



**HAL**  
open science

# Caractérisation biochimique du complexe AFL et identification des partenaires de LEC2 lors du développement de la graine

Céline Boulard

► **To cite this version:**

Céline Boulard. Caractérisation biochimique du complexe AFL et identification des partenaires de LEC2 lors du développement de la graine. Biologie moléculaire. Université Paris-Saclay, 2017. Français. NNT : 2017SACLS402 . tel-02356789v1

**HAL Id: tel-02356789**

**<https://theses.hal.science/tel-02356789v1>**

Submitted on 9 Nov 2019 (v1), last revised 14 Nov 2019 (v2)

**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.

NNT : 2017SACLS402

THESE DE DOCTORAT  
DE L'UNIVERSITE PARIS-SACLAY,  
préparée à l'Université Paris-Sud

ÉCOLE DOCTORALE N° 567  
Sciences du Végétal : du Gène à l'Ecosystème

Spécialité de doctorat biologie

Par

**Mme Céline Boulard**

Caractérisation biochimique du complexe AFL et identification  
des partenaires de LEC2 lors du développement de la graine

**Thèse présentée et soutenue à Versailles, le 8 novembre 2017:**

**Composition du Jury :**

Mme Buitink Julia	Directrice de recherche INRA	Rapporteur
Mme Gallardo Karine	Directrice de recherche INRA	Rapporteur
Mr Colcombet Jean	Chargé de recherche INRA	Examinateur
Mr Roudier François	Professeur ENS Lyon	Président du Jury
Mme Zubieta Chloe	Chargée de recherche CNRS	Examinatrice
Mr Dubreucq Bertrand	Directeur de recherche INRA	Directeur de thèse





## Remerciements

Pour commencer, je tiens à remercier les membres du jury d'avoir accepté de juger mes travaux, Buitink Julia, Gallardo Karine, Colcombet Jean, Roudier François, Zubieta Chloe.

Merci également aux membres du comité de thèse d'avoir suivi mes travaux, Sandrine Lafranchi, Hakim Mireau et François Roudier.

Je remercie ensuite Loïc Lepiniec, pour m'avoir accepté dans son équipe et permis de développer une thématique qui me tient à cœur.

Un grand merci à Bertrand pour avoir encadré mes travaux pendant ces quatre années. Avoir un encadrant capable de garder son calme en toutes circonstances a été un atout pour moi, qui en suis bien incapable. Merci d'avoir pris le temps de discuter, d'écouter toutes mes hypothèses et d'avoir considéré toutes mes envies d'évolution du projet.

Je souhaite aussi remercier toute l'équipe « Développement et Qualité de la graine » : à Sébastien, Johanne, Damaris, Abdou, Chloé, Guillaume, Adrian pour l'intérêt porté à mes travaux et aux développements de mon projet. Merci à Martine pour sa disponibilité et pour les coups de pouce. Un grand merci tout particulier à Alexandra et Delphine, toujours présentes lorsque j'ai eu besoin d'aide, même pour des tâches aussi rébarbatives que des prélèvements de graines en développement. Vous avez été d'un grand soutien par votre aide et votre écoute, merci pour les moments de partage quotidiens.

Tous mes remerciements également à l'équipe voisine « Physiologie de la graine » : Annie, Helen, Isabelle, Loïc, Jeff, Béatrice, Adeline, Gwendal, Boris, Christine, Isabelle, Édith, pour leurs gentillesse, leurs présences et les bons moments de partage. Je souhaite plus particulièrement remercier François, mon voisin de bureau, pour son humour et son optimisme à toute épreuve; Elodie et Anne, pour m'avoir aidée et soutenue après le soir du 18 décembre 2014 ; j'ai été très touchée par votre altruisme, qui me restera gravé longtemps.

Merci aux membres des plate-formes de l'IJPB, toujours là pour aider. En cytologie, merci à Olivier, Katia et Adeline pour votre humour et les coups de pouce apportés lors de mes observations. A la plate-forme du végétal, merci aux serristes : Lilian, Bruno et Amélie, pour l'attention portée à mes plantes. A la plate-forme de biochimie, merci à Gwendal et Martine, accueillants, disponibles, curieux, avec qui je suis toujours contente de discuter.

Je tiens également à faire un petit clin d'œil à Laura, une stagiaire motivée que j'ai eu le plaisir d'encadrer et à Stéphanie et Marianne pour tous nos échanges.

Merci enfin à Pascale et Sandrine, sans qui je n'aurai pas appris autant de savoir-faire et n'aurai pas fait ce doctorat. Merci de m'avoir soutenu lors de mon envie d'évolution et de m'avoir ouvert les yeux sur mes capacités à continuer en thèse.





## Sommaire

I	Introduction .....	1
I.1	Introduction .....	1
I.1.1	Le développement de la graine .....	1
I.1.1.1	La fécondation .....	1
I.1.1.2	Les différents tissus de la graine .....	2
I.1.1.3	L'embryogenèse .....	2
I.1.1.4	La maturation .....	3
I.1.1.4.1	Le remplissage de la graine .....	3
I.1.1.4.1.1	Les protéines .....	3
I.1.1.4.1.2	Les glucides .....	4
I.1.1.4.1.3	Les lipides .....	4
I.1.1.5	La maturation tardive et la dessiccation .....	4
I.2	Revue .....	6
I.2.1	Introduction .....	6
I.2.2	Revue: Regulation and evolution of the interaction of the seed B3 transcription factors with NF-Y subunits. ....	7
I.3	Conclusion .....	38
I.4	Objectif de la thèse .....	41
II	Recherche de partenaires avec a priori .....	42
II.1	Introduction .....	42
II.2	Les LAFL interagissent avec le promoteur de l' <i>OLEOSINE1</i> .....	43
II.2.1	Formation d'un complexe AFL sur l'ADN .....	43
II.2.1.1	Validation de la formation du complexe in vitro .....	44
II.2.1.2	La dynamique de formation du complexe .....	44
II.2.2	Rôle d'un NF-YC dans le complexe LEC1, ABI3, LEC2 (LAL) .....	45
II.2.2.1	La famille des NF-Ys .....	45
II.2.2.2	Action de NF-YC2 sur le complexe LAL .....	46
II.2.3	Discussion .....	47
II.3	Caractérisation de l'interaction LEC1/LEC2 .....	50
II.3.1	Introduction .....	50
II.3.2	Article de résultats: .....	51
II.3.3	Résultats complémentaires .....	82
II.3.3.1	L'import et l'export nucléaire .....	82
II.3.3.2	Données préliminaires sur le trafic de LEC2 .....	83
II.3.3.3	Discussion .....	83
II.4	Des bZIPs font-ils partie du complexe LAFL ? .....	85
II.4.1	La famille des facteurs de transcriptions bZIPs .....	85
II.4.2	Structure .....	86
II.4.3	Régulation des bZIPs .....	86
II.4.4	Les bZIPs identifiés ayant une interaction avec certains AFL ou NF-Y .....	86
II.4.4.1	Interaction des bZIPs avec les NF-Y .....	86
II.4.4.2	Interaction des bZIPs avec un B3 .....	87
II.4.5	Résultats .....	88
II.4.5.1	Dimérisation des bZIPs .....	88
II.4.5.2	Action des bZIPs en dimère, seuls, sur le promoteur de l' <i>OLEOSINE1</i> .....	89

II.4.5.3	Implication des bZIPs dans le complexe LEC1, ABI3, NF-YC2, LEC2, (LANL) sur le promoteur de proOLE1 .....	89
II.4.6	Conclusion, discussion et perspectives .....	90
II.4.6.1	Dimérisation des bZIPs et activités avec les LANL .....	90
II.4.6.2	Formation du complexe .....	92
II.4.6.3	Action des bZIPs sur le promoteur .....	93
II.4.6.4	Conclusions et perspectives .....	94
II.5	Conclusion, discussion et perspectives .....	95
III	Recherche de partenaires sans a priori .....	97
III.1	Introduction .....	97
III.2	Recherche de partenaire .....	97
III.2.1	Fonctionnalité des LEC2 étiquetées .....	97
III.2.2	Capture de complexe protéique lié à LEC2 <i>in vivo</i> .....	99
III.3	Conclusion, discussion et perspectives .....	101
IV	Conclusion et discussion générale .....	104
IV.1	La formation du complexe LAL .....	104
IV.2	La recherche de nouveaux partenaires des AFL .....	105
IV.3	Localisation, expression et régulation de LEC1 et LEC2 .....	106
IV.4	Structure des protéines .....	108
IV.5	Les séquences promotrices .....	109
IV.6	Conclusion .....	109
V	Matériels et Méthodes .....	111
V.1	Matériels .....	111
V.1.1	Le matériel végétal .....	111
V.1.2	Les bactéries .....	111
V.1.3	Les levures .....	112
V.1.4	Les vecteurs .....	112
V.2	Méthodes .....	114
V.2.1	Biologie moléculaire .....	114
V.2.1.1	Réaction de recombinaison Gateway BP simple .....	114
V.2.1.2	Réaction de recombinaison Gateway BP pour recombinaison LR double ....	114
V.2.1.3	Réaction de recombinaison Gateway BP inversée .....	115
V.2.1.4	Réaction de recombinaison Gateway LR simple .....	115
V.2.1.5	Réaction de recombinaison Gateway LR double .....	115
V.2.1.6	Réaction TOPO .....	116
V.2.1.7	Extraction d'ADN génomique .....	116
V.2.2	Biologie végétale .....	116
V.2.2.1	Stérilisation et la culture du matériel végétal .....	116
V.2.2.2	Transformation des plantes d' <i>Arabidopsis thaliana</i> par <i>Agrobacterium tumefaciens</i> .....	117
V.2.2.3	Sélection des transformants .....	117
V.2.3	Analyse standard des protéines .....	117
V.2.3.1	Gel SDS-PAGE .....	117
V.2.3.2	Western blot .....	117
V.2.4	Constructions lignées LEC2 étiquetés .....	118
V.2.4.1	Construction de pGWB2 proLEC2-1Kb .....	118
V.2.4.2	Constructions de LEC2 et GFP témoin étiquetés HS en N et C terminal	118

V.2.4.3	Constructions de LEC2 et témoin étiquetés GFP et GS en N et C terminal ..	119
V.2.5	Constructions des vecteurs d'expressions pour l'étude des bZIPs .....	119
V.2.6	Expression et purification des protéines .....	119
V.2.6.1	Étude des interactions par EMSA .....	119
V.2.6.2	Production des anticorps spécifiques .....	120
V.2.6.3	Étude des interactions par thermophorèse .....	120
V.2.7	Fabrication des outils spécifique de LEC1 et LEC2 .....	121
V.2.7.1	Production des anticorps .....	121
V.2.7.2	Purification des anticorps .....	121
V.2.8	Observation et localisation .....	122
V.2.8.1	Immunolocalisation et la co-immunolocalisation sur matériel fixé .....	122
V.2.8.1.1	Préparation des échantillons .....	122
V.2.8.1.2	Coupe et immunolocalisation .....	122
V.2.8.2	Localisation de GFP:LEC2 in vivo .....	123
V.2.8.3	Observation confocale de GFP:LEC2 in vivo, étude du trafic de la protéine .	123
V.2.8.4	Observations confocales .....	123
V.2.9	Études des interactions .....	124
V.2.9.1	Interactions ADN / protéine .....	124
V.2.9.1.1	Interaction ADN / protéine par EMSA .....	124
V.2.9.1.2	Interaction ADN / protéine par Thermophorèse .....	125
V.2.9.1.3	Interaction ADN / protéine par expression transitoire chez <i>Physcomitrella patens</i> .....	125
V.2.9.2	Interactions protéine / protéine .....	126
V.2.9.2.1	Interaction protéine / protéine par double hybride .....	126
V.2.9.2.2	Interaction protéine / protéine par TAP TAG et Immunoprécipitation	127
V.2.9.2.2.1	Préparation de l'échantillon .....	127
V.2.9.2.2.2	Purification en tandem par affinité (TAP-tag) .....	128
V.2.9.2.2.3	Immunoprécipitation .....	129
V.2.9.2.2.4	Identification des partenaires de LEC2 par spectrométrie de masse .....	129
V.2.9.2.3	Identification des protéines en spectrométrie de masse .....	131
V.2.9.2.3.1	Migration courte en gel SDS-PAGE .....	131
V.2.9.2.3.2	Digestion trypsique sur gel .....	131
V.2.9.2.3.3	Dessalage des protéines digérées .....	131
V.2.9.2.3.4	Analyse en spectrométrie de masse .....	132
VI	Annexes .....	I

## Listes des abréviations

2S2	ALBUMINE 2S
ABA	acide abscissique
abi	acide abscissique insensible
ABI3	ACIDE ABSCISIC INSENSITIVE3
ADN-T	acide désoxyribonucléique transféré
ADNc	acide désoxyribonucléique complémentaire
AFL	ABI3, FUS3, LEC2
AIP2	« ABI3- INTERACTING PROTEIN 2 »
ARM	armadillo
<i>At</i>	<i>Arabidopsis thaliana</i>
bZIP	« BASIC LEUCINE ZIPPER »
CBF	« CAAT BOX BINDING FACTOR »
CL	corps lipidiques
Col-0	Columbia - 0
D86K	acide aspartique-86-lysine
DAF	« day after fertilization »
DAP	« day after pollinisation »
FAD	« FATTY ACID DESATURASE »
FT	facteur de transcription
FUS3	FUSCA3
GA	gibbérrellines
GFP	« GREEN FLUORESCENCE PROTEIN »
GS	protéine G Streptavidine
GTP/GDP	guanosine triphosphate/ guanosine diphosphate
HAP	histone
HFM	« histone fold motif »
HS	histidine- strepII
Hsp	« HEAT SHOCK PROTEIN »
IBB	« importine $\beta$ binding »
KD	constante de dissociation
L1L	LEC1 1 Like, LEAFY COTYLEDON 1 Like
LAF1	LEC1, ABI3, FUS3, LEC2
LAL	LEC1, ABI3, LEC2
LANL	LEC1, ABI3, NF-YC2, LEC2
LEA	« LATE EMBRYOGENESIS ABONDANT »
LEC1	« LEAFY COTYLEDON 1 »
LEC2	« LEAFY COTYLEDON 2 »
NES	« nuclear export sequence »
NF-Y	« NUCLEAR FACTOR Y »
NLS	« nuclear localisation sequence »
<i>Pp</i>	<i>Physcomitrella patens</i>
proOLE1	promoteur de l'OLEOSINE1
proOLE2	promoteur de l'OLEOSINE2
proOLE3	promoteur de l'OLEOSINE3
PSV	« PROTEIN STORAGE VESICULE »
RAV1	« RELATED TO ABI3/VP1 »
SSP	« seed storage protein »
TAG	triacylglycérol
TEV	« TOBACCO ETCH VIRUS »
Ws	Wassilewskija

## Liste des figures

<b>Figure 1</b> : Structure de la graine d' <i>Arabidopsis thaliana</i> au stade cœur de développement	p2
<b>Figure 2</b> : Schéma comparé du développement de graines d' <i>Arabidopsis</i> (dicotylédone) et de maïs (monocotylédone)	p3
<b>Figure 3</b> : Accumulation des différentes réserves au cours du développement de la graine d' <i>A. thaliana</i> au stade embryogénèse et aux différentes étapes de la maturation	p4
<b>Figure 4</b> : Structure des protéines AFL B3 et d'un complexe NF-Y	p42
<b>Figure 5</b> : Etude de l'interaction, par la technique du gel retard, des protéines LEC1, LEC2, ABI3 seules ou en combinaison sur la séquence promotrice de <i>proOLE1</i>	p44
<b>Figure 6</b> : Tests d'activité de LEC2 et ABI3 avec le promoteur de l' <i>OLEOSINE1</i> en présence ou absence de NF-YC2 et de LEC1 (WT) ou LEC1 muté D86K	p46
<b>Figure 7</b> : Séquences et structures de NF-Y B et C	p48
<b>Figure 8</b> : Séquence et structure de la protéine LEC2	p82
<b>Figure 9</b> : Localisation intracellulaire de GFP:LEC2	p83
<b>Figure 10</b> : Résumé des hypothèses relatives au transport et à la régulation de LEC2 dans la cellule	p84
<b>Figure 11</b> : Classement des 75 bZIPs identifiés chez <i>Arabidopsis thaliana</i>	p85
<b>Figure 12</b> : Structure d'un dimère de bZIPs avec une interaction avec l'ADN	p86
<b>Tableau 1</b> : Tableau récapitulatif des données connues des bZIPs ayant une interaction avec un ou plusieurs LANL	p87
<b>Tableau 2</b> : Tableau récapitulatif des résultats de la formation de dimères de bZIPs en double hybride	p88
<b>Figure 13</b> : Activité des bZIPs avec le promoteur de l' <i>OLEOSINE1</i>	p89
<b>Figure 14</b> : Activité des bZIPs avec le promoteur de l' <i>OLEOSINE1</i> en présence de ABI3, NF-YC2, LEC1, LEC2 (LANL)	p90
<b>Tableau 3</b> : Résumé des résultats obtenus sur l'étude des bZIPs en relation avec les LANL, avec le promoteur de l' <i>OLEOSINE1</i>	p91
<b>Figure 15</b> : Analyses de séquences du domaine B2	p92
<b>Figure 16</b> : Résumé des connaissances sur le complexe LAFL	p95
<b>Tableau 4</b> : Récapitulatif des étiquettes et descriptif des étiquettes choisies pour la capture et la purification de complexe lié à LEC2	p98
<b>Figure 17</b> : Tests d'activité de LEC2 sur le promoteur de l' <i>OLEOSINE1</i>	p98
<b>Figure 18</b> : Restauration phénotypique du mutant <i>lec2-1</i> par différentes constructions permettant d'étiqueter LEC2	p99
<b>Figure 19</b> : Génotypage des transformants LEC2 - étiquetés.	p100
<b>Tableau 5</b> : Résumé des identifications des protéines par spectrométrie de masse après purification des complexes protéiques par immunoprécipitation GFP	p101



## **I Introduction**

### **I.1 Introduction**

La capacité de fabriquer des graines est un trait fondamental et spécifique des spermaphytes. Les graines ont un rôle important (i) pour la dissémination des plantes, car elles permettent leur dispersion et leur survie dans des conditions environnementales hostiles (ii) dans l'industrie agroalimentaire, car elles sont la principale ressource pour la nutrition animale et humaine. La compréhension des contrôles génétiques du développement et de la maturation des graines reste donc un domaine de recherche essentiel pour la biologie végétale et l'agriculture.

Les plantes à graines, ou spermaphytes, sont composées de deux grands groupes: les gymnospermes et les angiospermes. Les angiospermes sont constituées d'un grand nombre de plantes (environ 370000 espèces). Le développement de l'albumen des angiospermes, après la fécondation, peut varier et permet de classer les graines en trois sous-groupes, les graines albuminées (ex : blé, orge, Brachypodium), les graines exalbuminées (ex: Brassicaceae) et les graines sans albumen ou endosperme réduit (ex : bananier, Nymphaeaceae, tomate) (Jiao *et al.* 2011). Les gymnospermes forment un groupe plus petit (un peu moins de 1000 espèces) dont les principaux représentants sont les conifères.

#### **I.1.1 Le développement de la graine**

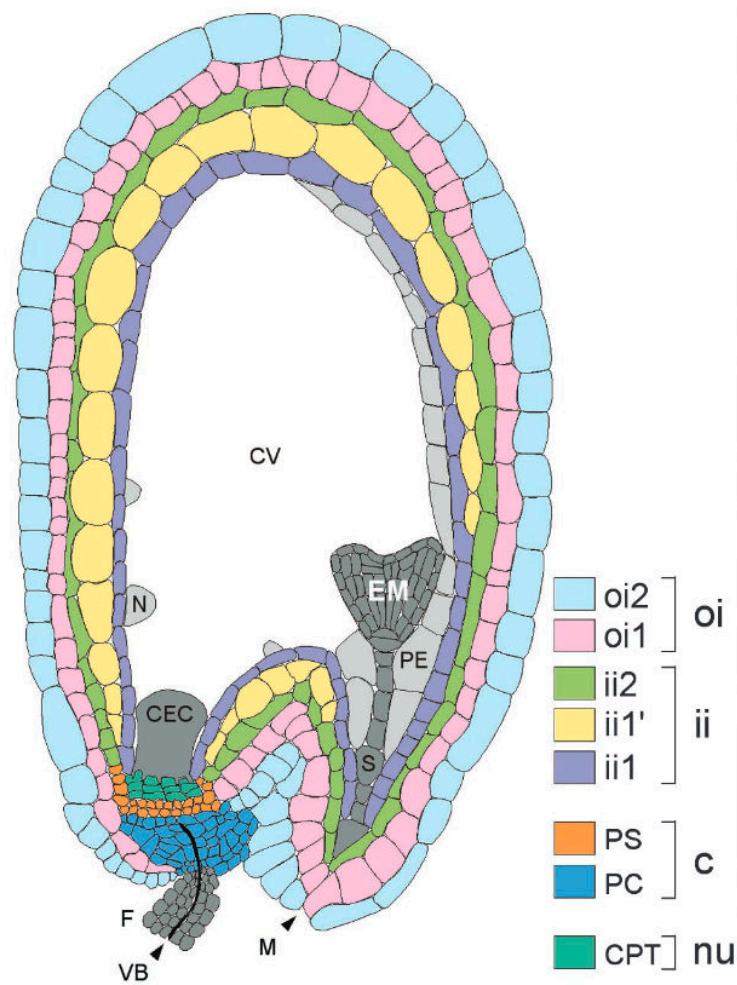
##### **I.1.1.1 La fécondation**

Le processus de formation d'une graine est essentiel dans le cycle de vie des plantes. Après la fécondation, le développement de la graine consiste en deux grandes étapes : l'embryogenèse, la maturation.

Au démarrage, ce processus consiste en une double fécondation par le contact d'un ovule contenant le sac embryonnaire, (ou gamétophyte femelle chez les angiospermes) et des grains de pollen (ou gamétophyte mâle) provenant des anthères et issus de deux divisions des microspores.

Le tube pollinique contenant les deux noyaux mâles va germer et se développer jusqu'à atteindre le sac embryonnaire par l'intermédiaire d'une des synergides. Les deux noyaux mâles vont ainsi atteindre et féconder, (i) l'oosphère qui donnera une cellule diploïde et formera l'embryon, (ii) les 2 noyaux du sac embryonnaire formant ainsi le zygote accessoire, triploïde, qui deviendra l'albumen.





**Figure 1 : Structure de la graine d'*Arabidopsis thaliana* au stade cœur de développement (d'après Debeaujon *et al* 2003).**

C, chalaze; CEC, albumen chalazal CV, vacuole centrale; EM, embryon; F, funicule; ii, téguments internes; M, micropyle; N, nodule; Nu, nucelle; Nu, nucelle; oi, téguments externes; PC, chalaze placentaire; PE, albumen périphérique; PS, couches à pigments; S, suspenseur; VB, faisceau vasculaire.

### **1.1.1.2 Les différents tissus de la graine**

À la suite de la fécondation, un ensemble de divisions cellulaires va permettre la mise en place des tissus : les téguments, l'albumen, puis le développement de l'embryon selon des axes de symétrie définis (Jenik *et al.* 2007; Laux *et al.* 2004). (Figure 1)

Les téguments (ou testa) protègent l'embryon et contribuent à la longévité de la graine (Finkelstein *et al.* 2008). Ils sont dérivés du tissu maternel et composés de différentes couches de cellules. Chez *Arabidopsis thaliana*, la testa est composée de 5 couches cellulaires (Leon-Kloosterziel *et al.* 1994; Schneitz *et al.* 1994) i) des téguments externes (oi : outer integuments) comprenant l'épiderme et le sous-épiderme qui accumulent l'amidon et sécrètent le mucilage (North *et al.* 2010), ii) des téguments internes (ii : inner integuments) qui sont composés de 2 couches cellulaires indifférenciées et adhérentes entre elles ainsi que de l'endothélium qui fait face à l'albumen et est riche en métabolites secondaires (tanins) (Debeaujon *et al.* 2003) (Figure 1).

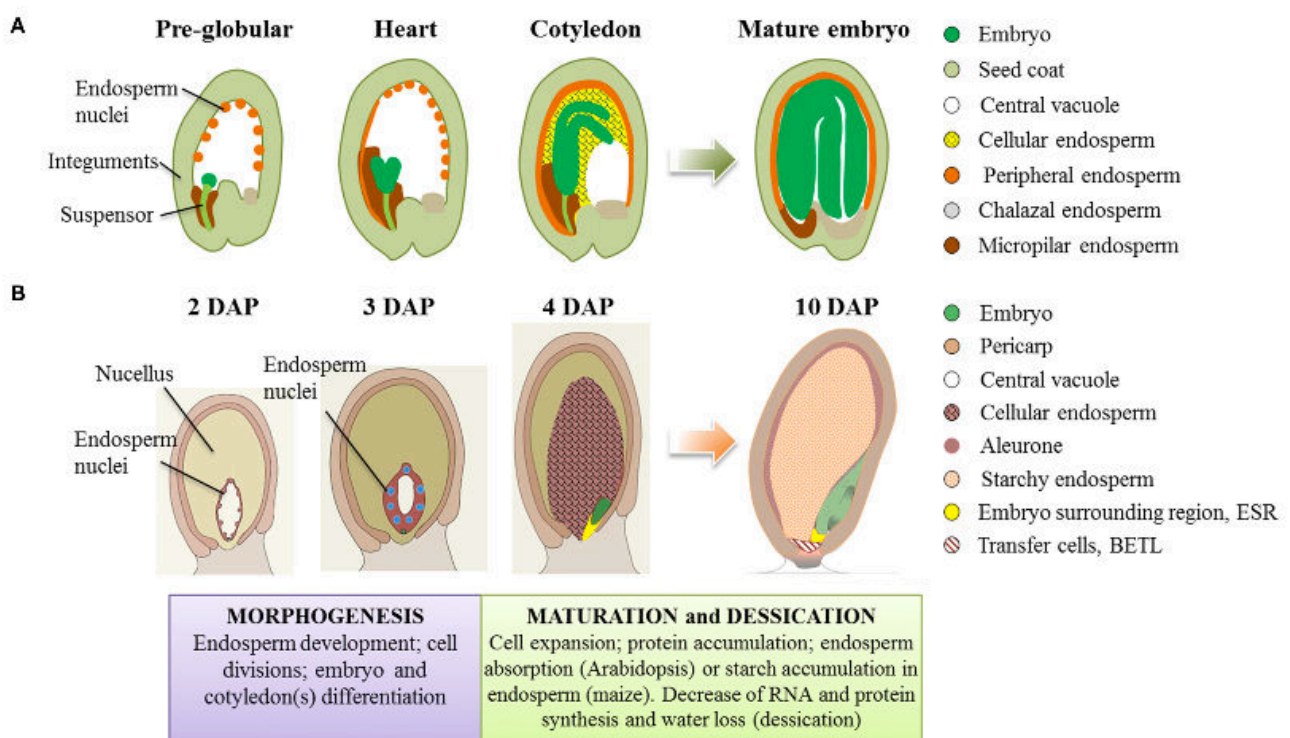
L'albumen se développe d'abord comme un syncytium, puis par une succession de mitoses sans cytokinèse. Au stade globulaire du développement, l'albumen syncytial entoure l'embryon, quelques noyaux se trouvent dans la chalaze (Sorensen 2002). L'albumen va alors se cellulariser, par une compartimentation du syncytium. L'albumen chalazal reste, lui, sous forme syncytiale jusqu'à un stade de développement tardif (Sorensen 2002). Chez les graines albuminées, l'albumen entoure complètement l'embryon. Il contient en général des glucides, de l'amidon et des réserves (protéines/lipides) nécessaires à la germination et à la croissance (Sabelli and Larkins 2009). Les grains de céréales ont une ou plusieurs couches d'aleurones entourant l'albumen, qui expriment les enzymes nécessaires à la dégradation de l'amidon (Sabelli and Larkins 2009).

Dans les graines exalbuminées, l'albumen est réduit à une monocouche autour de l'embryon à la fin de son développement et les réserves de glucides, de lipides et de protéines sont situées très majoritairement dans l'embryon (Baud *et al.* 2002; Vicente-Carbajosa and Carbonero 2005).

Des exemples de différents développements de l'albumen sont décrits dans la figure 2.

### **1.1.1.3 L'embryogenèse**

L'embryogenèse comprend deux étapes, l'embryogenèse précoce et l'organogenèse embryonnaire. Chez les dicotylédones, au cours de l'embryogenèse précoce, plusieurs mitoses permettent la formation de l'embryon globulaire, avec l'acquisition de symétries et de polarités (van Dop *et al.* 2015; Ten Hove *et al.* 2015). À ce stade, l'embryon a acquis la polarité apico-basale et initie le développement des différentes parties constitutives de l'embryon: l'épiderme, le suspenseur, l'hypophyse, l'hypocotyle et les cotylédons. Au stade cœur de l'embryon, les différents organes sont formés. À la fin du stade cœur de l'embryon,



**Figure 2 : Schéma comparé du développement de graines d'Arabidopsis (dicotylédone) et de maïs (monocotylédone) (d'après Locascio *et al* 2014).**

(A) Représentation schématique du développement de la graine d'Arabidopsis. Les étapes de développement de l'embryon sont indiquées. (B) Représentation schématique du développement des graines de maïs. Les étapes indiquent les jours après la pollinisation (DAP).

le méristème apical caulinaire se développe alors jusqu'au stade torpille. Le développement de l'embryon est mis en place sous une symétrie bilatérale, contrôlée par l'auxine (Jenik *et al.* 2007; Laux *et al.* 2004).

Chez les monocotylédones, le développement embryonnaire semble moins structuré et les plans de symétrie sont moins liés à un nombre précis de divisions cellulaires (Guillon *et al.* 2012; Dumas and Rogowsky 2008). Aussi, le développement embryonnaire est principalement décrit chez des dicotylédones où il est particulièrement structuré. Le développement comparé de l'embryon de monocotylédone et de dicotylédone est présenté en figure 2.

