 86–96. [[CrossRef](#)]
73. Galinha, C.; Hofhuis, H.; Luijten, M.; Willemsen, V.; Blilou, I.; Heidstra, R.; Scheres, B. PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature* **2007**, *449*, 1053–1057. [[CrossRef](#)]
74. Laux, T.; Mayer, K.F.X.; Berger, J.; Jürgens, G. The WUSCHEL gene is required for shoot and floral meristem integrity in *Arabidopsis*. *Development* **1996**, *122*, 87–96. [[CrossRef](#)]
75. Zuo, J.; Niu, Q.W.; Frugis, G.; Chua, N.H. The WUSCHEL gene promotes vegetative-to-embryonic transition in *Arabidopsis*. *Plant J.* **2002**, *30*, 349–359. [[CrossRef](#)] [[PubMed](#)]

76. Jha, P.; Ochatt, S.J.; Kumar, V. *WUSCHEL*: A master regulator in plant growth signaling. *Plant Cell Rep.* **2020**, *39*, 431–444. [[CrossRef](#)] [[PubMed](#)]
77. Mordhorst, A.P.; Hartog, M.V.; El Tamer, M.K.; Laux, T.; De Vries, S.C. Somatic embryogenesis from *Arabidopsis* shoot apical meristem mutants. *Planta* **2002**, *214*, 829–836. [[CrossRef](#)]
78. Bouchabké-Coussa, O.; Obellianne, M.; Linderme, D.; Montes, E.; Maia-Grondard, A.; Vilaine, F.; Pannetier, C. *WUSCHEL* overexpression promotes somatic embryogenesis and induces organogenesis in cotton (*Gossypium hirsutum* L.) tissues cultured in vitro. *Plant Cell Rep.* **2013**, *32*, 675–686. [[CrossRef](#)] [[PubMed](#)]
79. Arroyo-Herrera, A.; Ku Gonzalez, A.; Canche Moo, R.; Quiroz-Figueroa, F.R.; Loyola-Vargas, V.M.; Rodriguez-Zapata, L.C.; Burgeff D'Hondt, C.; Suárez-Solís, V.M.; Castaño, E. Expression of *WUSCHEL* in *Coffea canephora* causes ectopic morphogenesis and increases somatic embryogenesis. *Plant Cell. Tissue Organ Cult.* **2008**, *94*, 171–180. [[CrossRef](#)]
80. Orłowska, A.; Kępczyńska, E. Identification of Polycomb Repressive Complex1, Trithorax group genes and their simultaneous expression with *WUSCHEL*, *WUSCHEL*-related *Homeobox5* and *SHOOT MERISTEMLESS* during the induction phase of somatic embryogenesis in *Medicago truncatula* Gaertn. *Plant Cell. Tissue Organ Cult.* **2018**, *134*, 345–356. [[CrossRef](#)]
81. Haecker, A.; Groß-Hardt, R.; Geiges, B.; Sarkar, A.; Breuninger, H.; Herrmann, M.; Laux, T. Expression dynamics of *WOX* genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. *Development* **2004**, *131*, 657–668. [[CrossRef](#)] [[PubMed](#)]
82. Tvorogova, V.E.; Fedorova, Y.A.; Potsenkovskaya, E.A.; Kudriashov, A.A.; Efremova, E.P.; Kvitkovskaya, V.A.; Wolabu, T.W.; Zhang, F.; Tadege, M.; Lutova, L.A. The *WUSCHEL*-related homeobox transcription factor MtWOX9-1 stimulates somatic embryogenesis in *Medicago truncatula*. *Plant Cell. Tissue Organ Cult.* **2019**, *138*, 517–527. [[CrossRef](#)]
83. Masiero, S.; Colombo, L.; Grini, P.E.; Schnittger, A.; Kater, M.M. The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell* **2011**, *23*, 865–872. [[CrossRef](#)]
84. Heck, G.R.; Perry, S.E.; Nichols, K.W.; Fernandez, D.E. AGL15, a MADS domain protein expressed in developing embryos. *Plant Cell* **1995**, *7*, 1271–1282. [[CrossRef](#)] [[PubMed](#)]
85. Rounsley, S.D.; Ditta, G.S.; Yanofsky, M.F. Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell* **1995**, *7*, 1259–1269. [[CrossRef](#)] [[PubMed](#)]
86. Wang, H.; Caruso, L.V.; Downie, A.B.; Perry, S.E. The embryo MADS domain protein AGAMOUS-LIKE15 directly regulates expression of a gene encoding an enzyme involved in gibberellin metabolism. *Plant Cell* **2004**, *16*, 1206–1219. [[CrossRef](#)]
87. Zhu, C.; Perry, S.E. Control of expression and autoregulation of AGL15, a member of the MADS-box family. *Plant J.* **2005**, *41*, 583–594. [[CrossRef](#)]
88. Harding, E.W.; Tang, W.; Nichols, K.W.; Fernandez, D.E.; Perry, S.E. Expression and maintenance of embryogenic potential is enhanced through constitutive expression of *AGAMOUS-LIKE15*. *Plant Physiol.* **2003**, *133*, 653–663. [[CrossRef](#)]
89. Thakare, D.; Tang, W.; Hill, K.; Perry, S.E. The MADS-domain transcriptional regulator *AGAMOUS-Like15* promotes somatic embryo development in *Arabidopsis* and soybean. *Plant Physiol.* **2008**, *146*, 1663–1672. [[CrossRef](#)]
90. Iwase, A.; Ishii, H.; Aoyagi, H.; Ohme-Takagi, M.; Tanaka, H. Comparative analyses of the gene expression profiles of *Arabidopsis* intact plant and cultured cells. *Biotechnol. Lett.* **2005**, *27*, 1097–1103. [[CrossRef](#)]
91. Iwase, A.; Mitsuda, N.; Koyama, T.; Hiratsu, K.; Kojima, M.; Arai, T.; Inoue, Y.; Seki, M.; Sakakibara, H.; Sugimoto, K.; et al. The AP2/ERF transcription factor *WIND1* controls cell dedifferentiation in *Arabidopsis*. *Curr. Biol.* **2011**, *21*, 508–514. [[CrossRef](#)] [[PubMed](#)]
92. Iwase, A.; Mita, K.; Nonaka, S.; Ikeuchi, M.; Koizuka, C.; Ohnuma, M.; Ezura, H.; Imamura, J.; Sugimoto, K. *WIND1*-based acquisition of regeneration competency in *Arabidopsis* and rapeseed. *J. Plant Res.* **2015**, *128*, 389–397. [[CrossRef](#)] [[PubMed](#)]
93. Iwafuchi-Doi, M. The mechanistic basis for chromatin regulation by pioneer transcription factors. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2019**, *11*, e1427. [[CrossRef](#)]
94. Zaret, K.S.; Carroll, J.S. Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev.* **2011**, *25*, 2227–2241. [[CrossRef](#)]
95. Mayran, A.; Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **2018**, *293*, 13795–13804. [[CrossRef](#)]
96. Zaret, K.S. Pioneering the chromatin landscape. *Nat. Genet.* **2018**, *50*, 167–169. [[CrossRef](#)]
97. Tao, Z.; Hu, H.; Luo, X.; Jia, B.; Du, J.; He, Y. Embryonic resetting of the parental vernalized state by two B3 domain transcription factors in *Arabidopsis*. *Nat. Plants* **2019**, *5*, 424–435. [[CrossRef](#)] [[PubMed](#)]
98. Kumar, V.; Van Staden, J. New insights into plant somatic embryogenesis: An epigenetic view. *Acta Physiol. Plant.* **2017**, *39*, 1–17. [[CrossRef](#)]
99. Klemm, S.L.; Shipony, Z.; Greenleaf, W.J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **2019**, *20*, 207–220. [[CrossRef](#)]
100. Peng, H.; Zhang, J. Plant genomic DNA methylation in response to stresses: Potential applications and challenges in plant breeding. *Prog. Nat. Sci.* **2009**, *19*, 1037–1045. [[CrossRef](#)]
101. Akhter, Z.; Bi, Z.; Ali, K.; Sun, C.; Fiaz, S.; Haider, F.U.; Bai, J. In response to abiotic stress, DNA methylation confers epigenetic changes in plants. *Plants* **2021**, *10*, 1096. [[CrossRef](#)]
102. Liu, J.; He, Z. Small DNA Methylation, big player in plant abiotic stress responses and memory. *Front. Plant Sci.* **2020**, *11*, 1977. [[CrossRef](#)]

103. Lee, K.; Seo, P.J. Dynamic epigenetic changes during plant regeneration. *Trends Plant Sci.* **2018**, *23*, 235–247. [[CrossRef](#)] [[PubMed](#)]
104. Birnbaum, K.D.; Roudier, F. Epigenetic memory and cell fate reprogramming in plants. *Regeneration* **2017**, *4*, 15–20. [[CrossRef](#)] [[PubMed](#)]
105. Lee, K.; Park, O.S.; Jung, S.J.; Seo, P.J. Histone deacetylation-mediated cellular dedifferentiation in *Arabidopsis*. *J. Plant Physiol.* **2016**, *191*, 95–100. [[CrossRef](#)]
106. Li, J.; Wang, M.; Li, Y.; Zhang, Q.; Lindsey, K.; Daniell, H.; Jin, S.; Zhang, X. Multi-omics analyses reveal epigenomics basis for cotton somatic embryogenesis through successive regeneration acclimation process. *Plant Biotechnol. J.* **2019**, *17*, 435–450. [[CrossRef](#)]
107. Osorio-Montalvo, P.; Sáenz-Carbonell, L.; De-la-Peña, C. 5-Azacytidine: A Promoter of Epigenetic Changes in the Quest to Improve Plant Somatic Embryogenesis. *Int. J. Mol. Sci.* **2018**, *19*, 3182. [[CrossRef](#)]
108. Mozgova, I.; Hennig, L. The polycomb group protein regulatory network. *Annu. Rev. Plant Biol.* **2015**, *66*, 269–296. [[CrossRef](#)]
109. Duarte-Aké, F.; Nic-Can, G.; De-la-Peña, C. Somatic embryogenesis: Polycomb complexes control cell-to-embryo transition. In *Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications*; Springer International Publishing: Cham, Switzerland, 2019; pp. 339–354, ISBN 9783030147600.
110. Ikeuchi, M.; Iwase, A.; Rymen, B.; Harashima, H.; Shibata, M.; Ohnuma, M.; Breuer, C.; Karina Morao, A.; de Lucas, M.; De Veylder, L.; et al. PRC2 represses dedifferentiation of mature somatic cells in *Arabidopsis*. *Nat. Plants* **2015**, *1*, 15089. [[CrossRef](#)] [[PubMed](#)]
111. Chanvivattana, Y.; Bishopp, A.; Schubert, D.; Stock, C.; Moon, Y.H.; Sung, Z.R.; Goodrich, J. Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. *Development* **2004**, *131*, 5263–5276. [[CrossRef](#)]
112. Mozgová, I.; Muñoz-Viana, R.; Hennig, L. PRC2 represses hormone-induced somatic embryogenesis in vegetative tissue of *Arabidopsis thaliana*. *PLoS Genet.* **2017**, *13*, e1006562. [[CrossRef](#)]
113. Han, S.K.; Wu, M.F.; Cui, S.; Wagner, D. Roles and activities of chromatin remodeling ATPases in plants. *Plant J.* **2015**, *83*, 62–77. [[CrossRef](#)]
114. Schneider, A.; Aghamirzaie, D.; Elmarakeby, H.; Poudel, A.N.; Koo, A.J.; Heath, L.S.; Grene, R.; Collakova, E. Potential targets of VIVIPAROUS1/ABI3-LIKE1 (VAL1) repression in developing *Arabidopsis thaliana* embryos. *Plant J.* **2016**, *85*, 305–319. [[CrossRef](#)] [[PubMed](#)]
115. Shen, Y.; Devic, M.; Lepiniec, L.; Zhou, D.X. Chromodomain, Helicase and DNA-binding CHD1 protein, CHR5, are involved in establishing active chromatin state of seed maturation genes. *Plant Biotechnol. J.* **2015**, *13*, 811–820. [[CrossRef](#)] [[PubMed](#)]
116. Brocklehurst, S.; Watson, M.; Carr, I.M.; Out, S.; Heidmann, I.; Meyer, P. Induction of epigenetic variation in *Arabidopsis* by over-expression of DNA METHYLTRANSFERASE1 (MET1). *PLoS ONE* **2018**, *13*, e0192170. [[CrossRef](#)]
117. Jackson, J.P.; Lindroth, A.M.; Cao, X.; Jacobsen, S.E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **2002**, *416*, 556–560. [[CrossRef](#)] [[PubMed](#)]
118. Cattaneo, P.; Graeff, M.; Marhava, P.; Hardtke, C.S. Conditional effects of the epigenetic regulator JUMONJI 14 in *Arabidopsis* root growth. *Development* **2019**, *146*. [[CrossRef](#)] [[PubMed](#)]
119. Deng, W.; Liu, C.; Pei, Y.; Deng, X.; Niu, L.; Cao, X. Involvement of the Histone Acetyltransferase AtHAC1 in the Regulation of Flowering Time via Repression of FLOWERING LOCUS C in *Arabidopsis*. *Plant Physiol.* **2007**, *143*, 1660–1668. [[CrossRef](#)]
120. Li, W.; Liu, H.; Cheng, Z.J.; Su, Y.H.; Han, H.N.; Zhang, Y.; Zhang, X.S. DNA methylation and histone modifications regulate de novo shoot regeneration in *Arabidopsis* by modulating WUSCHEL expression and auxin signaling. *PLoS Genet.* **2011**, *7*, e1002243. [[CrossRef](#)] [[PubMed](#)]
121. Karami, O.; Rahimi, A.; Mak, P.; Horstman, A.; Boutilier, K.; Compier, M.; van der Zaal, B.; Offringa, R. An *Arabidopsis* AT-hook motif nuclear protein mediates somatic embryogenesis and coinciding genome duplication. *Nat. Commun.* **2021**, *12*. [[CrossRef](#)] [[PubMed](#)]
122. Jones-Rhoades, M.W.; Bartel, D.P.; Bartel, B. microRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **2006**, *57*, 19–53. [[CrossRef](#)] [[PubMed](#)]
123. Bartel, D.P. MicroRNAs: Target recognition and regulatory functions. *Cell* **2009**, *136*, 215–233. [[CrossRef](#)]
124. Reinhart, B.J.; Weinstein, E.G.; Rhoades, M.W.; Bartel, B.; Bartel, D.P. MicroRNAs in plants. *Genes Dev.* **2002**, 1616–1626. [[CrossRef](#)] [[PubMed](#)]
125. Wójcik, A.M.; Gaj, M.D. miR393 contributes to the embryogenic transition induced in vitro in *Arabidopsis* via the modification of the tissue sensitivity to auxin treatment. *Planta* **2016**, *244*. [[CrossRef](#)]
126. Wójcikowska, B.; Wójcik, A.M.; Gaj, M.D. Epigenetic regulation of auxin-induced somatic embryogenesis in plants. *Int. J. Mol. Sci.* **2020**, *21*, 2307. [[CrossRef](#)] [[PubMed](#)]
127. Wang, F.-X.; Shang, G.-D.; Wu, L.-Y.; Xu, G.-Z.; Zhao, X.-Y.; Wang, J.-W. Chromatin accessibility dynamics and a hierarchical transcriptional regulatory network structure for plant somatic embryogenesis. *SSRN Electron. J.* **2020**, 1–16. [[CrossRef](#)]
128. Schmidt, E.D.L.; Guzzo, F.; Toonen, M.A.J.; De Vries, S.C. A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *Development* **1997**, *124*, 2049–2062. [[CrossRef](#)]
129. Kumar, V.; Van Staden, J. Multi-tasking of SERK-like kinases in plant embryogenesis, growth, and development: Current advances and biotechnological applications. *Acta Physiol. Plant.* **2019**, *41*, 31. [[CrossRef](#)]
130. Hofmann, F.; Schon, M.A.; Nodine, M.D. The embryonic transcriptome of *Arabidopsis thaliana*. *Plant Reprod.* **2019**, *32*, 77–91. [[CrossRef](#)] [[PubMed](#)]

131. Jin, F.; Hu, L.; Yuan, D.; Xu, J.; Gao, W.; He, L.; Yang, X.; Zhang, X. Comparative transcriptome analysis between somatic embryos (SEs) and zygotic embryos in cotton: Evidence for stress response functions in SE development. *Plant Biotechnol. J.* **2014**, *12*, 161–173. [[CrossRef](#)]
132. Romani, F.; Moreno, J.E. Molecular mechanisms involved in functional macroevolution of plant transcription factors. *New Phytol.* **2020**, *230*, 1345–1353. [[CrossRef](#)] [[PubMed](#)]
133. Bui, L.T.; Pandzic, D.; Youngstrom, C.E.; Wallace, S.; Irish, E.E.; Szövényi, P.; Cheng, C.L. A fern *AINTEGUMENTA* gene mirrors *BABY BOOM* in promoting apogamy in *Ceratopteris richardii*. *Plant J.* **2017**, *90*, 122–132. [[CrossRef](#)] [[PubMed](#)]
134. Tao, Z.; Shen, L.; Gu, X.; Wang, Y.; Yu, H.; He, Y. Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. *Nature* **2017**, *551*, 124–128. [[CrossRef](#)] [[PubMed](#)]

5. Objectifs de la thèse

Le développement de la graine est soumis à un contrôle très fin par l'action d'hormones et la régulation de l'expression des gènes par des facteurs de transcription. La maturation, en particulier, est contrôlée par un groupe de facteurs de transcription, les LAFL, *i.e.* LEC1, ABI3, FUS3 et LEC2, décrits comme des régulateurs maîtres, dont le rôle est essentiel pour le développement de la graine. Lorsque l'un de ces facteurs est muté, la maturation ne s'effectue pas correctement et les graines n'acquièrent pas la capacité de tolérance à la dessiccation. À l'inverse, leur surexpression induit des phénotypes forts, comme l'accumulation ectopique de réserves, ou la formation spontanée d'embryons somatiques, induisant ainsi un « état graine », c'est-à-dire d'embryogenèse et de maturation, dans d'autres organes. Même si les LAFL ne sont pas impliqués directement dans l'embryogenèse précoce durant le développement normal de la graine, ils sont néanmoins nécessaires à l'initiation et au maintien du caractère embryogène des cellules végétales. Lorsqu'ils sont mutés, les plantes ne répondent plus aux protocoles d'embryogenèse somatique.

Ce projet de thèse a eu pour but de mieux comprendre la fonction d'un des membres des LAFL : LEC2 (LEAFY COTYLEDON 2). Ce facteur de transcription, membre de la famille des B3, comme ses partenaires ABI3 et FUS3, est un acteur essentiel du développement de la graine et des processus d'embryogenèse somatique. Cependant, certaines de ses fonctions et les mécanismes impliqués restent incompris. Des données préliminaires sur la séquence de LEC2, en lien avec le fait que sa surexpression induit des phénotypes très variables, notamment au niveau de la formation spontanée d'embryons somatiques, nous laissaient penser que ce facteur de transcription est soumis à de fortes régulations post-traductionnelles contrôlant son activité.

Une partie du projet a donc eu pour but de mettre en place un outil de mutagenèse *in planta*, ciblant spécifiquement le locus *LEC2*, de manière à induire des mutations ponctuelles aléatoires dans la séquence et, ainsi, affecter les modifications post-traductionnelles le ciblant. Cette approche de mutagenèse sans *a priori* menée sur *LEC2* a été réalisée en parallèle d'une approche complémentaire avec *a priori*. Cette seconde approche d'analyse des relations structure/fonction a eu pour objectif d'identifier et caractériser les différents domaines de LEC2 afin de mieux comprendre son mode d'action. Le travail réalisé a permis de mettre en évidence une possible nouvelle fonction de LEC2 comme facteur pionnier. Cette hypothèse a alors été testée dans le cadre d'une dernière partie.

Chapitre 1 :
Approche sans *a priori*

Chapitre 1 : Approche sans *a priori*

1. Introduction

La surexpression de *LEC2*, comme celle de *LEC1* et *FUS3*, peut induire la formation spontanée d'embryons somatiques sans induction hormonale chez *Arabidopsis* et d'autres espèces (Lotan et al., 1998; Stone et al., 2001; Liu et al., 2018). Cependant, ce phénotype est extrêmement rare, assez aléatoire et n'a été observé que quelques fois au laboratoire (B. Dubreucq et S. Baud, communication personnelle). La rareté de ce phénotype suggère que des mécanismes de régulation bloquent l'état embryogène induit par la surexpression de *LEC2*, ou que *LEC2* est suffisamment régulé pour ne pas induire cet état de manière constitutive. En effet, il a été montré que *LEC2* est soumis à de très fortes régulations à différents niveaux : transcriptionnel (sur son promoteur), épigénétique et post-traductionnel (Berger et al., 2011; Lepiniec et al., 2018; Duarte-Aké et al., 2019; Miller et al., 2019). En permettant l'expression de *LEC2* sous contrôle d'un promoteur constitutif comme le promoteur *CaMV35S* (Amack and Antunes, 2020), on s'affranchit des régulations transcriptionnelles et chromatiniennes exercées sur le promoteur endogène *pLEC2*. Cependant, le phénotype d'apparition spontanée des embryons somatiques restant rare et aléatoire, nous faisons l'hypothèse que *LEC2* est également soumis à de fortes régulations post-traductionnelles empêchant l'apparition d'un tel phénotype. En effet, la formation d'embryons sur toute la plante, comme observé dans de rares cas chez des surexprimeurs de *LEC2* est délétère pour la plante, qui ne peut pas assurer normalement son développement végétatif et pourrait donc se « protéger » en bloquant, une trop forte accumulation de *LEC2* dans les tissus.

Dans ce chapitre, j'ai testé cette hypothèse en mettant en place un outil de mutagenèse *in planta*, ciblée spécifiquement au locus *LEC2*, en contexte surexprimeur, afin de générer de nouveaux variants du facteur de transcription qui échapperaient à ces régulations post-traductionnelles. La recherche des variants a été réalisée en menant un crible sur la capacité des plantes à former des embryons somatiques de manière spontanée et de façon plus fréquente qu'observée jusqu'ici.

2. Résultats

2.1. Génération et validation des surexpresseurs de LEC2

2.1.1. Surexpression de LEC2*SH

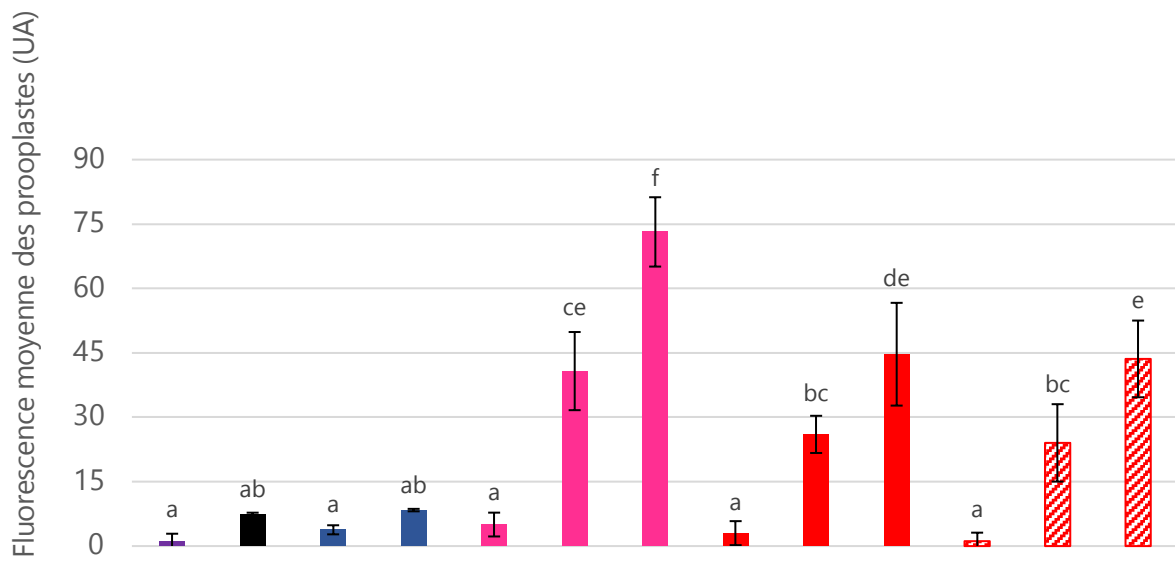
Le but de cette approche « sans *a priori* » a été de réaliser une mutagenèse aléatoire, ciblée spécifiquement à un locus *LEC2* transgénique. En effet, afin de s'affranchir d'effets délétères que pourrait causer la mutagenèse sur la fonction de LEC2, nous avons décidé de ne pas cibler la version endogène du facteur de transcription, mais une autre séquence, éditée, de *LEC2* notée *LEC2**, dont la séquence protéique est identique à la séquence endogène, mais dont la séquence nucléotidique a été modifiée de manière optimisée pour l'expression chez *Arabidopsis*. Un tag histidine (SH) est présent en partie C-terminale de la protéine, permettant une éventuelle quantification de l'accumulation de la protéine dans les plantes par western-blot. La séquence de *LEC2*SH* a été générée par C. Boulard, équipe SEEDEV avant mon arrivée au laboratoire. *LEC2*SH* a été placé sous contrôle d'un promoteur constitutif *pCaMV35S* permettant la surexpression ectopique du facteur de transcription et la réalisation d'un crible sur la capacité des plantes à former des embryons somatiques de manière spontanée. Une construction a également été réalisée avec une protéine GFP fusionnée en partie N-terminale de LEC2, cette fusion pouvant permettre de tester la localisation et/ou la stabilité du facteur de transcription avant et après mutagenèse.

La surexpression de cette version éditée en fond sauvage permet, ainsi, de réaliser un crible après mutagenèse sur la capacité des plantes à former des embryons somatiques, sans que la fonction de la copie endogène de LEC2 soit modifiée pour autant, ce qui pourrait affecter l'embryogenèse zygotique.

Des plantes *Arabidopsis thaliana* sauvages, écotype Col-0, ont donc été transformées avec deux constructions : *pCaMV35S::LEC2*SH* et *pCaMV35S::GFP:LEC2*SH* dans le but de surexprimer une version éditée de LEC2, localisable grâce à la GFP. Avant de soumettre ces plantes à la mutagenèse, la fonctionnalité de LEC2*SH a été testée en expression transitoire dans des protoplastes de mousses avec le promoteur de l'*OLEOSIN 1* (*pOLE1*) et l'effet de la surexpression de *LEC2*SH* et *GFP:LEC2*SH* dans les plantes a été observé.

2.1.2. Validation fonctionnelle de LEC2*SH dans la mousse

L'activité transcriptionnelle de la fusion protéique GFP:LEC2*SH a été testée dans un système d'expression transitoire dans des protoplastes de mousse *Physcomitrium patens*. Cette méthode, développée au sein de l'équipe SEEDEV, permet de tester l'activité d'un facteur de transcription, ici



<i>pOLE1::GFP</i>		+		+		+	+		+	+		+	+
AtLEC1			+	+			+			+			+
AtLEC2					+	+	+						
GFP:AtLEC2								+	+	+			
GFP:LEC2*SH											+	+	+

Figure 1-1 : La version éditée de LEC2, notée LEC2*SH est fonctionnelle dans les protoplastes de mousse. Mesure de la fluorescence moyenne de la GFP dans les protoplastes de mousses transfectés avec les différentes versions du facteur de transcription LEC2 (avec ou sans AtLEC1) et sa cible *pOLE1::GFP*. Les analyses statistiques ont été réalisées grâce à un test ANOVA avec correction de Welsh et un risque $\alpha=0,05$.

GFP:LEC2*SH, avec un promoteur cible, ici celui de l' *OLEOSIN 1* (*pOLE1*) fusionné à la *GFP* (Thévenin et al., 2012; Baud et al., 2016). La liaison du facteur de transcription sur le promoteur cible induit ainsi l'expression de la *GFP* dans les protoplastes, dont la fluorescence est détectée par cytométrie en flux.

L'activité du facteur de transcription GFP:LEC2*SH a été comparée à celle du facteur de transcription natif (AtLEC2), ainsi qu'à une fusion de la *GFP* au facteur de transcription sauvage (GFP:AtLEC2) validée précédemment par l'équipe et qui permet une restauration du mutant *lec2* (Boulard et al., 2018). De façon attendue, AtLEC2 est capable d'activer la transcription de la *GFP* placée sous contrôle de *pOLE1*, et en présence de AtLEC2 et AtLEC1, la fluorescence de la *GFP* est plus intense, en raison de l'effet coopératif de AtLEC1 sur l'activité transcriptionnelle de AtLEC2 (Baud et al., 2016). Les résultats (Figure 1-1) montrent que la fluorescence moyenne des protoplastes transfectés avec la combinaison *pOLE1::GFP* + *GFP:LEC2*SH* est moins forte que celle des protoplastes exprimant *pOLE1::GFP* + *AtLEC2*. Cependant, la fluorescence moyenne observée pour la combinaison *pOLE1::GFP* + *GFP:LEC2*SH* est équivalente à celle observée pour la combinaison *pOLE1::GFP* + *GFP:AtLEC2*. De plus, lorsque l'on co-transfecte *AtLEC1* avec *pOLE1::GFP* et *AtLEC2*, la fluorescence moyenne des protoplastes augmente significativement. On observe la même tendance pour les protoplastes transformés avec *pOLE1::GFP* + *AtLEC1* + *GFP:AtLEC2* ou *GFP:LEC2*SH*, bien que le niveau d'expression quantifié soit plus faible. Il apparaît donc que la version éditée LEC2*SH fusionnée à la *GFP* a une activité moins forte que celle de la protéine native AtLEC2 sur le promoteur de l' *OLEOSIN 1* en complexe ou non avec AtLEC1. Cependant, cette activité est du même ordre de grandeur que celle de la fusion GFP:AtLEC2. Or, il a déjà été montré que chez *Arabidopsis*, la fusion traductionnelle GFP:AtLEC2 sous contrôle du promoteur *pLEC2* est fonctionnelle et est capable de restaurer le phénotype sauvage en fond mutant *lec2* (Boulard et al., 2018). Il semble donc que la version éditée de LEC2, notée LEC2*SH, soit fonctionnelle, bien que sa séquence en acides nucléiques ait été modifiée.

2.1.3. Caractérisation des surexpresseurs de LEC2*SH in planta

2.1.3.1. Observations phénotypiques

Le développement des transformants primaires a été observé régulièrement et une centaine de plantules portant une des deux constructions ont été transférées en serre. Que ce soit en culture *in vitro* ou en serre, certaines plantes ont présenté des phénotypes anormaux, comme un défaut d'organisation du méristème apical caulinaire (MAC), l'apparition d'amas de cellules indifférenciées (cals) et à plus faible fréquence, l'apparition spontanée d'embryons somatiques (de l'ordre de 1 à

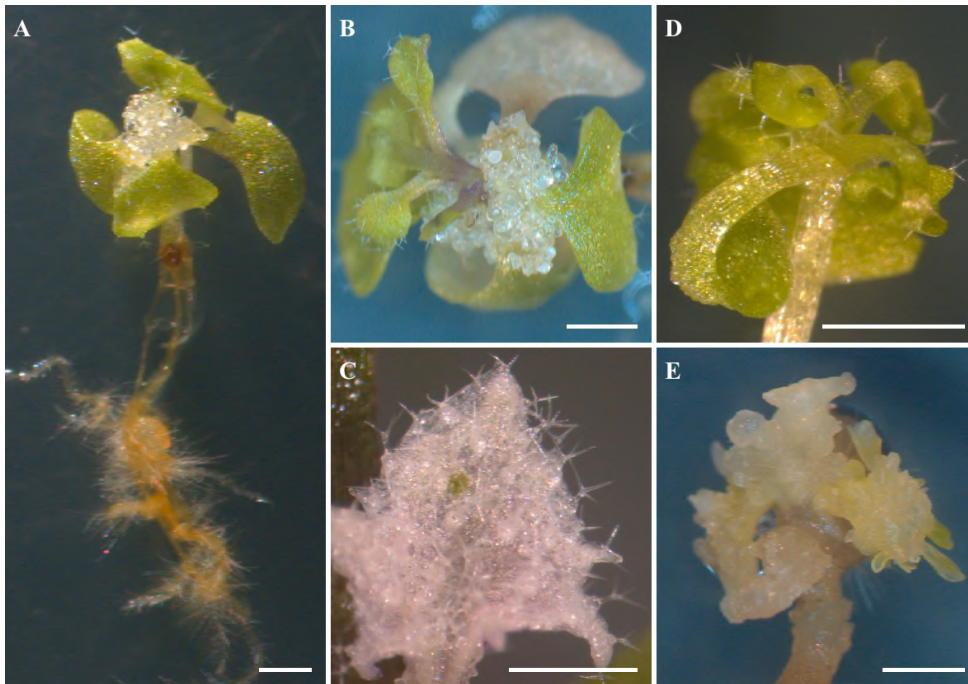


Figure 1-2 : Environ 1% des plantules surexprimant *LEC2*SH* cultivées *in vitro* présentent un phénotype. Les plantules *pCaMV35S::LEC2*SH* ont des phénotypes particuliers déjà décrits comme des racines épaisses et foncées pouvant former A-C) du cal translucide au niveau du méristème apical caulinaire, D) des méristèmes désorganisés ou encore E) des embryons somatiques formés spontanément à très faible fréquence. Les images ont été obtenues à la loupe binoculaire. Barres d'échelle : 1 mm.

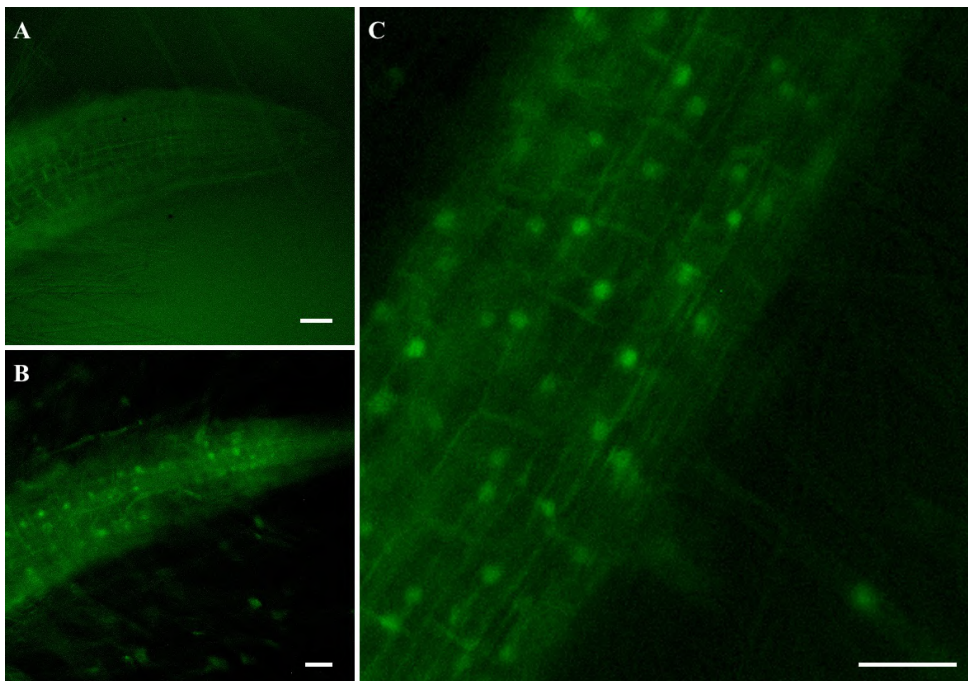


Figure 1-3 : GFP::LEC2 est localisé dans les noyaux des racines. Accumulation de la protéine GFP:LEC2*SH dans des plantules cultivées *in vitro* observée au microscope stéréoscopique à fluorescence. Les plantules semblent exprimer la GFP au niveau des noyaux des cellules racinaires. A) Plantule *pCaMV35S::LEC2*SH* (témoin négatif de fluorescence) ; B-C) plantules *pCaMV35S::GFP:LEC2*SH*. Barres d'échelle : 50 μ m.

2% des plantes observées – Figure 1-2). Les plantes cultivées en serre à un stade plus avancé ont également montré des phénotypes comme de la fasciation des tiges, des fleurs stériles, des feuilles à l'aspect vert plus foncé, plus épaisses et brillantes que ce qui est observé d'ordinaire chez *Arabidopsis thaliana*, ressemblant à des feuilles de plantes grasses (Figure 1-4). Ces phénotypes sont conformes à ceux déjà observés (Stone et al., 2001; Santos Mendoza et al., 2005) et attendus, compte tenu du grand nombre de plantes manipulées, confirmant la probable fonctionnalité de LEC2*SH avec ou sans la fusion avec la GFP.

2.1.3.2. Localisation de la GFP dans les plantules

Afin de confirmer l'expression de LEC2*SH, les plantules cultivées *in vitro* exprimant *pCaMV35S::GFP:LEC2*SH* ont été observées en microscopie stéréoscopique à fluorescence (Axiozoom). De la fluorescence liée à la GFP a été détectée en particulier au niveau des racines, possiblement dans les noyaux des cellules ainsi que dans ceux des poils absorbants (Figure 1-3). La localisation nucléaire n'a pas pu être confirmée grâce à une co-localisation avec le DAPI colorant le noyau, car les plantules étaient observées en culture *in vitro* et nous ne voulions pas en sacrifier.

2.1.3.3. Expression de *GFP:LEC2*SH* dans les plantes

Des extractions d'ARNm et des analyses de RT-PCR ont également été réalisées sur des plantes *pCaMV35S::LEC2*SH* présentant une variation de phénotype (Figure 1-4A). Les analyses de RT-PCR ont montré que *LEC2*SH* est bien exprimé chez les plantes transgéniques et pas dans la plante sauvage Col-0 (Figure 1-4B, C). Une RT-PCR a également été réalisée avec des amorces amplifiant la copie native *AtLEC2* à une jonction intron-exon confirmant que les échantillons n'étaient pas contaminés par de l'ADN génomique (Figure 1-4D). Une légère amplification est détectée dans les plantes transformées, mais le témoin positif d'ADN génomique présente une bande de taille plus grande correspondant à la copie endogène avec un intron. La bande détectée dans les plantes *pCaMV35S::LEC2*SH* correspond probablement à l'hybridation résiduelle des amorces spécifiques de *AtLEC2* sur l'ADN rétro-transcrit de *LEC2*SH*.

2.1.3.4. Conclusion

Les observations effectuées apportent une preuve de la fonctionnalité de *pCaMV35S::LEC2*SH* et *pCaMV35S::GFP:LEC2*SH*. En effet, des phénotypes de croissance déjà décrits ont été observés et la protéine GFP:LEC2*SH est détectée dans les racines, dans ce qui semble être les noyaux. De plus, l'analyse de l'activité transcriptionnelle en protoplastes de mousses a montré que GFP:LEC2*SH est capable de se lier à sa cible *pOLE1* et d'activer l'expression de la GFP. Malgré le grand nombre de plantes observées (plusieurs centaines), seulement une plantule a montré l'apparition d'embryons

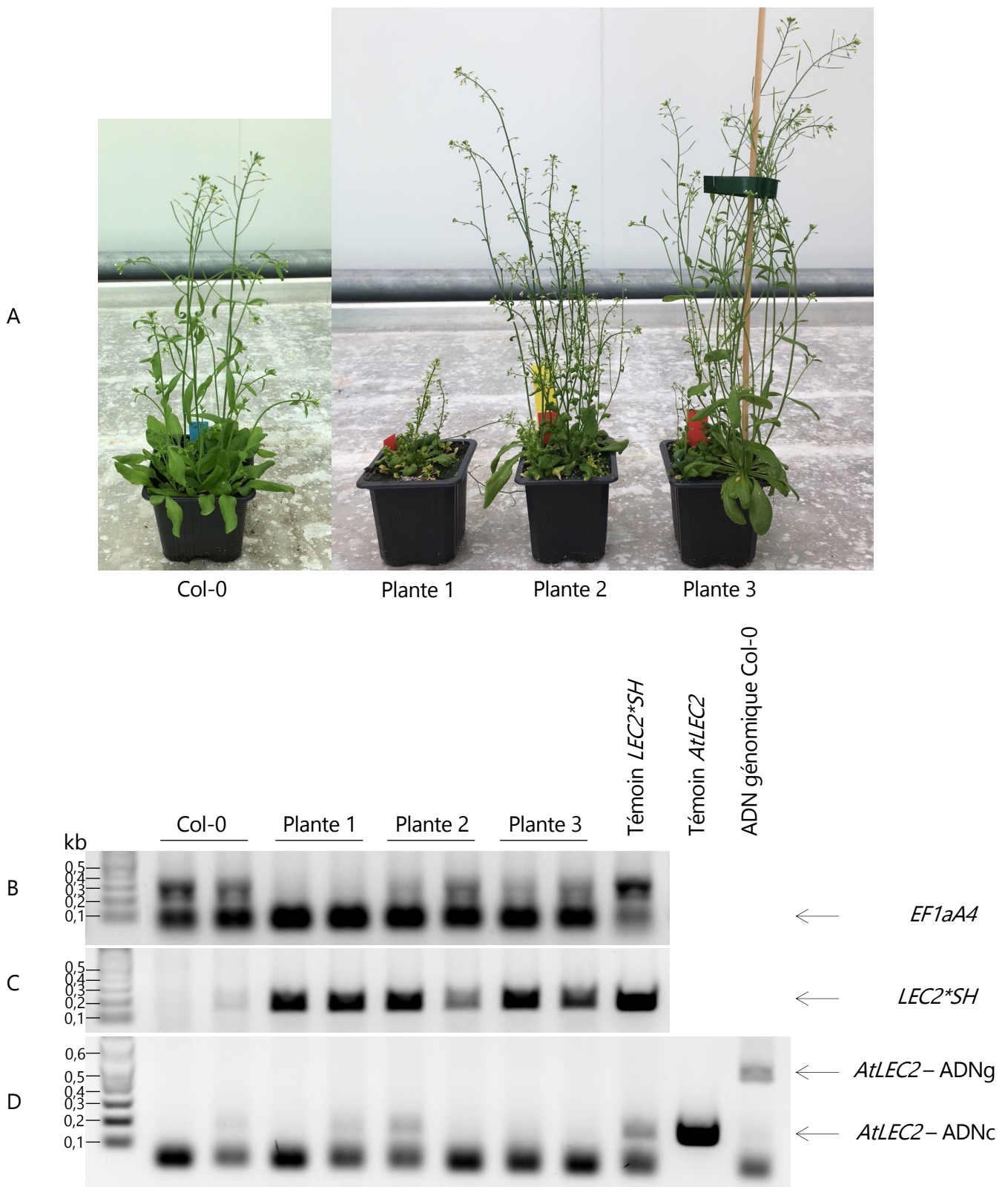


Figure 1-4 : Vérification de l'expression de *LEC2*SH* dans des plantes montrant une gradation de phénotype. A) Phénotypes des plantes *pCaMV35S::LEC2*SH* utilisées pour la mutagenèse CRISPR-X. B-D) Analyse en RT-PCR de l'accumulation du transcrite *LEC2*SH*. B) *EF1aA4* (témoin positif) ; C) *LEC2*SH* ; D) *AtLEC2*.

somatiques (Figure 1-2E), nous confirmant que ce phénotype surexprimeur est rare et aléatoire et nous encourage à générer de nouveaux variants de *LEC2*, qui pourraient avoir une meilleure capacité à induire l'embryogenèse somatique spontanée. Ces plantes constituent le matériel génétique nécessaire à la réalisation de notre crible pour mettre en évidence des phénotypes dominants, de gain de fonction, dans des plantules après mutagenèse du locus transgénique.

2.2. Choix d'un outil de mutagenèse *in planta*

Le but de l'approche de mutagenèse est d'induire des mutations aléatoires, mais ciblées spécifiquement au locus *LEC2* (ou précisément *LEC2*SH* dans notre cas) et ainsi générer de nouveaux variants du facteur de transcription présentant un gain de fonction, *i.e.* la formation d'embryons somatiques de manière spontanée et plus fréquente. Notre objectif était d'utiliser un système de mutagenèse qui, comme une mutagenèse chimique de type EMS, pourrait induire des mutations ponctuelles sur la séquence d'ADN mais ici uniquement ciblées au locus *LEC2* et non sur le génome entier.

La technologie CRISPR-Cas9 a été adaptée d'un système de défense immunitaire, présent chez les procaryotes et les archées, permettant de cliver des séquences d'ADN de phages (Wiedenheft et al., 2012). Elle implique une enzyme, la CRISPR-Cas9, capable de reconnaître une séquence d'ADN, ciblée grâce à un ARN guide, comportant une séquence spécifique PAM (*Protospacer Adjacent Motif*) (Jinek et al., 2012). Généralement la séquence PAM reconnue par la CRISPR-Cas9 est une séquence de type *NGG*, où *N* peut être n'importe quel nucléotide. La CRISPR-Cas9 se positionne sur l'ADN ciblé par l'ARN guide et induit une coupure double brin. Cette coupure est alors réparée par la machinerie de réparation de l'ADN qui peut induire des erreurs, généralement des insertions ou des délétions. Le système CRISPR-Cas9 a été adapté pour l'utilisation chez les eucaryotes et est maintenant un outil de mutagenèse incontournable y compris en biologie végétale, car il permet de produire des lignées mutantes bien plus rapidement que les systèmes de mutagenèse classique et contribue à générer efficacement de la variabilité génétique (Nogué et al., 2016).

Comme le système CRISPR-Cas9 permet d'introduire des insertions ou des délétions, les mutations générées conduisent généralement à un changement de phase, l'apparition d'un codon STOP prématuré et ainsi une perte de fonction par *knock-out* (KO) du gène ciblé. Depuis que CRISPR-Cas9 a été mis au point, de nombreuses évolutions ont été faites, afin de l'améliorer ou de générer différents types de modifications des cibles, en utilisant des mutants de l'enzyme ayant une activité différente, ou en fusionnant la Cas9 avec d'autres enzymes (Zhang et al., 2019). On peut, par exemple, induire des mutations ponctuelles ciblées en fusionnant une enzyme nCas9 (nickaseCas9),

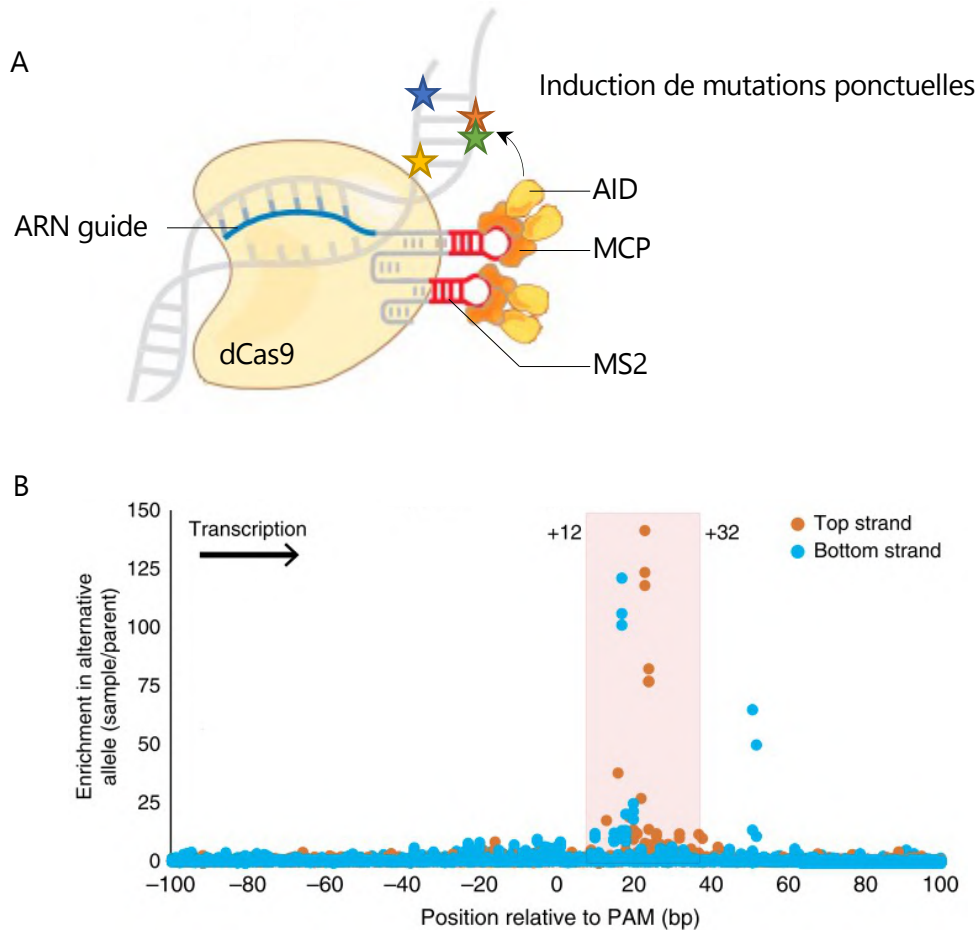


Figure 1-5 : Schématisation du système de mutagenèse CRISPR-X, capable d'induire des mutations ponctuelles aléatoires sur une fenêtre de 100 pb autour du locus. A) Le système CRISPR-X est composé d'une enzyme CRISPR-dCas9, d'un ARN guide contenant deux structures en épingle à cheveux *MS2* pour recruter une cytidine déaminase (AID) fusionnée à une protéine liant *MS2* (MCP). B) Enrichissement en mutations ponctuelles par rapport à la position de la séquence PAM des ARN guides chez l'animal. La flèche indique le sens de transcription.

Modifié d'après Konermann et al., 2015 (A) et Hess et al., 2016 (B).

capable de générer des coupures simple brin sur l'ADN, ou dCas9 (deadCas9), catalytiquement inactive, avec une cytidine ou une adénine déaminase. L'utilisation d'une cytidine ou d'une adénine déaminase combinée avec une enzyme Cas9, nCas9 ou dCas9 permet la modification ponctuelle et irréversible d'une cytosine (C) ou d'une adénine (A) en uracile (U) reconnue comme une thymine (T) par la machinerie de réplication. Cette approche d'édition de base (*base editing*) a été validée avec différents types d'enzymes et chez différentes espèces, que ce soit chez les animaux ou les plantes, par exemple le riz, la tomate, le maïs, Arabidopsis, la mousse, etc. (Shimatani et al., 2017; Zong et al., 2017; Gaillochet et al., 2020).

Nous avons voulu tirer parti des récentes avancées technologiques basées sur le système de mutagenèse CRISPR-Cas9 pour réaliser une mutagenèse sans *a priori* ciblée spécifiquement au locus *LEC2*SH*, notamment le système « CRISPR-X » développé par Hess et al. en 2016.

2.3. Mise en place de l'approche « CRISPR-X »

2.3.1. Présentation de la méthode

CRISPR-X (Hess et al., 2016) est un outil de mutagenèse aléatoire, ciblée à un locus donné, mettant en jeu une enzyme dCas9 (catalytically deadCas9) catalytiquement inactive, *i.e.* incapable de couper l'ADN (Qi et al., 2013; Komor et al., 2016), une cytidine déaminase (AID) (Ren et al., 2018) fusionnée à une protéine MCP (*MS2 COAT PROTEIN*), qui reconnaît spécifiquement des structures en double tige-boucle *MS2* présentes sur les ARN guides (Li et al., 2017) (Figure 1-5A). À travers ce système, la cytidine déaminase est capable d'induire des mutations ponctuelles, localisées au locus d'intérêt, sur une fenêtre d'une centaine de paires de bases autour de l'ARN guide (Figure 1-5B). Les mutations ponctuelles pouvant être générées sont d'une grande variété de substitutions : C>T, mais aussi G>A, G>C, C>A... CRISPR-X est un outil de mutagenèse très efficace dans les cellules animales mais il n'a jamais été testé dans un organisme végétal. Malgré le risque associé à l'utilisation de cette méthode très innovante chez les plantes, nous avons choisi de tester CRISPR-X car la diversité des mutations générées et la fenêtre de mutations autour de l'ARN guide conviennent parfaitement à notre approche de mutagenèse aléatoire *in planta* sans *a priori*.

2.3.2. Choix des ARN guides

Plusieurs domaines fonctionnels de LEC2 ont été caractérisés : le domaine B3 de fixation à l'ADN, le domaine B2 de co-activation ou de co-régulation et deux domaines PEST prédits, qui sont des séquences peptidiques riches en prolines, acides glutamiques, serines et thréonines, servant de signal protéolytique nécessaire à la dégradation par le protéasome (Rechsteiner and Rogers, 1996;

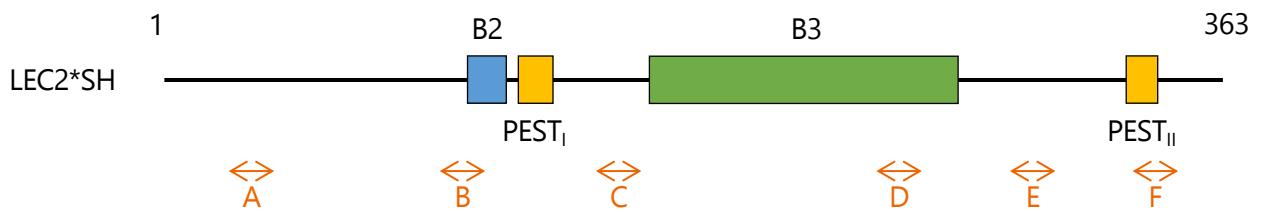


Figure 1-6 : Représentation schématique de la structure de LEC2 avec ses différents domaines et position des ARN guides. Le domaine B3 de liaison à l'ADN est représenté en vert et le domaine B2 d'interaction protéine-protéine en bleu. Les domaines PEST prédits sont indiqués en jaune. Les ARN guides sont représentés par les flèches oranges.

Maruri-López et al., 2014). Des données récentes générées dans le cadre de mon projet ont également mis en évidence des domaines nécessaires à l'activité du facteur de transcription sur le promoteur de l'*OLEOSIN 1* dans des protoplastes de mousse, notamment dans les parties N et C-terminales (voir [Chapitre 2](#)). Partant du principe que le système CRISPR-X est capable d'induire des mutations, chez l'animal, sur une fenêtre allant jusqu'à 100 pb autour de l'ARN guide, 6 ARN guides ont été définis sur la base de leur position le long de la séquence de *LEC2*SH* et d'un faible nombre de sites potentiels « hors-cible » (*off-targets*) possibles dans le génome d'*Arabidopsis thaliana*. Le choix des ARN guides a été réalisé de manière à couvrir la plus grande partie possible du gène et, notamment, des zones potentiellement soumises à une régulation (prédictions de modifications post-traductionnelles, d'interaction protéine-protéine...). La structure globale de *LEC2* et la position des ARN guides sont schématisées dans la Figure 1-6. Les ARN guides ont été dessinés en utilisant le site internet CRISPOR (<http://crispor.tefor.net/crispor.py>), voir [Matériel et méthodes](#). Les séquences sont données en Annexe 1.

Deux ARN guides ont été définis en partie 5' du gène : l'ARN guide A débute en position 89 après l'ATG et couvre la partie N-terminale de *LEC2*. L'ARN guide B débute en position 307 et couvre le domaine B2 de co-activation ou de co-régulation de *LEC2*, qui permet notamment son interaction avec le facteur de transcription *LEC1*. Les ARN guides C et D sont respectivement positionnés à partir des bases 471 et 759 et couvrent en partie le domaine B3 de fixation à l'ADN. Dans la partie 3' du gène, l'ARN guide E, débutant 895 pb après l'ATG est situé entre le domaine B3 et le deuxième domaine PEST. Enfin, l'ARN guide F couvre le domaine PEST et l'extrémité 3' de *LEC2*.

2.3.3. Transformation des plantes par le système CRISPR-X et crible

La technique de clonage *GoldenBraid* (Sarrion-Perdigones et al., 2011, 2013), dont le principe est détaillé dans le Matériel et méthodes, permet de combiner facilement plusieurs unités transcriptionnelles dans un même vecteur et, ainsi, de transformer des plantes avec un unique plasmide porteur de toutes les unités transcriptionnelles nécessaires, facilitant la sélection, la stœchiométrie des surexpressions et limitant les problèmes de mutations induites par l'insertion de plusieurs ADN-T. Nous avons donc mis en place une stratégie basée sur le système *GoldenBraid* pour cloner les différentes unités composant le système CRISPR-X dans un seul et même vecteur. Des plantes surexprimant *LEC2*SH* (voir partie [2.1. Génération et validation des surexprimeurs de *LEC2*](#)) ont été transformées avec un plasmide comportant :

- les ARN guides A et B sous contrôle du promoteur *pU6-26* (Vazquez-Vilar et al., 2016) et fusionnés avec un Scaffold2MS2 pour former une structure en double tige-boucle ;

- la dCas9 sous contrôle du promoteur constitutif *pRPS5A* (Tsutsui and Higashiyama, 2016) ;
- la fusion MCP:AID sous contrôle du promoteur constitutif *pAtUbq10* ;
- la DsRed sous contrôle du promoteur *pCsVMV* permettant la sélection des transformants primaires.

Pour le premier essai de CRISPR-X, j'ai transformé les plantes avec une combinaison de deux ARN guides (A et B) sur les 6 générés, afin de ne pas saturer le gène de mutations dans le cas où la mutagenèse serait très efficace.

Le crible envisagé pour détecter les plantes comportant des mutations étant la formation d'embryons somatiques à forte fréquence à 10-15 jours de culture sans induction hormonale, environ 230 transformants primaires ainsi sélectionnés ont été semés sur milieu Arabidopsis et cultivés *in vitro*. Après environ un mois, aucune apparition d'embryons somatiques n'a été observée sur les plantules à part une. Le crible n'a donc pas permis de mettre en évidence de nouveaux variants de LEC2 dont la surexpression induit l'apparition d'embryons somatiques à une plus forte fréquence.

Afin de vérifier si des mutations ont été induites par CRISPR-X mais sans faire apparaître de phénotype, un grand nombre de plantes a été séquençé au locus *LEC2*SH*. La méthode de séquençage a permis d'obtenir plus de 3 623 000 séquences de 100 pb sur 1 kb de gène. Nous pensions que l'ensemble des séquences mutées et non mutées donnerait un corpus de données permettant des comparaisons, nous n'avons donc pas réalisé de contrôle négatif en séquençant au même moment l'ADN de plantules non soumises à mutagenèse. Malgré l'absence de ce contrôle solide, les résultats de séquençage ont été analysés avec le logiciel *CLC Genomic Workbench* (Qiagen) en comparaison avec la séquence de référence de *LEC2*SH*. L'analyse rapide n'a pas permis d'identifier de mutation. En parallèle de nos analyses, l'équipe de Fabien Nogué (Réparation de l'ADN et ingénierie des génomes, IJPB) a testé mes constructions du système CRISPR-X dans la mousse *Physcomitrium patens* (Collonnier et al., 2017) et a montré qu'aucune mutation n'était induite par la mutagenèse dans ce système.

En absence de phénotype observé lors du crible, de mutation détectée en première approche par analyse des résultats de séquençage profond, et devant l'inefficacité du système dans la mousse, nous avons décidé d'interrompre l'approche CRISPR-X.

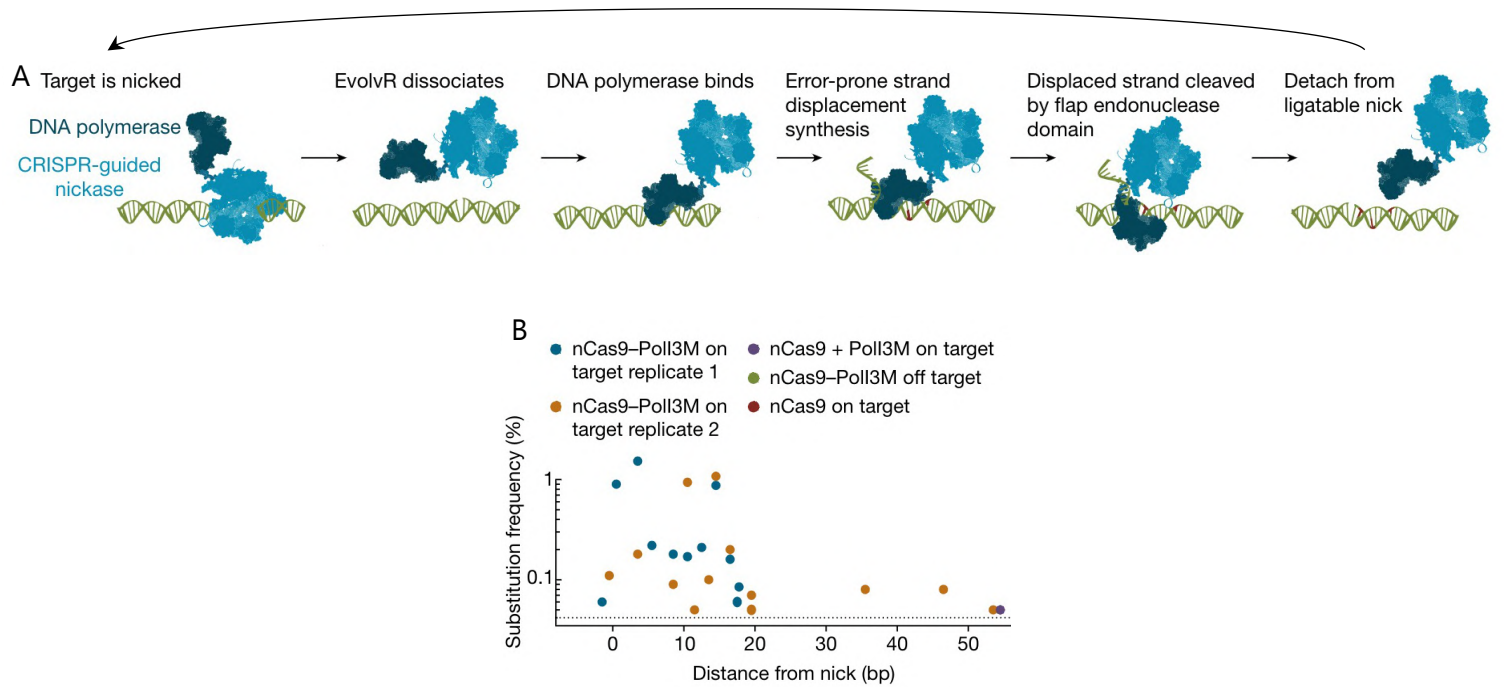


Figure 1-7 : Schématisation du système de mutagenèse EvolvR. A) Le système EvolvR est composé d'une CRISPR-nCas9 fusionnée à une ADN polymérase infidèle. L'association du complexe EvolvR sur sa cible permet l'induction d'une coupure simple brin par la CRISPR-nCas9. L'ADN polymérase infidèle répare la coupure en induisant des mutations et clive l'ancien brin. B) La fusion entre la nCas9 et la polymérase permet d'induire des substitutions de nucléotides sur une fenêtre de 20 pb autour de l'ARN guide à forte fréquence (points bleus et oranges), contrairement à lorsque les enzymes ne sont pas fusionnées (violet).

Modifié d'après Halperin et al., 2018.

2.3.4. Discussion mutagenèse CRISPR-X

La méthode de mutagenèse CRISPR-X semblait idéale pour notre approche de mutagenèse sans *a priori*, *in planta*, ciblée aléatoirement au locus *LEC2*SH*, cependant, il semble que cette technique ne soit pas efficace dans un système végétal, du moins pas chez *Arabidopsis thaliana* ou *Physcomitrium patens*. Bien que les différents éléments composant CRISPR-X aient tous été montrés comme étant fonctionnels chez les plantes, la dCas9 d'une part (Piatek et al., 2015), la cytidine déaminase (AID) d'autre part (Ren et al., 2018) ou encore le complexe MCP-MS2 (Li et al., 2017), leur association ne conduit pas, du moins dans notre cas, à l'induction de mutations *in vivo*. De plus, il semble que la dCas9 utilisée ici ne soit pas la plus efficace chez les plantes. En effet, une version de la dCas9 avec des introns (dCas9i) est plus stable et plus efficace pour la liaison à ses cibles et le recrutement de partenaires (A. Dubois, RDP Lyon, communication personnelle). Il est possible que la structure de la fusion traductionnelle ne permette plus l'activité enzymatique ou encore que les mécanismes de réparation des génomes végétaux après action de la déaminase soient plus efficaces et ne permettent pas que les mutations soient transmises. Ce point sera discuté ci-après. Comme la méthode de mutagenèse CRISPR-X ne semble pas efficace chez les plantes, nous avons décidé de mettre en place une nouvelle stratégie de mutagenèse, tout aussi innovante.