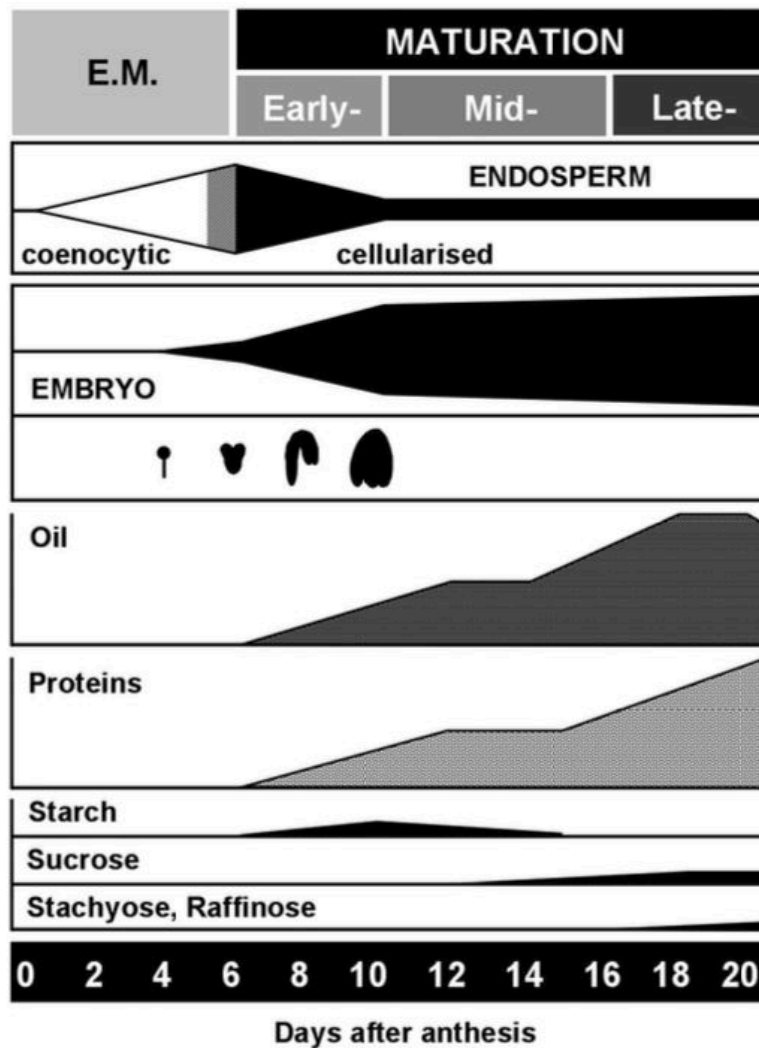
#### ***1.1.1.4 La maturation***

Au stade torpille, l'organogenèse est achevée et la phase de maturation commence. Cette étape couvre la période de développement de 7 à 20 DAF chez *Arabidopsis* et se compose de la croissance de l'embryon, de l'accumulation des réserves (lipides, protéines, glucides), l'acquisition de la tolérance à la dessiccation et de la dormance (Baud *et al.* 2002). De 7 à 12 DAF, la concentration d'amidon diminue et la synthèse des réserves lipidiques et protéiques est amorcée (Figure 3). Une fois que l'embryon a acquis sa taille mature (12-16 DAF), les réserves lipidiques et protéiques sont accumulées majoritairement dans l'embryon. Pendant la maturation tardive (17 à 20 DAF), la synthèse des réserves de lipides diminue tandis que la synthèse protéique augmente jusqu'au dernier stade. L'embryon devient quiescent et acquiert une tolérance à la dessiccation (12-15 DAF) (Ooms *et al.* 1993b; Righetti *et al.* 2015). La graine perd alors de l'eau (de 32% à 10% chez *Arabidopsis*) et accumule des sucres (Baud *et al.* 2002; Baud *et al.* 2008).

##### ***1.1.1.4.1 Le remplissage de la graine***

###### ***1.1.1.4.1.1 Les protéines***

Chez la graine d'*Arabidopsis thaliana*, les principales protéines de stockage (SSP : Seed Storage Protein) sont l'ALBUMINE 2S, ou NAPINE et la GLOBULINE 12S, ou CRUCIFERINE. Les vésicules de réserves protéiques (PSV : Protein Storage Vesicles) contiennent d'autres protéines moins abondantes, comme les protéines de choc thermique (Hsp : Heat Shock Protein) et les protéines LEA (Late Embryogenesis Abundant) (Wehmeyer *et al.* 1996; Tunnacliffe and Wise 2007). Ces protéines sont synthétisées dans le réticulum endoplasmique granuleux et transportées dans les corps protéiques par la voie de transport des vésicules. La maturation des protéines se produit par protéolyse directement dans la vésicule de stockage des protéines (North *et al.* 2010; Baud *et al.* 2008). Les corps protéiques peuvent représenter jusqu'à un tiers du poids total de la graine. L'accumulation des SSPs commence au début de la phase de maturation de l'embryon et se poursuit pendant la maturation tardive, le contenu en protéines dans les graines augmentant jusqu'au



**Figure 3 : Accumulation des différentes réserves au cours du développement de la graine d'*A. thaliana* au stade embryogénèse et aux différentes étapes de la maturation (d'après Baud *et al* 2008).**

Les embryons sont représentés aux stades: globulaire, cœur, torpille et U retourné. Représentation du remplissage de la graine en huile, protéines, amidon et glucides au cours du développement. Le volume relatif de l'endosperme et de l'embryon au sein de la graine est présenté tout au long du processus de développement.

dernier stade de développement, contrairement aux lipides (Figure 3) (Baud *et al.* 2002). Les LEA et les Hsp ont probablement un rôle dans l'acquisition de la tolérance à la dessiccation, mais leurs rôles précis n'ont pas encore été définis (Wehmeyer *et al.* 1996; Tunnacliffe and Wise 2007).

#### *1.1.1.4.1.2 Les glucides*

Le contenu de la graine est constitué de différents composés glucidiques tels que les glucides solubles (hexose, saccharose) et les polymères insolubles (amidon). L'accumulation de l'amidon, principalement située dans le tégument et la radicule de l'embryon, est transitoire chez les Brassicaceae, elle commence pendant l'embryogenèse et se termine à la fin de la maturation précoce. À l'inverse, elle peut être majeure chez les céréales. Les hexoses sont présents au cours des premiers stades de développement (3 DAF) et sont trouvés à l'état de traces au stade mature (Figure 3). Le rapport saccharose / hexose joue un rôle important pour les voies de signalisation qui contrôlent le développement de l'embryon (Baud *et al.* 2002; Gutierrez *et al.* 2007; Baud *et al.* 2008).

#### *1.1.1.4.1.3 Les lipides*

Dans les graines de Brassicaceae, les lipides sont stockés essentiellement sous forme de triacylglycérols (TAG) dans des organelles de réserve nommées corps lipidiques (CL). Ces CLs, possèdent un cœur de lipides neutres recouvert d'une hémi-membrane dans laquelle sont enchâssées des protéines contribuant à la structure du CL. Ces protéines sont principalement les OLEOSINES, la CALEOSINE et la STEROLEOSINE, qui sont ancrées dans la membrane de l'organelle (Jolivet *et al.* 2013). Les CLs sont localisés dans le cytoplasme et sont impliqués dans la mobilisation des TAGs lors de la germination pour fournir de l'énergie à la jeune plantule. La teneur en lipides d'*Arabidopsis thaliana* peut atteindre 40% de la masse de la graine sèche (Baud *et al.* 2002; Baud *et al.* 2008; North *et al.* 2010). La composition en acides gras dans les CLs évolue au cours du développement de l'embryon. Ainsi, les acides gras mono et polyinsaturés se trouvent en petites quantités à 7 DAF et représentent près de 80% de la composition en acides gras dans la graine mature, alors que les acides gras saturés tendent à diminuer avec le développement de l'embryon. À 12 DAF, la composition en acides gras semble être constante chez *Arabidopsis*. L'accumulation des lipides se termine lors de l'établissement de la tolérance à la dessiccation, à la fin de l'étape de maturation (Figure 3) (Baud *et al.* 2002).

#### **1.1.1.5 La maturation tardive et la dessiccation**

L'acquisition de la tolérance à la dessiccation est une étape importante pour la graine, permettant sa survie dans des conditions environnementales difficiles. Ce processus est établi lors de la maturation tardive.

L'acide abscissique (ABA) et les gibbérellines (GAs) sont des phytohormones qui ont des rôles importants dans le développement de la graine et de l'embryon. L'ABA est impliqué dans la réponse au stress, le maintien de la dormance et la tolérance à la dessiccation (Nambara and Marion-Poll 2003). Les GAs sont impliquées dans la germination (Pimenta Lange and Lange 2006; Curaba *et al.* 2004).

Le ratio ABA-GA est important pour la croissance de l'embryon, le contrôle de la maturation et de la germination. Les deux hormones ont un effet antagoniste sur le développement et l'ABA peut réduire le niveau d'expression des GAs (Finkelstein *et al.* 2002; Gutierrez *et al.* 2007; Finch-Savage and Leubner-Metzger 2006).

Lors de la perception de l'ABA par ses récepteurs, des événements précoces impliquant des protéines G et des canaux ioniques font appel à des messagers secondaires (calcium, espèces réactives de l'oxygène et lipides). Ces événements induisent notamment des réactions de phosphorylation et de déphosphorylation de protéines dont le rôle dans les voies de signalisation cellulaire des plantes est essentiel (Kersten *et al.* 2009). Les protéines phosphatases type 2C (PP2Cs) sont par exemple, impliquées dans la perception du signal en condition de stress. De nombreux mutants ABA insensibles (*abi*) ont été identifiés par leur perte de sensibilité à l'ABA : les gènes *abi1* et *abi2* codent pour des protéines de la famille des PP2Cs, ils ont été identifiés comme des régulateurs négatifs de la voie de transduction du signal de l'ABA (Luan 2003; Ludwikow 2015).

ABI5 est un facteur de transcription de la famille des bZIPs. Il est fortement impliqué dans la maturation tardive et est un régulateur positif de la voie de signalisation de l'ABA. Il a aussi été démontré qu'ABI5 régule l'expression des LEAs (Carles *et al.* 2002). ABI3 (ABCISSIC ACID INSENSITIVE 3), un facteur de transcription à domaine B3 d'une importance clef dans le contrôle de la dormance, de la transduction du signal ABA, de la résistance à la dessiccation et de l'accumulation des réserves, est également impliqué dans la sensibilité à l'ABA. ABI3 fait partie de la famille des AFL (ABI3, FUS3, LEC2) que nous décrirons plus en détails dans les chapitres suivants.

## I.2 Revue

### I.2.1 Introduction

Les facteurs de transcription ont pour rôle le contrôle de l'expression de gènes à des stades particuliers du développement de la plante et de la graine (Gong *et al.* 2004).

De nombreux facteurs de transcription (FTs) sont impliqués dans le développement de la graine. Certains contrôlent la prolifération et l'expansion cellulaire, comme par exemple TRANSPARENT TESTA GLABRA2 (TTG2) et APETALA 2 (AP2) qui ont un rôle dans le contrôle de la taille de la graine (Orozco-Arroyo *et al.* 2015). D'autres FTs sont impliqués dans la synthèse d'acides gras, tels que WRINKLED1 (WRI1), en contrôlant l'expression des gènes codants des enzymes du métabolisme lipidique impliqués dans le remplissage de la graine (Baud and Lepiniec 2009).

Les facteurs de transcription LAFL : LEAFY COTYLEDON1, ACID ABSCISIC INSENSITIVE3, FUSCA3, LEAFY COTYLEDON2, régulent différents aspects du développement et ont été identifiés comme facteurs de transcription maîtres du développement de la graine et, particulièrement, de l'embryon (Santos-Mendoza *et al.* 2008). Les LAFL sont étudiés depuis de nombreuses années et le chapitre suivant (article de revue) vise à regrouper les connaissances accumulées sur l'étude des LAFL dans le contrôle du développement de la graine : leur structure et leur fonction, leur rôle de régulateur transcriptionnel et leur mode d'action.

Dans un second chapitre de la revue, l'étude de l'évolution des gènes LAFL dans le règne végétal est présentée afin de caractériser l'ancêtre commun des LAFL et l'évolution des facteurs au cours de l'évolution des plantes.





## **I.2.2 Revue: Regulation and evolution of the interaction of the seed B3 transcription factors with NF-Y subunits.**

Accepté dans BBA Gene Regulatory Mechanisms

Boulard C., Fatihi A., Lepiniec L. and Dubreucq B.\*

Institut Jean-Pierre Bourgin (IJPB), INRA, AgroParisTech, ERL-CNRS, Saclay Plant Sciences (SPS),  
Université Paris-Saclay, RD10, F-78026 Versailles, France

\*To whom correspondence should be addressed.

E-mail [bertrand.dubreucq@inra.fr](mailto:bertrand.dubreucq@inra.fr); Tel.: 33-1 30 83 30 66

Key words : transcription factor, NF-YB, regulation, seed, LEC

### **Abstract**

The LAFL genes (LEC1, ABI3, FUS3, LEC2) encode transcription factors that regulate different aspects of seed development, from early to late embryogenesis and accumulation of storage compounds. These transcription factors form a complex network, with members able to interact with various other players to control the switch between embryo development and seed maturation and, at a later stage in the plant life cycle, between the mature seed and germination.

In this review, we first summarize our current understanding of the role of each member in the network in the light of recent advances regarding their regulation and structure / function relationships. In a second part, we discuss new insights concerning the evolution of the LAFL genes to address the more specific question of the conservation of LEAFY COTYLEDONS 2 in both dicots and monocots and the putative origin of the network. Last we examine the current major limitations to current knowledge and future prospects to improve our understanding of this regulatory network.

### **1- Introduction :**

After fertilization, seed development is divided into two main phases, embryogenesis and maturation. In dicotyledonous plants, during early embryogenesis, several mitoses together with auxin flows lead to the acquisition of symmetries (lateral, radial, ...) and polarities (van Dop *et al.* 2015; ten Hove *et al.* 2015). The embryo then develops in relation to these axes of symmetry (Jenik *et al.* 2007; Laux *et al.* 2004; Bassel *et al.* 2014). When organogenesis is completed, the maturation phase starts. This stage comprises embryo growth, the accumulation of reserves (lipids, proteins, and carbohydrates), and the acquisition of desiccation tolerance and dormancy (Baud *et al.* 2008). At the end of its development, the seed loses water to reach 5 to 10% of seed weight and the embryo becomes quiescent.

Spatial and temporal resolution of mRNA accumulation during seed development, showing very precise expression patterns for specific gene sets, has suggested that transcriptional regulation is the basis of functional differentiation (Belmonte *et al.* 2013). On the same line, the role of transcriptional regulatory network on the control of seed metabolism and development has also been shown (Sreenivasulu *et al.* 2008). Mutations of some highly conserved transcription factors (TFs) have been shown to deeply impact their embryogenesis and ectopic expression of certain TFs can activate seed genes and confer seed metabolism or traits to vegetative tissues, supporting the key role of transcriptional control by these TFs during seed development (Braybrook *et al.* 2006).

Bioinformatics studies identified 1653 *Arabidopsis* genes coding for TFs (<https://www.araport.org>) grouped in 38 different families (Gong *et al.* 2004; Qu and Zhu 2006; Riechmann *et al.* 2000). Of these 38 families, 11 were shown to be plant-specific, such as AP2-EREBP, NAC, WRKY or B3-type (Riechmann *et al.* 2000; Qu and Zhu 2006).

Among these TFs, 1327 are expressed during seed development (Le *et al.* 2010). A few of them are involved in determining seed size by controlling cell proliferation, expansion or differentiation in the integuments or the endosperm (table 1); these include *AUXIN RESPONSE FACTOR 2 (ARF2)*, *TRANSPARENT TESTA GLABRA 2 (TTG2)*, *SEEDSTICK (STK)*, *APETALA2 (AP2)*, *AGAMOUS LIKE 62 (AGL62)* or *MINISEED 3 (MINI3)* (Orozco-Arroyo *et al.* 2015). Other TFs involved in embryo development have essential roles in the normal development of the seed, such as the TFs of the B3 DNA-Binding domain family *LEAFY COTYLEDON 2 (LEC2)*, *ABSCISIC ACID INSENSITIVE 3 (ABI3)*, and *FUSCA3 (FUS3)*, together with the NF-YB *LEAFY COTYLEDON 1 (LEC1)* and *LEAFY COTYLEDON-Like1 (L1L)*, as described in previous reviews (Vicente-Carbajosa and Carbonero 2005; Braybrook and Harada 2008; Santos-Mendoza *et al.* 2008). These TFs are collectively referred to as

the LAFL proteins. Their involvement in different regulatory pathways was shown by genetic analyses and molecular studies of their expression patterns at different stages of seed development.

All these TFs recognize specific DNA cis-elements, such as RY sequences for the B3-type TF, (Yamasaki *et al.* 2013), or AW grounds for the AP2-EREBP family (Maeo *et al.* 2009). The MYB family has an affinity for AC-rich sequence motifs (Barthole *et al.* 2014; Prouse and Campbell 2012; Kelemen *et al.* 2015), bZIPs recognize G-Boxes or ABRE elements (ABA Response element) (Cheng *et al.* 2014; Jakoby *et al.* 2002), and MADS boxes recognize the CArG-Box (de Folter and Angenent 2006). All these elements are listed in table 2. Specific binding of these transcription factors on the target DNA cis-element can lead to either transcription of the downstream sequence, or to its repression, as is the case of the ASIL1 trihelix member that recognizes GT elements and represses gene expression (Gao *et al.* 2009).

Although the LAFL genes show similar and overlapping functions, their expression patterns differ during seed development (Sreenivasulu and Wobus 2013; To *et al.* 2006). LEC1 and LEC2 are expressed during the first period of embryonic development, from 4 to 15 DAP and 3 to 14 DAP, respectively (Fig 1). LEC1 is expressed to higher levels than LEC2, peaking at 8 DAP and 10 DAP, respectively. LEC1 and LEC2 messenger RNAs are found in both the embryo and the endosperm (Kroj *et al.* 2003; Lotan *et al.* 1998). LEC1 like (L1L) is expressed throughout development of the embryo, weakly during embryogenesis and more strongly during the maturation phase, both in the embryo and the endosperm (Kwong *et al.* 2003; Lotan *et al.* 1998). Nevertheless, the ectopic expression of L1L complements the phenotype of *lec1*, suggesting a redundant function of both proteins. ABI3 is expressed from 3 DAP up to the end of the maturation phase, and *in situ* hybridization has shown localization in the entire embryo (To *et al.* 2006). The expression of FUSCA3 starts at 3 DAP and lasts until of the end of the maturation phase (Kroj *et al.* 2003), and is exclusively localized in the epidermis (Gazzarrini *et al.* 2004). These specific expression patterns may explain the maintenance of similar genes, with redundant protein activities (subfunctionalization). In addition, they may be involved in different and specific complexes depending on their tissue accumulation, as is the case for the MYB/bHLH/WD40 complexes (Lepiniec *et al.* 2006).

Last, at the end of seed development, the acquisition of desiccation tolerance is an important step for the seed, allowing it to survive under unfavourable environmental conditions. This process is established during late maturation. Abscisic acid (ABA) and gibberellins (GAs) are phytohormones that

have significant antagonistic roles in seed development and germination (Curaba *et al.* 2004; Pimenta Lange and Lange 2006; Finkelstein *et al.* 2008). GAs are involved in embryo development and germination, while ABA is important for the acquisition of tolerance to desiccation and inhibition of germination (Phillips *et al.* 1997; Finkelstein *et al.* 2002; Finkelstein *et al.* 2008; Gutierrez *et al.* 2007). In the seed, ABA represses GA synthesis and induces dormancy (Kanno *et al.* 2010).

## **2- The LAFL transcription factors**

### **2.1- AFL-B3 functions.**

The LAFL proteins control different aspects of seed development, from the formation of the embryo to the accumulation of storage proteins and lipids. Mutations in each *LAFL* genes results in related but distinct phenotypes, suggesting overlapping but specific functions in seed development and reserves accumulation (Meinke *et al.* 1994, To *et al.*, 2006, Roscoe *et al.* 2015). In particular, the cotyledons of the corresponding mutants are strongly affected in the accumulation of these storage compounds and display several characteristic traits of leaves, such as trichomes on their surface and anthocyanin accumulation (Stone *et al.* 2001; Meinke *et al.* 1994). It has to be noted that the complex phenotypes observed in *lafl* mutants (from reserve accumulation to anthocyanin and lack of trichomes) could be the common effects of mutations affecting globally organ or cell fate. Under this perspective, if the cotyledon turns to true leaves, it is just normal that it harbours trichomes and do not accumulate reserves. In the same way, accumulation of anthocyanins is often observed during leaf stress. LAFLs would in that case be involved in the specific switch between embryogenesis and maturation, their absence leading to a direct jump into a “post germination stage” (Keith *et al.* 1994; Yamamoto *et al.* 2014).

The LAFL regulators are involved in the accumulation of lipid and protein reserves during embryo development. LEC2 regulates fatty acid (FA) biosynthesis through the activation of the WRI1 gene, which encodes an AP2-type transcription factor (Angeles-Nunez and Tiessen 2011; Baud *et al.* 2007) that controls the expression of several genes coding for enzymes of FA biosynthesis (Maeo *et al.* 2009). It also controls the expression of MYB118 and MYB115, which are involved in the biosynthesis of specific FA (omega-7 monounsaturated) in the endosperm (Barthole *et al.* 2014; Troncoso-Ponce *et*

*et al.* 2016). The NF-YB LEC1 and L1L genes are also involved in FA biosynthesis, probably through the activation of LEC2, ABI3 and FUS3 (Baud *et al.* 2007; Mu *et al.* 2008). Interestingly, LEC1 and L1L have been shown to interact with NF-YC2 and bZIP67 to regulate FA biosynthesis (Yamamoto *et al.* 2009b; Mendes *et al.* 2013). Similar interactions also control the biosynthesis of storage proteins (Baud *et al.* 2016; Kroj *et al.* 2003; Yamamoto *et al.* 2009). For instance, ABI3 can form regulatory complexes with bZIPs 10, 25 and 53 to activate the expression of the genes encoding abundant LEA storage proteins, cruciferins and oleosins (Monke *et al.* 2012; Lara *et al.* 2003; Alonso *et al.* 2009). On the same line, we have shown that LEC1, LEC2, ABI3 and NF-YC2 can synergistically activate the expression of the OLEOSIN 1 gene (Baud *et al.* 2016). In conclusion, the LAFL genes are involved in different aspects of seed development and reserve accumulation in an intricate network, as shown in figure 2 that highlights this complexity.

In addition LEC2 and LEC1 have the intriguing ability to induce somatic embryogenesis when ectopically expressed in vegetative or embryonic tissues (Gaj *et al.* 2005; Stone *et al.* 2008; Stone *et al.* 2001), suggesting a broader function in determining embryonic fate. This ability to induce somatic embryogenesis could be related to auxin metabolism (Bai *et al.* 2013). Indeed, LEC1 and LEC2 control the expression of YUCCA genes, which are involved in auxin biosynthesis (Stone *et al.* 2008; Wojcikowska *et al.* 2013; Cheng *et al.* 2006; Junker *et al.* 2012). Furthermore, auxin also activates the expression of FUS3 (Gazzarrini *et al.* 2004). ABI3 is involved in embryogenesis, early and late maturation (Parcy *et al.* 1997; Kanno *et al.* 2010). Abscisic Acid (ABA) has a positive role in the expression of ABI3 (Cheng *et al.* 2014; Lopez-Molina *et al.* 2002), which is also involved the establishment of desiccation tolerance (Lopez-Molina *et al.* 2002). Therefore, seed embryogenesis and maturation are intricate processes involving several interacting plant hormones, the effects of which remain difficult to unravel. This complex regulation is exemplified by different regulatory loops involving LAFL members, ABA and gibberellic acids (GA). LEC2 and FUS3 both repress GA biosynthesis directly and indirectly, FUS3 up regulates ABA accumulation, while GA has been shown to repress FUS3 expression (Gazzarrini *et al.* 2004; Curaba *et al.* 2004; Braybrook *et al.* 2006; Wang *et al.* 2004). Last, the transcription factor WRKY41, which induces the expression of ABI3 during seed maturation, is negatively regulated by ABA (Ding *et al.* 2014; Kanno *et al.* 2010).

Although the LAFL contribute to the same developmental programs, they display various specificities (figure 1). For instance, LEC2, LEC1 and ABI3 have synergistic activities on the expression of an

oleosin promoter, but not with FUS3 (Baud *et al.* 2016). The different expression patterns of LEC2, LEC1 and ABI3 (To *et al.* 2006; Cheng *et al.* 2014; Kroj *et al.* 2003), together with genetic analyses, suggest that LEC1 and LEC2 are primarily initiators of embryonic fate, whereas ABI3 would be an effector driving efficient reserve storage and acquisition of desiccation tolerance during late maturation. In this scheme, FUS3, which is more specifically expressed in the epidermis, could be involved in the maintenance of the epidermis in an “embryo-compatible stage”.

## **2.2- B3 protein structure:**

The plant-specific B3 DNA-binding domain is composed of approximately 100 amino acids, and is found in various characterized or putative transcription factors. (Wang *et al.* 2012; Yamasaki *et al.* 2004). It has been functionally characterized in ABI3 and VP1 in Arabidopsis and maize, respectively (Giraudat *et al.* 1992; Suzuki *et al.* 1997). The first structure of this domain was established for RAV1, which belongs to the REM sub-family (Yamasaki *et al.* 2004). The B3 domain consists of 7  $\beta$  sheets and 2  $\alpha$  helices in a  $\beta$ -barrel-like structure (Yamasaki *et al.* 2004; Waltner *et al.* 2005). The RAV1 B3 domain shows structural similarities with endonucleases EcoRII-N and C-Bfil, suggesting divergence from a common ancestor (Golovenko *et al.* 2014; Yamasaki *et al.* 2004; Yamasaki *et al.* 2013). It recognizes the 6 bp DNA consensus sequence CAYSYR (also called RY), where Y is C or T, S is G or C, R is A or G (Yamasaki *et al.* 2004). Nevertheless, other B3 transcription factors can recognize a wide range of RY-type cis elements (Yamasaki *et al.* 2013; Baud *et al.* 2016).

The comparison of the structure of RAV 1 B3-domain with an endonuclease has allowed precise prediction of the interaction between B3 domains and RY DNA cis elements (Yamasaki *et al.* 2004; Golovenko *et al.* 2014), but resolution of the B3 structure in the context of an entire protein and its target DNA is still required to confirm these predictions.

## **2.3- NF-Y protein structure.**

LEC1 is homologous to the CCAAT box - binding factor (CBF) A subunit also known as HAP3 or NF-YB (Kirkbride *et al.* 2013; Lotan *et al.* 1998). CBF and HAP are orthologous complexes initially found in mammals and yeast, respectively, renamed NF-Y for all eukaryotes (Mantovani 1999). These transcription factors bind DNA in trimeric complexes composed of different NF-YA (HAP2/CBF-B), NF-YB (HAP3/CBF-A) and NF-YC (HAP5/CBF-C) subunits (Mantovani 1999; Dolfini *et al.* 2012). In plants, the complexes are composed of the NF-YA subunits, which bind to DNA (10 genes), NF-YB (13 genes) and NF-YC (13 genes) (Mantovani 1999; Riechmann *et al.* 2000; Siefers *et al.* 2009). NF-YB and NF-YC possess highly conserved domains, which are similar to the histone fold motif (HFM) (Hackenberg *et al.* 2012; Lee *et al.* 2003; Mantovani 1999). NF-YAs consist of a B / C-binding domain, allowing the interaction with other members of the NF-Y complex, and a DNA binding domain (Hackenberg *et al.* 2012; Siefers *et al.* 2009).

Plant NF-YB homologues were first characterized in corn and oilseed rape by sequence similarities with the mammalian B subunits (Albani and Robert 1995; Li *et al.* 1992) and then functionally in *Arabidopsis*, as LEC1 (NF-YB9) and L1L (NF-YB6) (Kwong *et al.* 2003; Lee *et al.* 2003; Lotan *et al.* 1998; Siefers *et al.* 2009). The kinetics of formation of the NF-Y complexes show that NF-YB and NF-YC first bind together with high affinity, NF-YA is then recruited, conferring on the trimer the capacity to bind DNA (Fleming *et al.* 2013; Hackenberg *et al.* 2012). NF-YC subunit appears to be an intermediate in the interaction between NF-YA and NF-YB (Siefers *et al.* 2009). Two-hybrid studies have shown that LEC1 and L1L (NF-YB subunits) preferentially interact with NF-YC, without interacting with NF-YA, providing some explanations for the specific function of the LEC1-type proteins (Calvenzani *et al.* 2012; Hackenberg *et al.* 2012). Nevertheless, AtL1L, as other NF-YB subunit can form a canonical Histone Fold Domain (HFD) with some NF-YA and C subunits (Gnesutta *et al.* 2017). A comparative study of amino acid sequences of rat, yeast and *Arabidopsis* NF-YB genes identified amino acids that are critical for the activity of LEC1 and L1L (Lee *et al.* 2003, Yang *et al.* 2005; Kwong *et al.* 2003). Among them, an aspartic acid in position 55 (D55) is a primary importance for LEC1 activity *in planta* (Lee *et al.* 2003) and Histine in position 50 could be involved in specific contacts with DNA in relation to this D55 (Gnesutta *et al.* 2017). It has been suggested that this specific D55 might also provide selectivity for different target DNA sequences (Nardone *et al.* 2016). NF-YB proteins do not bind to DNA in sequence specific process, this function being usually ensured by NF-YA subunits that specifically allows binding to target DNA motifs, but have non sequence specifi interactions with



DNA (Romier *et al.* 2003; Hackenberg *et al.* 2012; Petroni *et al.* 2012). In plants, the high number of NF-Y genes was already mentioned, suggesting more diverse functions compared to mammals. Although LEC1 and L1L interact with NF-YA and C, as do other NF-YB members (Gnesutta *et al.* 2017), recent evidences suggest they also probably interact with other transcription factors to bind DNA, such as B3 or bZIP proteins (Baud *et al.* 2016; Yamamoto *et al.* 2009). Among this line, transcriptomic analysis, have recently pointed the large variability of gene sets regulated by LEC1 and its putative partners (Pelletier *et al.* 2017). The role and the function of heteromers between LEC1/NF-YC and other transcription factor is not fully understood and the structure of the heteromer could be, or not, slightly different from the published NF-Y A/B/C complex, both in term of protein protein and protein/DNA interactions.

#### **2.4- Regulation of LAFL expression:**

Regulation of the expression of the LAFL genes is now well described (To *et al.* 2006; Santos-Mendoza *et al.* 2008; Parcy *et al.* 1997; Monke *et al.* 2012). The LAFLs form regulatory loops in which LEC1 and LEC2 can activate their own expression as well as the expression of FUS3 and ABI3, depending on the tissues considered. ABI3 and FUS3 also regulate their own transcription positively. However, localization and expression data are based so far on mRNA accumulation, while protein regulation (*e.g.*, post-transcriptional control and protein lifetime) could certainly be important. For instance, ABI3 itself appears to be regulated by degradation through the proteasome pathway (Lynch *et al.* 2012; Kirchler *et al.* 2010). The AIP2 E3 ligase is found in all tissues of the plant, its expression is induced by ABA applications, and it contains a RING motif that can polyubiquitinate ABI3 on its C terminus *in vitro* (Zhang *et al.* 2005). Similarly, AIP2 is able to target FUS3 for degradation in the protoderm of the developing embryo at a specific phase during embryogenesis (Duong *et al.* 2017). Furthermore, AKIN10, a Snf1-related kinase, was shown to physically interact and phosphorylate FUS3 to regulate developmental phase transitions (Tsai and Gazzarrini 2012). Identification of the partners involved *in vivo* in the regulation of the LAFL target genes would certainly be a crucial element for better understanding the regulations involved.

As ectopic expression of LEC1 and LEC2 can trigger embryo development and other abnormal phenotypes (Suzuki *et al.* 2007; Gao *et al.* 2009; Jia *et al.* 2013; Chen *et al.* 2015), it is necessary to strictly control LAFL genes expression to avoid unwanted expression of embryonic genes in other organs or stages of development. Thus, coordinated repression of the LAFLs is triggered during the seed maturation to germination phase transition, and the inactivation of LEC1 and LEC2 is maintained during the following stages of plant development. It has been proposed that miRNAs prevent expression of differentiation-promoting transcription factors and promote proper embryonic patterning (Nodine and Bartel 2010). miRNAs were shown to target LEC1, LEC2 and FUS3 directly or indirectly and to suppress their activities in the embryo proper, during embryo development, until they are required (Willmann *et al.* 2011; Vashisht and Nodine 2014).

The transcription factors ASIL1 and ASIL2 inhibit LAFL expression during embryo development and after germination (Gao *et al.* 2009; Willmann *et al.* 2011). Similarly the VALs (HSI/HSL) B3-type TFs are also involved in the negative regulation of embryonic development (Schneider *et al.* 2016; Jia *et al.* 2014; Suzuki and McCarty 2008) and mediate repression of the LAFL network during germination (Jia *et al.* 2013; Suzuki *et al.* 2007). VAL1, VAL2 and VAL3 belong to the B3-domain-containing family of transcription factors, but contain other domains including an EAR (ERF Associated Repression) repression domain, a PHD-L (Plant HomeoDomain-Like) domain (in VAL1 and VAL2) and a CW-Zf domain that was shown, in VAL1, to interact with H3K4me3 marks (Suzuki and McCarty 2008; Schneider *et al.* 2016). HDA19, a histone deacetylase, was shown to directly interact with the CW-Zf domain of VAL2 to repress the expression of seed maturation genes, including LEC1 and LEC2 (Zhou *et al.* 2013). The H3K27me3 repressive marks found on LAFL loci in vegetative tissues, suggest that LAFL transcriptional repression is epigenetically regulated. The Polycomb Repressive Complex (PRC2) was shown to regulate the expression of AFLs (Kohler *et al.* 2003; Liu *et al.* 2016). Although the precise mechanisms involved are not fully understood, the fine functional dissection of LEC2 promoter has highlighted that the transcriptional regulation of LEC2 is linked to a precise repressive element and controlled by repressive Histone H3K27 tri-methylation marks (Berger *et al.* 2011).

### **3- LAFL Networks and Evolution:**

#### **3.1- Evolution of B3 AFLs**

The phylogeny of the B3 superfamily has been investigated and shows 5 subfamilies: Auxin Response Factors (ARF), Abscisic Acid Insensitive3 (ABI3), High level of Sugar Inducible (HSI)/called VAL, Related to ABI3/VP1 (RAV) and Reproductive Meristem (REM) (Romanel *et al.* 2009). In Arabidopsis, 118 genes with a B3 domain have been found, and the 3 AFL-B3 genes involved in seed maturation programs (LEC2, ABI3 and FUS3) all belong to the same phylogenetic group.

Interestingly, a B3-domain-containing protein was found in the green alga *Chlamydomonas reinhardtii*, suggesting a common ancestor for the ABI3 and HSI families (Romanel *et al.* 2009; Wang *et al.* 2012). ABI3/VP1 has been identified in many clades, and is considered as a probable ancestor for the B3 family. In the moss *Physcomitrella patens* only ABI3-like genes have been identified; like in angiosperms, they activate PpLEA promoters in the presence of ABA (Marella *et al.* 2006) (Figure 3), suggesting that the ancestral role of ABI3 was linked to desiccation tolerance (Khandelwal *et al.* 2010). FUS3 was identified in both gymnosperm and angiosperm genomes. Functional copies have been characterized in eudicots and in monocots (Moreno-Risueno *et al.* 2008; Luerksen *et al.* 1998) and FUS3 could have emerged from duplication of ABI3 followed by neo- and/or sub-functionalization, as it exhibits a much more specific expression pattern (epidermis) than ABI3 (Gazzarrini *et al.* 2004). Since FUS3 was only found in seed plants, it would have appeared quite late in the LAFL family (Han *et al.* 2017).

LEC2 was first identified in Arabidopsis (Stone *et al.* 2001) and has been functionally characterized, but only in dicot species such as brassicacea, castor bean (Kim *et al.* 2013) or cacao (Zhang *et al.* 2014; Li *et al.* 2010). It has been proposed that LEC2 is a duplication of FUS3 that occurred either before or after the monocot/dicot separation (Li *et al.* 2010). Nevertheless, recent studies do not support this hypothesis. First, it has been suggested that FUS3 originated later than ABI3 and LEC2 (Han *et al.* 2017). Moreover, LEC2 and ABI3 can form a complex with LEC1 on the OLEOSIN 1 promoter, which is not the case for FUS3 (Baud *et al.* 2016). Nevertheless, one cannot exclude that LEC2 and FUS3 have partially diverged in cereals, as suggested for maize B3 genes, with specialization for both LEC2 and FUS3 either in the embryo or in the endosperm (Grimault *et al.* 2015). Additional work needs to be performed to investigate the ability of FUS3 to bind to LEC1 or L1L.

Rather, it has been proposed that LEC2 recently arose from a duplication of ABI3, after the monocot and dicot split or, alternatively, LEC2 may have been lost in monocots after duplication (Han *et al.* 2017). Indeed, the existence of a LEC2 homolog in monocot has been suggested in grasses (*Oryza sativa*, *Brachypodium distachyon* and *hordeum vulgare*) (Carbonero *et al.* 2016) but this is still controversial according to differences in gene structure, phylogenetic position and expression patterns (Han *et al.* 2017).

### **3.2- Evolution of NF-YB LEC1**

As LEC1 and L1L have demonstrated specific functions in seed development and the control of embryonic fate, the origin of these two members was investigated. LEC1 and L1L contain a specific aspartate (D) residue instead of a Lysine (K) found in the B domain of the other NF-YB proteins, which is important for interactions with other transcription factors, including B3 AFLs (Lee *et al.* 2003; Baud *et al.* 2016). A positive selection for the presence of this D residue was suggested in LEC1-type genes, contributing to the maintenance of the distinction between LEC1 and non-LEC1-type genes (Cagliari *et al.* 2014). LEC1-type genes have been identified in the lycophyte *Selaginella moellendorffii* but not in the moss *Physcomitrella patens* suggesting that LEC1-type genes originated after the emergence of vascular plants (Cagliari *et al.* 2014; Xie *et al.* 2008). It has been proposed that LEC1 genes originated from an ancestral NF-YB, after several rounds of duplication, neo-functionalization (change from D to K) and /or sub-functionalization (expression in the seed). This neo-functionalization would have allowed LEC1-type genes to acquire specific functions in tolerance to desiccation in vegetative organs of nonseed plants (figure 3). The function of LEC1-type genes in gymnosperms development is still a matter of debate, but could be also related to the response to water stress (Xie *et al.* 2008). The LEC1-type genes form a monophyletic lineage in the NF-YB family that shows an asymmetric divergence in the sequences when compared to non-LEC1-type, suggesting a significantly different evolution rate following gene duplication (Yang *et al.* 2005). The common role of LEC1-type genes both in seed development and the response to water stress could be explained by co-option of the already existing water-deficiency-resistance trait, probably based on abscisic acid (ABA). It would probably also have associated the early neo-functionalized LEC1-type genes with the seed formation

program that is, to many aspects, an elaborated form of drought resistance for the mother plant's progeny (Xie *et al.* 2008; True and Carroll 2002).

It is remarkable that LEC1-type genes and B3 AFLs are both groups of genes initially involved in the tolerance to desiccation. LEC1 and LEC1-like are surprisingly capable of interactions with proteins that do not belong to the NF-Y family (Hobo *et al.* 1999; Huang *et al.* 2015; Yamamoto *et al.* 2009; Baud *et al.* 2016). It would be interesting to investigate if the association between LEC1 and the B3 members of the AFL family is a specific characteristic acquired early during evolution, in which case it would have initially been involved in the resistance to water stress in vegetative organs, or if it appeared later, in seed plants, specifically for the formation of the seed.

#### **4) Conclusions and prospects**

1) The interactions between LAFL proteins, as well as with other transcription factors and hormones, form an intricate and robust regulatory network controlling not only seed reserve accumulation but also early seed development.

2) The core of the network is based on the LAFL genes, which are considered to be master regulators and share partially overlapping functions.

3) The network is regulated at the transcriptional and post-translational levels, and probably by other not fully understood regulatory processes (chromatin based, control of subcellular localisation, other members of the complexes).

The possibility that a LAFL complex would not only transcriptionally activate target gene expression, but also modify chromatin, as some putative partners share similarities with chromatin remodelers, is very exciting. This would pave the way to consider these factors as pioneer transcription factors *i.e.* factors that contribute to open chromatin or access silent chromatin and initiate cell-fate changes, as they are also involved in many aspects of embryonic fate (Zaret and Carroll 2011; Drouin 2014, Lu *et al.* 2014, Oldfield, *et al.* 2014).

Filling these knowledge gaps is a goal to achieve for better comprehension of the fine mechanisms of seed maturation. This will require more precise analyses (*i.e.* at the cellular level) of specific gene

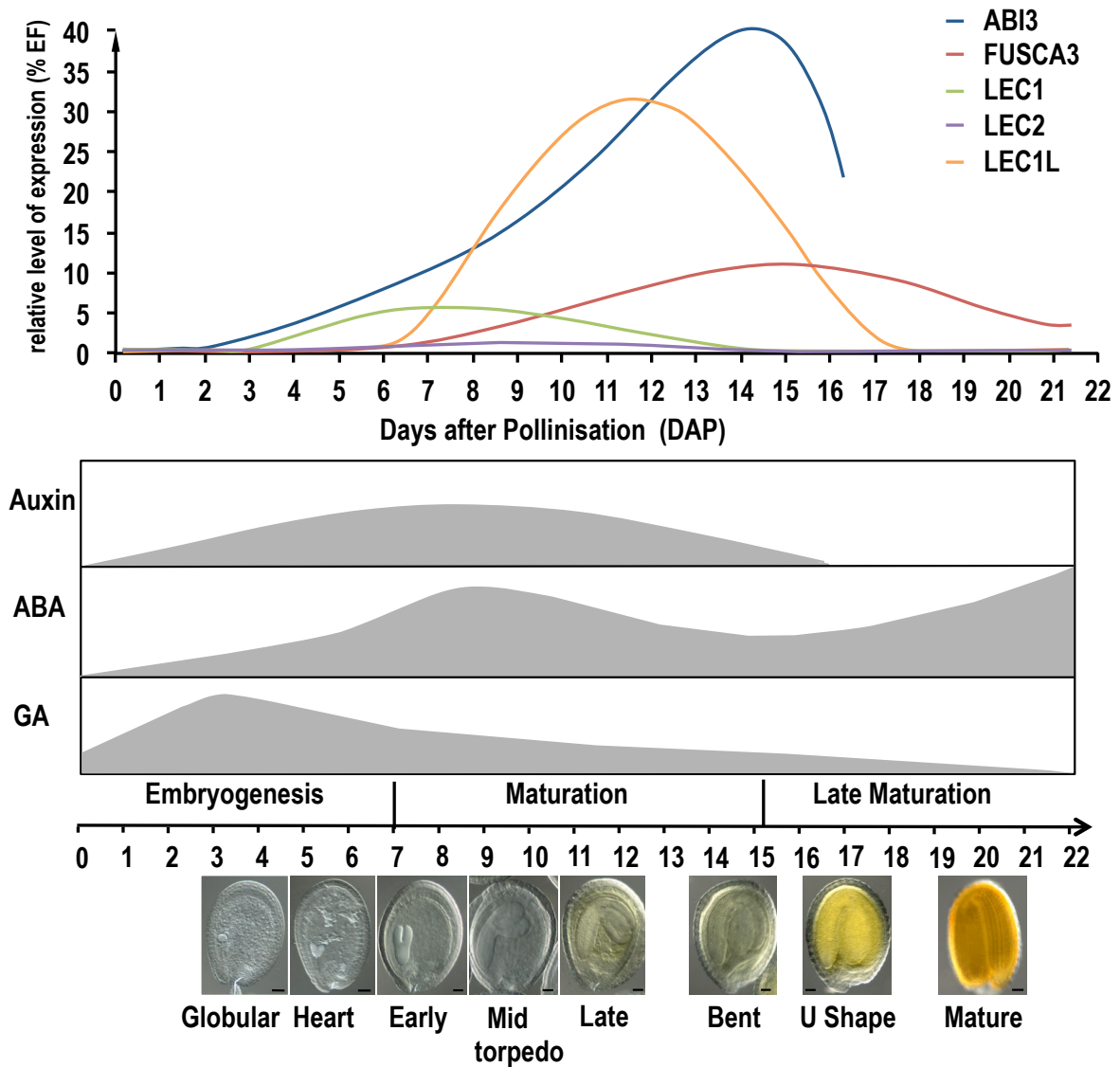
expression, chromatin state, but also classical hormone and metabolome content to bridge transcriptional regulation and reserve accumulation.

This network has evolved since the emergence of land plants, and then probably expanded in seed plants from pre-existing mechanisms of water stress tolerance, to achieve a high level of complexity in the current seed plants (Schallau *et al.* 2008). Over recent years, a great effort was made to characterize the direct and indirect targets of the transcription factors involved, but knowledge about the underlying biochemical mechanisms is still fragmentary. Considerable work will be necessary to confirm certain interactions observed *in vitro* between proteins, and to establish unequivocally (i.e with the tools of protein biochemistry) which partners are involved in planta in a given complex. Further characterization of the protein complexes will be useful for understanding the fine-tuning of reserve accumulation in seeds, and should be useful for increasing or modifying seed reserves content.

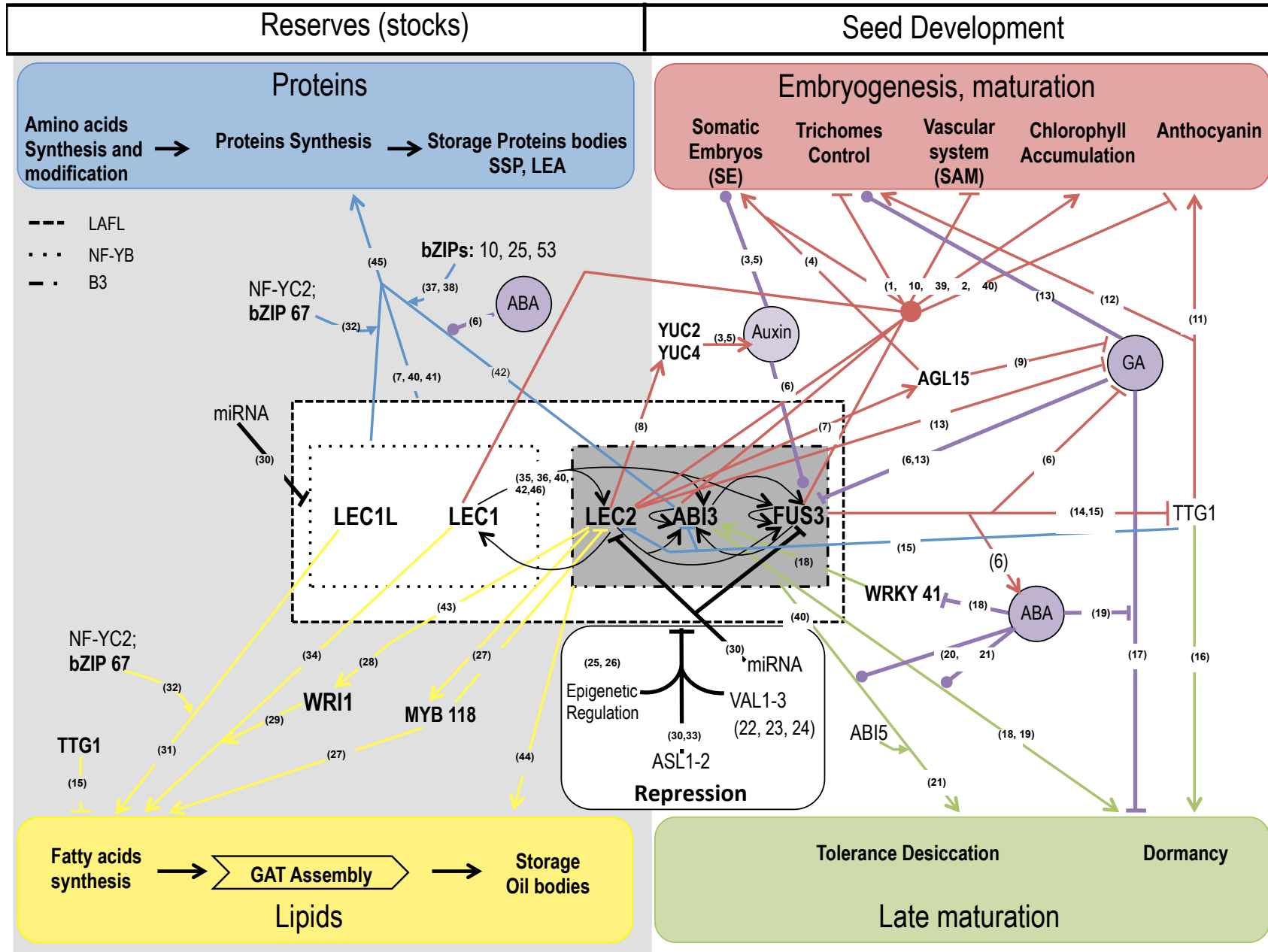
Facing the complexity of the network involved, mixing major transcription factors and hormonal regulations, modelling will certainly be needed to bridge quantitative (transcriptomic, metabolomic, ) and functional data (protein-protein interactions, complex identification) to unravel the control of seed maturation by LAFs. Alternatively, the simplification of the networks controlling seed maturation has been proposed, aiming at separating the distinct functions of the AFLs by either mathematical or biological simplification (Devic and Roscoe 2016). The expected benefits would be to identify the minimal gene set (close to the ancestor's gene set) and the core architecture of the network.

### **Acknowledgements**

We thank Mark Tepfer for critical reading of the manuscript. This work was supported by ANR projects CERES (ANR-BLAN-1238), and ERA-CAPS ABCEEDS. The IJPB benefits from the support of the Labex Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS) grant.

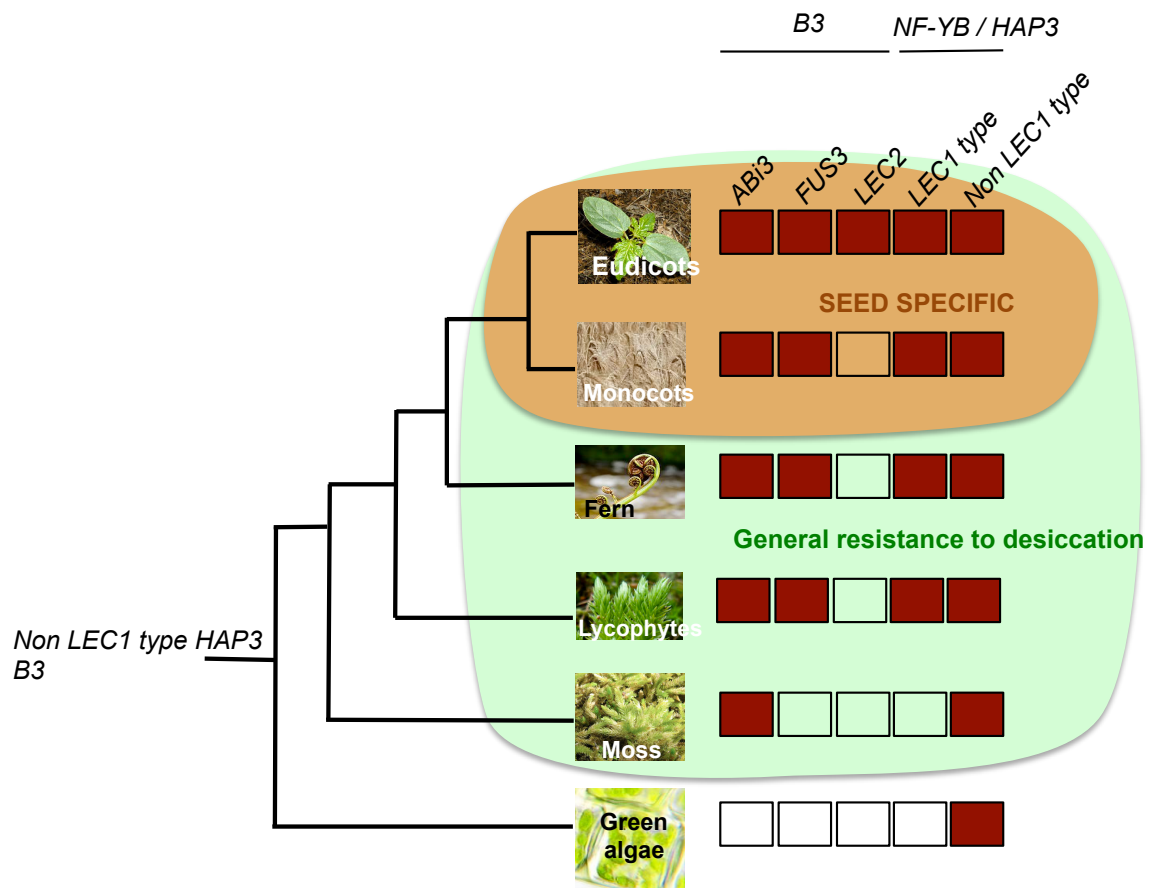


**Figure 1:** Expression of AFLs during seed development of *Arabidopsis thaliana*. mRNA expressions were carried out during development correlate with levels of hormones, Auxin, Abscissic Acid and Gibberrellins in whole silique, embryos were observed during development by Normarski under microscopy. (Kanno *et al.* 2010)





**Figure 2:** The LAFL network during seed development involvement and regulation of LAFLs on the seed development stages.1 (Gaj *et al.* 2005), 2 (Meinke *et al.* 1994), 3 (Wojcikowska *et al.* 2013), 4 (Thakare *et al.* 2008), 5 (Cheng *et al.* 2006) 6 (Gazzarrini *et al.* 2004), 7 (Braybrook *et al.* 2006), 8 (Stone *et al.* 2008), 9 (Wang *et al.* 2004), 10 (West *et al.* 1994), 11 (Walker *et al.* 1999), 12 (Payne *et al.* 2000), 13 (Curaba *et al.* 2004), 14 (Tsuchiya *et al.* 2004), 15 (Chen *et al.* 2015), 16 (Leon-Kloosterziel *et al.* 1994), 17 (Raz *et al.* 2001), 18 (Ding *et al.* 2014), 19 (Kanno *et al.* 2010), 20 (Cheng *et al.* 2014), 21 (Lopez-Molina *et al.* 2002), 22 (Suzuki *et al.* 2007), 23 (Jia *et al.* 2013), 24 (Schneider *et al.* 2016), 25 (Berger *et al.* 2011), 26 (Jia *et al.* 2014), 27 (Barthole *et al.* 2014), 28 (Baud *et al.* 2007), 29 (Maeo *et al.* 2009), 30 (Willmann *et al.* 2011), 31 (Mendes *et al.* 2013), 32 (Yamamoto *et al.* 2009), 33 (Gao *et al.* 2009), 34 (Mu *et al.* 2008), 35 (To *et al.* 2006), 36 (Santos-Mendoza *et al.* 2008), 37 (Lara *et al.* 2003), 38 (Alonso *et al.* 2009), 39 (Stone *et al.* 2001), 40 (Parcy *et al.* 1997), 41 (Kroj *et al.* 2003), 42 (Monke *et al.* 2012), 43 (Angeles-Nunez and Tiessen 2011), 44 (Roscoe *et al.* 2015), 45 (Baud *et al.* 2016), 46 (Santos Mendoza *et al.* 2005)



**Figure 3:** The LAF genes during evolution of plants.

Red and white boxes represent presence or absence of the genes respectively.

In green : general resistance to desiccation mechanisms. In orange : specialisation to seeds.

Gene	Locus	Mutants characterized	Expression				Functional category	Family	Reference(s)
			Embryo	Endosperm	Seed coat	seedling			
LEC2	AT1G28300	<i>lec2-1, lec2-3, lec2-4, lec2-5, lec2-10, lec2-11</i>	X	X			TF	B3	1, 2, 3, 4
LEC1	AT1G21970	<i>lec1-1, lec1-2, lec1-10, lec1-11</i>	X	X			TF	NF-Y	2, 5, 6, 7
ABI3	AT3G24650	<i>abi3-1, abi3-3, abi3-4, abi3-10, abi3-11</i>	X		X		TF	B3	2, 5, 6, 8, 9
FUS3	AT3G26790	<i>fus3-1, fus3-2, fus3-3</i>	X		X	X	TF	B3	2, 7, 10, 11, 12
LEC1L	AT5G47670		X				TF	NF-Y	13
WRI1	AT3G54320	<i>wri1-1, wri1-3, wri1-4, wri1-5</i>	X			X	TF	AP2/EREBP	14, 15
MYB118	AT3G27785	<i>myb118-1</i>	X	X		X	TF	R2-R3-type MYB	4, 16
AGL15	AT5G13790	<i>agl15-3, agl15-4</i>	X			X	TF	MADS	17, 18
VAL1	AT2G30470	<i>val1-2, val1-2</i>	X			X	TF	B3-PHD -CW-EAR	19, 20, 21
VAL2	AT4G32010	<i>val2-1, val2-2</i>	X			X	TF	B3-PHD -CW-EAR	19, 20, 21
VAL3	AT4G21550	<i>val3-1, val3-2, val3-3</i>	X			X	TF	B3 -CW-EAR	19, 20, 21
ASIL1	AT1G54060	<i>asil1-1, asil1-2</i>	X			X	TF	Trihelix	22, 23
ASIL2	AT3G14180	<i>asil2-1</i>	X			X	TF	Trihelix	23
TTG1	AT5G24520	<i>ttg1-1, ttg1-9</i>	X		X	X	TF	WD40	24, 25
NF-YC2	AT1G56170		X			X	TF	NF-Y	26, 27, 28
bZIP10	AT4G02640		X			X	TF	leucine zipper	29, 30, 31, 32
bZIP25	AT3G54620		X			X	TF	leucine zipper	29, 30, 31, 32
bZIP 28	AT3G10800	<i>bzip28-2</i>	X			X	TF	leucine zipper	28, 29
bZIP53	AT3G62420		X			X	TF	leucine zipper	29, 30, 31
bZIP67	AT3G44460	<i>bzip67-1, bzip67-2</i>	X				TF	leucine zipper	29
ABI5	AT2G36270	<i>abi5-1, abi5-2, abi5-3</i>	X	X		X	TF	leucine zipper	29, 33
WRKY41	AT4G11070	<i>wrky41-1, wrky41-2</i>	X			X	TF	w-box	34
YUC2	AT4G13260	<i>yuc2</i>	X			X	Enzyme	oxido-reductase	35, 36
YUC4	AT5G11320	<i>yuc4</i>	X			X	Enzyme	oxido-reductase	35, 36

**Table1:** Summary of transcription factors involved in the development of the seed. <https://www.arabidopsis.org/>, 1 (Stone *et al.* 2001), 2 (Meinke *et al.* 1994), 3 (Angeles-Nunez and Tiessen 2011), 4 (Barthole *et al.* 2014), 5 (Lotan *et al.* 1998), 6 (West *et al.* 1994), 7 (Parcy *et al.* 1997), 8 (Giraudat *et al.* 1992), 9 (To *et al.* 2006), 10 (Keith *et al.* 1994), 11 (Bäumlein *et al.* 1994), 12 (Gazzarrini *et al.* 2004), 13 (Kwong *et al.* 2003), 14 (Cernac and Benning 2004), 15 (Baud *et al.* 2007), 16 (Zhang *et al.* 2009), 17 (Thakare *et al.* 2008), 18 (Braybrook *et al.* 2006), 19 (Suzuki *et al.* 2007), 20 (Suzuki and McCarty 2008), 21 (Jia *et al.* 2013), 22 (Gao *et al.* 2009), 23 (Willmann *et al.* 2011), 24 (Walker *et al.* 1999), 25 (Leon-Kloosterziel *et al.* 1994), 26 (Petroni *et al.* 2012), 27 (Hackenberg *et al.* 2012), 28 (Liu and Howell 2010), 29 (Jakoby *et al.* 2002), 30 (Ehlert *et al.* 2006), 31 (Llorca *et al.* 2015), 32 (Lara *et al.* 2003) 33 (Finkelstein and Lynch 2000), 34 (Ding *et al.* 2014), 35 (Cheng *et al.* 2006), 36 (Stone *et al.* 2008).