2.4. Approche « EvolvR »

2.4.1. Présentation de la méthode

La méthode de mutagenèse EvolvR (Halperin et al., 2018) consiste en l'utilisation d'une nCas9, ayant une activité de coupure simple brin sur l'ADN cible, fusionnée à une ADN polymérase mutée, PolI3M, dite « infidèle », car elle induit des mutations au moment de la polymérisation (Figure 1-7). Ainsi, la nCas9 se lie à l'ADN ciblé grâce aux ARN guides, elle induit une coupure simple brin, se dissocie et la polymérase se lie à l'ADN clivé et synthétise un nouveau brin tout en incorporant des erreurs de nucléotides. L'ancien brin est clivé, puis le complexe nCas9:PolI3M, ou EvolvR, se détache. Tant que la nCas9 est capable de reconnaître sa cible grâce aux ARN guides, EvolvR peut agir et provoquer des mutations au locus d'intérêt. Comme CRISPR-X, EvolvR n'a pas été développé chez les plantes mais chez les bactéries. La nCas9 est fonctionnelle chez les plantes et est utilisée notamment dans les systèmes de *base editing* avec une cytidine ou une adénine déaminase comme évoqué ci-dessus. Ne sachant pas si la PolI3M, qui est une ADN polymérase bactérienne, est efficace chez les plantes, nous avons décidé la tester en comparaison avec une ADN polymérase d'*Arabidopsis*, REV1 (AT5G44750) (Takahashi et al., 2005). Avant de transformer EvolvR chez *Arabidopsis* dans des plantes surexprimant *LEC2*SH*, nous avons décidé de tester le système que

j'ai cloné dans la mousse *Physcomitrium patens* avec l'équipe de Fabien Nogué. Malheureusement, aucune des deux polymérasés mutées, Poll3M et REV1, n'a permis d'induire un nombre de mutations significatif dans la mousse pour répondre à nos critères.

2.4.2. Discussion mutagenèse EvolvR

Comme CRISPR-X, le système EvolvR est composé de différents éléments, dont chacun indépendamment est fonctionnel dans les systèmes végétaux. Cependant, il semble, encore une fois, que EvolvR ne soit pas capable d'induire un grand nombre de mutations, ou, du moins, que nous ne puissions pas détecter un grand nombre de mutations après mutagenèse. Nous ne savons pas comment expliquer que ces outils soient incapables d'induire des mutations chez les végétaux. Une hypothèse pour expliquer cette inefficacité, serait que les systèmes de réparation des végétaux sont « trop efficaces » pour que des mutations soient conservées et transmises aux cellules filles. Nous pouvons penser que CRISPR-X, comme EvolvR permet d'accumuler des mutations au locus ciblé, mais que celles-ci sont détectées très rapidement par la machinerie de réparation de l'ADN et rectifiées avant que nous puissions les identifier. Un autre élément nous poussant à croire que ces systèmes de mutagenèse ne sont pas efficaces chez les plantes, contrairement aux cellules animales ou chez les procaryotes, est que ces méthodes très innovantes n'ont jamais encore été publiées chez les végétaux. Bien que les articles présentant CRISPR-X et EvolvR datent respectivement de 2016 et 2018, aucune publication ne propose une adaptation ou une preuve de concept chez les organismes végétaux, alors que des groupes très actifs travaillent spécifiquement sur ces outils de mutagenèse très prometteurs, partout dans le monde.

3. Discussion et dernière approche envisagée

Étant donné que les systèmes de mutagenèse CRISPR constituent un outil pour mon projet de thèse et non un projet de recherche à part entière, nous avons décidé de ne pas poursuivre ce travail de compréhension et de mise au point d'une mutagenèse basée sur des systèmes CRISPR dans le cadre de mon doctorat.

La fusion entre une nCas9 ou une dCas9 avec une cytidine ou une adénine déaminase, forme un système de mutagenèse ciblée très efficace de *base editing* chez les plantes, on parle alors de CBE ou de ABE pour *cytidine* ou *adenine base editors* (Shimatani et al., 2017; Zong et al., 2017; Li et al., 2019; Guyon-Debast et al., 2021). Afin de poursuivre notre approche sans *a priori* visant à générer de nouveaux variants de LEC2 par mutagenèse aléatoire, nous avons envisagé d'utiliser un système de ce type. Les CBE/ABE permettent d'induire des mutations ponctuelles de manière très précise et

ciblée. Les mutations induites sont définies par les ARN guides et sont générées dans une fenêtre très réduite en amont ou en aval du PAM (*Protospacer Adjacent Motif*). Dans le cadre d'une mutagenèse ciblée au locus *LEC2*SH*, l'utilisation de *base editors* aurait permis d'induire des mutations, mais avec *a priori*, en fonction d'un nombre fini de localisations des PAM dans la séquence de *LEC2*. D'autres enzymes nCas9 ont été générées et permettent d'augmenter le nombre de cibles grâce à des PAM moins restrictifs de type NG et non NGG (Nishimasu et al., 2018; Veillet et al., 2020). Il faudrait dans ce cas utiliser un grand nombre d'ARN guides pour induire des mutations, qui finalement seraient assez peu nombreuses sur la séquence de *LEC2*SH*. En effet, le nombre de motifs *NG* ou *NGG* dans la séquence de *LEC2* est assez réduit et ne permettrait pas d'une part de couvrir l'ensemble de la séquence de mutations et d'autre part, fait perdre l'intérêt d'une induction de mutations aléatoires et sans *a priori* sur *LEC2*.

Dans l'état actuel de nos connaissances des systèmes de mutagenèse CRISPR, aucun outil ne nous permettrait de générer des variants aléatoires de *LEC2* de manière assez efficace et sans *a priori* chez les plantes. Nous avons donc décidé de ne pas poursuivre dans cette voie et avons envisagé une autre manière de mettre en place notre recherche de variants de *LEC2* induisant la formation d'embryons somatiques. Une possibilité serait de créer une banque de variants de *LEC2* cette fois non pas par mutagenèse *in planta*, mais *in silico*. En effet, plusieurs entreprises de synthèse à façon proposent maintenant des services de génération de banques de plusieurs millions de variants. Les types de mutations sont à définir selon la question biologique posée. Il pourrait s'agir de banques de variants comportant chacun un acide aminé muté par séquence, une combinaison des 22 acides aminés possibles sur chaque position le long de la séquence à partir de 2 400 \$ pour 48 variants. On peut également avoir une combinatoire de mutations synthétisées sur 1 kb de séquence avec une banque de 10^8 variants à partir de 3 500 \$ ou encore une banque combinatoire de variants pouvant avoir chacun des 22 acides aminés à chacune des positions. Cette dernière option permet de générer plus de 10^{10} variants pour un coût minimum de 10 000 \$. Ces différentes stratégies permettraient alors de :

- recevoir une banque de mutants *mLEC2*,
- la cloner dans un vecteur d'expression sous contrôle d'un promoteur fort avec ou sans fusion avec la GFP,
- transformer des plantes sauvages
- réaliser un crible reposant sur l'apparition d'embryons somatiques en culture *in vitro* des plantules transformants primaires.

Les plantules présentant des embryons somatiques seraient alors séquencées au locus *mLEC2* afin d'identifier les mutations présentes sur la séquence. Si une ou plusieurs mutation est détectée au locus, alors, une nouvelle séquence de LEC2 sera générée par synthèse à façon, afin d'étudier plus en détails les effets de la mutation, notamment sur la capacité des plantes à induire l'embryogenèse somatique en contexte surexprimeur, sous contrôle du promoteur endogène, mais aussi les éventuels effets sur le développement de la graine. Après discussion de la faisabilité de cette approche lors de mon comité de suivi de deuxième année de thèse, bien que la technologie soit disponible et les tarifs abordables, nous nous sommes rendus à l'évidence qu'il était trop tard dans mon projet pour mettre en place une banque de mutants et aller jusqu'à l'identification et l'analyse des mutations dans la durée restante de mon contrat doctoral. C'est pourquoi, nous avons décidé d'arrêter définitivement l'approche de recherche de variants de LEC2 par mutagenèse sans *a priori* dans le cadre de ma thèse. Cependant, il est possible que ce travail soit poursuivi par quelqu'un d'autre de l'équipe, car nous restons persuadés qu'une telle approche, sans *a priori*, en générant un grand nombre de variants, permettrait de mieux comprendre les mécanismes qui permettent à LEC2 de réguler à la fois le développement de la graine et l'embryogenèse somatique. Ce type d'approche a déjà été développé, notamment sur l'histone H4 (Corcoran et al., 2022) : des plantes ont été transformées par un système CRISPR qui mute le gène d'intérêt et, en même temps, le remplace par un variant généré aléatoirement. Ce système a permis d'identifier de nouveaux variants et d'étudier leur fonction sur la régulation de la floraison. L'approche reste donc intéressante.

4. Quel promoteur pour surexprimer *LEC2* ?

Cette partie sans *a priori* du projet de thèse reposait sur le fait que malgré la publication de l'apparition d'embryons somatiques chez des surexprimeurs de *LEC2* (Stone et al., 2001), ce phénotype n'est apparu que très rarement au cours des différents travaux menés au sein de l'équipe. Cependant, le promoteur utilisé pour surexprimer *LEC2*, était celui du *CaMV35S* utilisé en routine au laboratoire et dans d'autres équipes mais que nous savions sensible à des régulation épigénétiques (Amack and Antunes, 2020). J'ai donc généré, des plantes surexprimant *LEC2* mais cette fois sous contrôle d'un autre promoteur dit « constitutif », celui du gène *AtUbg10* (AT4G05320) (Sarrion-Perdigones et al., 2013).

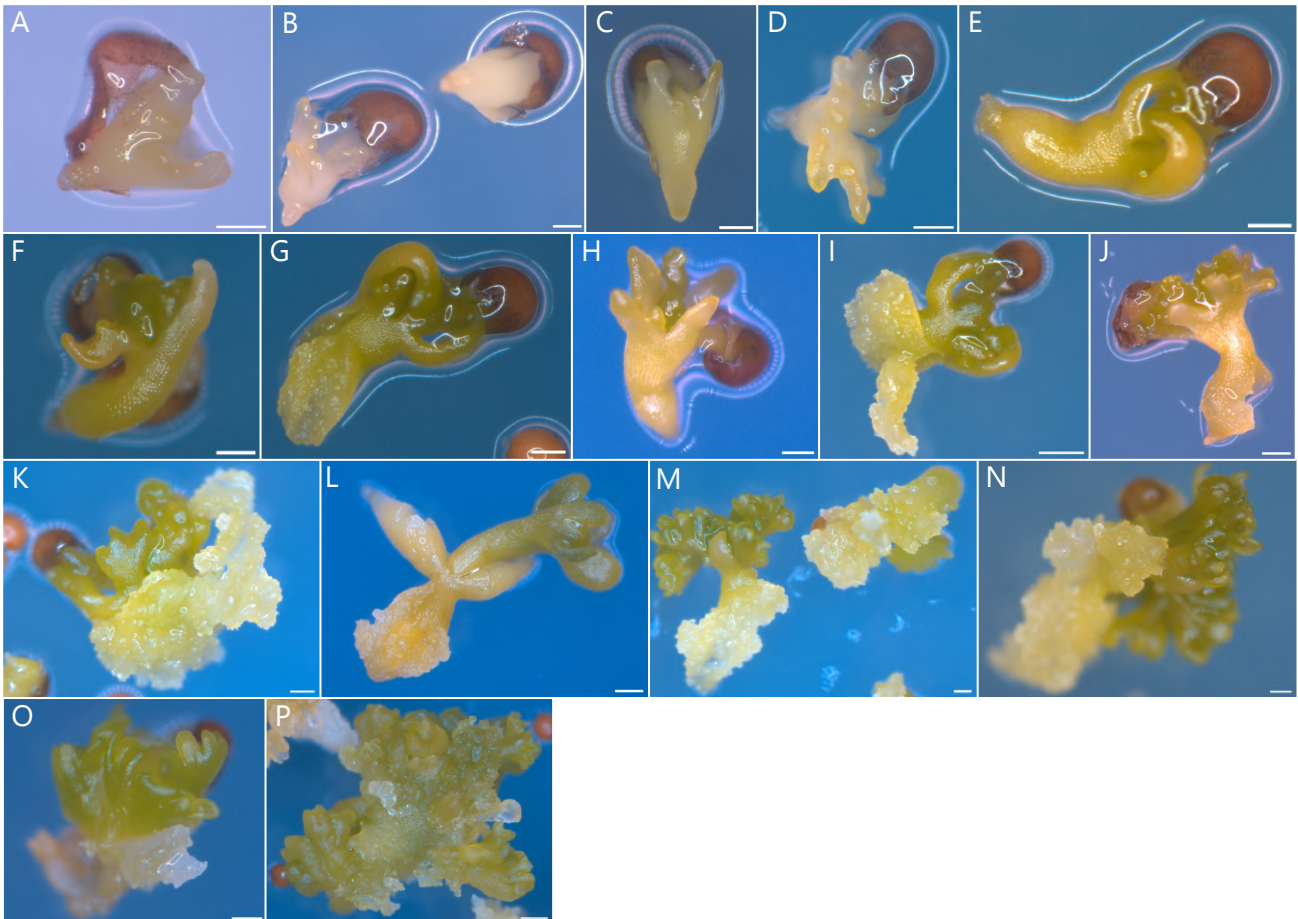


Figure 1-8 : La surexpression de *LEC2* sous contrôle du promoteur *pAtUBQ10* en fond mutant *lec2-1* induit des phénotypes anormaux dès la germination, ainsi que le développement d'embryons somatiques et de cal sur les plantules. Plantules cultivées *in vitro* A-C) J+5, D-E) J+7, F-G) J+10, H-J) J+12, K) J+17, L-M) J+21, N-O) J+24, P) J+33 après germination. Barres d'échelle : 200 μ m.

4.1. Phénotypes surexpresseurs de *AtLEC2* sous contrôle de *pAtUbq10*

De manière très intéressante et étonnante, la totalité des transformants primaires obtenus (plus de 60) avec une construction *pAtUbq10::GFP:AtLEC2:tRbcSE9* en fond mutant *lec2-1* ont donné des phénotypes anormaux pendant la croissance de plantules cultivées *in vitro* sur un milieu de culture classique, sans ajout d'hormone et ce, dès la germination (Figure 1-8). Les plantules ont montré un développement végétatif extrêmement perturbé, avec une agglomération d'embryons somatiques et/ou de cal, comme décrit dans la littérature dans le cas de surexpresseurs de LEC1 ou LEC2 (Lotan et al., 1998; Stone et al., 2001). Un seul transformant primaire portant la même construction en fond sauvage (Ws) a pu être observé et a également donné un phénotype anormal. De nouvelles plantes de génotypes *lec2-1* et Ws ont alors été transformées et les graines sèches, identifiées grâce à un marqueur de sélection DsRed (coloration rouge des graines), ont été disséquées afin d'observer les embryons sans que le processus de germination ne soit initié. Des transformants primaires en fond Ws et *lec2-1* exprimant *pAtUbq10::GFP:AtLEC2:tRbcSE9* ont été comparés avec d'autres transformants portant la construction *pAtUbq10::GFP:LEC2 M104-Q275:tRbcSE9*, une version non fonctionnelle de LEC2 (voir [Chapitre 2](#)), utilisée ici comme contrôle négatif. Les embryons surexprimant *LEC2 M104-Q275* en fond Ws et *lec2-1* ont montré un phénotype normal, avec, dans le cas du mutant *lec2-1* une absence de complémentation, caractérisée par une accumulation de chlorophylle et d'anthocyanes visibles sur les cotylédons, dont la forme est plus épaisse et pointue que celle de l'embryon Ws (Figure 1-9A et B). De même, l'axe de l'embryon *lec2-1* est plus large que celui du Ws. L'embryon de génotype sauvage surexprimant *AtLEC2* sous contrôle de *pAtUbq10* montre un phénotype assez différent de l'embryon de type Ws surexprimant *LEC2 M104-Q275*, avec des cotylédons plus larges et une pointe racinaire très arrondie (Figure 1-9A et C). Enfin, si l'architecture globale de l'embryon Ws surexprimant *AtLEC2* semble tout de même correspondre à ce qui est attendu, celle d'embryons de type *lec2-1* surexprimant *AtLEC2* est totalement modifiée avec une gradation de phénotypes, allant d'un aspect quasi normal, à des structures pouvant faire penser à celle d'un « volant de badminton », exactement comme celle de certaines plantules observées en culture *in vitro* (Figures 1-9D et 1-8A,B). Ces observations montrent que la surexpression de *AtLEC2* sous contrôle de *pAtUbq10* induit des défauts de développement dès l'embryogenèse et non seulement après germination. L'axe de polarité apico-basal semble conservé et on peut imaginer que la pointe observée en partie basse de l'embryon correspond à la pointe racinaire et que les cotylédons sont fusionnés et forment la majeure partie de l'embryon.

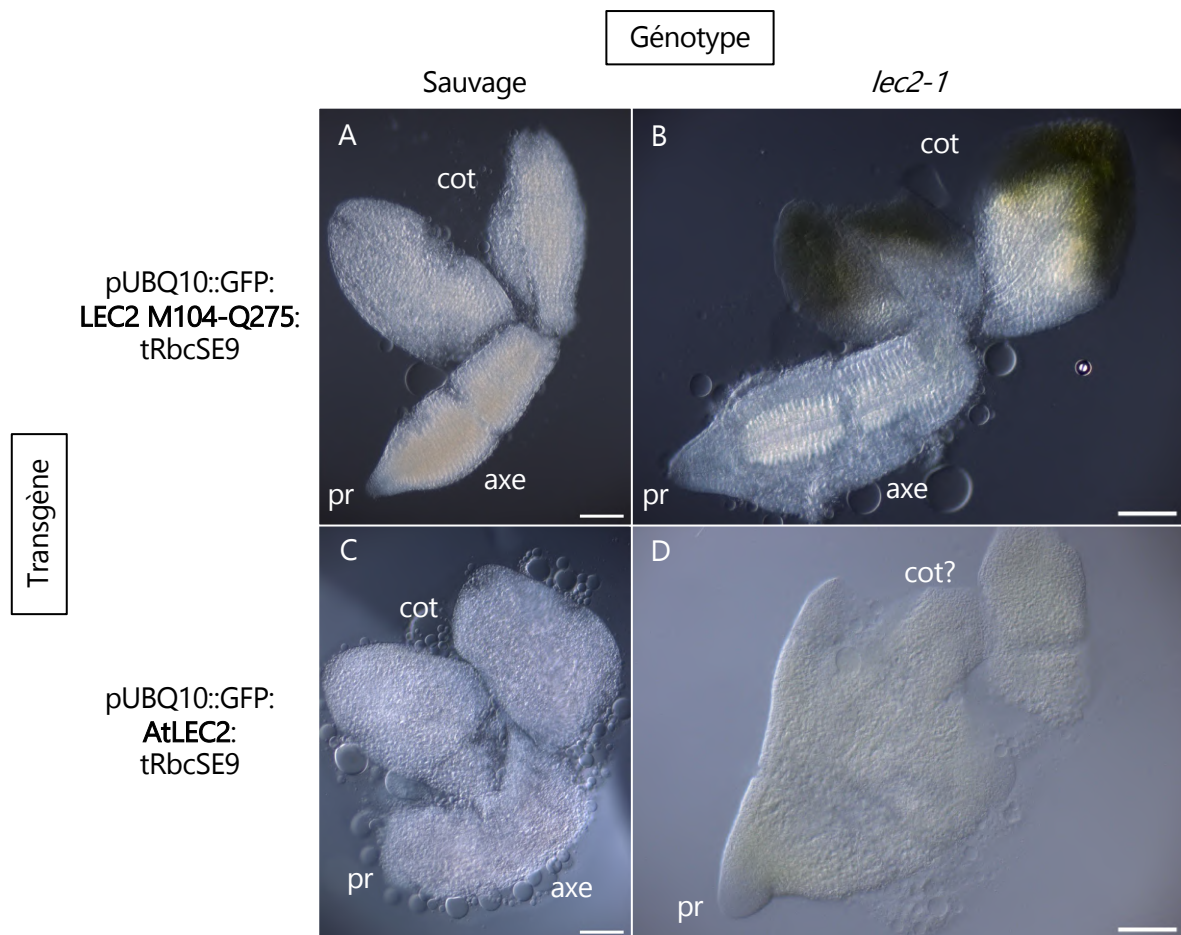


Figure 1-9 : La surexpression de *LEC2* sous contrôle de *pAtUBQ10* induit des défauts sévères dès l'embryogenèse. Embryons matures A) sauvage et B) *lec2-1* surexprimant LEC2 M104-Q275, une version non fonctionnelle de LEC2. Embryons matures C) sauvage et D) *lec2-1* surexprimant AtLEC2. Barre d'échelle : 50 μ m.

Abréviations : cot, cotylédons ; pr, pointe racinaire.

4.2. Le phénotype surexpresser est plus fort chez le mutant *lec2-1*

Par deux fois, des plantes sauvages Ws et mutantes *lec2-1* ont été transformées avec la construction *pAtUbq10::GFP:AtLEC2:tRbcSE9*. Dans les deux cas, le nombre de transformants primaires obtenus pour les deux génotypes est très différent et les mutants *lec2-1* ont donné beaucoup plus de transformants que les plantes sauvages. Au-delà du nombre de transformants récoltés, une importante différence de phénotype a pu être observée chez les plantes Ws, qui semblent nettement moins impactées que les mutants *lec2* (Figure 1-9C et D). Le faible nombre de transformants primaires chez le sauvage pourrait être lié à un effet embryonnaire létal de la surexpression de LEC2 (non visible ou moins important quand aucune copie endogène de AtLEC2 n'est présente). Dans ce cas, les transformants primaires obtenus sont ceux pour lesquels l'expression du transgène est la plus faible, ce qui pourrait expliquer pourquoi le phénotype est moins sévère pour ces plantes.

4.1. L'utilisation de *pAtUbq10* vaut-elle mieux que celle de *pCaMV35S*?

L'utilisation du promoteur *pAtUbq10* plutôt que *pCaMV35S* permet d'induire un phénotype extrêmement fort, avec l'apparition d'embryons somatiques sur les plantules, comme décrit dans la bibliographie. La surexpression d'un transgène, en particulier avec le *pCaMV35S* peut induire du *silencing* (Rajeevkumar et al., 2015). Il est donc possible que l'expression de *LEC2* sous contrôle de ce promoteur fort soit bloquée par des mécanismes épigénétiques comme le *silencing* transcriptionnel ou post-transcriptionnel des gènes (TGS ou PTGS), empêchant l'apparition de phénotype développemental fort. Néanmoins l'accumulation de réserves spécifiques des graines dans les feuilles d'*Arabidopsis* a été démontrée avec ce même promoteur (Mendoza et al., 2005). Dans le cas de l'apparition d'embryons somatiques sur les plantules, les rares fois où ce phénotype a été observé avec un *pCaMV35S* (Figure 1-2), il semble que les plantules se développaient normalement dans un premier temps, avec la formation de racines et de cotylédons, mais que le développement végétatif devenait ensuite anormal. Au contraire, l'utilisation du *pAtUbq10* induit des changements majeurs dès l'embryogenèse, montrant que ce promoteur permet la surexpression de *LEC2* dès la formation de l'embryon. Nous ne savons pas à quel stade de développement de la graine ou de la plantule le *pCaMV35S* induit l'expression du transgène, mais il est probable que ce soit de manière plus tardive qu'avec le *pAtUbq10* (Tsutsui and Higashiyama, 2016). Cependant, le défaut du *pAtUbq10* est que le phénotype surexpresser de LEC2 est tellement fort qu'il est très difficile d'amener une plante à fleurs et d'obtenir des graines. Dans notre cas, seules deux plantes ont pu être transférées en serre et ont donné des graines, probablement parce

que le transgène était moins exprimé, à cause peut-être du locus d'insertion de l'ADN-T dans le génome. Une alternative pour éviter l'apparition de phénotypes trop forts, serait de tester une collection d'autres promoteurs constitutifs ou un système d'expression inductible à la dexaméthasone (dex), par exemple. De cette manière, la protéine ne serait localisée dans les noyaux qu'après traitement des plantes à la dexaméthasone. Ainsi, on pourrait s'affranchir de l'effet délétère de la surexpression de *LEC2* au cours de l'embryogenèse et du développement précoce par un traitement adapté par le corticoïde de synthèse.

Chapitre 2 :
Approche avec *a priori*

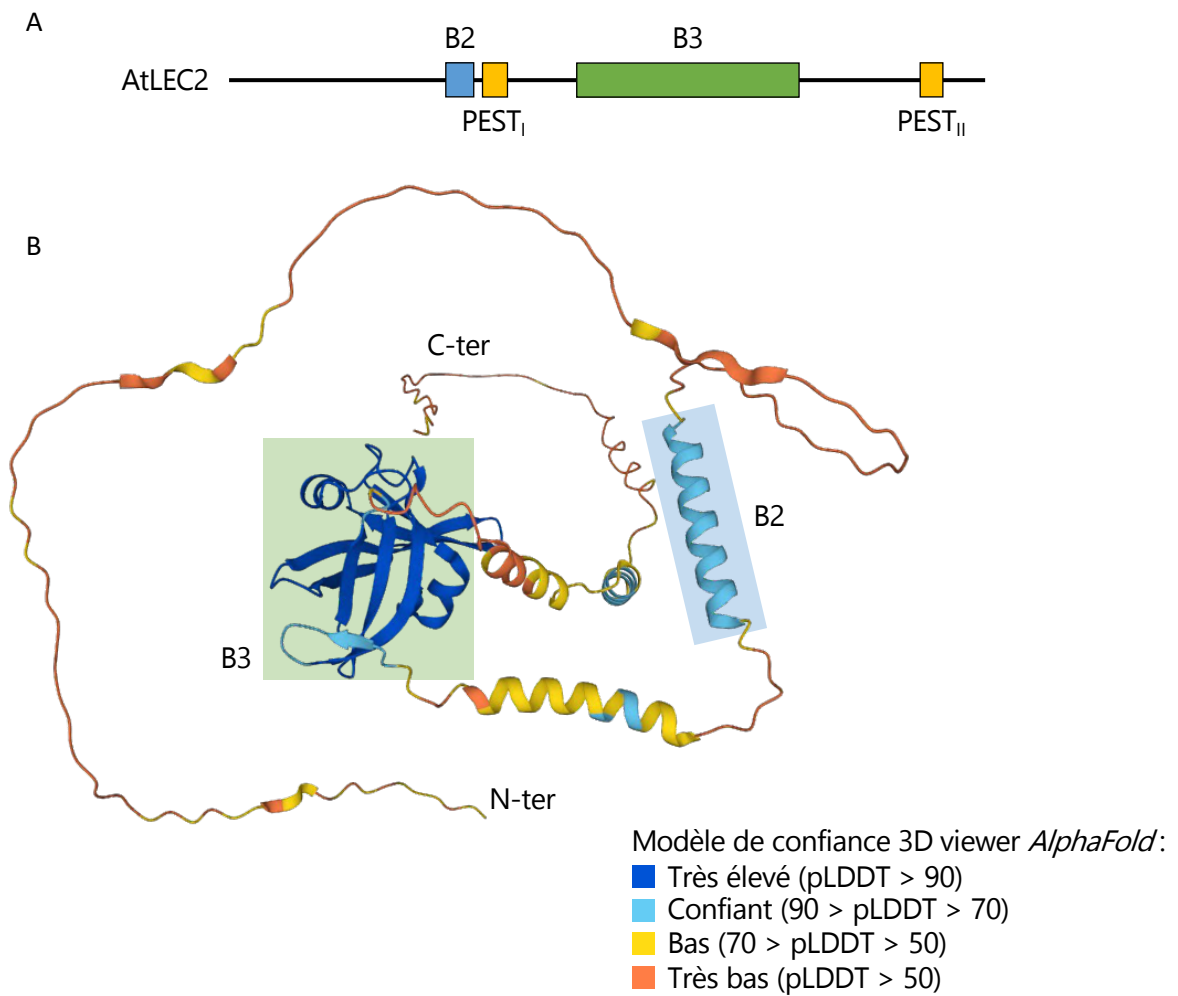


Figure 2-1 : La structure de la protéine AtLEC2, en dehors du domaine B3, a été très peu caractérisée. A) Structure schématique des domaines de AtLEC2. Le domaine B3 de liaison à l'ADN est représenté en vert et le domaine B2 d'interaction protéine-protéine en bleu. Les domaines PEST prédits sont indiqués en jaune. B) Prédiction de la structure de la protéine AtLEC2 donnée par l'algorithme *AlphaFold* (Jumper *et al.*, 2021, Varadi *et al.*, 2022). Le domaine B3, encadré en vert a une structure prédite avec une grande probabilité, avec notamment 7 feuillets β et 2 hélices α très conservés chez les facteurs de transcription à domaine B3. Le domaine B2, encadré en bleu est prédit avec un seuil de confiance assez élevé comme étant une hélice α . Le reste de la protéine, notamment les parties N et C-terminales sont très peu structurées.

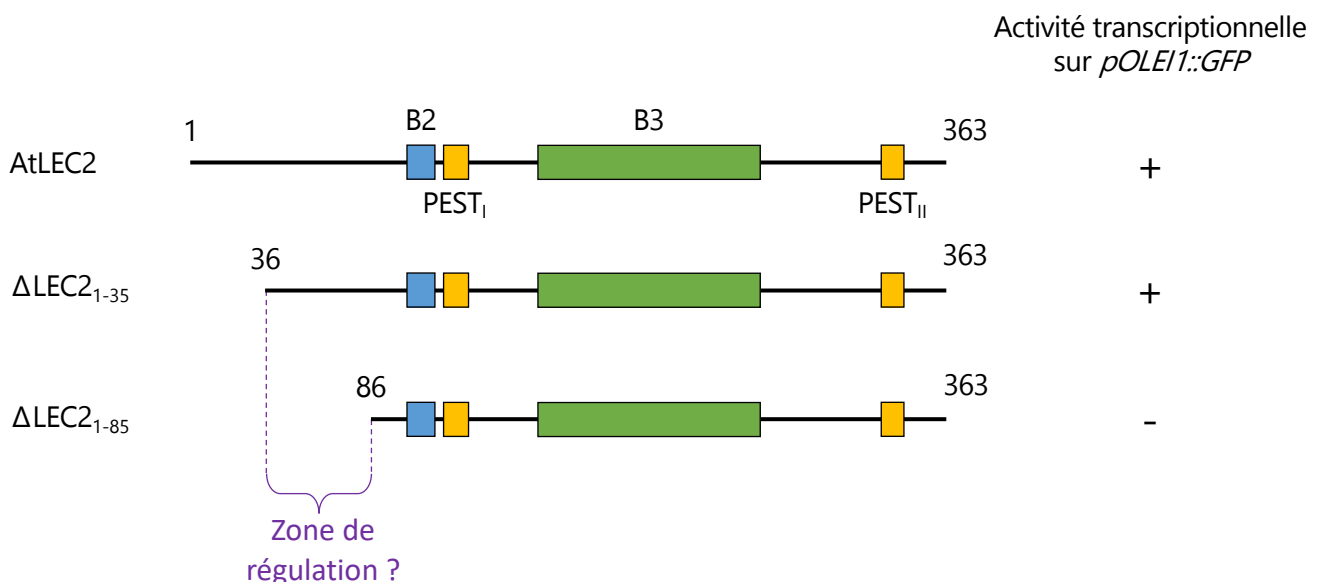


Figure 2-2 : Une partie de l'extrémité N-terminale de LEC2, située entre les acides aminés D36 à L87 est nécessaire à son activité transcriptionnelle dans des protoplastes de mouche.

Chapitre 2 : Approche avec *a priori*

1. Introduction

Il existe très peu de données de structure de la protéine LEC2 (Figure 2-1). On sait que LEC2, comme ABI3 et FUS3, possède un domaine B3 de liaison à l'ADN dont la séquence de 110 acides aminés et la structure sont très conservés dans cette famille de facteurs de transcription (Stone et al., 2001; Carbonero et al., 2016; Sasnauskas et al., 2018). Le domaine B3, qui permet les interactions ADN-protéines comporte 7 feuillets β et 2 hélices α dont la structure a été résolue pour la première fois avec la protéine RAV1 (Yamasaki et al., 2004). LEC2 compte également un domaine B2, basique, nécessaire pour des interactions protéine-protéine et dont la séquence est très conservée, notamment avec un motif consensus TKxxARxxRxxAxR retrouvé chez ABI3 et LEC2 dans plusieurs espèces (Boulard et al., 2018). Le consensus du domaine B2 n'est pas retrouvé chez FUS3. Même si la structure de ce domaine n'a pas été élucidée, l'algorithme *AlphaFold* (Jumper et al., 2021; Varadi et al., 2022) prédit toujours une hélice α avec un indice de confiance assez élevé sur plusieurs espèces (Figure 2-1B). Deux domaines PEST ont également été prédits en N et C-terminal du domaine B3, ainsi que des acides aminés prédits comme étant hautement phosphorylables. Ainsi, hormis pour le domaine B3, la structure de la protéine LEC2 n'a jamais été résolue.

1.1. Données préliminaires : une partie de l'extrémité N-terminale est nécessaire à l'activité transcriptionnelle de LEC2 dans la mousse

L'approche avec *a priori* que j'ai menée sur LEC2 au cours de ma thèse a été initiée lors de mon stage de M1, qui avait alors pour but d'obtenir la version la plus courte possible de LEC2, qui soit toujours fonctionnelle et ce, dans le but de résoudre la structure de la protéine. Les résultats que j'ai obtenus ont montré qu'une partie de l'extrémité N-terminale de LEC2, située entre les acides aminés D36 et L86, est nécessaire à son activité transcriptionnelle sur le promoteur *pOLE1* en expression transitoire dans des protoplastes de mousse, sans pour autant que cette séquence soit conservée à travers les espèces (Figure 2-2). Voir Chapitre 1, paragraphe [2.1.2 Validation fonctionnelle de LEC2*SH dans la mousse](#) et Matériel et méthodes, paragraphe [2.1.5.2. Transformation transitoire de *Physcomitrium patens*](#) pour le mode opératoire de cette technique.

La fonctionnalité d'un facteur de transcription peut être affectée par quatre éléments :

- sa localisation nucléaire (signaux d'entrée ou de sortie du noyau, *i.e.* NLS et NES),

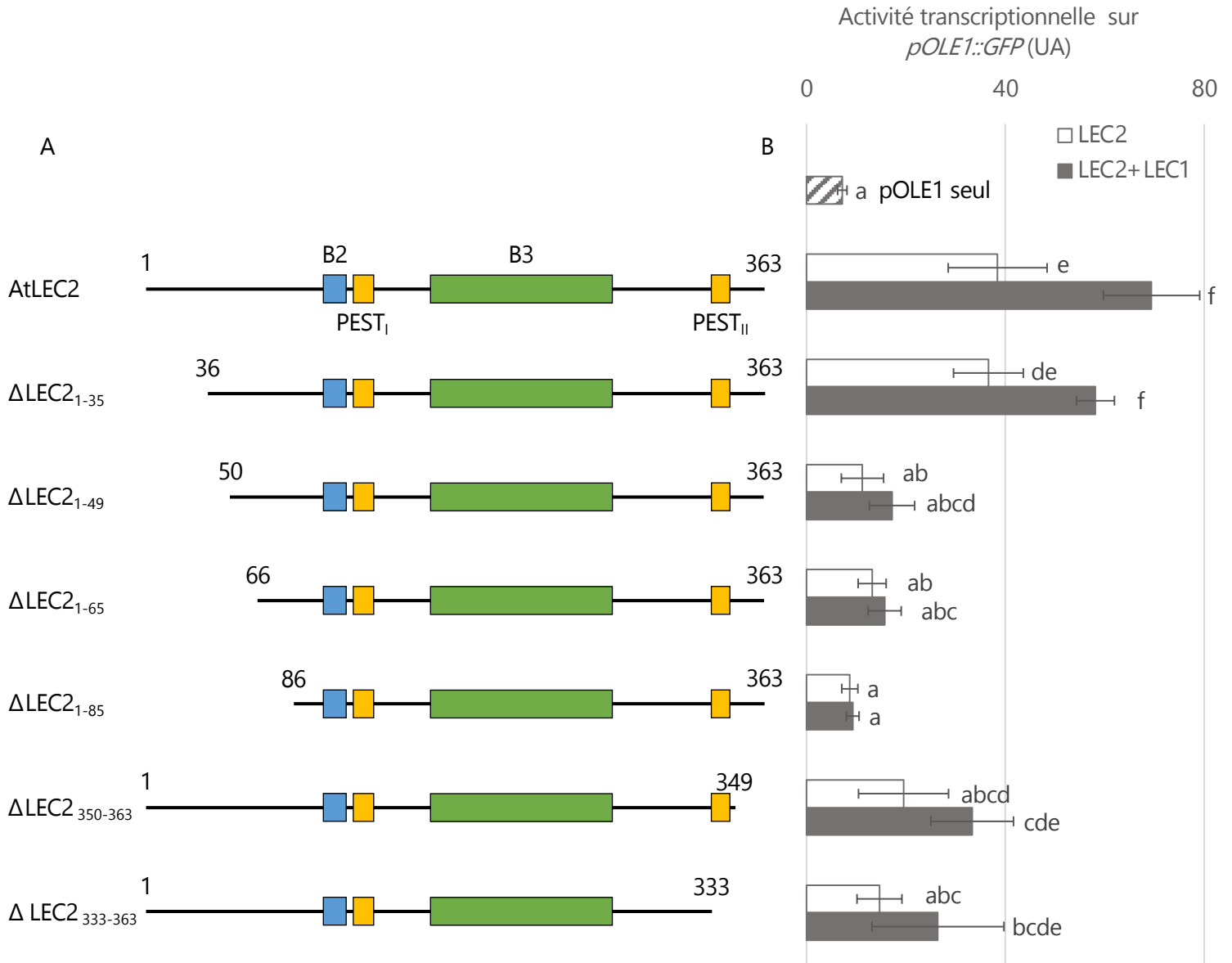


Figure 2-3 : Des délétions des extrémités N et C-terminales de AtLEC2 réduisent drastiquement son activité transcriptionnelle sur *pOLE1::GFP* dans les protoplastes de mousse. A) Représentation schématique de la protéine AtLEC2 et des différentes délétions dont l'activité transcriptionnelle a été testée. B) Mesure de la fluorescence de la GFP dans les protoplastes de mousse transformés, reflétant l'activité transcriptionnelle en unité arbitraire (UA) de AtLEC2 ou ses variants, seuls ou en complexe avec LEC1 sur *pOLE1::GFP*. Les groupes statistiques ont été obtenus grâce à un test ANOVA avec un risque $\alpha=0,05$.

- sa liaison à l'ADN,
- son activité transcriptionnelle (recrutement de la machinerie transcriptionnelle)
- son interaction protéine-protéine avec d'autres facteurs ou coactivateurs transcriptionnels.

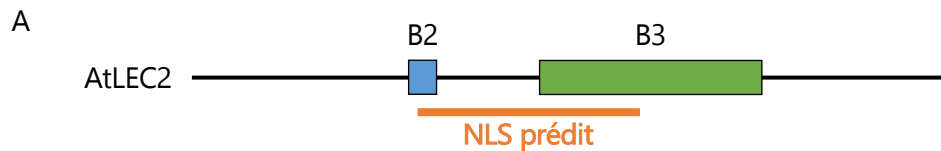
Bien que les domaines B2 et B3 ne soient pas touchés par cette délétion, nous en avons déduit que i) l'un de ces quatre éléments pouvait être impacté et que ii) la partie N-terminale de LEC2 possède une zone de régulation essentielle à son activité, qui n'avait jusqu'alors pas été mise en évidence, pouvant affecter la structure de la protéine, sa localisation ou encore son activité transcriptionnelle seule ou en complexe avec d'autres protéines.

1.2. Affinement de la zone de régulation en N-ter de LEC2

Afin de déterminer plus précisément les limites de la zone nécessaire à l'activité transcriptionnelle de LEC2, plusieurs délétions supplémentaires ont été réalisées entre les acides aminés D36 et L86 et leur fonctionnalité a été testée sur *pOLE1::GFP* en expression transitoire dans des protoplastes de mousse (Figure 2-3). L'intensité de fluorescence mesurée dans les protoplastes reflète l'activité transcriptionnelle des facteurs de transcription sur leur promoteur cible (Thévenin et al., 2012). L'activité de différents variants de LEC2 est mesurée sur *pOLE1::GFP* lorsque le facteur est seul ou en complexe avec AtLEC1 et comparée au témoin négatif, correspondant aux protoplastes transformés avec *pOLE1::GFP* seul et aux témoins positifs *pOLE1::GFP* + AtLEC2 et *pOLE1::GFP* + AtLEC2 + AtLEC1. Il a été montré que AtLEC1 n'active pas seul la transcription via *pOLE1* mais a un effet coopératif sur AtLEC2 et l'activité transcriptionnelle de AtLEC2, reflétée par la fluorescence de la GFP est toujours plus forte lorsque AtLEC1 est présent (Baud et al., 2016).

Les résultats obtenus (Figure 2-3) montrent que la délétion des 35 premiers acides aminés de AtLEC2 (Δ LEC2₁₋₃₅) n'impacte pas significativement son activité transcriptionnelle. Cependant, les délétions entre les acides aminés S50 et Y66 (Δ LEC2₁₋₄₉ et Δ LEC2₁₋₆₅) réduisent sévèrement l'activité transcriptionnelle du facteur, même si AtLEC1 a toujours un léger effet coopératif. Lorsque la protéine débute au niveau de L86 (Δ LEC2₁₋₈₅), elle perd toute son activité transcriptionnelle sur *pOLE1::GFP*, seule ou en complexe avec AtLEC1 et le signal détecté dans les protoplastes est du même niveau que celui du bruit de fond de *pOLE1::GFP*. Des délétions en partie C-terminale de AtLEC2 (Δ LEC2₃₄₇₋₃₆₃ et Δ LEC2₃₃₃₋₃₆₃) réduisent également significativement l'activité transcriptionnelle de LEC2 seul ou en complexe avec AtLEC1, mais plus légèrement.

Il semble donc que les parties N et C-terminales de AtLEC2 soient importantes pour son activité transcriptionnelle sur *pOLE1::GFP* dans les protoplastes de mousse. Les domaines B2 et B3 ne sont pas impactés par ces délétions, à moins que l'absence des extrémités N et C-terminales ne



B

cNLS Mapper Result

Predicted NLSs in query sequence	
MDNFLPFPSSNANSVQELSMDPNNNRSHFTTVPTYDHHQAQPHHFLPPFS	50
YPVEQMAAVMNPQPVYLSECYPQIPVTQTGSEFGSLVGNPCLWQERGGFL	100
DPRMTKMARI RNRKNAMR SRNNS PNSSPSELVDSKRQLMMLNLKNNVQI	150
SDKKDSYQQSTFDNKKLRVLCEKELKNSDV GSLGRIVL PKRDAEANLPKL	200
SDKEGIVVQMRDVF SMQSWSFYKFWNNKSRMYVLENTGEFVKQNGAEI	250
GDFLTIYEDESKNLYFAMNGNSGKQNEGRENESRERNHYEAMLDYIPRD	300
EEEASIAMLIGNLNDHYPIPNDLMDLTTDLQHHQATSSSMPPEDHAYVGS	350
SDDQVSFNDFEWW	363

Predicted monopartite NLS		
Pos.	Sequence	Score

Predicted bipartite NLS		
Pos.	Sequence	Score
109	RINRKNAMR SRNNSPNSSPSELVDSKRQL	3.2
137	RQLMMLNLKNNVQISDKKDSYQQSTFDNKKLRVLCE	3
150	ISDKKDSYQQSTFDNKKLRVLCEKELKNSDV	3.5
185	RIVLPKRDAEANLPKLSDKEGIVVQMRDVF SM	3.1

Figure 2-4 : Un domaine de localisation nucléaire est prédit entre les domaines B2 et B3 de AtLEC2 (de R109 à M216). A) Représentation schématique de la structure de AtLEC2, avec le NLS prédit représenté en orange. B) Résultats donnés par l'outil de prédiction *cNLS Mapper* (Kosugi et al., 2009). Le NLS prédit est représenté par les acides aminés en rouge.

provoque une déstructuration totale de la protéine. Dans le cas des délétions Δ LEC2₁₋₄₉ et Δ LEC2₁₋₆₅, LEC1 a toujours un effet coopératif sur LEC2, suggérant que l'interaction protéine-protéine via le domaine B2 n'est pas altérée et que la structure de la protéine n'est pas modifiée au point qu'elle ne soit plus fonctionnelle. Si nous reprenons les différents éléments pouvant affecter la fonctionnalité d'un facteur de transcription, nous pouvons alors poser l'hypothèse que les parties N et C-terminales de LEC2 peuvent être importantes pour la localisation de LEC2 ou pour son activité transcriptionnelle. Dans ce chapitre, je vais donc m'attacher à caractériser les domaines de localisation nucléaire et d'activation de la transcription de LEC2, qui jusqu'ici n'ont jamais été mis en évidence. Pour cela, j'associerai les données préliminaires présentées ci-dessus, à des outils de prédictions basés sur la séquence de la protéine ainsi que des validations fonctionnelles dans différents systèmes d'expression : le tabac, Arabidopsis, ou encore la levure et la mousse.

2. Mise en évidence du signal de localisation nucléaire (NLS) de LEC2

Pour exercer sa fonction de régulateur de l'expression des gènes, un facteur de transcription doit être localisé dans le noyau. Afin d'assurer une régulation précise des processus cellulaires, comme l'expression des gènes, la transduction des signaux ou encore la progression du cycle cellulaire, les eucaryotes ont développé une régulation sélective du transport bidirectionnel des macromolécules entre le noyau et le cytoplasme (Lange et al., 2007). Le moyen d'import nucléaire le plus décrit est celui mettant en jeu l'hétérodimère d'importines α et β et les signaux de localisation nucléaire (NLS) contenus dans les peptides. Afin d'entrer dans le noyau, les protéines formées dans le réticulum endoplasmique, peuvent se lier à l'importine α via leur NLS, qui est généralement composé d'acides aminés basiques. Le complexe alors formé par l'hétérodimère d'importines α/β et la protéine d'intérêt par l'intermédiaire de son NLS va alors être transloqué dans le noyau via les pores nucléaires. Un facteur de transcription peut donc entrer dans le noyau grâce à un NLS ou bien via son interaction avec d'autres protéines de transport nucléaire.

2.1. Un NLS est prédit en partie centrale de AtLEC2

Grâce au programme de prédiction des signaux de localisation nucléaire *cNLS mapper* (Kosugi et al., 2009) (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi), un domaine de localisation nucléaire de LEC2 a été prédit et serait formé de deux parties, dont une entre les acides aminés R109 et V179, et l'autre entre R184 et M216 (Figure 2-4). Ce NLS serait alors situé au niveau des domaines B2 et B3 de la protéine LEC2, mais pas dans les parties N et C-terminales du facteur de transcription (Figure 2-4A). Si cette prédiction est confirmée, il semblerait donc que la baisse

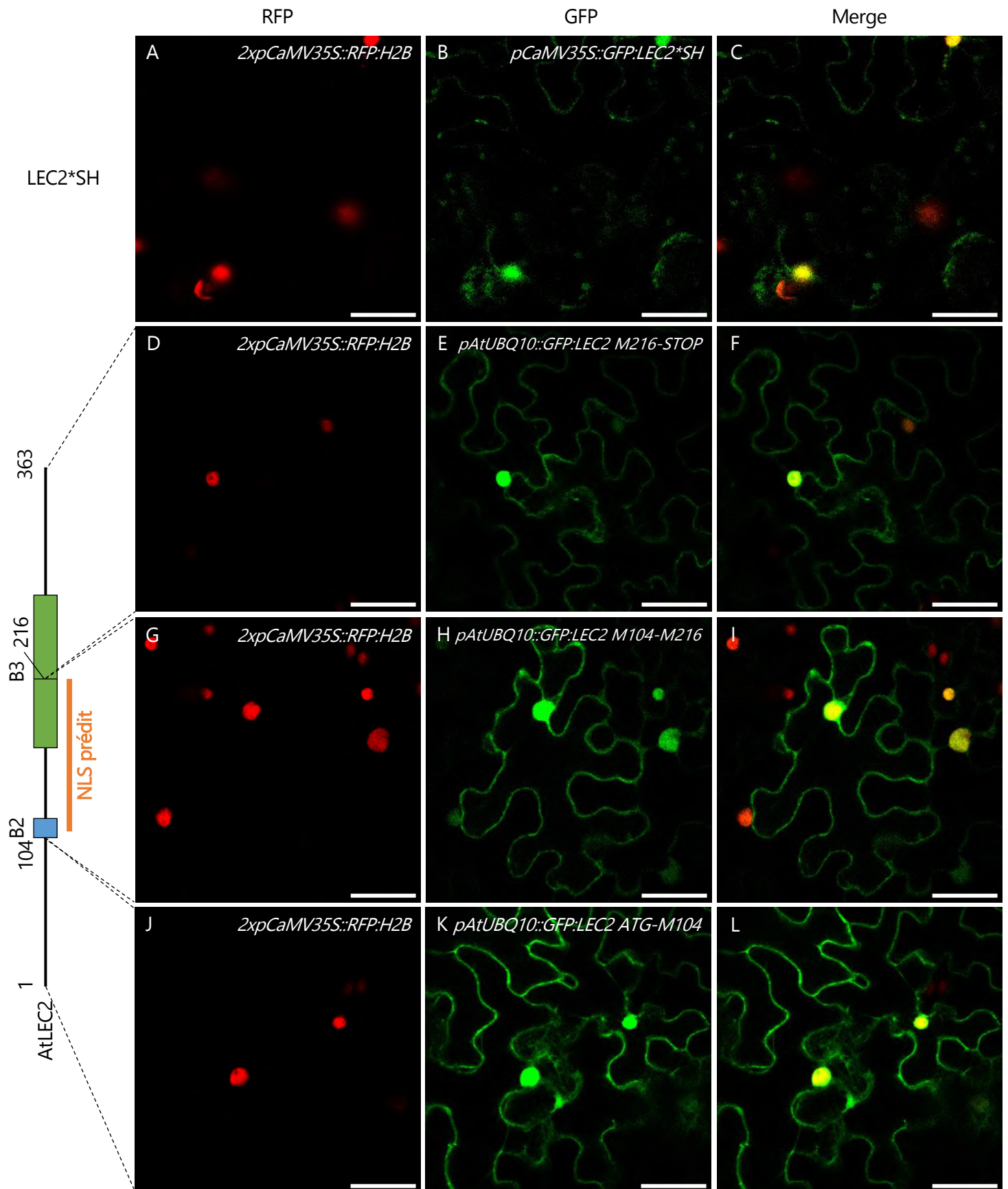


Figure 2-5 : L'expression transitoire de différents variants de LEC2 dans les feuilles de *Nicotiana benthamiana* ne permet pas de confirmer la position du domaine de localisation nucléaire (NLS) du facteur de transcription. Localisation de A-C) pCaMV35S::GFP:LEC2*SH ; D-F) pUBQ10::GFP:LEC2 M216-STOP ; G-I) LEC2 M104-M216 ; J-L) LEC2 ATG-M104. La première colonne (A ; D ; G ; J) correspond aux noyaux marqués en rouge dans la lignée de tabac RFP:H2B (Martin *et al.*, 2009) (580-620 nm). La seconde colonne (B ; E ; H ; I) montre la localisation de la GFP (495-525 nm). La dernière colonne (C ; F ; I ; L) montre la superposition des deux premières images. Barres d'échelle : 50 μ m.

d'activité transcriptionnelle de LEC2 sans ses parties N ou C-ter ne soit pas liée à un défaut de localisation.

2.2. Validation expérimentale de la prédiction du NLS de AtLEC2

Afin de valider la prédiction du programme *cNLS mapper*, un système d'expression transitoire dans des feuilles de tabac (*Nicotiana benthamiana*) a été utilisé. Pour localiser le noyau des cellules de tabac, une lignée portant un marqueur nucléaire rouge RFP:H2B a été choisie (Martin et al., 2009). La partie centrale de LEC2 (LEC2 M104-M216), prédite comme étant un NLS, ainsi que les extrémités N et C-terminales (LEC2 ATG-M104 et LEC2 M216-STOP) ont été amplifiées par PCR et clonées en fusion avec la GFP sous contrôle d'un promoteur constitutif *pAtUbq10*, afin de tester leur localisation cellulaire. Les résultats obtenus (Figure 2-5) montrent que le témoin positif AtLEC2 a bien une localisation nucléaire, confirmée par la colocalisation avec les noyaux rouges. De même, LEC2 M104-M216, qui contient le domaines B2, une partie du domaine B3 ainsi que le NLS prédit, a également une localisation nucléaire. Cependant, les témoins négatifs LEC2 ATG-M104 et LEC2 M216-STOP ne comportant pas de NLS prédit ont aussi une localisation nucléaire. Les résultats obtenus en expression transitoire dans les feuilles de tabac ne permettent donc pas de valider la présence de domaine de localisation nucléaire en partie centrale de AtLEC2.

2.3. Localisation de variants de LEC2 *in planta*

Des plantes d'*Arabidopsis thaliana* mutées au locus LEC2, notées *lec2-1* ont été transformées avec différents variants de LEC2 présentant une activité transcriptionnelle réduite dans les protoplastes de mousse (cf. données préliminaires) mais de plus grande taille que les fragments utilisés dans le système d'expression de tabac du paragraphe précédent. Ces variants ont été clonés sous contrôle du promoteur de AtLEC2 (*pLEC2*) et en fusion avec la GFP en partie N-terminale de la protéine. Ces constructions transformées en fond mutant *lec2-1* permettent de tester la capacité des variants de LEC2 à restaurer le phénotype sauvage. La fusion avec la GFP permet également d'observer la localisation du variant de LEC2 en comparaison avec AtLEC2.

AtLEC2 s'exprime dans l'embryon entre les stades globulaire et torpille (Figures Intro-19 et 20). Les embryons en développement au stade cœur, observés au microscope confocal, montrent une localisation de AtLEC2 au noyau, confirmée par la colocalisation avec le DAPI (Figure 2-6A-C). Les délétions de la partie N-terminale de LEC2 (Δ LEC2₁₋₈₅) et de la partie C-terminale (Δ LEC2₂₇₅₋₃₆₃) (Figure 2-6D-I) ne semblent pas affecter la localisation nucléaire du facteur de transcription. La localisation de Δ LEC2₁₋₈₅ et de Δ LEC2₂₇₅₋₃₆₃ n'étant pas différente de celle de AtLEC2, cela confirme que la délétion d'une des extrémités N ou C-terminale n'impacte pas la localisation et donc que le

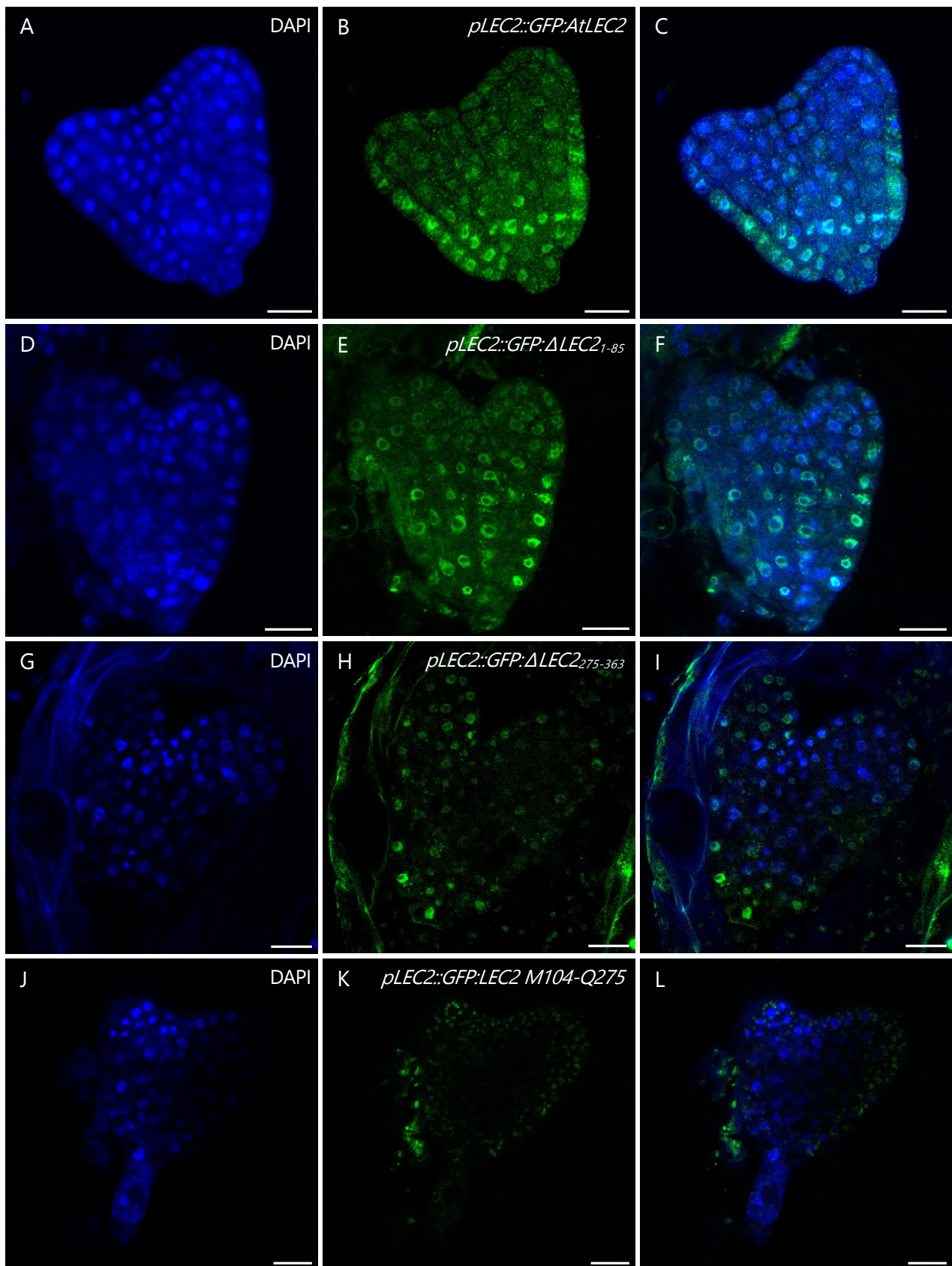


Figure 2-6 : La délétion d'une extrémités N ou C-terminale de LEC2 n'affecte pas sa localisation au cours du développement de la graine d'*Arabidopsis thaliana*. A-C) Embryon *lec2-1* au stade cœur complémenté avec A-C) $pLEC2::GFP:AtLEC2$; D-E) $\Delta LEC2_{1-85}$; G-H) $\Delta LEC2_{275-363}$; J-L) $LEC2 M104-Q275$. La première colonne (A ; D ; G ; J) correspond aux noyaux marqués au DAPI (450-590 nm). La seconde colonne (B ; E ; H ; I) montre la localisation de la GFP (495-525 nm). La dernière colonne (C ; F ; I ; L) montre la superposition des deux premières images. Barres d'échelle : 15 μ m.

NLS n'est situé ni en N-ter, ni en C-ter de LEC2. Cependant, lorsque des extrémités N et C-terminales sont toutes les deux délétées (LEC2 M104-Q275), la GFP n'est pas détectée dans le noyau.

2.4. Conclusion et discussion

Les données de prédiction obtenues situent le NLS de LEC2 en partie centrale de la protéine, entre les domaines B2 et B3. Cependant, l'expression transitoire dans des feuilles de tabac *Nicotiana benthamiana* n'a pas permis de confirmer cette hypothèse. En effet, il semble que quel que soit le peptide fusionné à la GFP, sa localisation soit nucléaire. Le noyau laisse entrer librement des molécules jusqu'à 60 kDa, or, bien que la GFP ait une localisation cytosolique, il s'agit d'une petite protéine de 27 kDa et, bien qu'elle n'ait pas de NLS dans sa séquence, il a été montré que cette petite protéine fluorescente est capable d'entrer librement dans le noyau (Böhm et al., 2006; Seibel et al., 2007). Il est donc probable que les fusions réalisées entre la GFP et les différents variants de LEC2, qui sont tous des petits peptides d'environ 100 acides aminés, aient un poids moléculaire inférieur à 60 kDa et entrent librement dans le noyau par diffusion. Pour montrer la localisation cellulaire d'un peptide fusionné à la GFP, il faudrait alors réaliser une protéine fusion de poids moléculaire bien plus élevé, par exemple en utilisant un hétéro/homo-dimère de GFP. Cependant, il a été montré qu'un hétérodimère comportant jusqu'à 6 GFP fusionnées possède toujours une localisation nucléaire (Seibel et al., 2007), il est donc probable qu'un hétérodimère de GFP ne suffise pas à régler ce problème. Lors des agro-infiltrations de feuilles de tabac, la construction d'intérêt est co-transformée avec un plasmide *p19* qui comporte un inhibiteur du *silencing*, et qui permet d'augmenter le niveau d'expression de la protéine (Vanhercke et al., 2013). Étant donné que la diffusion libre dans le noyau dépend de la concentration de molécules dans le cytosol, nous avons essayé d'agro-infiltrer les feuilles de tabac sans *p19*. Cependant, l'absence de *p19* n'a pas perturbé l'expression transitoire des protéines et la localisation observée était toujours nucléaire, y compris pour les témoins négatifs sans NLS putatif. Une autre option, afin de confirmer la prédiction de NLS de LEC2 en expression transitoire dans des feuilles de tabac, serait de fusionner le NLS putatif ainsi que les contrôles négatifs sans NLS prédits avec une protéine connue n'ayant pas de localisation nucléaire. Ainsi, le NLS seul pourrait suffire à la translocation de la protéine dans le noyau à l'inverse des contrôles négatifs. Malheureusement, par manque de temps, je n'ai pas pu mettre en place cette alternative. Néanmoins, l'expression de différents variants de LEC2 *in planta* chez *Arabidopsis thaliana* a permis de montrer que la localisation nucléaire de LEC2 n'est pas impactée lorsque la partie N ou C-terminale de LEC2 est délétée. Ainsi, il semble donc que le NLS de LEC2 ne soit pas situé dans ces extrémités et cela constitue donc une preuve indirecte de sa présence en partie

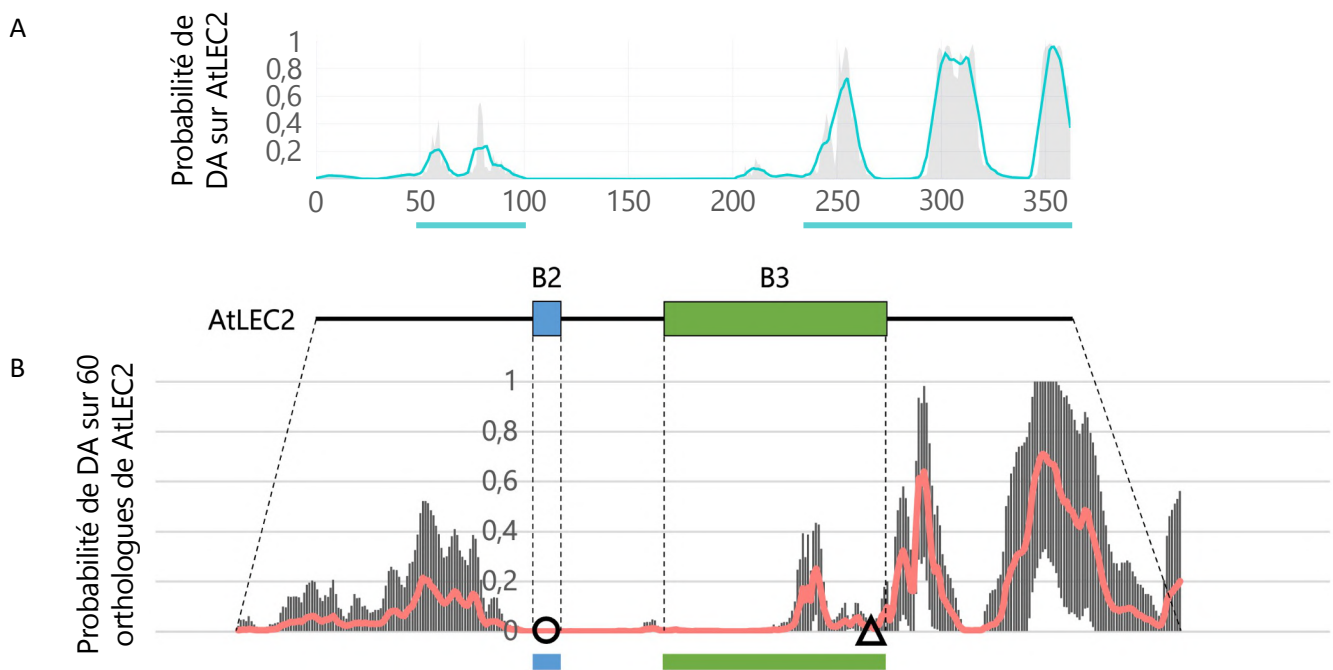


Figure 2-7 : Des domaines d'activation de la transcription (DA) sont prédits grâce à l'algorithme *Adpred* (Erijman et al., 2020) en N et C-terminal de AtLEC2 et ils semblent fortement conservés. A) La courbe bleue représente la probabilité de présence de domaine d'activation de la transcription sur la séquence de AtLEC2 représentée en abscisse. B) La courbe rouge représente la prédiction de domaines d'activation de 60 orthologues de LEC2 moyennée et normalisée sur la séquence de AtLEC2. Les repères rond et triangulaire sont les points de référence situés dans le B2 et à la fin du B3 utilisés pour la normalisation des prédictions.

centrale du facteur de transcription. Néanmoins, lorsque les parties N et C-terminales sont toutes les deux délétées, la GFP n'est plus détectée au noyau. Cela peut être lié à la conformation de la protéine, qui ne serait pas correcte et qui ne permettrait alors pas au NLS de se lier à l'importine α , ou bien la protéine pourrait être dégradée avant de rentrer au noyau. En effet, il semble que la GFP soit détectée au niveau de vésicules à l'extérieur des noyaux.