	Target element	Consensus sequence	Reference(s)
bZIP67; bZIP53; BZIP10; bZIP25; ABI5	G-BOX / ABRE	cACGTGTC	1, 2,
VAL; ARF2	RY elements	GCATGC	3
LEC2, ABI3	RY elements	GCATGC	4
FUS3	RY elements	cGCATGCG	4
LEC1; L1L	CCAAT-box-binding factor CBF	CCAAT	5, 6
WRI1	AW motifs	CnTnG n7CG	7
ASIL1	GT box	GTGATT	8
MYB118	AC rich sequence	TAACGG	9, 10
AGL15	CArg-Box (C-A/T-rich-G)	CCTAATTAGG	11

**Table 2:** DNA sequences recognized by transcription factors involved in the development of the seed. 1 (Peng and Weselake 2011), 2(Jakoby *et al.* 2002), 3) (Yamasaki *et al.* 2013),4) (Baud *et al.* 2016) 5) (Fleming *et al.* 2013), 6 (Mantovani 1998), 7 (Maeo *et al.* 2009), 8 (Gao *et al.* 2009),9 (Barthole *et al.* 2014), 10 (Prouse and Campbell 2012), 11 (de Folter and Angenent 2006).

## References

- Albani D, Robert LS (1995) Cloning and characterization of a Brassica napus gene encoding a homologue of the B subunit of a heteromeric CCAAT-binding factor. *Gene* 167 (1-2):209-213
- Alonso R, Onate-Sanchez L, Weltmeier F, Ehlert A, Diaz I, Dietrich K, Vicente-Carbajosa J, Droge-Laser W (2009) A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. *The Plant cell* 21 (6):1747-1761. doi:10.1105/tpc.108.062968
- Angeles-Nunez JG, Tiessen A (2011) Mutation of the transcription factor LEAFY COTYLEDON 2 alters the chemical composition of Arabidopsis seeds, decreasing oil and protein content, while maintaining high levels of starch and sucrose in mature seeds. *J Plant Physiol* 168 (16):1891-1900. doi:10.1016/j.jplph.2011.05.003
- Bai B, Su YH, Yuan J, Zhang XS (2013) Induction of somatic embryos in Arabidopsis requires local YUCCA expression mediated by the down-regulation of ethylene biosynthesis. *Mol Plant* 6 (4):1247-1260. doi:10.1093/mp/sss154
- Barthole G, To A, Marchive C, Brunaud V, Soubigou-Taconnat L, Berger N, Dubreucq B, Lepiniec L, Baud S (2014) MYB118 represses endosperm maturation in seeds of Arabidopsis. *The Plant cell* 26 (9):3519-3537. doi:10.1105/tpc.114.130021
- Bassel GW, Stamm P, Mosca G, Barbier de Reuille P, Gibbs DJ, Winter R, Janka A, Holdsworth MJ, Smith RS (2014) Mechanical constraints imposed by 3D cellular geometry and arrangement modulate growth patterns in the Arabidopsis embryo. *Proc Natl Acad Sci U S A* 111 (23):8685-8690. doi:10.1073/pnas.1404616111
- Baud S, Dubreucq B, Miquel M, Rochat C, Lepiniec L (2008) Storage reserve accumulation in Arabidopsis: metabolic and developmental control of seed filling. *Arabidopsis Book* 6:e0113. doi:10.1199/tab.0113
- Baud S, Kelemen Z, Thevenin J, Boulard C, Blanchet S, To A, Payre M, Berger N, Effroy-Cuzzi D, Franco-Zorrilla JM, Godoy M, Solano R, Thevenon E, Parcy F, Lepiniec L, Dubreucq B (2016) Deciphering the Molecular Mechanisms Underpinning the Transcriptional Control of Gene Expression by Master Transcriptional Regulators in Arabidopsis Seed. *Plant Physiol* 171 (2):1099-1112. doi:10.1104/pp.16.00034
- Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B (2007) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. *Plant J* 50 (5):825-838. doi:10.1111/j.1365-313X.2007.03092.x
- Bäumlein H, Miséra S, Luerksen H, Kölle K, Horstmann C, Wobus U, Müller AJ (1994) the FUS3 gene of Arabidopsis thaliana is a regulator of gene expression during late embryogenesis *Plant J*
- Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, Le BH, Drews GN, Brady SM, Goldberg RB, Harada JJ (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci U S A* 110 (5):E435-444. doi:10.1073/pnas.1222061110.
- Boulard C, Bardet M, Chardot T, Dubreucq B, Gromova M, Guillermo A, Miquel M, Nesi N, Yen-Nicolay S, Jolivet P (2015) The structural organization of seed oil bodies could explain the

- contrasted oil extractability observed in two rapeseed genotypes. *Planta* 242 (1):53-68. doi:10.1007/s00425-015-2286-4
- Berger N, Dubreucq B, Roudier F, Dubos C, Lepiniec L (2011) Transcriptional regulation of *Arabidopsis* LEAFY COTYLEDON2 involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *Plant Cell* 23 (11):4065-4078. doi:10.1105/tpc.111.087866
- Braybrook SA, Harada JJ (2008) LECs go crazy in embryo development. *Trends in plant science* 13 (12):624-630. doi:10.1016/j.tplants.2008.09.008
- Braybrook SA, Stone SL, Park S, Bui AQ, Le BH, Fischer RL, Goldberg RB, Harada JJ (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc Natl Acad Sci USA* 103 (9):3468-3473
- Cagliari A, Turchetto-Zolet AC, Korbes AP, Maraschin Fdos S, Margis R, Margis-Pinheiro M (2014) New insights on the evolution of Leafy cotyledon1 (LEC1) type genes in vascular plants. *Genomics* 103 (5-6):380-387. doi:10.1016/j.ygeno.2014.03.005
- Calvenzani V, Testoni B, Gusmaroli G, Lorenzo M, Gnesutta N, Petroni K, Mantovani R, Tonelli C (2012) Interactions and CCAAT-Binding of *Arabidopsis thaliana* NF-Y Subunits. *PLoS One* 7 (8):e42902. doi:10.1371/journal.pone.0042902
- PONE-D-12-15283 [pii]
- Carbonero P, Iglesias-Fernandez R, Vicente-Carbajosa J (2016) The AFL subfamily of B3 transcription factors: evolution and function in angiosperm seeds. *J Exp Bot*. doi:10.1093/jxb/erw458
- Cernac A, Benning C (2004) WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*. *The Plant Journal* 40 (4):575-585. doi:10.1111/j.1365-313X.2004.02235.x
- Chen M, Zhang B, Li C, Kulaveerasingam H, Chew FT, Yu H (2015) TRANSPARENT TESTA GLABRA1 Regulates the Accumulation of Seed Storage Reserves in *Arabidopsis*. *Plant Physiol* 169 (1):391-402. doi:10.1104/pp.15.00943
- Cheng Y, Dai X, Zhao Y (2006) Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*. *Genes Dev* 20 (13):1790-1799. doi:10.1101/gad.1415106
- Cheng ZJ, Zhao XY, Shao XX, Wang F, Zhou C, Liu YG, Zhang Y, Zhang XS (2014) Abscisic acid regulates early seed development in *Arabidopsis* by ABI5-mediated transcription of SHORT HYPOCOTYL UNDER BLUE1. *Plant Cell* 26 (3):1053-1068. doi:10.1105/tpc.113.121566
- Curaba J, Moritz T, Blervaque R, Parcy F, Raz V, Herzog M, Vachon G (2004) AtGA3ox2, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in *Arabidopsis*. *Plant Physiol* 136 (3):3660-3669. doi:10.1104/pp.104.047266
- de Folter S, Angenent GC (2006) trans meets cis in MADS science. *Trends in plant science* 11 (5):224-231. doi:10.1016/j.tplants.2006.03.008
- Devic M, Roscoe T (2016) Seed maturation: Simplification of control networks in plants. *Plant Sci* 252:335-346. doi:10.1016/j.plantsci.2016.08.012

- Ding ZJ, Yan JY, Li GX, Wu ZC, Zhang SQ, Zheng SJ (2014) WRKY41 controls Arabidopsis seed dormancy via direct regulation of ABI3 transcript levels not downstream of ABA. *Plant J* 79 (5):810-823. doi:10.1111/tpj.12597
- Dolfini D, Gatta R, Mantovani R (2012) NF-Y and the transcriptional activation of CCAAT promoters. *Crit Rev Biochem Mol Biol* 47 (1):29-49. doi:10.3109/10409238.2011.628970
- Drouin J (2014) Minireview: pioneer transcription factors in cell fate specification. *Mol Endocrinol* 28 (7):989-998. doi:10.1210/me.2014-1084
- Duong S, Vonapartis E, Li CY, Patel S, Gazzarrini S (2017) The E3 ligase ABI3-INTERACTING PROTEIN2 negatively regulates FUSCA3 and plays a role in cotyledon development in Arabidopsis thaliana. *J Exp Bot*. doi:10.1093/jxb/erx046
- Ehlert A, Weltmeier F, Wang X, Mayer CS, Smeekens S, Vicente-Carbajosa J, Droge-Laser W (2006) Two-hybrid protein-protein interaction analysis in Arabidopsis protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. *Plant J* 46 (5):890-900. doi:10.1111/j.1365-313X.2006.02731.x
- Finkelstein R, Reeves W, Ariizumi T, Steber C (2008) Molecular aspects of seed dormancy. *Annu Rev Plant Biol* 59:387-415. doi:10.1146/annurev.arplant.59.032607.092740
- Finkelstein RR, Gampala SS, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. *Plant Cell* 14 Suppl:S15-45
- Finkelstein RR, Lynch TJ (2000) The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. *Plant Cell* 12 (4):599-609
- Fleming JD, Pavesi G, Benatti P, Imbriano C, Mantovani R, Struhl K (2013) NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereopositioned with growth-controlling transcription factors. *Genome Res* 23 (8):1195-1209. doi:10.1101/gr.148080.112
- Gaj MD, Zhang S, Harada JJ, Lemaux PG (2005) Leafy cotyledon genes are essential for induction of somatic embryogenesis of Arabidopsis. *Planta* 222 (6):977-988
- Gao MJ, Lydiate DJ, Li X, Lui H, Gjetvaj B, Hegedus DD, Rozwadowski K (2009) Repression of seed maturation genes by a trihelix transcriptional repressor in Arabidopsis seedlings. *Plant Cell* 21 (1):54-71. doi:10.1105/tpc.108.061309
- Gazzarrini S, Tsuchiya Y, Lumba S, Okamoto M, McCourt P (2004) The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones gibberellin and abscisic acid. *Dev Cell* 7 (3):373-385. doi:10.1016/j.devcel.2004.06.017
- Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* 4 (10):1251-1261. doi:10.1105/tpc.4.10.1251
- Gnesutta N, Saad D, Chaves-Sanjuan A, Mantovani R, Nardini M (2017) Crystal Structure of the Arabidopsis thaliana L1L/NF-YC3 Histone-fold Dimer Reveals Specificities of the LEC1 Family of NF-Y Subunits in Plants. *Mol Plant* 10 (4):645-648. doi:10.1016/j.molp.2016.11.006
- Golovenko D, Manakova E, Zakrys L, Zaremba M, Sasnauskas G, Grazulis S, Siksnys V (2014) Structural insight into the specificity of the B3 DNA-binding domains provided by the co-crystal

- structure of the C-terminal fragment of Bfil restriction enzyme. *Nucleic Acids Res* 42 (6):4113-4122. doi:10.1093/nar/gkt1368
- Gong W, Shen YP, Ma LG, Pan Y, Du YL, Wang DH, Yang JY, Hu LD, Liu XF, Dong CX, Ma L, Chen YH, Yang XY, Gao Y, Zhu D, Tan X, Mu JY, Zhang DB, Liu YL, Dinesh-Kumar SP, Li Y, Wang XP, Gu HY, Qu LJ, Bai SN, Lu YT, Li JY, Zhao JD, Zuo J, Huang H, Deng XW, Zhu YX (2004) Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. *Plant Physiol* 135 (2):773-782. doi:10.1104/pp.104.042176
- Grimault A, Gendrot G, Chaignon S, Gilard F, Tcherkez G, Thevenin J, Dubreucq B, Depege-Fargeix N, Rogowsky PM (2015) Role of B3 domain transcription factors of the AFL family in maize kernel filling. *Plant Sci* 236:116-125. doi:10.1016/j.plantsci.2015.03.021
- Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C (2007) Combined networks regulating seed maturation. *Trends in plant science* 12 (7):294-300. doi:10.1016/j.tplants.2007.06.003
- Hackenberg D, Wu Y, Voigt A, Adams R, Schramm P, Grimm B (2012) Studies on differential nuclear translocation mechanism and assembly of the three subunits of the Arabidopsis thaliana transcription factor NF-Y. *Mol Plant* 5 (4):876-888. doi:10.1093/mp/ssr107
- Han JD, Li X, Jiang CK, Wong GKS, Rothfels CJ, Rao GY (2017) Evolutionary Analysis of the LAFL Genes Involved in the Land Plant Seed Maturation Program. *FrontPlant Sci.* doi:doi:10.3389/fpls;2017.00439
- Hobo T, Kowyama Y, Hattori T (1999) A bZIP factor, TRAB1, interacts with VP1 and mediates abscisic acid-induced transcription. *Proc Natl Acad Sci USA* 96 (26):15348-15353. doi:10.1073/pnas.96.26.15348
- Huang M, Hu Y, Liu X, Li Y, Hou X (2015) Arabidopsis LEAFY COTYLEDON1 Mediates Postembryonic Development via Interacting with PHYTOCHROME-INTERACTING FACTOR4. *Plant Cell* 27 (11):3099-3111. doi:10.1105/tpc.15.00750
- Jakoby M, Weisshaar B, Droge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F (2002) bZIP transcription factors in Arabidopsis. *Trends in plant science* 7 (3):106-111
- Jenik PD, Gillmor CS, Lukowitz W (2007) Embryonic patterning in Arabidopsis thaliana. *Annu Rev Cell Dev Biol* 23:207-236. doi:10.1146/annurev.cellbio.22.011105.102609
- Jia H, McCarty DR, Suzuki M (2013) Distinct roles of LAFL network genes in promoting the embryonic seedling fate in the absence of VAL repression. *Plant Physiol* 163 (3):1293-1305. doi:10.1104/pp.113.220988
- Jia H, Suzuki M, McCarty DR (2014) Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. *Wiley interdisciplinary reviews Developmental biology* 3 (1):135-145. doi:10.1002/wdev.126
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473 (7345):97-100. doi:10.1038/nature09916



- Jolivet P, Acevedo F, Boulard C, d'Andrea S, Faure JD, Kohli A, Nesi N, Valot B, Chardot T (2013) Crop seed oil bodies: from challenges in protein identification to an emerging picture of the oil body proteome. *Proteomics* 13 (12-13):1836-1849. doi:10.1002/pmic.201200431
- Junker A, Rohn H, Schreiber F (2012) Visual analysis of transcriptome data in the context of anatomical structures and biological networks. *Front Plant Sci* 3:252. doi:10.3389/fpls.2012.00252
- Kanno Y, Jikumaru Y, Hanada A, Nambara E, Abrams SR, Kamiya Y, Seo M (2010) Comprehensive hormone profiling in developing Arabidopsis seeds: examination of the site of ABA biosynthesis, ABA transport and hormone interactions. *Plant Cell Physiol* 51 (12):1988-2001. doi:10.1093/pcp/pcq158
- Keith K, Kraml M, Dengler NG, McCourt P (1994) *fusca3*: A Heterochronic Mutation Affecting Late Embryo Development in Arabidopsis. *Plant Cell* 6 (5):589-600
- Kelemen Z, Sebastian A, Xu W, Grain D, Salsac F, Avon A, Berger N, Tran J, Dubreucq B, Lurin C, Lepiniec L, Contreras-Moreira B, Dubos C (2015) Analysis of the DNA-Binding Activities of the Arabidopsis R2R3-MYB Transcription Factor Family by One-Hybrid Experiments in Yeast. *PLoS One* 10 (10):e0141044. doi:10.1371/journal.pone.0141044
- Khandelwal A, Cho SH, Marella H, Sakata Y, Perroud PF, Pan A, Quatrano RS (2010) Role of ABA and ABI3 in desiccation tolerance. *Science* 327 (5965):546. doi:10.1126/science.1183672
- Kim HU, Jung SJ, Lee KR, Kim EH, Lee SM, Roh KH, Kim JB (2013) Ectopic overexpression of castor bean LEAFY COTYLEDON2 (LEC2) in Arabidopsis triggers the expression of genes that encode regulators of seed maturation and oil body proteins in vegetative tissues. *FEBS open bio* 4:25-32. doi:10.1016/j.fob.2013.11.003
- Kirchler T, Briesemeister S, Singer M, Schutze K, Keinath M, Kohlbacher O, Vicente-Carbajosa J, Teige M, Harter K, Chaban C (2010) The role of phosphorylatable serine residues in the DNA-binding domain of Arabidopsis bZIP transcription factors. *European journal of cell biology* 89 (2-3):175-183. doi:10.1016/j.ejcb.2009.11.023
- Kirkbride RC, Fischer RL, Harada JJ (2013) LEAFY COTYLEDON1, a key regulator of seed development, is expressed in vegetative and sexual propagules of *Selaginella moellendorffii*. *PLoS One* 8 (6):e67971. doi:10.1371/journal.pone.0067971
- Kohler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev* 17 (12):1540-1553
- Kroj T, Savino G, Valon C, Giraudat J, Parcy F (2003) Regulation of storage protein gene expression in Arabidopsis. *Development* 130 (24):6065-6073. doi:10.1242/dev.00814
- Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ (2003) LEAFY COTYLEDON1-LIKE Defines a Class of Regulators Essential for Embryo Development. *Plant Cell* 15 (1):5-18
- Lara P, Onate-Sanchez L, Abraham Z, Ferrandiz C, Diaz I, Carbonero P, Vicente-Carbajosa J (2003) Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. *J Biol Chem* 278 (23):21003-21011. doi:10.1074/jbc.M210538200

- Laux T, Wurschum T, Breuninger H (2004) Genetic regulation of embryonic pattern formation. *Plant Cell* 16 Suppl:S190-202. doi:10.1105/tpc.016014
- Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamura JK, Harada JJ, Goldberg RB (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci U S A* 107 (18):8063-8070. doi:10.1073/pnas.1003530107
- Lee H, Fischer RL, Goldberg RB, Harada JJ (2003) Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci U S A* 100 (4):2152-2156
- Leon-Kloosterziel KM, Keijzer CJ, Koornneef M (1994) A Seed Shape Mutant of Arabidopsis That Is Affected in Integument Development. *Plant Cell* 6 (3):385-392
- Lepiniec L, Debeaujon I, Routaboul J-M, Baudry A, Pourcel L, Nesi N, Caboche M (2006) Genetics and biochemistry of seed flavonoids. *Ann Rev Plant Biol* 57 (1):405-430
- Li XY, Mantovani R, Hooft van Huijsduijnen R, Andre I, Benoist C, Mathis D (1992) Evolutionary variation of the CCAAT-binding transcription factor NF-Y. *Nucleic Acids Res* 20 (5):1087-1091
- Li Y, Jin K, Zhu Z, Yang J (2010) Stepwise origin and functional diversification of the AFL subfamily B3 genes during land plant evolution. *J Bioinform Comput Biol* 8 Suppl 1:33-45
- Liu J, Deng S, Wang H, Ye J, Wu HW, Sun HX, Chua NH (2016) CURLY LEAF Regulates Gene Sets Coordinating Seed Size and Lipid Biosynthesis. *Plant Physiol* 171 (1):424-436. doi:10.1104/pp.15.01335
- Liu JX, Howell SH (2010) bZIP28 and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in Arabidopsis. *Plant Cell* 22 (3):782-796. doi:10.1105/tpc.109.072173
- Llorca CM, Berendzen KW, Malik WA, Mahn S, Piepho HP, Zentgraf U (2015) The Elucidation of the Interactome of 16 Arabidopsis bZIP Factors Reveals Three Independent Functional Networks. *PLoS One* 10 (10):e0139884. doi:10.1371/journal.pone.0139884
- Lopez-Molina L, Mongrand S, McLachlin DT, Chait BT, Chua NH (2002) ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. *Plant J* 32 (3):317-328
- Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93 (7):1195-1205.
- Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K., and Zhang, Y. (2016). Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell* 165, 1375-1388.
- Luerssen H, Kirik V, Herrmann P, Misera S (1998) FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. *Plant J* 15 (6):755-764.
- Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A (2005) MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. *Proc Natl Acad Sci U S A* 102 (48):17531-17536. doi:10.1073/pnas.0508418102

- Lynch T, Erickson BJ, Finkelstein RR (2012) Direct interactions of ABA-insensitive(ABI)-clade protein phosphatase(PP)2Cs with calcium-dependent protein kinases and ABA response element-binding bZIPs may contribute to turning off ABA response. *Plant Mol Biol* 80 (6):647-658. doi:10.1007/s11103-012-9973-3
- Maeo K, Tokuda T, Ayame A, Mitsui N, Kawai T, Tsukagoshi H, Ishiguro S, Nakamura K (2009) An AP2-type transcription factor, WRINKLED1, of *Arabidopsis thaliana* binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. *Plant J* 60 (3):476-487. doi:10.1111/j.1365-313X.2009.03967.x
- Mantovani R (1998) A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res* 26 (5):1135-1143
- Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene* 239 (1):15-27. doi:S0378-1119(99)00368-6 [pii]
- Marella HH, Sakata Y, Quatrano RS (2006) Characterization and functional analysis of ABSCISIC ACID INSENSITIVE3-like genes from *Physcomitrella patens*. *Plant J* 46 (6):1032-1044. doi:doi:10.1111/j.1365-313X.2006.02764.x
- Meinke DW, Franzmann LH, Nickle TC, Yeung EC (1994) Leafy Cotyledon Mutants of *Arabidopsis*. *Plant Cell* 6 (8):1049-1064. doi:10.1105/tpc.6.8.1049
- Mendes A, Kelly AA, van Erp H, Shaw E, Powers SJ, Kurup S, Eastmond PJ (2013) bZIP67 regulates the omega-3 fatty acid content of *Arabidopsis* seed oil by activating fatty acid desaturase3. *The Plant cell* 25 (8):3104-3116. doi:10.1105/tpc.113.116343
- Mizzotti C, Ezquer I, Paolo D, Rueda-Romero P, Guerra RF, Battaglia R, Rogachev I, Aharoni A, Kater MM, Caporali E, Colombo L (2014) SEEDSTICK is a master regulator of development and metabolism in the *Arabidopsis* seed coat. *PLoS Genet* 10 (12):e1004856. doi:10.1371/journal.pgen.1004856
- Monke G, Seifert M, Keilwagen J, Mohr M, Grosse I, Hahnel U, Junker A, Weisshaar B, Conrad U, Baumlein H, Altschmied L (2012) Toward the identification and regulation of the *Arabidopsis thaliana* ABI3 regulon. *Nucleic Acids Res* 40 (17):8240-8254. doi:10.1093/nar/gks594
- Moreno-Risueno MA, Gonzalez N, Diaz I, Parcy F, Carbonero P, Vicente-Carbajosa J (2008) FUSCA3 from barley unveils a common transcriptional regulation of seed-specific genes between cereals and *Arabidopsis*. *Plant J* 53 (6):882-894. doi:10.1111/j.1365-313X.2007.03382.x
- Mu J, Tan H, Zheng Q, Fu F, Liang Y, Zhang J, Yang X, Wang T, Chong K, Wang XJ, Zuo J (2008) LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in *Arabidopsis*. *Plant Physiol* 148 (2):1042-1054. doi:10.1104/pp.108.126342
- Nardone V, Chaves-Sanjuan A, Nardini M (2016) Structural determinants for NF-Y/DNA interaction at the CCAAT box. *Biochim Biophys Acta*. doi:10.1016/j.bbagr.2016.09.006
- Nordine MD, Bartel DP (2010) MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes Dev* 24 (23):2678-2692. doi:10.1101/gad.1986710
- North H, Baud S, Debeaujon I, Dubos C, Dubreucq B, Grappin P, Jullien M, Lepiniec L, Marion-Poll A, Miquel M, Rajjou L, Routaboul JM, Caboche M (2010) *Arabidopsis* seed secrets unravelled after a decade of genetic and omics-driven research. *Plant J* 61 (6):971-981. doi:10.1111/j.1365-313X.2009.04095.x

- Ohto MA, Floyd SK, Fischer RL, Goldberg RB, Harada JJ (2009) Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in Arabidopsis. *Sex Plant Reprod* 22 (4):277-289. doi:10.1007/s00497-009-0116-1
- 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. *Mol Cell* 55, 708-722.
- Orozco-Arroyo G, Paolo D, Ezquer I, Colombo L (2015) Networks controlling seed size in Arabidopsis. *Plant Reprod* 28 (1):17-32. doi:10.1007/s00497-015-0255-5.
- Parcy F, Valon C, Kohara A, Misera S, Giraudat J (1997) The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. *Plant Cell* 9 (8):1265-1277. doi:10.1105/tpc.9.8.1265
- Payne CT, Zhang F, Lloyd AM (2000) GL3 encodes a bHLH protein that regulates trichome development in Arabidopsis through interaction with GL1 and TTG1. *Genetics* 156 (3):1349-1362
- 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. *Proc Natl Acad Sci U S A* 114, E6710-E6719.
- Peng FY, Weselake RJ (2011) Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis. *BMC Genomics* 12:286. doi:10.1186/1471-2164-12-286
- Petroni K, Kumimoto RW, Gnesutta N, Calvenzani V, Fornari M, Tonelli C, Holt BF, 3rd, Mantovani R (2012) The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. *Plant Cell* 24 (12):4777-4792. doi:10.1105/tpc.112.105734
- Phillips J, Artsaenko O, Fiedler U, Horstmann C, Mock HP, Muntz K, Conrad U (1997) Seed-specific immunomodulation of abscisic acid activity induces a developmental switch. *Embo J* 16 (15):4489-4496. doi:10.1093/emboj/16.15.4489
- Pimenta Lange MJ, Lange T (2006) Gibberellin biosynthesis and the regulation of plant development. *Plant Biol (Stuttg)* 8 (3):281-290. doi:10.1055/s-2006-923882
- Prouse MB, Campbell MM (2012) The interaction between MYB proteins and their target DNA binding sites. *Biochim Biophys Acta* 1819 (1):67-77. doi:10.1016/j.bbagr.2011.10.010
- Qu LJ, Zhu YX (2006) Transcription factor families in Arabidopsis: major progress and outstanding issues for future research. *Curr Opin Plant Biol* 9 (5):544-549. doi:10.1016/j.pbi.2006.07.005
- Raz V, Bergervoet JH, Koornneef M (2001) Sequential steps for developmental arrest in Arabidopsis seeds. *Development* 128 (2):243-252
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290 (5499):2105-2110

- Romanel EA, Schrago CG, Counago RM, Russo CA, Alves-Ferreira M (2009) Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. *PLoS One* 4 (6):e5791. doi:10.1371/journal.pone.0005791
- Romier C, Cocchiarella F, Mantovani R, Moras D (2003) The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J Biol Chem* 278 (2):1336-1345. doi:10.1074/jbc.M209635200
- Roscoe TT, Guilleminot J, Bessoule JJ, Berger F, Devic M (2015) Complementation of Seed Maturation Phenotypes by Ectopic Expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in Arabidopsis. *Plant Cell Physiol* 56 (6):1215-1228. doi:10.1093/pcp/pcv049
- Santos Mendoza M, Dubreucq B, Miquel M, Caboche M, Lepiniec L (2005) LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves. *FEBS Lett* 579 (21):4666-4670. doi:10.1016/j.febslet.2005.07.037
- Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L (2008) Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *Plant J* 54 (4):608-620. doi:10.1111/j.1365-313X.2008.03461.x
- Schallau A, Kakhovskaya I, Tewes A, Czihal A, Tiedemann J, Mohr M, Grosse I, Manteuffel R, Baumlein H (2008) Phylogenetic footprints in fern spore- and seed-specific gene promoters. *The Plant Journal* 53 (3):414-424. doi:doi:10.1111/j.1365-313X.2007.03354.x
- Schneider A, Aghamirzaie D, Elmarakeby H, Poudel AN, Koo AJ, Heath LS, Grene R, Collakova E (2016) Potential targets of VIVIPAROUS1/ABI3-LIKE1 (VAL1) repression in developing Arabidopsis thaliana embryos. *Plant J* 85 (2):305-319. doi:10.1111/tpj.13106
- Schneider A, Aghamirzaie D, Elmarakeby H, Poudel AN, Koo AJ, Heath LS, Grene R, Collakova E (2016) Potential targets of VIVIPAROUS1/ABI3-LIKE1 (VAL1) repression in developing Arabidopsis thaliana embryos. *Plant J* 85 (2):305-319. doi:10.1111/tpj.13106
- Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ (2006) The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. *Development* 133 (2):251-261. doi:10.1242/dev.02194
- Schutze K, Harter K, Chaban C (2008) Post-translational regulation of plant bZIP factors. *Trends Plant Sci* 13 (5):247-255. doi:10.1016/j.tplants.2008.03.002
- Siefers N, Dang KK, Kumimoto RW, Bynum WEt, Tayrose G, Holt BF, 3rd (2009) Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiol* 149 (2):625-641. doi:10.1104/pp.108.130591
- Sreenivasulu N, Usadel B, Winter A, Radchuk V, Scholz U, Stein N, Weschke W, Strickert M, Close TJ, Stitt M, Graner A, Wobus U (2008) Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. *Plant physiology* 146 (4):1738-1758. doi:10.1104/pp.107.111781
- Sreenivasulu N, Wobus U (2013) Seed-development programs: a systems biology-based comparison between dicots and monocots. *Annual review of plant biology* 64:189-217. doi:10.1146/annurev-arplant-050312-120215

- Stone S, Kwong L, Yee K, Pelletier J, Lepiniec L, Fischer R, Goldberg R, Harada J (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc Natl Acad Sci U S A* 98 (20):11806 - 11811
- Stone SL, Braybrook SA, Paula SL, Kwong LW, Meuser J, Pelletier J, Hsieh TF, Fischer RL, Goldberg RB, Harada JJ (2008) Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc Natl Acad Sci U S A* 105 (8):3151-3156. doi:0712364105 [pii]  
10.1073/pnas.0712364105
- Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ (2001c) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc Natl Acad Sci U S A* 98 (20):11806-11811
- Suzuki M, Kao CY, McCarty DR (1997) The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity.
- Suzuki M, McCarty DR (2008) Functional symmetry of the B3 network controlling seed development. *Curr Opin Plant Biol* 11 (5):548-553. doi:10.1016/j.pbi.2008.06.015
- Suzuki M, Wang HH, McCarty DR (2007) Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiol* 143 (2):902-911. doi:10.1104/pp.106.092320
- Swaminathan K, Peterson K, Jack T (2008) The plant B3 superfamily. *Trends in plant science* 13 (12):647-655. doi:10.1016/j.tplants.2008.09.006
- Ten Hove CA, Lu KJ, Weijers D (2015) Building a plant: cell fate specification in the early Arabidopsis embryo. *Development* 142 (3):420-430. doi:10.1242/dev.111500
- Thakare D, Tang W, Hill K, Perry SE (2008) The MADS-domain transcriptional regulator AGAMOUS-LIKE15 promotes somatic embryo development in Arabidopsis and soybean. *Plant Physiol* 146 (4):1663-1672. doi:10.1104/pp.108.115832
- To A, Valon C, Savino G, Guillemot J, Devic M, Giraudat J, Parcy F (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell* 18 (7):1642-1651. doi:10.1105/tpc.105.039925
- Troncoso-Ponce MA, Barthole G, Tremblais G, To A, Miquel M, Lepiniec L, Baud S (2016) Transcriptional Activation of Two Delta-9 Palmitoyl-ACP Desaturase Genes by MYB115 and MYB118 Is Critical for Biosynthesis of Omega-7 Monounsaturated Fatty Acids in the Endosperm of Arabidopsis Seeds. *Plant Cell* 28 (10):2666-2682. doi:10.1105/tpc.16.00612
- True JR, Carroll SB (2002) Gene co-option in physiological and morphological evolution. *Annu Rev Cell Dev Biol* 18:53-80. doi:10.1146/annurev.cellbio.18.020402.140619
- Tsai AY, Gazzarrini S (2012) AKIN10 and FUSCA3 interact to control lateral organ development and phase transitions in Arabidopsis. *Plant J* 69 (5):809-821. doi:10.1111/j.1365-313X.2011.04832.x
- Tsuchiya Y, Nambara E, Naito S, McCourt P (2004) The FUS3 transcription factor functions through the epidermal regulator TTG1 during embryogenesis in Arabidopsis. *The Plant Journal* 37 (1):73-81. doi:10.1046/j.1365-313X.2003.01939.x

- van Dop M, Liao CY, Weijers D (2015) Control of oriented cell division in the Arabidopsis embryo. *Curr Opin Plant Biol* 23:25-30. doi:10.1016/j.pbi.2014.10.004
- Vashisht D, Nodine MD (2014) MicroRNA functions in plant embryos. *Biochem Soc Trans* 42 (2):352-357. doi:10.1042/BST20130252
- Vicente-Carbajosa J, Carbonero P (2005) Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int J Dev Biol* 49 (5-6):645-651. doi:10.1387/ijdb.052046jc
- Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC (1999) The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. *Plant Cell* 11 (7):1337-1350.
- Waltner JK, Peterson FC, Lytle BL, Volkman BF (2005) Structure of the B3 domain from Arabidopsis thaliana protein At1g16640. *Protein Sci* 14 (9):2478-2483. doi:10.1110/ps.051606305
- Wang H, Caruso LV, Downie AB, Perry SE (2004) The embryo MADS domain protein AGAMOUS-Like 15 directly regulates expression of a gene encoding an enzyme involved in gibberellin metabolism. *Plant Cell* 16 (5):1206-1219. doi:10.1105/tpc.021261 tpc.021261 [pii]
- Wang Y, Deng D, Zhang R, Wang S, Bian Y, Yin Z (2012) Systematic analysis of plant-specific B3 domain-containing proteins based on the genome resources of 11 sequenced species. *Mol Biol Rep* 39 (5):6267-6282. doi:10.1007/s11033-012-1448-8
- West M, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, Harada JJ (1994) LEAFY COTYLEDON1 Is an Essential Regulator of Late Embryogenesis and Cotyledon Identity in Arabidopsis. *Plant Cell* 6 (12):1731-1745. doi:10.1105/tpc.6.12.1731
- Willmann MR, Mehalick AJ, Packer RL, Jenik PD (2011) MicroRNAs regulate the timing of embryo maturation in Arabidopsis. *Plant Physiol* 155 (4):1871-1884. doi:10.1104/pp.110.171355
- Wojcikowska B, Jaskola K, Gasiorek P, Meus M, Nowak K, Gaj MD (2013) LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of Arabidopsis, via YUCCA-mediated auxin biosynthesis. *Planta* 238 (3):425-440. doi:10.1007/s00425-013-1892-2
- Xie Z, Li X, Glover BJ, Bai S, Rao GY, Luo J, Yang J (2008) Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes. *Mol Biol Evol* 25 (8):1581-1592.
- Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T (2009) Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J* 58 (5):843-856.
- Yamamoto A, Yoshii M, Murase S, Fujita M, Kurata N, Hobo T, Kagaya Y, Takeda S, Hattori T (2014) Cell-by-cell developmental transition from embryo to post-germination phase revealed by heterochronic gene expression and ER-body formation in Arabidopsis leafy cotyledon mutants. *Plant & cell physiology* 55 (12):2112-2125. doi:10.1093/pcp/pcu139
- Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Osanai T, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2004) Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1. *Plant Cell* 16 (12):3448-3459. doi:10.1105/tpc.104.026112

- Yamasaki K, Kigawa T, Seki M, Shinozaki K, Yokoyama S (2013) DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends in plant science* 18 (5):267-276. doi:10.1016/j.tplants.2012.09.001
- Yang J, Xie Z, Glover BJ (2005) Asymmetric evolution of duplicate genes encoding the CCAAT-binding factor NF-Y in plant genomes. *New Phytol* 165 (2):623-632
- Zaret KS, Carroll JS (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 25 (21):2227-2241. doi:10.1101/gad.176826.111
- Zhang X, Garreton V, Chua N-H (2005) The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation. *Genes Dev* 19 (13):1532-1543
- Zhang Y, Cao G, Qu LJ, Gu H (2009) Involvement of an R2R3-MYB transcription factor gene AtMYB118 in embryogenesis in Arabidopsis. *Plant Cell Rep* 28 (3):337-346. doi:10.1007/s00299-008-0644-4
- Zhang Y, Clemens A, Maximova SN, Guiltinan MJ (2014) The Theobroma cacao B3 domain transcription factor TcLEC2 plays a dual role in control of embryo development and maturation. *BMC Plant Biol* 14:106. doi:10.1186/1471-2229-14-106
- Zhou Y, Tan B, Luo M, Li Y, Liu C, Chen C, Yu CW, Yang S, Dong S, Ruan J, Yuan L, Zhang Z, Zhao L, Li C, Chen H, Cui Y, Wu K, Huang S (2013) HISTONE DEACETYLASE19 interacts with HSL1 and participates in the repression of seed maturation genes in Arabidopsis seedlings. *Plant Cell* 25 (1):134-148. doi:10.1105/tpc.112.096313



### **I.3 Conclusion**

Les graines montrent plusieurs transitions cruciales dans leur développement (fécondation/zygote, embryogenèse/maturation, maturation/dormance, dormance/germination) qui impliquent l'action coordonnée de facteurs de transcription qui agissent sous le contrôle d'hormones et par la formation de complexes avec des partenaires spécifiques. Les facteurs de transcription à domaine B3, ainsi que des membres de la famille des NF-Y ont un rôle important dans la régulation du développement de la graine. Ces facteurs sont rassemblés dans un groupe nommé LAFL, composé des AFL pour les membres des facteurs de transcription à domaine B3 de ACIDE ABSCISIC INSENSITIVE3 (ABI3), FUSCA3 (FUS3), LEAFY COTYLEDON2 (LEC2) et LEAFY COTYLEDON1 (LEC1) et LEAFY COTYLEDON1-LIKE (L1L) pour les membres des NF-YB.

Les données accumulées depuis de nombreuses années, montrent l'implication des LAFL dans les différentes étapes du développement de la graine : expression pendant les stades de développement de l'embryon de l'embryogenèse à la maturation tardive, rôle dans la différenciation cellulaire, la mise en place des réserves (Chapitre I.2.2, figure 2 p. 21). Ces implications peuvent être classées en 3 grands types : en tant que régulateurs transcriptionnels, avec une action en réseau et via des complexes protéiques.

#### **1) Régulateurs transcriptionnels : les cibles des LAFL**

Les LAFL sont des acteurs essentiels pour la mise en place des réserves protéiques et lipidiques nécessaires au développement futur de la plante, notamment lors des synthèses des acides gras, des protéines des corps protéiques ou des protéines qui structurent les corps lipidiques (Chapitre I.2.2, figure 2 p. 21). Pourtant, les LAFL ont des profils d'expression variés ; certains facteurs de transcription agissent tout au long du développement de l'embryon, notamment *ABI3*, qui s'exprime dès les premiers jours de développement et jusque dans la résistance à la dessiccation en lien avec l'ABA lors de la maturation tardive. D'autres, comme *LEC1* et *LEC2*, s'expriment plus faiblement et agissent sur les stades de développement plus précoces allant de l'embryogénèse à la maturation (Chapitre I.2.2, figure 1 p. 20), suggérant une action essentielle dès le tout début du développement, alors que les réserves commencent seulement à s'accumuler. Les LAFL sont donc aussi des acteurs maîtres du développement de l'embryon et vont influencer la différenciation cellulaire puis la mise en place ou la dégradation des métabolites par le contrôle direct, au niveau transcriptionnel, de l'expression de certains gènes (Santos Mendoza *et al.* 2005; Braybrook *et al.* 2006; Yamamoto *et al.* 2010).

## 2) Le réseau de régulation des LAFL

Un réseau complexe joue sur l'expression ou la répression des LAFL (Chapitre I.2.2, figure 2 p. 21). Les LAFL ont la capacité de déclencher l'expression de leurs partenaires, comme pour LEC1 qui permet l'expression de *LEC2*, *FUS3* et *ABI3* et d'induire leurs propres expressions comme pour *FUS3* et *ABI3*. LEC1 et LEC2 activent l'expression de *ABI3* et *FUS3*, leurs profils d'expression et leurs rôles dans la boucle de régulation leur confèreraient donc le statut d'initiateurs du réseau. Ces boucles de régulation permettent ainsi de favoriser l'expression de toute ou partie des LAFL et révèlent l'importance qu'ont ces protéines lors du développement de la graine.

Néanmoins leurs actions sont extrêmement bien contrôlées : (a) il existe une forte régulation négative de l'expression des gènes, qui a été révélée (i) par l'action de PRC2 sur la chromatine et les modifications épigénétiques des histones au niveau des gènes *ABI3*, *FUS3*, *LEC2* (Shen *et al.* 2015; Kohler and Hennig 2010) (ii) par l'action des protéines VAL qui régulent négativement l'expression des LAFL, (iii) par régulation traductionnelle et l'action de micro ARNs qui vont permettre la répression de l'expression de la protéine par le système DICER / RISC (Fatihi *et al.* 2016), (b) par la régulation de *ABI3* et *FUS3* par la voie du protéasome et l'action d'enzymes de type E3 ligase (Zhang *et al.* 2005; Duong *et al.* 2017). Enfin l'activité des facteurs de transcription est aussi régulée positivement ou négativement par les hormones : acide abscissique, gibbérellines et auxine.

La régulation des activités des LAFL n'est cependant pas complètement résolue : peu d'études sur les protéines, hormis pour *ABI3* et *FUS3*, ont été menées pour étudier ces mécanismes de régulation, notamment par la présence de modifications post-traductionnelles régulant l'activité, l'import ou l'export nucléaire des protéines, ou celles initiant la dégradation, déterminant ainsi le temps de demi-vie des protéines.

## 3) Les complexes protéiques : les partenaires des LAFL

Il a été suggéré que les LAFL formeraient un complexe sur les promoteurs des gènes impliqués dans la mise en place des réserves tels que l'*albumine 2S* et l'*OLEOSINE1* (Kroj *et al.* 2003; Baud *et al.* 2016). Des études ont aussi démontré que les facteurs de transcription bZIPs pourraient intervenir dans l'activité des LAFL sur ces promoteurs ou l'intervention d'autres protéines, telle que NF-YC2. L'interaction entre ces partenaires reste cependant partiellement caractérisée.

Ces résultats révèlent la difficulté d'identifier réellement l'action de ces facteurs de transcription : ils sont impliqués dans de nombreuses voies, sont régulés et régulent l'expression d'autres gènes, leurs actions étant liées aux hormones ou à la présence

d'autres facteurs. Les études menées sont principalement réalisées sur l'expression génique laissant ouverte la question d'une interaction protéique, sans réellement avoir étudié les actions, les régulations ou les interactions d'un point de vue de la biochimie des protéines.

#### I.4 Objectif de la thèse

Comme nous l'avons vu précédemment, les facteurs LAFL (LEC1 ABI3, FUS3, LEC2) sont impliqués dans de nombreuses voies de régulation au cours des différentes étapes du développement et dans la mise en place des réserves dans l'embryon.

Des études récentes ont suggéré la formation d'un complexe protéique composé d'ABI3, LEC2 et LEC1 sur le promoteur de l'*OLEOSINE1* qui agirait sur l'expression du gène cible de façon synergique. De plus il a été montré que NF-YC2 pourrait interagir avec LEC2 par l'intermédiaire de LEC1, ouvrant ainsi la perspective de la formation d'un plus gros complexe (Baud *et al.* 2016).

L'objectif de mon projet de thèse est d'identifier et de caractériser les partenaires potentiels de ce complexe composé d'ABI3, LEC1 et LEC2, qui activent spécifiquement la transcription de gènes cibles au cours des différentes phases du développement des graines.

Le projet est constitué de 2 tâches principales :

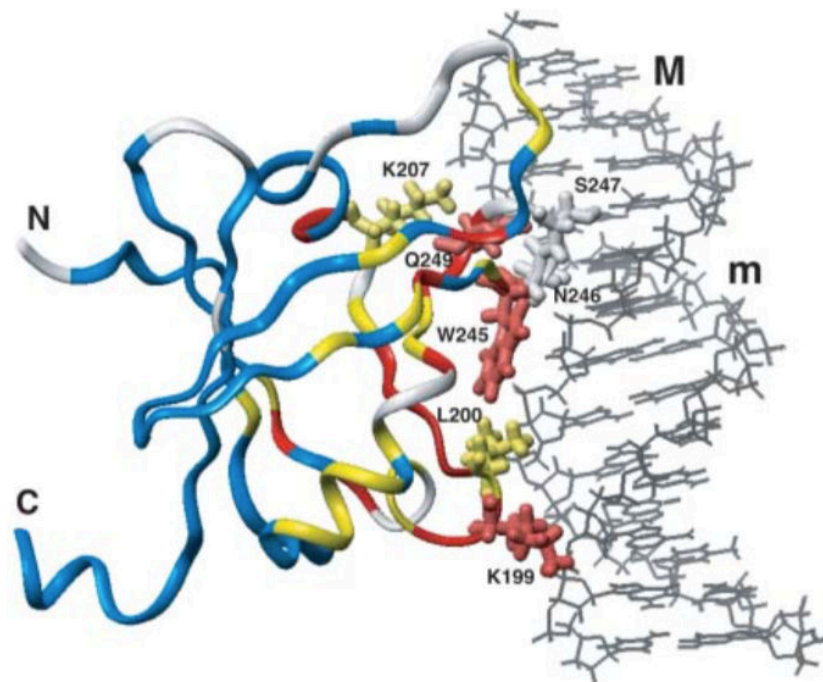
1) La recherche de partenaires du complexe avec a priori et leur caractérisation :

Cette partie repose sur la caractérisation de l'interaction entre LEC2 et LEC1 et de la recherche de nouveaux partenaires à partir de données issues de la littérature.

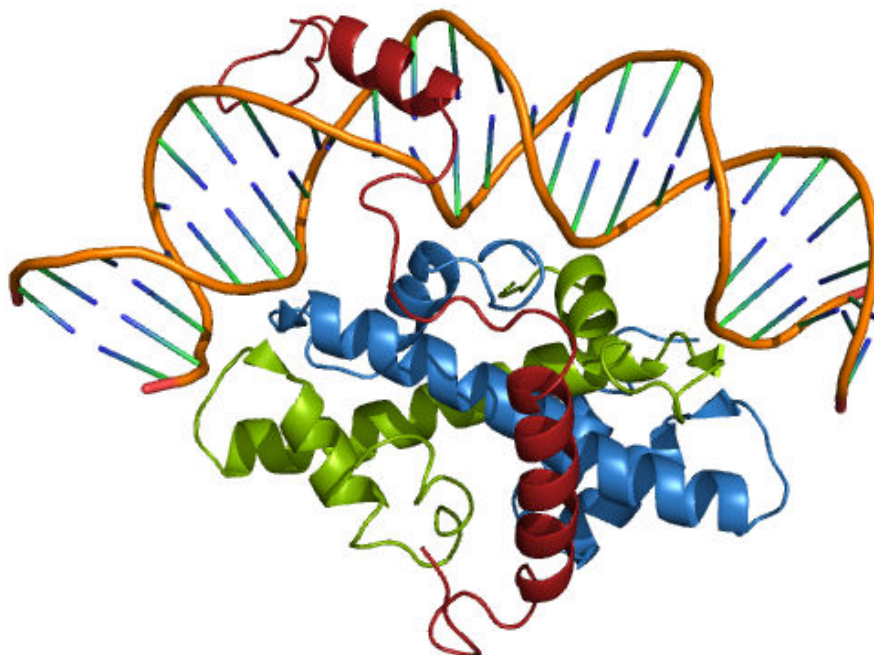
2) La recherche de partenaires associés à LEC2 sans a priori :

Cette partie repose sur la purification de complexes protéiques en utilisant des copies étiquetées fonctionnelles, *in vivo* au cours du développement de la graine d'*Arabidopsis thaliana*.

A



B



**Figure 4 : Structure des protéines AFL B3 et d'un complexe NF-Y (d'après Yamasaki *et al* 2004 et Nardone *et al* 2016).**

(A) structure 3D du domaine B3 de RAV1 en contact avec l'ADN, l'interaction entre la protéine et l'ADN serait prédite dans les sillons majeurs et mineurs. (B). Structure 3D du complexe NFY sur l'ADN, NF-YA (rouge) interagit avec l'ADN et avec le dimère NF-B (bleu) et NF-YC (vert) .

## II Recherche de partenaires avec a priori

### II.1 Introduction

Les LAFL sont des facteurs de transcription (FTs) qui appartiennent à deux différentes familles : les facteurs à domaine B3 et les NF-Y, les structures des domaines spécifiques de chaque famille sont à présent modélisées. De plus, les profils d'expression des gènes au cours du développement sont caractérisés. Néanmoins, les partenaires et les interactions entre les protéines AFL ne sont que partiellement élucidés.

ABI3, FUS3 et LEC2 (les AFL) appartiennent à la famille des facteurs de transcription à domaine B3. Ce domaine est très conservé et constitué de 7 feuillets  $\beta$  et de 2 hélices  $\alpha$ , structuré en  $\beta$ -barrel-like (Yamasaki *et al.* 2004; Waltner *et al.* 2005). Les FTs à domaine B3 se fixent sur les éléments cis de l'ADN, les boîtes RY (Chapitre I.2.2 ; Figure 4). LEC1 et LEC1-LIKE font partie de la famille des NF-Ys (NF-YB9 et NF-YB6 respectivement). Ce sont des NF-YBs qui sont caractérisés par leur incapacité à se lier spécifiquement à l'ADN. Les NF-Ys ont des domaines structurés en hélices  $\alpha$ , (Hackenberg *et al.* 2012). Les NF-YB trimérisent avec les autres membres des NF-Y : NF-YA et NF-YC (Mantovani 1999). Le trimère formé interagit avec l'ADN par reconnaissance d'éléments cis : les boîtes CCAAT (Lee *et al.* 2003) (Figure 4).

Les profils d'expressions de *LEC1* et *LEC2* sont similaires. Ils s'expriment dès le premier jour du développement de l'embryon entre 4 et 15 DAP et 3 et 14 DAP, respectivement. *LEC1* a une expression plus importante que *LEC2* et les ARNm sont localisés dans l'embryon et dans les cellules de l'albumen (Lotan *et al.* 1998; Kroj *et al.* 2003, Chapitre I.2.2, figure 1 p. 20). Les autres membres des B3, *ABI3* et *FUSCA3* sont exprimés dès l'embryogenèse à 3 DAP jusqu'à la fin de la phase de maturation. Les données d'hybridation *in situ* montrent une localisation d'*ABI3* dans tout l'embryon (To *et al.* 2006) alors que *FUS3* est localisé exclusivement dans l'épiderme (Kroj *et al.* 2003; Gazzarrini *et al.* 2004, Chapitre I.2.2, figure 1 p. 20). Enfin, *LEC1-LIKE (L1L)* est exprimé tout au long de la formation de l'embryon, faiblement lors de l'embryogenèse et plus fortement lors de la phase de maturation et est localisé dans l'embryon et les cellules de l'albumen (Kwong *et al.* 2003, Chapitre I.2.2, figure 1 p. 20).

Comme présenté précédemment, les AFLs semblent former des complexes pour induire l'expression des gènes cibles. Ainsi *LEC1* a une activité synergique (dans le sens où *LEC1*, qui n'a pas d'activité d'induction seule, augmente l'activité d'autres facteurs) avec *LEC2* ou *ABI3* et en présence des 3 protéines avec le promoteur de l'*OLEOSINE1*. *LEC1* et *LEC1-LIKE* sont les seuls NF-YB contenant un changement d'acide aminé en position 86 (K86D), l'acide aspartique 86 correspond à l'acide aspartique 55 retrouvé par alignement de séquence de *LEC1* et *LEC1-LIKE* d'arabidopsis et *HAP3* de levure (Kirkbride *et al.* 2013), qui



leur confère cette activité d'effet synergique avec ces FTs à domaine B3. De plus LEC1 co-immunoprécipite *in vitro* avec LEC2 en présence de NF-YC2 et du promoteur proOLE1 (Baud *et al.* 2016).

Les interactions entre les NF-Y B et C ont été testées par double hybride et révèlent que L1L interagit avec 9 NF-YC, alors que LEC1 ne montre qu'une légère interaction avec ces mêmes NF-YC, notamment NF-YC2 (Calvenzani *et al.* 2012). Des études d'interaction dans des protoplastes suggèrent la formation d'un complexe avec une synergie de l'activité transcriptionnelle entre NF-YB (6 ou 9) et NF-YC2, qui augmente en présence de bZIPs (Yamamoto *et al.* 2009; Mendes *et al.* 2013).

Les bZIPs sont des facteurs de transcription à domaine leucine zipper qui forment des dimères et interagissent notamment avec les boîtes « G » de l'ADN (Jakoby *et al.* 2002). Certaines de ces protéines ont une activité synergique sur l'activation des promoteurs des gènes de protéines de réserve ou les enzymes de la voie de biosynthèse des lipides en présence de LEC1, LEC1-LIKE, NF-YC2 et ABI3. (Lara *et al.* 2003; Alonso *et al.* 2009; Yamamoto *et al.* 2009; Mendes *et al.* 2013).

L'étude de recherche de partenaires avec a priori a porté sur la caractérisation biochimique de l'interaction entre LEC1 et LEC2 dont les profils d'expressions sont similaires et qui ont une activité synergique sur le même promoteur, ainsi que sur la recherche de nouveaux partenaires associés au complexe LEC1, LEC2 et ABI3. En particulier nous avons essayé de caractériser l'action de NF-YC2 sur l'activité transcriptionnelle du complexe et l'intervention de bZIPs dans la formation et l'activité du complexe, sur proOLE1.

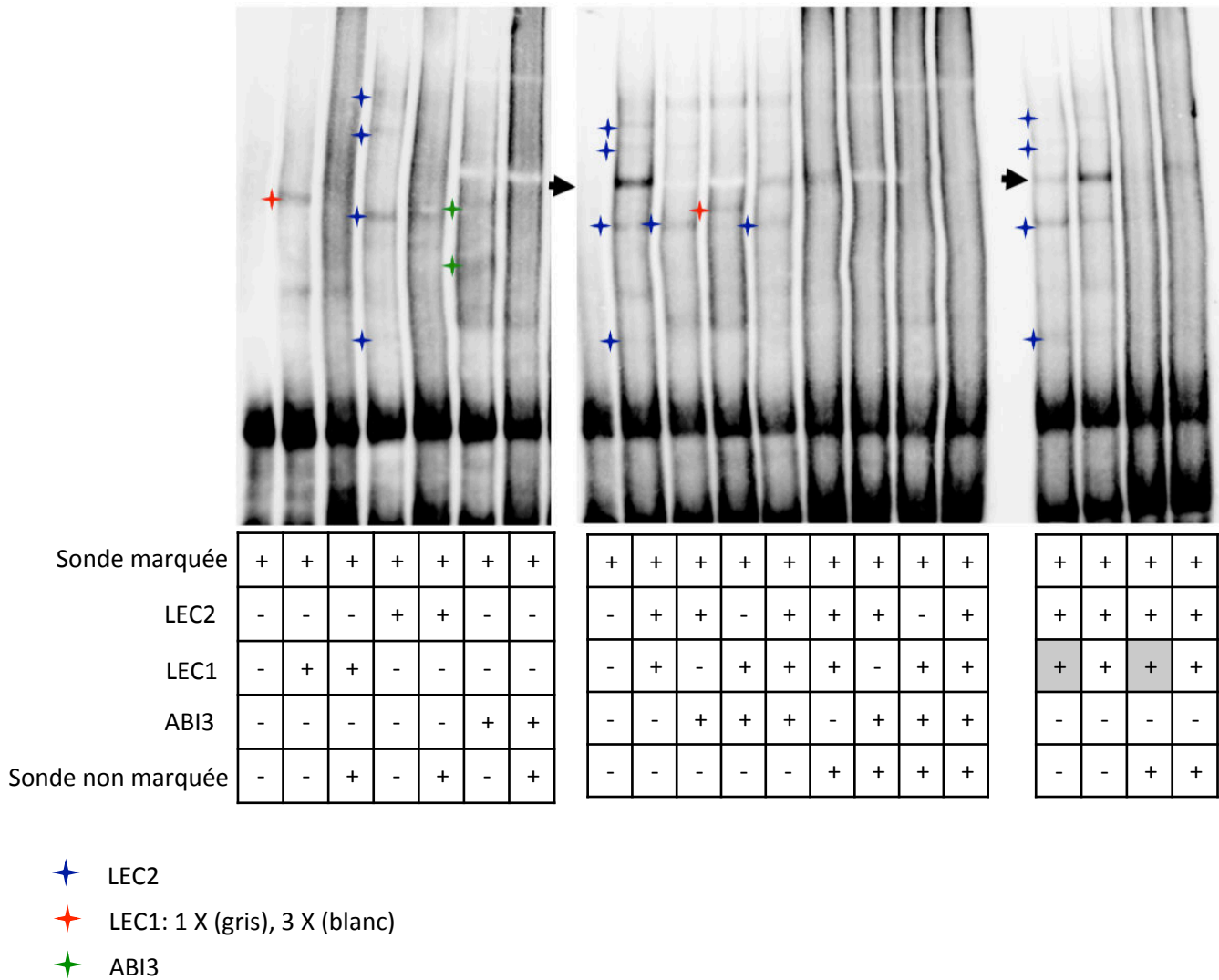
## **II.2 Les LAFL interagissent avec le promoteur de l'OLEOSINE1**

### **II.2.1 Formation d'un complexe AFL sur l'ADN**

Les LAFL (LEC1, ABI3, FUS3, LEC2) sont impliqués dans de nombreuses voies du développement de l'embryon et notamment dans la mise en place des réserves lipidiques et protéiques, en contrôlant la mise en place et le remplissage des réserves. (Chapitre I.2)

Les expériences réalisées dans l'équipe sur les LAFL et leurs affinités pour proOLE1, ont confirmé que, individuellement, les LAFL sont affins pour les boîtes RY du promoteur proOLE1 (Baud *et al.* 2016). La formation d'un complexe a été étudiée en expression dans des protoplastes de mousse (*P.patens*) et a montré que si ABI3, FUS3 et LEC2 ont une activité transcriptionnelle sur le promoteur proOLE1, FUS3 montre moins d'efficacité que LEC2 et ABI3. Associés LEC1, LEC2 et ABI3 montrent une augmentation de l'activité transcriptionnelle synergique. FUS3 ne montre pas la même activité, ce qui laisse suggérer la formation d'un complexe sur proOLE1 composé des 3 protéines LEC1, ABI3 et LEC2 (LAL) d'où FUS3 est exclu, bien qu'il soit un B3 de la même famille et avec des phénotypes





**Figure 5 : Étude de l'interaction, par la technique du gel retard, des protéines LEC1, LEC2, ABI3 seules ou en combinaison sur la séquence promotrice de l'OLEOSINE1 de 100 pb marquée par la biotine.** Les retards observés pour chacune des protéines seules sont représentés par des étoiles, les retards obtenus avec LEC2 et LEC1 sont localisés avec la flèche noire. (Donnée supplémentaire S6, Baud et al 2016). Un à trois µg de protéines ont été incubés avec 20 fmol de sonde biotinylée dans un volume réactionnel final de 20µL

similaires de la graine chez le mutant (Baud *et al.* 2016). Cette différence suggère que l'effet observé est spécifique à LEC1, LEC2 et ABI3 et que d'autres B3 proches ne rentrent pas dans le même complexe qui régule ce promoteur.

Les expériences réalisées dans les protoplastes de *Physcomitrella patens* (*Pp*) ne permettent pas d'exclure l'intervention de protéines exprimées dans la mousse. Il n'est pas possible de valider la formation d'un complexe dans ce type d'expérience. Pour cela, les interactions directes ADN / protéines *in vitro* de ces facteurs de transcription sur proOLE1 ont été testées par des expériences de gel retard.