En conclusion, je n'ai pas pu confirmer de façon directe la localisation du NLS de LEC2, mais ces résultats ont tout de même permis de montrer que la perte de fonctionnalité du facteur de transcription sur *pOLE1::GFP* dans les protoplastes de mousse n'est pas liée à un défaut de localisation nucléaire.

3. Mise en évidence des domaines d'activation de la transcription de LEC2

Comme indiqué précédemment, un activateur transcriptionnel contient généralement un domaine de liaison à l'ADN, un domaine de localisation nucléaire, un domaine d'interaction avec d'autres protéines et un domaine d'activation de la transcription (DA). Les domaines d'activation sont assez peu connus : on sait qu'ils sont impliqués dans l'initiation et le niveau de transcription des gènes et peuvent agir de différentes manières pour cela : interaction et recrutement des coactivateurs transcriptionnels, de la machinerie basale de la transcription (ARN pol II, complexe Mediator, facteurs de transcription généraux, *TATA-binding protein*), mais ils peuvent également jouer un rôle dans le remodelage de la chromatine et dans les complexes de modification des histones, ainsi que dans l'interaction avec les nucléosomes (Erkina and Erkin, 2016). Les domaines d'activation sont interchangeables à travers les espèces eucaryotes et peuvent être fonctionnels chez les animaux, comme les levures ou les plantes (Yanagisawa, 2001; Erkina and Erkin, 2016). Cependant, aucune séquence consensus ni aucun motif structural commun n'ont été mis en évidence, mis à part la présence de résidus hautement hydrophobes et acides et il semble que la liaison du facteur de transcription à sa cible rende la structure des DA plus stable.

3.1. Des domaines d'activation de la transcription sont prédits en N et C-ter de LEC2

Bien que les domaines d'activation aient une séquence très peu conservée, un algorithme basé sur un réseau de neurones, *ADpred* (<https://adpred.fredhutch.org/>) a été développé afin de prédire la présence de domaines d'activation acides dans des facteurs de transcription (Erijman et al., 2020). Dans le cas de AtLEC2, *ADpred*, prédit un premier domaine en partie N-terminale, en amont du domaine B2, et un autre domaine en partie C-terminale débutant à la fin du domaine B3 (Figure 2-7A). Les séquences en N et C-terminal de LEC2 n'étant pas conservées à travers les espèces, nous

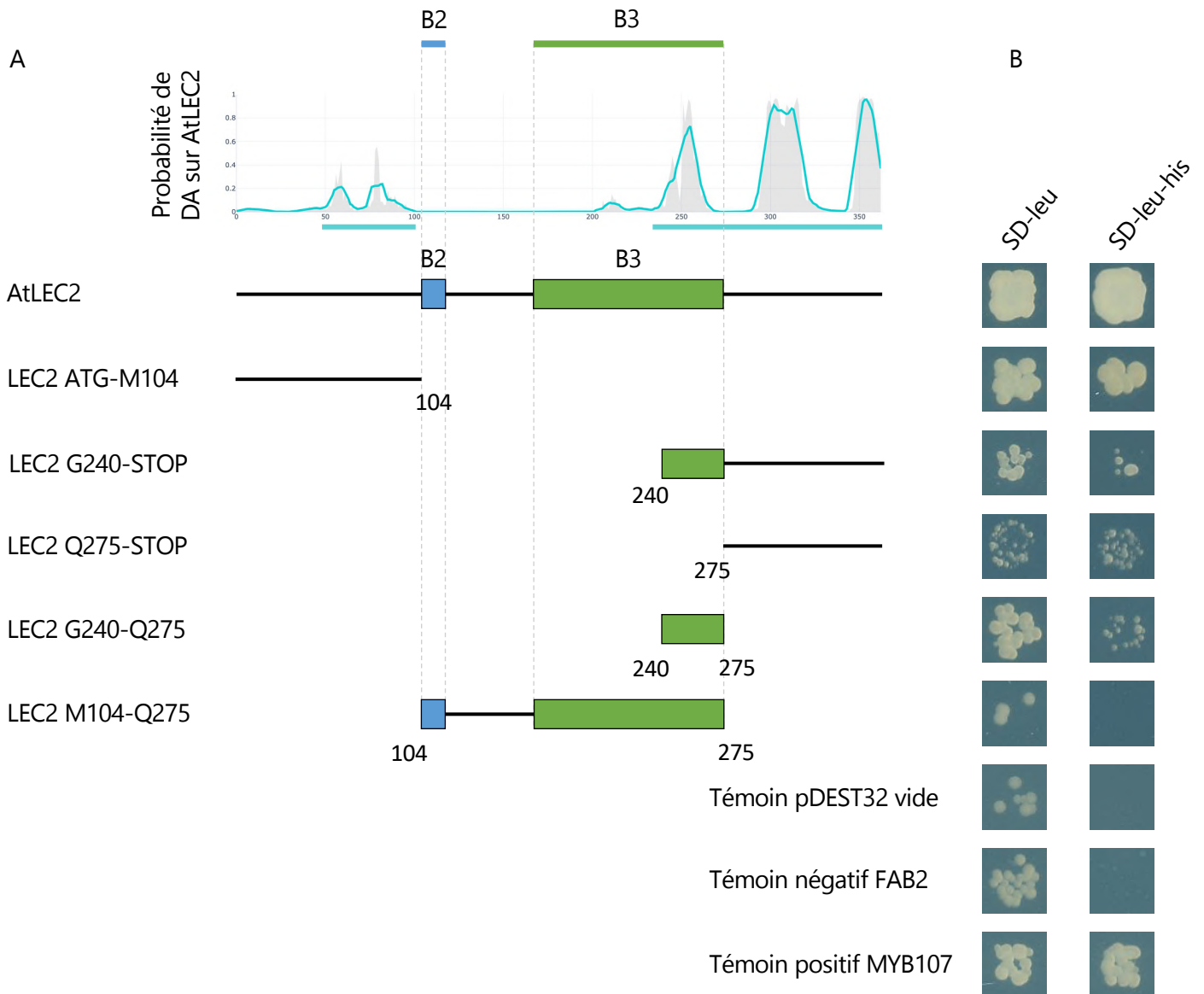


Figure 2-8 : La fonctionnalité des domaines d'activation de la transcription prédits par l'algorithme *Adpred* (Erijman et al., 2020) est confirmée par un test d'activité transcriptionnelle dans la levure. A) Représentation schématique des délétions de AtLEC2 clonées dans pDEST32, qui permet l'expression dans la levure de la fusion du domaine d'activation transcriptionnelle putatif au domaine de liaison à l'ADN de GAL4. B) Test d'activité transcriptionnelle sur milieu sélectif. Le milieu SD-leu permet la sélection des clones de levure portant le vecteur pDEST32. Le milieu SD-leu-his permet la sélection des clones portant une autotrophie à l'histidine apportée par l'activité transcriptionnelle du fragment, testée sur $LYS2::GAL1_{UAS}-GAL1_{TATA}-HIS3$.

avons voulu tester si *ADpred* prédit également des domaines d'activation dans une soixantaine d'orthologues de LEC2 issus de différentes espèces de plantes dicotylédones (liste des séquences en Annexe 5). Grâce à l'aide d'Ariel Erijman (développeur d'*ADpred*) et de Jasmine Burguet (équipe Modélisation et Imagerie Numérique, IJPB), nous avons obtenu des données de prédiction de domaines d'activation pour 60 orthologues de LEC2. Comme les séquences n'ont pas toutes la même longueur, nous les avons normalisées, sur la base de repères conservés situés dans le domaine B2 et dans le domaine B3, de façon à obtenir une prédiction consensus regroupant les 60 orthologues de AtLEC2 (Figure 2-7B). Les résultats obtenus montrent que les domaines d'activation prédits en N et C-terminal sont conservés chez 60 orthologues, bien que la séquence soit différente. Ces résultats coïncident avec l'idée que les domaines d'activation de la transcription peuvent être conservés à travers les espèces sans que leur séquence soit pour autant la même (Erkina and Erkine, 2016). D'autre part, cela suggère que les domaines sont hautement conservés en N et C-terminal chez LEC2.

3.2. Validation expérimentale de la prédiction des DA de AtLEC2

Afin de confirmer les données fournies par *ADpred*, j'ai réalisé des tests d'activation transcriptionnelle en levure (*Saccharomyces cerevisiae*) sur différents fragments de AtLEC2 prédits comme étant des domaines d'activation de la transcription. Les domaines d'activation prédits de LEC2 ont été amplifiés par PCR et clonés en fusion avec le domaine de liaison à l'ADN de *GAL4*. Les vecteurs obtenus ont été transformés dans une souche de levure AH109, dont le génotype exact est décrit dans le chapitre Matériel et méthodes, comportant un gène de synthèse de l'histidine (*HIS3*) sous contrôle de l'UAS et de la TATA-box de *GAL1*. Ainsi, le domaine de liaison de *GAL4* se lie à l'UAS de *GAL1* et, si le domaine d'activation est fonctionnel, alors il est capable d'activer la transcription de *HIS3* et de conférer une autotrophie à l'histidine. Comme témoins négatifs de ce test d'activation de la transcription dans la levure, le plasmide pDEST32 vide a été testé, ainsi que la $\Delta 9$ stearoyl-ACP desaturase *FAB2* (Kazaz et al., 2020), qui n'est pas un facteur de transcription. Le témoin positif utilisé est MYB107, un facteur de transcription déjà montré comme ayant une activité transcriptionnelle dans la levure (To et al., 2020). Les résultats obtenus montrent que la partie N-terminale de AtLEC2 (LEC2 ATG-M104) est capable d'activer la transcription, de même que la partie C-terminale (LEC2 G240-STOP) (Figure 2-8). De plus, le fragment compris entre G240 et Q275, c'est-à-dire situé à la fin du domaine B3 est capable d'activer seul la transcription, ainsi que la partie C-terminale non structurée Q275-STOP. En revanche, le domaine d'activation prédit entre M104 et Q275 n'est pas capable d'activer seul la transcription dans la levure lorsque le fragment

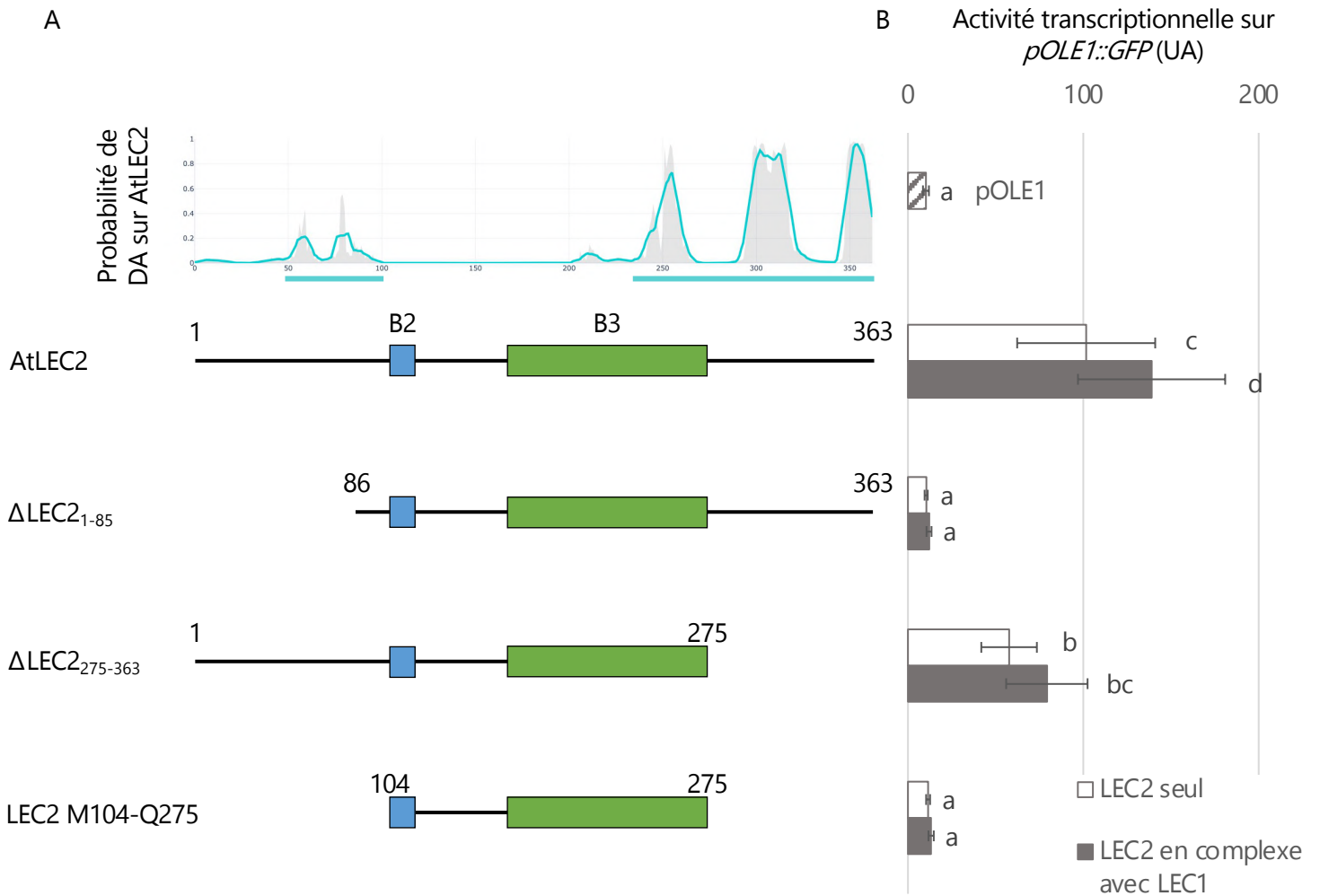


Figure 2-9 : Lorsque les domaines d'activation de la transcription de LEC2 sont délétes, son activité transcriptionnelle est fortement réduite. A) Représentation schématique des délétions de LEC2 par rapport aux domaines d'activation prédits par *Adpred* (Erijman et al., 2020). B) Activité transcriptionnelle des différentes versions de LEC2 seul ou en complexe avec LEC1 sur *pOLE1::GFP* mesurée dans les protoplastes de mousse. Les groupes significatifs ont été obtenus grâce à un test de Kruskal-Wallis avec un risque $\alpha=0,05$.

testé comporte le domaine B2 et le domaine B3. En conclusion, les données prédites par *ADpred* ont pu être validées fonctionnellement dans la levure.

3.3. L'absence de DA réduit l'activité transcriptionnelle de AtLEC2 dans la mousse

3.3.1. Activité transcriptionnelle de LEC2 sur *pOLE1*

Au regard des nouvelles données fonctionnelles obtenues, l'activité transcriptionnelle de différents variants de LEC2 sur *pOLE1::GFP* a de nouveau été testée dans des protoplastes de mousse. Ce travail a été réalisé dans le cadre du stage de Master 1 de Jean Michaux, que j'ai encadré. Les résultats montrent que le domaine d'activation de la transcription situé en partie N-terminale de AtLEC2 est nécessaire à l'activité transcriptionnelle sur *pOLE1* dans la mousse (Figure 2-9). En effet, comme montré précédemment dans les résultats préliminaires (Figure 2-3), lorsque les 86 premiers acides aminés de LEC2 sont délétés, le facteur perd toute son activité transcriptionnelle bien que les domaines d'activation de la transcription situés en C-ter de LEC2 soient toujours présents. Cependant, lorsque la partie non structurée située en partie C-terminale de LEC2, *i.e.* à partir de Q275, est délétée, l'activité transcriptionnelle de Δ LEC2₂₇₅₋₃₆₃ est significativement réduite bien que toujours présente. Cela sous-entend que le domaine d'activation situé en N-ter serait nécessaire et suffisant pour activer la transcription via *pOLE1*. Cependant, la délétion Δ LEC2₂₇₅₋₃₆₃ ne nous permet pas d'affirmer cette hypothèse, car le domaine situé à la fin du B3, entre M216 et Q275 est toujours présent et possiblement actif. Enfin, la présence de ce dernier domaine n'est pas suffisante pour l'activité transcriptionnelle de LEC2 lorsque ses domaines N et C-ter sont délétés. En effet, LEC2 M104-Q275 n'a aucune activité transcriptionnelle sur *pOLE1* dans les protoplastes de mousse.

3.3.2. Activité transcriptionnelle de LEC2 sur d'autres promoteurs cibles

La mesure de l'activité transcriptionnelle de LEC2 quantifiée dans les protoplastes de mousse est une expérience réalisée de manière courante dans l'équipe et a permis de montrer que les LAFL forment un complexe pour activer la transcription de leur cible (Thévenin et al., 2012; Baud et al., 2016). Cependant, nous avons pour habitude d'utiliser uniquement *pOLE1* qui permet d'obtenir de bons résultats quantitatifs dans la mousse. Au cours de ma thèse, j'ai voulu tester l'activité transcriptionnelle de LEC2 sur d'autres promoteurs cibles fusionnés à la GFP pour généraliser les résultats obtenus sur *pOLE1*.

À partir de données bibliographiques, de résultats obtenus précédemment au laboratoire et l'analyse de la séquence des promoteurs, notamment la présence d'une boîte RY en 5' du gène, j'ai testé l'activité de LEC2 seul ou en complexe sur trois promoteurs différents : *pAGL15* (AT5G13790),

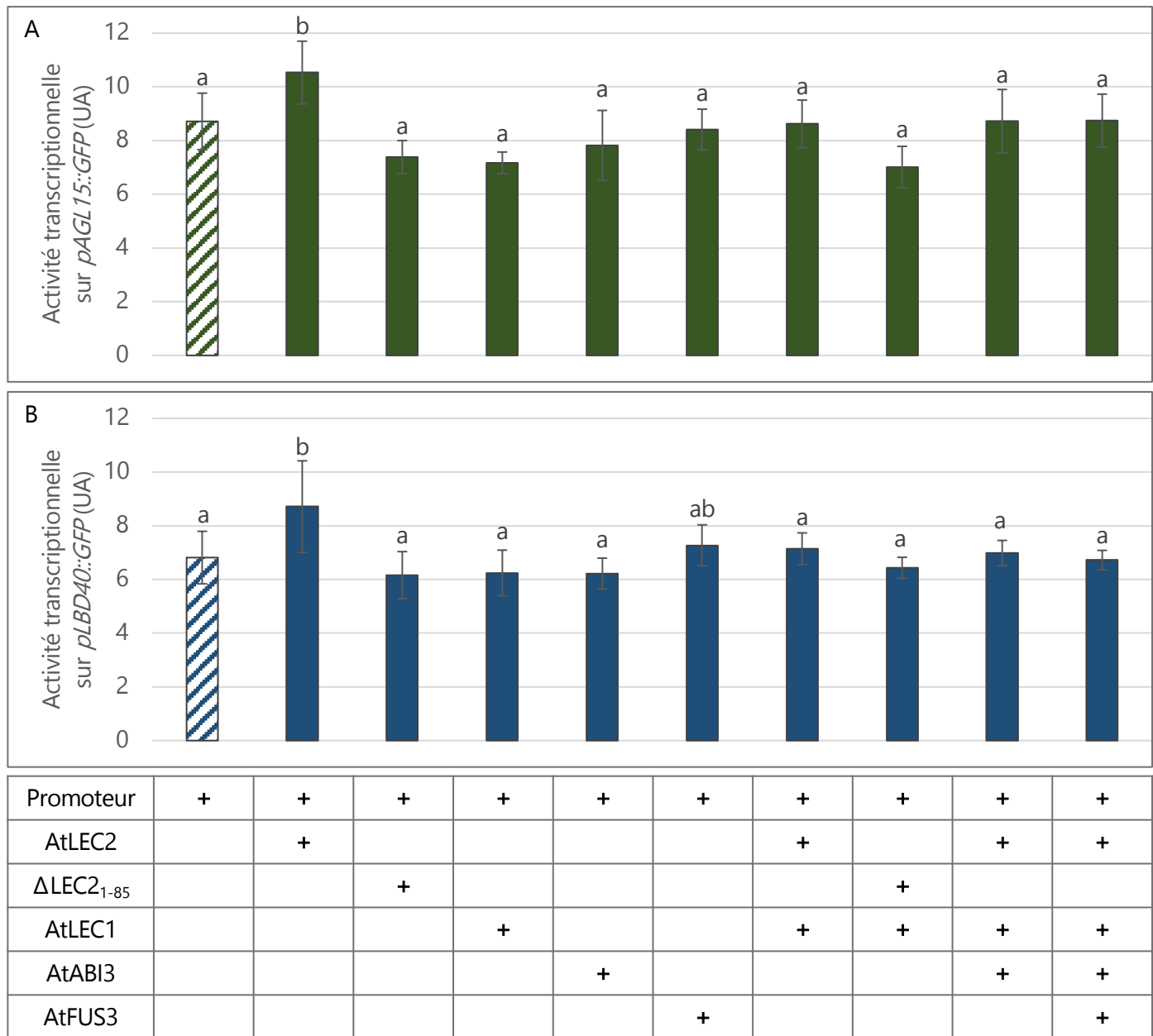


Figure 2-10 : AtLEC2 perd son activité transcriptionnelle sur les promoteurs *pAGL15* et *pLBD40* lorsque le domaine d'activation N-ter est délété. A) Activité transcriptionnelle des LAFL seuls ou en complexe et de Δ LEC2₁₋₈₅ seul ou avec LEC1 sur *pAGL15::GFP* et B) *pLBD40::GFP* dans les protoplastes de mousse. Les groupes significatifs ont été obtenus grâce à un test ANOVA avec un risque $\alpha=0,05$.

pYUC5 (AT5G43890) et *pLBD40* (AT1G67100). *AGL15* est un facteur de transcription impliqué dans le développement de la graine et l'embryogenèse somatique et constitue une cible importante de LEC2 (Heck et al., 1995; Wójcik et al., 2020), *YUC5* est impliqué dans le métabolisme de l'auxine et les gènes *YUC* sont dérégulés chez des surexprimeurs de LEC2 (Wójcikowska et al., 2013). Enfin, *LOB40* est un des gènes graine spécifique les plus exprimés chez les surexprimeurs de LEC2 (Braybrook et al., 2006). Si les résultats obtenus avec *pYUC5* ne sont pas concluants (non présentés), j'ai pu montrer que AtLEC2 a une activité transcriptionnelle sur *pAGL15* et *pLBD40* (Figure 2-10). Toutefois, les niveaux de fluorescence de la GFP obtenus, reflétant l'activité transcriptionnelle des facteurs de transcription sur leur cible sont bien inférieurs à ceux observés en routine avec *pOLE1*. Par ailleurs, LEC2 a une activité transcriptionnelle sur ces deux promoteurs, mais le complexe des LAFL n'a aucune activité coopérative comme dans le cas de *pOLE1*. La co-transfection des protoplastes avec LEC1, ABI3 ou FUS3 semble même empêcher l'activité de LEC2 sur ces promoteurs, la seule augmentation significative de la fluorescence correspondant à celle des protoplastes transfectés avec LEC2 seul et le promoteur cible. Enfin, Δ LEC2₁₋₈₅ n'a aucune activité transcriptionnelle significative sur *pAGL15* et *pLBD40* (Figure 2-10), confirmant les résultats obtenus précédemment sur *pOLE1*.

3.4. Conclusion et discussion

L'algorithme *ADpred* (Erijman et al., 2020) a permis de mettre en évidence la présence de domaines d'activation de la transcription situés en N et C-terminal de LEC2. L'algorithme a été développé à partir d'analyses fonctionnelles de domaines d'activation synthétiques fonctionnels ou non dans la levure et a été testé sur de nombreux facteurs de transcription de levure, de drosophile, d'humain et de virus et il est remarquable que la prédiction fonctionne aussi bien sur des protéines de plantes (A. Erijman, communication personnelle). La fonctionnalité de ces domaines a été confirmée dans la levure grâce à un test d'activation de la transcription, ainsi qu'en expression transitoire dans des protoplastes de mousse sur *pOLE1*, *pAGL15* et *pLBD40*.

Le domaine d'activation situé en N-ter de LEC2 semble nécessaire et suffisant pour activer la transcription de la GFP sous contrôle du *pOLE1*. Selon l'hypothèse de départ de ce chapitre, la perte de fonctionnalité de LEC2 lorsque ses extrémités N ou C-terminales sont délétées, n'est donc pas liée à sa localisation nucléaire, ni *a priori* à son domaine de liaison ou son domaine d'interaction protéine-protéine, mais plutôt à ses domaines d'activation de la transcription.



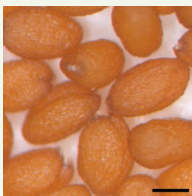


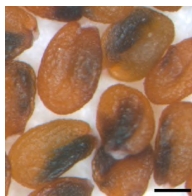
Cette partie a également permis de montrer que LEC2 a une activité transcriptionnelle quantifiable en expression transitoire dans des protoplastes de mousse sur d'autres promoteurs que le *pOLE1*

généralement utilisé au laboratoire. Δ LEC2₁₋₈₅ perd son activité transcriptionnelle sur les trois types de promoteurs testés. Le choix de *pAGL15* et *pLBD40* n'est pas anodin, car si *OLE1* est un gène impliqué dans la formation des corps lipidiques (Siloto et al., 2006; Miquel et al., 2014) et est donc une cible de LEC2 dans un contexte de maturation de la graine, *AGL15* est exprimé pendant l'embryogenèse et est impliqué dans la signalisation hormonale (Heck et al., 1995; Zheng et al., 2009) et *LBD40* est un facteur de transcription putatif impliqué dans la formation des frontières des organes (Xu et al., 2016a). Ces résultats suggèrent donc que LEC2 peut activer la transcription de gènes liés à des processus de maturation comme *OLE1* et que son activité est plus forte lorsqu'il est en complexe avec ABI3, FUS3 et LEC1 (Baud et al., 2016; Boulard et al., 2018). À l'inverse, LEC2 est capable d'activer la transcription de gènes liés à des processus d'embryogenèse comme *AGL15* et *LBD40*, mais n'a pas besoin des autres membres des LAFL, qui ont plutôt, au contraire, tendance à diminuer son activité transcriptionnelle. En effet, LEC2 seul active la transcription de la *GFP* sous contrôle de *pAGL15* et *pLBD40*, mais n'a plus aucune activité en complexe. L'activité de LEC2 mesurée étant tout de même très faible sur *pAGL15* et *pLBD40*, en comparaison avec les résultats obtenus généralement avec *pOLE1*, cela suggère que LEC2 a peut-être besoin de partenaires différents des LAFL pour son activité transcriptionnelle sur les gènes liés à l'embryogenèse, ou que l'activité transcriptionnelle sur des gènes régulateurs ou de transduction comme *LBD40* ou *AGL15* ne nécessite pas des niveaux aussi importants que les protéines de réserves. Cela va dans le sens de notre hypothèse, développée dans la revue présentée en introduction – partie 4 : [Revue : « Contrôle génétique et moléculaire de l'embryogenèse somatique »](#), que LEC2 pourrait avoir plusieurs rôles, un dans la régulation des gènes de maturation, et l'autre à des gènes d'embryogenèse. Ces différences de rôles peuvent être fonction d'une variété de gènes ciblés par LEC2 ainsi que de ses partenaires.

4. Fonctionnalité de LEC2 sans domaine d'activation *in planta*

Les expériences précédentes ont permis de mieux caractériser les domaines fonctionnels de LEC2 dans un contexte hétérologue. La validation de ces domaines *in planta* est une vérification importante de l'activité de ces domaines pour affiner nos connaissances sur le mode d'action de LEC2. Afin de savoir si LEC2 sans domaine d'activation de la transcription est capable de restaurer le phénotype sauvage dans des plantes *lec2-1*, de nombreux variants de LEC2 sans domaine d'activation, délétés en différents points des parties N et C-terminales ont été clonés sous contrôle du promoteur *pLEC2* et en fusion avec la GFP, comme expliqué dans la partie [2.3. Localisation de variants de LEC2 *in planta*](#). La restauration du phénotype sauvage dans des plantes *lec2-1* a été

A

Fond génétique	Ws	<i>lec2-1</i> (homozygote)	<i>lec2-1</i> (homozygote)	<i>lec2-1</i> (homozygote)	<i>lec2-1</i> (homozygote)	<i>lec2-1</i> (homozygote)
t-DNA	/	/	pLEC2::GFP: AtLEC2: tRbcSE9	pLEC2::GFP: Δ LEC2 ₁₋₈₅ : tRbcSE9	pLEC2::GFP: Δ LEC2 ₂₇₅₋₃₆₃ : tRbcSE9	pLEC2::GFP: LEC2 M104-Q275: tRbcSE9
Insertion	/	/	Homozygote	Homozygote	Hétérozygote	Hétérozygote
Ségrégation	0% [<i>lec2</i>]	100% [<i>lec2</i>]	0% [<i>lec2</i>]	0% [<i>lec2</i>]	25% [<i>lec2</i>]	100% [<i>lec2</i>]
Phénotype des graines						
Complémentation	/	/	Oui	Oui	Oui	Non

B

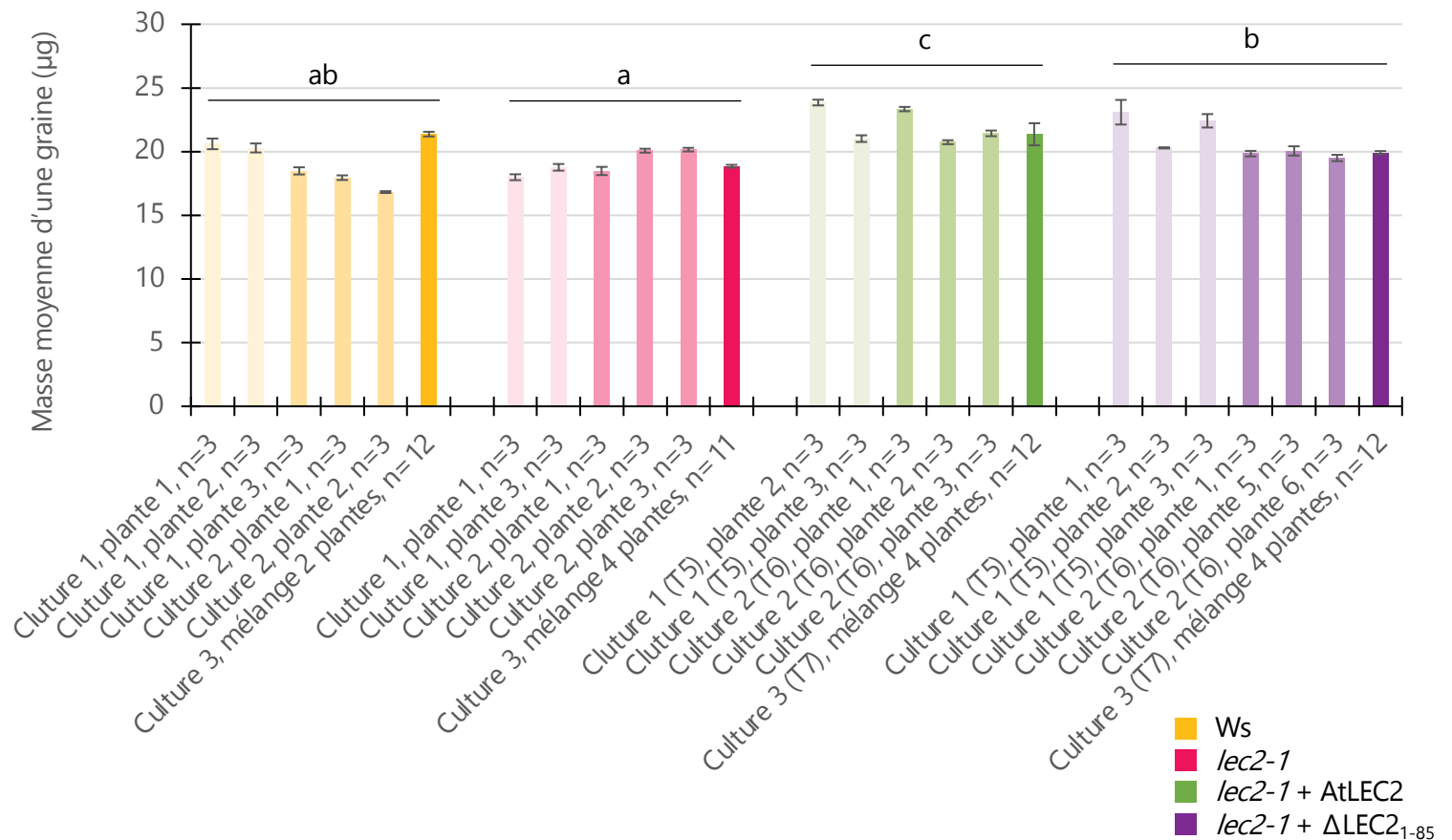


Figure 2-11 : LEC2 sans domaine d'activation restaure le phénotype sauvage. A) Tableau de complémentation du mutant *lec2* par différents variants avec ou sans domaine d'activation de la transcription, géotypes et phénotypes observés. B) Masse moyenne d'une graine, calculée à partir d'échantillons de 50 graines. Les différentes nuances de couleurs représentent des cultures indépendantes. Au minimum 3 échantillons de 50 graines ont été prélevés sur 2 à 4 plantes, issues de 3 cultures différentes. La déviation standard représente l'écart-type standard divisé par la racine carrée du nombre d'échantillons. Les lettres représentent les groupes significatifs obtenus grâce à un test ANOVA avec un risque $\alpha=0,05$, réalisé sur l'ensemble des valeurs issues de chaque géotype et des différentes cultures. Barres d'échelle : 200 μ m.

observée. La teneur des graines en protéines a été quantifiée en spectroscopie infrarouge (NIRS) et par SDS-PAGE. L'accumulation d'OLE1 et la mobilisation des corps lipidiques ont également été testées par western-blot et microscopie confocale, et la teneur et la composition en lipides a été testée par NIRS et par chromatographie en phase gazeuse (GC). Ces analyses ont été réalisées tout au long de la thèse, et en fonction de l'avancée du projet et des résultats obtenus dans d'autres parties, ce ne sont pas toujours les mêmes variants qui ont été étudiés et les résultats décrits ci-dessous ne sont pas forcément présentés dans l'ordre chronologique. La majorité des analyses ont été réalisées sur des délétions de la partie N-ter, notamment Δ LEC2₁₋₈₅ car il s'agit du variant perdant toute son activité transcriptionnelle dans la mousse tout en gardant sa localisation nucléaire *in planta*.

4.1. LEC2 sans domaine d'activation restaure le phénotype sauvage *in planta*

Comme nous l'avons vu dans la partie introductive, le phénotype des mutants *lec2* est caractérisé par la présence de trichomes sur les cotylédons qui sont ainsi modifiés en feuilles, l'accumulation d'anthocyanes, visibles sur les cotylédons matures et sur les graines, ainsi qu'une teneur en lipides plus faible liée à un défaut de maturation des graines (Meinke et al., 1994). Pour tester la complémentation du mutant *lec2-1* par les variants de LEC2 sans domaine d'activation, le premier phénotype observable est celui des graines. Alors que les graines du mutant *lec2-1* ont une tache brune spécifique de l'accumulation d'anthocyanes dans les cotylédons, le contrôle positif de transformation du mutant *lec2-1* par la version native de LEC2, notée AtLEC2, montre bien une restauration du phénotype sauvage des graines (Figure 2-11). Il en va de même pour les variants de LEC2 dont les domaines d'activation présents en N ou en C-terminal de la protéine ont été délétés. Les graines des mutants *lec2-1* exprimant Δ LEC2₁₋₈₅ et Δ LEC2₂₇₅₋₃₆₃ sous contrôle de *pLEC2* ont un phénotype sauvage. Ces variants de LEC2 sans domaine d'activation en N ou en C-ter et présentant un défaut d'activité transcriptionnelle dans les protoplastes de mousse semblent donc capables de restaurer le phénotype visuel des graines *lec2-1*. En revanche, le variant de LEC2 sans N-ter ni C-ter, noté LEC2 M104-Q275 semble incapable de restaurer le phénotype visuel des graines, qui ont un phénotype [*lec2*]. Si au moins une partie des délétions testées restaurent le phénotype sauvage, sont-elles capables de restaurer le contenu en réserves des graines ?

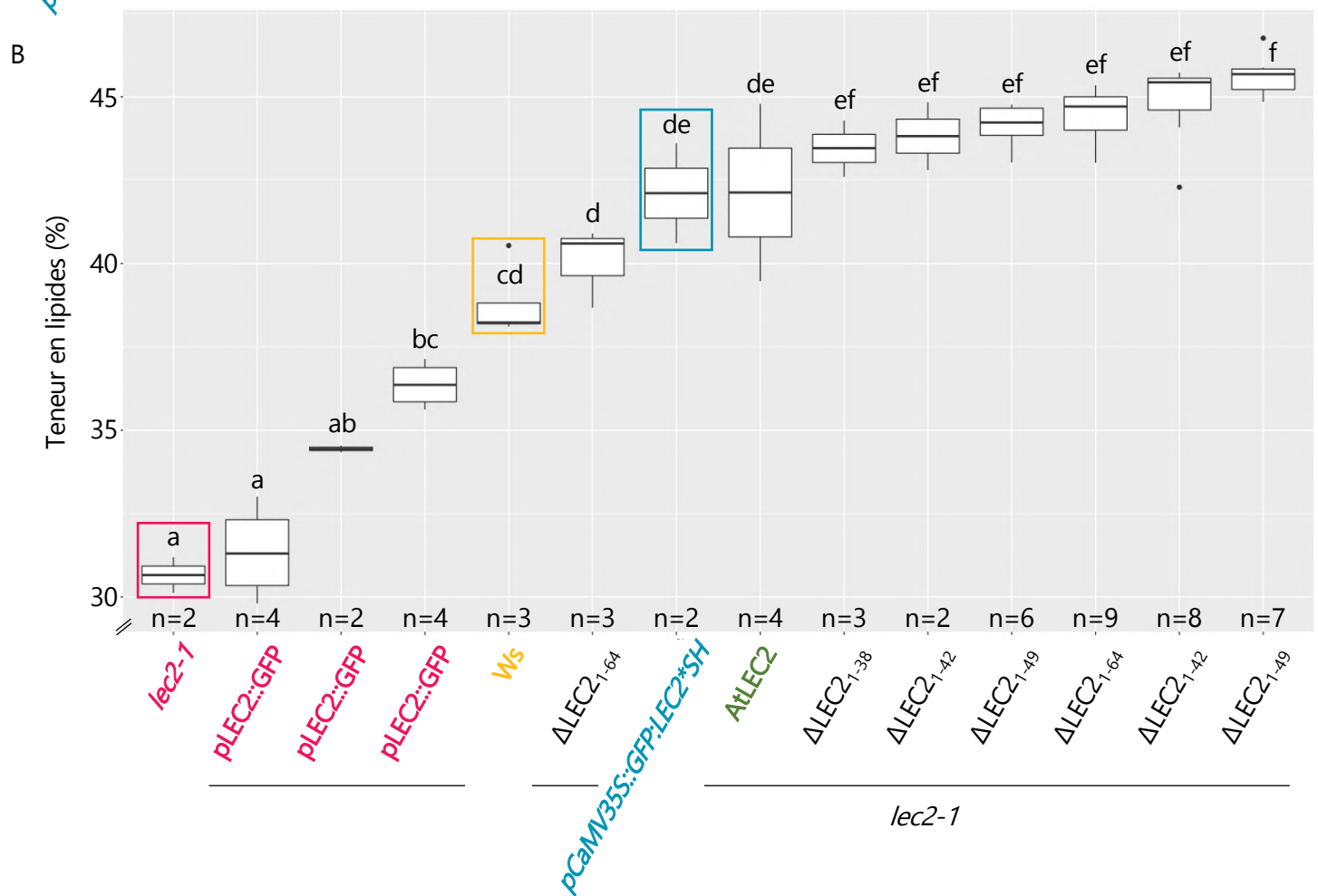
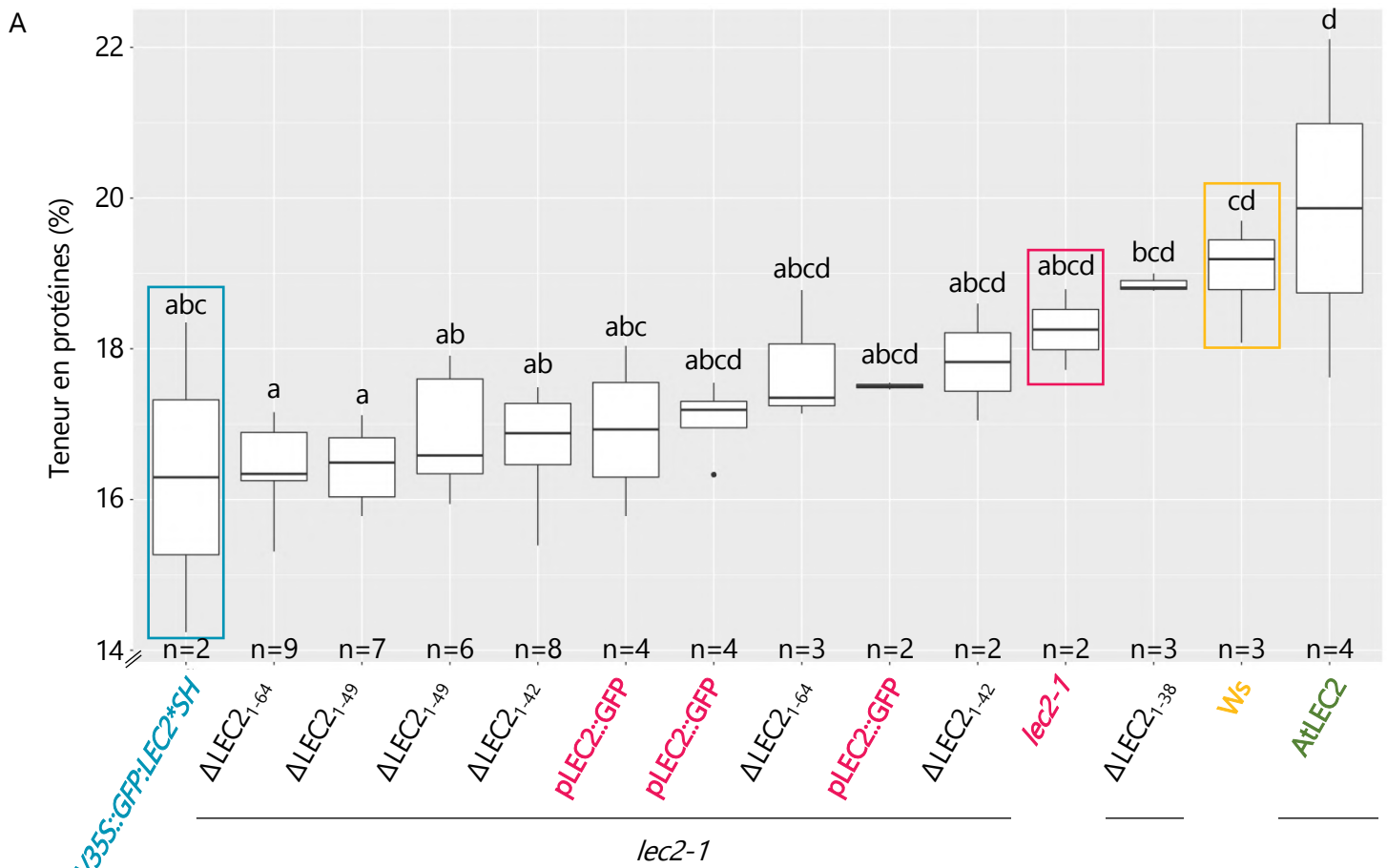


Figure 2-12 : Les différents variants de LEC2 délétés en N-ter semblent réverter le mutant *lec2-1* en terme de teneur en lipides dans les graines. A) La teneur en protéines et B) en lipides des différents génotypes a été mesurée en NIRS. Pour plus de clarté, les génotypes contrôles sont repérés : en rose pour le mutant *lec2-1* et *lec2-1* non complémenté, en jaune pour le sauvage *Ws*, vert pour le mutant *lec2-1* complémenté avec *AtLEC2* et en bleu pour *pCaMV35S::GFP:LEC2*SH*. Les groupes significatifs ont été obtenus grâce à un test ANOVA avec un risque $\alpha=0,05$.

4.2. Effet des délétions sur le contenu en réserves des graines

4.2.1. Masse des graines sèches

Des échantillons de 50 graines ($n \geq 3$) ont été prélevés sur plusieurs plantes issues de trois cultures indépendantes de quatre génotypes : *Ws*, *lec2-1* et *lec2-1* complémenté par AtLEC2 ou LEC2 sans domaine d'activation en N-ter, Δ LEC2₁₋₈₅, sous contrôle de *pLEC2*. La masse moyenne d'une graine a été déterminée à partir de la moyenne de 50 graines (Figure 2-11B). Nous pouvons constater que les graines issues de plantes *lec2-1* sont les plus légères et que celles des mutants *lec2-1* complémentés par AtLEC2 ou Δ LEC2₁₋₈₅ sont significativement plus lourdes que celles du mutant.

4.2.2. Analyse des teneurs en protéines et lipides par spectroscopie infrarouge

Des graines du mutant *lec2-1* exprimant plusieurs variants de LEC2 présentant des délétions dans la partie N-terminale sous contrôle du promoteur *pLEC2* ont été analysées par spectroscopie infrarouge (NIRS) afin de déterminer leur teneur en protéines et en lipides. Pour que cette expérience soit optimale, il faut que les différents génotypes testés soient tous issus de la même culture, traités de la même manière et que les graines analysées soient récoltées sur au moins 6 individus par génotype. Les résultats présentés dans la Figure 2-12 n'ont pas été obtenus en respectant tous ces critères : tous les génotypes testés n'ont pas été cultivés au même moment et le nombre d'individus indépendants pour chaque génotype n'est pas toujours supérieur ou égal à 6. Cependant, en première approche, l'expression d'un variant de LEC2 sans partie N-terminale semble bien permettre de restaurer la teneur en protéines et lipides du mutant *lec2*. Concernant la teneur en protéines (Figure 2-12A), il semble qu'il n'y ait pas de différence significative entre les différents lots de graines testés, et même les valeurs obtenues pour les différents contrôles *pCaMV35S::GFP:LEC2*SH*, *lec2-1* et *Ws* ne sont pas significativement différentes. Cependant, la teneur en lipides des graines montre des différences entre les génotypes contrôles (Figure 2-12B). En effet, la teneur en lipides des graines *lec2-1* est significativement inférieure à celle des graines *Ws* et *pCaMV35S::GFP:LEC2*SH*, bien que les résultats ne soient pas significativement différents entre le sauvage et le surexpresser de LEC2*SH. Les résultats obtenus montrent également que la teneur en lipides des différents génotypes mutants *lec2-1* complémentés avec les variants en N-ter de LEC2 est significativement différente de celle du mutant *lec2-1* et des génotypes contrôles *lec2-1* exprimant la GFP seule sous contrôle de *pLEC2*. Ainsi, bien que l'analyse NIRS n'ait pas été réalisée dans des conditions optimales et que les résultats doivent être confirmés en la répétant dans de bonnes conditions, il semble que les délétions en N-ter de LEC2 restaurent le phénotype visuel des graines *lec2* ainsi que le contenu des graines en lipides.

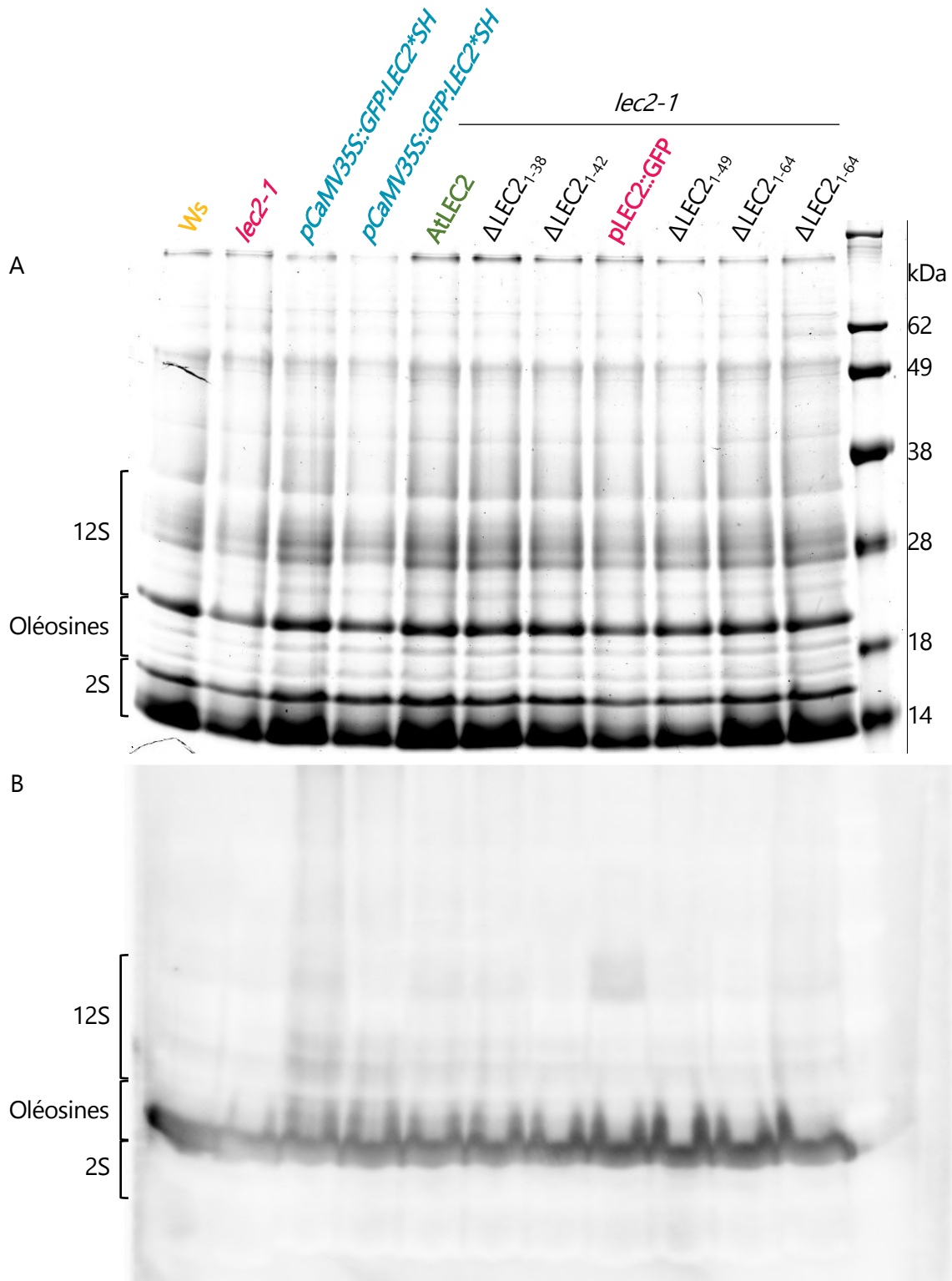


Figure 2-13 : La composition des graines en protéines de réserve ne semble pas modifiée dans les mutants *lec2-1* exprimant différents variants de LEC2 délétés en N-ter, de même que l'accumulation d'OLE1. A) Gel d'électrophorèse de protéines de réserves (SDS-PAGE), coloré au bleu de Coomassie et B) western-blot anti-OLE1, réalisés sur des extraits de protéines de réserve de graines d'environ 4 graines. Pour plus de clarté, les génotypes contrôles sont repérés : en rose pour le mutant *lec2-1* et *lec2-1* non complémenté, en jaune pour le sauvage Ws, vert pour le mutant *lec2-1* complémenté avec *AtLEC2* et en bleu pour *pCaMV35S::GFP:LEC2*SH*. La position des protéines 12S, 2S et des oléosines est représentée.

4.2.3. Analyse de la composition en protéines de réserve et expression de OLE1

Afin de poursuivre l'étude de la restauration de fonction, une analyse de la composition en protéines de réserve a été menée par séparation des protéines sur gel (SDS-PAGE) suivie d'un western-blot avec un anticorps anti-OLE1 afin de détecter l'accumulation d'OLE1 dans les graines (Figure 2-13). Les profils de protéines de réserves observés sur chaque génotype semblent ne pas montrer de différence de quantité ou de composition des protéines dans les graines testées. En effet, les mêmes bandes sont présentes sur chaque piste du gel SDS-PAGE (Figure 2-13A) et aucun génotype ne montre de bandes manquantes ou supplémentaires. Sur certaines pistes, l'intensité des bandes semble légèrement différente, mais peut être liée aux variations de la quantité de matériel déposé sur le gel. Le western-blot, montre que OLE1 est accumulée dans les graines de chacun des génotypes testés (Figure 2-13B).

Enfin, la mobilisation et la taille des oléosomes a également été observée par microscopie confocale sur des embryons imbibés 1h, 24h et 48h et colorés au rouge NIL. Aucune différence n'a été observée au sein des différents génotypes testés (données non montrées). Comme il ne semble pas y avoir de différence au niveau des protéines de réserve, de l'accumulation des oléosines et de la formation des corps lipidiques entre le mutant *lec2* et le sauvage *Ws*, on ne peut pas conclure sur l'effet des variants de LEC2 délétés en N-ter sur la restauration du phénotype *lec2-1* à ce niveau.

4.2.4. Analyse de la teneur en lipides par chromatographie en phase gazeuse

Afin de valider les observations faites en spectroscopie infrarouge (NIRS) sur les teneurs en lipides des graines, une analyse des lipides contenus dans les graines a été réalisée par chromatographie en phase gazeuse (GC). Les acides gras ont été extraits des graines et leur quantité a été mesurée, indiquant de manière très claire, que le mutant *lec2-1* a une teneur plus faible que le sauvage *Ws*, confirmant les résultats de NIRS et les données issues de la bibliographie (Figure 2-14). Les résultats montrent également que la teneur en acides gras contenus dans les graines des mutants *lec2-1* complétés par le contrôle *AtLEC2* et le variant sans domaines d'activation Δ LEC2₁₋₈₅ est semblable à celle du sauvage et significativement différente du mutant *lec2-1*. Ces résultats confirment donc les données de NIRS et nous permettent d'affirmer que les variants de LEC2 sans domaine d'activation permettent de restaurer la fonction de *lec2-1* au moment de la maturation de la graine.

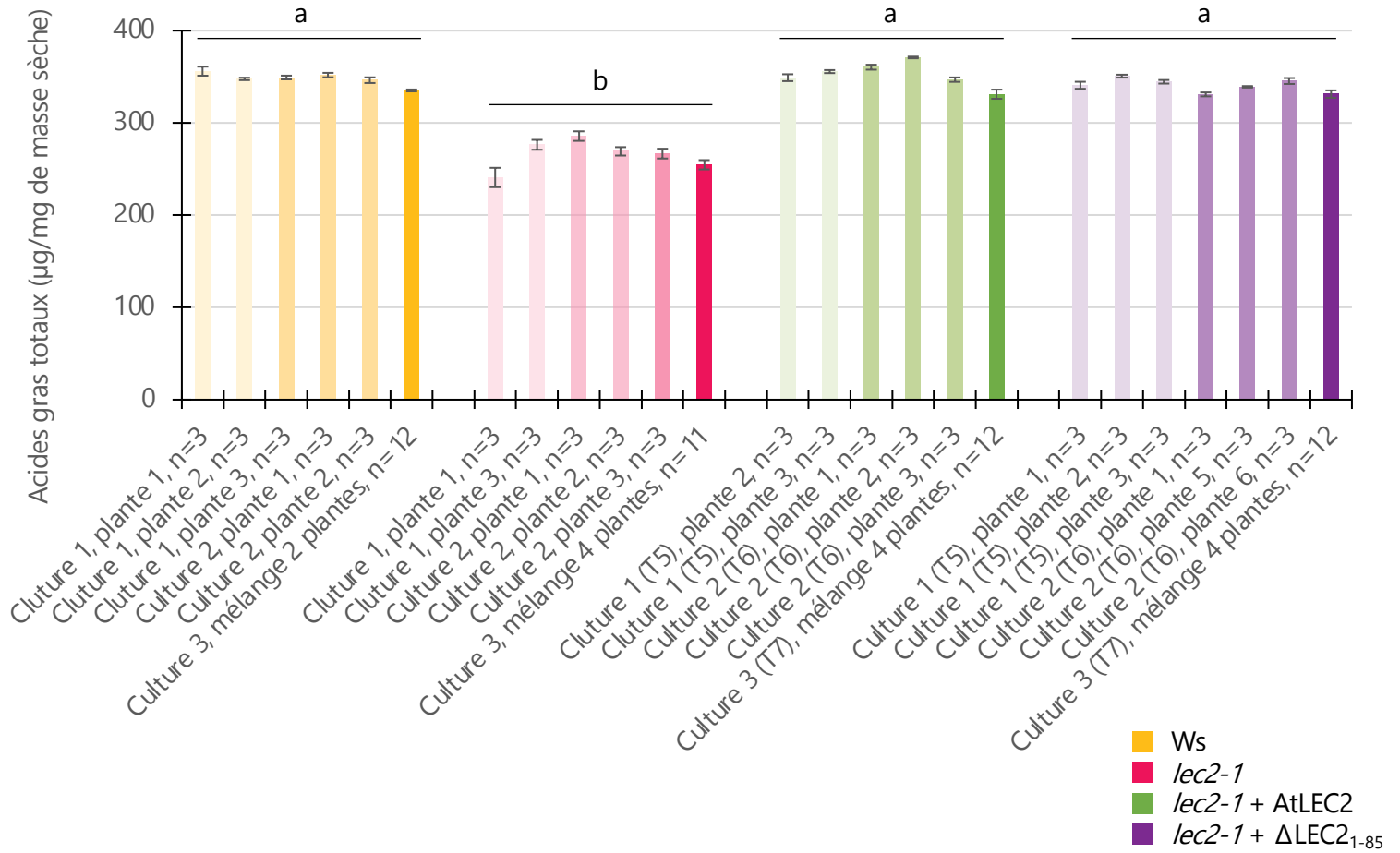


Figure 2-14 : L'expression de Δ LEC2₁₋₈₅ en fond mutant *lec2-1* restaure le phénotype sauvage au niveau de la teneur en acides gras totaux des graines. B) Teneur en acides gras totaux en µg/mg de graine sèche, déterminée par chromatographie en phase gazeuse (GC). Les différentes nuances de couleurs représentent des cultures indépendantes. Au minimum 3 échantillons de 50 graines ont été prélevés sur 2 à 4 plantes, issues de 3 cultures différentes. La déviation standard représente l'écart-type standard divisé par la racine carrée du nombre d'échantillons. Les lettres représentent les groupes significatifs obtenus grâce à un test ANOVA avec un risque $\alpha=0,05$, réalisé sur l'ensemble des valeurs issues de chaque génotype et des différentes cultures.

4.3. Conclusion et discussion

Dans cette partie, nous avons démontré que différents variants de LEC2, dont le domaine d'activation de la transcription a été délété, sont capables de restaurer le phénotype sauvage chez le mutant *lec2-1*, alors que ces mêmes variants ont une activité transcriptionnelle fortement altérée, voire nulle, dans des protoplastes de mousse. En effet, comme montré dans la Figure 2-11, l'expression de LEC2 sans domaine d'activation en N-ter (Δ LEC2₁₋₈₅) ou en C-ter (Δ LEC2₂₇₅₋₃₆₃) sous contrôle de *pLEC2* chez le mutant *lec2-1* montrent clairement un phénotype sauvage. Le variant LEC2 M104-Q275, quant à lui, ne permet pas de restaurer le phénotype sauvage des graines et ne semble donc pas être fonctionnel *in planta*. Cependant, une grande partie de la protéine a été délétée chez ce variant, qui ne contient plus que son domaine B2 d'interaction protéine-protéine, son domaine B3 de liaison à l'ADN et probablement son NLS, même si ce dernier point n'a pas pu être prouvé dans la partie 2. Avec de telles délétions, la structure de la protéine peut potentiellement être affectée. Comme discuté dans la partie 2.4, le variant LEC2 M104-Q275 n'est pas détecté au noyau dans les embryons testés (Figure 2-6J-L). Il est donc possible que, comme les variants délétés en N ou en C-terminal, LEC2 M104-Q275 soit toujours capable de se lier à l'ADN (domaine B3) et d'avoir des interactions protéine-protéine (domaine B2) *in planta*, mais ne soit plus capable d'aller au noyau pour assurer sa fonction. Néanmoins, LEC2 M104-Q275 n'a pas montré d'activité transcriptionnelle dans la levure non plus, bien que le domaine d'activation présent en fin de domaine B3 soit toujours présent dans cette protéine tronquée (Figure 2-8). Une expérience de simple hybride va être réalisée afin de se savoir si LEC2 M104-Q275 est capable de se lier à *pOLE1* dans la levure grâce à son domaine B3.

Une quantification de la teneur des graines en lipides et protéines par spectroscopie infrarouge (NIRS) a été réalisée en première approche sur des variants de LEC2 délétés en partie N-terminale et, bien que les conditions de cultures et de réplicats ne soient pas réunies pour que les résultats soient incontestables, il semble que la teneur en lipides des mutants complémentés soit plus proche de celle de graines sauvages que mutantes. Les observations faites au NIRS ont été confirmées pour les lipides de manière bien plus robuste par chromatographie en phase gazeuse, prouvant que la maturation de la graine est restaurée dans les mutants *lec2-1* complémentés. L'analyse du contenu par électrophorèse SDS-PAGE et par western-blot ne permet pas de voir des différences entre les différents génotypes testés. Si la réversion phénotypique visuelle est très nette, il semble donc que le contenu des graines en composés de réserve soit également restauré.