### **II.2.1.1 Validation de la formation du complexe *in vitro***

Afin de confirmer la formation d'un complexe protéique composé d'ABI3, LEC1 et LEC2 sur proOLE1, les protéines ont été testées individuellement et en complexe à 2 et 3 protéines, sur une séquence minimum du promoteur composée des 3 boîtes G-box et 3 RY. LEC2 et ABI3 forment des complexes ADN / protéines, ainsi que LEC1 (Figure 5, Annexe 1A). Néanmoins LEC1 est un NF-YB n'ayant pas de domaine d'interaction spécifique avec l'ADN, ceci laisse penser à des interactions aspécifiques du domaine HFM (Histone Fold Motif) de LEC1. Aucun retard n'est observé lorsque ABI3 est associée avec une ou les deux autres protéines.

Les expériences ont montré la formation d'un complexe LEC1/LEC2 avec la sonde, qui valide la formation d'un complexe protéique entre ces deux facteurs de transcription (Figure 5). De plus la variation de la quantité de LEC1 lors de la formation de complexe révèle une plus forte intensité du retard sur gel quand la quantité de LEC1 augmente, ce qui suggère une importance de la stœchiométrie de LEC1 dans l'interaction avec LEC2 et l'ADN (Figure 5 panneau de droite) (Baud *et al.* 2016).

### **II.2.1.2 La dynamique de formation du complexe**

Les expériences *in vitro* ont montré la formation d'un complexe entre LEC1 et LEC2 sur proOLE1, ce complexe étant dépendant de la stœchiométrie de LEC1. Des expériences de double hybride ont mis en évidence la capacité de LEC1 à dimériser (Baud *et al.* 2016). LEC2 pourrait donc interagir avec LEC1 ou un dimère de LEC1.

Les différentes combinaisons de complexes contenant ABI3 n'ont pas montré la formation d'une interaction double ou triple avec l'ADN *in vitro*. Pourtant LEC1 et ABI3, exprimés dans des protoplastes de mousse, induisent une activité transcriptionnelle de façon synergique. Le plan d'expérience de cette étude ne tient cependant pas compte de la cinétique d'incorporation de chacune des protéines et la formation du complexe sur l'ADN pourrait en effet reposer sur une chronologie précise. Dans ce cas, le complexe ABI3, LEC1 et LEC2 sur l'ADN ne serait observable que dans des conditions particulières lors de la formation du



complexe, répondant à une cinétique de présence des protéines sur l'ADN selon un ordre précis. Cette hypothèse pourrait être testée avec des expériences de gel retard basées sur une combinatoire d'addition des protéines pour tester l'effet sur la formation du complexe.

## **II.2.2 Rôle d'un NF-YC dans le complexe LEC1, ABI3, LEC2 (LAL)**

### **II.2.2.1 La famille des NF-Ys**

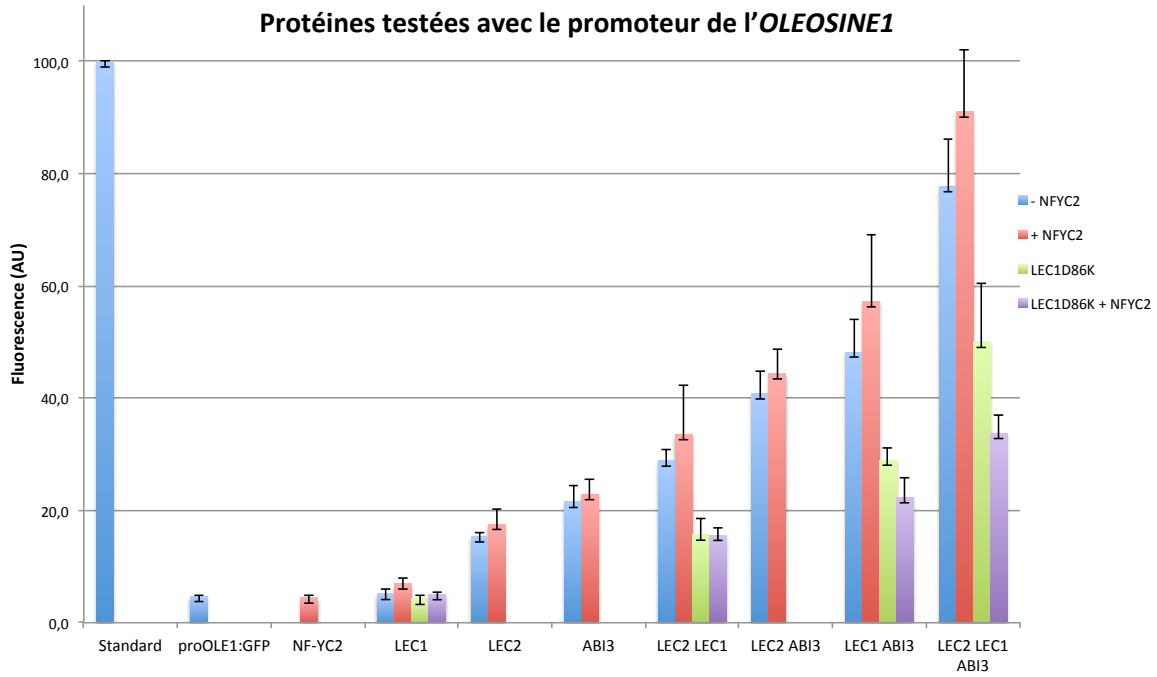
La famille des facteurs de transcription NF-Ys est conservée chez les eucaryotes, elle est retrouvée chez l'Homme, la levure, la mousse, les plantes. Ces facteurs de transcription se fixent sur les éléments cis CAAT-box de l'ADN (Mantovani 1999; Zhang *et al.* 2015; Kirkbride *et al.* 2013).

Les NFY sont largement représentés chez *Arabidopsis thaliana*. Ils interviennent, notamment dans les étapes du développement, en réponse à un stress par exemple (Calvenzani *et al.* 2012; Gnesutta *et al.* 2017). La famille des NF-Ys est composée de 3 membres, A, B, C, codés par 10, 13 et 13 gènes respectivement, tous contenant des domaines fortement conservés. NF-YA est composé, dans sa partie N terminale, d'un domaine d'interaction à l'ADN et d'interaction protéine – protéine. Les NF-YB et C ont, eux, un domaine conservé similaire au domaine Histone Fold Motif (HFM) (Mantovani 1999; Hackenberg *et al.* 2012; Lee *et al.* 2003). Les NF-Y sont structurés en 3/4 hélices alpha et 2 boucles et forment un trimère selon une cinétique de formation précise : NF-Y B et C forment un dimère par l'intermédiaire des 2 domaines HFM, le dimère est ensuite redirigé au noyau et formera un trimère avec NF-YA qui permet l'interaction avec des séquences spécifiques d'ADN (Mantovani 1999; Hackenberg *et al.* 2012). Le domaine HFM est nécessaire et suffisant pour la formation du dimère et l'interaction de NF-YA avec la boîte CAAT (Nardini *et al.* 2013; Gnesutta *et al.* 2017).

La formation d'un trimère avec NF-YA n'est pas exclusive, des études ayant en effet démontré que NF-YB et C forment des trimères avec d'autres facteurs de transcription. Le facteur de transcription MADS-box 18 de riz (Masiero *et al.* 2002), les protéines CONSTANS (CO) et CONTANS-LIKE (COL) forment ainsi des trimères avec un dimère NF-YB/C et se fixent sur la séquence spécifiquement reconnue par CO et COL, permettant l'expression des gènes cibles (Ben-Naim *et al.* 2006; Gnesutta *et al.* 2017).

Comme rappelé précédemment, LEC2 interagit avec LEC1 (NF-YB9) *in vitro* sur proOLE1. LEC1 et LEC1-LIKE sont des membres de la famille de NF-Y, B9 et B6 respectivement. Leurs activités diffèrent de celles des autres NF-YB par la présence d'un acide aminé acide aspartique (D) en position 86 à la place d'une lysine (K) pour les autres NF-YBs, qui leur confère la capacité d'augmenter l'activité transcriptionnelle de ABI3 et LEC2 sur le promoteur de l'OLEOSINE1 (Baud *et al.* 2016). Une étude systématique des interactions des 13 NF-YB

A



B

	Ratio	Ratio	Ratio	Ratio	Student t-Test	Significativité
LEC1	5,794	4,05	5,71	4,94		
LEC1-NF-YC2	8,090	6,22	7,55	5,86	0,035	*
LEC2	14,287	15,39	15,48	16,21		
LEC2-NF-YC2	16,688	21,27	15,32	16,98	0,15	NS
ABI3	23,739	21,78	23,19	17,33	0,15	NS
ABI3-NF-YC2	25,713	24,59	20,08	21,11		
LEC2-LEC1	29,710	27,18	27,49	31,17	0,33	NS
LEC2-LEC1-NF-YC2	35,786	35,05	21,55	42,03		
LEC2-ABI3	35,874	42,47	39,64	45,44	0,28	NS
LEC2-ABI3-NF-YC2	41,531	41,87	43,29	50,85		
LEC1-ABI3	53,384	48,64	39,88	51,00	0,23	NS
LEC1-ABI3-NF-YC2	64,875	47,65	46,40	70,02		
LEC2-LEC1-ABI3	84,682	75,18	67,24	84,01	0,11	NS
LEC2-LEC1-ABI3-NF-YC2	95,947	85,39	78,63	103,98		
LEC1D86K	3,56	4,54	4,67		0,13	NS
LEC1D86K NF-YC2	4,60	5,49	5,20			
LEC2-LEC1D86K	16,49	12,44	18,15		0,94	NS
LEC2-LEC1D86K-NF-YC2	16,96	14,78	14,88			
LEC1D86K-ABI3	30,82	26,56	29,46		0,047	*
LEC1D86K-ABI3 -NF-YC2	25,10	18,51	23,30			
LEC2-LEC1D86K ABI3	34,50	56,52	54,89	54,2	0,05	*
LEC2-LEC1D86K ABI3-NF-YC2	34,04	30,40	36,86			

	Ratio	Ratio	Ratio	Ratio	Student t-Test	Significativité
LEC1	5,794	4,05	5,71	4,94	0,18	NS
LEC1D86K	3,56	4,54	4,67			
LEC2-LEC1	29,710	27,18	27,49	31,17	0,0008	***
LEC2-LEC1D86K	16,49	12,44	18,15			
LEC1-ABI3	53,384	48,64	39,88	51,00	0,0032	**
LEC1D86K-ABI3	30,82	26,56	29,46			
LEC2-LEC1-ABI3	84,682	75,18	67,24	84,01	0,0058	**
LEC2-LEC1D86K-ABI3	34,50	56,52	54,89	54,2		
LEC1-NF-YC2	8,090	6,22	7,55	5,86	0,04	*
LEC1D86K-NF-YC2	4,60	5,49	5,20			
LEC2-LEC1-NF-YC2	35,786	35,05	21,55	42,03	0,017	*
LEC2-LEC1D86K-NF-YC2	16,96	14,78	14,88			
LEC1-ABI3-NF-YC2	64,875	47,65	46,40	70,02	0,0049	**
LEC1D86K-ABI3-NF-YC2	25,10	18,51	23,30			
LEC2-LEC1-ABI3-NF-YC2	95,947	85,39	78,63	103,98	0,0004	***
LEC2-LEC1D86K-ABI3-NF-YC2	34,04	30,40	36,86			
LEC2	14,287	15,39	15,48	16,21		
LEC2-LEC1D86K	16,49	12,44	18,15		0,82	NS
LEC2-LEC1D86K-NF-YC2	16,96	14,78	14,88		0,8	NS
ABI3	23,739	21,78	23,19	17,33		
LEC1D86K ABI3	30,82	26,56	29,46		0,014	*
LEC1D86K ABI3 NF-YC2	25,10	18,51	23,30		0,75	NS
LEC2-ABI3	35,874	42,47	39,64	45,44		
LEC2-LEC1D86K ABI3	34,50	56,52	54,89	54,2	0,15	NS
LEC2-LEC1D86K ABI3 NF-YC2	34,04	30,40	36,86		0,057	NS

Figure 6 : Tests d'activité de LEC2 et ABI3 avec le promoteur de l'OLEOSINE1 en présence ou absence de NF-YC2 et de LEC1 (WT) ou LEC1 muté D86K, par la méthode d'expression transitoire chez la mousse *Physcomitrella patens*.

(A) En bleu les résultats avec les protéines sauvages correspondant aux résultats de références (Baud *et al* 2016), en rouge en présence de NF-YC2, en vert LEC1 (WT) est remplacé par LEC1 D86K, en violet activité testée en présence de LEC1 D86K et NF-YC2. (B) Tableau récapitulatif des t-tests N=4 réplicats technique, \* : < 0,05, \*\* : <0,01, \*\*\* : <0,001. (NS) : non significatif.

et C d'*Arabidopsis* en double hybride, a permis de mettre en évidence les interactions de LEC1 et LEC1-Like avec des membres NF-YCs, comme NF-YC2 (Calvenzani *et al.* 2012; Hackenberg *et al.* 2012). Par ailleurs, en système hétérologue, NF-YC2, en présence de LEC1 ou LEC1-LIKE, active l'expression de gènes de réserve (Yamamoto *et al.* 2009). Enfin, NF-YC2 immunoprécipite LEC1 et LEC2 dans des protoplastes d'*Arabidopsis thaliana* (Baud *et al.* 2016).

LEC1 semble avoir une affinité significative avec NF-YC2 notamment dans le contrôle de l'expression de gènes de réserve et interagit avec LEC2 sur *proOLE1*. NF-YC2 pourrait donc être un partenaire du complexe LAL et interagir en trimère avec un facteur de transcription B3 spécifique, en l'occurrence LEC2 ou ABI3.

### **II.2.2.2 Action de NF-YC2 sur le complexe LAL**

Afin de valider cette hypothèse et de comprendre l'interaction de NF-YC2 avec le complexe LAL, son activité a été étudiée en expression transitoire chez la mousse, en présence des protéines LAL et du promoteur de l'*OLEOSINE1* (Figure 6).

NF-YC2 seul n'a pas d'activité sur le *proOLE1*, mais induit faiblement l'activité avec LEC1. En présence de LEC1 muté D86K avec ou sans NF-YC2, l'activité transcriptionnelle retombe au niveau du bruit de fond. NF-YC2 et LEC1 semblent donc former un dimère NF-YB/C sur *proOLE1*. Les deux NF-Y sont constitués d'un domaine HFM, le dimère pourrait donc se fixer à l'ADN de façon aspécifique. De plus, l'acide aminé D86 de LEC1 est bien impliqué dans cette induction minimale. Un autre membre de la famille NF-Y endogène de *Physcomitrella patens* pourrait, ici, également jouer un rôle dans l'activité du dimère sur la sonde.

L'association de NF-YC2 au complexe LEC1, ABI3, LEC2 ne montre pas d'effet significatif au seuil de 5%. Cependant, l'absence d'activité significative de NF-C2 n'exclut cependant pas l'hypothèse que NF-YC2 pourrait être un membre du complexe *in planta* avec un rôle restant à caractériser comme un rôle physique de la stabilité du complexe.

L'acide aminé D86 de LEC1 est essentiel pour l'activité synergique de LEC1 avec les B3 de type LEC2 ou ABI3. Le mutant LEC1D86K n'a plus cette activité avec LEC2, avec ou sans NF-YC2, le niveau d'activité transcriptionnelle étant au même niveau que LEC2 seul. La formation du trimère putatif LEC1D86K, NF-YC2, LEC2 n'a pas d'effet significativement différent de LEC2 seul. NF-YC2 n'intervient pas directement sur l'activité de LEC2.

Cependant, la mutation LEC1D86K n'a pas le même effet sur ABI3 : l'activité transcriptionnelle n'est pas totalement réduite à l'activité d'ABI3 seule. D'autres acides aminés de LEC1 pourraient être impliqués dans l'interaction avec ABI3. Lorsque LEC1D86K et NF-YC2 sont en présence d'ABI3, l'activité transcriptionnelle est réduite à l'activité d'ABI3 seule. La présence de NF-YC2 pourrait réduire l'interaction de LEC1 avec ABI3. NF-YC2



n'aurait d'interaction directe qu'avec LEC1 en absence de l'acide aminé essentiel de LEC1 (D86) et il éliminerait totalement l'effet de LEC1 sur ABI3.

La présence de NF-YC2 a un effet réducteur plus fort que LEC1D86K seule, l'activité sur le promoteur étant limitée à une activité inférieure à celle de LEC2 et ABI3 associés. NF-YC2 semble donc avoir une importance majeure dans l'interaction de LEC1 avec les B3 et plus particulièrement avec ABI3. Il n'est cependant pas possible de conclure sur l'effet direct de NF-YC2 sur les B3 car les changements observés ne sont pas significatifs au seuil de 5%.

### II.2.3 Discussion

NF-YC2 immunoprécipite avec LEC1 et LEC2 exprimés en système hétérologue (Baud *et al.* 2016). NF-YC2 semble donc former un trimère avec LEC1 et LEC2.

Les résultats préliminaires n'étant pas tous significatifs, les interprétations restent hypothétiques. NF-YC2 a un très léger effet sur le complexe LAL sur *proOLE1* et pourrait intervenir dans l'activité du complexe LAL par l'intermédiaire de LEC1 et, plus particulièrement, lorsque LEC1 est liée à ABI3 (Figure 6). Lorsque ABI3 est en présence du LEC1 mutée D86K, l'activité transcriptionnelle est supérieure à l'activité d'ABI3 seule, contrairement à LEC2 qui, en présence de cette même protéine, n'a plus aucun effet synergique. En présence des 2 B3, le dimère inactif (LEC1D86K/NF-YC2) a un effet de réduction de l'activité transcriptionnelle, qui est inférieure à l'effet LEC2 + ABI3.

Il existe des protéines homologues des NF-Ys chez la mousse *Physcomitrella patens*. Ainsi 2 NF-YA, 9 NF-YB, 12 NF-YC y ont été identifiés, dont certaines sont impliquées dans la réponse au stress (Zhang *et al.* 2015). Parmi les AFL, 5 copies d'ABI3 (*PpABI3 1 à 5*) ont été identifiées chez la mousse, mais pas d'homologue de *LEC2* et *FUS3* (Li *et al.* 2010). Une étude a montré que *PpABI3A* peut interagir avec des dimères de NF-YB/C et avec NF-YC seul, permettant une activité synergique sur le promoteur de protéine *LEA1* chez la mousse. *AtABI3* et *PpABI3A* ont la même activité sur le promoteur *LEA1*, *AtABI3* étant cependant plus actif que *PpABI3A*. Lorsque ABI3 est en trimère avec un dimère NF-Y B/C, l'activité transcriptionnelle augmente chez *Physcomitrella* avec NF-YB4/C1 et chez *Arabidopsis* avec LEC1 et NF-YC1. Néanmoins, *PpABI3* en dimère avec NF-YC1 est plus actif que le trimère (Yotsui *et al.* 2013). Des protéines NF-YB « non LEC1 » pourraient être impliquées dans une régulation négative de la transcription lorsqu'elles forment un trimère avec un NF-YC et ABI3.

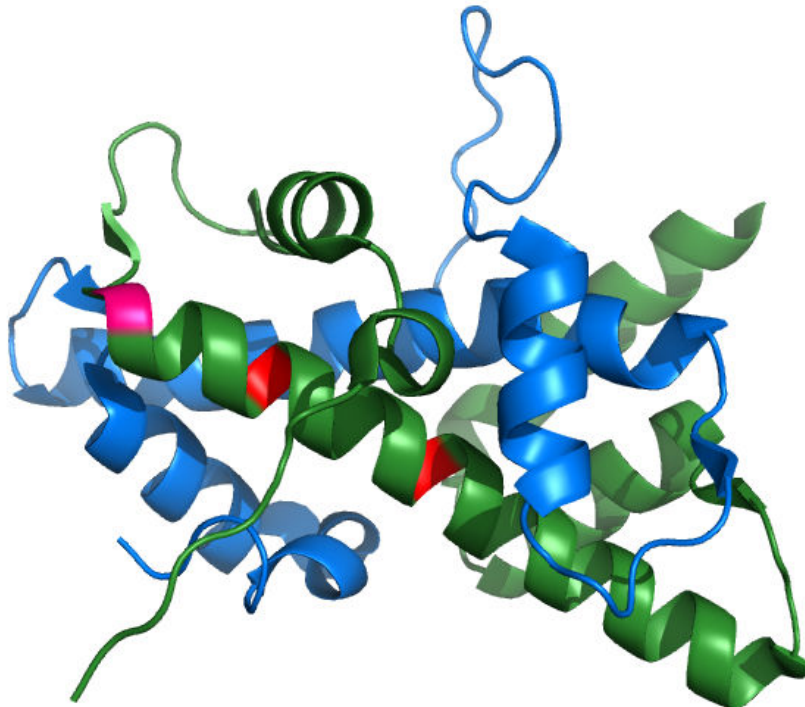
Des interactions avec des dimères NF-YB/C et d'autres protéines que NF-YA ont été démontrées. Elles ont pour conséquence une activité synergique sur l'activité transcriptionnelle des cibles autres que les séquences promotrices reconnues par les NF-Y (Masiero *et al.* 2002; Ben-Naim *et al.* 2006; Gnesutta *et al.* 2017). Par exemple, il a été



A

	1	10	20	30	40	50	60
LEC1	MERGAPFSHYQLPKSISELNLDQHSNNPTPMTSSVVVAGAGDKNNGIVVQQQPPCVAREQ						
LEC1-LIKE	MERGGFHGYRKL SVNNTTSP PGLAAN-FLMAEGSMRPPEFNQPNKTSNGGEEECTVREQ						
	α-1		L1	α-2			L2
	61	70	80	90	100	110	120
LEC1	DQYMPIANVIRIMRKTLP SHAKISD DAKETIQECVSEYISFVTGEANERCQREQRKTI TA						
LEC1-LIKE	DRFMPIANVIRIMRRILPAHAKISD SKETIQECVSEYISFITGEANERCQREQRKTI TA						
	α-3		α-4				
	121	130	140	150	160	170	180
LEC1	EDILWAMSKLGF DNYVDPLTVFINRYREIETDRGSALRGEPPSLRQTYGGNGIGFHG PSH						
LEC1-LIKE	EDVLWAMSKLGFDDYIEPLTLYLHRYRELEGERGVSCSAGSVSMTNGLVVKRPNGTMTEY						
	-----						
	181	190	200	210	220	230	238
LEC1	GLPPPGPYGYGMLDQSMVMGGGRYYQNGSSGQDESSVGGGSSSSINGMPAFDHYGQYK						
LEC1-LIKE	GAYGPVPGIHMAQYHYRHQNGFVFSGNEPNSKMSGSSSGASGARVEVFPTQQHKY---						

B



**Figure 7 : Séquences et structures de NF-Y B et C.**

(A) Alignement des séquences de LEC1 (NF-YB9) et LEC1-Like (NF-YB6), en rouge les acides aminés conservés, en bleu l'acide aspartique 86 (D86) conférant l'activité LEC1, les boîtes noires représentent les domaines conservés des NF-YBs. (Lee et al 2002 ; Hackenberg et al 2012), (B) Structure 3D de LEC1-Like (vert) et NF-YC3 (bleu) (Gnesutta et al 2017), en rouge les acides aminés conservés chez LEC1 et LEC1-Like présents sur l'hélice  $\alpha$ -2, en magenta D86.

montré que MADS-box 18, du riz, dimérise avec le NF-YB1 et est capable de former un trimère avec NF-YC. NF-YB1 de riz a une forte similarité avec LEC1 mais ne contient pas d'acide aminé spécifique d'une activité LEC1 (Masiero *et al.* 2002). D'autres protéines semblent former un trimère avec les dimères NF-YB/C. Notamment CONSTANS (CO) et CONSTANS-Like (COL), qui sont des facteurs de transcription impliqués dans la floraison (Ben-Naim *et al.* 2006; Gnesutta *et al.* 2017). Ces protéines ont la faculté de former des trimères avec les dimères NF-YB/C qui ont une plus grande affinité pour les éléments cis reconnus par CO et COL. La présence de NF-YA permettrait de créer une compétition et, ainsi, induire une régulation de la transcription des gènes cibles (Ben-Naim *et al.* 2006; Gnesutta *et al.* 2017). Ces interactions semblent être de type protéine-protéine et ont pour conséquence l'augmentation de l'affinité pour les éléments cis (Mantovani 1999).

Des bZIPs formant un trimère avec les NF-YB et C ont aussi été identifiées. Par exemple bZIP67, en présence de LEC1 ou LEC1-LIKE et NF-YC2, ce qui permet d'augmenter l'activité transcriptionnelle de gènes de réserves (Mendes *et al.* 2013; Yamamoto *et al.* 2009). bZIP28, qui est impliqué dans la réponse au stress, interagit avec NF-YB3 et NF-YC2 (Liu and Howell 2010). Il n'a cependant pas été démontré que les bZIPs remplacent le NF-YA et, dans la négative, les bZIPs pourraient jouer le rôle de cofacteurs et augmenter l'affinité des NF-Y pour leurs cibles (Liu and Howell 2010; Mantovani 1999). Les NF-Ys, LEC1 et NF-YC2, interagissent donc avec de nombreux partenaires qui pourraient jouer un rôle dans l'effet synergique de l'activité transcriptionnelle des B3 (LEC2 et ABI3) sur *proOLE1*.

Afin de valider l'hypothèse d'une modulation de l'activité des B3 sur le promoteur via la composition du dimère de NF-YB/C, d'autres partenaires non LEC1 pourraient être étudiés avec NF-YC2 qui a une légère tendance à être associé au complexe LAL. NF-YC2 et LEC1 en double hybride dimérisent avec d'autres membres, notamment les NF-YB 1 et 4 pour NF-YC2 et les NF-YC3 et 9 pour LEC1 (Calvenzani *et al.* 2012; Hackenberg *et al.* 2012). L'activité de ces autres partenaires potentiels pourrait être aussi testée pour la synergie de LEC1 ou LEC1 muté, avec des B3.

Le dimère de NF-YB/C se forme grâce au domaine HFM, principalement à travers les hélices  $\alpha$  2 du domaine (Mantovani 1999; Nardini *et al.* 2013) (Figure 7). Afin d'étudier l'interaction entre LEC1 et NF-YC2 d'une part et l'affinité de NF-YC2 avec les B3 d'autre part, des mutations ciblées dans les domaines pourraient être réalisées. En particulier, des études chez l'Homme ont montré que les acides aminés Arg108 et 115 chez NF-YB et Arg93 et Asp100 pour NF-YC seraient impliqués dans la dimérisation et la stabilisation du complexe (Hackenberg *et al.* 2012). Des acides aminés retrouvés uniquement chez LEC1 et LEC1-Like, qui semblent impliqués dans l'interaction des hélices  $\alpha$ -2, tels que I91 et les acides

aminés environnants pourraient également être étudiés de plus près (Figure 7). Par ailleurs, l'histidine 80, (publiée précédemment comme étant l'histidine 50 ; H50) qui est spécifique de LEC1 et L1L, et identifiée par alignement de séquence de LEC1 et LEC1-LIKE d'arabidopsis et HAP3 de levure (Kirkbride *et al.* 2013), semble avoir, dans l'analyse de structure de complexe entre L1L, NF-YA et NF-YB, une importance particulière pour les interactions avec l'ADN (Gnesutta *et al.* 2016), sans que cet acide aminé ne montre cependant une importance particulière *in planta* (Lee *et al.* 2003). Néanmoins NF-Y B et C semblent interagir avec d'autres protéines plus grosses que les NF-YAs (22 à 37 KDa), ayant des domaines de liaison à l'ADN comme CO (41,9 KDa), COL (51,2 KDa) (Ben-Naim *et al.* 2006; Gnesutta *et al.* 2017) ou les facteurs de transcription à domaine B3 tels que LEC2 (41,7KDa) et ABI3 (79,5 KDa), dont la structure et le domaine de liaison à l'ADN diffèrent beaucoup des NF-YA. Il est à noter que la structure du complexe NF-Y a été résolue sur la base de la fixation à la CAAT box. Ceci suggère une modification des interactions protéine - protéine et protéine - ADN possibles pour les NF-Y en fonction des partenaires. Afin de comprendre ces interactions et le mode d'action sur l'ADN, la résolution de la structure du complexe B3-NF-YB/C en présence d'un ADN ne contenant pas la CAAT box mais les sites de fixation de LEC2/ABI3 semble indispensable.

## II.3 Caractérisation de l'interaction LEC1/LEC2

### II.3.1 Introduction

Cent dix-huit gènes à domaine B3 ont été recensés chez *Arabidopsis thaliana* (Swaminathan *et al.* 2008). Une protéine contenant un domaine B3 a été trouvée chez l'algue verte, 5 ABI3 chez *P.patens*, suggérant que le domaine B3 est un domaine conservé dans le règne végétal (Romanel *et al.* 2009; Zhang *et al.* 2015). La famille des CCAAT box - binding factor (CBF) connue aussi sous le nom HAP ou NF-Y, a été retrouvée chez les mammifères et identifiée chez de nombreux eucaryotes : levure, humain, plante, mousse (Lotan *et al.* 1998; Kirkbride *et al.* 2013; Mantovani 1999). L'étude du rôle des facteurs de transcription dans le développement de la plante reste d'actualité, notamment les LAFL, fortement impliqués dans le développement de la graine (Santos-Mendoza *et al.* 2008).

Parmi les LAFL, les facteurs de transcription ABI3, LEC1 et LEC2 activent ensemble, de manière synergique, la transcription de *proOLE1* laissant supposer la formation d'un complexe sur *proOLE1* permettant la transcription du gène. De plus, il a été démontré que NF-YC2 immunoprécipite LEC1 et LEC2, *in vitro*, suggérant une interaction directe entre LEC1 et LEC2, qui a été confirmée par l'étude de l'interaction entre LEC1 et LEC2 sur *proOLE1* par gel retard (Baud *et al.* 2016).