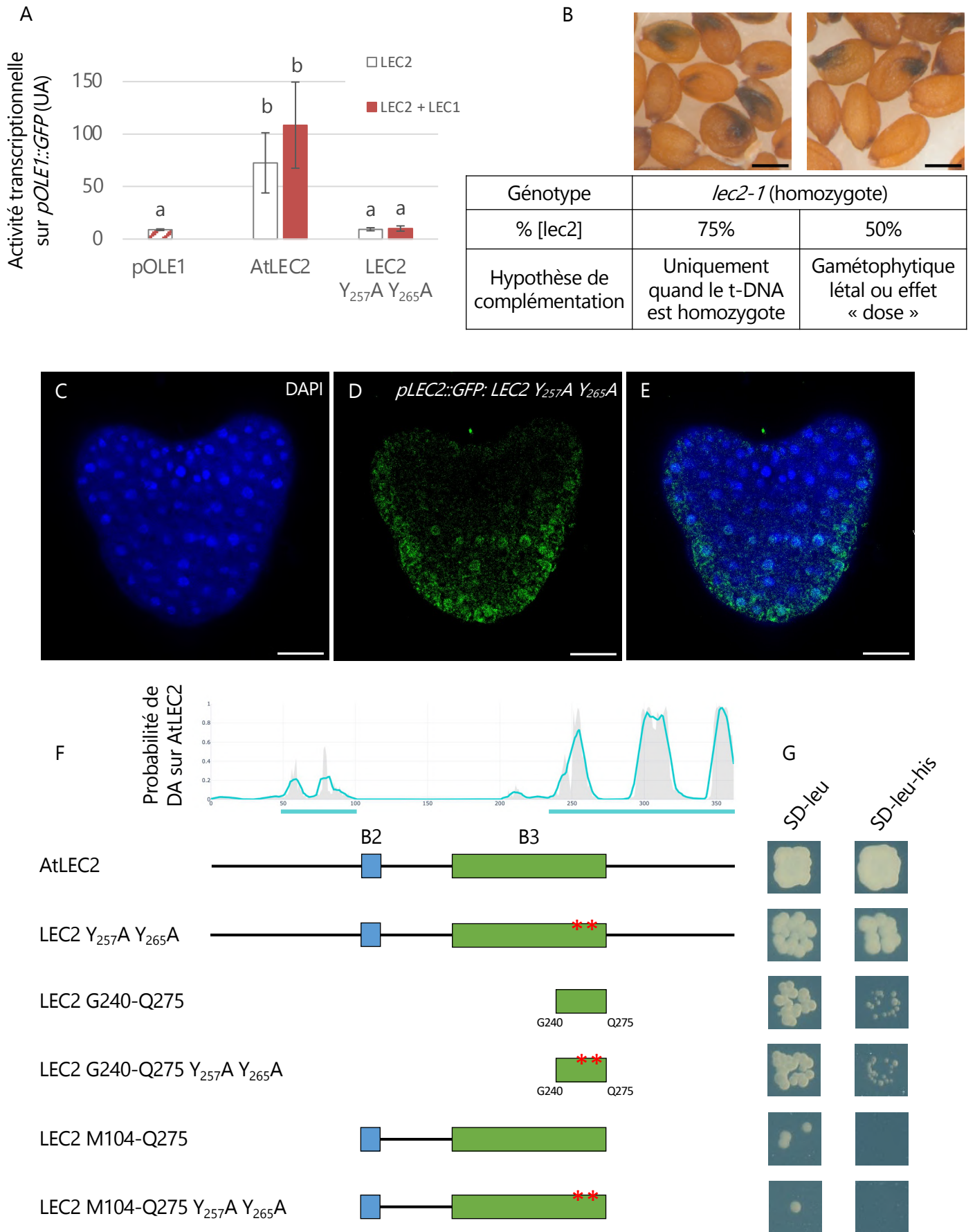


Figure 2-15 : LEC2 $Y_{257}A Y_{265}A$ perd sa fonctionnalité dans la mouche et *in planta*. A) Activité transcriptionnelle de AtLEC2 et LEC2 $Y_{257}A Y_{265}A$ sur *pOLE1::GFP* seuls ou en complexe avec LEC1 dans les protoplastes de mouche. B) Complémentation du mutant *lec2-1* avec le variant LEC2 $Y_{257}A Y_{265}A$. C-E) Localisation de *pLEC2::GFP:LEC2 Y₂₅₇A Y₂₆₅A* dans un embryon au stade cœur. F) Représentations schématiques de LEC2 et autres variants avec la mutation $Y_{257}A Y_{265}A$ et G) test d'activité transcriptionnelle dans la levure. Barres d'échelle : B) 200 μ m ; C-E) 15 μ m.

LEC2 étant un facteur de transcription, afin d'avoir une meilleure idée de la fonctionnalité de LEC2 sans domaine d'activation, nous allons, dans le chapitre suivant, réaliser une analyse transcriptomique de mutants *lec2-1* complétés par AtLEC2 en comparaison avec Δ LEC2₁₋₈₅. Si les transcriptomes obtenus sont identiques, alors cela confirmera que LEC2 n'a pas besoin de domaine d'activation pour exercer sa fonction *in planta*, bien qu'elle soit nécessaire dans les protoplastes de mousse.

5. Cas particulier : le variant LEC2 Y₂₅₇A Y₂₆₅A

Les données de séquence de LEC2 ont montré une forte conservation de deux tyrosines situées en position 257 et 265. Ces deux tyrosines sont également présentes et conservées chez ABI3, mais pas chez FUS3. De manière intéressante, chez AtLEC2, les deux tyrosines sont positionnées précisément au niveau du domaine d'activation de la transcription prédit à la fin du domaine de liaison à l'ADN (B3), en position extérieure relativement au sillon de l'ADN. La tyrosine fait partie du groupe des acides aminés à chaînes latérales hydrophobes. Elle possède un hétérocycle portant un groupement hydroxyle, ce qui fait d'elle un acide aminé phosphorylable. Afin de tester l'importance des deux tyrosines sur l'activité transcriptionnelle des facteurs à domaine B3, et notamment de LEC2, elles ont été modifiées par mutagenèse dirigée en alanines, l'acide aminé à chaîne latérale hydrophobe le plus neutre possible, donnant le variant LEC2 Y₂₅₇A Y₂₆₅A. De la même manière que pour les variants de LEC2 portant des délétions dans les extrémités N et/ou C-terminales, la fonctionnalité de LEC2 Y₂₅₇A Y₂₆₅A a été testée dans les protoplastes de mousses et *in planta*, en fond mutant *lec2-1*. La localisation de la protéine *in planta* a été observée grâce à une fusion avec la GFP et l'activité transcriptionnelle de différents variants de LEC2 présentant la mutation Y₂₅₇A Y₂₆₅A a été testée dans la levure.

Dans les protoplastes de mousses, la protéine LEC2 Y₂₅₇A Y₂₆₅A montre une perte totale d'activité transcriptionnelle sur *pOLE1::GFP* en complexe ou non avec LEC1 (Figure 2-15A). *In planta*, l'expression de LEC2 Y₂₅₇A Y₂₆₅A en fusion avec la GFP et sous contrôle du promoteur *pLEC2*, en fond mutant *lec2-1*, a montré que la mutation LEC2 Y₂₅₇A Y₂₆₅A ne permet pas de restaurer complètement le phénotype sauvage des graines (Figure 2-15B). En effet, les différents lots de graines observés, issus de transformants indépendants et dont les plantes mères étaient homozygotes pour la mutation *lec2-1*, présentent des ségrégations avec 75% ou 50% de phénotype [*lec2*]. Cette ségrégation suggère un effet « dose » de LEC2 Y₂₅₇A Y₂₆₅A, qui ne serait alors pas capable de réverter le phénotype soit à l'état homozygote ou soit à un niveau d'expression élevé. LEC2 Y₂₅₇A Y₂₆₅A n'a donc aucune activité transcriptionnelle sur *pOLE1::GFP* dans les protoplastes

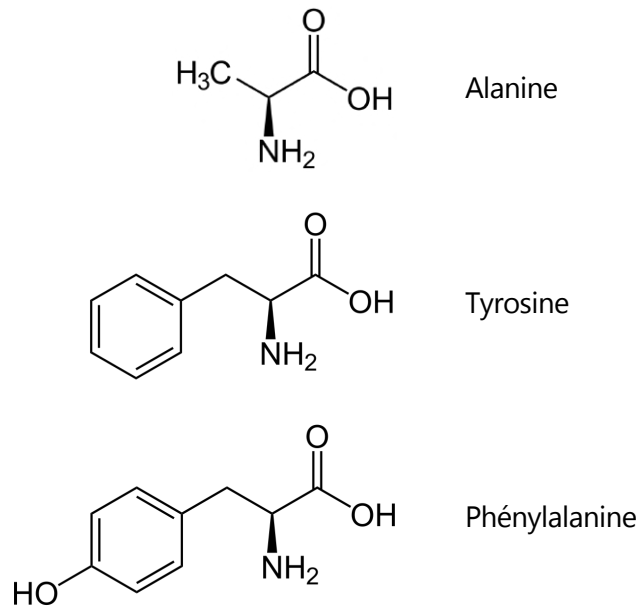


Figure 2-16 : Structure de l'alanine (A), de la tyrosine (Y) et de la phénylalanine (F), membres du groupe des acides aminés à chaîne hydrophobe. (Source : wikipedia.org).



Figure 2-17 : Structure 3D du domaine B3 de AtLEC2 lié à l'ADN. Le domaine B3 est représenté en vert, les tyrosines en position 257 et 265 sont colorées en orange et indiquées par la flèche rouge, l'ADN est représenté sous forme de double hélice.

de mousse et est incapable de compléter totalement le mutant *lec2-1*. Cependant, contrairement à LEC2 M104-Q275, qui n'est fonctionnel ni dans la mousse, ni dans la plante, l'observation en microscopie confocale d'embryons exprimant *GFP:LEC2 Y₂₅₇A Y₂₆₅A* sous contrôle de *pLEC2* a montré que ce variant possède toujours un signal nucléaire. Cela indique ainsi que le NLS n'est pas touché et que le facteur de transcription est toujours capable d'entrer au noyau. Néanmoins, cela n'explique pas sa perte de fonction *in planta* et dans la mousse. D'autre part, un test d'activation de la transcription dans la levure a montré que les mutations *Y₂₅₇A Y₂₆₅A* n'affectent en rien la capacité de LEC2 à activer la transcription lorsque la protéine est entière, ni même lorsque l'activité du fragment LEC2 G240-Q275 seul est testée (Figure 2-15F,G). Néanmoins, LEC2 M104-Q275 *Y₂₅₇A Y₂₆₇A* n'a pas d'activité transcriptionnelle, tout comme LEC2 M104-Q275 sans la mutation. Cette version mutée du facteur de transcription est moins active *in planta* et dans la mousse, bien que ni son domaine d'activation, ni son domaine de localisation nucléaire, ni, *a priori*, son domaine d'interaction protéine-protéine ne soient perturbés. Nous pouvons faire l'hypothèse que le domaine B3 et/ou la capacité de liaison à l'ADN de LEC2 sont affectés par la mutation LEC2 *Y₂₅₇A Y₂₆₅A*. Afin de tester cette hypothèse, la capacité de fixation à l'ADN du variant LEC2 *Y₂₅₇A Y₂₆₅A* devra être étudiée au cours d'une expérience de simple hybride dans la levure sur *pOLE1*. La ségrégation du phénotype [*lec2-1*] *in planta* pourrait suggérer que la liaison à l'ADN est moins efficace, mais pas tout à fait nulle, ce qui pourrait expliquer une restauration partielle du phénotype sauvage. Dans le système hétérologue des protoplastes de mousses, LEC2 *Y₂₅₇A Y₂₆₅A* n'a probablement pas tous ses partenaires présents *in planta* et qui pourraient améliorer l'affinité de LEC2 pour ses cibles, par exemple. Cela pourrait donc expliquer pourquoi LEC2 *Y₂₅₇A Y₂₆₅A* n'est pas totalement inactif *in planta*, comme c'est le cas dans la mousse.

Enfin, l'effet de la substitution des tyrosines (Y) en alanines (A) peut être discutée. En effet, comme décrit précédemment, l'alanine est un acide aminé neutre, mais sa structure ne comporte pas de cycle aromatique comme c'est le cas pour la tyrosine (Figure 2-16). L'encombrement de l'alanine est donc bien différent de celui de la tyrosine et peut ainsi modifier la structure et/ou le repliement de la protéine. Nous devrions donc réaliser une autre mutagenèse des deux tyrosines en position 257 et 265 de AtLEC2 en les substituant en phénylalanine (F) qui a la même structure que la tyrosine hormis le groupement hydroxyle. La substitution d'une tyrosine en phénylalanine est également très intéressante car la phénylalanine n'est pas phosphorylable. Ainsi, si les tyrosines *Y₂₅₇* et *Y₂₆₅* de LEC2 sont soumises à des phosphorylations pouvant moduler (positivement ou négativement) l'activité de la protéine LEC2, en modifiant, par exemple, sa capacité de liaison à l'ADN, une phénylalanine pourrait donc améliorer ou au contraire altérer la fonction de LEC2. D'un point de

vue structural, Y257 et Y265 avec I256, L264 et F266 forment un patch hydrophobe faisant partie d'un noyau, lui aussi hydrophobe, de la protéine. Une modification de cette partie de la protéine pourrait déstabiliser le domaine B3 entraînant une perte de liaison à l'ADN et, *in fine*, sa perte de fonction *in planta* (Figure 2-17) (G. Sasnauskas, communication personnelle).

6. Discussion générale sur l'approche avec *a priori* réalisée sur LEC2

L'approche avec *a priori* menée dans cette partie de mon projet avait pour objectif d'obtenir de meilleures connaissances sur la structure et la fonction du facteur de transcription LEC2. Elle a été initiée par l'observation que, lorsque des domaines non structurés et non conservés à travers les espèces sont délétés en parties N et C-terminales de la protéine, son activité transcriptionnelle est fortement réduite voire nulle dans des protoplastes de mousse. En alliant des données de prédiction et leur validation dans différents systèmes d'expression, j'ai pu montrer que LEC2 a un NLS probablement situé entre le domaine B2 et le domaine B3. J'ai également montré que des domaines d'activation de la transcription sont prédits et fonctionnels dans les parties non structurées situées en N et C-terminal de LEC2. J'ai alors confirmé que l'absence de ces domaines d'activation est responsable de la baisse d'activité transcriptionnelle de LEC2, quantifiée en expression transitoire dans des protoplastes de mousse sur le promoteur cible *pOLE1*, mais également sur d'autres cibles : *pAGL15* et *pLOB40*, ce qui montre que ce mode de fonctionnement n'est pas spécifique à *pOLE1*. Cependant, de manière très surprenante, les variants de LEC2 sans domaine d'activation de la transcription semblent fonctionnels *in planta*. Le résultat le plus notable étant pour Δ LEC2₁₋₈₅, dont la partie N-terminale a été délétée, qui n'est plus du tout fonctionnel dans la mousse, mais qui restaure parfaitement (et à tous les niveaux : développement, couleur de la graine, qualité et quantité des réserves) le phénotype sauvage *in planta*. Cela pose la question de l'importance et du rôle des domaines d'activation de LEC2 pour sa fonction.

On sait que LEC1 (NF-YB9) est un facteur de transcription membre du groupe des NF-Y et que les NF-Y ont un rôle de facteurs pionniers (Nardini et al., 2013; Oldfield et al., 2014). En effet, NF-YB et NF-YC forment un dimère capable d'écarter les nucléosomes en mimant un dimère H2A-H2B. NF-YA quant à lui assure la liaison à la *CCAAT box* via son hélice α . LEC1 est actuellement un des seuls facteurs pionniers à avoir été identifiés chez les plantes (Tao et al., 2017; Lai et al., 2018). Comme évoqué dans le chapitre introductif, lors du processus de vernalisation, LEC1 est capable de se lier à l'ADN nucléosomal du locus *FLC*, qui est *silencé* et rend alors la chromatine accessible pour LEC2 et FUS3. Ces derniers peuvent alors se lier, via leur domaine B3, aux motifs RY présents dans des cibles *CME* (*Cold Motif Element*) de *FLC* et recruter FRI ainsi que des protéines comme EFS, une

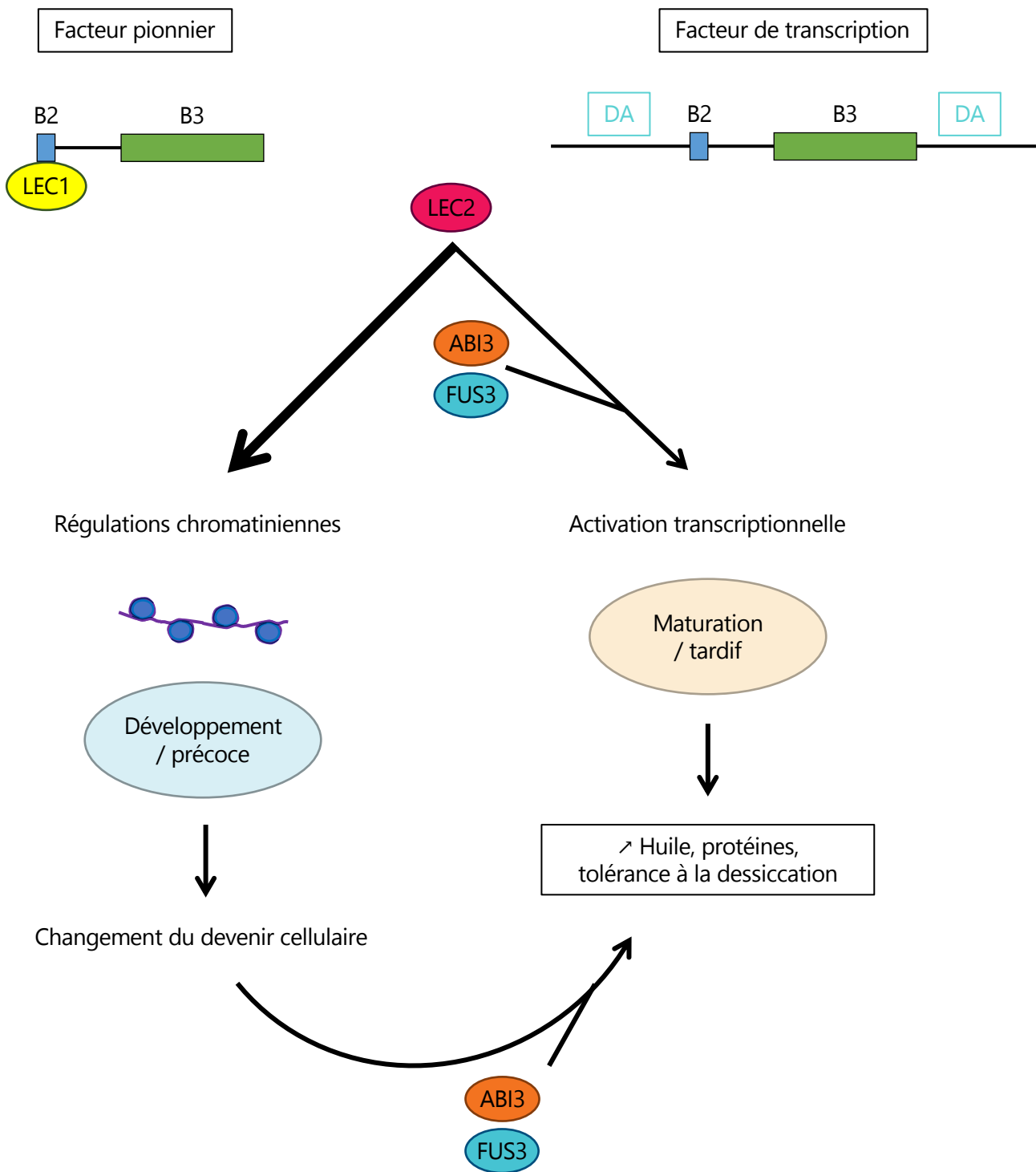


Figure 2-18 : LEC2 aurait une double fonction de facteur pionnier et de facteur de transcription. LEC2 en tant que facteur de transcription induirait l'expression de gènes liés à la maturation de la graine en lien avec ABI3 et FUS3. En tant que facteur pionnier, la fonction de LEC2 ne nécessiterait pas de domaines d'activation. En complexe avec LEC1, il se lierait à ses cibles et induirait un remodelage de la chromatine, rendant accessibles des gènes liés au développement précoce. Cette activité de pionnier induirait un changement du devenir cellulaire induisant l'expression de ABI3 et FUS3 dont l'expression contrôle la maturation de la graine.

méthyltransférase. Cette dernière est capable d'induire des marques actives H2K36me3 et ainsi ouvrir la chromatine au locus *FLC* qui pourra être transcriptionnellement actif (Tao et al., 2017, 2019). Nous savons également que LEC1 forme un complexe avec NF-YC et se lie au domaine B2 de LEC2 (Baud et al., 2016; Boulard et al., 2018). Nous faisons donc l'hypothèse que LEC2 pourrait remplacer NF-YA et former un complexe avec LEC1 (NF-YB9) et NF-YC. Dans ce cas, LEC2 assurerait la liaison du complexe NF-Y à l'ADN grâce à son domaine B3 en ciblant, non pas la *CCAAT box*, comme c'est le cas pour NF-YA, mais des motifs RY ou G-box. Du fait de la spécificité de ses domaines de liaison, LEC2 pourrait alors guider le complexe LEC1/NF-YC sur différentes cibles et ainsi faire partie d'un complexe pionnier avec LEC1. Dans ce cas, LEC2 n'aurait pas besoin de son activité transcriptionnelle, mais uniquement de sa fonction de liaison à l'ADN, ce qui pourrait expliquer pourquoi LEC2 sans domaine d'activation n'a aucune activité transcriptionnelle dans la mousse mais garde toute sa fonctionnalité *in planta*. Nous avons vu que les domaines d'activation de la transcription de LEC2 sont fortement conservés à travers les espèces, ce qui suggère qu'ils doivent cependant être importants pour la fonction de LEC2. C'est pour ces raisons que nous faisons d'hypothèse d'une double fonction de LEC2, avec et sans domaines d'activation de la transcription (Figure 2-18). En effet, LEC2 pourrait avoir une première fonction d'activateur transcriptionnel, qui est celle que l'on connaît déjà, comme régulateur maître du développement de la graine en activant la transcription de gènes de maturation, en coopération avec *ABI3* et *FUS3* notamment. Mais LEC2 pourrait également avoir une activité de facteur/complexe pionnier, avec LEC1 et NF-YC, en apportant une spécificité de cibles grâce à son domaine B3 de liaison à l'ADN et permettant alors la possible induction de modifications au locus (de la chromatine, de l'ADN, ...). Dans ce cas, la fonction de LEC2 en tant que pionnier serait liée au développement précoce de l'embryon et à un changement de devenir cellulaire ou d'état, en permettant, par exemple, l'ouverture de la chromatine au niveau de certaines cibles sans pour autant activer directement leur transcription. Notre hypothèse est que cette activité de LEC2 comme pionnier serait sa fonction la plus précoce ou peut-être principale, permettant d'induire un état « graine » en fixant le complexe NF-Y au niveau de cibles développementales, induisant ainsi l'ouverture de la chromatine et en permettant également la transcription des gènes *ABI3* et *FUS3*. Leur expression conduirait alors à l'accumulation de réserves dans la graine. Si cette hypothèse est vraie, alors LEC2 devrait être capable de se lier à des cibles sans pour autant activer leur transcription de manière directe.

Pour tester cette hypothèse, nous avons donc réalisé une analyse de transcriptome de graines en développement, au stade globulaire-cœur (stade d'expression de LEC2), sur un génotype sauvage (Ws) et sur le mutant *lec2-1*, complémenté avec AtLEC2 ou Δ LEC2₁₋₈₅. Le but de cette analyse de

transcriptome est de comparer les gènes dont l'expression est activée par AtLEC2 ou Δ LEC2₁₋₈₅, et ainsi vérifier si LEC2 sans domaine d'activation est capable d'induire la transcription des mêmes cibles que AtLEC2. De plus, si les transcriptomes de AtLEC2 et Δ LEC2₁₋₈₅ sont identiques, cela constitue une dernière preuve de la fonctionnalité, *in planta*, de LEC2 sans domaine d'activation de la transcription. Enfin, nous avons réalisé une expérience de ChIP-seq sur les mêmes échantillons afin d'identifier des cibles liées par LEC2 avec ou sans domaine d'activation. La comparaison des résultats obtenus en transcriptome d'une part et en ChIP-seq d'autre part, devrait alors permettre d'identifier des *loci* liés mais non activés par LEC2 et d'éventuelles différences entre AtLEC2 et Δ LEC2₁₋₈₅. Enfin, il est possible que le contrôle de l'expression de certaines cibles de LEC2 soit indirect et passe par le facteur ABI3. Afin de tester cette hypothèse et savoir si la délétion de la partie N-terminale de LEC2 affecte sa fonction, il faudrait tester son effet en fond double mutant *lec2 abi3*.

Chapitre 3 : Double fonction de LEC2

A

Génotypes comparés	Nombre de gènes différentiellement exprimés
<i>Ws</i> vs. <i>lec2-1</i> + <i>AtLEC2</i>	371
<i>Ws</i> vs. <i>lec2-1</i> + Δ LEC2 ₁₋₈₅	48
<i>lec2-1</i> + <i>AtLEC2</i> vs. <i>lec2-1</i> + Δ LEC2 ₁₋₈₅	1039

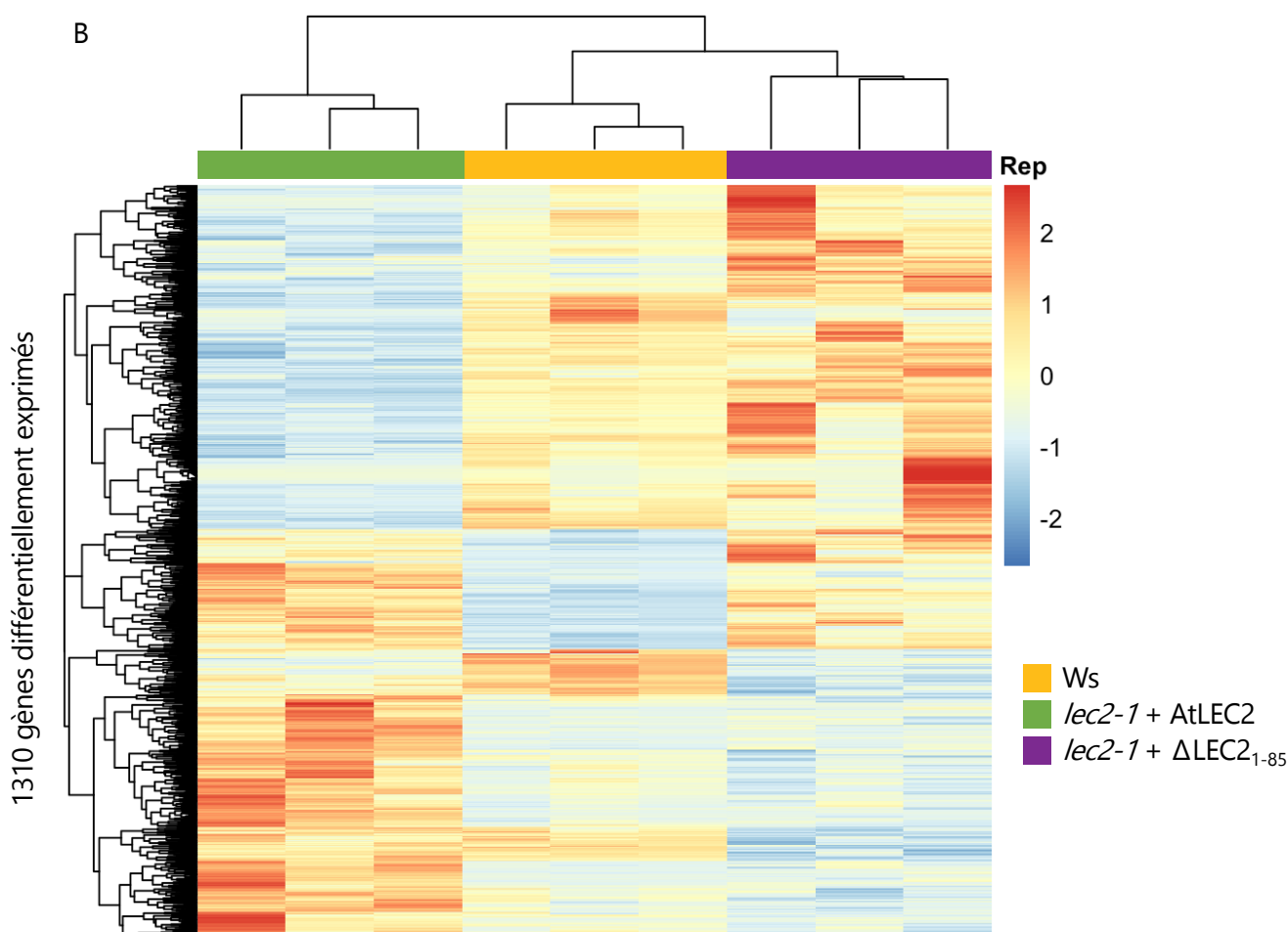


Figure 3-1 : Le transcriptome de graines *lec2-1* complémentées par LEC2 sans domaine d'activation (Δ LEC2₁₋₈₅) est très proche du sauvage *Ws*. A) Nombre de gènes différentiellement exprimés entre les différents génotypes testés. B) Carte de chaleur (*heatmap*) des gènes différentiellement exprimés entre les différentes conditions : graines sauvages *Ws* (jaune), et graines de mutants *lec2-1* exprimant *AtLEC2* (vert) ou Δ LEC2₁₋₈₅ (violet). Les gènes considérés comme différentiellement exprimés sont ceux ayant une *q-value* < 0,05 après normalisation et analyse avec DESeq2 (Love et al., 2014).

Chapitre 3 : Double fonction de LEC2

1. Introduction

L'approche avec *a priori* décrite dans le Chapitre 2 a permis de mettre en évidence la présence de domaines d'activation de la transcription (DA) qui ne semblent pas indispensables à l'activité de LEC2 *in planta*. Nous faisons l'hypothèse que LEC2 puisse avoir une fonction à la fois de facteur de transcription nécessitant ses domaines d'activation, et à la fois de facteur pionnier, ne nécessitant pas ses DA. L'objectif de ce chapitre est de tester cette hypothèse, en alliant l'analyse de différents jeux de données de RNA-seq, CHIP-seq et ATAC-seq, obtenus dans le cadre du projet de thèse, mais également issus de la littérature (Mendoza et al., 2005; Pelletier et al., 2017; Wang et al., 2020; Gao et al., 2022).

2. La perte d'un domaine d'activation ne semble pas impacter le transcriptome des graines en développement

La restauration du phénotype sauvage lors de l'expression d'un variant de LEC2 sans domaine d'activation de la transcription en fond mutant *lec2-1* a été démontrée de différentes manières : phénotype des graines, localisation nucléaire du facteur de transcription, composition en lipides de réserves, etc. Néanmoins, nous avons voulu savoir si, sans domaine d'activation en partie N-ter, Δ LEC2₁₋₈₅ active la transcription des mêmes cibles que la version endogène du facteur de transcription. Pour cela, une expérience de RNA-seq a été mise en place. Des graines au stade cœur de l'embryon ont été prélevées, puis les ARN totaux ont été extraits, sur trois génotypes différents : Ws, et *lec2-1*, complémenté avec une construction *pLEC2::GFP:AtLEC2:trbcSE9* ou *pLEC2::GFP: Δ LEC2₁₋₈₅:trbcSE9*. Les résultats obtenus, c'est-à-dire le nombre de fragments (*reads*) alignés sur chaque gène, ont été normalisés puis la différence d'expression des gènes dans les différentes conditions a été analysée grâce au package R d'analyse d'expression différentielle *DESeq2* (Love et al., 2014). La comparaison des transcriptomes de graines Ws d'une part, avec des graines de mutant *lec2-1* complémenté avec *AtLEC2* ou Δ LEC2₁₋₈₅ d'autre part, montre que les mutants complémentés sont chacun assez proches du sauvage avec environ 370 et 50 gènes différentiellement exprimés par rapport à Ws, respectivement (*q-value* < 0,05 ; Figure 3-1). Bien que les transcriptomes des deux mutants complémentés montrent plus de différence (environ 1000 gènes différentiellement exprimés), il semble que LEC2 sans domaine d'activation (Δ LEC2₁₋₈₅) ait la

même activité sur l'accumulation des transcrits dans les échantillons étudiés que la copie endogène présente dans le sauvage *Ws*. Ces résultats confirment donc ceux obtenus au cours du Chapitre 2 : l'absence de domaine d'activation en partie N-ter n'impacte pas l'accumulation des ARN dans la graine, restaure parfaitement le phénotype sauvage dans un mutant *lec2*.

3. Action transcriptionnelle de LEC2 sur ses cibles

Nous avons vu que les variants de LEC2 avec et sans domaine d'activation de la transcription sont tous deux capables de restaurer le phénotype sauvage. Ainsi, l'absence de domaine d'activation de LEC2 ne modifie pas le transcriptome des graines. Or, nous faisons l'hypothèse que LEC2 puisse avoir une activité de facteur pionnier avec LEC1. Nous avons donc cherché à savoir si LEC2 est capable de se lier à des cibles, sans pour autant activer leur transcription. Pour cela, nous avons réalisé une expérience de ChIP-seq sur des graines en développement, au stade cœur, sur les géotypes *Ws* et mutants *lec2-1* complétés avec *AtLEC2* ou Δ *LEC2*₁₋₈₅, de manière à identifier les *loci* sur lesquels le facteur de transcription se lie. La réalisation d'une expérience de ChIP-seq ciblant un facteur de transcription dans des graines en développement est difficile à mettre en œuvre, en raison de la quantité de matériel nécessaire pour extraire la chromatine et de la faible expression du facteur de transcription. Le ChIP-seq a été réalisé sur une copie de LEC2 étiquetée avec la GFP (voir matériel et méthodes), à partir de plusieurs centaines de graines en développement extraites des siliques. Les résultats de séquençage obtenus ne sont malheureusement pas de très bonne qualité, probablement en raison d'une trop faible quantité de chromatine utilisée pour l'étape d'immunoprécipitation. De ce fait, une grande quantité d'informations est probablement perdue, avec des faux positifs ou faux négatifs, *i.e.* des *loci* ciblés par LEC2 mais pour lesquels on ne discernerait pas de signal. Néanmoins, en attendant d'avoir les résultats d'un réplicat de l'expérience (en cours), nous avons décidé d'utiliser nos résultats de ChIP-seq avec beaucoup de précaution et en filtrant les valeurs les plus significatives, pour une première approche (*fold change* > 10, c'est-à-dire au moins 10 fois plus de signal par rapport au contrôle négatif).

L'objectif de l'analyse est d'identifier des gènes qui seraient liés par LEC2 mais non activés transcriptionnellement. A cette fin, nous avons donc comparé ces résultats de ChIP-seq, réalisé sur GFP:LEC2 avec des données transcriptomiques publiées, obtenues sur des plantules surexprimant *LEC2:GR* sous contrôle d'un *pCaMV35S* et traitées à la dexaméthasone identifiant les cibles primaires ne nécessitant pas de synthèse protéique (Mendoza et al., 2005), ainsi que des données de ChIP-seq réalisé sur LEC1 dans des graines en développement, au stade cotylédons retournés

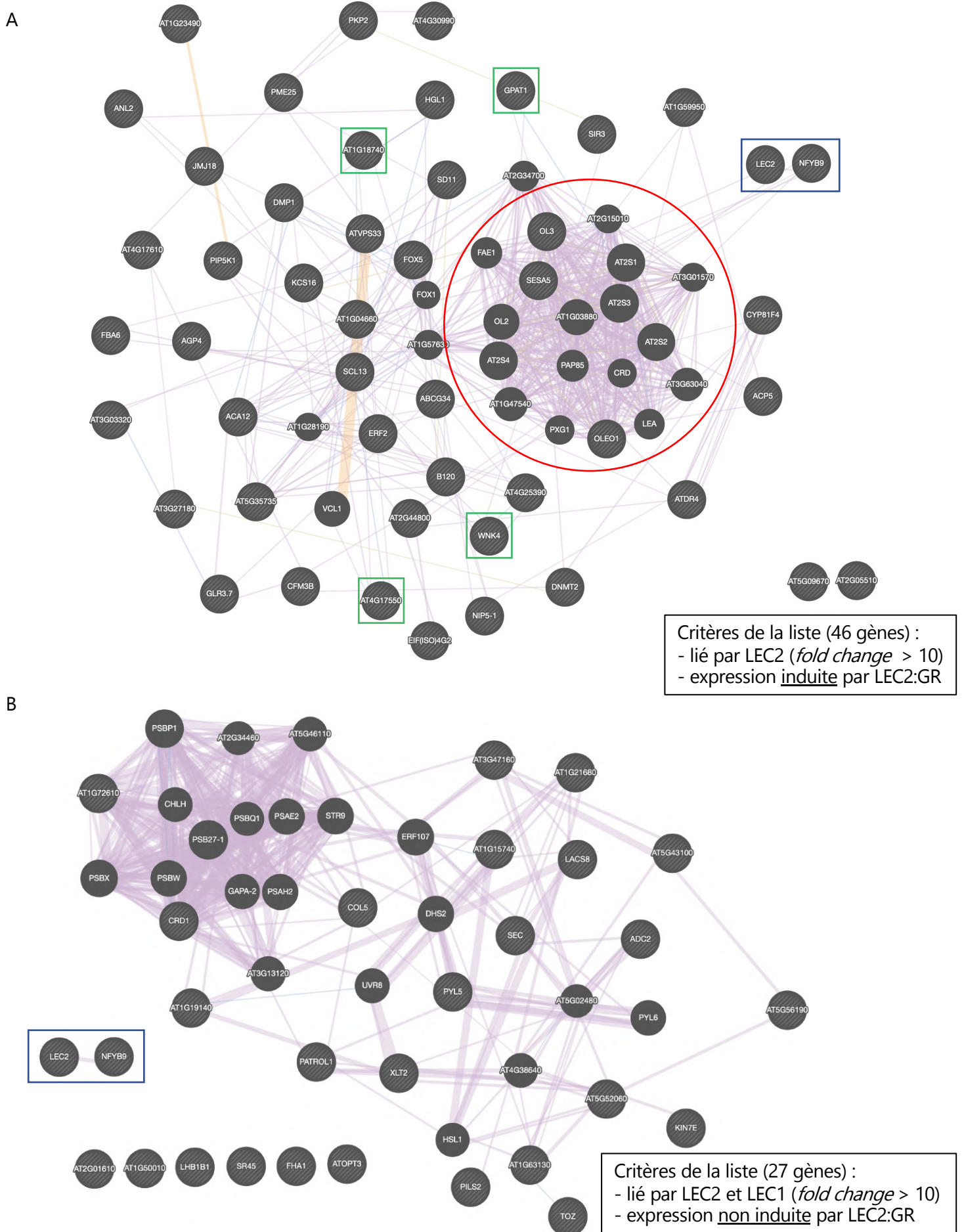


Figure 3-2 : LEC2 se lie sur certains *loci* et active leur transcription, mais également sur d'autres sans activer leur transcription. A) Les 46 gènes identifiés comme étant ciblés par LEC2 et dont l'expression est induite par LEC2:GR semblent former un réseau impliqué dans la maturation de la graine entouré en rouge. Les gènes ciblés également par LEC1 sont encadrés en vert. B) Les 27 gènes identifiés comme étant ciblés par LEC2 et LEC1 mais dont l'expression n'est pas induite par LEC2 semble former un autre réseau. Les réseaux ont été prédits par l'algorithme *GeneMANIA* (Ward-Farley et al., 2010). Les cercles hachurés correspondent aux gènes présents dans les listes. Les gènes non hachurés sont ceux ayant une interaction connue. LEC2 et LEC1 (NF-YB9) ont été ajoutés aux listes afin d'apparaître sur les réseaux et sont encadrés en bleu. Les lignes violettes représentent des co-expressions, les lignes oranges des prédictions d'interaction et les lignes bleues la co-localisation des protéines.

(Pelletier et al., 2017). L'ensemble de ces jeux de données nous permet ainsi d'observer si des gènes sont ciblés par les deux variants de LEC2, avec ou sans LEC1 et si l'expression de ces gènes est induite ou non par *LEC2:GR* et savoir si LEC2 seul ou avec LEC1 se lie à certaines cibles sans activer leur expression.

3.1. LEC2 se lie à des cibles et active leur transcription

Dans un premier temps, l'analyse croisée des données de CHIP-seq réalisé sur LEC2 avec les données transcriptomiques de *pCaMV35S::LEC2:GR* permet d'obtenir une liste de 46 gènes liés par LEC2 et dont l'expression est induite par LEC2:GR. Cette liste, considérée comme une liste de gènes témoins de l'activité transcriptionnelle directe de LEC2, a été soumise à l'algorithme *GeneMANIA* (<https://genemania.org/>) (Wardle-Farley et al., 2010), qui permet de prédire des réseaux biologiques et des interactions génétiques. *GeneMANIA* a mis en évidence un réseau structuré de gènes spécifiques de la maturation de la graine, avec plusieurs gènes impliqués dans la voie de biosynthèse des lipides et des protéines comme les *OLEOSIN* (*OLE* ou *OLEO*) et *SEED STORAGE ALBUMIN* (*SESA* ou *AT2S*), qui sont donc ciblés et dont l'expression est induite par LEC2 (Figure 3-2A). Parmi ces 46 gènes, seulement 4, de fonction inconnue, sont également liés par LEC1 (Pelletier et al., 2017).

3.2. LEC2 se lie à des cibles sans activer leur transcription

De la même manière, nous avons identifié 312 gènes semblant être liés par LEC2, mais dont l'expression n'est pas induite en réponse à *LEC2:GR*. Parmi ces gènes, 27 sont également ciblés par LEC1 (Figure 3-2B). Un réseau différent du premier semble émerger de cette liste de 27 gènes, avec des fonctions qui ne semblent pas être liées à l'accumulation des réserves. *GeneMANIA* ne semble pas prédire de co-expression des gènes, de colocalisation des protéines, ni d'interactions entre *LEC1*, *LEC2* et les gènes issus de cette liste.

En conclusion, il semble y avoir un certain nombre de gènes sur lesquels LEC2 se fixe, mais dont l'expression n'est pas induite et qui ne font pas partie d'un « réseau maturation » typique. LEC2 pourrait donc bien avoir i) un rôle de fixation et d'activation transcriptionnelle de certains gènes, notamment liés à la maturation de la graine, mais aussi ii) un rôle de fixation, sans activation transcriptionnelle sur d'autres gènes. Cela conforte notre hypothèse de double fonction de LEC2 en tant que facteur de transcription et de facteur pionnier, capable d'ouvrir ou fermer la chromatine au niveau de ses cibles. Si cette hypothèse est vraie, alors il faut s'attendre à ce qu'une différence d'accessibilité de l'ADN soit observée lorsque le facteur est muté.

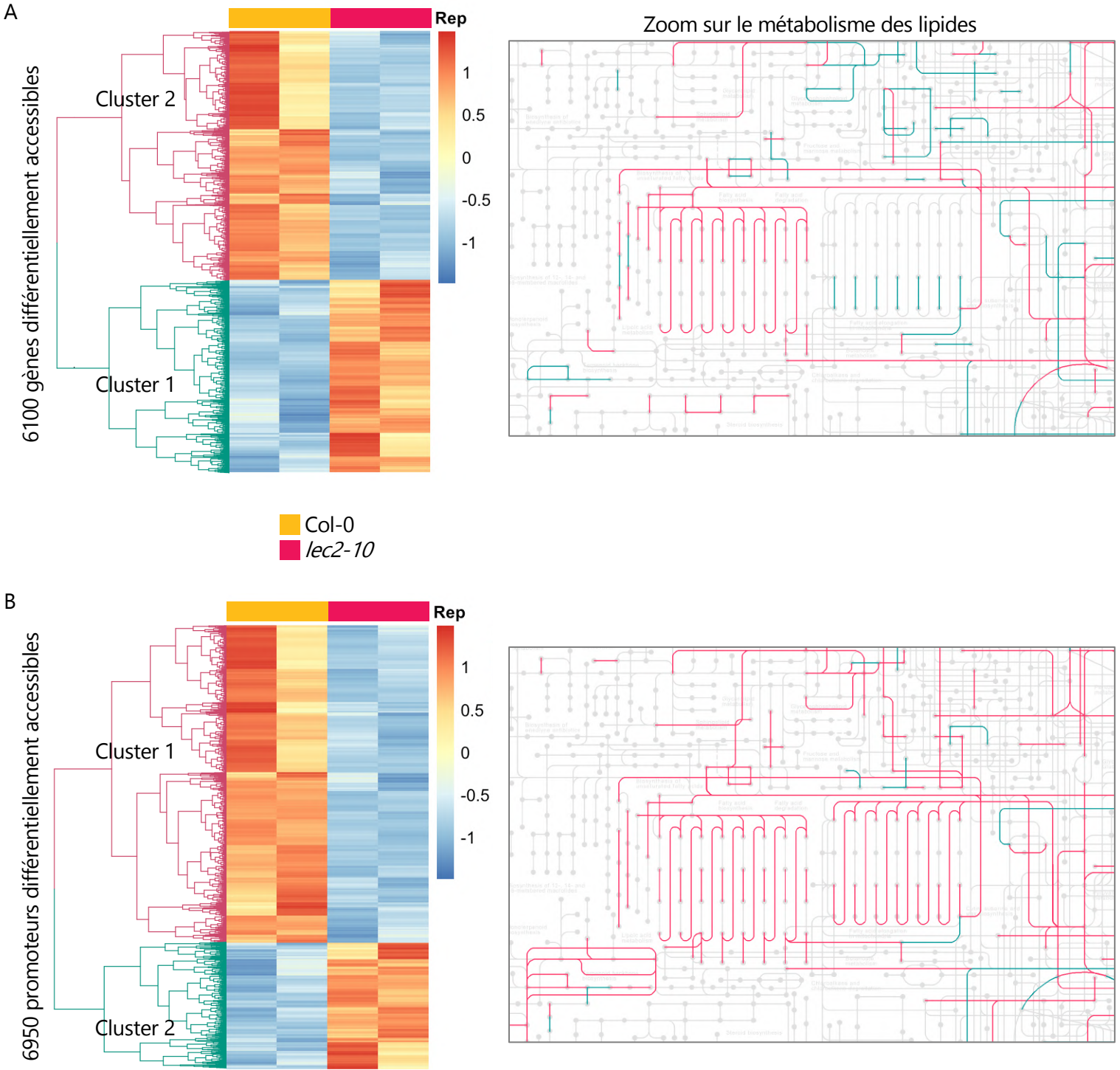


Figure 3-3 : Il existe des différences d'accessibilité de certains *loci* entre le mutant *lec2-10* et le sauvage Col-0, notamment au niveau de gènes impliqués dans la voie de biosynthèse des lipides. Cartes de chaleur (*heatmaps*) représentant A) les 6100 gènes et B) les 6950 promoteurs (1kb en amont du gène) présentant une différence d'accessibilité liée à l'ouverture ou la fermeture de la chromatine entre les deux génotypes ($q\text{-value} < 0,05$). Dans les deux cas, deux clusters sont visibles, regroupant d'une part les gènes dont la chromatine est fermée dans les graines du mutant *lec2-10* et ouverte chez le sauvage et inversement. Les cartes métaboliques zoomées sur le métabolisme des lipides représentent les gènes issus des deux clusters (respectivement représentés en rose et en bleu) dans chaque condition et ont été obtenues grâce à l'algorithme *KEGG Mapper* (Kanehisa and Sato, 2020).

4. Action transcriptionnelle de LEC2 et d'ouverture de la chromatine sur ses cibles

Si LEC2 se fixe sur certains *loci* et permet l'ouverture ou la fermeture de la chromatine sans activer la transcription des gènes, alors, ces gènes doivent montrer une différence d'ouverture ou de fermeture de la chromatine entre des graines sauvages et des graines de mutant *lec2*. Nous avons donc utilisé des données d'ATAC-seq disponibles dans la littérature et qui permettent précisément de répondre à la question de l'ouverture de la chromatine (Wang et al., 2020; Gao et al., 2022). Ces expériences d'ATAC-seq ont été réalisées sur des feuilles et des embryons matures de géotypes sauvage (Col-0) et *lec2-10* (SALK_015228, un mutant d'ADN-T différent de que celui utilisé au cours du projet de thèse).

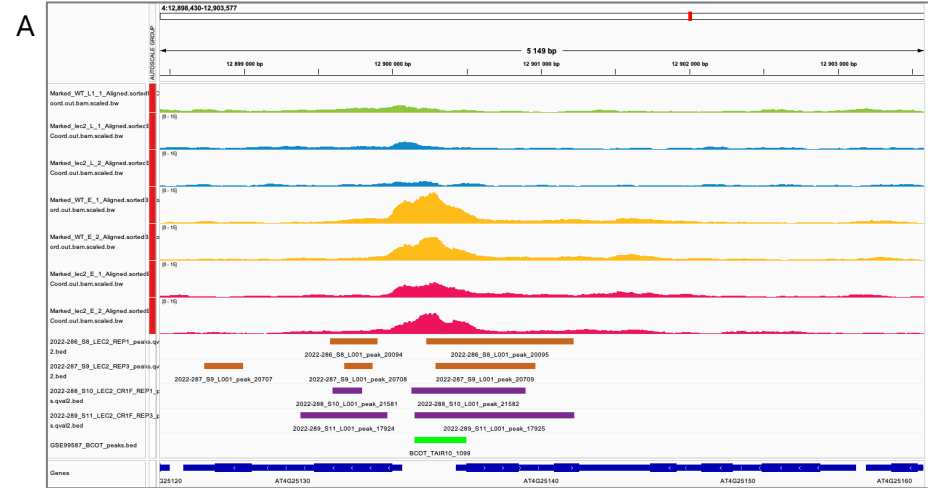
4.1. Certains *loci* sont différenciellement accessibles chez le mutant *lec2*

Les données d'ATAC-seq obtenues à partir d'embryons matures de mutant *lec2-10* et de sauvage ont été normalisées puis soumises au package *DESeq2* (Love et al., 2014), afin d'identifier les régions différenciellement accessibles au niveau des gènes d'une part et au niveau des promoteurs (*i.e* 1kb en amont des gènes) d'autre part. L'analyse d'expression différentielle *DESeq2* a permis de montrer qu'environ 6100 gènes et 6950 promoteurs montrent une différence significative d'accessibilité entre les deux géotypes ($q\text{-value} < 0,05$; Figure 3-3). Il semble donc que LEC2 ait bien un effet sur l'ouverture et/ou la fermeture de la chromatine. L'utilisation de l'algorithme *KEGG Mapper* (Kanehisa and Sato, 2020) a permis de montrer, par exemple, que certains *loci* accessibles chez le sauvage et non chez le mutant *lec2-10* (*clusters roses*) sont impliqués spécifiquement dans le métabolisme des lipides.

4.2. Analyse croisée des données de RNA-seq, ChIP-seq et ATAC-seq

4.2.1. LEC2 se lie à l'ADN et active la transcription

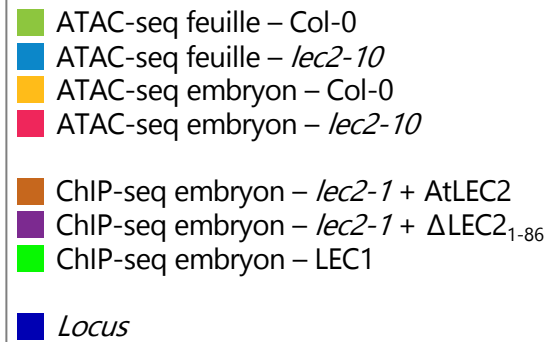
Pour certains gènes, LEC2 agit comme un facteur de transcription classique : il se lie à l'ADN et active la transcription. Dans cette situation, LEC2 n'induit pas de modifications de la chromatine : aucune différence d'accessibilité n'est visible dans les feuilles et les embryons du mutant *lec2-10* et du sauvage Col-0. C'est le cas, par exemple de l'*OLEOSIN1* (*OLE1*, AT4G25140 ; Figure 3-4A) : l'expression d'*OLE1* est induite par LEC2:GR, les données de ChIP-seq indiquent que LEC2 se fixe au promoteur, de même que LEC1, mais l'accessibilité de la chromatine ne semble pas altérée dans les embryons du mutant.



AT4G25140 (OLEOSIN 1)

Expression : induite par LEC2:GR

Chromatine : pas d'effet de *lec2-10*



AT5G51210 (OLEOSIN 3)

Expression : induite par LEC2:GR

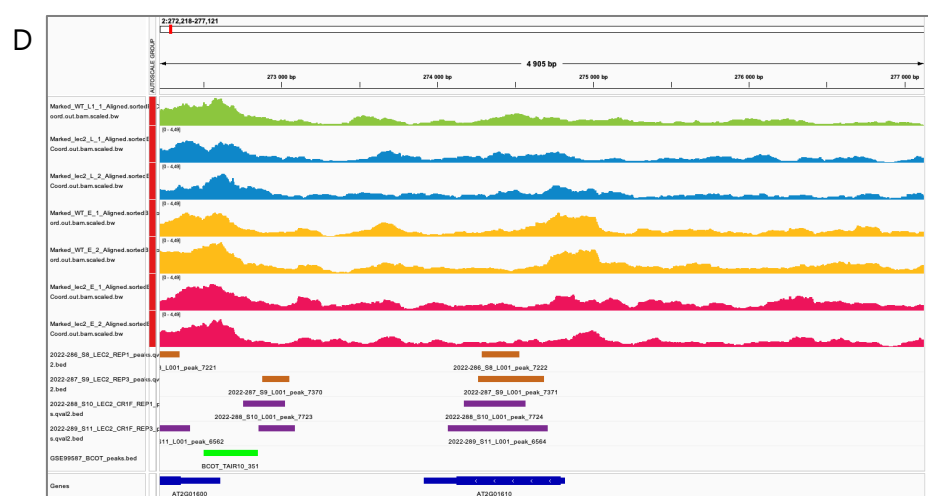
Chromatine : effet de *lec2-10*



AT1G71090 (PIN-LIKES 2)

Expression : non induite par LEC2:GR

Chromatine : effet de *lec2-10*



AT2G01610 (RING/U-box superfamily protein)

Expression : non induite par LEC2:GR

Chromatine : pas d'effet de *lec2-10*

Figure 3-4 : LEC2 a différents effets sur les *loci* auxquels il se lie. Représentation *IGV Integrative Genomics Viewer* (Robinson et al., 2011) des données d'ATAC-seq sur des feuilles et des embryons matures sauvages et mutants *lec2-10* (Wang et al., 2020; Gao et al., 2022), de ChIP-seq sur des graines au stade cœur *lec2-1* complémentés avec *AtLEC2* ou Δ *LEC2*₁₋₈₅ et de ChIP-seq sur LEC1 dans des embryons au stade cotylédons retournés (Pelletier et al., 2017). *Loci* A) dont l'expression est activée, B) dont l'expression est activée et la chromatine modifiée par LEC2, C) dont l'expression n'est pas activée mais la chromatine modifiée, D) dont ni l'expression ni la chromatine ne sont impactées par LEC2.

4.2.2. LEC2 se lie à l'ADN, ouvre la chromatine et active la transcription

Pour d'autres gènes, LEC2 agit comme un facteur de transcription mais a également un effet sur l'ouverture de la chromatine. Par exemple, dans le cas de l'*OLEOSIN3* (AT5G51210 ; Figure 3-4B), LEC2:GR induit l'expression d'*OLE3*, les données de CHIP-seq montrent que LEC2 se fixe au niveau du promoteur du gène avec LEC1 et, lorsque *lec2* est muté, on observe une différence significative ($q\text{-value} < 0,05$) de l'accessibilité de la chromatine par rapport au sauvage.

4.2.3. LEC2 se lie à l'ADN, ouvre la chromatine mais n'active pas la transcription

Comme évoqué précédemment, LEC2 avec ou sans LEC1 cible un certain nombre de *loci* mais n'active pas leur transcription (Figure 3-2C). C'est le cas de *PIN-LIKES 2* (*PILS2*, AT1G71090 ; Figure 3-4C) dont l'expression n'est pas induite par LEC2:GR, bien que le locus soit ciblé par LEC2 au niveau du gène et par LEC1 au niveau du promoteur. Dans le mutant *lec2-10*, la chromatine est fermée au niveau de ce gène.

4.2.4. LEC2 se lie à l'ADN, n'ouvre pas la chromatine et n'active pas la transcription

Enfin, en dernier cas de figure, certains *loci* sont liés par LEC2, mais étonnamment, leur transcription n'est pas induite par LEC2:GR et l'accessibilité de la chromatine n'est pas différente entre le mutant *lec2-10* ou le sauvage. C'est le cas de AT2G01610, un gène codant pour une protéine impliquée dans l'ubiquitination et exprimée tout au long du cycle de vie de la plante (Figure 3-4D). Il est à noter que sur la liste de 27 gènes liés par LEC2 et LEC1 et dont l'expression n'est pas induite par LEC2:GR, 10 ne montrent aucune différence d'accessibilité en ATAC-seq que ce soit dans le gène ou dans le promoteur.

Les quatre exemples présentés ici sont représentatifs des différents profils que l'on peut obtenir en comparant les données d'ATAC-seq et de CHIP-seq. D'autres exemples sont visibles en Annexe 6.

5. Conclusion et discussion

Toutes ces données suggèrent que LEC2 puisse agir de différentes manières sur ses cibles : il peut avoir une activité transcriptionnelle ou non et modifier l'ouverture de la chromatine ou non. Notre hypothèse de départ était que LEC2 puisse avoir une fonction avec activation transcriptionnelle de ses cibles et une fonction sans activité transcriptionnelle. L'analyse croisée des données de RNA-seq, CHIP-seq et ATAC-seq semble montrer que cette hypothèse est vraisemblable. En effet, il existe bien des gènes se comportant comme attendu sous notre hypothèse : leur expression est induite

par LEC2:GR et leur chromatine est fermée chez le mutant *lec2-10*. Cependant ces résultats ne sont que préliminaires étant donné la qualité du ChIP-seq réalisé. En effet, au-delà des exemples cités dans ce travail, on ne sait pas sur combien de gènes LEC2 pourrait avoir cette fonction de pionnier, ni quelles fonctions biologiques sont concernées. Si un réseau spécifique de la maturation dont l'expression est induite par LEC2, apparaît, les données actuelles ne permettent pas de discerner un réseau spécifique de gènes ciblés par LEC2 mais dont l'expression n'est pas induite. L'exemple de *PIN-LIKES2* présenté en partie 4.2.3, pour qui LEC2 se lie, ouvre la chromatine mais n'active pas la transcription est représentatif du type de trace que l'on peut obtenir. Bien que *PILS2* soit impliqué dans la voie de signalisation de l'auxine, aucun réseau lié à l'embryogenèse ou la signalisation hormonale n'a pu être identifié jusqu'ici. De meilleurs résultats de ChIP-seq sur LEC2 nous permettraient d'avoir une idée plus exhaustive de l'ensemble des gènes et de la ou les fonction(s) biologique(s) concernées.

Dans le cas des gènes ciblés, mais dont ni l'expression, ni l'accessibilité ne semblent impactées par LEC2, plusieurs hypothèses s'offrent à nous : i) il est possible que ces *loci* ciblés soient des faux-positifs dus à la qualité de l'expérience de ChIP-seq ; ii) il est également possible que LEC2 se fixe de manière aspécifique sur des cibles qui pourraient être constitutivement accessibles ou bien rendues accessibles par d'autres facteurs/protéines ; enfin, iii) LEC2 pourrait interagir avec d'autres partenaires pour stabiliser leur(s) interaction(s) avec ces cibles. Par exemple, LEC2 pourrait former un complexe avec d'autres partenaires, non identifiés, et assurer la fixation de ce complexe à l'ADN, mais sans avoir lui-même d'activité transcriptionnelle ou de modification de l'accessibilité des cibles. Néanmoins, dans le cas de AT2G01610, présenté précédemment (Figure 3-4D), aucune boîte RY, le site spécifique de fixation de LEC2, n'a été identifiée sur le promoteur.

Les données issues de la littérature, le transcriptome de LEC2:GR (Mendoza et al., 2005), le ChIP-seq réalisé sur LEC1 (Pelletier et al., 2017) et l'ATAC-seq réalisé sur des feuilles et embryons sauvages et mutant *lec2-10* (Wang et al., 2020; Gao et al., 2022) ont été une grande ressource pour tester notre hypothèse de double fonction de LEC2. L'accessibilité de ces données est un outil précieux pour le type d'analyses que nous avons mené. Néanmoins, il ne faut pas oublier que tous les résultats traités ici n'ont été obtenus ni dans les mêmes conditions et ni au même stade de développement de l'embryon. Il a été montré que l'ouverture de la chromatine par un facteur pionnier sur ses cibles peut être maintenue au cours du temps (Karagianni et al., 2020). Il est néanmoins possible que l'effet de LEC2 sur l'ouverture de la chromatine au moment de son expression (*i.e.* pendant le développement précoce de la graine), ne soit plus visible sur l'embryon mature, stade auquel l'ATAC-seq a été réalisé. En conséquence, lorsqu'un locus est ciblé par LEC2

mais qu'aucun effet n'est visible sur la chromatine dans le mutant *lec2-10*, cela peut être dû à la différence de stade entre les expériences. De plus, d'après les données de ChIP-seq, LEC2 et LEC1 ne semblent pas se lier toujours ensemble aux mêmes cibles. Cependant, le fait que les expériences n'aient pas été réalisées au même stade exactement et que le ChIP-seq de LEC2 ne soit pas optimal ne permettent pas de conclure sur le fait que LEC2 agisse parfois avec LEC1 et parfois seul ou avec d'autres partenaires.

Discussion et perspectives

Discussion générale

1. La position du NLS de LEC2 n'a pas pu être confirmée

Les résultats obtenus dans le Chapitre 2, partie 2. « [Mise en évidence du signal de localisation nucléaire \(NLS\) de LEC2](#) », n'ont pas permis de confirmer la présence du NLS en partie centrale de la protéine. L'identification du NLS putatif de LEC2 a été déterminée grâce à un algorithme de prédiction : *cNLS mapper* (Kosugi et al., 2009). Cependant, d'autres algorithmes, comme *NLSradamus* (Nguyen Ba et al., 2009) ne prédisent pas de NLS dans la séquence de LEC2. On peut donc mettre en doute l'utilisation de ce type d'algorithme ou penser que LEC2 n'a pas de NLS, mais entre au noyau grâce à son interaction avec d'autres protéines, via son domaine B2 par exemple. Chez ABI3, le NLS et le domaine B2 sont superposés. Des mutations ponctuelles dans cette zone chez ABI3 affectent la localisation nucléaire et donc la fonction d'activateur transcriptionnel (Marella and Quatrano, 2007). De même, des mutations dans les acides aminés conservés du domaine B2 de LEC2 empêchent l'interaction avec LEC1 (NF-YB9) (Boulard et al., 2018). On sait que les sous-unités B et C du complexe NF-Y s'assemblent dans le cytoplasme pour rentrer dans le noyau via l'interaction du dimère avec une importine 13 (Kahle et al., 2005). De plus NF-YA dispose d'un NLS non canonique (ncNLS), plus long qu'un NLS classique et entre dans le noyau grâce à une importine β (Kahle et al., 2005). On pourrait alors envisager que LEC2, à la manière de NF-YA, entre dans le noyau par l'intermédiaire d'un ncNLS non identifié.

La fusion traductionnelle entre différents variants de LEC2 et la GFP a montré que lorsque les parties N-terminale ou C-terminale de LEC2 sont éliminées, la localisation du facteur de transcription n'est pas affectée dans l'embryon, ce qui nous laisse penser que le NLS, ou tout autre partie de la protéine lui permettant d'aller au noyau, n'est pas présente dans ces extrémités non structurées. Toutefois, lorsque les parties N- et C-terminales de LEC2 sont délétées simultanément, la protéine (LEC2 M104-Q275) n'est plus localisée au noyau. On ne peut donc pas exclure que ces domaines de LEC2 soient nécessaires à sa localisation, même si notre hypothèse est que dans ce cas d'élimination des deux extrémités, la structure de la protéine est trop affectée, ce qui pourrait également impacter son transport dans le noyau.

Enfin, pour tenter de confirmer la position d'un NLS potentiel du facteur de transcription, il est nécessaire d'améliorer notre expérience d'expression transitoire dans les feuilles de tabac (*Nicotiana benthamiana*). Le produit de la fusion traductionnelle entre une GFP et un petit peptide

est trop petit et entre très probablement dans le noyau par simple diffusion. Il faudrait donc améliorer cette expérience en réalisant une triple fusion : entre la GFP, le peptide à tester (NLS putatif) et une protéine plus grosse dont on sait qu'elle n'est pas localisée au noyau. Il faut cependant s'assurer que cette dernière protéine n'ait pas d'autre signal d'adressage, par exemple un signal d'adressage au réticulum endoplasmique qui pourrait entrer en compétition avec le NLS.