Afin de caractériser l'interaction entre LEC1 et LEC2, le chapitre suivant porte (sous la forme d'un article soumis) sur la formation du complexe, son effet sur la liaison à l'ADN, ainsi que la localisation cellulaire de chacune des protéines. Ces études ont permis de caractériser d'un point de vue biochimique l'interaction des protéines LEC1 et LEC2.

### **II.3.2 Article de résultats:**

**Understanding molecular interactions in the LAFL complex that controls seed development.**

Boulard C.<sup>1</sup>, Thévenin J.<sup>1</sup>, Tranquet O.<sup>2</sup>, Laporte V.<sup>2</sup>, Lepiniec L.<sup>1</sup> and Dubreucq B.<sup>1\*</sup>

**Running title: LEC1 (NF-YB9) controls LEC2 activity through the conserved B2 domain**

<sup>(1)</sup> Institut Jean-Pierre Bourgin (IJPB), INRA, AgroParisTech, ERL-CNRS, Saclay Plant Sciences (SPS), Université Paris-Saclay, RD10, F-78026 Versailles, France

<sup>(2)</sup> UR1268 BIA, INRA Angers-Nantes rue de la geraudiere 44316 Nantes cedex 3

\* Author for correspondence

**Key words:** embryo, transcription, nucleus, maturation, protein complex

**Running title:** LEC1 increases LEC2 activity through the B2 domain.

**Word count:** 5416, excluding references and legends to figures

**Abstract:**

The LAFL transcription factors LEC2, ABI3, FUS3 and LEC1 are major inducers of seed development. LEC2, ABI3 and FUS3 are closely related proteins that contain a B3-type DNA binding domain. We have previously shown that LEC1 (a NF-YB type protein) can increase LEC2 and ABI3 but not FUS3 activity. Interestingly, LEC2 and ABI3 contain a B2 domain, the function of which remains elusive. Consistent with a possible complex between LEC2 and LEC1, we showed that LEC1 and LEC2 partially co-localised in the nucleus of developing embryos. By comparing protein sequences from various species, we identified within the B2 domains a set of highly conserved residues (*i.e.* TKxxARxxRxxAxxR). We showed that this domain directly interacts with LEC1 in yeast. Mutations of the conserved amino acids of the B2 domain abolished this interaction both in yeast and in moss protoplasts. Conversely, the mutations of key aminoacids for the function of LEC1 *in planta* (D86K) prevented the interaction with LEC2. Last, we demonstrated *in vitro* (by thermophoresis) that the addition of LEC1 increases significantly (four fold) the affinity of LEC2 for DNA target sequences. Taken together, these results provide molecular evidences for the binding of LEC1 to transcription factors containing B2-domain such as LEC2 or ABI3, to form specific heteromers with higher affinity to DNA, involved in the control of the seed maturation program.

**Introduction:**

During seed development, the zygote passes through a series of stages that are under tight transcriptional controls. Several transcription factors that form a complex and intricate network (for review see (Fatih *et al.* 2016; Devic and Roscoe 2016; Carbonero *et al.* 2016) have been characterized. Among Them, the LAFLs are encoded by 4 genes, namely *LEAFY COTYLEDON 1* (*LEC1*)(Lee *et al.* 2003), *ASCISSIC ACID INSENSITIVE 3* (*ABI3*)(Giraudat *et al.* 1992), *FUSCA3* (*FUS3*)(Luerssen *et al.* 1998) and *LEAFY COTYLEDONS 2* (*LEC2*)(Stone *et al.* 2001). Genetic and molecular analyses have demonstrated that they control many aspects of seed development such as the establishment of reserves (To *et al.* 2006; Roscoe *et al.* 2015; Kroj *et al.* 2003; Bäumllein *et al.* 1994), the acquisition of tolerance to desiccation (Ooms *et al.* 1993; Khandelwal *et al.* 2010), the

inhibition of trichomes growth on the cotyledons or the repression of anthocyanin synthesis (Meinke *et al.* 1994). Besides their implication in seed development, these genes are involved in somatic embryogenesis (Gaj *et al.* 2005; Jia *et al.* 2013) and, more broadly, in the initiation and maintenance of embryonic fate (Feeney *et al.* 2013; Jia *et al.* 2013).

LEC2, ABI3 and FUS3 are members of the plant specific family of B3 transcription factors (Li *et al.* 2010). They contain a conserved DNA binding domain (the B3 domain) and are of variable amino acid length. ABI3 contains four domains (A1, B1, B2 and B3), LEC2 containing two (B2 and B3) and FUS3 only one, the B3 domain (Han *et al.* 2017). It has been proposed that LEC2 and FUS3 derive from an ABI3-like putative common ancestor (Han *et al.* 2017). The B3 domain is involved in binding to a core DNA cis-element (CATG) called RY motif, with variable extension such as cGCATGCG for FUS3 and gCATGc for LEC2 and ABI3 (Baud *et al.* 2016). Previous analyses suggested that the B2 domain could act as putative nuclear localisation signals, but mutations in this domain do not alter ABI3 localisation (Marella and Quatrano 2007).

LEC1 is a member of the NF-YB protein family that is highly conserved among eukaryots (Gusmaroli *et al.* 2001; Mantovani 1999). It has no specific DNA binding activity but interacts together with NF-YC and NF-YA, which possesses the DNA binding capabilities, to form a complex that binds the CCAAT cis-element involved in the initiation of transcription (Dolfini *et al.* 2012). LEC1 and LEC1 like (L1L), its close paralog possess specific amino acid residues compared to other NF-YB (Lee *et al.* 2003). They can interact with other transcription factors such as bZIPs (Yamamoto *et al.* 2009; Gnesutta *et al.* 2017).

In a previous work we have shown that LEC1 can increase the activity of LEC2 to trigger the activity of a target promoter and that it can co-immunoprecipitated with LEC2 in the presence of NF-YC2 and target DNA sequence. Nevertheless, little is known on the molecular interactions between those transcription factors and biochemical data remain elusive. Here, we have investigated the localisation of LEC1 and LEC2 with specific antibodies to prove partial co-localisation and showed their direct interactions through the B2 domain in two hybrid experiments. Using thermophoresis, we demonstrated that this interaction changes the affinity of LEC2 to its DNA target sequence. Taken together these results allow drawing a model of interaction of LEC1 with the B2-domain of LEC2, increasing the affinity of the complex for target DNA sequences and thus, its activity *in planta*.

## Methods.

### Antibodies production and purification

The antibodies were produced from recombinant protein expressed in *E. coli*. The DE3 rosetta™ bacteria were transformed with the pETG-10A vector allowing the expression of the proteins with a N term histidine tag. LEC1 and LEC2 production were induced with 0.5 mM IPTG, at 37 °C for 3 h. The bacteria were pelleted by centrifugation (5000 g, 15 min at 4 °C), and the pellet was resuspended in a lysis buffer (1X PBS, 1% triton). The bacterial solution was lysed with a sonicator (Sonics, USA) at 60% power over 1 minute through steps of 10 seconds of sonication followed by 10 seconds of incubation on ice. After centrifugation of the lysate (45 min 15000 g at 4 °C), the pellet containing the inclusion bodies was washed with (i) PBS, 2% triton, (ii) PBS 1X 1M NaCl, 1M urea and (iii) 1X PBS, successively.

Antibody production was carried out on the technical platform of the Biopolymers Interaction Assemblies of INRA ([https://www6.angers-nantes.inra.fr/bia\\_eng/Technical-Platforms/Protein-Antibodies](https://www6.angers-nantes.inra.fr/bia_eng/Technical-Platforms/Protein-Antibodies)). Animal experiments were carried out at the INRA facilities, which are authorized by the Local Veterinary Department (authorization no. 44502). Two “New Zealand” rabbits were immunized for LEC1 and eight “Wistar” rats were immunized for LEC2 by repeated intraperitoneal injections with 300 µg of LEC1 and 100 µg of LEC2 respectively, emulsified in Freund's adjuvant (complete for the initial injection, incomplete for the others). Immunizations were repeated four times with two weeks interval. One week after the last injection, animals were bled by cardiac puncture under generalized anaesthesia. Serum were prepared by centrifugation (2000g-20min) and stored at -20°C.

The antibodies obtained were purified by immuno-affinity. Twenty-five µg of the recombinant proteins required for immunization were separated on a SDS-PAGE 8% gel. After transferring the proteins onto a PVDF membrane and staining the latter with red culvert (P7767, Sigma), the coloured strips corresponding to the recombinant proteins were cut out and incubated in the saturation buffer TBST 5% milk (20mM tris, 150mM NaCl, pH 7,4, 0,5% tween, 5% milk) for 1 h at 4 °C. The serum was incubated with the membranes overnight at 4 °C. After 4 washings (15 minutes) of the membranes with TBST buffer (20mM tris, 150mM NaCl, pH 7,4, 0,5% tween), the bound antibodies were eluted several times with 500mM glycine, 500mM NaCl, 0.1% tween 20, 1% BSA, pH 3.0. The eluates were immediately neutralized with 0,1 volume of 1M tris base solution. The purified antibodies were



concentrated on Amicon Ultra 10000 MWCO Millipore™ filtration units and stored at -20 °C until further use.

## **Cytological analyses**

### **Preparation of the samples:**

The developing siliques of Arabidopsis plants were picked (2 to 14 DAP) and incubated for 2 h at 4 °C. in the MTSB fixing solution (50 mM PIPES, 5 mM MgSO<sub>4</sub>, 5 mM EGTA pH 6.9, 4% formaldehyde), washed for 30 minutes in 0.5X MTSB buffer without formaldehyde and 15min in 1X PBS buffer.

The samples were dehydrated using increase concentrations of ethanol (30 to 97% in 1X PBS buffer for 30 minutes) and stained for 10 minutes with 0.01% toluidine blue in ethanol at 40 °C. The inclusion steps in wax were carried at 40 °C as follows: (i) 10 minutes in 97% ethanol, (ii) overnight incubation in a 97% ethanol wax (305413, sigma) solution, (ii) last incubation of 2 hours in wax. The wax embedded samples were cut with a Microm™ HM355S microtome with blades 35/80mm (MX35, Thermo electron corporation) to a thickness of 8 µm and recovered on polysine coated slides (20715, Thermo scientific). The slides were de-waxed and samples were rehydrate in successive decreasing (100% to 30% in 1x PBS buffer) ethanol solutions, and then treated with 1M sodium citrate pH 6.0 until boiling in the microwave for epitope unmasking and rinsed twice with 1X PBS buffer. Slides were saturated one hour with 1% BSA, 1X PBS, and incubated overnight at 4 °C in a humid atmosphere with the specific antibody diluted 1:500th in the saturation solution. After 4 washings of 3 minutes in PBST (1 × PBS, 0.1% Tween 20), the slides were incubated 3H at 37 °C. in a humid atmosphere with the 1:100<sup>th</sup> goat Alexa labelled secondary antibody (Anti-rabbit Alexa 488 ref: A11008, Anti-rabbit Alexa 647 ref: A21245 , Anti mouse Alexa 488 ref: A11001, Anti rat, Alexa 488 ref: 11006). After two 10-minutes wash in a 1X PBS, the slides were mounted in H1000 vectashield (Vector laboratories INC) mounting medium containing 1 µg / ml DAPI and then sealed with varnish. The slides were stored at 4 °C until observation (carried out as described hereafter).

### **Confocal observations.**

The developing siliques of the GFP:LEC2 lines were harvested from 3 to 8 DAP. The embryos were extracted from the seeds under the binocular and placed on slides in a mounting medium (50% glycerol). Embryos were observed using a LEICA SP8 TCP spectral confocal laser microscope (Leica

Microsystems, Wetzlar Germany) equipped with an argon laser, 405nm and 633nm laser diodes and hybrid photodetectors. Emissions bands were 415-459 nm for DAPI, 503-527 nm for Alexa 488, 659-675-459 nm for Alexa 647 for fixed observations, emissions bands of 426-464 nm for DAPI, 490-522 nm for GFP for *in vivo* observations. The same equipment was used for observations of immunohisto detection of LEC1 and LEC2.

#### **Construction of GFP: LEC2 lines:**

The *GFP* and *LEC2* cDNAs were amplified by PCR with the Phusion High-Fidelity polymerase (New England Biolabs, Ipswich, UK). The oligonucleotides were designed to position double Gateway recombination cassettes (Thermo Fisher,) according to the following pairs: (AttB1 AttB5r), (AttB5, attB2) respectively.

The cDNA were BP-gateway recombined in p221 donor vectors according to the manufacturer's recommendations and then LR-gateway recombined into the desired destination vector.

#### **Construction of pGWB2:pLEC2 (-1Kb) destination vector:**

In order to express the tagged proteins in plants, the cDNA were placed under the control of the endogenous promoter, in the modified pGWB2 vector. The p35S promoter, initially present in the vector, was replaced by the 1kb length *LEC2* promoter by enzymatic digestion. Enzyme restriction sites HindIII (R014S, Biolabs) and XbaI (R0145S, Biolabs) were placed at the 5' and 3' ends, respectively, of the *LEC2* promoter amplified by PCR, and then inserted into the Topo Blunt vector (ThermoFisherScientific). PCR product and the vector pGWB2 were digested successively by 2 units of each enzymes and then purified on gel using the Promega Wizar SV clean UP system kit. The ligation reaction was carried out according to a 9:1 PCR / vector product ratio with 800 units of T4 ligase (New England Biolabs), at 16 °C, overnight. DB3.1 bacteria were transformed with the ligation product and colonies obtained PCR checked with primers located in the *LEC2* promoter sequence.

#### **Plasmids constructions**

*LEC1*, *LEC1D86K* and *LEC2* cDNAs in pDONR207 were directly LR-gateway recombined in pCor104 PMDC140 R1R2. *LEC2* T105G, K106S, A108T, R109N, I110C, K113S was de novo synthesized (Geneart, DNA strings, Thermofisher, Regensburg, Germany). The sequences were PCR amplified to

introduce attB1 attB2 Gateway recombination sequences at position 5 'and 3' of the cDNA respectively (see sup table for oligonucleotides sequence) BP-Gateway in pDONR207 after verification of the sequence and then recombined LR-recombined in pCor104 PMDC140 R1R2. For Yeast two-Hybrid analyses, the cDNAs of *LEC1* and *LEC1D86K* contained in pDONR207 were directly LR-Gateway recombined in pDEST22 and pDEST32 vectors (ThermoFisher Waltham, USA). *LEC2* was mutated in the B2 domain by site-directed mutagenesis according to the Agilent quick exchange protocol according to the manufacturer's instructions. *LEC2* and *LEC2 TGA*, deleted from domain B3, were PCR amplified (see sup table for oligonucleotides used), verified by sequencing, BP-Gateway recombined in pDONR207. The inserts were then LR-Gateway recombined in the double hybrid vectors pDEST22 or pDEST32.

#### **Plant transformation:**

The destination vectors were transferred to *Agrobacterium tumefaciens* C58C1pMP90 by electroporation and selected according to their antibiotic resistance. The *lec2-1* mutant (col-0) was transformed by agro-infiltration of the flower buds as described in Bechtold *et al.*, (1993). Primary transformants were selected on Murashige and Skoog medium containing hygromycin (50 µg.mL<sup>-1</sup>) in growth chambers (16h day, 8h night at 25 °C). The analyses were conducted using plants homozygous for T-DNA insertions.

#### **Yeast two Hybrids experiment:**

The yeast strain AH109 (Clontech, Mountain view, USA) was cultured on YPDA medium (Clontech, Mountain view, USA) at 28 °C. A 50 mL culture reaching an OD 600 between 1 and 2 was collected, washed with sterile water, recuperate by centrifugation (maximum speed for 15 seconds) and suspended in 500 µL of 0.1 M lithium acetate, 500 ng of each plasmid construct were mixed with 10 µL of carrier DNA (15632-01, Invitrogen), 50 µL of cell solution of yeast and 36 µL of 40% [v / v] lithium acetate of polyethylene glycol 4000. After incubation for 30 min at 28 °C, the transformation mixture was subjected to a heat shock at 42 °C. for 25 minutes. After centrifugation (at maximum speed for 15 s), the pellet was suspended in 200 µL of sterile water and plated on synthetic dextrose-Trp-Leu medium (Clontech, Mountain view, USA). To analyse the interactions, co-transformed yeast cells were transferred into 100 µL of YPDA medium, incubated overnight at 28 °C. in sterile water. Five µL of

transformed yeasts were tested on a selective medium. Colonies growing on a medium lacking His or His and Ala were considered for positive interactions in the presence of various concentrations of 3-aminotriazole.

#### **Transient expression in moss:**

The study of the interaction of transcription factors on a promoter sequence by transient expression in *physcomitrella patens* moss was carried out as previously described by (Thevenin *et al.* 2012). Briefly the transcriptions factors were recombined by simple gateway reaction in pBS Tpp-A vector that allows their expressions under the control of a rice actin promoter. The 250 bp length promoter sequence of *OLEOSIN 1*, containing 2 box RY and 2 box GBOX (Baud *et al.* 2016), was recombined by simple gateway reaction in the vector pBS Tpp-B that allows control of GFP expression. Thus, an interaction between the transcription factors on the promoter sequence will induce promoter activation and GFP expression. The GFP expression was measured by flow cytometry using a PARTEC CyFlow Space instrument (Partec S.A.R.L., Sainte Geneviève des Bois France) as described in Thevenin *et al.* (2012).

#### **Expression in *E coli* and purification of recombinant proteins:**

LEC2 and LEC1 were produced from the vector pET-trx1a (Baud *et al.* 2016). The protein production in transformed *E. coli* (DE3 rosetta) was induced with 0.5 mmol.L<sup>-1</sup> of IPTG at 37 °C for 3h. Purification was adapted from the protocol of (Choudhury *et al.* 2009). The bacterial pellet recovered by centrifugation of 15 min 5000 g at 4 °C was taken up in the solubilisation buffer of the inclusion bodies (50 mM tris HCL pH 8, 300 mM NaCl, 8 M urea, 5 mM imidazole, complete protease inhibitor cocktail Ultra mini EDTA free rock). After lysis of the bacteria to the sonicator (10 times 30 seconds at 60% power), the lysate was incubated overnight at 4 °C with slow stirring and then centrifuged for 45 min 20000 g at 4 °C. The histidine-tagged proteins recovered in the supernatant were incubated for 2 h with the nickel-nitrilotriacetic acid agarose resin (Ni-NTA) (Qiagen), previously equilibrated with the solubilization buffer. The purification was carried out at 4 °C, the resin was washed with 3 column volumes of buffer containing increasing concentrations of imidazole (50 mM tris-HCl pH 8, 300 mM NaCl, 8M urea, 20 mM-50 mM imidazole). The protein was eluted with 5 column volumes of the buffer

containing 250 mM imidazole. The eluate was immediately dialyzed with 3 baths of 1X PBS buffer and then concentrated with Amicon Ultra 10000 MWCO filtration units (Millipore). Supplementary figure 4.

### **Quantifying DNA / protein interactions by Thermophoresis.**

The 108 bp double stranded probe of the *proOLE1* promoter contains 3 Gboxes (underlined) and 3 RY-Like domains indicated in bold **CAAACGIGTCTTTGAATAGACTCCTCTCGTTAA CACATGCAGCGGCTGCATGGTGACGCCATTAACACGTGGCCTACAATTGCATGATGTCTCCATT GACACGTGACT**. The probe was prepared with two oligonucleotides of 80bp that have 48 overlapping bp, labelled with cyanine 5 in 5' end, synthesized by Sigma.

(CAAACGTGTCTTTGAATAGACTCCTCTCGTTAAACACATGCAGCGGCTGCATGGTGACGCCATTAACACGTGGCCTACA and AGTCACGTGTCAATGGAGACATCATGCAATGTTAGGCCACGTGTTAATGGCGTCACCATGCAGCCGCTGCATGTGTTA-3'). An equimolar mixt of each oligonucleotide was heated for 5 min at 95 °C. and was then treated with Klenow according to the supplier's recommendations (Biolabs) The probe was purified on a column with the MinElute Reaction Cleanup Kit (Qiagen). The existence of a single probe population was verified on gel (5% acrylamide in 0.25% v / v glycerol and 0.5% Tris-borate / EDTA, migration from 60 minutes to 120 V), and revealed by fluorescence of cyanine 5 using a Typhoon FLA9500 fluorescence reader according to the Cyanine 5 method with the following parameters: laser at 635 nm, LPR filter and 100 µm resolution (Thermofisher).

The DNA / protein interaction was measured by thermophoresis on a Monolith NT.115 (NanoTemper). The recombinant LEC2 and LEC1 proteins were re-suspended in a 1 × PBS buffer, 0.2 mM n-Dodecyl β-D-maltoside (DDM), concentrated around to 650 µmol.L-1. LEC2 or equimolar mixt of LEC1 and LEC2 were diluted 16 time in a 1: 1 in cascade, in 1 × PBS, 2mM DDM. The cyanine-labeled probe was diluted to a final concentration of 20 nmol.L-1 in binding buffer (10 mM Tris, 150 mM KCl, 1 mM dithiothreitol, 0.05% Nonidet P-40 [v / v] And 0.5 mg of poly [dl-dC], pH 7.5) and distributed in each tube containing the protein. The reaction was incubated for 20 minutes at room temperature in the dark and then loaded into standard capillaries. The capillaries were scanned with an excitation in the red with a LED power at 20%. The thermophoresis measurements of the probe were carried out at 40% laser power at 25 °C constant, with laser on for 30 seconds followed by laser off for 30 seconds several time. The Nanotemper MO. Control V1.4.3 software was used for data analysis

### **Analyses of the recombinant proteins**

The purified proteins were analyzed on SDS-PAGE gel according to (Laemmli 1970). The SDS-PAGE gel (6% acrylamide concentration gel, 12% acrylamide separation gel) was loaded with the sample included in Laemmli 1X final buffer (50mM Tris HCl pH 6.8, 2% SDS, 10% Glycerol, 0.1% BBP, 25mM DTT). Electrophoresis was conducted at 110V constant in 1X migration buffer (25mM tris, 192mM glycine, 0.05% SDS). The molecular weight of proteins is estimated using invitrogen molecular weight marker MARK 12 TM. The gel is then stained with coomassie blue R250, with a staining solution (0,08% coomassie blue G250, 20% ethanol, 1,6% H<sub>3</sub>PO<sub>4</sub>, 8% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>), disclosed with the decolorizing solution in water.

The proteins separated by SDS-PAGE were transferred in buffer (25mM tris, 190mM glycine, 0.02% SDS, 20% ethanol) to a PVDF membrane (GE Healthcare) using a Transblot Transfer System SD Semi dry Biorad. The transfer was carried out at 15V constant for 1 hour. The membrane was then saturated with TBS-T-saturation buffer (20 mM tris, 150 mM NaCl pH 7.40, 5% tween20, 5% milk), 1 h at room temperature and then incubated with the primary antibody (anti-his tag HRP monoclonal antibody – Thermofisher scientific) diluted in TBS-T-milk 1/2000<sup>e</sup>, for 3 h at room temperature. The membrane was finally washed four times in TBS-T for 10 min, followed by a single wash in TBS. The membrane was revealed by chemiluminescence according to the instructions of the chemiluminescent development kit (Thermo Scientific) with the Fujifilm LAS-4000 imaging system. An Invitrogen MagicMark™ Western protein molecular weight standard was loaded onto the gel to estimate band size.

### **Sequences used for alignments:**

Sequences used for multiple alignments (figure 4) can be retrieved at <https://www.ncbi.nlm.nih.gov/protein/> with the following ID numbers: At LEC2 : OAP12706, Th LEC2 : XP\_010540454, Aa LEC2 : KFK44694, Gr LEC2 : XP\_012484717, Et LEC2 : XP\_006415675, Cr LEC2 : XP\_006304381, Bn LEC2 : NP\_001302874, Tc LEC2 : EOY28206, At ABI3 : BAD94272, Os ABI3 : XP\_015629311, Bd ABI3 : XP\_010232723, Et ABI3 : XP\_006418703, Zm ABI3 : ARS22483, Tc ABI3 : XP\_017972070, Mt ABI3 : XP\_013448806, At FUS3 : OAP04174, Tc FUS3 XP\_007031991, Et FUS3 : XP\_006395520, Bn FUS3 : XP\_009152025, Cr FUS3 : XP\_006293066, Po FUS3 : XP\_006373111.

## Results

### LEC2 and LEC1 partially co-localized *in vivo*.

We aimed at characterizing the accumulation pattern of LEC2 protein during embryo development using a reporter GFP-LEC2 translational fusion under the control of the endogenous 1kb *LEC2* promoter (*pLEC2::GFP-LEC2*) (figure 1-A). First we demonstrated that the translational fusion was functional *in planta* since it could rescue the *lec2-1* loss of function allele. Consistent with this result, the translational fusion was also active in moss protoplasts (sup figure 5). From the globular stage on, the fluorescence was mainly detected in the bottom part of the embryo, and in a few cells of the cellularized endosperm. After the heart transition, GFP fluorescence increased in the axis and in the cotyledons with a maximum intensity in the outer layers of the embryo *i.e.* mainly the inner layer adjacent to the epidermis and the epidermis itself (figure 1-A d, e). The fluorescence declined rapidly after the torpedo stage. This pattern is fully consistent with the previously characterized expression of *LEC2* (To *et al.* 2006) and suggests a partial nuclear localization.

To investigate further the localisation of native LEC2 and LEC1, we carried out similar experiments by immunohistochemistry using a specific primary antibody raised against LEC2 and LEC1 (figure 1-B and C). The antibodies showed no signal in the *lec1-10* and *lec2-1* loss of function alleles used for control, respectively (figure 1-B, m,n and 1-C, m,n). The signal obtained with LEC2 antibodies was very similar to the pattern of *pLEC2::GFP-LEC2* accumulation (figure 1-B and sup figure 1). Looking more specifically to nuclei (Figure 1-B, f to l), fluorescence was detected as dots, excluded from the nucleolus, and predominantly peri-nucleolar, with a slight increase of the signal in the nuclei at late globular to heart stage (figure 1-B, k,l). LEC1 accumulation was detected as early as on globular stage (figure 1-C, a-e) with a signal dispersed between the cytoplasm and the nucleus. Fluorescence was detected in most cells without specific cell domains from the early globular to the young torpedo. LEC1 also showed a dotted accumulation pattern in the nucleus from both the embryo and endosperm cells (figure 1-C, f to l). Last a clear but partial co-localisation of LEC1 and LEC2 was found in the nucleus (figure 2)

### The addition of LEC1 increases the affinity of LEC2 for target DNA

To go further in the comprehension of LEC2 and LEC1 interactions we quantified the affinity of LEC2 with the *OLE1* promoter by thermophoresis. By measuring the motion of molecules in a microscopic temperature gradient, thermophoresis is a method of choice to measure molecular interactions *in vitro*. We first tested LEC2 alone with the fluorescent probe. The result obtained (figure 3-A) allowed the determination of an estimated KD value of 274.6 +/- 31.6  $\mu$ M. This result is an estimation of the real KD of LEC2 since the highest level of the plateau could not be reached. The low affinity obtained may explain this result. As expected, LEC1 alone did not show any interaction with the DNA probe (figure 3-A). Nevertheless when an equimolar ratio of LEC1 was added to LEC2, the estimated KD for DNA dropped to 74.3 +/- 23.9  $\mu$ M (figure 3-A, B), suggesting a four-fold increase in DNA affinity.

### **LEC2 and ABI3 share specific amino acid residues in their B2 domain**

Among the LAFL genes, LEC2, ABI3 and FUS3 belong to the B3 family and share high similarities in their B3 DNA binding domain. Besides these similarities, they also exhibit differences including the conserved domains (figure 4-A). These differences could explain the functional specificity of ABI3 and LEC2, which can synergistically activate an *oleosin promoter* with LEC1, whereas FUS3 cannot (Baud *et al.* 2016). Therefore, we focused the analyses on the B2 domain which is present in ABI3 and LEC2 but not in FUS3 (Han *et al.* 2017). The B2 domains from different plant families were aligned showing a few highly conserved aa residues in LEC2/ABI3 proteins (figure 4-B, red residues) and aa residues conserved in both LEC2, ABI3 and FUS3 (figure 4-B, blue residues). All LEC2 and ABI3 proteins analysed carry a TKxxARxxRxxAxR consensus sequence in their B2 domain, suggesting a conserved function for this sequence which is not present in FUS3.

### **The B2 domain of LEC2 and D86 residues of LEC1 are involved in the interaction in yeast**

In order to further characterize the interaction between LEC1 and LEC2, we performed a yeast two-hybrid analysis. The full length LEC2 sequence alone gave a false positive activation. Therefore, a truncated version of LEC2, without the B3 domain (LEC2 R<sub>168</sub>STOP) but with the conserved residues between LEC2 and ABI3 was used to test possible interactions with LEC1. As shown in table 1, the truncated version of LEC2 was able to bind LEC1 (see sup figure 2 for details). When the conserved residues of the B2 domain were mutated (*lec2.m12*: T<sub>105</sub>G A<sub>108</sub>T R<sub>109</sub>N, R<sub>168</sub>-STOP), the interaction decreased. Moreover, it has been previously shown that the LEC1 D<sub>86</sub> residue is required *in planta*



and allows synergistic effect with LEC2 and ABI3 (Lee *et al.* 2003; Baud *et al.* 2016). When expressed together with LEC2 L<sub>166</sub>STOP, the *lec1D*<sub>86</sub>K mutation inhibits the interaction. The combine effect of the two mutations (*i.e.* LEC1 D<sub>86</sub> and LEC2 m12-B3) was even more drastic. Altogether, these data suggest that LEC2 is able to interact with LEC1 through its B2 domain and that the LEC1 D<sub>86</sub> residue is important for this interaction.

### **The role of the B2 domain is confirmed in moss protoplasts**

We investigated whether the conserved residues in the B2 domain are involved in the ability of LEC2 to synergistically activate *proOLE1* in moss protoplasts. When expressed alone, LEC2, but not LEC1, was able to activated *proOLE1::GFP* expression (figure 5). The addition of LEC1 to LEC2 significantly increases the activity of *proOLE1* expression, but has no effect on FUS3. Mutations of the B2 domain (figure 5, *lec2m1*: T<sub>105</sub>G, K<sub>106</sub>S, A<sub>108</sub>T, R<sub>109</sub>N, I<sub>110</sub>C, K<sub>113</sub>S) did not affect the activity of LEC2, but clearly inhibit the induction by LEC1. These results confirm that the conserved residues identified in the B2 domain are involved in the interaction between LEC1 and LEC2 and thus, in the enhancement of LEC2 activity caused by LEC1.