2. La réalisation d'un crible sur l'apparition d'embryons somatiques est-elle pertinente ?

L'approche de mutagenèse que nous voulions développer avait pour but d'augmenter la fréquence d'apparition d'embryons somatiques, laquelle est extrêmement faible lorsque l'on surexprime *LEC2* sous contrôle de *pCaMV35S*. Or, nous avons vu que lorsque l'on utilise un autre promoteur constitutif comme le *pAtUbq10*, le phénotype est extrêmement fort et apparaît dès l'embryogenèse zygotique. Une version inductible du promoteur *pAtUbq10*, permettrait de « déclencher » la surexpression de *LEC2* une fois le développement végétatif entamé et de s'affranchir des défauts de développement de l'embryon. Une autre alternative serait d'utiliser un promoteur fort dont l'expression n'a pas lieu dans la graine, mais par exemple dans les feuilles, et qui ne soit pas soumis au *silencing* comme le *pCaMV35S*. Il semble que l'effet de *LEC2* sur l'apparition d'embryons somatiques soit une question de temporalité et de niveau d'expression. Par exemple, si l'on utilise le promoteur *pCaMV35S*, la surexpression est trop tardive et probablement soumise au *silencing*, ce qui n'induit pas ou très peu de phénotypes. Dans le cas de *pAtUbq10*, l'expression de *LEC2* est trop forte et se fait certainement trop tôt, affectant de façon importante l'embryon et, par la suite, aboutissant à des plantes couvertes d'embryons somatiques et/ou de cal dès les premiers jours de leur développement végétatif. Des résultats obtenus précédemment dans l'équipe, ont montré que l'expression de *LEC2* sous contrôle de *pOLE1* induit une accumulation d'huile pendant la croissance de la plante (Boulard et Dubreucq, non publié). *LEC2* sous contrôle de *pOLE1* induit sa propre expression, des cellules restent dans un état « embryonnaire » et la plante accumule des réserves, mais n'induit cependant pas d'ES pour autant. La temporalité d'expression implique de questionner « l'allumage » et « l'extinction » de l'expression de *LEC2* car, dans le cas de l'utilisation de *pAtUbq10* comme de *pOLE1*, le phénotype est continu et la plante n'arrête jamais de produire des embryons somatiques, du cal ou de l'huile. Quelques rares plantes surexprimant *LEC2* sous contrôle de *pAtUbq10* ont montré la formation d'organes différenciés (racines ou feuilles). Dans ce cas, il est probable que l'expression du transgène dans une cellule soit devenue moins forte, en raison d'une régulation inconnue et que cette cellule, ainsi que ses descendantes puissent alors suivre un

programme de développement végétatif et non plus d'embryogenèse en continu, menant alors à la formation d'organes.

Maintenant que nous savons que le caractère d'induction spontanée de l'embryogenèse somatique semble finalement plus lié au profil d'expression plutôt qu'à la séquence de LEC2, est-il nécessaire de générer de nouveaux variants du facteur de transcription, qui, en contexte surexpresser, seraient capables de former des embryons somatiques à une fréquence plus élevée ? Est-ce qu'un promoteur spécifique, de feuille par exemple, pourrait permettre la formation d'embryons somatiques sur les plantes, de manière efficace ? Dans ce cas, l'approche de départ qui visait à induire des mutations ponctuelles ciblées au locus peut probablement être envisagée et un crible pourrait être réalisé sur l'apparition spontanée d'embryons somatiques sur les plantules en culture *in vitro*. Néanmoins, il est peut-être nécessaire de modifier la question de départ et ne pas réaliser un crible en contexte surexpresser, mais sauvage. Dans ce cas, est-ce que des variants de LEC2, en contexte d'induction hormonale de l'ES, permettraient une meilleure réponse au traitement ? Le type de crible, qui serait ici réalisé sur un grand nombre de variants en culture *in vitro*, est assez lourd et compliqué à mettre en place et ne nous paraît pas envisageable avec les moyens dont nous disposons. En effet, ce type d'approche a déjà été réalisé au sein de l'équipe au cours d'un crible chimique dont le but était de tester l'effet de différentes molécules sur l'efficacité d'induction de l'ES et nous savons qu'il est difficile à réaliser à grande échelle (Thévenin et al., non publié).

On sait que lorsque *lec2* est muté, les plantes perdent leur capacité d'induction de l'ES (Gaj et al., 2005). A l'inverse, un variant qui complèterait le phénotype [*lec2*] dans la graine pourrait-il être récalcitrant pour l'embryogenèse somatique, en découplant les fonctions liées à l'embryogenèse somatique et zygotique ? Une telle approche serait extrêmement difficile à mettre en place dans le cadre d'une mutagenèse de type CRISPR, car les mutations sont induites à l'état hétérozygote. Or, si l'on cherche un phénotype de perte de fonction de l'embryogenèse somatique, il est nécessaire d'avoir la mutation à l'état homozygote. Si l'on considère le grand nombre de variants à tester et à obtenir à l'état homozygote, cela représente une charge de travail et un nombre de plantes T2 considérable à manipuler, sans compter la difficulté de réaliser un crible en contexte d'induction de l'ES, comme évoqué ci-dessus.

La réponse à la question : « la réalisation d'un crible sur l'apparition d'embryons somatiques a-t-elle vraiment un sens ? » doit donc être modulée.

- En contexte surexpresser fort, le crible n'a plus lieu d'être, car la surexpression de LEC2 induit vraiment la formation d'embryons somatiques de manière spontanée et on ne peut donc pas cribler sur l'amélioration de ce phénotype.
- L'identification de variants répondant mieux aux protocoles d'induction de l'ES serait une alternative, mais néanmoins très difficile à mettre en place.
- La réalisation du crible prévu initialement (apparition spontanée d'embryons somatiques) semble être l'alternative la plus facile à mettre en place, mais seulement en utilisant un promoteur plus spécifique, de feuille, par exemple, permettant l'expression ectopique de *LEC2* ou une construction inductible avec un promoteur fort comme le *pAtUBQ10*.

Si l'activité de LEC2 sous contrôle d'un promoteur fort est liée à sa temporalité et à son niveau d'expression, nous pouvons émettre une autre hypothèse : l'apparition d'embryons somatiques en réponse à l'expression ectopique de LEC2 dépend peut-être de l'accessibilité chromatinienne des gènes liés aux processus d'embryogenèse. En d'autres termes, si *pAtUBQ10* permet une expression très précoce de LEC2, *i.e.* au cours de l'embryogenèse, alors les cibles d'embryogenèse sont certainement accessibles à ces stades de développement. Mais si *pCaMV35S* permet une expression plus tardive au cours du développement des plantules, alors *LEC2* est bien exprimé, mais la chromatine peut être fermée au niveau des cibles. De la même façon, l'expression de *LEC2* sous contrôle de *pOLE1*, ne débute qu'au moment de la maturation, donc après l'embryogenèse, et dans un contexte chromatinien différent. Dans ce cas, *LEC2* est exprimé en continu sous contrôle de *pOLE1*, sur lequel il forme une boucle d'activation positive, mais les cibles liées à l'embryogenèse ne sont pas disponibles et les plantes ne font pas d'embryogenèse somatique. Ainsi, si l'on envisage un crible d'apparition d'embryons somatiques en contexte surexpresser inductible, ou spécifiquement dans les feuilles, il est probable que l'on n'obtienne aucun résultat, car les cibles d'embryogenèse ne sont pas accessibles dans les feuilles. Les cibles ne pourraient être rendues accessibles qu'en réponse à une induction hormonale.

3. Embryogenèse somatique ou zygotique ? LEC2, responsable d'un « état embryonnaire »

Peut-on découpler la fonction embryogenèse somatique de la fonction embryogenèse zygotique ? Comme évoqué plus tôt, les processus d'ES sont altérés dans un mutant *lec2* (Gaj et al., 2005), mais pourrait-on avoir un variant pour lequel l'embryogenèse zygotique serait normale, mais qui serait affecté dans sa capacité à induire de l'ES ? Chez le cotonnier, l'analyse de génotypes récalcitrants aux processus d'embryogenèse somatique a montré une hyperméthylation du génome (Guo et al.,

2020). À l'inverse, chez les géotypes les plus permissifs, l'expression de *GhLEC1*, *GhLEC2* et *GhFUS3* est plus forte (Zheng et al., 2014). Est-ce que le variant de LEC2 sans domaine d'activation (*i.e.* Δ LEC2₁₋₈₅) est toujours capable de former des embryons somatiques en contexte surexprimeur ou en contexte d'induction de l'ES ? On sait également que l'expression ectopique de LEC2 dans les feuilles induit l'accumulation de composés de réserves spécifiques des graines (Mendoza et al., 2005). LEC2 est donc capable de modifier un métabolisme de feuille en métabolisme embryonnaire. Lorsque *LEC2* est sous contrôle d'un promoteur ubiquitaire et que son expression est plus forte et continue pendant l'embryogenèse, l'architecture de l'embryon est totalement modifiée et les plantules sont couvertes d'embryons somatiques et/ou de cal.

Toutes ces observations suggèrent que ce facteur de transcription est responsable d'un « état embryonnaire » plutôt qu'un classique régulateur de la biosynthèse de réserves au moment de la maturation. L'expression de *LEC2* est activée au cours du développement de la graine et réprimée au moment de la germination et tout au long du développement végétatif (Lepiniec et al., 2018). Au cours des processus d'embryogenèse somatique, l'expression de *LEC2* et des autres *LAF1* est activée en réponse aux signaux hormonaux et à l'action de facteurs de transcription comme BBM (Horstman et al., 2017), conduisant à la formation d'embryons. De plus, lorsque l'expression de LEC2 n'est pas réprimée à la fin du développement de la graine, la plante ne suit pas un programme de développement végétatif, mais, en quelque sorte, conserve un caractère embryonnaire. Si l'on définit une graine comme un organe dans lequel un embryon se forme et des réserves sont accumulées, alors on peut dire que LEC2 est responsable d'un « état embryonnaire ». L'embryogenèse serait considérée comme un état précoce et la maturation un état tardif, ce dernier ne pouvant être induit que par ou après l'état précoce. Dans la revue « [Contrôle génétique et moléculaire de l'embryogenèse somatique](#) », nous faisons l'hypothèse d'une double fonction de LEC2 : une au cours de l'embryogenèse zygotique et l'autre au cours de l'embryogenèse somatique. Finalement, nous pouvons penser que LEC2 n'a pas de double fonction concernant l'EZ ou l'ES, mais bien une seule : celle d'induire un « état embryogène ».

4. Les domaines d'activation du facteur de transcription LEC2

4.1. Pourquoi avoir un FT avec des DA dispensables ?

Dans le Chapitre 2, partie 3 « [Mise en évidence des domaines d'activation de la transcription de LEC2](#) », nous avons identifié les domaines d'activation de la transcription (DA) de LEC2 et testé leur fonctionnalité. Au cours de la partie 4, « [Fonctionnalité de LEC2 sans domaines d'activation in](#)

[planta](#) », nous avons montré que LEC2 sans DA complémente le phénotype du mutant *lec2-1*. Ainsi, LEC2 sans domaine d'activation perd son activité transcriptionnelle dans les protoplastes de mousse mais semble fonctionnel chez *Arabidopsis*. Ces résultats posent la question de la conservation des domaines d'activation sur le facteur de transcription : pourquoi sont-ils toujours présents s'ils ne sont pas nécessaires à la fonction de LEC2 ?

Le Chapitre 2 nous a conduits à l'hypothèse que LEC2 pourrait jouer un rôle de facteur pionnier en complexe avec LEC1 (NF-YB9) et NF-YC. Dans ce cas, LEC2 prendrait la place de NF-YA dont la fonction est de lier le complexe à l'ADN. Ainsi, LEC2 apporterait la spécificité de cibles grâce à son domaine B3 tout en étant lié à LEC1 grâce à son domaine B2. Le complexe NF-Y/LEC2, en tant que pionnier ne jouerait pas un rôle d'activateur transcriptionnel, mais d'ouverture ou de fermeture de la chromatine au niveau de cibles particulières en ciblant des boîtes RY ou éventuellement d'autres motifs cis-régulateurs jusqu'alors inaccessibles à la machinerie transcriptionnelle. Pour cela, la fonctionnalité de LEC2 ne dépend pas de la présence d'un domaine d'activation. Néanmoins, si LEC2 n'a pas perdu ses domaines d'activation de la transcription au cours de l'évolution, cela suggère qu'il joue à la fois un rôle pionnier et un rôle d'activateur transcriptionnel à domaine B3. En d'autres termes, si les DA sont toujours présents, on peut faire l'hypothèse que la fonction de pionnier ne doit pas être l'unique fonction de LEC2. Mais dans ce cas, pourquoi LEC2 sans domaine d'activation en N-ter ou en C-ter complémente-t-il parfaitement le phénotype [*lec2*] ? Il faut noter qu'un domaine d'activation fonctionnel subsiste dans le domaine B3. En effet, le variant de LEC2 M104-Q275, dont les domaines N- et C-ter ont été délétés, ne complémente pas le phénotype [*lec2*], mais cela peut être lié à un défaut de localisation et/ou de conformation de la protéine, qui n'est pas localisée au noyau pour ce variant. Il n'est pas exclu également que LEC2 M104-Q275 ait une activité transcriptionnelle *in planta*, s'il avait accès au noyau. Il serait donc intéressant de générer un variant de LEC2, dont les extrémités N- et C-ter sont toujours présentes, mais dont les domaines d'activation N- et C-ter seraient mutés pour les rendre non fonctionnels. La protéine aurait ainsi une structure semblable à la version endogène de LEC2, mais avec seulement un domaine d'activation situé dans le domaine B3, qui serait fonctionnel. Cette expérience nous permettrait de savoir si ce domaine seul autorise la fonction d'activateur transcriptionnel de LEC2 *in planta*. Un autre moyen de vérifier si le DA situé dans le domaine B3 confère à lui seul une partie de l'activité transcriptionnelle de LEC2, serait de muter uniquement ce domaine tout en gardant les extrémités N- et C-ter. La mutation Y₂₅₇A Y₂₆₅A localisée au niveau de ce domaine pourrait répondre à cette question. Cette mutation rend LEC2 non fonctionnel tant dans la mousse que chez *Arabidopsis*, mais ne semble pas affecter la fonction d'activation transcriptionnelle elle-même. En

effet, dans la levure, la protéine entière LEC2 Y₂₅₇A Y₂₆₅A, comme le fragment LEC2 G240-Q275 Y₂₅₇A Y₂₆₅A seul, ont tous deux une activité transcriptionnelle (Figure 2-15F,G). Nous pensons donc que cette mutation affecte plutôt la liaison à l'ADN que l'activité transcriptionnelle. Cette hypothèse reste à prouver au cours d'une expérience de simple hybride, par exemple, réalisée avec le promoteur *pOLE1* dans la levure. Il est cependant possible qu'on ne puisse pas dissocier l'activité du DA portée dans le domaine B3 et l'activité de liaison de ce dernier.

4.2. Conservation de LEC2 à travers les espèces

La présence de DA non nécessaires à l'activité transcriptionnelle du facteur de transcription pourrait-elle être liée à une évolution récente chez les dicotylédones ? Pour rappel, aucun orthologue de LEC2 n'a jamais été identifié chez les monocotylédones (Boulard et al., 2017)) et il semble que LEC2 dans sa structure « DA-B2-B3-DA » soit spécifique aux plantes dicotylédones. Un projet mené récemment dans l'équipe (Thévenin et Dubreucq, non publié) a eu pour but d'identifier des orthologues de facteurs B3 chez les céréales, afin de comprendre pourquoi LEC2 semble être absent chez les monocotylédones. Plusieurs protéines ont été identifiées sur la base de la présence d'un domaine B2 et d'un domaine B3 similaires à ceux de LEC2 et ABI3, et leur activité transcriptionnelle a été testée avec *pOLE1* dans des protoplastes de mousse. Les résultats ont montré qu'aucun de ces orthologues putatifs n'active ce promoteur, qu'ils soient seuls ou en complexe avec les autres membres des LAFL dans la mousse. Ceci suggère que ces protéines ne sont pas fonctionnelles. Cependant, au vu des résultats que j'ai obtenus au cours de cette thèse, il semble que le test d'activité transcriptionnelle avec *pOLE1* dans les protoplastes de mousse ne soit pas le meilleur outil pour identifier un orthologue de LEC2, pour sa fonction pionnière en particulier. Il serait judicieux de tester si ces orthologues putatifs, placés sous contrôle de *pLEC2*, sont capables de compléter le phénotype du mutant *lec2 in planta*, par exemple. Il serait également intéressant d'observer si les orthologues putatifs de LEC2, ou une autre protéine à domaines B2 et B3, sont surexprimés dans les transcriptomes d'espèces monocotylédones au cours de l'embryogenèse somatique. Ces expériences permettraient à la fois de statuer sur l'existence d'un orthologue de LEC2 chez les monocotylédones, d'apporter ou non une preuve que les domaines d'activation présents en N- et C-ter sont une évolution récente apparue chez les dicotylédones et que la fonction ancestrale de LEC2 est une activité de facteur pionnier et non d'activateur transcriptionnel. On peut alors imaginer que la présence de domaines d'activation de la transcription chez les orthologues de LEC2 soit une spécificité des dicotylédones. En effet, les graines de céréales accumulent de l'amidon dans l'albumen, contrairement aux dicotylédones, qui stockent de l'huile et/ou des protéines dans les cotylédons. Ainsi, les domaines d'activation jouent

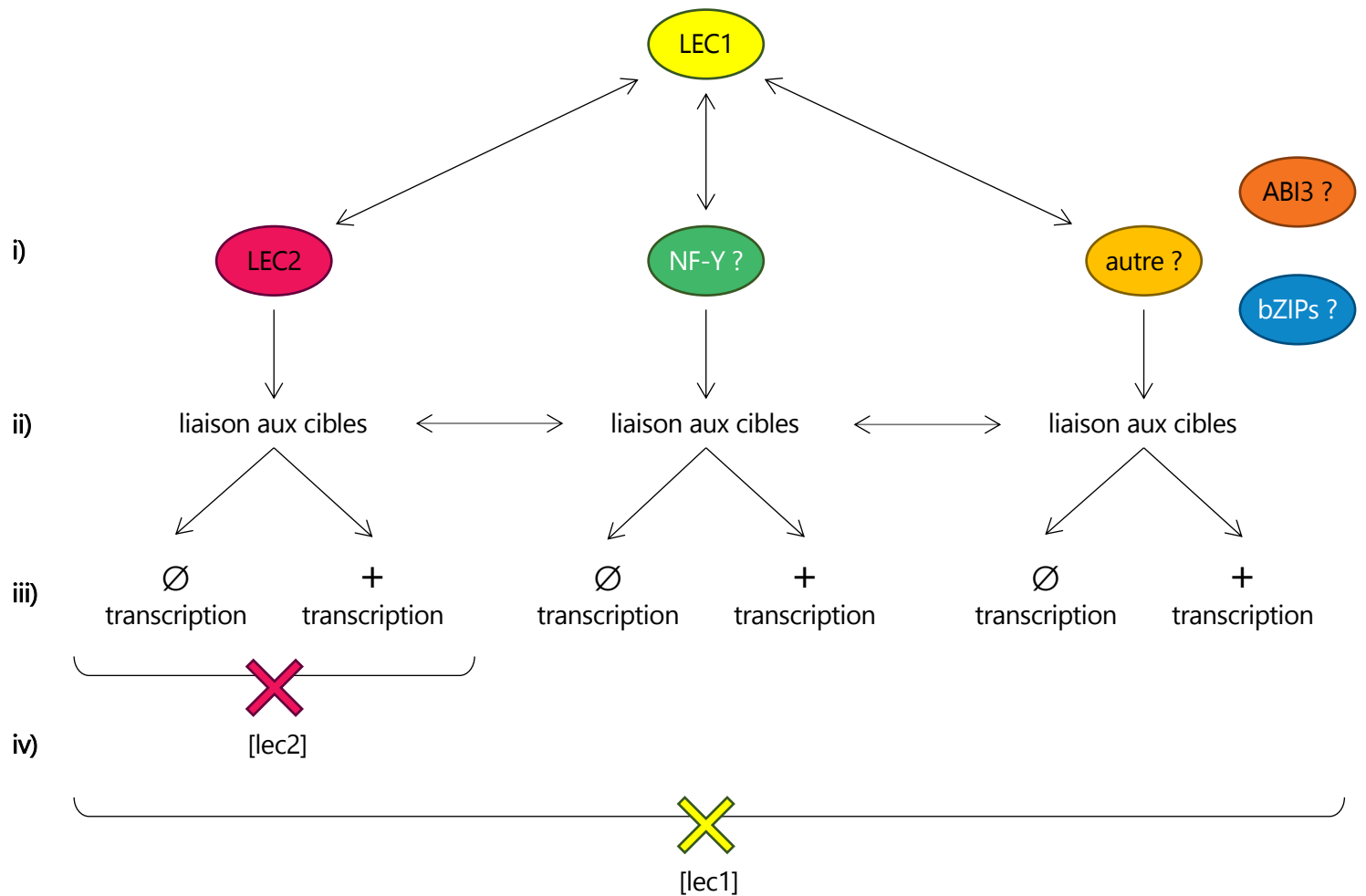


Figure Discussion : LEC1 interagirait avec plusieurs partenaires, dont LEC2, pour activer la transcription ou modifier la structure de la chromatine de certaines cibles. i) Les différents partenaires de LEC1 lui donneraient sa spécificité en se liant à l'ADN. ii) Certains *loci* pourraient être ciblés par plusieurs partenaires de LEC1. iii) Comme LEC2, les partenaires de LEC1 pourraient avoir une fonction d'activateur transcriptionnel ou non sur les cibles. iv) Lorsque *lec2* est muté, LEC1 perdrait son activité sur les cibles de LEC2, mais garderait son activité spécifique d'autres partenaires. Le mutant *lec1* perd toute son activité, c'est pourquoi le phénotype [lec1] serait plus fort que le phénotype [lec2].

peut-être un rôle spécifique à un « effet AFL-B3 », lié à l'accumulation d'huile et de certaines protéines. Les céréales n'auraient ainsi pas besoin d'un LEC2 pourvu de domaines d'activation ?

Enfin, la fonction de LEC2 est étroitement liée à celle de LEC1. Il semble que LEC2 soit nécessaire à la fonction de LEC1, que ce soit dans le cadre de son activité transcriptionnelle pour l'accumulation d'huile ou dans le cadre de son activité de facteur pionnier. LEC1 a un effet coopératif sur l'activité transcriptionnelle de LEC2 (Baud et al., 2016), mais finalement, peut-être est-ce LEC2 qui permet l'activité de LEC1 ? LEC2 pourrait alors assurer la liaison spécifique de LEC1 à ses cibles grâce à son domaine B3. On sait que les mutants *lec1* et *lec2* ont des phénotypes très similaires, mais que les mutants *lec1* sont plus affectés que les mutants *lec2*. Il est possible que lorsque, LEC2 est muté, LEC1 perde une grande part de son activité transcriptionnelle et de facteur pionnier, mais garde une activité partielle grâce à son interaction avec d'autres facteurs de transcription comme ABI3, le complexe NF-Y, ou encore bZIP67 (Yamamoto et al., 2009; Baud et al., 2016) ce qui expliquerait pourquoi le mutant *lec2* est moins affecté que le mutant *lec1* (Figure Discussion). Ainsi, en lien avec ce qui a été discuté précédemment, LEC2 aurait un rôle plus ancien chez les plantes à graines, dans la reconnaissance des cibles de LEC1 pour son activité de facteur pionnier, ne nécessitant aucune activité transcriptionnelle mais simplement un domaine B2 d'interaction protéine-protéine nécessaire à son interaction avec LEC1, et un domaine B3 de liaison à l'ADN pour la reconnaissance des sites spécifiques. Au cours de l'évolution, des domaines d'activation de la transcription auraient pu apparaître chez les plantes dicotylédones, afin d'améliorer la fonction de LEC2 et de LEC1 comme activateurs transcriptionnels des gènes de biosynthèse de réserves et de gènes spécifiques des graines en développement. C'est ainsi que LEC2, en lien avec LEC1, jouerait à la fois un rôle d'activateur transcriptionnel sur certaines cibles et un rôle de pionnier uniquement, induisant l'ouverture de la chromatine sur d'autres cibles.

Perspectives

Ce projet de thèse a permis de mettre en évidence une probable fonction de LEC2 comme facteur pionnier. En effet, il semble que LEC2 soit capable d'assurer au moins une partie de ses fonctions sans domaine d'activation de la transcription. Néanmoins, afin de confirmer pleinement cette hypothèse, une expérience de complémentation du mutant *lec2* est nécessaire, en s'assurant que tous les domaines d'activation sont bien inactifs. En effet, il est envisageable que le domaine d'activation localisé en fin de domaine B3 soit suffisant pour assurer seul une partie de l'activité transcriptionnelle de LEC2, ce qui peut sembler contradictoire avec la conservation des domaines N et C-terminaux.

Par ailleurs, l'analyse de données de ChIP-seq, ATAC-seq et RNA-seq semble montrer que LEC2 est capable de se lier à certaines cibles, d'induire l'ouverture de la chromatine et ce sans forcément activer la transcription des gènes concernés suggérant ainsi que LEC2 puisse agir comme facteur pionnier. Des résultats de ChIP-seq de meilleure qualité sur LEC2 devraient pouvoir apporter des preuves supplémentaires de cette activité que nous pensons avoir identifiée.

Le variant Δ LEC2₁₋₈₅, semble compléter toutes les fonctions liées à l'embryogenèse zygotique chez le mutant *lec2-1*. La surexpression de ce variant sous contrôle de *pAtUBQ10* permettrait de savoir si les fonctions liées à l'embryogenèse somatique sont également assurées par ce variant et ainsi confirmer que LEC2 sans domaine d'activation est fonctionnel *in planta* pour ce rôle.

Deux variants de LEC2 sont non fonctionnels *in planta*. Une expérience de simple hybride dans la levure permettrait de vérifier si LEC2 Y₂₅₇A Y₂₆₅A, qui n'est pas fonctionnel dans la mousse et qui ne complète pas le mutant *lec2*, mais dont l'activité transcriptionnelle n'est pas affectée dans la levure, est capable de se lier à l'ADN. On pourrait alors statuer sur l'importance de ces résidus dans la fonction de LEC2. De même, la capacité de liaison à l'ADN du variant LEC2 M104-Q275, dont les domaines N- et C-ter ont été éliminés, qui n'a ni activité transcriptionnelle dans la levure, ni dans la plante et qui n'est pas localisé au noyau, devrait être testée en simple hybride pour vérifier si le domaine B3 permettant une liaison à l'ADN est toujours fonctionnel ou pas.

Enfin, si LEC2 joue un rôle très important au cours du développement de la graine, notamment avec LEC1 et si l'absence de domaines d'activation de la transcription ne semble pas impacter sa fonctionnalité pour le contrôle de la maturation, il se peut qu'une partie de l'activité de LEC2 soit réalisée par d'autres partenaires, comme ABI3. Ainsi, l'expression d'un variant de LEC2 sans domaine

d'activation de la transcription en fond double mutant *lec2 abi3* devrait permettre de préciser le rôle d'ABI3 dans ce système complexe d'expression, en particulier concernant le contrôle de l'accumulation des réserves dans la graine par les AFL-B3.

Matériel et méthodes

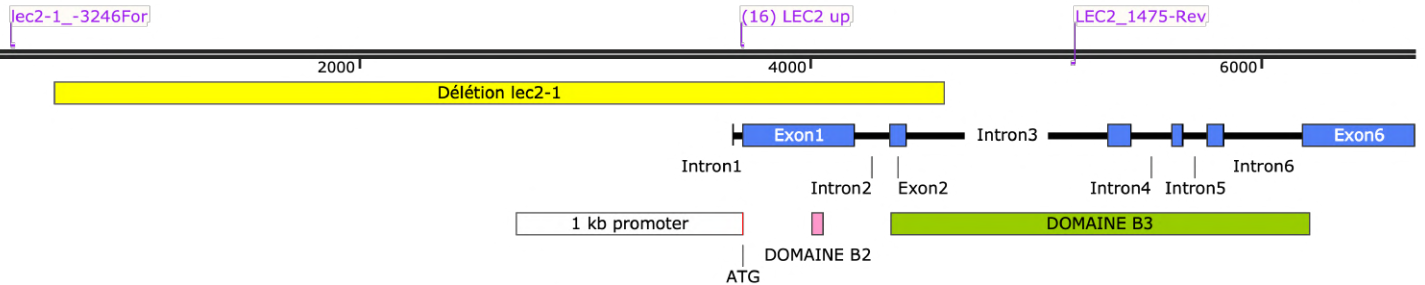


Figure M_M-1 : Structure de l'allèle mutant *lec2-1*. Une délétion de 3,9 kb (représentée en jaune) est située entre -3 048 et 895 pb en amont et en aval de l'ATG (en rouge). Le promoteur de 1 kb est représenté en blanc, les domaines B2 et B3 en rose et en vert respectivement. Les introns et exons sont représentés en traits noirs et barres bleues. Les amorces utilisées pour le génotypage sont indiquées en violet (séquences en Annexe 2).

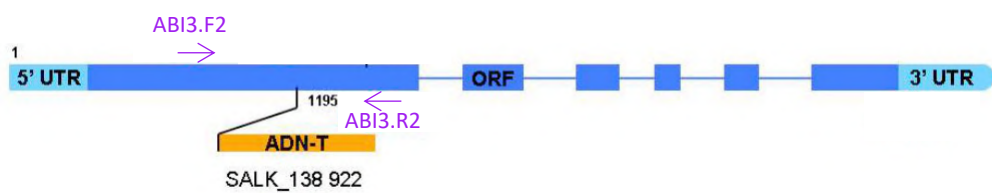


Figure M_M-2 : Structure de l'allèle mutant *abi3-1* (adapté de Barthole, 2013). Les introns et exons sont représentés en bleus, respectivement par les traits et les barres. Les amorces utilisées pour le génotypage sont indiquées en violet (séquences en Annexe 2).

Matériel et méthodes

1. Matériel

1.1. Matériel végétal

Lors de ce projet, des plantes *Arabidopsis thaliana* sauvages, d'écotype Col-0 (Columbia) et Ws (Wassilewskija) ont été cultivées. L'allèle mutant *lec2-1* (Figure M_M-1) est issu d'une collection de mutants ADN-t d'*Arabidopsis thaliana*, écotype Ws, et comporte une délétion de 3,9 kb, commençant 3 048 pb en amont de l'ATG et s'arrêtant 895 pb après l'ATG (Meinke et al., 1994; Stone et al., 2001). L'allèle mutant *abi3-11* provient de la lignée SALK_138922 en fond Col-0 et possède une insertion d'ADN-T, située dans le premier exon (Figure M_M-2). Pour l'expression transitoire dans des protoplastes de mousse, la souche sauvage *Physcomitrium patens* Gransden type sauvage a été utilisée. Pour les expériences d'expression transitoire dans les feuilles de tabac, une lignée de *Nicotiana benthamiana*, dont les noyaux sont marqués en rouge grâce à une transformation stable de *2xpCaMV35S::RFP:H2B::t35S*, a été utilisée (Martin et al., 2009).

1.2. Souches bactériennes

L'amplification et l'extraction d'ADN plasmidique a été réalisée grâce à la culture et la transformation de bactéries *Escherichia coli* électro-compétentes de souche DH10b (F- *endA1 deoR+ recA1 galE15 galK16 nupG rpsL Δ(lac)X74 φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) StrR λ*). La souche *E. coli* DB3.1 (F-, *gyrA462, endA1, Δ(sr1-recA), mcrB, mrr, hsdS20(rB-, mB-), supE44, ara14, galK2, lacY1, proA2, rpsL20(Smr), xyl5, Δleu, mtl1* – Invitrogen), également électro-compétente, a été utilisée pour l'amplification de vecteurs Gateway non recombinés comportant un marqueur de sélection *ccdB* (Bernard and Couturier, 1992), car elle comporte une mutation affectant un gène codant pour une gyrase (*gyrA462*) (Miki et al., 1992). Pour la transformation stable et transitoire des plantes (*Arabidopsis* et tabac), la souche C58C1 pMP90 d'*Agrobacterium tumefaciens* a été utilisée (Koncz and Schell, 1986).

1.3. Souche de levure

Les tests d'activité transcriptionnelle ont été réalisés sur la souche de levure *Saccharomyces cerevisiae* AH109 (*MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1_{UAS}-GAL1_{TATA}-HIS3, GAL2_{UAS}-GAL2_{TATA}-ADE2, URA3::MEL1_{UAS}-MEL1_{TATA}-lacZ*) (James et al., 1996).

1.4. Vecteurs plasmidiques

1.4.1. Vecteurs Gateway™

Vecteur	Type de vecteur	Expression	Sites Gateway™	Résistance	Sélection des vecteurs vides
pDONR207	Entrée	Bactéries	attP1, attP2	Gentamycine	ccdB
pDONR221	Entrée	Bactéries	attP1, attP5	Kanamycine	ccdB
pDONR221	Entrée	Bactéries	attP5, attP2	Kanamycine	ccdB
pDONR221	Entrée	Bactéries	attP1, attP2	Kanamycine	ccdB
pBS Tpp-A	Destination	Bactéries, mousse	attR1, attR2	Carbénicilline	ccdB
pBS Tpp-B	Destination	Bactéries, mousse	attR1, attR2	Carbénicilline	ccdB
pGWB2	Destination	Bactéries, plantes	attR1, attR2	Kanamycine, Hygromycine	ccdB
pΔGR	Destination	Bactéries, plantes	attR1, attR2	Kanamycine	ccdB
pDEST32	Destination	Bactéries, levures	attR1, attR2	Kanamycine	ccdB

1.4.2. Vecteurs GoldenBraid 2.0

Vecteur	Type de vecteur	Expression	Site de restriction	Résistance	Sélection des vecteurs vides
pUPD	Entrée	Bactéries	<i>BsmBI</i>	Ampicilline	LacZ
pUPD2	Entrée	Bactéries	<i>BsmBI</i>	Chloramphénicol	LacZ
pDGB3_α1	Destination	Bactéries, plantes	<i>BsaI</i>	Kanamycine	LacZ
pDGB3_α2	Destination	Bactéries, plantes	<i>BsaI</i>	Kanamycine	LacZ
pDGB3_Ω1	Destination	Bactéries, plantes	<i>BsmBI</i>	Spectinomycine	LacZ
pDGB3_Ω2	Destination	Bactéries, plantes	<i>BsmBI</i>	Spectinomycine	LacZ

2. Méthodes

2.1. Culture et transformation des plantes

2.1.1. Stérilisation des graines sèches

Les graines sont stérilisées en surface de façon courante par du chlore en solution hydro-alcoolique. La solution de stérilisation est préparée en dissolvant une pastille de chlore (Orlav, Hydrachim) dans 45 mL d'eau distillée, puis diluée au 1/10^{ème} dans de l'éthanol 96% (v/v). Environ 20 mg de graines sont mis à incuber pendant 8 min sous agitation dans 1 mL de la solution de stérilisation. Les graines sont ensuite rincées dans 1 mL d'éthanol absolu pendant 8 min sous agitation et séchées pendant une nuit sous flux d'air stérile.

2.1.2. Stérilisation des siliques immatures

De la même manière que pour la stérilisation des graines, une solution de stérilisation est préparée, mais diluée dans au 1/10^{ème} dans de l'eau (v/v) et non pas de l'éthanol 96%. Les siliques sont mises à incuber pendant 3 min dans la solution de stérilisation, puis rincées 3 fois à l'eau stérile. Les siliques peuvent être utilisées directement sans séchage.

2.1.3. Culture *in vitro*

Toutes les étapes de culture *in vitro* de plantes (*Arabidopsis* ou mousses) sont réalisées en conditions stériles sous une hotte à flux laminaire horizontal. Le matériel utilisé est stérilisé au stérilisateur à billes Steri 250 (Keller, Suisse) ou au bec bunsen (pinces et spatules), à l'éthanol absolu pour le matériel le plus fragile ou au four Pasteur (filtres et broyeurs).

2.1.3.1. *Arabidopsis thaliana*

Les graines stérilisées sont semées sur du milieu *Arabidopsis* Gamborg B5 (Duchefa, <http://www.duchefa.com>) 3,16 g/L additionné d'agar 7 g/L, de saccharose 10 g/L, de pourpre de bromocrésol (BCP) 8 g/L. Le milieu peut être également supplémenté d'un ou plusieurs antibiotiques aux concentrations nécessaires à la sélection des graines transformées. Les graines semées sont stratifiées à 4°C à l'obscurité pendant 48h, puis placées en chambre de culture (16h de lumière à une intensité de 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$, 18°C, 60% d'humidité relative). Les boîtes de Petri sont scellées avec du ruban adhésif Micropore™ (3M, Saint-Paul, Minnesota, États-Unis)

2.1.3.2. *Physcomitrium patens*

La liste des milieux et solutions utilisés pour la culture et la transformation des protoplastes de mousse est donnée en Annexe 7. Les mousses (*Physcomitrium patens*) Grandsden type sauvage sont cultivées *in vitro* sur du milieu PPNH₄ solide complété avec du NH₄-tartrate 2,7 mM, favorisant la croissance des chloronemas et améliorant le rendement lors de l'isolation de protoplastes (Ashton et al., 1979; Thévenin et al., 2012). Le milieu de culture est recouvert de disques de cellophane (Cellulose type 325P, A.A. Packaging Limited, Preston, Lancashire, Royaume-Uni) stériles afin de faciliter le prélèvement des protonemas. Les boîtes de Petri sont scellées avec du ruban adhésif Micropore™ Ref 1530-0 (3M, Saint-Paul, Minnesota, États-Unis). Les mousses sont cultivées en conditions contrôlées : 24°C, 70% d'humidité avec un régime lumineux de 16h de lumière, 8h d'obscurité. Des spores sont mises à germer afin d'obtenir une culture de mousses à repiquer chaque semaine. Les mousses sont prélevées et broyées dans de l'eau stérile en utilisant un homogénéiseur Polytron PT1200E (Fisher Scientific, Hampton, New Hampshire, États-Unis) à axes

thermorésistants (7 mm x 82 mm), puis filtrées à travers un filtre de 140 µm, et rincées avec de l'eau stérile. Les mousses ainsi fragmentées sont alors mises en suspension dans de l'eau stérile et 2,5 mL de solution de mousses sont inoculées sur une nouvelle boîte de Petri dans les conditions décrites ci-dessus.

2.1.4. Culture en serre et chambre de culture

2.1.4.1. *Arabidopsis thaliana*

Les plantes sont cultivées soit à partir de germinations de graines sèches, ou après stratification à 4°C pendant 48h sur du papier filtre Whatman® (Maidstone, Royaume-Uni), soit à partir de plantules cultivées *in vitro* pendant une dizaine de jours et repiquées sur du terreau enrichi en NPK (mélange de tourbe de sphaigne et de sable à pH=5,5, TREF). Les plantes sont cultivées en serre, dans des conditions d'humidité relative variable, photopériode minimum de 13h assurée par un éclairage additionnel à l'aide d'ampoules Philips-E40 spectre rouge – HPS 400W SONT PIA PLUS, 120 µE.m-2.sec-1, 16h à 20°C / 8h à 15°C) et irrigués par une solution minérale nutritive enrichie en azote (pH=6,5 ; Lesaint et Coïc, 1983).

2.1.4.2. *Nicotiana benthamiana*

Les graines de *Nicotiana benthamiana* sont semées sur terreau enrichi en NPK (mélange de tourbe de sphaigne et de sable à pH=5,5, TREF) en barquettes. Les jeunes plantules âgées d'une semaine sont repiquées en pots carrés. Les plants sont cultivés en serre durant quatre à cinq semaines (humidité relative variable ; photopériode minimum de 13h assurée par un éclairage additionnel à l'aide d'ampoules Philips-E40 spectre rouge – HPS 400W SONT PIA PLUS, 120 µE.m-2.sec-1, 16h à 20°C / 8h à 15°C) et irrigués par une solution minérale nutritive enrichie en azote (pH=6,5 ; Lesaint et Coïc, 1983).

2.1.5. Transformation stable et transitoire

2.1.5.1. Transformation stable d'*Arabidopsis thaliana*

Les plantes sont transformées par trempage des fleurs dans une solutions d'agrobactéries (*floral dip*). Deux jours avant la transformation, un clone d'agrobactérie portant le vecteur d'intérêt est prélevé et ensemencé dans 5 mL de LB liquide supplémenté des antibiotiques nécessaires à la sélection du clone portant le vecteur et les plasmides *helpers*. La préculture est incubée 28°C avec agitation pendant une nuit. La veille de la transformation, 500 µL de la préculture sont ensemencés dans 50 mL de LB liquide supplémenté des antibiotiques nécessaires à la sélection du clone. La culture est incubée à 28°C avec agitation pendant une nuit. Le jour de la transformation, la culture

d'agrobactéries est centrifugée à 4 000 g pendant 10 minutes, puis resuspendue dans 75 mL de tampon d'infiltration (saccharose 5% (p/v), MgCl₂ 0,2% (p/v), Silwet L77 15% (v/v)). Les plantes en fleurs, âgées d'environ 1 mois, sont alors transformées en immergeant les hampes florales dans la solution d'agrobactéries pendant 5 minutes. Les graines ayant intégré l'ADN-t d'intérêt, *i.e.* les transformants primaires, sont sélectionnés selon l'outil de sélection présent dans le vecteur : résistance à un antibiotique ou fluorescence rouge des graines.

2.1.5.2. Transformation transitoire de *Physcomitrium patens*

Les protoplastes de mousses sont transformés selon le protocole développé dans l'équipe (Thévenin et al., 2012). Dans un premier temps, les protoplastes sont isolés : après 6-7 jours de croissance, le protonéma de 2 boîtes est récolté et incubé avec une agitation douce par basculement à 25 tr/min, à l'obscurité, pendant 3 heures, dans un erlenmeyer contenant 10 mL de mannitol 8,5% (p/v, Sigma-Aldrich, Saint-Louis, Missouri, États-Unis) et 10 mL de driselase à 2% (p/v, Sigma-Aldrich). La préparation est filtrée à travers 3 tamis superposés (140 µm, 80 µm et 40 µm), puis rincée avec 10 mL de mannitol 8,5%. La solution de protoplastes est centrifugée trois fois à 60 g, à température ambiante, pendant 5 minutes. Après chaque centrifugation, le surnageant est jeté et le culot est doucement remis en suspension dans 10 mL de mannitol 8,5%. Après la dernière centrifugation, le culot est remis en suspension dans la solution MMM à température ambiante à la concentration finale de 0,5.10⁶ protoplastes/mL (voir liste des solutions en Annexe 7). Une fois, isolés, 300 µL de protoplastes en suspension sont transfectés simultanément avec 5 µg de chaque plasmide (un ou plusieurs facteurs de transcription et un promoteur cible fusionné à la GFP, respectivement clonés dans des vecteurs pBS TPP-A et B). La transformation est médiée par 300 µL de PEG 4000 40% (p/v), et un choc thermique à 45°C pendant 7 min dans un bain-marie. Dans chaque tube, 6,5 mL de milieu liquide (PPNH₄ et mannitol 8,5%) sont rapidement ajoutés. Les protoplastes ainsi transfectés sont placés à 24°C (température ambiante), à l'obscurité, pendant 48h, jusqu'à l'[analyse de la fluorescence au cytomètre de flux](#).

2.1.5.3. Transformation transitoire de *Nicotiana benthamiana*

Des feuilles de tabacs sont infiltrées avec une solution d'agrobactéries afin de permettre l'expression transitoire des gènes d'intérêt. Deux jours avant l'agro-infiltration, un clone d'agrobactérie portant le vecteur est prélevé et ensemencé dans 4 mL de LB liquide supplémenté des antibiotiques nécessaires à la sélection du vecteur et des plasmides *helpers*. La préculture est incubée à 28°C, avec agitation, pendant une nuit. La veille de la transformation, 500 µL de la préculture d'agrobactéries sont ensemencés dans 10 mL de LB liquide supplémenté des antibiotiques nécessaires. La culture est incubée à 28°C avec agitation pendant une nuit. Le jour de

l'agro-infiltration, la culture d'agrobactéries est centrifugée à 5 000 g, à 10°C, pendant 10 min, puis re-suspendue dans 10 mL de milieu d'infiltration (milieu de bouturage 1,3% (p/v), saccharose 4% (p/v), pH=5,7). La solution est de nouveau centrifugée à 5 000 g, à 10°C, pendant 10 min, puis re-suspendue dans 10 mL de milieu d'infiltration. La DO_{500 nm} de la solution d'agrobactéries est ajustée à une DO_{500 nm}=0,5. Afin d'améliorer l'expression du transgène, les feuilles peuvent être co-infiltrées avec des agrobactéries portant le vecteur *helper p19* (Vanhercke et al., 2013) permettant l'expression d'un inhibiteur de *silencing*. La solution d'agrobactéries est placée dans une seringue de 1 mL (sans aiguille) et infiltrée sur la face abaxiale d'une feuille de tabac âgé de 4-5 semaines, en appliquant une pression, de part et d'autre de la nervure centrale. Les feuilles sont observées après 3 jours en [microscopie confocale](#).

2.2. Biologie moléculaire

2.2.1. PCR

2.2.1.1. PCR haute-fidélité

Si l'on souhaite amplifier un fragment d'ADN avec le moins d'erreur possible, par exemple pour amplifier un fragment d'ADN pour le cloner dans un vecteur, il est nécessaire d'utiliser une Taq polymérase très efficace qui n'induit pas d'erreurs et possédant une activité de relecture. La matrice utilisée peut être de l'ADN génomique ou bien un plasmide contenant le fragment d'intérêt, dilué alors au 1/100^{ème} ou au 1/1000^{ème}. Pour les PCR haute-fidélité, la Phusion™ High-Fidelity DNA Polymerase (2 U/μL) (ThermoScientific, Waltham, Massachusetts, États-Unis) est utilisée, selon les conditions décrites ci-dessous :

Mélange PCR	Volume
ADN	1 μL
Phusion™ HF Buffer 5X	10 μL
dNTP 5 mM	2 μL
Amorce 1 (10 pmol/μL)	2,5 μL
Amorce 2 (10 pmol/μL)	2,5 μL
Phusion™ High-Fidelity DNA Polymerase (2 U/μL)	0,5 μL
H ₂ O qsp 50 μL	31,5 μL

Programme PCR	Température	Temps
Dénaturation	98°C	30 sec
35 cycles	Dénaturation	98°C
	Hybridation	55°C
	Élongation	72°C
Élongation finale	72°C	10 min

2.2.1.2. PCR sur colonies

Au cours des clonages, la vérification des clones ayant intégré le bon vecteur est réalisée par PCR sur colonies. Pour cela, un clone est piqué avec une pointe de cône et dilué dans 10-20 μL d'eau et

1 µL de cette dilution est utilisée comme matrice. Les PCR sur colonies sont réalisées avec une Taq polymérase produite au laboratoire, selon les conditions suivantes :

Mélange PCR	Volume
ADN	1 µL
DreamTaq Buffer 10 X (ThermoScientific)	2 µL
dNTP 2 mM	0,5 µL
Amorce 1 (10 pmol/µL)	1 µL
Amorce 2 (10 pmol/µL)	1 µL
Taq polymérase « maison »	0,5 µL
H ₂ O qsp 20 µL	14 µL

Programme PCR	Température	Temps	
Dénaturation	94°C	5 min	
35 cycles	Dénaturation	94°C	30 sec
	Hybridation	55°C	30 sec
	Élongation	72°C	1 min/kb
Élongation finale	72°C	10 min	

2.2.1.3. PCR de génotypage

La liste des amorces utilisées pour le génotypage des plantes est donnée en Annexes 2 et 3. Lorsque les amorces utilisées sont efficaces, il est possible de le réaliser avec la Taq polymérase « maison » dans les conditions décrites ci-dessus pour les PCR sur colonies. Cependant, lorsque l'amplification de l'ADN d'intérêt est plus difficile, comme c'est le cas pour le mutant *lec2-1*, par exemple, le génotypage peut être réalisé avec une Taq polymérase commerciale plus efficace. Ici, la DreamTaq DNA Polymerase (5 U/µL) (ThermoScientific) est utilisée dans les conditions suivantes :

Mélange PCR	Volume
ADN	1 µL
DreamTaq Buffer 10 X	2 µL
dNTP 5 mM	0,8 µL
Amorce 1 (10 pmol/µL)	0,4 µL
Amorce 2 (10 pmol/µL)	0,4 µL
DreamTaq DNA Polymerase (5 U/µL)	0,2 µL
H ₂ O qsp 20 µL	15,2 µL

Programme PCR	Température	Temps	
Dénaturation	94°C	5 min	
35 cycles	Dénaturation	94°C	30 sec
	Hybridation	55°C	30 sec
	Élongation*	72°C	1 min/kb*
Élongation finale	72°C	10 min	

* Dans le cas du génotypage du mutant *lec2-1*, 1 min 40 sec d'élongation.

2.2.2. Gel d'électrophorèse

Afin de déterminer la taille d'un fragment d'ADN ou d'ARN, une électrophorèse sur gel d'agarose est réalisée. Les acides nucléiques sont déposés avec du tampon de charge 0,5 X (bleu de bromophénol 0,25% p/v (Sigma), EDTA 0,1 M (v/v), glycérol 40% (v/v)) sur un gel d'agarose 1% (p/v, Sigma), TAE 0,5 X (v/v, ThermoScientific), 10 µg/µL de bromure d'éthydiu – BET (Euromedex, Souffelweyersheim, France). Le gel est placé dans une cuve d'électrophorèse Mupid® One dans du tampon d'électrophorèse TAE 0,5 X (ThermoScientific). La migration s'effectue à 135 V pendant 25 min. Le marqueur de taille 1 kb DNA Ladder (ThermoScientific) est utilisé pour déterminer la taille

des fragments. La lecture du gel se fait par éclairage UV avec un Gel Doc XR+ System (Bio-Rad, Hercules, Californie, États-Unis).

2.2.3. Clonage Gateway™

La stratégie de clonage Gateway™ permet d'introduire un fragment d'ADN d'intérêt dans un vecteur d'expression plasmidique de façon orientée et sans utiliser d'enzymes de restriction mais par recombinaison. La réaction de recombinaison a lieu entre des sites Gateway™ (attB/attP, attL/attR...) présents en bordure du fragment d'ADN et dans le vecteur de destination. Les clonages Gateway™ ont été réalisés en suivant les instructions du développeur, Invitrogen (Thermo Scientific). La liste des vecteurs utilisés est donnée dans la partie [1.4.1 Vecteurs Gateway™](#) de ce chapitre.

2.2.3.1. Réactions de recombinaison

L'insert d'intérêt est amplifié par PCR et des bordures Gateway™ AttB sont ajoutées en amont et en aval. La première réaction de recombinaison, notée BP, permet d'insérer l'amplicon dans un vecteur d'entrée, grâce à la recombinaison entre les sites AttB de l'insert et AttP contenues dans le vecteur. La recombinaison BP est réalisée en incubant 3 µL de produit de PCR avec 150 ng de plasmide et 1 µL de Gateway™ BP Clonase™ II (Invitrogen) à 25-28°C pendant 1 à 2h, ou sur la nuit. La réaction est stoppée en ajoutant 0,5 µL de protéinase K (2 µg/µL) et en incubant à 37°C 10 minutes. Une partie du produit de réaction BP (~2 µL) est transformé dans des bactéries électro-compétentes (voir partie [2.2.3. Culture et transformation des bactéries](#)). La réaction de recombinaison LR a lieu entre un vecteur d'entrée et un vecteur de destination et permet de transférer l'insert du premier vecteur vers le second via leurs sites de recombinaison AttL et AttR respectivement. La réaction LR se fait dans les mêmes conditions que la BP décrite ci-dessus : 100-300 ng de vecteur d'entrée, 150 ng de vecteur de destination et 1 µL de Gateway™ LR Clonase™ II (Invitrogen) sont incubés à 25°C pendant 1 à 2h, puis la réaction est stoppée avec de la protéinase K de la même manière et les bactéries électro-compétentes sont transformées.

2.2.3.2. Clonages Gateway™ Chapitre 1

[pCaMV35S::GFP:LEC2*SH dans pGWB2](#)

À partir des constructions disponibles dans l'équipe SEEDDEV, la *GFP* clonée dans pDONR221 et *LEC2*SH* dans pDONR207, une nouvelle construction *pCaMV35S::GFP:LEC2*SH* a été clonée dans pGWB2. *LEC2*SH* est une version modifiée de *LEC2* obtenue par Céline Boulard par synthèse à façon, dont la séquence en acides nucléiques est différente de *AtLEC2* natif, mais dont la séquence protéique reste inchangée. Cela a permis l'introduction du promoteur 35S permettant l'expression

ectopique nécessaire pour le crible ainsi que de la GFP pour le suivi de la localisation de la protéine. Les sites Gateway™ attB5 et attB2 ont été ajoutés au fragment *LEC2*SH* par PCR haute-fidélité. *LEC2*SH* a ainsi pu être recombinaison dans le vecteur pDONR221 par réaction BP. Une réaction de LR-multiple, a permis d'assembler les fragments *GFP* et *LEC2*SH* dans un vecteur pBS Tpp-A permettant l'expression dans des protoplastes de mousse (*Physcomitrium patens*) sous le contrôle du promoteur de l'actine de riz pour tester sa fonctionnalité. Le fragment *GFP:LEC2*SH* cloné dans pBS Tpp-A a ensuite été recombinaison dans un vecteur pDONR207 grâce à une réaction de BP-reverse, avant d'être recombinaison par LR dans le vecteur binaire pGWB2 permettant l'expression dans *Arabidopsis thaliana* sous le contrôle d'un promoteur constitutif 35S.

2.2.3.3. Clonages Gateway™ Chapitre 2

Variants de LEC2 pour l'expression transitoire dans les protoplastes de mousse

Différents variants de LEC2 ont été amplifiés par PCR haute-fidélité, permettant l'ajout des extrémités AttB, puis clonés par recombinaison BP dans un vecteur pDONR207 avant une recombinaison LR dans le vecteur pBS Tpp-A permettant l'expression dans des protoplastes de mousse (*Physcomitrium patens*) sous le contrôle du promoteur de l'actine de riz.

Variants et fragments de LEC2 pour l'expression dans la levure

De la même manière que décrit précédemment, les variants ou fragments de LEC2 ont été amplifiés par PCR et clonés dans pDONR207, puis une recombinaison LR a permis de les cloner dans le vecteur pDEST22 permettant la fusion entre un domaine d'activation de la transcription putatif (ici les variants de LEC2) et le domaine de liaison à l'ADN de *Gal4* de levure.

2.2.4. Clonage GoldenBraid 2.0

Le GoldenBraid 2.0 un système de clonage itératif permettant d'assembler plusieurs unités transcriptionnelles les unes à la suite des autres, dans un vecteur d'expression binaire, par une succession de réactions de digestion, grâce à des enzymes de restriction de type IIS, puis de ligation (Sarrion-Perdigones et al., 2011, 2013). Contrairement aux enzymes de restriction de type II, qui reconnaissent et coupent directement un site spécifique, les enzymes de restriction de type IIS sont capables de reconnaître un site de restriction et de couper quelques nucléotides en amont, ou en aval. Les enzymes utilisées pour les réactions GoldenBraid sont *BsaI* et *BsmBI* selon le vecteur d'intérêt (voir voir tableau partie [1.1.1. Vecteurs GoldenBraid 2.0](#)).

2.2.4.1. Réactions de digestion-ligation

Dans un premier temps, les fragments d'ADN à cloner doivent être domestiqués : les éventuels sites de restriction *BsaI* et *BsmBI* présents sont retirés par PCR (avec des amorces portant des mutations)

ou par synthèse à façon en créant une mutation ponctuelle silencieuse. Des séquences spécifiques appelées préfixe et suffixe sont ensuite ajoutées en bordure du fragment. Ces séquences sont définies selon les consignes données par les développeurs (<https://gbcloning.upv.es/>). Elles contiennent les sites de restriction des enzymes *Bsal* et *BsmBI*, ainsi qu'une suite de nucléotides précise qui, une fois coupée par l'enzyme, laissera une extrémité flanquante cohésive, qui pourra être liguée à l'extrémité d'un autre fragment.

La première réaction du clonage, la réaction de domestication, permet d'insérer le fragment domestiqué dans un vecteur d'entrée pUPD2 grâce à une réaction de digestion par l'enzyme *BsmBI*, puis une ligation par l'enzyme T4 ligase. Ainsi, chaque partie de la future unité transcriptionnelle que l'on veut former est insérée dans un vecteur d'entrée pUPD ou pUPD2. Ensuite, lors de l'assemblage multipartite, les différentes composantes de la future unité transcriptionnelle (promoteur, séquence codante, étiquette, terminateur...) domestiqués dans pUPD ou pUPD2 sont assemblées dans un vecteur d'expression grâce aux extrémités cohésives des bordures générées après digestion. Les vecteurs de destination utilisés sont pDGB3_α1, pDGB3_α2, pDGB3_Ω1, pDGB3_Ω2, voir tableau partie [1.1.1. Vecteurs GoldenBraid 2.0](#) de ce chapitre. Les clonages sont réalisés la façon suivante : deux unités transcriptionnelles contenues dans des vecteurs pDGB3_α1 et pDGB3_α2 sont assemblées dans un vecteur pDGB3_Ω1 ou pDGB3_Ω2 puis, de la même manière, les deux unités transcriptionnelles dans pDGB3_Ω1 et pDGB3_Ω2 sont assemblées dans des vecteurs pDGB3_α1 ou pDGB3_α2 et ainsi de suite.

Les mélanges réactionnels sont composés d'environ 40 ng de chaque ADN de départ (amplicon PCR haute-fidélité ou plasmide), 75 ng de vecteur d'entrée ou de destination selon l'étape du clonage, 10 U d'enzyme de restriction *BsmBI/Esp3I* 10 U/μL (ThermoFisherScientific ER0452) ou *Bsal* 5000 U (NewEngland BioLabs R0535L), 1 U d'ADN T4 ligase 1 U/μL (ThermoFisherScientific EL0016), du tampon T4 ligase correspondant (1X final) et d'eau stérile pour un volume final de 10 μL. La réaction de digestion-ligation est faite dans un thermocycleur programmé de la façon suivante :

Programme		Température	Temps
Digestion		37°C	10 min
25 cycles	Digestion	37°C	2 min
	Ligation	16°C	5 min
Digestion		37°C	10 min
Inactivation des enzymes		80°C	20 min

2.2.4.2. Clonages GoldenBraid 2.0 Chapitre 1

Certaines sous-unités ont été synthétisées à façon afin de les domestiquer : MCP:AID, Scaffold2MS2 (Twist Bioscience, San Francisco, Californie, États-Unis), ARN guides (Eurofins Genomics, Les Ulis, France). La dCas9 provient du kit GoldenBraid 2.0 (<https://gbcloning.upv.es/>), les sous-unités promotrices et terminatrices ainsi que les plasmides ont été fournis par l'équipe de Jean-Denis FAURE (DIPOL, IJPB).

Clonages CRISPR-X

- [Cas particulier : les ARN guides](#)

Les ARN guides ont été conçus grâce au site internet CRISPOR (<http://crispor.tefor.net/crispor.py>) qui permet de déterminer les meilleurs candidats, à partir d'une séquence d'intérêt, ici la séquence de *LEC2*SH*. Les ARN guides ont été sélectionnés en fonction de leur emplacement sur la séquence de *LEC2*SH*, ainsi que sur la base d'un faible nombre de mutations hors-cibles (« *off-targets* ») potentielles dans le génome d'*Arabidopsis thaliana* (Annexe 1). Le site internet du GoldenBraid (<https://gbcloning.upv.es/do/crispr/>) a été utilisé afin de domestiquer les ARNs guides proposés par le site CRISPOR, c'est-à-dire leur apporter les bons préfixes et suffixes nécessaires au clonage. Les ARNs guides sont commandés sous la forme d'oligonucléotides sens et antisens (Eurofins Genomics, Les Ulis, France). Ils sont dilués à une concentration de 1 µM et hybridés pendant 30 min à température ambiante pour former un insert d'ADN double brin avant la réaction de digestion-ligation. Les séquences des ARN guides utilisés sont données en Annexe 1.

- [Clonage des parts et unités transcriptionnelles](#)

Domestication dans pUPD ou pUPD2		
Parts	Vecteur	Référence ou provenance
pU6-26	pUPD	(Vazquez-Vilar et al., 2016)
ARN guides	/	
Scaffold2MS2	pUPD2	
pAtUbq10	pUPD	(Sarrion-Perdigones et al., 2013)
MCP:AID	pUPD2	
tRbcSE9	pUPD2	Lionel Gissot, équipe DIPOL, IJPB
pRPS5A	pUPD2	Lionel Gissot, équipe DIPOL, IJPB
dCas9 (D10A H840A)	pUPD2	(Vazquez-Vilar et al., 2016)
Premier niveau d'unités transcriptionnelles		
Unité transcriptionnelle	Vecteur	Référence ou provenance
pU6-26::ARN guide::Scaffold2MS2	pDGB3_α1/2	

pAtUbq10::MCP:AID:tNOS	pDGB3_α2	
pRPS5A::dCas9:tRbcSE9	pDGB3_α1	Lionel Gissot, équipe DIPOL, IJPB
pCsVMV::DsRed:tNOS	pDGB3_α2	Lionel Gissot, équipe DIPOL, IJPB
Deuxième niveau d'unités transcriptionnelles		
pU6-26::ARN guide1::Scaffold2MS2:pU6-26::ARN guide2::Scaffold2MS2	pDGB3_Ω1	
pRPS5A::dCas9:tRbcSE9;pCsVMV::DsRed:tNOS	pDGB3_Ω2	
Troisième niveau d'unités transcriptionnelles		
pU6-26::ARN guide::Scaffold2MS2:pU6-26::ARN guide2::Scaffold2MS2:pRPS5A::dCas9:tRbcSE9;pCsVMV::DsRed:tNOS	pDGB3_α1	
Quatrième niveau d'unités transcriptionnelles		
pU6-26::ARN guide::Scaffold2MS2:pU6-26::ARN guide2::Scaffold2MS2:pRPS5A::dCas9:tRbcSE9;pCsVMV::DsRed:tNOS;pAtUbq10::MCP:AID:tNOS	pDGB3_Ω1	

Clonages EvolvR

Les polymérases infidèles REV1 et Poll3M ont été synthétisées à façon (Twist Bioscience, San Francisco, Californie, États-Unis).

Domestication dans pUPD ou pUPD2		
Parts	Vecteur	Référence ou provenance
pRPS5A	pUPD2	Lionel Gissot, équipe DIPOL, IJPB
pAtUbq10	pUPD	(Sarrion-Perdigones et al., 2013)
nCas9	pUPD2	
REV1	pUPD2	
Poll3M	pUPD2	
tRbcSE9	pUPD2	Lionel Gissot, équipe DIPOL, IJPB
Premier niveau d'unités transcriptionnelles		
Unité transcriptionnelle	Vecteur	Référence ou provenance
pRPS5A::nCas9:REV1:tRbcSE9	pDGB3_α1	
pAtUbq10::nCas9:REV1:tRbcSE9	pDGB3_α1	
pRPS5A::nCas9:Poll3M:tRbcSE9	pDGB3_α1	

2.2.4.3. Clonages GoldenBraid 2.0 Chapitre 2

Expression de variants de LEC2 *in planta* sous contrôle de *pLEC2*

Domestication dans pUPD ou pUPD2		
Parts	Vecteur	Référence ou provenance
pLEC2 (-500 pb)	pUPD2	
GFP	pUPD2	Lionel Gissot, équipe DIPOL, IJPB
LEC2 et variants	pUPD2	
tRbcSE9	pUPD2	Lionel Gissot, équipe DIPOL, IJPB
Premier niveau d'unités transcriptionnelles		
Unité transcriptionnelle	Vecteur	Référence ou provenance
pLEC2::LEC2:tRbcSE9	pDGB3_α1	
pLEC2::GFP:LEC2:tRbcSE9	pDGB3_α1	
pNOS::Hygro:tNOS	pDGB3_α2	Bertrand Dubreucq
pCsVMV::DsRed:tNOS	pDGB3_α2	Lionel Gissot, équipe DIPOL, IJPB
Deuxième niveau d'unités transcriptionnelles		
pLEC2::LEC2:tRbcSE9::pNOS::Hygro:tNOS	pDGB3_Ω1	
pLEC2::GFP:LEC2/Δ:tRbcSE9::pNOS::Hygro:tNOS	pDGB3_Ω1	
pLEC2::GFP:LEC2/Δ:tRbcSE9::pCsVMV::DsRed:tNOS	pDGB3_Ω1	

Expression transitoire dans le tabac de variants de LEC2

Domestication dans pUPD ou pUPD2		
Parts	Vecteur	Référence ou provenance
pAtUbq10	pUPD	(Sarrion-Perdigones et al., 2013)
GFP	pUPD2	Lionel Gissot, équipe DIPOL, IJPB
LEC2 et variants	pUPD2	
tRbcSE9	pUPD2	Lionel Gissot, équipe DIPOL, IJPB
Premier niveau d'unités transcriptionnelles		
Unité transcriptionnelle	Vecteur	
pAtUbq10::GFP:LEC2/Δ:tRbcSE9	pDGB3_α1	

2.2.5. Culture et transformation des bactéries

Les bactéries électro-compétentes (*Escherichia coli* et *Agrobacterium tumefaciens*) sont transformées par électroporation avec le plasmide d'intérêt : 1-2 ng de plasmide et 40 µL d'eau stérile sont ajoutés dans un tube contenant 10 µL de bactéries électro-compétentes ($1,5 \cdot 10^6$ cfu/µg

d'ADN) concentrées dans du glycérol 10% (v/v). Le mélange est transféré dans une cuvette d'électroporation et soumis à un choc électrique de 1,25 kV pendant 3 à 5 ms. La membrane bactérienne est ainsi déstabilisée et l'ADN peut entrer dans la cellule. Le mélange est remis en suspension dans 1 mL de milieu LB liquide et incubé à 37°C avec agitation (300 tr/min) pendant 1h pour *Escherichia coli* ou à 28°C pour *Agrobacterium tumefaciens* avant étalement sur milieu solide. Les bactéries sontensemencées sur une boîte de Petri contenant du milieu LB gélosé (15 g/L agar) supplémenté d'un ou plusieurs antibiotiques aux concentrations requises afin de sélectionner les clones qui ont intégré le vecteur souhaité. Les boîtes sont mises à incuber à 37°C pendant une nuit pour *E. coli* et à 28°C, 48h pour *A. tumefaciens*.

2.2.5.1. Concentrations des molécules de sélection des clones

Molécule	Concentration stock	Dilution finale
Ampicilline	100 mg/mL	100 µg/mL
Carbénicilline	100 mg/mL	100 µg/mL
Céfotaxime	200 mg/mL	200 µg/mL
Chloramphénicol	30 mg/mL	30 µg/mL
Gentamycine	50 mg/mL	50 µg/mL
Hygromycine	100 mg/mL	100 µg/mL
Rifampicine	25 mg/mL	25 µg/mL
Spectinomycine	50 mg/mL	50 µg/mL
X-Gal	40 mg/mL	40 µg/mL
IPTG	1 M	0,5 mM

2.2.6. Culture et transformation des levures

Les levures non transformées sont cultivées sur du milieu YPD solide (YPD 50 g/L Clontech, agar 20 g/L, pH=6,5) à 28°C pendant 48h. La veille de la transformation, une colonie est prélevée et cultivée à 28°C avec agitation (300 tr/min), une nuit, dans 10 mL de milieu YPD liquide (50 g/L, pH=6,5). La $DO_{600\text{ nm}}$ de la préculture liquide est mesurée, de manière à ensemencer $5 \cdot 10^6$ cellules dans 50 mL de milieu YPD liquide, dans un Erlenmeyer ($DO_{600} = 0,1 = 1 \cdot 10^6$ cellules/mL) à 28°C avec agitation. La $DO_{600\text{ nm}}$ est de nouveau mesurée après 2 générations (4-5h), la transformation peut être réalisée lorsque la culture atteint une densité de 1 à $2 \cdot 10^7$ cellules/mL. Les levures sont alors centrifugées à 3 000 g pendant 5 min, lavées dans 25 mL d'eau stérile, puis de nouveau centrifugées à 3 000 g pendant 5 minutes. Le culot est repris dans 1 mL de LiAc 100 mM, puis centrifugé pendant quelques secondes à 16 000 g. Le surnageant est retiré et le culot est resuspendu dans 400-500 µL de LiAc 100 mM, homogénéisé au vortex et 50 µL de levures sont transférés dans un nouveau tube pour chaque transformation.

Un mélange est réalisé, comprenant 500-1 000 ng de plasmide à transformer, 10 µL d'ADN carrier et de l'eau, qsp 84 µL. Dans le tube comprenant les 50 µL de levures à transformer, 240 µL de PEG 50% (p/v) sont ajoutés, puis 36 µL de LiAc 1M et enfin les 84 µL du mélange de plasmide et d'ADN carrier. Le mélange est homogénéisé au vortex pendant 1 min, puis les levures sont incubées à 28°C pendant 30 min, puis homogénéisées de nouveau et placées à 42°C pendant 25 min afin d'être transformées par choc thermique. Les tubes sont placés dans la glace, puis centrifugés à 16 000 g pendant quelques secondes afin de culotter les levures, qui sont alors reprises dans 200 µL d'eau stérile et étalées sur du milieu minimum supplémenté des acides aminés nécessaires à la sélection des levures ayant intégré le vecteur conférant une auxotrophie à un acide aminé. Ici, le milieu de sélection utilisé est SD-leu (26,7 g/L Minimal SD base Clontech, 0,69 g - Leu DO Supplement Clontech, agar 20 g/L). Les levures transformées sont alors cultivées à 28°C pendant 48h.

2.2.7. Extraction d'acides nucléiques

2.2.7.1. Extraction d'ADN plasmidique

Les clones d'*E. coli* portant le plasmide d'intérêt sont prélevés et mis en culture dans 8 mL de milieu LB liquide contenant le ou les antibiotique(s) nécessaire(s) à la sélection, à 37°C, pendant une nuit. Les bactéries sont centrifugées pendant 10 minutes à 5 000 g, puis les plasmides sont extraits selon le protocole NucleoSpin® Plasmid Macherey-Nagel (mini-prep). L'extraction est basée sur une lyse alcaline suivie d'une purification sur une colonne contenant une membrane de silice sur laquelle les acides nucléiques se lient spécifiquement. Les plasmides sont élués dans 50 µL d'eau stérile. Une mini-prep permet d'obtenir une concentration d'ADN d'environ 100-250 ng/µL. Pour obtenir une plus forte quantité d'ADN, nécessaire à la transformation transitoire des protoplastes de mousses (voir chapitre [Transformation transitoire de *Physcomitrium patens*](#)), l'extraction de plasmide est faite à partir d'une culture de 50 mL avec un kit NucleoBond® Plasmid, qui permet d'obtenir des rendements d'environ 1 000 ng/µL.

2.2.7.2. Extraction d'ADN de plantes

Extraction rapide

Pour l'extraction d'ADN servant à du génotypage uniquement, un protocole d'extraction rapide est utilisé. Des morceaux de feuille d'environ 0,2 cm² sont prélevés et placés dans des tubes de transfert de 1,2 mL contenant une bille métallique, puis plongés dans l'azote liquide et broyés à 25 vibrations/secondes pendant 45 secondes en utilisant un broyeur vibrant. La poudre obtenue est centrifugée pendant à 4 000 g, à 4°C, pendant 5 minutes, puis re-suspendue dans 200 µL de tampon d'extraction (Tris HCl 200 mM pH=7,5, NaCl 250 mM, EDTA 25 mM, 0,5% SDS) et centrifugée à 4

000 g, à 15°C, pendant 10 min. 100 µL de surnageant sont prélevés et transférés dans une plaque 96 puits contenant 100 µL d'isopropanol. Les échantillons sont incubés à température ambiante durant 10 min, puis centrifugés à 4 000 g, à 15°C, pendant 30 min. Le surnageant est éliminé et les culots d'ADN sont lavés avec 100 µL d'éthanol 75%, puis centrifugés à 4 000 g pendant 5 minutes. Le surnageant est éliminé et les culots sont séchés pendant 15 min à 65°C et re-suspendus dans 100 µL de TE (Tris 10 mM pH=8,0, EDTA 1 mM pH=8,0) à 4°C, avec agitation, pendant une nuit.

Extraction au CTAB

Le matériel végétal est broyé dans de l'azote liquide avec du sable de Fontainebleau, ainsi que du PVPP (Polyvinylpyrrolidone – Sigma Aldrich, Saint-Louis, Missouri, États-Unis) s'il s'agit de graines : le PVPP piège les tanins qui réduisent l'efficacité d'extraction des acides nucléiques. La poudre est mise en suspension dans un volume de tampon d'extraction au CTAB, bromure de cetyltriméthylammonium (2% CTAB (p/v), NaCl 1,4 M, EDTA 20 mM, Tris-HCl 100 mM, 0,2% β-mercaptoéthanol (v/v)), préalablement chauffé à 65°C. Le volume de tampon est défini en fonction de la quantité d'échantillon et un tube avec fermeture à vis est utilisé afin d'empêcher l'ouverture liée aux échappements de gaz. Après avoir homogénéisé au vortex et dégazé les tubes, les échantillons sont incubés à 65°C pendant 30 min avec agitation. Après incubation, un volume de chloroforme (CHCl₃) est ajouté. Les échantillons sont homogénéisés au vortex puis centrifugés à 16 000 g pendant 20 min. Trois phases sont alors visibles : la phase aqueuse contenant les acides nucléique, un gâteau protéique et la phase organique contenant les solvants. La phase aqueuse est prélevée et transférée dans un nouveau tube. Un volume de phénol chloroforme 24:1 (v/v) est ajouté puis les échantillons sont centrifugés à 16 000 g pendant 5 min. La phase aqueuse est prélevée et transférée dans un nouveau tube, puis les acides nucléiques sont précipités en ajoutant 0,5 volume d'isopropanol et 10% (v/v) de NaCl 5M, avant de centrifuger à 16 000 g pendant 20 min. Le surnageant est retiré et le culot contenant les acides nucléiques est lavé en ajoutant 500 µL d'éthanol 70%, puis en centrifugeant à 16 000 g pendant 5 min. Enfin, l'éthanol est retiré, puis le culot séché et remis en suspension dans 50 µL d'eau.

2.2.7.3. Extraction d'ARN de plantes

Extraction d'ARN avec le kit RNeasy® Plant Mini Kit (Qiagen)

Le matériel végétal est broyé dans de l'azote liquide avec du sable de Fontainebleau, puis les ARN totaux sont extraits selon le protocole RNeasy® Plant Mini Kit (Qiagen, Hilden, Allemagne).

Extraction d'ARN sur de petites quantités de matériel avec le kit PicoPure™ (ThermoFischerScientific)

Entre 2 et 5 mg de matériel est congelé dans l'azote liquide, puis broyé en présence de PVPP s'il s'agit de graines. Les ARN sont extraits en ajoutant 100 µL de tampon XB contenu dans le kit PicoPure™, suivi d'une incubation à 42°C, avec une agitation de 500 rpm, pendant 30 min, après avoir brièvement homogénéisé au vortex. Après incubation, les échantillons sont centrifugés à 3 000 g pendant 2 minutes. Le surnageant est prélevé, puis transféré dans un nouveau tube contenant 100 µL d'éthanol 70% (v/v), en s'assurant de bien mélanger en pipettant et refoulant. La colonne est préconditionnée en ajoutant 250 µL de tampon CB, incubée à température ambiante pendant 5 min et centrifugée à 16 000 g pendant 1 min. Le mélange d'extraits cellulaires et d'éthanol est transféré dans une colonne préconditionnée, puis centrifugé à 100 g pendant 2 min pour lier les acides nucléiques à la membrane, avant d'être centrifugé à 16 000 g pendant 1 min pour retirer le surnageant. 100 µL de tampon de lavage W1 sont ajoutés avant une centrifugation à 8 000 g pendant 1 min. Le traitement à la DNase est réalisé en ajoutant 10 µL de DNaseI et 30 µL de tampon RDD à la colonne et en incubant à température ambiante pendant 15 min. 40 µL de tampon de lavage W1 sont ajoutés, puis les colonnes sont centrifugées à 8 000 g pendant 15 sec. La membrane est lavée deux fois en ajoutant 100 µL de tampon de lavage W2, éliminé par centrifugation à 8 000 g pendant 1 min, puis à 16 000 g pendant 2 min. Le liquide passé à travers la colonne est retiré, puis la membrane est séchée en centrifugeant à 16 000 g pendant 1 min. La colonne de purification est transférée dans un nouveau tube et l'ARN extrait est élué en ajoutant 12 µL de tampon d'éluion EB directement sur la membrane, puis en incubant à température ambiante pendant 1 min. La colonne est centrifugée à 1 000 g pendant 1 min afin que le tampon d'éluion se propage bien sur la membrane, puis à 16 000 g pendant 1 min pour éluer l'ARN. L'éluât peut alors être dosé au NanoDrop ou avec une puce Agilent et est conservé à -20°C ou -80°C pour une conservation plus longue.

2.2.7.4. Rétrotranscription des ARN en ADNc

Les ARN totaux extraits peuvent être rétrotranscrits en ADN complémentaire d'après le protocole de la RevertAid H Minus Reverse Transcriptase (ThermoScientific), en utilisant un oligo-dT pour l'initiation. Ici, aucun traitement à la DNase n'a été réalisé mais une vérification par PCR de l'absence d'ADN génomique a été effectuée.

2.3. Biochimie

2.3.1. Analyse des protéines totales et western-blot

2.3.1.1. Extraction des protéines de réserve de graines d'*Arabidopsis thaliana*

Les protéines sont extraites à partir de 50 graines placées dans un tube Eppendorf de 2 mL à vis, avec une bille en céramique et 250 µL de tampon d'échantillon LDS NuPAGE™ 2X (Novex®), puis broyées deux fois de suite en utilisant un FastPrep® réglé sur 6,5 M/sec pendant 30 sec. La mousse est réduite en centrifugeant brièvement les tubes puis les graines sont de nouveau broyées une fois dans les mêmes conditions. Les échantillons sont centrifugés à 13 000 g à température ambiante pendant 5 min pour éliminer la mousse formée, puis le liquide est prélevé et transféré dans un nouveau tube de 1,5 mL à vis avant d'être de nouveau centrifugés à 13 000 g à température ambiante pendant 10 min. Un culot doit se former et éventuellement une phase lipidique superficielle. Le surnageant est transféré dans un nouveau tube 1,5 mL et évitant de prélever la phase lipidique. Si l'extrait ne semble pas clair, la dernière étape est répétée. Les extraits protéiques ainsi obtenus sont conservés à -20°C.

2.3.1.2. Dépôt sur gel SDS-PAGE – électrophorèse de protéines

Avant le dépôt sur gel, les protéines extraites sont centrifugées à 13 000 g pendant 5 min. Des gels commerciaux NuPAGE™ 12% Bis-Tris Gel 10 ou 12 puits (Invitrogen) sont utilisés. Le gel est placé dans une cuve de migration en présence de tampon de migration MOPS SDS NuPAGE® 1 X (Novex®). Les puits du gel sont lavés avec du tampon MOPS à la pipette. Avant le dépôt, les protéines sont chauffées à 50°C pendant 10 min. L'équivalent de 4 graines (~ 20 µL) est déposé dans chaque puits, ainsi que 5 µL de marqueur de taille SeaBlue® Pre-Stained Standard 1X (Novex®). Avant le début de la migration, 500 µL de d'antioxydant NuPAGE™ (fourni avec le gel) sont déposés dans la partie haute de la cuve. La migration est réalisée à 230 V, 45 mA pour 1 gel ou 90 mA pour 2 gels, jusqu'à ce que le bleu de charge atteigne le boudin formé par le gel, mais sans le dépasser (~1h30). Le gel est coloré sur la nuit à température ambiante dans 100 mL solution de bleu de Coomassie : 80% (v/v) acide orthophosphorique (acide orthophosphorique (2,5%, v/v), sulfate d'ammonium (10%, p/v)) et bleu de Coomassie (0,4% bleu de Coomassie dilué dans de l'éthanol (p/v)). Le gel est rincé avec de l'eau ultrapure jusqu'à ce qu'il soit bien décoloré et stocké à 4°C en attendant la lecture. Le gel est scanné et les bandes quantifiées en utilisant le logiciel *ImageQuantTL* (GE Healthcare Life Sciences, Uppsala, Suède).

2.3.1.3. Western-blot anti-OLE1

Une électrophorèse de protéines SDS-PAGE est réalisée de la même manière que décrit ci-dessus, mais en déposant 100 fois moins de matériel : dilution au 1/100^{ème} des protéines extraites, dans du NuPAGE™ tampon d'échantillon LDS NuPAGE™ 2X (Novex®). Après migration, le gel n'est pas coloré avec une solution de bleu de Coomassie, mais préparé pour le transfert sur membrane de nitrocellulose à l'aide du kit de transfert Trans-Blot Turbo RTA Mini 0,2 µm Nitrocellulose (Bio-Rad, Hercules, Californie, États-Unis). Le gel est placé dans du tampon de transfert (1X Trans-Blot Turbo Transfert Buffer, éthanol absolu 20% (v/v)), tandis qu'une membrane de nitrocellulose (fournie dans le kit) est trempée dans de l'éthanol absolu, puis dans du tampon de transfert. Le montage est réalisé de la manière suivante : dans une des cassettes du Trans-Blot Turbo Transfer System (Bio-Rad), quelques feuilles de papier de transfert (fournies dans le kit) et préalablement imbibées dans le tampon de transfert sont placées en évitant les bulles d'air, la membrane est positionnée au-dessus du papier, puis le gel et enfin une dernière couche de papier de transfert. Le rouleau fourni avec la machine de transfert est utilisé entre chaque couche pour retirer les bulles d'air. Le transfert est réalisé à 15 V, 30 mA, pendant 30 min. Après le transfert, la membrane est incubée dans une solution de saturation (lait demi-écrémé en poudre Régilait 5%, TBS 1X, Tween20 0,05%) à 4°C avec agitation pendant une nuit. Le lendemain, après un rinçage rapide puis 10 min de lavage avec agitation dans du TBS-T 0,05% (TBS 1X, Tween20 0,05%), la membrane est incubée avec l'anticorps primaire* dilué dans 10 mL de solution de saturation, à température ambiante avec agitation pendant 1h30. La membrane est rincée rapidement puis lavée 3 fois pendant 10 min avec agitation dans du TBS-T 0,05%. La membrane est alors incubée avec l'anticorps secondaire** dilué dans 10 mL de solution de saturation de la même manière que pour l'anticorps primaire. La membrane est de nouveau rincée rapidement, puis lavée 2 fois pendant 10 min avec agitation dans du TBS-T 0,05%, suivi d'un dernier lavage pendant 10 min avec agitation dans du TBS 1X (sans Tween20). La révélation est réalisée avec le kit Clarity™ Western ECL Substrate (Bio-Rad) : 750 µL de chaque solution sont mélangés et déposés sur pochette plastifiée, puis la membrane est placée face contre le liquide et incubée pendant 5 min, puis révélée avec un ImageQuant™ LAS 4000 (GE Healthcare, Uppsala, Suède).

* anticorps primaire : anti-OLE1 « 644253, lapin 6442 à S3ci, sérum 3, 23/07/2003 » (Sabine D'Andréa, équipe DYSCOL, IJPB). Dilution 1/50 000^{ème}. (D'Andréa et al., 2007)

** anticorps secondaire : anti-IgG lapin (équipe DYSCOL, IJPB). Dilution 1/50 000^{ème}. (D'Andréa et al., 2007)

2.3.2. Extraction des lipides et analyse par chromatographie en phase gazeuse

2.3.2.1. Extraction et transestérification des acides gras totaux

Les acides gras totaux sont extraits à partir de lots de 50 graines, entières. Chaque échantillon est incubé au bain-marie (94°C, 1h30) dans 1,32 mL d'un mélange de méthanol/acide sulfurique/toluène (100:2,5:30, v/v/v) en présence de C17:0 (un acide gras absent chez les plantes et faisant ici office d'étalon interne). Après refroidissement des échantillons dans la glace, 500 µL d'hexane et 1,5 mL d'eau salée (0,9% NaCl, m/v) sont successivement ajoutés dans chaque tube. Une phase organique contenant les esters méthyliques se forme alors après agitation et centrifugation des tubes à 2 000 g, à 4°C pendant 10 min. Celle-ci est prélevée puis concentrée si nécessaire sous flux de diazote dans des fioles à fond plat jusqu'à obtention d'un volume équivalent à environ 90 µL. La concentration des échantillons est réalisée proportionnellement à la quantité de lipides qu'ils contiennent.

2.3.2.2. Chromatographie en phase gazeuse

Les acides gras transestérifiés sont analysés en routine par chromatographie en phase gazeuse sur un appareil Agilent 7890B. Pour cela, 1 µL de chaque échantillon est injecté dans une colonne (EC-WAX, 30 m x 0,53 mm). D'abord soumise à une température de 160°C, celle-ci est ensuite progressivement amenée à 190°C (+ 40°C/min), puis à 230°C (+ 4°C/min). La température finale est maintenue durant 2 min. La détection des différents esters méthyliques est assurée par un détecteur à ionisation de flamme (Browse et al., 1986). L'identification de la nature des acides gras est rendue possible par comparaison de leurs temps de rétention avec ceux de standards commerciaux. L'interface numérique du logiciel *OpenLab CDS ChemStation Edition* (Agilent) permet d'ajuster les paramètres et d'intégrer les pics obtenus.

2.3.3. ChIP-seq

2.3.3.1. Extraction de la chromatine

Des graines en développement (300 mg par échantillon) au stade globulaire-cœur (4 à 5 jours après fécondation) issues de 3 génotypes (Ws, pLEC2::GFP:AtLEC2 et pLEC2::GFP:Δ_{N-ter}LEC2 en fond *lec2-1*) sont prélevées sans les siliques, congelées dans l'azote liquide, puis conservées à -80°C. Le matériel végétal est broyé deux fois avec un *Tissue Lyser II* (Qiagen, Hilden, Allemagne) réglé sur 30 vibrations/sec pendant 1 minute afin d'extraire les noyaux, puis mis en suspensu dans 10 mL de tampon de fixation (HEPES 10 mM pH=7,6, saccharose 0,5 M, KCl 5 mM, MgCl₂ 5 mM, EDTA 5 mM, inhibiteur de protéase Roche cOmplete™ – 2 tablettes pour 100 mL, β-mercaptoéthanol 14 mM, Triton X-100 0,6%). La suspension cellulaire est doublement fixée au DSG 2,5 mM pendant 1 heure

avec agitation continue sur une roue, puis au formaldéhyde 1% pendant 5 minutes. La réaction de fixation est rapidement arrêtée en ajoutant de la glycine 0,2 M. À partir de cette étape et sauf mention contraire, les échantillons sont constamment placés à 4°C. Afin d'augmenter l'efficacité d'extraction des noyaux, les échantillons sont homogénéisés dans la glace en utilisant un *douncer* et filtrés avec un tamis cellulaire de 70 µm *ClearLine®* ou du *Miracloth*. Les noyaux sont ensuite culottés par centrifugation à 3 000 g à 4°C pendant 10 minutes en utilisant un rotor oscillant puis re-suspendus délicatement dans 300 µL de tampon d'isolation nucléaire (HEPES 10 mM pH=7,6, saccharose 0,5 M, KCl 5 mM, MgCl₂ 5 mM, EDTA 5 mM, inhibiteur de protéase Roche cOmplete™ – 2 tablettes pour 100 mL). À partir de cette étape et jusqu'à la fin du protocole, des tubes *Eppendorf DNA LoBind 1,5 mL (Safe-Lock)* sont utilisés afin de limiter la liaison de l'ADN et des protéines aux parois du tube. Afin de séparer les noyaux des chloroplastes restants, les organites sont séparés sur gradient de Percoll (600 µL de gradient de Percoll 15% pH=8-9, HEPES 10 mM pH=8,8, saccharose 0,5 M, MgCl₂ 5 mM, KCl 5 mM, EDTA 5 mM), par centrifugation à 2 000 g à 4°C pendant 5 minutes en utilisant un rotor oscillant.

2.3.3.2. Fragmentation de la chromatine par sonication

Le culot de noyaux est resuspendu dans 950 µL de tampon de lyse nucléaire (Tris-HCl 50 mM pH=7,5, SDS 0,1%, EDTA 10 mM), puis transférés dans des tubes *Covaris® milliTUBES 1mL AFA Fiber* et soumis à 15 minutes de sonication en utilisant un *Covaris® S220 Focused-ultrasonicator* (*Covaris®*, Woburn, Massachusetts, Etats-Unis). Afin de tester l'efficacité de la fragmentation, 50 µL de chromatine sont prélevés avant et après sonication, puis soumis à une réaction réverse de fixation afin de séparer l'ADN et les protéines, en ajoutant du Tris-HCl 5 mM pH=6,5, EDTA 1,25 mM, NaCl 20 mM. Les échantillons sont incubés à 65°C avec agitation à 500 rpm sur la nuit.

Le lendemain, 48 µg de RNaseA sont ajoutés, puis les échantillons sont mis à incuber à 65°C pendant 1 heure. Enfin, 20 µg de protéinase K sont ajoutés et les échantillons sont de nouveau placés à 65°C durant 1 heure. La réaction est purifiée avec un kit *MinElute® Reaction Cleanup Kit* (Qiagen, Hilden, Allemagne), puis l'éluât est déposé sur un gel d'électrophorèse (agarose 1,5%, TAE 0,5 X, BET 10 µg/µL) et mis à migrer à 100 V pendant 1 à 2 heures. Pour les échantillons soniqués, une taille comprise entre 200 et 500 pb est attendue. S'il reste des fragments de plus grande taille, les échantillons peuvent être soniqués une seconde fois pendant 5 minutes. Une fois la sonication complète, les débris sont culottés par centrifugation à 16 000 g, à 15°C, pendant 5 minutes et le surnageant est transféré dans de nouveaux tubes.

2.3.3.3. Immunoprécipitation de la chromatine

Pour une meilleure efficacité, des billes magnétiques *Dynabeads™ Protein A* et *Protein G* (30 mg/mL, Invitrogen) sont utilisées. Toutes les étapes se font à 4°C, sauf mention contraire. Pour chaque immunoprécipitation, 100 µg de chaque type de bille est mélangé pour une quantité totale de 200 µg de billes couplées A et G par échantillon. Les billes sont lavées avec 1 mL de tampon de dilution de ChIP (Tris-HCl 15 mM pH=7,5, NaCl 150 mM, Triton X-100 1%, EDTA 1 mM pH=8,0) puis placées sur un portoir aimanté. Le tampon est retiré puis les billes sont re-suspendues un volume adéquat (prévoir 20 µL de suspension de billes par échantillon) et 2 µg d'anticorps dirigé contre la GFP (réf. ab290, Abcam, Cambridge) sont ajoutés. Les billes sont pré-incubées avec l'anticorps à 4°C sur une roue, pendant 1 à 2 heures. Le SDS contenu dans le tampon de lyse nucléaire doit être dilué au moins 10 fois car il peut limiter l'interaction entre l'anticorps et sa cible. Par conséquent, 300 µL de chromatine fragmentée sont immunoprécipités dans 2,7 mL de tampon de dilution de ChIP et 100 µL de billes couplées aux anticorps. L'immunoprécipitation est réalisée à 4°C sur une roue, sur la nuit. Afin de garder un contrôle négatif d'immunoprécipitation (*input*), 1/10^{ème} du volume de chromatine utilisé pour l'immunoprécipitation (*i.e.* 30 µL) est prélevé et gardé à 4°C. Le reste de chromatine fragmentée peut être conservé à -80°C.

2.3.3.4. Lavages de l'immunoprécipitation

Des lavages successifs sont réalisés afin d'éliminer la chromatine non immunoprécipitée et les interactions non spécifiques. Chaque étape de lavage est réalisée deux fois en utilisant un portoir aimanté : une fois rapide, en re-suspendant les billes à la pipette, une autre fois en incubant les billes 15 minutes à 4°C sur une roue. Il est très important d'éviter de faire des bulles, car elles limitent l'attraction des billes par l'aimant. Le premier lavage est réalisé avec du tampon de dilution de ChIP, puis avec du tampon à faible teneur en sel (SDS 0,1%, Triton X-100 1%, EDTA 2 mM, Tris-HCl 20 mM pH=8,9, NaCl 150 mM), un troisième (facultatif) est fait avec du tampon à haute teneur en sel (SDS 0,1%, Triton X-100 1%, EDTA 2 mM, Tris-HCl 20 mM pH=8,9, NaCl 150 mM). Ce lavage est réalisé dans le cas d'une immunoprécipitation des histones, mais pas pour lorsqu'il s'agit de facteurs de transcription, car la forte teneur en sel peut déstabiliser l'interaction des facteurs de transcription avec l'ADN, qui est assez faible. Les échantillons sont ensuite lavés avec un tampon de lavage LiCl (LiCl 0,25 M, NP40/Ipepal 1%, deoxycholate 1%, EDTA 1 mM pH=8,0, TrisHCl 20 mM pH=8,0). Enfin, un dernier lavage est réalisé de la même manière, mais à température ambiante avec du TE. Les billes sont transférées dans un nouveau tube.

2.3.3.5. Élu­tion de la chromatine immunoprécipitée

Après le dernier lavage, le TE est retiré et les billes sont re-suspendues (sans faire de bulles) dans 100 µL de tampon d'élu­tion (NaHCO₃ 0,1 M, SDS 1%) puis mises à incuber à 65°C avec agitation à 500-700 rpm pendant 10 minutes. Le surnageant est retiré et transféré dans de nouveaux tubes, puis les billes sont éluées une deuxième fois de la même manière. Les 30 µL de chromatine fragmentée prélevée avant l'immunoprécipitation (*input*) sont également repris dans 170 µL de tampon d'élu­tion. Afin d'arrêter la fixation des protéines sur l'ADN, du NaCl 25 mM est ajouté aux 200 µL de chromatine éluée ainsi qu'aux tubes *input*. Les échantillons sont mis à incuber à 65°C sur la nuit. Le lendemain du TrisHCl 5 mM pH=6,5, EDTA 1,25 mM pH=8,0 et 10 µg de protéinase K sont ajoutés aux échantillons qui sont mis à incuber à 40-50°C pendant 1 à 2 heures. La réaction est purifiée en utilisant le kit *MinElute® Reaction Cleanup Kit* (Qiagen), mais en ajoutant 550 µL de tampon ERC et en éluant avec 13 µL de tampon EB, car une partie du volume élué est utilisé pour le dosage de la chromatine.

2.3.3.6. Dosage de la chromatine immunoprécipitée

Après élu­tion, la chromatine est dosée en utilisant un *NanoDrop™ 3300* (ThermoFisherScientific, Wilmington, Delaware, États-Unis) permettant le dosage de picogrammes d'ADN par µL. Le dosage est réalisé en suivant le protocole *PicoGreen® Assay Protocol* (ThermoFisherScientific, Wilmington, Delaware, États-Unis). La chromatine non immunoprécipitée (*input*) est dosée au *NanoDrop™ 2000*.

2.3.3.7. Librairie et séquençage

Pour chaque répliques biologiques, 100 pg d'ADN immunoprécipité (IP) ou génomique (*input*) ont été utilisés pour préparer des bibliothèques avec *MicroPlex Library Preparation Kit v2* (Diagenode, Seraing, Belgique). La qualité des bibliothèques a été validée à l'aide de l'analyseur de fragments Agilent/Advanced Analytical. Les librairies multiplexées ont été séquençées à l'aide d'un système NextSeq 500 (Illumina) avec des fragments de 41 paires de paires. Les *reads* ont été mis en correspondance avec l'assemblage du génome tair10 à l'aide de *Bowtie2 v2.3.4.1* avec des paramètres très sensibles p10 et en autorisant une erreur d'appariement. Le marquage des doublons a été effectué à l'aide de *PicardTools MarkDuplicates v2.8.1* et le filtrage des régions génomiques présentant une couverture aberrante ou une faible complexité de séquence (Quadrana et al., 2016) a été effectué à l'aide de *samtools v1.7* pour éliminer les biais de séquençage. Les pics ont été déterminés à l'aide de *MACS2 v2.1.1* en mode étroit, au format BAMPE, avec une taille de génome fixée à 10⁷, un paramètre keep-dup fixé à 5, et en utilisant l'option B.

Les pics ChIP-seq ont été associés à un gène lorsqu'ils se chevauchaient complètement, partiellement ou lorsqu'ils se trouvaient à l'intérieur du gène. Si aucun chevauchement n'a été trouvé, le pic a été attribué au gène annoté le plus proche dans une fenêtre de 5 kb. Dans les rares cas où un pic était à égale distance de deux gènes, les deux annotations ont été conservées. Le logiciel *Integrative Genomics Viewer – IGV* (Robinson et al., 2011) a été utilisé pour visualiser la distribution du signal ChIP-seq sur des gènes représentatifs.

2.4. Cytologie et imagerie

2.4.1. Axiozoom

L'expression et la localisation de la GFP dans des plantules cultivées *in vitro* a été observée grâce à un Axio Zoom.V16 (Zeiss, Oberkochen, Allemagne). La GFP est observée à une longueur d'onde d'excitation de 488 nm et une longueur d'onde d'émission entre 495 et 535 nm.

2.4.2. Cytométrie en flux

Une fois les protoplastes de mousse transformés et incubés à l'obscurité pendant 48h (voir [Transformation transitoire de *Physcomitrium patens*](#)), le surnageant de milieu PPNH₄ liquide est retiré et la solution de protoplastes est concentrée avec 800 µL de mannitol 8,5% ou de milieu PPNH₄ liquide. La solution de protoplastes est filtrée au travers d'un filtre 50 µm (CellTrics®, Partec, Görlitz, Allemagne). La fluorescence de la GFP est détectée pour chaque protoplaste à l'aide d'un cytomètre de flux (Partec CyFlow® Space, laser 488 nm 20 mW) et les données sont traitées avec le logiciel FlowMax®. La fluorescence GFP est détectée avec un filtre FITC 527 nm/30 nm (canal FL1, 275 V) comme décrit dans (Thévenin et al., 2012).

2.4.3. Microscopie confocale

Les acquisitions ont été réalisées avec un microscope confocal inversé SP5 ou droit SP8 (Leica Microsystems, Wetzlar, Germany). Selon les réglages suivants pour les longueurs d'onde d'excitation et d'émission :

Molécule fluorescente	Longueur d'onde d'excitation	Longueur d'onde d'émission
GFP	488 nm (Argon)	495-535 nm
RFP	561 nm (Diode 561)	580-620 nm
DAPI	405 nm (UV)	450-490 nm
Plastes	488 nm (Argon)	650-700 nm
Rouge Nil (oléosomes)	488 nm (Argon)	550-650 nm

2.4.3.1. Observation des embryons d'*Arabidopsis thaliana*

Des embryons au stade cœur sont prélevés en ouvrant des siliques âgées de 4-5 jours, les graines sont déposées sur une lame avec une goutte de 50% glycérol, contenant du DAPI (4',6-diamidino-2-phénylindole) dilué au 1/1000^{ème}. Les embryons sont extraits des graines en appuyant avec des pinces fines sur la lamelle, puis laissés à l'obscurité pendant 30 minutes afin que le DAPI entre dans les cellules. La GFP est détectée, ainsi que les noyaux grâce à la fluorescence du DAPI.

2.4.3.2. Observation des cellules de *Nicotiana benthamiana*

Trois jours après [la transformation](#), un morceau de feuille de tabac agro-infiltrée d'environ 1 cm² est prélevé au niveau de la zone d'infiltration, mais en dehors du point d'infiltration, puis placé sur une lame, face abaxiale sur le dessus. Le montage est réalisé dans une goutte d'eau et une lamelle est placée au contact de la face abaxiale de la feuille. L'observation de la fluorescence de la GFP, de la RFP et des plastes est faite directement en microscopie confocale.

2.4.3.3. Observation des oléosomes dans les embryons d'*Arabidopsis thaliana*

Des graines sèches sont mises à imbiber sur papier filtre Whatman® (Maidstone, Royaume-Uni) humide, pendant 1h, 24h ou 48h, puis les embryons sont extraits des graines à l'aide de pinces fines et d'un scalpel. Les embryons sont placés sur une lame et le montage est réalisé dans du glycérol 50%, rouge Nil 0,1% (v/v) (Sigma-Aldrich). La lamelle est placée délicatement, puis les lames sont mises à incuber à l'obscurité pendant au moins 30 min (Miquel et al., 2014). La taille et la mobilisation des corps lipidiques sont observées au niveau de l'axe et des cotylédons.

2.5. Analyses bio-informatiques

2.5.1. Séquençage et analyse des données

2.5.1.1. Séquençage avec Genoscreen (Lille)

Pour la vérification des séquences lors des clonages, le service de séquençage Sanger proposé par l'entreprise Genoscreen (Lille, France) est utilisée. Les produits de PCR ou les plasmides sont envoyés avec des amorces permettant d'amplifier le fragment d'intérêt. La qualité des électrophorégrammes reçus est analysée avec le logiciel « 4Peaks » (Nucleobytes BV) et la séquence obtenue est alignée avec la séquence de référence en utilisant le logiciel « DNA Strider » (Marck, 1988). Pour les alignements multiples, le programme « CLUSTALW » (<https://www.genome.jp/tools-bin/clustalw>) est utilisé.

2.5.1.2. Séquençage avec BGI (Hong-Kong)

DNBSEQ™ PCR free library and sequencing

Pour le séquençage profond des plantules surexprimant *LEC2*SH* et soumises à la mutagenèse CRISPR-X, l'ADN génomique des feuilles de 30 plantules différentes a été extrait au CTAB, puis le locus *LEC2*SH* a été amplifié par PCR haute-fidélité et rassemblé en un seul échantillon. Environ 10 µg de produit PCR de 1 kb de long a été envoyé à l'entreprise BGI, située à Hong-Kong. Le service « *PCR free library and sequencing* » par la technologie *DNBSEQ™* a été choisi pour séquencer en profondeur le produit de PCR obtenu et ainsi détecter d'éventuelles mutations ponctuelles induites par la mutagenèse CRISPR-X. Le séquençage a permis d'obtenir 37 millions de séquences de 100 pb. Les résultats obtenus ont été analysés grâce au logiciel *CLC Genomic Workbench* (Qiagen, Hilden, Allemagne), en comparaison à la séquence de référence *LEC2*SH*. De l'ADN non soumis à mutagenèse et séquençé dans les mêmes conditions doit servir de contrôle pour comparer la séquence « mutée » et la séquence « sauvage ».

DNBSEQ™ Eukaryotic Strand-specific Transcriptome Resequencing

Pour l'analyse du transcriptome des graines de mutant *lec2-1* complémenté avec différentes versions de *LEC2*, l'ARN de graines en développement au stade torpille-cœur extrait avec le kit PicoPure™ a été envoyé pour séquençage au BGI à Hong-Kong. Pour chaque génotype, 3 échantillons contenant 1 µg d'ARN avec une concentration ≥ 40 ng/µL ont été envoyés. Une bibliothèque de 20 millions de fragments de 100 pb a été réalisée et les fragments ont été alignés sur le génome d'*Arabidopsis thaliana*. Les résultats ont été analysés en utilisant la méthode *DESeq2* (Love et al., 2014) afin de déterminer les gènes différentiellement exprimés entre chaque groupe testé, sur la base d'une *q-value* $< 0,05$. La valeur du *log2FoldChange* > 1 ou < -1 , c'est-à-dire deux fois plus ou deux fois moins exprimé dans une condition par rapport à une autre a également été utilisée.

2.5.1.3. Séquençage du CHIP-seq

Voir partie [2.3.3.7. Librairie et séquençage](#).

2.5.1.4. Données d'ATAC-seq

Les données d'ATAC-seq ont analysées de la même manière que les données de RNA-seq grâce au package R *DESeq2* (Love et al., 2014). Le seuil utilisé pour déterminer un gène différentiellement représenté d'une condition à l'autre est une *q-value* $< 0,05$.

2.5.1.5. Heatmaps (cartes de chaleur)

Les heatmaps ont été réalisés grâce au package R *pheatmap* sur les gènes dont l'expression différentielle a été déterminée grâce à *DESeq2* ($q\text{-value} < 0,05$).

2.5.2. Normalisation des données ADpred

Dans un premier temps, deux points de repères ont été définis sur chaque séquence d'orthologues de LEC2 (voir Annexe 5), au niveau d'acides aminés très fortement conservés dans le domaine B2 et dans le domaine B3. Les points repères de toutes les séquences ont été mis en correspondance en réalisant une mise à l'échelle linéaire et une translation sur un nouvel axe. Sur cet axe résultant de la translation, la distance entre les deux points de référence peut être définie arbitrairement. Ici, on a choisi le nombre moyen d'acides aminés entre les deux points dans les 60 séquences testées. Après normalisation des données individuelles, la moyenne des profils a été calculée afin d'obtenir un profil moyen, correspondant à la probabilité de présence d'un domaine d'activation sur chaque séquence. Un script a été développé sur R par Jasmine Burguet (équipe Modélisation et Imagerie Numérique, IJPB).

Références

Références bibliographiques

- Aida, M., Ishida, T., and Tasaka, M.** (1999). Shoot apical meristem and cotyledon formation during *Arabidopsis* embryogenesis: interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes. *Development* **126**: 1563–1570.
- Alizadeh, M., Hoy, R., Lu, B., and Song, L.** (2021). Team effort: Combinatorial control of seed maturation by transcription factors. *Current Opinion in Plant Biology* **63**: 102091.
- Amack, S.C. and Antunes, M.S.** (2020). CaMV35S promoter – A plant biology and biotechnology workhorse in the era of synthetic biology. *Current Plant Biology* **24**: 100179.
- Ashton, N.W., Cove, D.J., and Featherstone, D.R.** (1979). The isolation and physiological analysis of mutants of the moss, *Physcomitrella patens*, which over-produce gametophores. *Planta* **144**: 437–442.
- Barthole, G.** (2013). Contribution à l'étude du contrôle transcriptionnel de la maturation de la graine d'*Arabidopsis*.
- Batista, R.A., Figueiredo, D.D., Santos-González, J., and Köhler, C.** (2019). Auxin regulates endosperm cellularization in *Arabidopsis*. *Genes Dev.* **33**: 466–476.
- Baud, S. et al.** (2016). Deciphering the molecular mechanisms underpinning the transcriptional control of gene expression by master transcriptional regulators in *Arabidopsis* seed. *Plant Physiology* **171**: 1099–1112.
- Baud, S., Boutin, J.-P., Miquel, M., Lepiniec, L., and Rochat, C.** (2002). An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiology and Biochemistry* **40**: 151–160.
- Baud, S., Dubreucq, B., Miquel, M., Rochat, C., and Lepiniec, L.** (2008). Storage Reserve Accumulation in *Arabidopsis*: Metabolic and Developmental Control of Seed Filling. *The Arabidopsis Book* **6**: e0113.
- Baud, S., Mendoza, M.S., To, A., Harscoët, E., Lepiniec, L., and Dubreucq, B.** (2007). WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*: Oil synthesis regulatory network in seeds. *The Plant Journal* **50**: 825–838.
- Baud, S., Wuillème, S., Lemoine, R., Kronenberger, J., Caboche, M., Lepiniec, L., and Rochat, C.** (2005). The AtSUC5 sucrose transporter specifically expressed in the endosperm is involved in early seed development in *Arabidopsis*: The AtSUC5 sucrose carrier is endosperm specific. *The Plant Journal* **43**: 824–836.
- Baumlein, H., Misera, S., Luerssen, H., Kolle, K., Horstmann, C., Wobus, U., and Muller, A.J.** (1994). The FUS3 gene of *Arabidopsis thaliana* is a regulator of gene expression during late embryogenesis. *Plant J* **6**: 379–387.
- Berger, F., Grini, P.E., and Schnittger, A.** (2006). Endosperm: an integrator of seed growth and development. *Current Opinion in Plant Biology* **9**: 664–670.
- Berger, N., Dubreucq, B., Roudier, F., Dubos, C., and Lepiniec, L.** (2011). Transcriptional Regulation of *Arabidopsis* LEAFY COTYLEDON2 Involves RLE , a cis -Element That Regulates Trimethylation of Histone H3 at Lysine-27. *The Plant Cell* **23**: 4065–4078.
- Bernard, P. and Couturier, M.** (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *Journal of Molecular Biology* **226**: 735–745.
- Bernard, V., Brunaud, V., and Lecharny, A.** (2010). TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. *BMC Genomics* **11**: 166.
- Böhm, C., Seibel, N.M., Henkel, B., Steiner, H., Haass, C., and Hampe, W.** (2006). SorLA Signaling by Regulated Intramembrane Proteolysis. *Journal of Biological Chemistry* **281**: 14547–14553.
- Boulard, C., Fatihi, A., Lepiniec, L., and Dubreucq, B.** (2017). Regulation and evolution of the interaction of the seed B3 transcription factors with NF-Y subunits. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1860**: 1069–1078.
- Boulard, C., Thévenin, J., Tranquet, O., Laporte, V., Lepiniec, L., and Dubreucq, B.** (2018). LEC1 (NF-YB9) directly interacts with LEC2 to control gene expression in seed. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1861**: 443–450.

- Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., and Colot, V.** (2017). DNA methylation dynamics during early plant life. *Genome Biology* **18**: 1–12.
- Bouyer, D., Roudier, F., Heese, M., Andersen, E.D., Gey, D., Nowack, M.K., Goodrich, J., Renou, J.-P., Grini, P.E., Colot, V., and Schnittger, A.** (2011). Polycomb Repressive Complex 2 Controls the Embryo-to-Seedling Phase Transition. *PLoS Genet* **7**: e1002014.
- Braybrook, S. and Harada, J.** (2008). LECs go crazy in embryo development. *Trends in Plant Science* **13**: 624–630.
- Braybrook, S.A., Stone, S.L., Park, S., Bui, A.Q., Le, B.H., Fischer, R.L., Goldberg, R.B., and Harada, J.J.** (2006). Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proceedings of the National Academy of Sciences* **103**: 3468–3473.
- Breuninger, H., Rikirsch, E., Hermann, M., Ueda, M., and Laux, T.** (2008). Differential Expression of WOX Genes Mediates Apical-Basal Axis Formation in the Arabidopsis Embryo. *Developmental Cell* **14**: 867–876.
- Browse, J., McCourt, P.J., and Somerville, C.R.** (1986). Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. *Analytical Biochemistry* **152**: 141–145.
- Buitink, J. and Leprince, O.** (2008). Intracellular glasses and seed survival in the dry state. *Comptes Rendus Biologies* **331**: 788–795.
- Burke, T.W., Willy, P.J., Kutach, A.K., Butler, J.E.F., and Kadonaga, J.T.** (1998). The DPE, a Conserved Downstream Core Promoter Element That Is Functionally Analogous to the TATA Box. *Cold Spring Harbor Symposia on Quantitative Biology* **63**: 75–82.
- Carbonero, P., Iglesias-Fernández, R., and Vicente-Carbajosa, J.** (2016). The AFL subfamily of B3 transcription factors: evolution and function in angiosperm seeds. *Journal of Experimental Botany* **68**: erw458.
- Carels, N.** (2015). A History of Genomic Structures: The Big Picture. In *Plant Biology and Biotechnology*, B. Bahadur, M. Venkat Rajam, L. Sahijram, and K.V. Krishnamurthy, eds (Springer India: New Delhi), pp. 131–178.
- Casanova-Sáez, R. and Voß, U.** (2019). Auxin Metabolism Controls Developmental Decisions in Land Plants. *Trends in Plant Science* **24**: 741–754.
- Chan, S.W.-L., Henderson, I.R., and Jacobsen, S.E.** (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nat Rev Genet* **6**: 351–360.
- Chen, B. et al.** (2022). BABY BOOM regulates early embryo and endosperm development. *Proc. Natl. Acad. Sci. U.S.A.* **119**: e2201761119.
- Chen, H., Li, H., Lu, X., Chen, L., Liu, J., and Wu, H.** (2019). Identification and Expression Analysis of GRAS Transcription Factors to Elucidate Candidate Genes Related to Stolons, Fruit Ripening and Abiotic Stresses in Woodland Strawberry (*Fragaria vesca*). *IJMS* **20**: 4593.
- Cheng, L., Shafiq, S., Xu, W., and Sun, Q.** (2018). EARLY FLOWERING IN SHORT DAYS (EFS) regulates the seed size in Arabidopsis. *Sci. China Life Sci.* **61**: 214–224.
- Choi, K., Kim, J., Hwang, H.-J., Kim, S., Park, C., Kim, S.Y., and Lee, I.** (2011). The FRIGIDA Complex Activates Transcription of *FLC*, a Strong Flowering Repressor in *Arabidopsis*, by Recruiting Chromatin Modification Factors. *The Plant Cell* **23**: 289–303.
- Cole, M., Chandler, J., Weijers, D., Jacobs, B., Comelli, P., and Werr, W.** (2009). *DORNROSCHE* is a direct target of the auxin response factor MONOPTEROS in the *Arabidopsis* embryo. *Development* **136**: 1643–1651.
- Collonnier, C., Guyon-Debast, A., Maclot, F., Mara, K., Charlot, F., and Nogué, F.** (2017). Towards mastering CRISPR-induced gene knock-in in plants: Survey of key features and focus on the model *Physcomitrella patens*. *Methods* **121–122**: 103–117.
- Corcoran, E.T., LeBlanc, C., Huang, Y.-C., Arias Tsang, M., Sarkiss, A., Hu, Y., Pedmale, U.V., and Jacob, Y.** (2022). Systematic histone H4 replacement in *Arabidopsis thaliana* reveals a role for H4R17 in regulating flowering time. *The Plant Cell* **34**: 3611–3631.
- Costa, L.M., Gutiérrez-Marcos, J.F., and Dickinson, H.G.** (2004). More than a yolk: the short life and complex times of the plant endosperm. *Trends in Plant Science* **9**: 507–514.
- Crevillén, P., Yang, H., Cui, X., Greeff, C., Trick, M., Qiu, Q., Cao, X., and Dean, C.** (2014). Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. *Nature* **515**: 587–590.

- D'Andréa, S., Jolivet, P., Boulard, C., Larré, C., Froissard, M., and Chardot, T.** (2007). Selective One-Step Extraction of *Arabidopsis thaliana* Seed Oleosins Using Organic Solvents. *J. Agric. Food Chem.* **55**: 10008–10015.
- Deng, W. and Roberts, S.G.E.** (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev* **19**: 2418–2423.
- Ding, X., Jia, X., Xiang, Y., and Jiang, W.** (2022). Histone Modification and Chromatin Remodeling During the Seed Life Cycle. *Front. Plant Sci.* **13**: 865361.
- Dolfini, D., Gatta, R., and Mantovani, R.** (2012). NF-Y and the transcriptional activation of CCAAT promoters. *Critical Reviews in Biochemistry and Molecular Biology* **47**: 29–49.
- Doll, N.M. and Ingram, G.C.** (2022). Embryo–Endosperm Interactions. *Annu. Rev. Plant Biol.* **73**: 293–321.
- Downes, B.P., Stupar, R.M., Gingerich, D.J., and Vierstra, R.D.** (2003). The HECT ubiquitin-protein ligase (UPL) family in *Arabidopsis*: UPL3 has a specific role in trichome development. *The Plant Journal* **35**: 729–742.
- Drews, G.N. and Koltunow, A.M.G.** (2011). The Female Gametophyte. *The Arabidopsis Book* **9**: e0155.
- Drews, G.N. and Yadegari, R.** (2002). Development and Function of the Angiosperm Female Gametophyte. *Annu. Rev. Genet.* **36**: 99–124.
- Duarte-Aké, F., Nic-Can, G., and De-la-Peña, C.** (2019). Somatic Embryogenesis: Polycomb Complexes Control Cell-to-Embryo Transition. In *Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications*, R. Alvarez-Venegas, C. De-la-Peña, and J.A. Casas-Mollano, eds (Springer International Publishing: Cham), pp. 339–354.
- Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L.** (2003). Radial Patterning of Arabidopsis Shoots by Class III HD-ZIP and KANADI Genes. *Current Biology* **13**: 1768–1774.
- Erijman, A., Kozłowski, L., Sohrabi-Jahromi, S., Fishburn, J., Warfield, L., Schreiber, J., Noble, W.S., Söding, J., and Hahn, S.** (2020). A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning. *Molecular Cell* **78**: 890–902.e6.
- Erkina, T.Y. and Erkine, A.M.** (2016). Nucleosome distortion as a possible mechanism of transcription activation domain function. *Epigenetics & Chromatin* **9**: 40.
- Farnsworth, E.** (2000). The Ecology and Physiology of Viviparous and Recalcitrant Seeds. *Annual Review of Ecology and Systematics* **3**: 107–138.
- Fatih, A., Boulard, C., Bouyer, D., Baud, S., Dubreucq, B., and Lepiniec, L.** (2016). Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds. *Plant Science* **250**: 198–204.
- Figueiredo, D.D., Batista, R.A., Roszak, P.J., Hennig, L., and Köhler, C.** (2016). Auxin production in the endosperm drives seed coat development in Arabidopsis. *eLife* **5**: e20542.
- Figueiredo, D.D., Batista, R.A., Roszak, P.J., and Köhler, C.** (2015). Auxin production couples endosperm development to fertilization. *Nature Plants* **1**: 15184.
- Figueiredo, D.D. and Köhler, C.** (2018). Auxin: a molecular trigger of seed development. *Genes & Development* **32**: 479–490.
- Figueiredo, D.D. and Köhler, C.** (2014). Signalling events regulating seed coat development. *Biochemical Society Transactions* **42**: 358–363.
- Finkelstein, R., Reeves, W., Ariizumi, T., and Steber, C.** (2008). Molecular Aspects of Seed Dormancy. *Annu. Rev. Plant Biol.* **59**: 387–415.
- Focks, N. and Benning, C.** (1998). *wrinkled1*: A Novel, Low-Seed-Oil Mutant of Arabidopsis with a Deficiency in the Seed-Specific Regulation of Carbohydrate Metabolism¹. *Plant Physiology* **118**: 91–101.
- Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D.** (2009). Polycomb Proteins Remain Bound to Chromatin and DNA during DNA Replication In Vitro. *Cell* **137**: 110–122.
- Gaillochet, C., Develtere, W., and Jacobs, T.B.** (2020). CRISPR Screens in Plants: Approaches, Guidelines, and Future Prospects. *The Plant Cell*: tpc.00463.2020.
- Gaj, M.D., Zhang, S., Harada, J.J., and Lemaux, P.G.** (2005). Leafy cotyledon genes are essential for induction of somatic embryogenesis of Arabidopsis. *Planta* **222**: 977–988.

- Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R., and Scheres, B.** (2007). PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. *Nature* **449**: 1053–1057.
- Gao, J. et al.** (2022). A robust mechanism for resetting juvenility during each generation in Arabidopsis. *Nat. Plants* **8**: 257–268.
- Garcês, H.M.P., Champagne, C.E.M., Townsley, B.T., Park, S., Malho, R., Pedroso, M.C., Harada, J.J., and Sinha, N.R.** (2007). Evolution of asexual reproduction in leaves of the genus *Kalanchoe*. *Proceedings of the National Academy of Sciences* **104**: 15578–15583.
- Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F., and Goodman, H.M.** (1992). Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* **4**: 1251–1261.
- Goldberg, R.B., de Paiva, G., and Yadegari, R.** (1994). Plant Embryogenesis: Zygote to Seed. *Science* **266**: 605–614.
- Grobei, M.A., Qeli, E., Brunner, E., Rehrauer, H., Zhang, R., Roschitzki, B., Basler, K., Ahrens, C.H., and Grossniklaus, U.** (2009). Deterministic protein inference for shotgun proteomics data provides new insights into *Arabidopsis* pollen development and function. *Genome Res.* **19**: 1786–1800.
- Guo, H., Fan, Y., Guo, H., Wu, J., Yu, X., Wei, J., Lian, X., Zhang, L., Gou, Z., Fan, Y., and Zeng, F.** (2020). Somatic embryogenesis critical initiation stage-specific ^m CHH hypomethylation reveals epigenetic basis underlying embryogenic redifferentiation in cotton. *Plant Biotechnology Journal* **18**: 1648–1650.
- Guyon-Debast, A., Alboresi, A., Terret, Z., Charlot, F., Berthier, F., Vendrell-Mir, P., Casacuberta, J.M., Veillet, F., Morosinotto, T., Gallois, J.L., and Nogué, F.** (2021). A blueprint for gene function analysis through Base Editing in the model plant *Physcomitrium* (*Physcomitrella*) *patens*. *New Phytologist* **230**: 1258–1272.
- Haberle, V. and Stark, A.** (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* **19**: 621–637.
- Hajheidari, M. and Huang, S.C.** (2022). Elucidating the biology of transcription factor–DNA interaction for accurate identification of cis-regulatory elements. *Current Opinion in Plant Biology* **68**: 102232.
- Halperin, S.O., Tou, C.J., Wong, E.B., Modavi, C., Schaffer, D.V., and Dueber, J.E.** (2018). CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**: 248–252.
- Hamamura, Y., Nagahara, S., and Higashiyama, T.** (2012). Double fertilization on the move. *Current Opinion in Plant Biology* **15**: 70–77.
- Han, J.-D., Li, X., Jiang, C.-K., Wong, G.K.-S., Rothfels, C.J., and Rao, G.-Y.** (2017). Evolutionary Analysis of the LAFL Genes Involved in the Land Plant Seed Maturation Program. *Frontiers in Plant Science* **8**: 439.
- Hand, M.L. and Koltunow, A.M.G.** (2014). The Genetic Control of Apomixis: Asexual Seed Formation. *Genetics* **197**: 441–450.
- Heath, J.D., Weldon, R., Monnot, C., and Meinke, D.W.** (1986). Analysis of storage proteins in normal and aborted seeds from embryo-lethal mutants of *Arabidopsis thaliana*. *Planta* **169**: 304–312.
- Heck, G.R., Perry, S.E., Nichols, K.W., and Fernandez, D.E.** (1995). AGL15, a MADS domain protein expressed in developing embryos. *The Plant Cell* **7**: 1271–1282.
- Hehenberger, E., Kradolfer, D., and Köhler, C.** (2012). Endosperm cellularization defines an important developmental transition for embryo development. *Development* **139**: 2031–2039.
- Heintzman, N.D. and Ren, B.** (2007). The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. *Cell. Mol. Life Sci.* **64**: 386–400.
- Hennig, L. and Derkacheva, M.** (2009). Diversity of Polycomb group complexes in plants: same rules, different players? *Trends in Genetics* **25**: 414–423.
- Hess, G.T., Frésard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., and Bassik, M.C.** (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nature Methods* **13**: 1036–1042.
- Horstman, A., Li, M., Heidmann, I., Weemen, M., Chen, B., Muñio, J.M., Angenent, G.C., and Boutilier, K.** (2017). The BABY BOOM transcription factor activates the LEC1-ABI3-FUS3-LEC2 network to induce somatic embryogenesis. *Plant Physiology*.

- Hu, Y., Zhou, L., Huang, M., He, X., Yang, Y., Liu, X., Li, Y., and Hou, X.** (2018). Gibberellins play an essential role in late embryogenesis of *Arabidopsis*. *Nature Plants* **4**: 289–298.
- Ingouff, M., Jullien, P.E., and Berger, F.** (2007). The Female Gametophyte and the Endosperm Control Cell Proliferation and Differentiation of the Seed Coat in *Arabidopsis*. *The Plant Cell* **18**: 3491–3501.
- Ingram, G.C.** (2010). Family life at close quarters: communication and constraint in angiosperm seed development. *Protoplasma* **247**: 195–214.
- Iwafuchi-Doi, M.** (2019). The mechanistic basis for chromatin regulation by pioneer transcription factors. *WIREs Systems Biology and Medicine* **11**: e1427.
- James, P., Halladay, J., and Craig, E.A.** (1996). Genomic Libraries and a Host Strain Designed for Highly Efficient Two-Hybrid Selection in Yeast. *Genetics* **144**: 1425–1436.
- Jha, P., Ochatt, S.J., and Kumar, V.** (2020). WUSCHEL: a master regulator in plant growth signaling. *Plant Cell Reports* **39**: 431–444.
- Jia, H., Suzuki, M., and McCarty, D.R.** (2014). Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. *WIREs Dev Biol* **3**: 135–145.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E.** (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**: 816–822.
- Jo, L., Pelletier, J.M., and Harada, J.J.** (2019). Central role of the LEAFY COTYLEDON1 transcription factor in seed development. *Journal of Integrative Plant Biology* **61**: 564–580.
- Joshi, C.P.** (1987). An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res* **15**: 6643–6653.
- Jumper, J. et al.** (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589.
- Kahle, J., Baake, M., Doenecke, D., and Albig, W.** (2005). Subunits of the Heterotrimeric Transcription Factor NF-Y Are Imported into the Nucleus by Distinct Pathways Involving Importin β and Importin 13. *Mol Cell Biol* **25**: 5339–5354.
- Kanehisa, M. and Sato, Y.** (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science* **29**: 28–35.
- Karagianni, P., Moulos, P., Schmidt, D., Odom, D.T., and Talianidis, I.** (2020). Bookmarking by Non-pioneer Transcription Factors during Liver Development Establishes Competence for Future Gene Activation. *Cell Reports* **30**: 1319–1328.e6.
- Kaufmann, K. and Airoldi, C.A.** (2018). Master Regulatory Transcription Factors in Plant Development: A Blooming Perspective. In *Plant Transcription Factors*, N. Yamaguchi, ed, *Methods in Molecular Biology*. (Springer New York: New York, NY), pp. 3–22.
- Kazaz, S., Barthole, G., Domergue, F., Ettaki, H., To, A., Vasselon, D., De Vos, D., Belcram, K., Lepiniec, L., and Baud, S.** (2020). Differential Activation of Partially Redundant $\Delta 9$ Stearoyl-ACP Desaturase Genes Is Critical for Omega-9 Monounsaturated Fatty Acid Biosynthesis During Seed Development in *Arabidopsis*. *The Plant Cell* **32**: 3613–3637.
- Kenrick, P. and Crane, P.R.** (1997). The origin and early evolution of plants on land. *Nature* **389**: 33–39.
- Kessler, S.A. and Grossniklaus, U.** (2011). She's the boss: signaling in pollen tube reception. *Current Opinion in Plant Biology* **14**: 622–627.
- Klemm, S.L., Shipony, Z., and Greenleaf, W.J.** (2019). Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**: 207–220.
- Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R.** (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**: 420–424.
- Koncz, C. and Schell, J.** (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *MGG Molecular & General Genetics* **204**: 383–396.
- Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H.** (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. U.S.A.* **106**: 10171–10176.

- Kouzarides, T.** (2007). Chromatin Modifications and Their Function. *Cell* **128**: 693–705.
- 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**: 5–18.
- Kyrchanova, O. and Georgiev, P.** (2021). Mechanisms of Enhancer-Promoter Interactions in Higher Eukaryotes. *IJMS* **22**: 671.
- Lai, W.K.M. and Pugh, B.F.** (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat Rev Mol Cell Biol* **18**: 548–562.
- Lai, X., Verhage, L., Hugouvieux, V., and Zubieta, C.** (2018). Pioneer Factors in Animals and Plants—Colonizing Chromatin for Gene Regulation. *Molecules* **23**: 1914.
- Lange, A., Mills, R.E., Lange, C.J., Stewart, M., Devine, S.E., and Corbett, A.H.** (2007). Classical Nuclear Localization Signals: Definition, Function, and Interaction with Importin α . *Journal of Biological Chemistry* **282**: 5101–5105.
- 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. *Proceedings of the National Academy of Sciences* **100**: 2152–2156.
- Lepiniec, L., Devic, M., Roscoe, T.J., Bouyer, D., Zhou, D.X., Boulard, C., Baud, S., and Dubreucq, B.** (2018). Molecular and epigenetic regulations and functions of the LAFL transcriptional regulators that control seed development. *Plant Reproduction* **31**: 291–307.
- Lewis, B.A. and Reinberg, D.** (2003). The mediator coactivator complex: functional and physical roles in transcriptional regulation. *Journal of Cell Science* **116**: 3667–3675.
- Li, J., Qin, R., Zhang, Y., Xu, S., Liu, X., Yang, J., Zhang, X., and Wei, P.** (2019). Optimizing plant adenine base editor systems by modifying the transgene selection system. *Plant Biotechnology Journal*: 1495–1497.
- Li, Z., Zhang, D., Xiong, X., Yan, B., Xie, W., Sheen, J., and Li, J.F.** (2017). A potent Cas9-derived gene activator for plant and mammalian cells. *Nature Plants* **3**: 930–936.
- Lin, J.-Y., Le, B.H., Chen, M., Henry, K.F., Hur, J., Hsieh, T.-F., Chen, P.-Y., Pelletier, J.M., Pellegrini, M., Fischer, R.L., Harada, J.J., and Goldberg, R.B.** (2017). Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. *Proceedings of the National Academy of Sciences* **114**: E9730–E9739.
- Liu, J., Deng, S., Wang, H., Ye, J., Wu, H.-W., Sun, H.-X., and Chua, N.-H.** (2016). *CURLY LEAF* Regulates Gene Sets Coordinating Seed Size and Lipid Biosynthesis. *Plant Physiol.* **171**: 424–436.
- Liu, Z., Ge, X.-X., Qiu, W.-M., Long, J.-M., Jia, H.-H., Yang, W., Dutt, M., Wu, X.-M., and Guo, W.-W.** (2018). Overexpression of the CsFUS3 gene encoding a B3 transcription factor promotes somatic embryogenesis in Citrus. *Plant Science* **277**: 121–131.
- Lotan, T., Ohto, M., Yee, K.M., West, M.A.L., Lo, R., Kwong, R.W., Yamagishi, K., Fischer, R.L., Goldberg, R.B., and Harada, J.J.** (1998). Arabidopsis LEAFY COTYLEDON1 Is Sufficient to Induce Embryo Development in Vegetative Cells. *Cell* **93**: 1195–1205.
- Louder, R.K., He, Y., López-Blanco, J.R., Fang, J., Chacón, P., and Nogales, E.** (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**: 604–609.
- Love, M.I., Huber, W., and Anders, S.** (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550+.
- Loyola-Vargas, V.M. and Ochoa-Alejo, N.** (2016). *Somatic Embryogenesis: Fundamental Aspects and Applications* V.M. Loyola-Vargas and N. Ochoa-Alejo, eds (Springer International Publishing).
- Lu, Q.S., Dela Paz, J., Pathmanathan, A., Chiu, R.S., Tsai, A.Y.L., and Gazzarrini, S.** (2010). The C-terminal domain of FUSCA3 negatively regulates mRNA and protein levels, and mediates sensitivity to the hormones abscisic acid and gibberellic acid in Arabidopsis. *Plant Journal* **64**: 100–113.
- Marck, C.** (1988). "DNA Strider": a "C" program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res* **16**: 1829–1836.
- Marella, H.H. and Quatrano, R.S.** (2007). The B2 domain of VIVIPAROUS1 is bi-functional and regulates nuclear localization and transactivation. *Planta* **225**: 863–872.

- Martin, K., Kopperud, K., Chakrabarty, R., Banerjee, R., Brooks, R., and Goodin, M.M.** (2009). Transient expression in *Nicotiana benthamiana* fluorescent marker lines provides enhanced definition of protein localization, movement and interactions *in planta*. *The Plant Journal* **59**: 150–162.
- Maruri-López, I., Rodríguez-Kessler, M., Rodríguez-Hernández, A.A., Becerra-Flora, A., Olivares-Grajales, J.E., and Jiménez-Bremont, J.F.** (2014). A maize spermine synthase 1 PEST sequence fused to the GUS reporter protein facilitates proteolytic degradation. *Plant Physiology and Biochemistry* **78**: 80–87.
- Mazzucotelli, E., Mastrangelo, A.M., Crosatti, C., Guerra, D., Stanca, A.M., and Cattivelli, L.** (2008). Abiotic stress response in plants: When post-transcriptional and post-translational regulations control transcription. *Plant Science* **174**: 420–431.
- Meinke, D.W., Franzmann, L.H., Nickle, T.C., and Yeung, E.C.** (1994). Leafy Cotyledon Mutants of Arabidopsis. *The Plant Cell* **6**: 1049–1064.
- Mendoza, M.S., Dubreucq, B., Miquel, M., Caboche, M., and Lepiniec, L.** (2005). LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in *Arabidopsis* leaves. *FEBS Letters* **579**: 4666–4670.
- Miki, T., Ae Park, J., Nagao, K., Murayama, N., and Horiuchi, T.** (1992). Control of segregation of chromosomal DNA by sex factor F in *Escherichia coli*. *Journal of Molecular Biology* **225**: 39–52.
- Miller, C., Wells, R., McKenzie, N., Trick, M., Ball, J., Fatihi, A., Dubreucq, B., Chardot, T., Lepiniec, L., and Bevan, M.W.** (2019). Variation in Expression of the HECT E3 Ligase *UPL3* Modulates LEC2 Levels, Seed Size, and Crop Yields in *Brassica napus*. *The Plant Cell* **31**: 2370–2385.
- Miquel, M., Trigui, G., d'Andréa, S., Kelemen, Z., Baud, S., Berger, A., Deruyffelaere, C., Trubuil, A., Lepiniec, L., and Dubreucq, B.** (2014). Specialization of Oleosins in Oil Body Dynamics during Seed Development in Arabidopsis Seeds. *Plant Physiology* **164**: 1866–1878.
- Mitra, A., Katak, S., Singh, A.N., Gaur, A., Razafindrabe, B.H.N., Kumar, P., Chatterjee, S., and Gupta, D.K.** (2021). Plant Stress, Acclimation, and Adaptation: A Review. In *Plant Growth and Stress Physiology*, D.K. Gupta and J.M. Palma, eds, *Plant in Challenging Environments*. (Springer International Publishing: Cham), pp. 1–22.
- Mitsuda, N. and Ohme-Takagi, M.** (2009). Functional Analysis of Transcription Factors in Arabidopsis. *Plant and Cell Physiology* **50**: 1232–1248.
- Molitor, A.M., Bu, Z., Yu, Y., and Shen, W.H.** (2014). Arabidopsis AL PHD-PRC1 Complexes Promote Seed Germination through H3K4me3-to-H3K27me3 Chromatin State Switch in Repression of Seed Developmental Genes. *PLoS Genetics* **10**.
- Mozgova, I. and Hennig, L.** (2015). The Polycomb Group Protein Regulatory Network. *Annu. Rev. Plant Biol.* **66**: 269–296.
- Mu, J., Tan, H., Zheng, Q., Fu, F., Liang, Y., Zhang, J., Yang, X., Wang, T., Chong, K., Wang, X.-J., and Zuo, J.** (2008). *LEAFY COTYLEDON1* Is a Key Regulator of Fatty Acid Biosynthesis in Arabidopsis. *Plant Physiology* **148**: 1042–1054.
- Müller, K., Bouyer, D., Schnittger, A., and Kermodé, A.R.** (2012). Evolutionarily Conserved Histone Methylation Dynamics during Seed Life-Cycle Transitions. *PLoS ONE* **7**.
- Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., Vorrhein, C., Moras, D., Romier, C., Bolognesi, M., and Mantovani, R.** (2013). Sequence-Specific Transcription Factor NF-Y Displays Histone-like DNA Binding and H2B-like Ubiquitination. *Cell* **152**: 132–143.
- Nguyen Ba, A.N., Pogoutse, A., Provar, N., and Moses, A.M.** (2009). NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* **10**: 202.
- Nishimasu, H. et al.** (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**: 1259–1262.
- Nogué, F., Mara, K., Collonnier, C., and Casacuberta, J.M.** (2016). Genome engineering and plant breeding: impact on trait discovery and development. *Plant Cell Reports* **35**: 1475–1486.
- Oestreich, K.J. and Weinmann, A.S.** (2012). Master regulators or lineage-specifying? Changing views on CD4+ T cell transcription factors. *Nat Rev Immunol* **12**: 799–804.

- Oldfield, A.J., Yang, P., Conway, A.E., Cinghu, S., Freudenberg, J.M., Yellaboina, S., and Jothi, R.** (2014). Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors. *Molecular Cell* **55**: 708–722.
- Olins, D.E. and Olins, A.L.** (2003). Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* **4**: 809–814.
- Pajoro, A. et al.** (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol* **15**: R41.
- Pelletier, J.M., Kwong, R.W., Park, S., Le, B.H., Baden, R., Cagliari, A., Hashimoto, M., Munoz, M.D., Fischer, R.L., Goldberg, R.B., and Harada, J.J.** (2017). LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proceedings of the National Academy of Sciences* **114**: E6710–E6719.
- Petroni, K., Kumimoto, R.W., Gnesutta, N., Calvenzani, V., Fornari, M., Tonelli, C., Holt, B.F., and Mantovani, R.** (2013). The Promiscuous Life of Plant NUCLEAR FACTOR Y Transcription Factors. *The Plant Cell* **24**: 4777–4792.
- Petruk, S., Sedkov, Y., Johnston, D.M., Hodgson, J.W., Black, K.L., Kovermann, S.K., Beck, S., Canaani, E., Brock, H.W., and Mazo, A.** (2012). TrxG and PcG Proteins but Not Methylated Histones Remain Associated with DNA through Replication. *Cell* **150**: 922–933.
- Piatek, A., Ali, Z., Baazim, H., Li, L., Abulfaraj, A., Al-Shareef, S., Aouida, M., and Mahfouz, M.M.** (2015). RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. *Plant Biotechnology Journal* **13**: 578–589.
- Plaschka, C., Larivière, L., Wenzek, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., Villa, E., and Cramer, P.** (2015). Architecture of the RNA polymerase II–Mediator core initiation complex. *Nature* **518**: 376–380.
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A.** (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**: 1173–1183.
- Quadrana, L., Bortolini Silveira, A., Mayhew, G.F., LeBlanc, C., Martienssen, R.A., Jeddloh, J.A., and Colot, V.** (2016). The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife* **5**: e15716.
- Radoeva, T. and Weijers, D.** (2014). A roadmap to embryo identity in plants. *Trends in Plant Science* **19**: 709–716.
- Rajeevkumar, S., Anunanthini, P., and Sathishkumar, R.** (2015). Epigenetic silencing in transgenic plants. *Front. Plant Sci.* **6**.
- Rechsteiner, M. and Rogers, S.W.** (1996). PEST sequences and regulation by proteolysis. *Trends in Biochemical Sciences* **21**: 267–271.
- Reidt, W., Wohlfarth, T., Ellerstrom, M., Czihal, A., Tewes, A., Ezcurra, I., Rask, L., and Baumlein, H.** (2000). Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. *Plant J* **21**: 401–408.
- Ren, B., Yan, F., Kuang, Y., Li, N., Zhang, D., Zhou, X., Lin, H., and Zhou, H.** (2018). Improved Base Editor for Efficiently Inducing Genetic Variations in Rice with CRISPR/Cas9-Guided Hyperactive hAID Mutant. *Molecular Plant* **11**: 623–626.
- Riechmann, J.L. et al.** (2000). *Arabidopsis* Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Science* **290**: 2105–2110.
- Robert, H.S., Grones, P., Stepanova, A.N., Robles, L.M., Lokerse, A.S., Alonso, J.M., Weijers, D., and Friml, J.** (2013). Local Auxin Sources Orient the Apical-Basal Axis in *Arabidopsis* Embryos. *Current Biology* **23**: 2506–2512.
- Robert, H.S., Grunewald, W., Sauer, M., Cannoot, B., Soriano, M., Swarup, R., Weijers, D., Bennett, M., Boutilier, K., and Friml, J.** (2015). Plant embryogenesis requires AUX/LAX-mediated auxin influx. *Development*: dev.115832.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P.** (2011). Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.
- Romier, C., Cocchiarella, F., Mantovani, R., and Moras, D.** (2003). The NF-YB/NF-YC Structure Gives Insight into DNA Binding and Transcription Regulation by CCAAT Factor NF-Y. *Journal of Biological Chemistry* **278**: 1336–1345.

- Roscoe, T.T., Guilleminot, J., Bessoule, J.-J., Berger, F., and Devic, M.** (2015). Complementation of Seed Maturation Phenotypes by Ectopic Expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in *Arabidopsis*. *Plant and Cell Physiology* **56**: 1215–1228.
- Rozsak, P. and Köhler, C.** (2011). Polycomb group proteins are required to couple seed coat initiation to fertilization. *Proc. Natl. Acad. Sci. U.S.A.* **108**: 20826–20831.
- Sainsbury, S., Bernecky, C., and Cramer, P.** (2015). Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**: 129–143.
- Salaün, C., Lepiniec, L., and Dubreucq, B.** (2021). Genetic and Molecular Control of Somatic Embryogenesis. *Plants* **10**: 1467.
- Sano, N., Rajjou, L., North, H.M., Debeaujon, I., Marion-Poll, A., and Seo, M.** (2016). Staying Alive: Molecular Aspects of Seed Longevity. *Plant and Cell Physiology* **57**: 660–674.
- Santos Mendoza, M., Dubreucq, B., Miquel, M., Caboche, M., and Lepiniec, L.** (2005). LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in *Arabidopsis* leaves. *FEBS Letters* **579**: 4666–4670.
- Santos-Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M., and Lepiniec, L.** (2008). Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *The Plant Journal* **54**: 608–620.
- Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juárez, P., Fernández-del-Carmen, A., Granell, A., and Orzaez, D.** (2011). GoldenBraid: An iterative cloning system for standardized assembly of reusable genetic modules. *PLoS ONE* **6**.
- Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., and Orzaez, D.** (2013). GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *Plant Physiology* **162**: 1618–1631.
- Sasnauskas, G., Manakova, E., Lapėnas, K., Kauneckaitė, K., and Siksnyš, V.** (2018). <scp>DNA</scp> recognition by *Arabidopsis* transcription factors <scp>ABI</scp> 3 and <scp>NGA</scp> 1. *The FEBS Journal* **285**: 4041–4059.
- Sawan, C. and Herceg, Z.** (2010). Histone Modifications and Cancer. In *Advances in Genetics* (Elsevier), pp. 57–85.
- Sayou, C. et al.** (2016). A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. *Nat Commun* **7**: 11222.
- Schmitz, R.J., Grotewold, E., and Stam, M.** (2022). Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *The Plant Cell* **34**: 718–741.
- Schuettengruber, B., Martinez, A.-M., Iovino, N., and Cavalli, G.** (2011). Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* **12**: 799–814.
- Seibel, N.M., Eljouni, J., Nalaskowski, M.M., and Hampe, W.** (2007). Nuclear localization of enhanced green fluorescent protein homomultimers. *Analytical Biochemistry* **368**: 95–99.
- Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S.E., Bastolla, U., and Gutierrez, C.** (2014). The Functional Topography of the *Arabidopsis* Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. *The Plant Cell* **26**: 2351–2366.
- Shimatani, Z. et al.** (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nature Biotechnology* **35**: 441–443.
- Siefers, N., Dang, K.K., Kumimoto, R.W., Bynum, W.E., Tayrose, G., and Holt, B.F.** (2009). Tissue-Specific Expression Patterns of *Arabidopsis* NF-Y Transcription Factors Suggest Potential for Extensive Combinatorial Complexity. *Plant Physiology* **149**: 625–641.
- Siloto, R.M.P., Findlay, K., Lopez-Villalobos, A., Yeung, E.C., Nykiforuk, C.L., and Moloney, M.M.** (2006). The Accumulation of Oleosins Determines the Size of Seed Oilbodies in *Arabidopsis*. *The Plant Cell* **18**: 1961–1974.
- Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordân, R., and Rohs, R.** (2014). Absence of a simple code: how transcription factors read the genome. *Trends in Biochemical Sciences* **39**: 381–399.
- Smale, S.T. and Baltimore, D.** (1989). The “initiator” as a transcription control element. *Cell* **57**: 103–113.

- Smith, L.M., Kelleher, N.L., and The Consortium for Top Down Proteomics** (2013). Proteoform: a single term describing protein complexity. *Nat Methods* **10**: 186–187.
- Sorenson, R.S., Deshotel, M.J., Johnson, K., Adler, F.R., and Sieburth, L.E.** (2018). *Arabidopsis* mRNA decay landscape arises from specialized RNA decay substrates, decapping-mediated feedback, and redundancy. *Proc. Natl. Acad. Sci. U.S.A.* **115**.
- Sprunck, S. and Groß-Hardt, R.** (2011). Nuclear behavior, cell polarity, and cell specification in the female gametophyte. *Sex Plant Reprod* **24**: 123–136.
- Stone, S.L., Braybrook, S.A., Paula, S.L., Kwong, L.W., Meuser, J., Pelletier, J., Hsieh, T.-F., Fischer, R.L., Goldberg, R.B., and Harada, J.J.** (2008). *Arabidopsis* LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proceedings of the National Academy of Sciences* **105**: 3151–3156.
- Stone, S.L., Kwong, L.W., Yee, K.M., Pelletier, J., Lepiniec, L., Fischer, R.L., Goldberg, R.B., and Harada, J.J.** (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proceedings of the National Academy of Sciences* **98**: 11806–11811.
- Suter, D.M.** (2020). Transcription Factors and DNA Play Hide and Seek. *Trends in Cell Biology* **30**: 491–500.
- Suzuki, M., Wang, H.H.-Y., and McCarty, D.R.** (2007). Repression of the *LEAFY COTYLEDON 1/B3* Regulatory Network in Plant Embryo Development by *VP1 / ABSCISIC ACID INSENSITIVE 3 - LIKE* B3 Genes. *Plant Physiology* **143**: 902–911.
- Swaminathan, K., Peterson, K., and Jack, T.** (2008). The plant B3 superfamily. *Trends in Plant Science* **13**: 647–655.
- Szakonyi, D. and Duque, P.** (2018). Alternative Splicing as a Regulator of Early Plant Development. *Front. Plant Sci.* **9**: 1174.
- Takahashi, S., Sakamoto, A., Sato, S., Kato, T., Tabata, S., and Tanaka, A.** (2005). Roles of *Arabidopsis AtREV1* and *AtREV7* in Translesion Synthesis. *Plant Physiology* **138**: 870–881.
- Tao, Z., Hu, H., Luo, X., Jia, B., Du, J., and He, Y.** (2019). Embryonic resetting of the parental vernalized state by two B3 domain transcription factors in *Arabidopsis*. *Nature Plants* **5**: 424–435.
- Tao, Z., Shen, L., Gu, X., Wang, Y., Yu, H., and He, Y.** (2017). Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. *Nature* **551**: 124–128.
- Thévenin, J. et al.** (2012). A new system for fast and quantitative analysis of heterologous gene expression in plants. *New Phytol* **193**: 504–512.
- To, A., Joubès, J., Thueux, J., Kazaz, S., Lepiniec, L., and Baud, S.** (2020). AtMYB92 enhances fatty acid synthesis and suberin deposition in leaves of *Nicotiana benthamiana*. *Plant J* **103**: 660–676.
- To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J., and Parcy, F.** (2006). A Network of Local and Redundant Gene Regulation Governs *Arabidopsis* Seed Maturation. *the Plant Cell Online* **18**: 1642–1651.
- Tsutsui, H. and Higashiyama, T.** (2016). pKAMA-ITACHI Vectors for Highly Efficient CRISPR/Cas9-Mediated Gene Knockout in *Arabidopsis thaliana*. *Plant Cell Physiol: pcw191*.
- Turchi, L., Carabelli, M., Ruzza, V., Possenti, M., Sassi, M., Peñalosa, A., Sessa, G., Salvi, S., Forte, V., Morelli, G., and Ruberti, I.** (2013). *Arabidopsis* HD-Zip II transcription factors control apical embryo development and meristem function. *Development* **140**: 2118–2129.
- Twell, D.** (2011). Male gametogenesis and germline specification in flowering plants. *Sex Plant Reprod* **24**: 149–160.
- Vanhercke, T., El Tahchy, A., Shrestha, P., Zhou, X.-R., Singh, S.P., and Petrie, J.R.** (2013). Synergistic effect of WR11 and DGAT1 coexpression on triacylglycerol biosynthesis in plants. *FEBS Letters* **587**: 364–369.
- Varadi, M. et al.** (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**: D439–D444.
- Vazquez-Vilar, M., Bernabé-Orts, J.M., Fernandez-del-Carmen, A., Ziarsolo, P., Blanca, J., Granell, A., and Orzaez, D.** (2016). A modular toolbox for gRNA–Cas9 genome engineering in plants based on the GoldenBraid standard. *Plant Methods* **12**: 10.

- Veillet, F., Perrot, L., Guyon-Debast, A., Kermarrec, M.P., Chauvin, L., Chauvin, J.E., Gallois, J.L., Mazier, M., and Nogué, F.** (2020). Expanding the CRISPR toolbox in *P. Patens* using SpCas9-NG variant and application for gene and base editing in solanaceae crops. *International Journal of Molecular Sciences* **21**.
- Verma, S., Attuluri, V.P.S., and Robert, H.S.** (2022). Transcriptional control of Arabidopsis seed development.: 19.
- Vu, L.D., Gevaert, K., and De Smet, I.** (2018). Protein Language: Post-Translational Modifications Talking to Each Other. *Trends in Plant Science* **23**: 1068–1080.
- Waese, J. et al.** (2017). ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. *Plant Cell* **29**: 1806–1821.
- Wang, F.-X., Shang, G.-D., Wu, L.-Y., Xu, Z.-G., Zhao, X.-Y., and Wang, J.-W.** (2020). Chromatin Accessibility Dynamics and a Hierarchical Transcriptional Regulatory Network Structure for Plant Somatic Embryogenesis. *Developmental Cell* **54**: 742-757.e8.
- Wang, H., Guo, J., Lambert, K.N., and Lin, Y.** (2007). Developmental control of Arabidopsis seed oil biosynthesis. *Planta* **226**: 773–783.
- Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S.** (2004). Hierarchical Recruitment of Polycomb Group Silencing Complexes. *Molecular Cell* **14**: 637–646.
- Warde-Farley, D. et al.** (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research* **38**: W214–W220.
- Wiedenheft, B., Sternberg, S.H., and Doudna, J.A.** (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331–338.
- Wójcik, A.M., Wójcikowska, B., and Gaj, M.D.** (2020). Current Perspectives on the Auxin-Mediated Genetic Network that Controls the Induction of Somatic Embryogenesis in Plants. *International Journal of Molecular Sciences* **21**: 1333.
- Wójcikowska, B., Jaskóła, K., Gąsior, P., Meus, M., Nowak, K., and Gaj, M.D.** (2013). LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of Arabidopsis, via YUCCA-mediated auxin biosynthesis. *Planta* **238**: 425–440.
- Xu, C., Luo, F., and Hochholdinger, F.** (2016a). LOB Domain Proteins: Beyond Lateral Organ Boundaries. *Trends in Plant Science* **21**: 159–167.
- Xu, W., Fiume, E., Coen, O., Pechoux, C., Lepiniec, L., and Magnani, E.** (2016b). Endosperm and Nucellus Develop Antagonistically in Arabidopsis Seeds. *Plant Cell* **28**: 1343–1360.
- Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S., and Hattori, T.** (2009). Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *The Plant Journal* **58**: 843–856.
- Yamasaki, K. et al.** (2004). Solution Structure of the B3 DNA Binding Domain of the Arabidopsis Cold-Responsive Transcription Factor RAV1[W]. *The Plant Cell* **16**: 3448–3459.
- Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., and Yokoyama, S.** (2013). DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends in Plant Science* **18**: 267–276.
- Yan, D., Duermeyer, L., Leoveanu, C., and Nambara, E.** (2014). The Functions of the Endosperm During Seed Germination. *Plant and Cell Physiology* **55**: 1521–1533.
- Yanagisawa, S.** (2001). The Transcriptional Activation Domain of the Plant-Specific Dof1 Factor Functions in Plant, Animal, and Yeast Cells. *Plant and Cell Physiology* **42**: 813–822.
- Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M.** (2013). VAL- and AtBM11-Mediated H2Aub Initiate the Switch from Embryonic to Postgerminative Growth in Arabidopsis. *Current Biology* **23**: 1324–1329.
- Yella, V.R. and Bansal, M.** (2017). DNA structural features of eukaryotic TATA-containing and TATA-less promoters. *FEBS Open Bio*: 11.
- Zaret, K.S.** (2018). Pioneering the chromatin landscape. *Nature Genetics* **50**: 167–169.
- Zhang, X., Garretton, V., and Chua, N.H.** (2005). The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation. *Genes and Development* **19**: 1532–1543.

- Zhang, Y., Lu, Y., El Sayyed, H., Bian, J., Lin, J., and Li, X.** (2022). Transcription factor dynamics in plants: Insights and technologies for in vivo imaging. *Plant Physiology* **189**: 23–36.
- Zhang, Y., Malzahn, A.A., Sretenovic, S., and Qi, Y.** (2019). The emerging and uncultivated potential of CRISPR technology in plant science. *Nat. Plants* **5**: 778–794.
- Zhang, Z. and Laux, T.** (2011). The asymmetric division of the *Arabidopsis* zygote: from cell polarity to an embryo axis. *Sex Plant Reprod* **24**: 161–169.
- Zhang, Z., Tucker, E., Hermann, M., and Laux, T.** (2017). A Molecular Framework for the Embryonic Initiation of Shoot Meristem Stem Cells. *Developmental Cell* **40**: 264-277.e4.
- Zheng, W., Zhang, X., Yang, Z., Wu, J., Li, F., Duan, L., Liu, C., Lu, L., Zhang, C., and Li, F.** (2014). AtWuschel Promotes Formation of the Embryogenic Callus in *Gossypium hirsutum*. *PLoS ONE* **9**: e87502.
- Zheng, Y., Ren, N., Wang, H., Stromberg, A.J., and Perry, S.E.** (2009). Global Identification of Targets of the *Arabidopsis* MADS Domain Protein AGAMOUS-Like15. *The Plant Cell* **21**: 2563–2577.
- Zhu, Q., Dabi, T., and Lamb, C.** (1995). TATA box and initiator functions in the accurate transcription of a plant minimal promoter in vitro. *Plant Cell* **7**: 1681–1689.
- Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.L., Wang, D., and Gao, C.** (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nature Biotechnology* **35**: 438–440.

Annexes

Annexes

Annexe 1 : Séquences des ARN guides utilisés pour la mutagenèse CRISPR-X

Annexe 2 : Amorces pour le génotypage des plantes

Annexe 3 : Autres amorces utilisées pour les clonages ou la vérification des clonages et des plantes transformées

Annexe 4 : Liste des variants et fragments de LEC2 clonés

Annexe 5 : Séquences des 60 orthologues de AtLEC2 utilisés pour la prédiction et la conservation des domaines d'activation de la transcription

Annexe 6 : Exemples de traces d'ATAC-seq et CHIP-seq obtenues pour plusieurs gènes et représentatives des profils pouvant être observés

Annexe 7 : Milieux et solutions utilisés pour la culture et la transformation des mousses

Annexe 1 : Séquences des ARN guides utilisés pour la mutagenèse CRISPR-X

Nom	Position	Séquence des oligonucléotides	Off-targets possibles
ARN guide A	LEC2*SH 89 rev	Sens : ATTGTGATCGTAAGTAGGCACGG Anti-sens : AAACCCGTGCCTACTTACGATCA	guide: ATGATCGTAAGTAGGCACGG TGG off-target: AAGTTCGTGAGTATGCACCG CGG * * * * CFD Off-target score: 0.095726 MIT Off-target score: 0.14 Position: Chr5:25927443-25927465:-
ARN guide B	LEC2*SH 307 for	Sens : ATTGGGGAGGATTCTCGATCCA Anti-sens : AAACTGGATCGAGGAATCCTCCC	guide: GGGGAGGATTCTCGATCCA AGG off-target: ATGGAGGATTCTCGAGCCA AGG ** * * CFD Off-target score: 0.020732 MIT Off-target score: 0.34 Position: Chr1:25589684-25589706:+ Distance from target: 15.693 Mbp
ARN guide C	LEC2*SH 471 rev	Sens : ATTGGTTATCGAAGGTAGACTGC Anti-sens : AAACGCAGTCTACCTTCGATAAC	guide: TGTTATCGAAGGTAGACTGC TGG off-target: TGTTATCAAAGGTA AACGC AGG * * * * CFD Off-target score: 0.188235 MIT Off-target score: 0.03 Position: Chr1:8700311-8700333:- Distance from target: 1.197 Mbp guide: TGTTATCGAAGGTAGACTGC TGG off-target: TGTTATCAAATGTGACTGT TGG * * * * CFD Off-target score: 0.053846 MIT Off-target score: 0.05 Position: Chr1:9899346-9899368:+ Distance from target: 0.002 Mbp
ARN guide D	LEC2*SH 759 rev	Sens : ATTGCTCATCCTCGTAGATGGTG Anti-sens : AAACCACCATCTACGAGGATGAG	guide: ACTCATCCTCGTAGATGGTG AGG off-target: TCTCATCCCGAAGATGGTT AGG * * * * CFD Off-target score: 0.346667 MIT Off-target score: 0.21 Position: Chr4:5299397-5299419:-
ARN guide E	LEC2*SH 895 for	Sens : ATTGTATGCTCGATTACATCCCA Anti-sens : AAACTGGGATGTAATCGAGCATA	guide: CTATGCTCGATTACATCCCA AGG off-target: CAATGCTTGATTACATACCA AGA * * * CFD Off-target score: 0.023994 MIT Off-target score: 0.24 Position: Chr1:9897249-9897271:- Distance from target: 0.000 Mbp
ARN guide F	LEC2*SH 1021 rev	Sens : ATTGCGTAAGCATGATCCTCTGG Anti-sens : AAACCCAGAGGATCATGCTTACG	guide: ACGTAAGCATGATCCTCTGG AGG off-target: ACGTACGGTGATCCTCAGG TGG * * * CFD Off-target score: 0.202020 MIT Off-target score: 0.22 Position: Chr1:9897103-9897125:+ Distance from target: 0.000 Mbp

Annexe 2 : Amorces pour le géotypage des plantes

Mutant	Sauvage/mutant	Amorce sens	Amorce antisens
<i>lec2-1</i>	Sauvage	(16) LEC2 up : CCGGGAAAAATGGATAACTT	LEC2_1475-Rev : CGAGGTAAAGTGGTAAACC
	Mutant	<i>lec2-1_3246</i> For : FCCTCATCTAAAGATTTGAGC	LEC2_1475-Rev : CGAGGTAAAGTGGTAAACC
<i>abi3-11</i>	Sauvage et mutant	ABI3.F2 : GATATGATGGAGACTTTTCGG	ABI3.R2 : CGGTGGAGATTCAAGTAAAG
Témoin de PCR	Sauvage et mutant	EF-Q1 : CTGGAGGTTTTGAGGCTGGTA	EF-Q2 : GCAAGGGTGAAAGCAAGAAGA

Annexe 3 : Autres amorces utilisées pour les clonages ou la vérification des clonages et des plantes transformées

Nom de l'amorce	Séquence (5'-3')	But
Bordures des vecteurs		
seq_LA	TCGCGTTAACGCTAGCATGGATCTC	Bordure vecteur Gateway (BP)
seq_LB	GTAACATCAGAGATTTTGAGACAC	Bordure vecteur Gateway (BP)
pUPD.F	GAATTGGGCCCAGCTCGCATGCTC	Bordure vecteur pUPD
pUPD.R	AGGTGACACTATAGAATACTCAAGC	Bordure vecteur pUPD
pUPD2 Fwd	GCCGCTTCTAGAGCGATG	Bordure vecteur pUPD2
pUPD2 Rev	CAGGGTGGTGACACCTTGCC	Bordure vecteur pUPD2
pGBD3F	TTGTGCCGAGCTGCCGGTCCG	Bordure vecteur pDGB3_α/Ω
pGBD3R	CCTTTTTCACGCCCTTTTAAATATCCG	Bordure vecteur pDGB3_α/Ω
PhyscoVectorA F2	TTGTGTGGAATTGTGAGCGG	Bordure vecteur pBS Tpp-A
pGWB2_FWD	CATGATTACGCCAAGCTTGC	Bordure vecteur pGWB2
pDest22_F	GATGAAGATACCCACCAA	Bordure vecteur pDEST22
pDest32_F	AACCGAAGTGCGCCAAGTGTCTG	Bordure vecteur pDEST32
M13-UP	GTA AACGACG GCCAGT	Bordure site de clonage multiple
M13-RP	GGAAACAGCTATGACCATG	Bordure site de clonage multiple
T3	ATTAACCCCTCACTAAAGGGA	Bordure site de clonage multiple
T7	GTAATACGACTCACTATAGGG	Bordure site de clonage multiple
Promoteurs		
LG1435_prUBQ10_1	TATTGCTTCACG CCTTAGC	Vérification pAtUbq10 - sens
prNOS 75 FW	ACGTTTGGAACTGACAGAACCG	Vérification pNOS – sens
prNOS 75 RV	CGGTTCTGTCAGTTCCAAACGT	Vérification pNOS – antisens
LG984pCMV_For199	GAAGAACCAGGCGAAGAAAA	Vérification pCsVMV – sens
LG985pCMV_Rev435	TGCCGA ACTTGGTTCTTAT	Vérification pCsVMV – antisens
LG1255promRPS5A_	TTGGAGTCAAAATCGGTTCC	Vérification pRPS5A – sens
LG1256promRPS5A_	GGCAGTGTCAAATGCAGAAA	Vérification pRPS5A – antisens
pLEC2-500-F-GBS1	GCGCCGTCTCGTCTCGGGAGCTAAAGG TACAAAGATTCTGTTTTTTC	Domestication GoldenBraid pLEC2 -500 pb, préfixe S1
pLEC2R-GBS2	GCGCCGTCTCGTCTCACATTTTTTCCC GGAGAGAGAGAGAGAGAGGGGAGAG AGATC	Domestication GoldenBraid pLEC2, suffixe S2

pAGL15_AttB1_1,3	GGGGACAAGTTTGTACAAAAAAGCAG GCTCAGATCCAGTTTTATAAGGGTGT TTGGAGGG	Amplification pAGL15 -1,3 kb, ajout AttB1
pAGL15_AttB2_Rev	GGGGACCACTTTGTACAAGAAAGCTG GGTTTTTCCCCAAAATTGAACAGAG AGAAGAAGG	Amplification pAGL15, ajout AttB2
pAGL15_370_For	CCGGTGTTACAAAGCTGCTACTAAAG TGTCG	Vérification pAGL15 – sens
pAGL15_370_Rev	CGACACTTTAGTAGCAGCTTTGTAAC ACCGG	Vérification pAGL15 – antisens
pYUC5_AttB1_For	GGGGACAAGTTTGTACAAAAAAGCAG GCTTATAGGTTTACTCTAAATAGCTC AAAAATAATAATTAAGCC	Amplification pYUC5 -1 kb, ajout AttB1
pYUC5_AttB2_Rev	GGGGACCACTTTGTACAAGAAAGCTG GGTCTTTAGGGGTGAGTTTGATCGAA TCAAGATTACAAGTAGTAAAATGG	Amplification pYUC5, ajout AttB2
pYUC5_550_For	CGGACTCTAATCAAAGTCCCATCAGA GACTTTGG	Vérification pYUC5 – sens
pYUC5_550_Rev	CCAAAGTCTCTGATGGGACTTTGATT AGAGTCCG	Vérification pYUC15 – antisens
pLBD40_AttB1_For	GGGGACAAGTTTGTACAAAAAAGCAG GCTATACTAGAGCCTATAGAAGTCAA AGTGATTCC	Amplification pLBD40 -1 kb, ajout AttB1
pLBD40_AttB2_Rev	GGGGACCACTTTGTACAAGAAAGCTG GGTTTTAGATGGGAAACAAAAACAA AATGTGTTTTGGGTTTGAGGG	Amplification pLBD40, ajout AttB2
pIAA30mBsmblAttB1	GGGGACAAGTTTGTACAAAAAAGCAG GCTGAGAGTTAATGACGTACATGCGC CGTGTTTAGCAG	Amplification pIAA30 -1 kb, ajout AttB1, mutation site BsmBI
pIAA30_AttB2_Rev	GGGGACCACTTTGTACAAGAAAGCTG GGTTTTTTTTTATTTCTTTTACTATT TCTCTCTCTATGATTTTATCTAAA ACAGGTG	Amplification pIAA30, ajout AttB2
pIAA30_640_For	CGAAAGGTAGTACTAGGAGTCCG	Vérification pIAA30 – sens
pIAA30_640_Rev	CGACTCCTAGTACTACCTTTCCG	Vérification pIAA30 – antisens
Protéines fluorescentes et sélection antibiotique		
GFP_F1	CGCCATGCCCGAAGGCTACG	Vérification GFP – sens
GFP_R1	CTTGTGCCCCAGGATGTTGC	Vérification GFP – antisens
LG2263_DsRed_607	TCCCACAACGAGGACTACAC	Vérification DsRed – antisens
Hygro-60-RV	CGCTGTGCGAACTTTTCGATC	Vérification Hygro – antisens
Hygro-500-FW	TGATCCCCATGTGTATCACTGGC	Vérification Hygro – sens
CRISPR		
MCP-AID270For	GGTGTAGAACTCCCCGTCGAGCATG G	Vérification MCP:AID – sens
MCP-AID750Rev	CCACGTGACCCTGTAACATCTGCCC	Vérification MCP:AID – antisens
LG1877deadCAS9_3	GGGAGCAGGCAGAAAACATT	Vérification dCas9 – sens
nCas9_GBS2-For	GCGCCGTCTCGCTCGAATGGATAAGA AGTACTCTATCGG	Domestication GoldenBraid nCas9, préfixe S2
nCas9_GBS3-Rev	GCGCCGTCTCGCTCAGGCTGCCTTAT CATCATCATCCTTGATGTCG	Domestication GoldenBraid nCas9, suffixe S3
LG1017_nCAS9_PmC	GTGAGTCCGAGAGAGAGAGC	Vérification nCas9 – antisens
nCas9_4080For	ATCGATCTCTCACAGCTCGG	Vérification nCas9 – sens

REV1-GBS3-For	GCGCCGTCTCGCTCGAGCCTCAGG	Domestication GoldenBraid, REV1, préfixe S3
REV1-GBS7-Rev	GCGCCGTCTCGCTCAAAGCTATCAGG GCCC	Domestication GoldenBraid, REV1, suffixe S7
Rev1_330Rev	CAGTGAAACCGTCCACCC	Vérification REV1 – antisens
Rev1_3040For	GGACTGGTACACCACCTAAGTG	Vérification REV1 – sens
Pol1_330Rev	CAATGACATCATCAGCCTCAACAC	Vérification Pol13M – antisens
Pol1_2365For	GGATCTGTATTTGAGCGTTACCAC	Vérification Pol13M – sens
LEC2		
(6) LEC2 UF	CTCTCTCTCTCTCCGGGAAA	Vérification LEC2 – sens
(16) LEC2 up	CCGGGAAAAATGGATAACTT	Vérification LEC2 – sens
(20) LEC2 372 lo	CTCATCATGGCGTTTTTCCT	Vérification LEC2 – antisens
(21) LEC2 up2	GCTTAGGGTTTTGTGTGAGA	Vérification LEC2 – sens
(23) LEC2 lo2	CCTCCCGAGTGACCCAAC	Vérification LEC2 – antisens
(24) LEC2 836 lo	TAACAGAAAAAGAAGAAAAATCA	Vérification LEC2 – antisens
(27) LEC2 MF	GCAGAAGCAAATCTTCCGAAGCTAT	Vérification LEC2 – sens
(28) LEC2 lo 645	TGAAAGACCAAGACTGCATA	Vérification LEC2 – antisens
(29) LEC2 MR	TTGTTAAAAAGTCTCCTATCTCAGCT CCAT	Vérification LEC2 – antisens
(30) LEC2 814F	TCGGGAAAACAAAATGAAGG	Vérification LEC2 – sens
(35) LEC2 low	CATATCACCACCACTCAAAG	Vérification LEC2 – antisens
LEC2edite270for	GGATCTGAGTTCGATCTCTCGTGGG	Vérification LEC2*SH - sens
LEC2edite750rev	GGTGTCTCGAGCACGTACATCC	Vérification LEC2*SH - antisens
LEC2edite370For	TCTCCTAACTCTTACCTTCTGAG	RT-PCR LEC2*SH - sens
LEC2edite518Rev	CCACATCAGAGTTCTTGAGTTCTT	RT-PCR LEC2*SH - antisens
LEC2 MF	GCAGAAGCAAATCTTCCGAAGCTAT	RT-PCR AtLEC2 - sens
LEC2 MR	TTGTTAAAAAGTCTCCTATCTCAGCT CCAT	RT-PCR AtLEC2 - antisens
Termineurs		
TerNOS_RV1	CATAGATGACACCGCGCGC	Vérification tNOS – antisens
LG1260termRBCSE9	GCGCCGTCTCGCTCGGTTGCTTTTCG TTCGTATCATCGG	Vérification et domestication GoldenBraid tRbcSE9, préfixe S1
LG1262termRBCSE9	TTCGATAGCGAAAACCGAAT	Vérification tRbcSE9 – antisens
35S ter Rev2	GCTCAACACATGAGCGAAAC	Vérification t35S – antisens
Linkers GoldenBraid 2.0		
LinkS3S4ligBSA1U	AGCCGCTGGTGCTGCTGCGGCCGCTG GGGCCGG	Linker GoldenBraid S3-S4 – sens
LinkS3S4ligBSA1L	CGAACCGGCCCCAGCGGCCGCAGCAG CACCAG	Linker GoldenBraid S3-S4 – antisens
LinkS5S6ligBSA1U	GCAGGCTGGTGCTGCTGCGGCCGCTG GGGCCGT	Linker GoldenBraid S5-S6 – sens
LinkS5S6ligBSA1L	TACCACGGGCCCCAGCGGCCGCAGCAG CACCAGC	Linker GoldenBraid S5-S6 – antisens
LinkerS6DUMMYFor	GGTATAGCCTCTACCAAAG	Linker GoldenBraid S6-S7 – sens
DUMMY-S5S7-UP	GCAGTAGCCTCTACCAAAG	Linker GoldenBraid S5-S7 – sens
DUMMY-S5S7-Down	AAGCCTTTGGTAGAGGCTA	Linker GoldenBraid S5-S7 – antisens

Annexe 4 : Liste des variants et fragments de LEC2 clonés

Variant	Amorces utilisées (5'-3')	Vecteur(s)	But	Référence texte
LEC2*SH (C. Boulard)	<p>Lec2editforwB1 : GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATG GATAACTTCCCTCCCTTTCCCATCTTCTAACGC</p> <p>LEC2editeB5for : GGGGACAAC TTTGTATACAAAAGTTGTAATGGAT AACTTCCCTCCCTTTCCCATCTTC</p> <p>LEC2editeB2rev : GGGGACCACTTTGTACAAGAAAGCTGGGTTTTAG TGGTGATGGTGATGATGTCC</p> <p>HISrevB2 : GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAG TGGTGATGGTGATGATGTCCACC</p>	pDONR207, pDONR221, pBS Tpp-A, pGWB2, pΔGR	Surexpression d'une version éditée de LEC2 <i>in planta</i> et test de fonctionnalité dans les protoplastes de mousse.	Chapitre 1 - Génération et validation des surexpresseurs de LEC2
ΔLEC2 ₁₋₃₅	<p>LEC2F106 (short) : GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGA TCATCATCAGGCTCAGCCT</p> <p>AttB2-stopNLEC2 : GGGGACCACTTTGTACAAGAAAGCTGGGTTCCACC ACCACTCAAAG</p> <p>LEC2-106F-GBS4 : GCGCCGTCTCGCTCGTTTCGATGGATCATCATCAG GCTCAGCCTCATCACTTCTTGCCCTCCG</p> <p>LEC2-Rev-GBS5 : GCGCCGTCTCGCTCACTGCacCCACCACTCAAAG TCGTTAAAGCTCACCTGATCATCGGATGAACCC</p>	pDONR207, pBS Tpp-A, pUPD2, pDGB3_α1, pDGB3_Ω1	Délétion en N-ter de LEC2. Expression transitoire dans les protoplastes de mousse et <i>in planta</i> .	Chapitre 2 - Introduction, Localisation de variants de LEC2 in planta, L'absence de DA réduit l'activité transcriptionnelle de AtLEC2 dans la mousse, Fonctionnalité de LEC2 sans domaines d'activation in planta
ΔLEC2 ₁₋₃₈	<p>LEC2_115F_AttB1 : GGGGACAAGTTTGTACAAAAAAGCAGGCTATGCA GGCTCAGCCTCATCACTTCTTGCCCTCCGTTTTCA TACCCGG</p> <p>AttB2-stopNLEC2 : GGGGACCACTTTGTACAAGAAAGCTGGGTTCCACC ACCACTCAAAG</p>			

	<p>LEC2-115F-GBS4 : GCGCCGTCTCGCTCGTTCGATGCAGGCTCAGCCT CATCACTTCTTGCCTCCGTTTTTCATACCCGG</p> <p>LEC2-Rev-GBS5 : GCGCCGTCTCGCTCACTGCacCCACCACTCAAAG TCGTTAAAGCTCACCTGATCATCGGATGAACCC</p>			
Δ LEC2 ₁₋₄₁	<p>LEC2_127F_AttB1 : GGGGACAAGTTTGTACAAAAAAGCAGGCTATGCA TCACTTCTTGCCTCCGTTTTTCATACCCGGTGG</p> <p>AttB2-stopNLEC2 : GGGGACCACTTTGTACAAGAAAGCTGGGTTCAAC ACCACTCAAAG</p> <p>LEC2-127F-GBS4 : GCGCCGTCTCGCTCGTTCGATGCATCACTTCTTG CCTCCGTTTTTCATACCCGGTGG</p> <p>LEC2-Rev-GBS5 : GCGCCGTCTCGCTCACTGCacCCACCACTCAAAG TCGTTAAAGCTCACCTGATCATCGGATGAACCC</p>			
Δ LEC2 ₁₋₄₅ * *Erreur de phase dans pDGB3_α1, utilisé comme témoin négatif <i>pLEC2::GFP</i>	<p>LEC2_136F_AttB1 : GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTT GCCTCCGTTTTTCATACCCGGTGGAGCAGATGGCG GCGG</p> <p>AttB2-stopNLEC2 : GGGGACCACTTTGTACAAGAAAGCTGGGTTCAAC ACCACTCAAAG</p> <p>LEC2-136F-GBS4 : GCGCCGTCTCGCTCGTTCGATGTTGCCTCCGTTT TCATACCCGGTGGAGCAGATGGCGGCGG</p> <p>LEC2-Rev-GBS5 : GCGCCGTCTCGCTCACTGCacCCACCACTCAAAG TCGTTAAAGCTCACCTGATCATCGGATGAACCC</p>			

<p>ΔLEC2₁₋₄₈</p>	<p>LEC2cr3f : GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATG TCATACCCGGTGGAGCAGATGGCG</p> <p>AttB2-stopNLEC2 : GGGGACCACTTTGTACAAGAAAGCTGGGTTCCACC ACCACTCAAAG</p> <p>LEC2-CR3F-GBS4 : GCGCCGTCTCGCTCGTTCGATGTCATACCCGGTG GAGCAGATGGCG</p> <p>LEC2-Rev-GBS5 : GCGCCGTCTCGTCACTGCacCCACCACTCAAAG TCGTTAAAGCTCACCTGATCATCGGATGAACCC</p>			
<p>ΔLEC2₁₋₆₄</p>	<p>LEC2cr6f : GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTA CTTATCGGAGTGTATCCTCAGATCCCGTTACG CAAACCGGAAGTGAATTTCGGTTCTCTGGTTGGTA ATCCT</p> <p>AttB2-stopNLEC2 : GGGGACCACTTTGTACAAGAAAGCTGGGTTCCACC ACCACTCAAAG</p> <p>LEC2-CR6F-GBS4 : GCGCCGTCTCGCTCGTTCGATGTACTTATCGGAG TGTTATCCTCAGATCCCGG</p> <p>LEC2-Rev-GBS5 : GCGCCGTCTCGTCACTGCacCCACCACTCAAAG TCGTTAAAGCTCACCTGATCATCGGATGAACCC</p>			
<p>ΔLEC2₁₋₈₅ / LEC2Δ_{AD N-ter}</p>	<p>LEC2cr1fGW : GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATG GTTGGTAATCCTTGTGGTGG</p> <p>AttB2-stopNLEC2 : GGGGACCACTTTGTACAAGAAAGCTGGGTTCCACC ACCACTCAAAG</p>			

	<p>LEC2-CR1F-GBS4 : GCGCCGTCTCGCTCGTTTCGATGGTTGGTAATCCT TGTTTGTGG</p> <p>LEC2-Rev-GBS5 : GCGCCGTCTCGCTCACTGCacCCACCACTCAAAG TCGTTAAAGCTCACCTGATCATCGGATGAACCC</p>			
Δ LEC2 ₃₄₈₋₃₆₃	<p>AttB1-NNLEC2 : GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATG GATAACTTCTTACC</p> <p>LEC2cr4R : GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAC GCGTGATCCTCAGGTGGCATTGATGAGG</p> <p>LEC2-ATGF-GBS4 : GCGCCGTCTCGCTCGTTTCGATGGATAACTTCTTA CCCTTTCCCTCTTCTAACGC</p> <p>LEC2-CR4R-GBS5 : GCGCCGTCTCGCTCACTGCACGTACGCGTGATCC TCAGGTGGCATTGATGAGG</p>	pDONR207, pBS Tpp-A, pUPD2, pDGB3_α1, pDGB3_Ω1	Délétion en C-ter de LEC2. Expression transitoire dans les protoplastes de mousse et <i>in</i> <i>planta</i> .	
Δ LEC2 ₃₃₄₋₃₆₃	<p>AttB1-NNLEC2 : GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATG GATAACTTCTTACC</p> <p>LEC2cr1rGW : GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAA TGGTGCTGAAGGTCAGTGG</p> <p>LEC2-ATGF-GBS4 : GCGCCGTCTCGCTCGTTTCGATGGATAACTTCTTA CCCTTTCCCTCTTCTAACGC</p> <p>LEC2-CR1R-GBS5 : GCGCCGTCTCGCTCACTGCACATGGTGCTGAAGG TCAGTGGTGAGGTC</p>			
Δ LEC2 ₂₇₅₋₃₆₃ / LEC2 Δ _{AD C-ter}	<p>AttB1-NNLEC2 : GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATG GATAACTTCTTACC</p>			

	<p>LEC2-Q275STOP-AttB2R : GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAT TTTCCCGAATTTCCATTCAT</p> <p>LEC2-ATGF-GBS4 : GCGCCGTCTCGCTCGTTCGATGGATAACTTCTTA CCCTTTCCCTCTTCTAACGC</p> <p>LEC2-Q275STOP-GBS7 : GCGCCGTCTCGCTCAAAGCTCATTTTCCCGAATT TCCATTCAT</p>			
LEC2 M104-Q275 / LEC2 Δ _{AD N et C-ter}	<p>LEC2-M104-AttB1F : GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATG ACGAAGATGGCAAGGATC</p> <p>LEC2-Q275STOP-AttB2R : GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAT TTTCCCGAATTTCCATTCAT</p> <p>LEC2_M104-GBS3F : GCGCCGTCTCGCTCGAGCCATGACGAAGATGGCA AGGATC</p> <p>LEC2-Q275STOP-GBS7 : GCGCCGTCTCGCTCAAAGCTCATTTTCCCGAATT TCCATTCAT</p>	pDONR207, pBS Tpp-A, pDEST32, pUPD2, pDGB3_α1, pDGB3_Ω1	Délétion en N et C-ter de LEC2. Expression transitoire dans les protoplastes de mousse, dans la levure et <i>in planta</i> .	Chapitre 2 - Localisation de variants de LEC2 in planta , Mise en évidence des domaines d'activation de la transcription de LEC2 , Fonctionnalité de LEC2 sans domaines d'activation in planta
LEC2 ATG-M104	<p>LEC2M104STOPAttB2 : GGGGACCACTTTGTACAAGAAAGCTGGGTCCCTAA CGCGGATCAAGAAAACCACCTCTCTCTTGCC</p> <p>LEC2-ATGF-GBS4 : GCGCCGTCTCGCTCGTTCGATGGATAACTTCTTA CCCTTTCCCTCTTCTAACGC</p> <p>LEC2_M104STOP_GBS7 : GCGCCGTCTCGCTCAAAGCCTAACGCGGATCAAG AAAACCACCTCTCTCTTGCC</p>	pDONR207, pDEST32, pUPD2, pDGB3_α1	Fragments de LEC2 pour tester la présence de NLS ou la fonctionnalité du domaine d'activation de la transcription. Expression dans la levure et transitoire dans le tabac.	Chapitre 2 - Validation expérimentale de la prédiction du NLS de LEC2 , Validation expérimentale de la prédiction des DA de AtLEC2

LEC2 M104-M216	LEC2_M104-GBS3F : GCGCCGTCTCGCTCGAGCCATGACGAAGATGGCA AGGATC LEC2_M216STOP_GBS7 : GCGCCGTCTCGCTCAAAGCTCAAGAGAAAACATC TCTCATCTG	pUPD2, pDGB3_α1	Fragments de LEC2 pour tester la présence de NLS. Expression transitoire dans le tabac.	Chapitre 2 - Validation expérimentale de la prédiction du NLS de LEC2
LEC2 M216-STOP	LEC2-M216-GBS3F GCGCCGTCTCGCTCGAGCCATGCAGTCTTGGTCT TTCAAATACAAG LEC2-Rev-GBS7R : GCGCCGTCTCGCTCAAAGCTCACCACTCAA GTCGTTAAAGC			
LEC2 G240-STOP	LEC2-G240ATG-AttB1 : GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATG GAATTTGTGAAGCAAATGGAGCTGAG AttB2-stopNLEC2 : GGGGACCACTTTGTACAAGAAAGCTGGGTTCACC ACCACTCAAAG			
LEC2 Q275-STOP	LEC2-Q275ATG-AttB1F : GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATG AATGAAGGAAGAGAAAATGAGTCGAGGG AttB2-stopNLEC2 : GGGGACCACTTTGTACAAGAAAGCTGGGTTCACC ACCACTCAAAG	pDONR207, pDEST32	Fragments de LEC2 pour tester la fonctionnalité des domaines d'activation de la transcription.	Chapitre 2 - Validation expérimentale de la prédiction des DA de AtLEC2
LEC2 G240-Q275	LEC2-G240ATG-AttB1 : GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATG GAATTTGTGAAGCAAATGGAGCTGAG LEC2-Q275STOP-AttB2R : GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAT TTTCCCGAATTTCCATTCAT			
LEC2 Y _{257A} Y _{265A}	LEC2-Y257A-Y265A : GGGAGACTTTTTAACAATAGCCGAGGACGAAAGCA AGAATCTCGCCTTCGCCATGAATGGAAATTCGG LEC2-Y257A-Y265r :	pDONR207, pBS TPP-A, pDEST32, pUPD2,	Effet de la mutation Y _{257A} Y _{265A} sur l'activité transcriptionnelle de LEC2, dans les protoplastes de mousse, dans la levure et <i>in planta</i> .	Chapitre 2 - Cas particulier : le variant LEC2 Y257A Y265A

	<p>CCGAATTTCCATTCATGGCGAAGGCGAGATTCTT GCTTTCGTCCCTCGGCTATTGTTAAAAAGTCTCC</p> <p>LEC2-ATGF-GBS4 : GCGCCGTCTCGCTCGTTCGATGGATAACTTCTTA CCCTTTCCTCTTCTAACGC</p> <p>LEC2-Rev-GBS5 : GCGCCGTCTCGCTCACTGCacCCACCACTCAAAG TCGTTAAAGCTCACCTGATCATCGGATGAACCC</p>	pDGB3_α1, pDGB3_Ω1		
<p>LEC2 M104-Q275 Y₂₅₇A Y₂₆₅A</p>	<p>LEC2-M104-AttB1F : GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATG ACGAAGATGGCAAGGATC</p> <p>LEC2-Q275STOP-AttB2R : GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAT TTTCCCGAATTTCCATTCAT</p>	pDONR207, pDEST32	Effet de la mutation Y ₂₅₇ A Y ₂₆₅ A sur les domaines d'activation de la transcription dans la levure.	
<p>LEC2 G240-Q275 Y₂₅₇A Y₂₆₅A</p>	<p>LEC2-G240ATG-AttB1 : GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATG GAATTTGTGAAGCAAATGGAGCTGAG</p> <p>LEC2-Q275STOP-AttB2R : GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAT TTTCCCGAATTTCCATTCAT</p>			

Annexe 5 : Séquences des 60 orthologues de AtLEC2 utilisés pour la prédiction et la conservation des domaines d'activation de la transcription

Le motif consensus TKxxARxxRxxAxR représentatif du domaine B2 (Boulard et al., 2018) est indiquée en rouge et la position du T par le premier nombre dans le nom de l'accession. Une tyrosine très conservée en fin du domaine B3 est représentée en bleu et sa position est indiquée par le deuxième nombre.

```
>XP_038688819_LEC2_[Tripterygium_wilfordii]-102-240
MEPFFPPPTTTNDTSSTATTPTQSQPHVATHDPALYSNIPPYYPTPQYFFWMVQNYNGSNEPAGPMIHAGNGLQGMLCSTQP
LLVPFTDHDQKRYTDAYMTKMARINRKRLARQSRSLNRSVSSSPNNSIQVQTRRDGRHEVQNGTVCVSRDLYNFYTPDNKRL
RELLKKELKNSDVGSLGRIVLPKREAQENLPLNTDKEGMQIVMRDVYSYKHWFSFKYKFWNSNNKSRMYVLENTVDFVKQNGLEI
GDSLTIYEDESKNLYFSIKKAKTQVDESSCENENKDNHHHHHLYMYQANRDQEEEEASLAILIEQLKHKEQQEEANTLMSLSID
GGTSSSRGKMRVDDHHSFNDYTYTGTGLVLPDPANHYNYKITIDDVNGKQEW
```

```
>XP_006304381_LEC2_[Capsella_rubella]-96-225
MDNFFSSSNANSVQFSLDNNLNSHFTTIPTYAHQPHHLLPPYPYVEQMAALMNPQTLYSSDGFPPQIPVQGTGSGFCSLV
SNTNFCGLGFLDPTKMARINRKRNAMIRSRNNSPNSPNDLVDSKRQLMMLDMKTSAQIPEKKDLYPHSTFDDKCLRVLCKVLL
KKSVDGALGRIVLPKREAEGKLPKLTNKEGIIIVEMRDVFSLSQSWFSKYKFWNSNNKSRMYVLENTGEFVKQNGLETGDTLTIYE
DESKNLYFAVTMSGKQNGGSGDESMEVNEIINNYDMNHYEELMNDYMQDEEEAAIAMLIGNLNDHYDPNPNDLMDLTIIDTDH
QHHQVQTASSSSSVHADVGSDDQASFSYAGW
```

```
>XP_006415675_LEC2_[Eutrema_salsugineum]-108-236
MDNFFLFPSSPNANSVQELSMDLNTRSHFSTVPTYDHHQAHHHYHMLPPYPNIACPVDPATAVMNPQPLHSSEAFSRFPFTRTG
SEFGSLVCNPLGRQERGGFLDPHMTKMARINRKKAMIRSRNNYCPNSSSNELVDSRRQVALNMKNNSSETAKKDLFRFSSFDN
KKLRVLLVKHLKNSDVGSLGRIVLPKREAEGNLPESLDSKEGIMVEMRDVDSVQSWFSKYKFWNSNNKSRMYVLENTGEFVKKNG
VETGDSLTIYEDESKNLYFSIRKYSKGQNEGREDESMEVNDMNFLDLTFDYIPNDEEEDSIAMLIGNLNDHYSIPNDLMDLT
TDLDQATSSSPHLHLSGDHATS DHVSSNDFVW
```

```
>KAF3486226.1_[Brassica_cretica]-100-243
MDNFFLFPSSPNANSVQELSMDLNKNRSHFSMAQPQHLLPPYSYVACPVLDQGTAMNHQPLHSSDAFPQIPVQGTGGEFGYLVC
KPGVQRQERGGFLDPHSTKMARINRKKAMIRSRNNSYLNSSSNELVDSRRQVALTMKNNAEIAARKDFYRFSSFDNKKLRVLLV
KHLKNSDVGSLGRIVLPKREAEGNLPESLDSKEGMVLQMRDVSQSWFSKYKLSLRTDKACINVKHRYWSNNKSRMYVLENTG
EFVKKNGVLMGDYLTIEYEDESKNLYFSIRKHPKQNDGREDESMEVIEMNFYEDIMFDYIPNGEDDSIAMLLGNLNEHYPPNP
DIMDLTVDLQHQATSSSPADHMSNDFLW
```

```
>XP_010460616_LEC2_isoform_X1_[Camelina_sativa]-105-234
MDNFFFPSSSNANSVQEFAMDANNLNSHLTTMPTYDHHQAEPHLLYSYPVEQMAAVMNPQPLYSSDAYPQIPVSQTGSEFC
SLVSNPNPCLRPERGGFLDPTKMARINRKNAMIRSRNNSPNSPDEVVDSKRQVMMLNMKNVQIPERKDHYPHSTFDNKK
LRVLCVKLLKNSDVGALGRIVLPKREAEGKLPKLSKEGIVLEMRDVSLSQSWFSKYKFWNSNNKSRMYVLENTGEFVRQNGVE
SGDTLTIYEDESKNLYFAVTKPGRQNGGRGVE SMKVNEMNHYEEIMYDFMPGDEEEASIAMLIGNLNDHYPNPNDLMDLTIIDH
QHHQAASSLPSVDYAHVGLDDQVSFDDIGW
```

```
>NP_564304_[Arabidopsis_thaliana]-105-234
MDNFFLFPSSPNANSVQELSDPNNRSHFTTVPTYDHHQAQPHHFLPPFSYPVEQMAAVMNPQPVYLSECYPQIPVPTQTGSEF
GSLVGNPCLWQERGGFLDPRMTKMARINRKNAMMRSRNNSPNSPSELVDSKRQLMMLNLKNNVQISDKKDSYQQSTFDNKK
LRVLCVKELKNSDVGSLGRIVLPKRDAEANLPKLSKEGIVVQMRDVSFMSQSWFSKYKFWNSNNKSRMYVLENTGEFVKQNGAE
IGDFLTIYEDESKNLYFAMNGNSGKQNEGRENESREERNHYEEAMLDIYIPRDEEEASIAMLIGNLNDHYPIPNLMDLTTDLQH
HQATSSSMPPEDHAYVGSDDQVSFNDFEWW
```

```
>XP_010499344_X1_[Camelina_sativa]-105-233
MDNFFFPSSSNANSVQEFAMDANNLNSHITTMPTLYHHQAEPHLLYSYPVEQMAAVMNPQQLYSSDAFPQIPVSQPGSEFC
SLVSNPNPCLRTERGGFLDPTKMARINRKNAMIRSRNNSPNSPDEVVESKRQVMMLNMKNVQIPVRTDLYPHSTFDNKK
LRVLCVKLLKNSDVGALGRIVLPKREAEGKLPKLSKEGIVLEMRDVSLSQSWFSKYKFWNSNNKSRMYVLENTGEFVRQNGVE
SGDTLTIYEDESKNLYFAVTKPGRQNGGRGVE SMEVNEMNHYEEIMYDFMPGDEEEASIAMLIGNLNDHYPNPNDLMDLTVDH
QHHQAASSLPSVDYPHVGLDDRVSFDDIWW
```

```
>AAL12004LEAFY_COTYLEDON_2_[Arabidopsis_thaliana]-105-234
MDNFFLFPSSPNANSVQELSDPNNRSHFTTVPTYDHHQAQPHHFLPPFSYPVEQMAAVMNPQPVYLSECYPQIPVPTQTGSEF
GSLVGNPCLWQERGGFLDPRMTKMARINRKNAMMRSRNNSPNSPSELVDSKRQLMMLNLKNNVQISDKKDSYQQSTFDNKK
LRVLCVKELKNSDVGSLGRIVLPKRDAEANLPKLSKEGIVVQMRDVSFMSQSWFSKYKFWNSNNKSRMYVLENTGEFVKQNGAE
IGDFLTIYEDESKNLYFAMNGNSGKQNEGRENESREERNHYEEAMLDIYIPRDEEEASIAMLIGNLNDHYPIPNLMDLTTDLQH
HQATSSMTPEDHAYVGSDDQVSFNDFEWW
```

>KAG2303789_[Brassica_carinata]-98-227
MDSFLPLSSSNPNVQELSMDFNNNSPGFSTFTPTYDHHHQAPHHLLPPYSYVACPVYQTAANKPQIPVIQTGSEFGSLVCD
PGFRQARCGFLDPTHAKMARINRKKAMIRSRNNSSPNSSNELVDSRRQVVLTMKNNAEIAARKDLRYFSSSHNKKLRVLLVK
QLKNSDVGSLGRIVLPKREAERNLPELSTKEGTIVEMRDADSMQNWSFKYKFWSSNNKSRMYVLENTGEFVTERVEIGDFLTI
YEDESKNLYFSVRKYEDKPNEGKEDESMEANDMNFYEDIAFDFIPKDEEEDSIAMLIIGNLNDNYPNPNLMDLPTDLHQHQA
TSSLPPDHYMTNPQHVGSSNDHMSFNDFVW

>KAG7655734_[Arabidopsis_suecica]-105-234
MDNFLPFPSSNANSVQELSMDFNNRSHFTTVPTYDHHHQAPHHFLPPFSYPVEQMAAVMNPQPVYLSECYPQIPVTQTGSEF
GSLVGNPCLWQERGGFLDPRMTAKMARINRKNAMMRSRNNSSPNSSSELVDSKRQLMMLNLKNNVQISDKKDSYQQSTFDNKK
LRVLCCEKELKNSDVGSLGRIVLPKRDAEANLPKLSKKEGIVVQMRDVFMSQSWFSKYKFWSSNNKSRMYVLENTGEFVKQNGAE
IGDFLTIYEDESKNLYFAMNGNSGKQNEGRENESRERNHYEEAMLDIYPRDEEEASIAMLIIGNLNDHYPIPNLMDLTTDLQH
HQATSSMPPEHDHAYVGSDDQVSFNDFEWW

>KAG7647807_[Arabidopsis_thaliana x Arabidopsis_arenosa]-105-234
MDNFLPFPSSNANSVQELSMDFNNRSHFTTVPTYDHHHQAPHHFLPPFSYPVEQMAAVMNPQPVYLSECYPQIPVTQTGSEF
GSLVGNPCLWQERGGFLDPRMTAKMARINRKNAMMRSRNNSSPNSSSELVDSKRQLMMLNLKNNVQISDKKDSYQQSTFDNKK
LRVLCCEKELKNSDVGSLGRIVLPKRDAEANLPKLSKKEGIVVQMRDVFMSQSWFSKYKFWSSNNKSRMYVLENTGEFVKQNGAE
IGDFLTIYEDESKNLYFAMNGNSGKQNEGRENESRERNHYEEAMLDIYPRDEEEASIAMLIIGNLNDHYPIPNLMDLTTDLQH
HQATSSMPPEHDHAYVGSDDQVSFNDFEWW

>CAG7864457_[Brassica_rapa]-96-225
MDNFLPFPSSNANFVQELSMDLNNRSLSTFTPTYDHHHQAPHSLOPYSYVACPVQDTAAMNPQIPVTQTRSEFGSLVCNPG
FGQARGGFLDPHTAKMARINRKKAMIRSRNNSSPNSSNELVGSRRQVVLTMKNNAEIAARKDLRYSSFDNKKLRVLLVKHL
KNSDVGSLGRIVLPKREAENLPELSTKEGMIVDMRDADSMQNWSFKYKFWSSNNKSRMYVLENTGEFVTEKRVEIGDFLTIYE
DESKNLYFSIRKHADKPNEGREDESMEANDMNFYEDIAFDFIPKDEEEDSIAMLIIGNLNDHYPIPNLMDLPTDLHQHQA
TSSLPPADYMTNPQYGGSSNDHMSFNDFVW

>XP_018489204_[Raphanus_sativus]-97-226
MDNFLPFPSSNANSVEELSMDLNNRSLSTFTPTYDHHHQAPPHLLPPYSYVASPVDQTAAMKQIPVIQTGSEYGLVYNP
GFRQARGGFLDPHTAKMARINRKKAMIRSRNNSSPNSSSTNEMVDSRRQVVLMSKNNAEIAARKDLRYSSFDNKKLRVLLVKHL
LKNSDVGSLGRIVLPKREAERNLPELSTKEGMIVEMRDADSMQNWSFKYKFWSSNNKSRMYVLENTGEFVTRKRVEIGDFLTIYE
EDESKNLYFSIRKYADKPNEGREDEWMEANDMNFYEDIAFDFIPKDEEEDSIAMLIIGNLNDHYPIPNLMDLPTDLHHHQAT
SSLPHVDYMTNPQHGGSSNDHMSFNDFVW

>XP_020866568_[Arabidopsis_lyrata_subsp._lyrata]-104-233
MDNFLPFPSSNANSVQELSMDLNNRSHFTTVPTYDHLAQPHHLMPPYSYPVEQMTAVTNPQPVYFSEGYQPVPVTKTGSEFG
SLVGNPCLRQERGGFLDPRMTAKMARINRKNAMIRSRNNSSPNSSSELVDSKRQVMMLNLKNNVQISDKKDSYQQSTFDNKKL
RVLCEKELKNSDVGSLGRIVLPKRDAEANLPKLSKKEGIVLEMRDVFMSQSWFSKYKFWSSNNKSRMYVLENTGEFVKQNGAEM
GDFLTIYEDESKNLYFAINGNSGKQNEGRENESRERNHYEDTMDLYISRDEEEASIAMLIIGNLNDHYPIPNLMDLATDLQYH
QATSSVPPEDHAYMGSSDDQVSFNDFEWW

>KAF8096561_[Sinapis_alba]-123-226
MDNSLPFPSSSNPNVQELSMDLNNRSHFTFTPTYDHHHQAPHHLMQPYSYIACPVQDMAAMNPQIPVIQTGSEFGYLVYNP
GFRQARCGFLDPHTAKMARINRKKAMIRSRNNSSPNSSNELVDSRRQVVLTMKNNAEIAARKDLRYFSSSDNNKKLRVLLVKHL
LKNSDVGSLGRIVLPKREAERNLPELSTKEGMIVEMRDADSMQNWSFKYKFWSSNNKSRMYVLENTGEFVTKKRVEIGDFLTIYE
EDESKNLYFSIRKYEDKPNEGREDESMEANDMNFYEDIAFDFIPKDEEEDSIAMLIIGNLNDQYPNPNLMDLPTDLHQHQA
TSSLPHVHYMTSPQHGGFSSDHMSFNDFVW

>KAG2269748_[Brassica_carinata]-96-223
MDNFLPFPSSNANSVQELSMDLNNRSLSTFTPTYDHHHQAPHHLLQPYSYVACPVQDTAAMNPQISVIQTGSEFGSLVCNPG
FRQARGGFLDPHTAKMARINRKKAMIRSRNNSSPNSSNELVGSRRQVVLTMKNNAEIAARKDLRYSSFDNKKLRVLLVKHL
KNSDVGSLGRIVLPKREAENLPELSTKEGMIVEMRDADSMQNWSFKYKFWSSNNKSRMYVLENTGEFVTEKRVEIGDFLTIYE
DESKNLYFSIRKHADKPNEGREDESMEANDMNFYEDIAFDFIPKDEEEDSIAMLIIGNLNDHYPIPNLMDLPTDLHQHQA
TSSLPPVDYMSNPQYGGSSNDHMSFNDFVW

>XP_013675255_[Brassica_napus]-96-225
MDNFLPFPSSNANSVEELSMDLNNRSLSTFTPTYDHHHQAPHHSLQPYSYVACPVQDTAAMNPQISVIQTGSEFGSLVCNPG
FRQARGGFLDPHTAKMARINRKKAMIRSRNNSSPNSSNELVGSRRQVVLTMKNNAEIAARKDLRYSSFDNKKLRVLLVKHL
KNSDVGSLGRIVLPKREAENLPELSTKEGMIVDMRDADSMQNWSFKYKFWSSNNKSRMYVLENTGEFVAEKRVEIGDFLTIYE
DESKNLYFSIRKHADKPNEGREDESMEANDMNFYEDIAFDFIPKDEEEDSIAMLIIGNLNDHYPIPNLMDLPTDLHQHQA
TSSLPPVDYMTNPQYGGSSNDHMSFNDFVW

>XP_013585180_[Brassica_oleracea_var._oleracea]-97-225
MDNFLPFPSSNANSVEELSMDLNNRSLSTFTPTYDHHHQAPHHSLQPYSYVACPVQDTAAMNPQISVIQTGSEFGSLVCNPG
FRQARGGFLDPHTAKMARINRKKAMIRSRNNSSPNSSNELVGSRRQVVLTMKNNAEIAARKDLRYSSFDNKKLRVLLVKHL
KNSDVGSLGRIVLPKREAENLPELSTKEGMIVEMRDADSMQNWSFKYKFWSSNNKSRMYVLENTGEFVAEKRVEIGDFLTIYE

DESKNLYFSIRKHADKPNEGREDESMEANDMNFYEDIAFDFIPKDEDEDSIAMLIGNLNDHYPNPNNLMDLPIDLHQHHQATS
SLPPVDYMTNPQYSGSSNDHMSFNDFVW

>XP_013714167_[Brassica_napus]-96-225
MDNFLPFSSSNANFVQELSMIDLNNRSLSTFPTYDHHQAQPHSLQPYSYVACPVDTAAMNPQIPVTQTGSEFGSLVCNPG
FGQARGGFLDPHTAKMARINRKKAMIRSRNNSPSSSNELVGSRRQVVLTMKNNAEIAARKDLRYRPSFDNKKLRVLLVKHL
KNSDVGSLGRILLPKREAEGNLPPELSTKEGMIVDMRDADSMQNWSFKYKFWSSNNKSRMYVLENTGQFVTEKRVEIGDFLTIE
DESKNLYFSIRKHADKPNEGREDESMEANDMNFYEDIAFDFIPKDEDEDSIAMLIGNLNDHYPNPNNRMDLPIDLHQHHQATS
LPPADYMTNPQYGGSSNDLMSFNDFVW

>KAG7598524_[Arabidopsis_suecica]-101-230
MDNFLPFSSSNANSVQELSMIDLNNRSHFTTVPPTYDHHQAQPHHLMPPYSYPVEQMMNPQPVYFSEGYPIPVVTQTGSEFGSLV
GNPCLRQERGGFLDPRMTKMARINRKNAMIRSRNNSPSSSELVDSKRQVMMNLKNNVQISDKKDSYQQSTFDNKKLRVLL
CEKELKNSDVGSLGRIVLPKRDAEANLQKLSKKEGIVLEMRDVFMSQSWSFKYKFWSSNNKSRMYVLENTGEFVKQNGAEMGDF
LTIYEDESKNLYFAINGNSGKQIEERENESRERNHYEDTMLDYISRDEEASIAMLIGNLNDHYPIPNLMDLTTDLHQHHQA
TSSVPSEDHAYVGSDDQVSFNDFEWW

>KAF3610465_[Brassica_cretica]-126-220
MDNFLPFSSSNANSVQELSMIDLNNRSLSTFPTYDHHQAQPHLLQPYSYVACPVDTAVMNPQISVIQTGSEFGSLVCNPG
FRQARGGFLDPHTAKMARINRKKAMIRSRNNSPSSSNELVGSRRQVVLTMKNNAEIAARKDLRYRPSFDNKKLRVLLVKHL
KNSDVGSLGRIVLPKRREAEGNLPPELSTKEGMIVEMRDADSMQNWFWSSNNKSRMYVLENTGEFVTEKRVEIGDFLTIEYEDSKN
LYFSIRKHADKPNEGREDESMEANDMNFYEDIAFDFIPKDEDEDSIAMLIGNLNDHYPNPNNLMDLPIDLHQHHQATSSSLPPV
DYMSNPQYGGSSNDHMSFNDFVS

>KAG5385288_[Brassica_rapa_subsp._trilocularis]-96-220
MDNFLPFSSSNANFVQELSMIDLNNRSLSTFPTYDHHQAQPHSLQPYSYVACPVDTAAMNPQIPVTQTGSEFGSLVCNPG
FGQARGGFLDPHTAKMARINRKKAMIRSRNNSPSSSNELVGSRRQVVLTMKNNAEIAARKDLRYRPSFDNKKLRVLLVKHL
KNSDVGSLGRIVLPKRREAEGNLPPELSTKEGMIVDMRDADSMQNWFWSSNNKSRMYVLENTGQFVTEKRVEIGDFLTIEYEDSKN
LYFSIRKHADKPNEGREDESMEANDMNFYEDIAFDFIPKDEDEDSIAMLIGNLNDHYPNPNNRMDLPIDLHQHHQATSLPPAD
YMTNPQYGGSSNDLMSFNDFVW

>XP_018490375_[Raphanus_sativus]-101-229
MDNLLPFSSSNANSVQELSMIDLNNRSHLSTMAQPHDLPPYSYVACPVLDQGTGAMNHQTFHLSDAFPHISVVQTGSEFGSLV
YKAGVRQERGVFLDPHTAKMARINRKKAMIRSRNNSPSSSNELVESRRQVSLTMKTNAEIAARKDLRYRPSFDNKKLRVLL
VKHLKNSDVGSLGRIVLPKRREAEGNLPPELTDKEGMVLEMKDVDSVQSWSFKYKYWSNNKSRMYVLENTGEFVKKNGVLMGDYL
TIYEDESKNLYFSIRKHPHKQNDKEDDESMEVNDVNYIEDIMFDYIPNDEDDSIAMLLGNLNEHYPNPNDLMSLTIDLHQHQ
ATSSSPPADHMSSNDFVW

>KAG2320722_[Brassica_carinata]-101-229
MDNLLPFSSSNANSVQELSMIDLNNRSHFSTMAQPHHLLPPYSYVACPVLDQGTGAMNHQPLHSSDAFPQIPVVQTESEFVSLV
YKPGVRQEIIGGFLDPHTAKMARINRKKAMIRSRNSPSSSNELVDSRRQVALTMKNNAEIAARKDLRYRPSFDNKKLRVLL
VKHLKNSDVGSLGRIVLPKRREAEGNLPPELSDKEGMVLEMRDVSQSWSFKYKYWSNNKSRMYVLENTGEFVKKNGVLMGDYL
TIYEDESKNLYFSIRKHPHKQNDREAEPMEVNDMNYIEDIMFDYIPNDEDDSIAMLLGNLNEHYPNPNDLMLNLTIDLHQHQ
ATSSSPPADHMSSNDFVW

>XP_010478197_[Camelina_sativa]-89-218
MDNFFPFSSSNANSVQEFAMDANNLSHLLYSYPMEQMAAVMNPQPLYSSDAFPQIPVSPQGEVCSLVSNPCLRPERRG
LFLDPHTAKMARINRKNAMIRSRNNSPSSPDEVVESKRQVMMNMRNNVQIPERKDLYPHSTFDNKKLRVSCVKKLKNSDVGA
LGRIVLPKRREAEGKLPKLSKKEGIVLEMRDVFSLQSWSFKYKFWSSNNKSRMYVLENTGEFVRQNGVESGDTLTIYEDESKNLY
FAVTKPGRQNGGRGAKSMEVNERNHYEIMYDFMPGDEEASIAMLIGNLNDHYPNPNDLMLGLTVDHQHYQAASSLLSVDYAH
VGSLLDQVSDDIW

>KAG5378817_[Brassica_rapa_subsp._trilocularis]-100-223
MDNFLPFSSSNANSVQELSMIDLNNRSHFMAQPQHLLPPYSYVACPALDQGTGMNHQPLHSSDAFPQIPVVQTTGGEFGYLVC
KPGVRQERGGFLDPHTAKMARINRKKAMLRSRNNSPSSSNELVDSRRQVALTMKNNAEIAARKDFYRFSFDNKKLRVLLV
KHLKNSDVGSLGRIVLPKRREAEGNLPPELSDKEGMVLEMRDVSQSWYWSNNKSRMYVLENTGEFVKKNGVLMGDYLTIEDE
SKNLYFSIRKHPHKQNDGREDESMEVIEEMNFYEDIMFDYIPNDEDDSIAMLLGNLNEHYPPNDLMDLTVNLDQHQATSSSP
PADHMSSNDFLW

>CAE5958712_[Arabidopsis_arenosa]-82-211
MDLNNRSHFTTVPPTYDHHQAQPHHLMPPYSYPVEQMMNPQPVYFSEGYPIPVVTQTGSEFGSLVGNPCLRQERGGFLDPRMTK
MARINRKNAMIRSRNNSPSSPSELVDSKRQVMMNLKNNVQISDKKDSYQQSTFDNKKLRVLCCEKELKNSDVGSLGRIVIP
KRDAEANLPKLSKKEGIVLEMRDVFMSQSWSFKYKFWSSNNKSRMYVLENTGEFVKQNGAEMGDFLTIYEDESKNLYFAINGNS
SKQNEGRENSRERNHYEDTMLDYISRDEEASIAMLIGNLNDHYPIPNLMDLTTDLHQHHQATSSVPSEDHAYVGSDDQV
SFNDFEWW

>XP_010499347_[Camelina_sativa]-105-207

MDNFFPSSSNANSVQEFAMDANNLNSHITTMPTLYHHQAEPHLLYSYPVEQMAAVMNPQHLLYSSDAFPQIPVSQPGSEFC
SLVSNPNPCLRTERGGFLDPTKMARINRKNAMIRSRNNSPSSPDEVVESKRQKLRVLCVKKLNSDVGALGRIVLPKREA
EGKLPKLSKDEGIVLEMRDVFSLQSWFSKYKFWSNKSRMYVLENTGEFVRQNGVESGDTLTIYEDESKNLYFAVTKPGRQNG
GRGVESEMEVNEMNHYEIMYDFMPGDEEEASIAMLIGNLNDHYPNPNLMDLGLTVDHQHHQAASSLPSVDYPHVGLSDDRVSFD
DIVW

>KAF8107878_[Sinapis_alba]-85-213

MDNLLPFSSNSNSVQELSMDLNNSHFSSTMVQPHLLPYSYVACQGTGAMNHQPFHSSDIAFPQIPVVQGTQGERGGFLDPH
STKMARINRKKAMIRSRNNSPSSSSELVDSRRQVALTMKNNAEIAARKDLYRFNTFDNKKLRVLLVKHLKNSDVGSLGRIV
LPKREAEGNLPESLDEKGMVLEMRDVSQSWFSKYKYSNNKSRMYVLENTGEFVKKNGVLMGDYLTIIYEDESKNLYFSIRK
HPHKQNDREDESMEVNDMNYEDVMFYI PNEDDSIAMLLGNLNEHYPNPNLMDLDTIDLQHHQATSSSPADHMSNDP
VW

>XP_017978733_[Theobroma_cacao]-138-271

MENSYPFTPTTTTTNTTLRQPDMGWSQNSEFSIYEPMLNKNFSSTPFSTASNSQFHHPPQFVAPSQSLQPPYRHPHIEQAGH
SLAYPMYFLLGQNGIEYGVSSSIHANNGVQGSFCMQTDGLSKEQERRIMDPYR TKVARIKRKLARQRSLSLQRNASSGASTQ
VDARRLTSSGADNSTDVNNSNDTKRDLYKFCTPDNKRLRVLLRKEKNSDVGSLGRIVLPKREAEGNLPESLDEKGIQVMIKD
VYSNQVWTLKYKFWSNKSRMYVLENTGDFVKQNGLEIGDSLTLTYEDESKNLYFSVIKLERIAANPPSNHQSDNENNYYSN
NDHHMYLPFTCQSRDEEETSLELLIEQLKHKEQEQETNDLMSLPMDAAYSHRLEPEAKSFLSDADGVASMETYTPIAATTLQT
SSTLLRGKAKSVDDFQLNNFDDCYGGLDMLPDVNHYNFSL

>XP_022777347_[Durio_zibethinus]-138-269

MDNSYIPFSPTTTTNTTLRQPNMGWSQNLEFSIYEPMSKNFSSTPLTTASNSQFHHPPQFMAPYQSLQPPYRHPHIEQAGQ
SLPYVYPFFMGRNGIEFGVCSNHNASNGVQGSFCIQTDGLSKEQERRILDPYR TKVARIKRKLARQRSLSLQRNASSEASTQ
VDARRLTNSGADTKVKNNDTKRDLYKFCTTENKRLRVLLSKELKNSDVGSLGRIVLPKREAENLPLETSDKEGIQVMIKD
SNQVWTLKYKFWTNNKSRMYVLENTGDFVKQNGLEIGDSLTLTYEDESKNLYFSILKVERVAVEPSKQHSNENNNHMYVFFT
CQSRDEEDTSLELLIEQLKHKEQEAENLMTLPMDAAYSHRLEPEARHLLSDSIVASRETYSRITTTPLQTSSTLLQKARS
VDIEINFDDCYGGLDMLPDVNHYNFSL

>XP_022720461_[Durio_zibethinus]-138-269

MENSYPFSPTTTTNTTLRQPNMGWSQNSGFCIYEPMSKNFSSTPSSTPSNSQFHHPSQLVAPYQSLQPLHRQPTQAGH
SLAYLVYPPFFLQNGIEFGESSIQASNGVQGSFCIQTDGLSEEQERRILDPYI TKVARIRKQARQRSLSLQRSASSGASTQ
VEERRLTSSGADTDVNNNDTKRDLYMFCTPDNKRLRMLLKKELKNSDVGSLGRIVLPKKEAEVNLPLETSDKEGILVMIKD
SNQVWTLKYKFWSNKSRMYVLENTGDFVKQNGLEIGDSLTLTYEDESKNLYFFILKVERFAAEASNNQHCDNVNNNNHMYL
TCQSRDEEETSLELLIEQLKHKEQEQEPNDFMTLPMDAAYSGRLPKEARHFLSDTFSSRETYSQITATTLQTSSTSI
RKGARSVDIIQINFDDCYGGLDMLPDVSLYNFSL

>XP_024951254_[Citrus_sinensis]-129-259

MENNFFPASNNNTATATANNNTINTFTNQQLSQPNMAYIQNSQLYLYPTYYHQNMQFPSTFFLDQSGGQSFVAAAAAPYVY
FWLAQNGIQNGGNNAI PAAAGSWSLQPNIRVPKEPQRVLDLPQK TKIARIKRKLARQKLSLQRNASSGASSNVPSSAHFD
ATRLPTYGVDTHRDLFNCTPDNKRLRVLLQKALKNSDVGSLGRIVLPKREAENLPLETSDKEGIIHIVIRDVYSDKHWT
LKFYWSNNKSRMYVLENTVDFVKQNGLEIGDSLTLTYEDESKNLYFSVKKVERPEAAEPSYKQNYDFFPAANQTP
TPIHQARDEEASLALLIEQLRHKEQEQEYTNLMTLPLDDHEEANNVVRASNNATS SGIHAHLAATSSSSHGAKMNNPEYD
HHQTVDDCYGLDMLPDVDTYNFSL

>KAF9840828_[Populus_deltoides]-128-260

MENVFPTLSTNSGNTTTTTNASSVMMGCPQNSQFSSIIYGPLSYSTPSSSLTQSNYFAPQYTTTPCETTQFPSTDSLDQRSQ
YTPAADPFYVPMVGQHA IEMQRSCILQPNALQLEQERRALDAYK TKVARCMRRLARQRRLLGRASSGASSTRITDASRLALHG
AGTDGQNSRTNTSKDLFEFLTPDNKKLRALLRKDLKNSDVGSLGRIVLPKREVEENLPLHDKEGILLVLRDIYSNQEWGLKL
KFWTNNKSRMYVLENTGEFVVRHGLETGDSLTLTYEDESKNLYFSITKVQRPESALAQTPHPKSHNCNHLTEHMQDGD
DEEYTS LALLAEQLENKEQEEANSLGAVPMDFSCSYTNELELNNPFNNISTYTPASAAAMQTSPPNGKMKAVDSDH
DVCYTGLGMLPDVHSYNFSL

>KAG6756038_[Populus_tomentosa]-129-260

MENVFPTLSTSSGNTTTTTNASSVMMGCPQNSQFSSIIYGPLSYSTPSSSLTQSNYFAPQYTTTPCETTQFPSTHALDQRS
QYTPAADPFYVPMVGQNA IEMQRSCILQPNALQLEQERRALDAYR TKVARCMRKLARQRRLLGKASSGASSTRIMDASRLALHG
ASTDGQNSRTNTSKDLYEFLTDQDNKKLRALLRKDLKNSDVGSLGRIVLPKREVEENLPLHDKEGILLMLRDIYSNQEWGLKL
KFWTNNKSRMYVLENTGEFVVRHGLETGDSLTLTYEDESKNLYFSITKVQRPESALAQTPHPTSHNCNHLTEHMQDGD
DEEYTS LAFLTEQLDNKEQEEANSLGAVPMDFSCSYTNELELNDPFNNMSTYTPASAAAMQTSPPNGKMKAVDSDH
DVCYTGLGMLPDVHSYNFSL

>KAG6754877_[Populus_tomentosa]-128-259

MENVFPTLSTSSGNTTTTTNANSVMMGCTQNSQVSSIIYGPLSYSTPSSSLTQSNYFAPQYTTTPCETTQFPSTHALVQRSQ
YTPAADPFYVPMVGQNA IEMQRSCILQPNALQLEQERRALDAYK TKVARCMRKLARQRRLLGKASSGASSTRIMDASRLALHG
STDGQNSRTNNSKELYEFLTDQDNKKLRALLRKDLKNSDVGSLGRIVLPKREVEENLPLHDKEGILLVLRDIYSNQEWGLKL
KFWTNNKSRMYVLENTGEFVVRHGLETGDSLTLTYEDESKNLYFSITKVQRPESALAQTPHPTSHNCNHLTEHMQDGD
DEEYTS ALLTKQLENKEQEEANSLGAVPMDFSCSYTNELELNNPFNNISTYTPASAAAMQTSPPNGKMKAVDSDH
DVCYTGLGMLPDVHSYNFSP

>XP_011043404_[Populus_euphratica]-127-259
 MENVFPTLSTSSGNTTTTNTNASNAIMGCPQNSPLSSIYGPSHPSSSSSLTHSNYFAPQYYTTPCETTQFPSTHALDQRSQY
 TPAPDFPYPVMVGQNAIEMQQRSCILQPNAFQLEQERRALDAYKTKVARCMRKLARQRRLGKASSGASSTRIMDASRLALHGAS
 TGGQNCRTNSASKDLYEFLTPDNKKLRALLRKDLKNSDVGSLGRIVLPKREVEENLPLHDKEGILLVLRDIYSNQEWGLKLLK
 FWTNNKSRMYVLENTGEFVKRHGLETGDSLTLYEDESKNLYFSITKVQRPESALAQTPPPRSHNCNHLTEHMWQDGDDEEYTSL
 ALLTEQLENKEQEEANSLGAVPMDFSCSYTNELELNNPFNNFSTYTQPASAAMQTSPPNGKMKAVDDSHDVEDDCYTGLGMLP
 DVHSYNFSL

>XP_024454872_[Populus_trichocarpa]-129-260
 MENLFTLFTSTGNTTNTNASNAIMGCPQNSQFSSMYGPLSYHSTVPSNSVTSQSNVPPQYYMPAPCESRQNPYTHSLEKPS
 HHTPSADPFRLAMAGQNAFEMQQRSCILQPNAFQLEQERRALDAYKTKVARCKRKLARQRSLSKTSIGASSTQIMDARRLALHG
 ASTCGQRSRSNTSKDLYEFLTPDNKRLRAVLRKDLKNSDVGSLGRIVLPKREVEENLPLVNDKEGILLFLRDVYSNQEWALKF
 KFWSNNKSRMYVLENTGEFVKQNGLETGDFLTLYEDESKNLYFSITKVGPASVPSQTPQPINHNSNCLYTPHMYQARDEEKS
 SLALLIEQLDQKEQEEANSLVAVPPDSAYTNNELTNNPFNNSSTYSQPASSAMQPSPPNGKMKAVDDSHDVEDDCYTGLGVLDPV
 YRFNFSL

>XP_011027804_[Populus_euphratica]-128-252
 MENLFTLFTSTGNTTNTNASNAIMGCPQNSQFSSMYGPLSYHSTVPSNSITQSSCVPPQYFMPAPCESWQNPYHMSLEKPSH
 HTPSTDFLLAMAGQNAFEMQQRSCILQPNAFQLEQERRALDAYKTKLARCKRKLARQRSLSKTSSTQIMDARRLALHGASTSGQRS
 RSNTSEDLYEFLTPDNKRLRAVLRKDLKNSDVGSLGRIVLPKREVEENLPLVNDKEGILLFLRDVYSNQEWALKFKFWSNNKS
 RMYILENTGEFVKRNGLETGDFLTLYEDESKNLYFSISRVGRSASVPSQTPQPINHNSNCLYTPHMCQARDEEQSSLALLIEQ
 LDQKEQEEADSLVAVPPDSAYMDSVQLTNNPFNNSSTYSQPASSAMQPSPPNGKMKAVDDSHDVEDDCYTGLGVLDPVYRYNFSL

>KAG5223544_[Salix_suchowensis]-121-252
 MENVFPTSSSTNSGSASNPTNASSVMGCPQNSQVPVYGPSSYPSTLSSLLTQSSYTPPCETTEFPAYSLEQRSQYTPSADP
 FYPVMAGPDAIEMQGPCILQPNAFQLEQERRALDANKTKVARCVRKLARQRRLGKASSGGSSARIKDTSRALALHGATIDSQNS
 RTNSKDLYEFLTPDNKKLRALLRKDLKNSDVGSLGRIVLPKREVEENLPLHDKEGILLVLRDVNSNQEWGLKFKFWANNKS
 RMYVLENTGEFVKLNGLETGDSLTLYEDESKNLYFSITKAQT PASALTQTPLPMSQNNNYLYTESTWQDRDEEYTSLALLTEQ
 LENEQEEANSLGAVPMDFSCSYTHEPELNNIRTYTQPASAAMQAS PANGKMKAVDDSHDVEDDCYTGLGMLPDVHSYNFSP

>KAB5534019_[Salix_brachista]-121-252
 MENVFPTSSSTNSGSTSNPTNASSVMGCPQNSQVPIYGPSSYPSTLSSLLTQSNYTPPCETTEFPPTYSLEQRSQYTPSADP
 FYPAMVGPDAIEMQRPCILQPNAFQLEQERRALDANKTKVARCVRKLARQRRLGKASSGGSSARVKDTSRALALHGANIDSQNS
 RTNSKDLYEFLTPDNKKLRALLRKDLKNSDVGSLGRIVLPKREVEENLPLHDKEGILLVLRDVNSNQEWGLKFKFWANNKS
 RMYVLENTGEFVKLNGLETGDSLTLYEDESKNLYFSITKAQT PASALTQTPLPMSQNNNYLYTEHTWQDRDEEYTSLALLTEQ
 LENEQEEANSLGAVPMDFSCSYTHEPELNDIRTYTQPASAAMQTS PANGKMKAVDDSHDVEDDCYTGLGMLPDVHSYNFSP

>KAF9672359_[Salix_dunii]-121-252
 MENVFPTSSSTNSGSTTNTNASSVMGCPQNSQVPIYGPSSYPSTLSSLLTQPNYTPPCETTEFPPTYSLEQRSQYTPSADP
 FYPVMVGPNAIEMQRPCILQPNAFQLEHERRALDANKTKVARCARKLARQRRLGKASSGASSARIKDTSRALALHGANIDSQNS
 RTNTSKDLYEFLTPDNKKLRALLRKDLKNSDVGSLGRIVLPKREVEENLPLHDKEGILLVLRDVNSNQEWGLKFKFWANNKS
 RMYVLENTGEFVKRSGLETGDSLTLYEDESKNLYFSITKAQT PASPLTQTPHPMSQNDNYLYTEHTWDRDEEYSSLALLTEQ
 LENEQEEANSLGAVPMDFSCSYTHEPELNNIRFTTQPASAAMQTS PANGKMKAVDDSHDADDCYTGLGMLPDVHSYNFSP

>XP_012092409_[Jatropha_curcas]-112-243
 MENLFSPLSTNITSNTTTITNTNATKPNMSFSQNSPLSSICGPLNYPSTSSPPDHSIPPPFQTQVHTSQFSFPYFVFLQN
 GIETQRSYLLQPNGFSIDQERRVLDAYKTKVARSKRKLARQRSLIKNSSSGVNSAQMDIKKPAFNGASTNVQDSQGNINKDLY
 TFCTLDNKKLRVLLTKELKNSDVGSLGRIVLPKRDTTEENLPIELSEKEGIQVVLIDVNTNQEWSLKFKYWSNNKSRMYVLENTG
 DFVKQNGMESGDSLTLYEDESKNLYFSIKKVETQEAEPFTNQSYLIPNACHARDEEEASFALLMEQLKHKEQEEANSLVT
 LMSDIGSSYRHKEEEKNGRFFNNVTSIGRTSSSAQSMDDHFDLDDCYNGLDILPDVNRYNFYD

>XP_024046749_[Citrus_clementina]-95-225
 MAYTQNSQLYLPTYHYHQNMQYPSTFPLDQSGGQSFAAAAAAAYPVYFVWLAQNGIQNGGGNNALPAAAGSWSLQPHIRVVK
 EPQRVVLDPQKTKIARIKRKLARQKSLSLQRNASSGSSNPVSSAHFDATRLPTYGVDTHRDLFNFCTPDNKRRLRVLLQKAL
 KNSDVGSLGRIILPKREAEENLPLSDKEGIHIVIRDVYSVKHWTLKFYWSNNKSRMYVLENTVDFVKQNGLEIGDSLTLYE
 DESKNLYFSVKKVERPEPAAEPSYKQNYDFFPAANQTPPTPIHQARDEEEASLALLIEQLRHKEQETYYNLMTLPLDDHEE
 ANNVVRASNNATSSGIIHAHLAATSSSSHGAKMNNPEYDHHQTVDDCYSCLDMLPDVDTYNFSI

>KHG23900_[Gossypium_arboreum]-104-233
 MENSYIPLFSTTATSSTPTITTTTTTHSPSNMMDWSQNSDFSMFEPMNSQFQHPSQLLSPNQSLQFPYNHSIEQAPVYPFFTQ
 KGFEFGVCKEQERRTNDPYRTRLARINRKLARQRSLQRNANSASTQVDARRLINSADTDTNKNKDTKKDLYRFCTADNKR
 RVLLKELKNSDVGSLGRIVLPKRDAEVNLPPLSDKEGIQVMIKDVYSNHMWTLYKFWNSNNKSRMYVLENTGDFVKQNGLGT
 GDSLTLYEDESKNLYFSITKVATVAVESNNQYYHNVNNGEEGNDNSNNMYLPFTSQSKDDEAASLELFMEQEPNDFIMT
 LPMDSTYGSHTMFPEETHRLPNNVDQMPSLRSRVDDISINFDDCYGGDLMLPEANQYNFSF

>XP_034912425_[Populus_alba]-125-256
 MENLFTLFTSTGNTTNTNASNAIMGCPQNSQFSSMYGPLSYHSTSSSITQSSCVPLYPMPAPCESRQNPYHMSLEKPSHHTP
 SADPFRLAMAGQNAFEMQQRSCILQPNAFQLEQERRALEAYKTKVARCKRKLARQRSLSKTSIGASSTQIMDARRLALHGASTC

GQSRSTNSKDLIEFLTPDNKRLRAVLRKDLKNSDVGSLGRIVLVPKREVEENLPLVNDKEGILLFLRDVYSNQEWALKFKFWS
NNKSRMYVLENTGEFVKQNWLETGDFLTLYEDESKNLYFSITKAGRSASVPSPTLLIEQLDQKEQEEANSLVAAPPDSAYTHN
EQLINNPLNNSSTFSQPASSAMQPSQNGKMKAVDDYHVDDCYTGLGVLPDVYRYNFSL

>OMO76390_[Corchorus_capsularis]-140-271
MENSYPFVTPPTTATTNNTSVRPPNVPWLQSGQNSEYPIYEPMNLDNFRFSTLDSASASNSNSNPNSQFLHHHTQTPNQS
LVPFRHPMEQAAGQSLAYPIYGYPFMGMQNGIEFGVQGCSCYMQPSVERRIPDSYR**TKVARIKRK**MARQSRSLTLQRNNSCPGAS
SHMDTRRLTSYGGDRSVKNNNNNDKRDLYKFFTPDNKRLRVLLKKEKLNKSDVGSGLGRIVLVPKREAEVNLPTLSDKEGIQVIFKD
VYSDQFWSLRYKFWLNNKSRMYVLENTGDFVVRQNGLEMGDSLTLTYDDESKNLYFSITKVGRIAAEPSPDIQSDNDNSNTNTSN
SNSNISNNNNNNNNINDMYMFPFTCYKDDEEETSLEILIEQLKQDNNDLMTLPMPEPTEDQARHEARSVDDVQVFNFDCCYGGLE
MLPGFSHYNFSL

>XP_016698142_[Gossypium_hirsutum]-134-265
MMENSYPFLLPTTSTATNTILRQPNMGCSQNSELSIYEPMNSKNFSSKPIITAPNSQFHHHTQFMAPYQSLRRLPYRHPPIEQTD
HSLAYPFFLGQNGIEFGSIIASNGVQHGSCFMQTDGVSKAERMLDPY**GSKAARIKRKL**ARQSRSLSLQRNASSGATALVDKRR
RLNSSGADNDVNINNDTKRDLYMFCTPDNKRLRVLLRKLKLNKSDVGSGLGRIVLVPKREAEVNLPTLSDKEGIQVVLKDVYSHQS
WPLKYKFWNSNNKSRMYVLENTADDFVKQNGLGIGDFLTLYEDEGKNLYFSVIKVERVAAAEPSPNNQPYDSQNNYMDMPFTGPLR
DEEETCLELLIEQLKHEHEESNDLMTLAMDATYSHGPPAADESQHFVSDANVELTATELQTSSTSLRGKARSVDEIQINFDCC
CYGLDMLPDVNLNFFLL

>XP_017607697_[Gossypium_arboreum]-132-264
MENSYPFLLPTTSTATNTILRQPNMGCSQNSELSIYEPMNSKNFSSKPIITAPNSQFHHHTQFMAPYQSLRRLPYRHPPIEQTDH
SLAYPFFLGQNGIEFGSIIASNGVQHGSCFMQTDGVSKAERMLDPY**GSKAARIKRKL**ARQSRSLSLQRNASSGATALVDKRR
LNSSGADNDVNINNDTKRDLYMFCTPDNKRLRVLLRKLKLNKSDVGSGLGRIVLVPKREAEVNLPTLSDKEGIQVVLKDVYSHQSW
PLKYKFWNSNNKSRMYVLENTADDFVKQNGLGIGDFLTLYEDEGKNLYFSVIKVERVAAAEPSPNNQPYDSQNNYMDMPFTGPLR
DEEETCLELLIEQLKHEHEESNDLMTLAMDATYSHGPPAADESQHFVSDANVELTATELQTSSTSLRGKARSVDEIQINFDCC
YGLDMLPDVNLNFFLL

>KAB2081998_[Gossypium_barbadense]-133-264
MENSYPFLLPTTSTATNTILRQPNMGCSQNSELSIYEPMNSKNFSSKPIITAPNSQFHHHTQFMAPYQSLRRLPYRHPPIEQTDH
SLAYPFFLGQNGIEFGSIIASNGVQHGSCFMQTDGVSKAERMLDPY**GSKAARIKRKL**ARQSRSLSLQRNASSGATALVDKRR
LNSSGADNDVNINNDTKRDLYMFCTPDNKRLRVLLRKLKLNKSDVGSGLGRIVLVPKREAEVNLPTLSDKEGIQVVLKDVYSHQSW
PLKYKFWNSNNKSRMYVLENTADDFVKQNGLGIGDFLTLYEDEGKNLYFSVIKVERVAAAEPSPNNQPYDSQNNYMDMPFTGPLR
DEEETCLELLIEQLKHEHEESNDLMTLAMDATYSHGPPAADESQHFVSDANVELTATELQTSSTSLRGKARSVDEIQINFDCC
YGLDMLPDVNLNFFLL

>TYJ34487_[Gossypium_mustelinum]-132-264
MENSYPFLLPTTSTATNTILRQPNMGCSQNSELSIYEPMNSKNFSSKPIITAPNSQFHHHTQFMAPYQSLRRLPYRHPPIEQTDH
SLAYPFFLGQNGIEFGSIIASNGVQHGSCFMQTDGVSKAERMLDPY**GSKAARIKRKL**ARQSRSLSLQRNASSGATALVDKRR
LNSSGADNDVNINNDTKRDLYMFCTPDNKRLRVLLRKLKLNKSDVGSGLGRIVLVPKREAEVNLPTLSDKEGIQVVLKDVYSHQSW
PLKYKFWNSNNKSRMYVLENTADDFVKQNGLGIGDFLTLYEDEGKNLYFSVIKVERVAAAEPSPNNQPYDSQNNYMDMPFTGPLR
DEEETCLELLIEQLKHEHEESNDLMTLAMDATYSHGPPAADESQHFVSDANVELTATELQTSSTSLRGKARSVDEIQINFDCC
YGLDMLPDVNLNFFLL

>GAV71053_[Cephalotus_follicularis]-120-243
MENLYTFFSNTSTDTTAMTNLRQPNIGCFINETSYPSPSTQFYTAPHYNYMPTHETLQNPYTYPIQQPDTSLALPMYPFM
QNGIQDGLQSPIHGSSFLQQNVSLMQEKKSIDAY**TKLARIKR**KQARQKSIRANKNSSASSSTSTQADVNERNNEGNRDLY
KFFTPDDKRLRVLLKKEKLNKSDVGSGLGRIVLVPKREAEVNLPIITDKEGIGIVLRDLYSIQEWFSFKFYWSNNKSRMYVLENTG
KFVRQNGLEMGDSLTLTYEDESCHLYFSIKKTEKPTAEPLYEQHYANQNNNYLYTPTSWQARDEEEASLALLIEQLKHEHQEV
NNLMALDCASSSSYMQTEEAQMDLSKQGTFTLLGAASSMQHSYPWHGNLDAVNDCQLNVDDYGGGLMPLDRL

>XP_039026020_[Hibiscus_syriacus]-237-368
MRSIDEAHLNQRWRQHSLKPTHWSDEYPKTTFRRPLVSHSTTPLATSFLFASLTLYMQFHGLFLSPQQIPLHLLTRFSCFLFHT
WTHPNGSILHPHDLIEPHFVLTQNRKEHTQSSSLSPFFPDHHCNSLHSSSHCTFVLMENSYIPFPTATTTTSTPTNS
TSNMGCSSNSNFSTIEPMSQFQHPSQLPPYQSLQLPYRYLMEQAPPMPYCFDFGVWKDQERRIMDPNR**TKMARINR**KLARQ
RSLSLQRNATPVASTQVDARRLIPSGINTYMRQNNNDKDHCFITPDNKRLRVLLKKEKLNKSDVGSGLGRIVLVPKREAEVNL
TLYDKEGIQVTLKDVYSNNEWTLKYKFWINNKSRMYVLENTGDFVKQNGMIRIGDSLTYEDESNNLYFSISKVATVAVESSN
QNQYNNNNNNNNSSHIYLASKDDDEASLELLIEQLKHEKPEPNLMMTLPMDTANNQRILPEETRHLFSHDSQMQTST
TLPGMSGTINFDDCYGLDMLPDVNLNFFLL

>TYI27395_[Gossypium_tomentosum]-135-265
MMENSYPFLLPTTSTATNTILRQPNMGCSQNSELSIYEPMNSKNFSSKPIITAPNSQFHHHTQFMAPYQSLRRLPYRHPPIEQTD
HSLAYPFFLGQNGIEFGSIIASNGVQHGSCFMQTDGVSKAERMLDPY**GSKAARIKRKL**ARQSRSLSLQRNASSGATALVDKRR
RLNSSGADNDVNINNDTKRDLYMFCTPDNKRLRVLLRKLKLNKSDVGSGLGRIVLVPKREAEVNLPTLSDKEGIQVVLKDVYSHQS
WPLKYKFWNSNNKSRMYVLENTADDFVKQNGLGIGDFLTLYEDEGKNLYFSVIKVERVAAAEPSPNNQPYDSQNNYMDMPFTGPLR
DEEETCLELLIEQLKHEHEESNDLMTLAMDATYSHGPPAADESQHFVSDANVELTATELQTSSTSLRGKARSVDEIQINFDCC
CYGLDMLPDVNLNFFLL

>XP_021635023_[*Hevea brasiliensis*]-129-260
MESPFSPSTDASKTTITTTNDSKAKMGCSQNSPFSSICGPLSYSTSSNSSTQSYHCPVECSMLPHFQSLQFPQAYPLGHAG
QYSPSAYPLYPFWFQNGIEIQKSCFLQPNGLSMEQERRVLDAYKTKVARSKRKLARQRSLSKNSSSGANSQVDTRRLSFG
ASNKEQDCQSFINRDLYKICTPDNKRLRVLLSKELKNSDVGSLGRIVLPKRGAEENLPILSDKEGIQVVRDVSNTKEWSLKF
KFWSNNKSRMYVLENTGDFVKQNGLEMGDSLTLYEDECKNLYFSIKKVPQVEVSYKQHSTNQNYLYIPHTYQARDEEEASL
QLLIEQLKHKEQQEANSVLTLSVDIASSYRRKEDERIGPLNNTSISFYQPAPAAVQSSSSPHSITRILDDHYIDDCYTGGLG
VLPDVNRYNFSL

>XP_021598372_[*Manihot esculenta*]-128-261
MENPFSPSIDTSNTTYDSESKMGGSQNSSTSTPPASSIYGPLSYSTSSCSSTQPYHCPVEYSMLPPFQSLQFPQAYPQGHG
GQYSPSTYPLYPCWFLQNGIEKSCFLQPNVSVIEQERRVLDAYKTKVARTKRKLARQRS LGKNASSGANLSQVDTRRLSLNGG
NNNGQAGLSNININRDLYKICTPDNKRLRVLLSKELKNSDVGSLGRIVLPKRGAEENLPILSDKEGMQVVRDLSNPKEWLKL
FKFWSNNKSRMYVLENTGDFVKQNGLEIGDSLTLYEDECKNLYFSIKKVKTPQIEVSYNQNSTNQNYLYIPHTYQARDEEEAS
LQLLIEQLNHKEQQEANSVLTLSMDIASSYRRKEEERLSPFNNTSISFYSPAPAAAMEASSPHNTMRILDDHYIDDCYSGL
DVLDPVNRYNFS

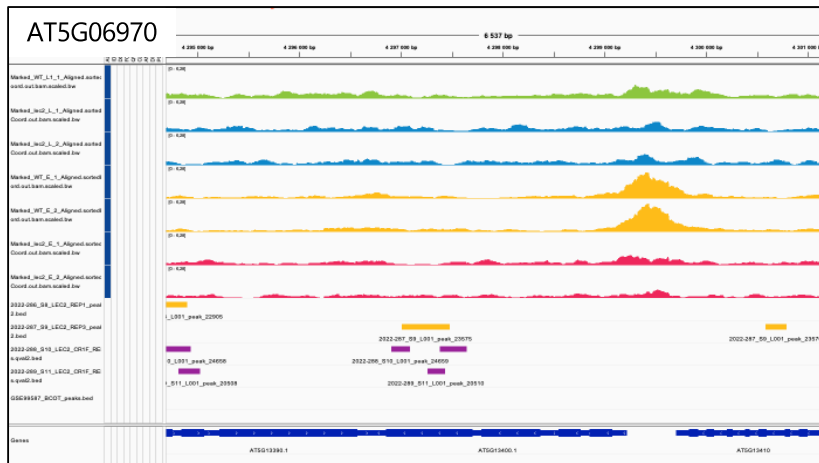
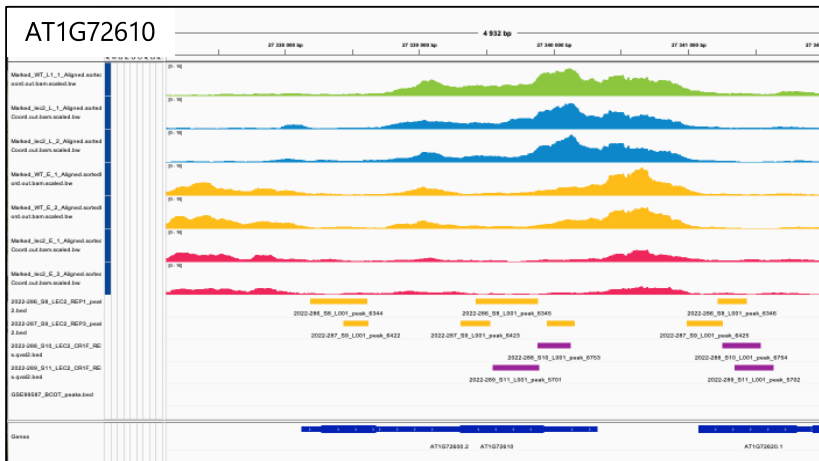
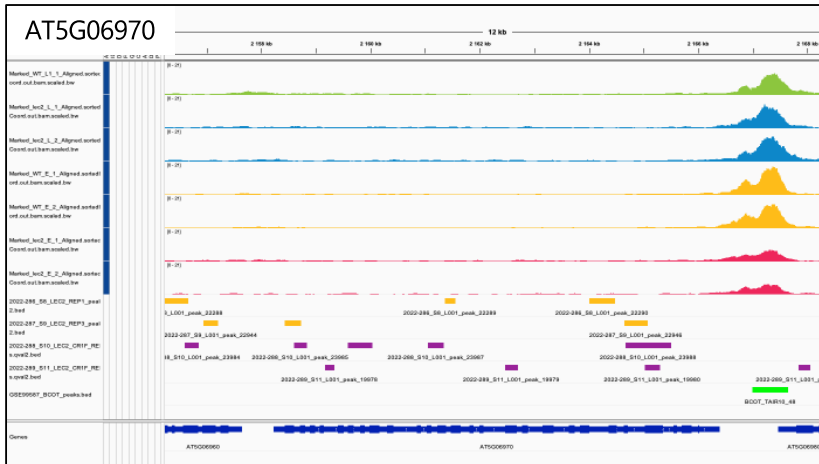
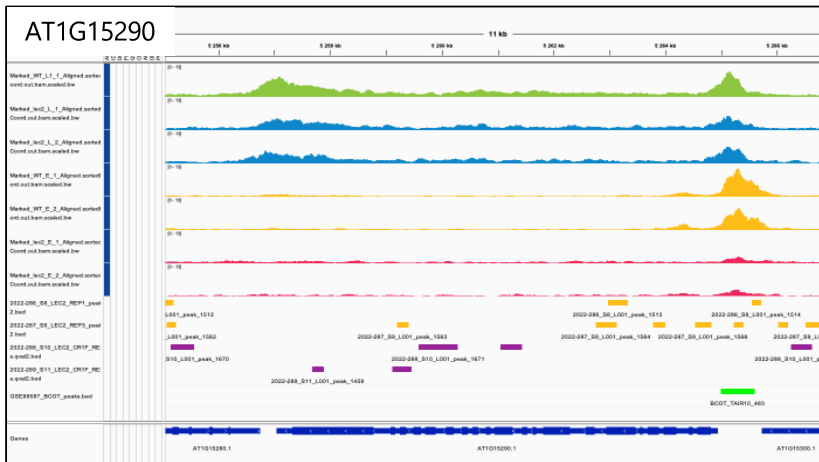
>NP_001310615_[*Ricinus communis*]-103-229
MDSFQNSQFSSIRPLCYSSSTSSNSSTQTYHCPAEDNMPPTTFQSPHFQOEYRPEQAGQNSTSAFMSMCPYWLSQNGTEIQTWL
LQONGLSIDQGKVLDAYKTKIARNKRKLARQRSLSRNSSSSANSIQVDTRRLAFNMQDSQSTSKRDLYKFCPTDNKRLRVLL
RKDLKTSDVGSLGRIVLPKREAEENLPILSDKEGILVAIRDVCSTKEWSLKYKYWSNNKSRMYVLENTGDFVKQNGMRIGDSL
TLYEDESKKLYFSVKKVEALEAEPSCQKQATNQNYIYIPHTYQARDEEEASLALLIEQLKHKEQQEENSLVLTLSMDIAPYR
KEEEKNYGPFNNTSISIYPQSTAAAMAIQPSSSAHSTMRVDDHYIDDFYTALDVLDPVNRYNFSL

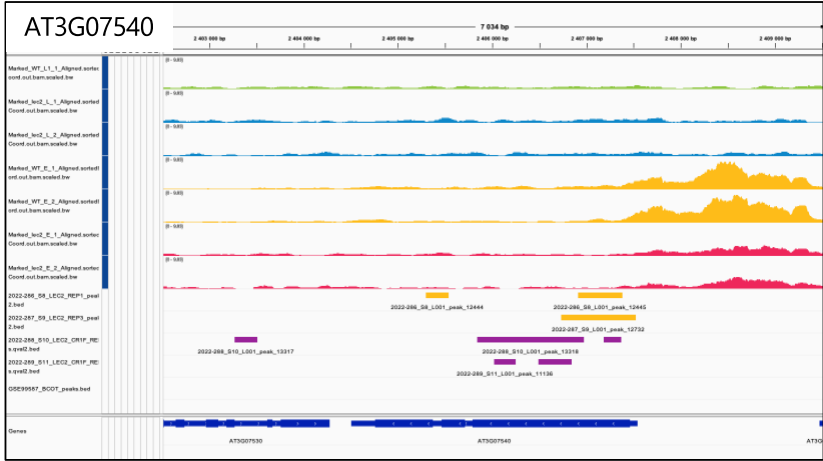
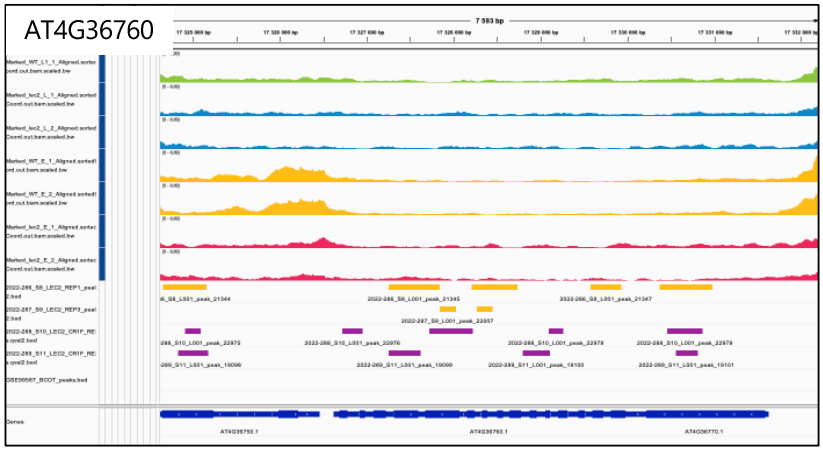
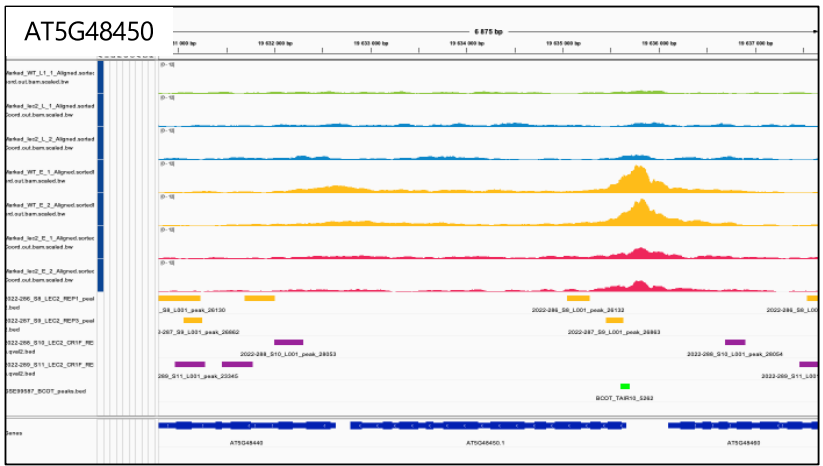
>XP_021664066_[*Hevea brasiliensis*]-103-228
MGCSQNSPLSSICGPLNHPSTSSNSSESYSYHCPAEYHMLPPFQSLQFPQAYPLEHAGQYSPSAYPLYPFWIWQNGIEIHKSCF
LQPNGLSIGQARRVLDAYKTKVARSKRKLARQRS PGKNFPGANSQVDIRRPSFVEDSQSTINRDLYKCTSDNKRLRALLS
KELKNSDVGSLGRIVLPKREAEENLPILSDKEGIQVVRDVSNSNKDWSLKFKFWCANNKSRMYVLENTGDFVKQNGLEIGDSL
LYEDECKNLYFSFKKLETPQILPSYKQRPTNYLYIPHTYQDRDEEEASLALLIEQLNRKEQQESNSVLTLPMDISSSYRHKKE
ERNGPFNNTSISINPEPAPAAAMQPTSMQVDDHYIDDCYTGGLVLPDVNSYNFSL

>XP_021667227_[*Hevea brasiliensis*]-103-228
MGCSQNSPLSSICGPLNHPSTSSNSSESYSYHCPAEYHMLPPFQSLQFPQAYPLEHAGQYSPSAYPLYPFWIWQNGIEIHKSCF
LQPNGLSIGQARRVLDAYKTKVARSKRKLARQRS PGKNFPGANSQVDIRRPSFVEDSQSTINRDLYKCTSDNKRLRALLS
KELKNSDVGSLGRIVLPKREAEENLPILSDKEGIQVVRDVSNSNKDWSLKFKFWCANNKSRMYVLENTGDFVKQNGLEIGDSL
LYEDECKNLYFSFKKLETPQILPSYKQRSTNYLYIPHTYQDRDEEEASLALLIEQLNHKEQQESNSVLTLSMDISSSYRHKKE
ERNGPFNNTSISINPEPAPAAAMQPTSMQVDDHYIDDCYTGGLVLPDVNIYNFSL

>Csa_6G517930_*Cucumis sativa*-111-244
MDDHNNPHQNHHHHHHELPPFPSTAVTAPTNTLHSPQNPCYLPPTLSQTCLASFSSAPQSFSSSHEPQPLAPRFPHSPTL
LSYGLDEPNIQSVFPKDDGKKALDAWSTKVARTNRRRLARQRTLTSTTRSSSSSSCGYHDDYSKGLTITHFPNTDSSHNRHL
FTFCTPDNKILRMLLKKELKNSDVGSLGRIVLPKREAEENLPFLTDREGLQIVVRDVSNNRRWTMKYKYWANRRSRMYVLENT
GFTSTN

>XP_021295029_[*Herrania umbratica*]-235-402
MTWRLFSIVPLQTNCDLLCKLNWTEALDMASEWQCSGADNGHREVQKTAKITLKEADIPSSISASLPYDSPFHLLVTDEPK
TATFKPKVAPTLQALSMVENSYPPTPTNTTLRQPNKGWSQNSEFSIYEPMLNKNFTSTPLSTASNSQFHHPQFEAPXQS
LQPPYRHPQAGHSLAYPMYFFXGQNGIEYGVSSIHANNGVQSGCFMQTDGLSKEQERRIMDPYR TKVARIKRKLARQRS
LSLQRNASSGASTQVDVRLISSGADNSTDVNNSNDTKRDLYKFRTPDNKRLRVLLRRELKNSDVGSLGRIVLPKREAEENLP
TLSDKEGIQVMIKDVYSNQVWTLKYKFWNNKSRMYVLENTGDFVKQNGLEIGDSLTLYEDESKNLVEQYFSIIKLERIAAE
PPSNHQHSDNETNYNSDHHMYLPFTCQSRDEEETSLELLIEQLKHKEQQEETNDLMSLMDAAYSHRLEPEAKSFLSDADG
VASMETYPPIAATTLQTSSTLLRGKAKSVDDFQLSNFDDCYGGLDMLDPVNHYNFSL





Annexe 7 : Milieux et solutions utilisés pour la culture et la transformation des mousses

Extrait du protocole « Principles and protocols to work with the moss *Physcomitrella patens* » Didier G. Schaefer, Institute of Ecology, Laboratory of Plant Cell Genetics, University of Lausanne, updated September 1998.

Macro elements: (100X individual solution, stored at 4°C)		final concentration	
Ca(NO ₃) ₂ 4H ₂ O solution	8 g in 100 mL of water	80 g/L (338 mM)	autoclavable
MgSO ₄ 7H ₂ O solution	2,5 g in 100 mL of water	25 g/L (100 mM)	autoclavable
FeSO ₄ 7H ₂ O solution	125 mg in 100 mL of water	1,25 g/L (4,5 mM)	filtered 0,2 µm

Micro elements 1000X:		final concentration	<u>for 20 mL</u>
CuSO ₄ 5H ₂ O solution (55 mg in 10 mL of water)		55 mg/L (0,22 mM)	200 µL
ZnSO ₄ 7H ₂ O solution (55 mg in 10mL of water)		55 mg/L (0,19 mM)	200 µL
H ₃ BO ₃ solution (61,4 mg in 10 mL of water)		614 mg/L (10 mM)	2 mL
MnCl ₂ 4H ₂ O solution (389 mg in 10 mL of water)		389 mg/L (2 mM)	200 µL
CoCl ₂ 6H ₂ O solution (55 mg in 10 mL of water)		55 mg/L (0,23 mM)	200 µL
KI solution (28 mg in 10 mL of water)		28 mg/L (0,17 mM)	200 µL
Na ₂ MoO ₄ 2H ₂ O solution (25 mg in 10 mL)		25 mg/L (0,1 mM)	200 µL
H ₂ O qsp 20 mL			16,8 mL

Autoclave in a 50 mL bottle and store at 4°C after opening.

KOH 10N:	<u>for 50 mL</u>
KOH	28 g
H ₂ O qsp 50mL	

Be careful add only 20 mL and complete after dissolution then autoclave in a 100 mL bottle.

Buffer phosphate 1000X:		final concentration	<u>for 100 mL</u>
KH ₂ PO ₄		250 g/L (1,84 mM)	25 g
pH=7 with KOH		10 N	around 10 mL
H ₂ O			qsp 100 mL

Autoclave in a 250 mL bottle and store at 4°C after opening.

Ammonium tartrate 1M:		final concentration	<u>for 50 mL</u>
(NH ₄) ₂ C ₄ H ₄ O ₆		184 g/L (1M)	9,2 g
H ₂ O			qsp 50mL

Autoclave in a 100 mL bottle and store at 4°C after opening.

Solid PPNH₄ medium:		final concentration	<u>for 500 mL</u>
Agar (Kalys)		6-7 g/L depend on batch	3 to 3,5 g
Ca(NO ₃) ₂ solution 100X		80 g/L	5 mL of 100X
MgSO ₄ solution 100X		25 g/L	5 mL of 100X
FeSO ₄ solution 100X		1,25 g/L	5 mL of 100X
Micro elements 1000X			0,5 mL of 1000X
Phosphate buffer 1000X			0,5 mL of 1000X
Ammonium tartrate 1M			1,35 mL of 1M
H ₂ O			qsp 500mL

Autoclave in a 1 L bottle and stored at RT, heat 8 min 500W (simmer) in a microwave before use.

8,5% mannitol:	final concentration	<u>for 500mL</u>
Mannitol (SIGMA M4125,182,17g/mol)	0,46 M	42,5 g
H ₂ O qsp 500 mL		470 mL

Mix at 4°C, pH=4,2 autoclaved in a 1 L bottle and store at RT.

Driselase 2% solution (-20°C):	final concentration	<u>for 50mL</u>
Driselase (Sigma D9515 -25G)	2%	1 g
Mannitol (Sigma M4125 -1kg)	8,5%	4,25 g

1-wrappe the 50mL falcon into parafilm and mix 45 min at 4°C, centrifuge 10 min at 10000 g.
2-pass the solution through a folded cellulose paper 30 µm (Fisherbrand 1611H).
3-sterilise with a 0,2 µm filter with syringe and aliquot by 10 mL in sterile 15mL tubes. Store at -20°C.

Tris-HCl 1M:		<u>for 100 mL</u>
Tris base		12,11 g
H ₂ O		qsp 100 mL

Adjust to pH=7,2 with HCl and autoclave in a 250 mL bottle.

PEG solution (-20°C):	final concentration	<u>for 30 mL</u>
Mannitol 7%	0,38 M	2,1 g
CaNO ₃	100 mM	0,708 g
PEG 4000 Sigma 81240 (p/v)	40%	12 g
TRIS 1M pH=7,2	10 mM	300 µL TRIS 1M
H ₂ O		15mL

Mix and heat at 50°C, autoclave in a 100 mL bottle. Make 10 mL sterile aliquots in tubes with warm solution. Sequential defrosting induces pH decrease. Store at -20°C. pH=7,5 before autoclave, pH=8 after autoclaved, inactive if pH < 6.

MMM solution (4°C):	final concentration	<u>for 100 mL</u>
Mannitol	8,5%	8,5 g
MgCl ₂	15 mM	0,305 g
MES	0.1%	0,1 g
H ₂ O		qsp 100mL
Adjust to pH=5,6 with KOH 1N		10 µL of KOH 10N

Autoclave in a 250 mL bottle, store at 4°C.

Liquid PPNH₄ medium + mannitol (without CaN₂O₆ and FeSO₄ to avoid precipitation):

	final concentration	<u>for 500mL</u>
Mannitol 6,6%	0,36 M	33 g
MgSO ₄ solution 100X	25 g/L	5 mL of 100X
Micro elements 1000X		0,5 mL of 1000X
Phosphate buffer 1000X		0,5 mL of 1000X
Ammonium tartrate 1M		1,35 mL of 1M
H ₂ O		qsp 500mL

Mix at 4°C, autoclave in a 1L bottle and store at RT.

Just before use, prepare in a new bottle		98mL of autoclaved solution
add Ca(NO ₃) ₂ from a 100X solution	80 g/L	1mL for 100 mL
add FeSO ₄ from a 100X solution	1,25 g/L	1mL for 100 mL