### **Discussion**

LEC1 (a NF-YB protein) together with the B3 regulators LEC2, ABI3 and FUS3 control seed development. LEC1 can act synergistically with LEC2 and ABI3, which both posses a B2-domain, to activate the expression of specific target genes. Moreover, we have previously shown that LEC1 can co-precipitate with LEC2 in the presence of NF-YC2 and a target DNA sequence.

Here we showed by immuno-histochemistry and translational fusion analyses that LEC2 and LEC1 partially co-localized, both in the nucleus and in the cytoplasm. This result is fully consistent with the *in vitro* interaction of the two proteins and suggests that the activity of the complex could be regulated by nuclear-cytoplasmic transport, as the classical eukaryotic NF-Y regulatory complexes. Although the B2 domain possesses a putative NLS signature sequence, point mutations deleting the NLS signal did not change the nuclear localisation of ABI3, nor is the B2 domain able to target a protein to the nucleus by itself (Marella and Quatrano 2007). Interestingly, some LEC1 homolog proteins (NF-YB) can form a dimer with NF-YC in the cytoplasm to enter the nucleus through a specific importin pathway (Kahle *et*

*al.* 2005). Thus, this putative regulatory mechanism remains to be investigated with LEC1. The co-localisation study of LEC1 and LEC2, showing some, but not always, co-localisation of the two proteins is on same line, suggests that LEC1 and LEC2 are involved together in the regulation of common target genes but also in other, and yet unknown, transcriptional complexes.

The important role of the B2 domain has been pointed out by characterizing specific *abi3* alleles (Bies-Etheve *et al.* 1999). Moreover, in VP1, an ABI3 homolog found in cereals, the B2 domain was shown to enhance the *in vitro* DNA binding activity of the transcription factor (Hill *et al.* 1996). Here, using site directed-mutagenesis, we have demonstrated that the interaction between LEC2 and LEC1, and the activation of LEC2 in moss protoplasts, involve a stretch of conserved amino acid residues (TKxxARxxRxxAxxR) of the B2 domain and a specific D86 residue of LEC1. Moreover thermophoresis experiments have demonstrated that LEC1 increases by for fold the affinity of LEC2 for target DNA sequence *in vitro*. These results are fully consistent with the *in vivo* synergistic effect of LEC1 on LEC2 and ABI3 in moss protoplasts.

It has been shown that NF-YB and NF-YC form a dimer that interacts through a specific region of NF-YA called A1A2 linker that binds 2  $\alpha$  helices (sup figure 3-A)(Nardone *et al.* 2016; Gnesutta *et al.* 2017). In LEC1, the interaction platform contains the D86 residue of major importance for the interaction with LEC2 and the activation of LEC2 (sup figure 3-B). Last, we showed that the NF-YA A1A2 linkers share similarities with the conserved B2 amino acids of LEC2 and ABI3 (sup figure 3-C). It has already been proposed that the CONSTANS protein, a plant transcription factor involved in regulation of flowering in Arabidopsis forms a complex with HAP protein (NF-Y homologs), interacting through a putative similar linker domain (sup figure 3-C) (Wenkel *et al.* 2006). Thus, we propose that LEC1 and LEC2 could interact in a trimeric form together with NF-YC using the LEC2 B2 domain as a substitute to the A1A2 linker. We have previously shown that NF-YC2, which is needed for LEC2 and LEC1 pulldown (Baud *et al.* 2016) is probably not directly interacting with LEC2 since i) LEC1 changes the KD of LEC2 *in vitro* without NF-YC2 and ii) LEC1 makes a supershift with LEC2 in EMSA experiments once again without NF-YC2 (Baud *et al.* 2016). It is possible that the addition of NY-YC2 would still increase the affinity of the LEC2-LEC1 complex for DNA, but this cannot be tested by thermophoresis as it would decrease the quantity of LEC2, which is already limiting. To our knowledge, no NF-YC2 mutant has been described with *lec* phenotypes, which would be expected if being a non-redundant regulator of the complex or responsible. One cannot discard the hypothesis

that another NF-YC would have redundant function. It will be interesting to look for the seed phenotype of multiple NF-YC mutants. Structure of the full putative complex (since the B2 domain was not crystallised together with the B3 domain) would open the way to a better understanding of the specificity of the LEC1 and L1L (NF-YB) proteins in plants and their ability to act in complex with, at least, B2 domain containing proteins.

From an evolutionary point of view we can hypothesize that LEC2 and ABI3 have gain the ability to bind LEC1, at least, through their specific B2 domain, that is not the case for FUS3. It was recently proposed that LEC2 and ABI3 derived from a common ancestor and that FUS3 originate late in the AFL family (Han *et al.* 2017). Alternatively, FUS3 may have lost the B2 domain (as it is much shorter than ABI3 and LEC2), keeping the DNA binding domain but not the ability to interact with LEC1. This would give some explanations to the apparent redundancy between B3 AFL, as they kept some but not all of their protein-protein interaction properties and thus are probably engaged in several transcriptional complexes. It would be interesting to further localize ABI3, simultaneously with LEC1 and LEC2 proteins at the intracellular level to monitor the ratio between co-localised and non-colocalised spots and estimate the putative cross talk between the different complexes during seed development.

Finally, ABI3 is a much larger protein than LEC2 and possesses two other domains (A1 and B1). It has been proposed that the A1, B1 and B2 domain of ABI3 forms a co-activator/co-repressor (COAR) domain that delineate a specific function in ABI3: the COAR would control desiccation tolerance, interaction with ABI5, and seed quiescence whereas the B3 would control earlier, and yet not fully understood aspects of seed maturation (Suzuki *et al.* 2014). We can now propose that ABI3 possesses 2 “super-domains”, the COAR for late maturation events and the SEEDEV domain, comprising the B2 and B3, involving interactions with LEC1 and other possible partners for earlier events (figure 3-B). To this respect, LEC2 could be considered as a specialized version of ABI3, with only the SEEDEV domain. The fact that both ABI3 and LEC2 have been conserved through the evolution probably implies that LEC2 is needed for specific but remaining to determine functions (embryogenesis, morphogenesis?). It has to be noticed that LEC2 is not found in monocots and one may suspect VP1, the ABI3 homolog in cereals, to carry both LEC2 and ABI3 functions in this clade.

## Références:

1. Fatihi, A., *et al.*, *Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds*. *Plant Sci*, 2016. **250**: p. 198-204.
2. Devic, M. and T. Roscoe, *Seed maturation: Simplification of control networks in plants*. *Plant Sci*, 2016. **252**: p. 335-346.
3. Carbonero, P., R. Iglesias-Fernandez, and J. Vicente-Carbajosa, *The AFL subfamily of B3 transcription factors: evolution and function in angiosperm seeds*. *J Exp Bot*, 2016.
4. Lee, H., *et al.*, *Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor*. *Proc. Natl. Acad. Sci. U S A*, 2003. **100**(4): p. 2152-2156.
5. Giraudat, J., *et al.*, *Isolation of the Arabidopsis ABI3 gene by positional cloning*. *Plant Cell*, 1992. **4**(10): p. 1251-61.
6. Luerksen, H., *et al.*, *FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana*. *Plant J.*, 1998. **15**(6): p. 755-64.
7. Stone, S.L., *et al.*, *LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development*. *Proc Natl Acad Sci U S A*, 2001. **98**(20): p. 11806-11.
8. To, A., *et al.*, *A network of local and redundant gene regulation governs Arabidopsis seed maturation*. *The Plant cell*, 2006. **18**(7): p. 1642-51.
9. Roscoe, T.T., *et al.*, *Complementation of Seed Maturation Phenotypes by Ectopic Expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in Arabidopsis*. *Plant & cell physiology*, 2015. **56**(6): p. 1215-28.
10. Kroj, T., *et al.*, *Regulation of storage protein gene expression in Arabidopsis*. *Development*, 2003. **130**(24): p. 6065-73.
11. Bäumllein, H., *et al.*, *the FUS3 gene of Arabidopsis thaliana is a regulator of gene expression during late embryogenesis* *Plant J*, 1994.
12. Ooms, J., *et al.*, *Acquisition of Desiccation Tolerance and Longevity in Seeds of Arabidopsis thaliana (A Comparative Study Using Abscisic Acid-Insensitive abi3 Mutants)*. *Plant Physiol.*, 1993. **102**(4): p. 1185-1191.
13. Khandelwal, A., *et al.*, *Role of ABA and ABI3 in desiccation tolerance*. *Science*, 2010. **327**(5965): p. 546.
14. Meinke, D.W., *et al.*, *Leafy Cotyledon Mutants of Arabidopsis*. *Plant Cell*, 1994. **6**(8): p. 1049-1064.
15. Gaj, M., *et al.*, *Leafy cotyledon genes are essential for induction of somatic embryogenesis of Arabidopsis* *Planta*, 2005. **222**(6): p. 977-988.
16. Jia, H., D.R. McCarty, and M. Suzuki, *Distinct roles of LAFL network genes in promoting the embryonic seedling fate in the absence of VAL repression*. *Plant Physiol*, 2013. **163**(3): p. 1293-305.
17. Feeney, M., *et al.*, *Following vegetative to embryonic cellular changes in leaves of Arabidopsis overexpressing LEAFY COTYLEDON2*. *Plant Physiol*, 2013. **162**(4): p. 1881-96.
18. Li, Y., *et al.*, *Stepwise origin and functional diversification of the AFL subfamily B3 genes during land plant evolution*. *J Bioinform Comput Biol*, 2010. **8 Suppl 1**: p. 33-45.
19. Han, J.D., *et al.*, *Evolutionary Analysis of the LAFL Genes Involved in the Land Plant Seed Maturation Program*. *Front.Plant Sci.*, 2017.

20. Baud, S., et al., *Deciphering the Molecular Mechanisms Underpinning the Transcriptional Control of Gene Expression by Master Transcriptional Regulators in Arabidopsis Seed*. *Plant Physiol*, 2016. **171**(2): p. 1099-112.
21. Marella, H.H. and R.S. Quatrano, *The B2 domain of VIVIPAROUS1 is bi-functional and regulates nuclear localization and transactivation*. *Planta*, 2007. **225**(4): p. 863-72.
22. Gusmaroli, G., C. Tonelli, and R. Mantovani, *Regulation of the CCAAT-Binding NF-Y subunits in Arabidopsis thaliana*. *Gene*, 2001. **264**(2): p. 173-85.
23. Mantovani, R., *The molecular biology of the CCAAT-binding factor NF-Y*. *Gene*, 1999. **239**(1): p. 15-27.
24. Dolfini, D., R. Gatta, and R. Mantovani, *NF-Y and the transcriptional activation of CCAAT promoters*. *Crit Rev Biochem Mol Biol*, 2012. **47**(1): p. 29-49.
25. Yamamoto, A., et al., *Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors*. *Plant J*, 2009. **58**(5): p. 843-56.
26. Gnesutta, N., et al., *Crystal Structure of the Arabidopsis thaliana L1L/NF-YC3 Histone-fold Dimer Reveals Specificities of the LEC1 Family of NF-Y Subunits in Plants*. *Mol Plant*, 2017. **10**(4): p. 645-648.
27. Thevenin, J., et al., *A new system for fast and quantitative analysis of heterologous gene expression in plants*. *New Phytol*, 2012. **193**(2): p. 504-12.
28. Choudhury, D., et al., *Production and recovery of recombinant propapain with high yield*. *Phytochemistry*, 2009. **70**(4): p. 465-72.
29. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4*. *Nature*, 1970. **227**(259): p. 680-5.
30. To, A., et al., *A network of local and redundant gene regulation governs Arabidopsis seed maturation*. *Plant Cell*, 2006. **18**(7): p. 1642-51.
31. Kahle, J., et al., *Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin beta and importin 13*. *Mol Cell Biol*, 2005. **25**(13): p. 5339-54.
32. Bies-Etheve, N., et al., *Importance of the B2 domain of the Arabidopsis ABI3 protein for Em and 2S albumin gene regulation*. *Plant Mol. Biol.*, 1999. **40**(6): p. 1045-54.
33. Hill, A., et al., *A conserved domain of the viviparous-1 gene product enhances the DNA binding activity of the bZIP protein EmBP-1 and other transcription factors*. *J Biol Chem*, 1996. **271**(7): p. 3366-74.
34. Nardone, V., A. Chaves-Sanjuan, and M. Nardini, *Structural determinants for NF-Y/DNA interaction at the CCAAT box*. *Biochim Biophys Acta*, 2016.
35. Wenkel, S., et al., *CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis*. *The Plant cell*, 2006. **18**(11): p. 2971-84.
36. Suzuki, M., et al., *Distinct functions of COAR and B3 domains of maize VP1 in induction of ectopic gene expression and plant developmental phenotypes in Arabidopsis*. *Plant Mol Biol*, 2014. **85**(1-2): p. 179-91.

## **Acknowledgments**

This work was mainly funded by ABCEEDS program (ERA- CAPS, FP7 #291864) and by CERES project (ANR-10-BLAN-1238). The IJPB benefits from the support of the Labex Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

## **Author contributions:**

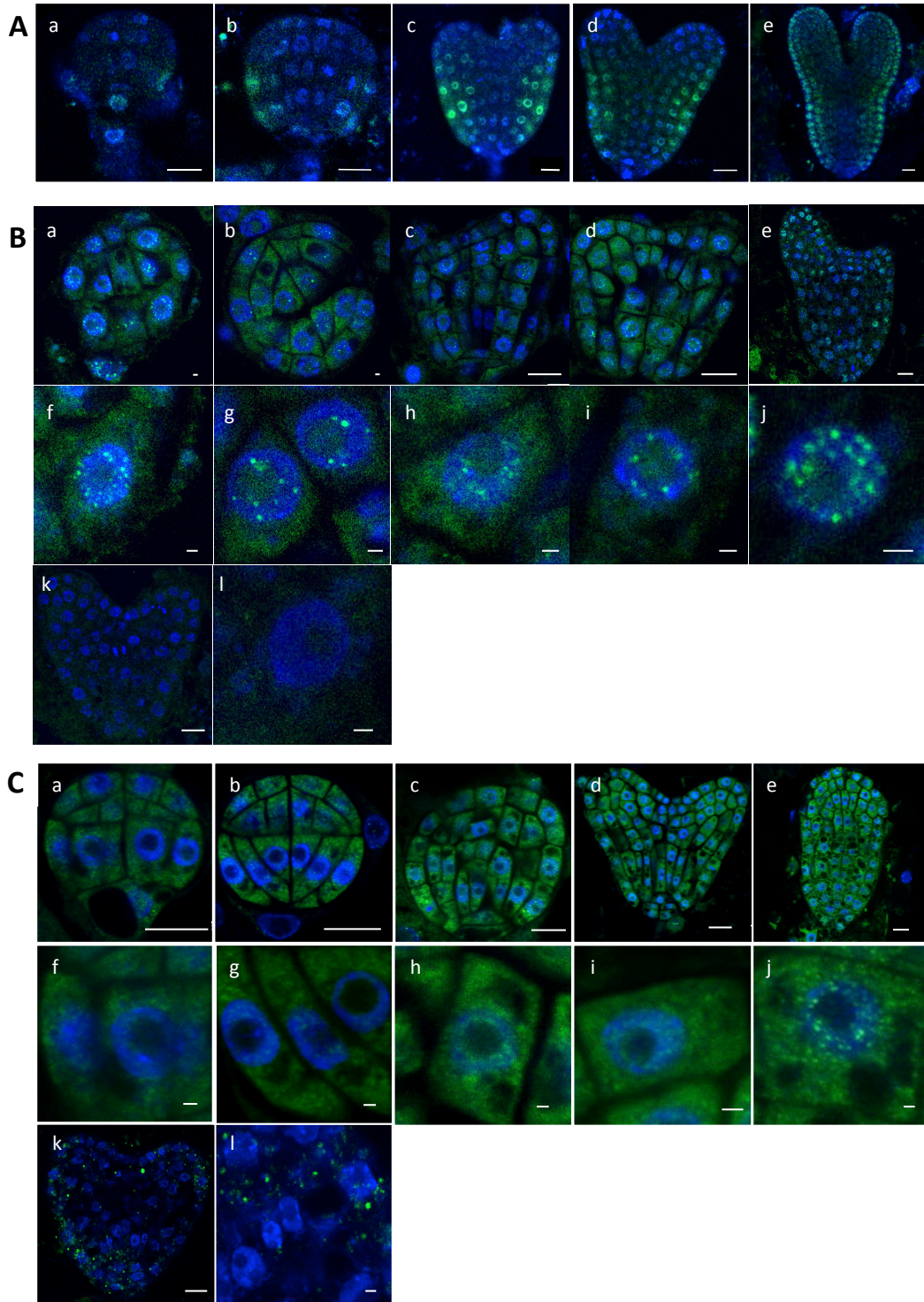
C.B. performed most of the experiments, J.T. and C.B. performed the experiences in physcomitrella, O.T. and V.L. produced the antibodies in small animals, L.L. and B.D. conceived and supervised the study. L.L., B.D. and C.B. participated in writing the paper. All authors approved the final manuscript.

## **Materials & Correspondence :**

The author responsible for correspondence and material requests is : Bertrand DUBREUCQ ([bertrand.dubreucq@inra.fr](mailto:bertrand.dubreucq@inra.fr))

## **Competing financial interest :**

The authors declare no competing financial interests.

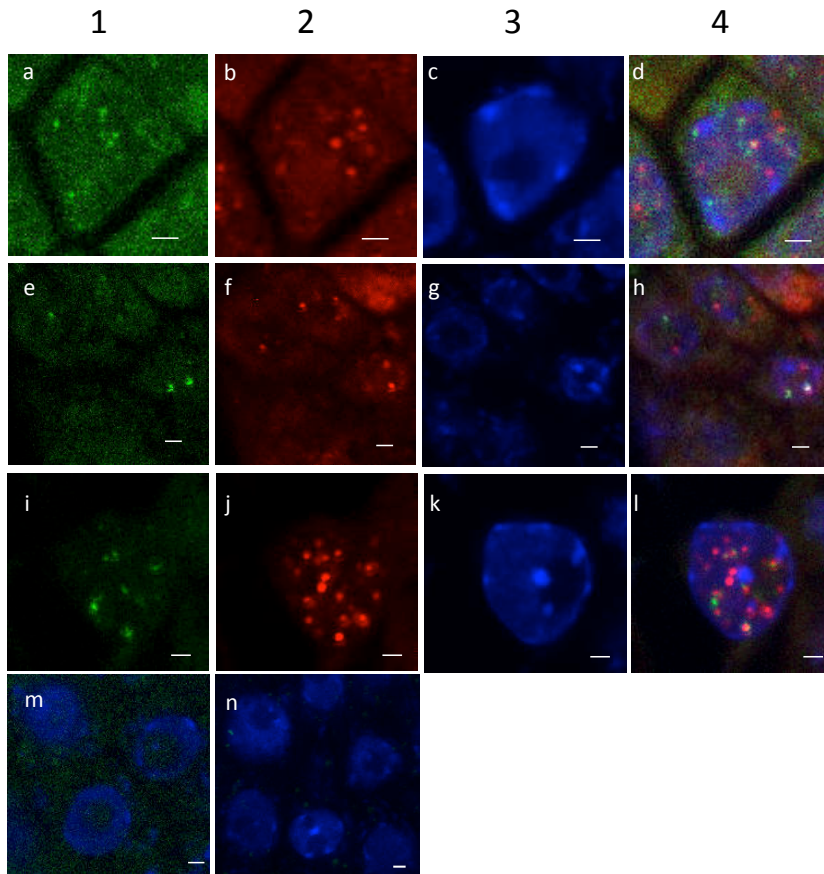


**Figure 1: LEC1 and LEC2 are found in both the nucleus and the cytosol.**

Nuclei are revealed by DAPI staining. **A-** Localization pattern of LEC2 *in planta* using pLEC2::GFP – LEC2 transformed embryos at different stages of development. (a-e) median longitudinal optical sections from embryos at early globular (a), globular (b), heart (c) and torpedo (e) stages of development. **B-** Immunolocalization of LEC2 at different stages of embryo development using a rat anti LEC2 specific antibody and secondary antibody bound to Alexa 488. **C-** Immunolocalization of LEC1 using a rabbit anti LEC1 specific antibody and secondary antibody bound to Alexa 488. For B and C: different stages of embryo development (a-e), (f-j) higher magnification of embryos at the same developmental stage than (a-e) respectively. (k-l) median longitudinal optical section of mutant embryo as control, *lec2-1* for B, and *lec1-10* for C

Bar = 10µm for a-e, l ; 1µm for f-i, l.





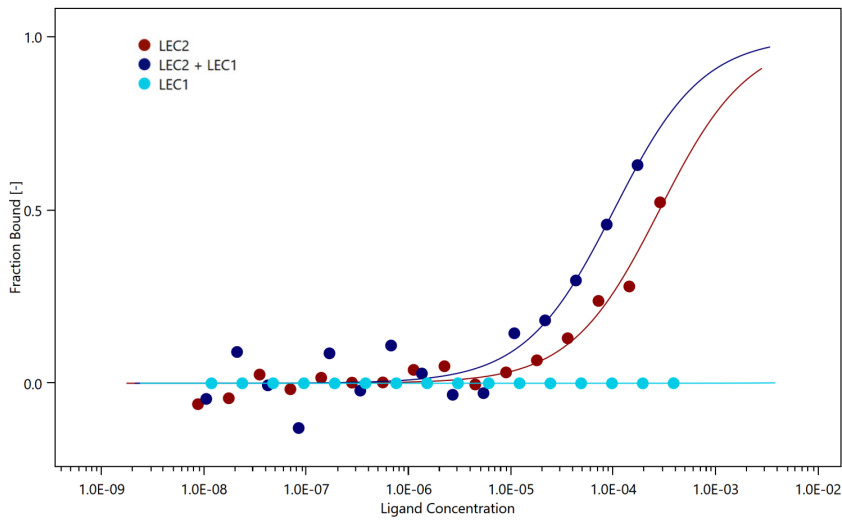
**Figure 2: Partial co-localisation of LEC1 and LEC2**

Immunolocalization of LEC proteins using a rat anti-LEC2 antibody and an anti-rat secondary antibody bound to Alexa 488 and a rabbit anti LEC1 antibody and an anti-rabbit secondary antibody bound to Alexa 647 respectively, at late stage of globular wild-type embryos. a-d and e-h embryo cells, i-l endosperm cells. Controls: *lec2-1 (m)*, *lec1-10 (n)*

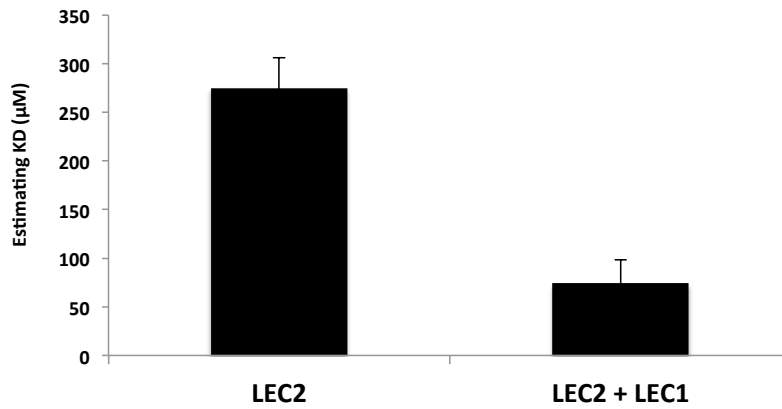
Columns (1), Alexa 488; (2), Alexa 647; (3), DAPI; (4), merged.

Bar = 1 $\mu$ m

A

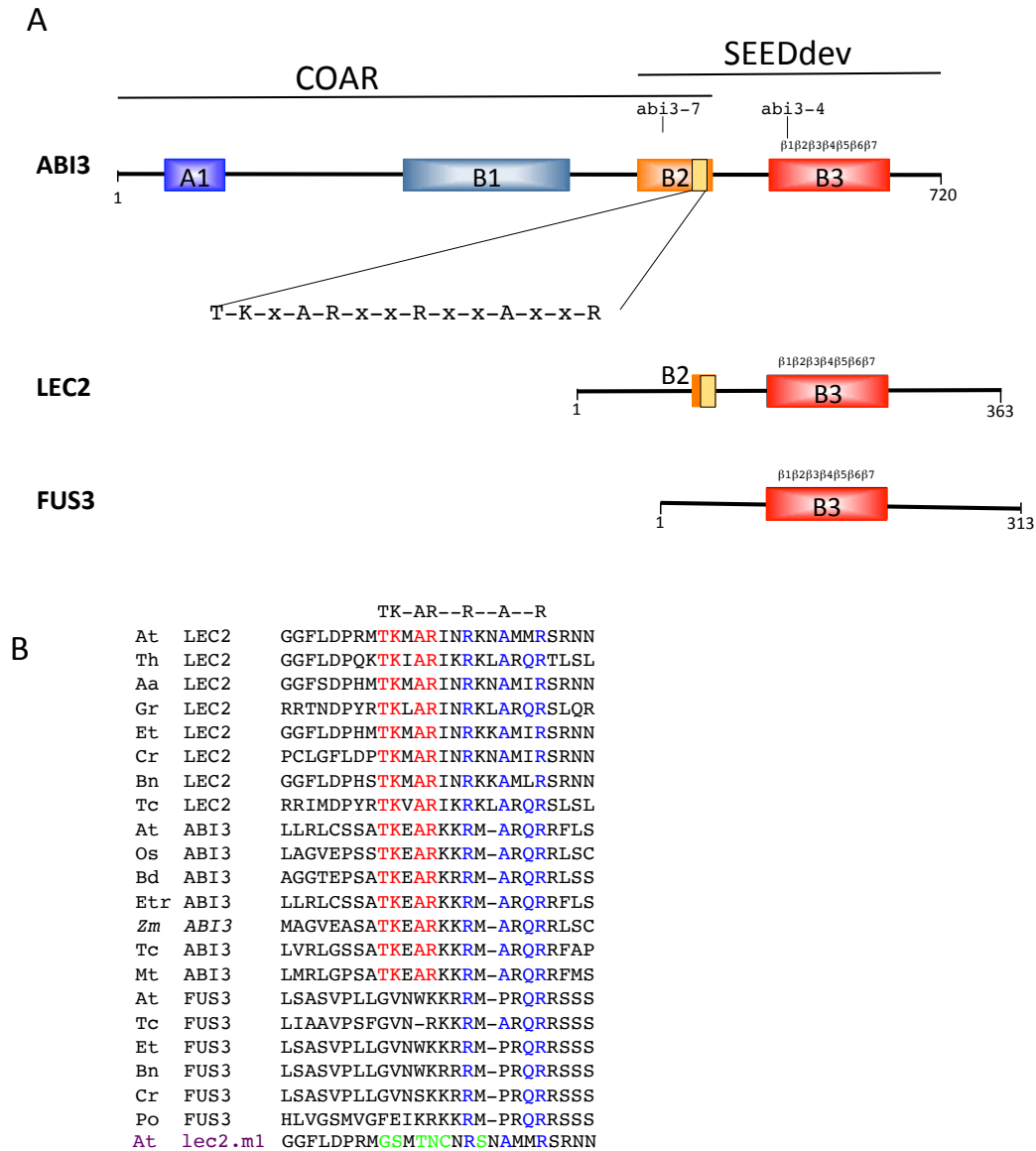


B



**Figure 3: LEC1 increases the affinity of LEC2 for target DNA *in vitro*.**

**A-** A representative curve of thermophoresis experiments carried out using *proOLE1* DNA sequence with LEC2 (red), LEC1 (cyan), or both LEC2 with LEC1 in equimolar quantities (blue). **B-** Effect of LEC1 on the estimated KD of LEC2 with the *proOLE1* (Mean of 5 measurements +/- sd).



**Figure 4: A conserved B2 domain is found in ABI3 and LEC2 homologous proteins.**

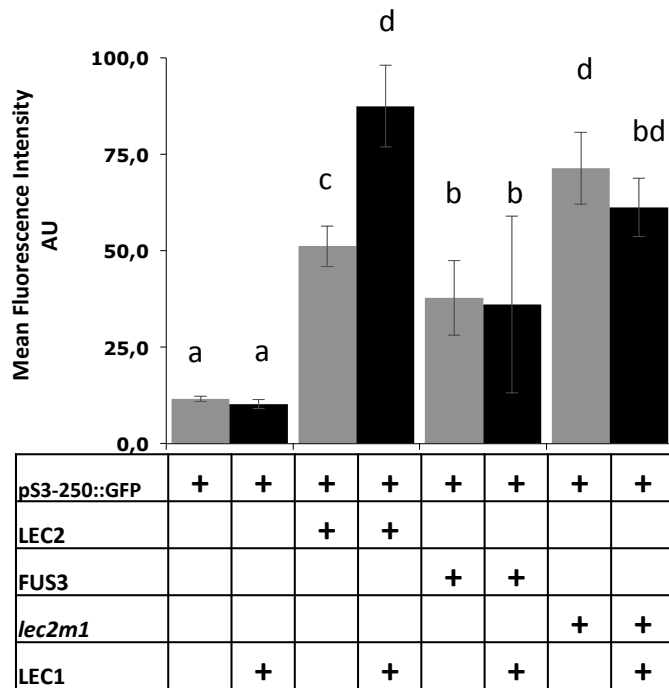
**A-** Structure of the LEC2, ABI3 and FUS3 proteins aligned on the B3 domains. COAR: co-activation/co-repression domains [36]. **B-** An alignment of B2 domains across species identifies conserved residues between LEC2 and ABI3 (red) and all AFL-B3 (blue) homologous proteins. The mutated residues in lec2.m1 are indicated in green (T<sub>105</sub>G, K<sub>106</sub>S, A<sub>108</sub>T, R<sub>109</sub>N, I<sub>110</sub>C, K<sub>113</sub>S).

At: *Arabidopsis thaliana*, Aa: *Abies alpina*, Bn: *Brassica napus*, Bd: *Brachypodium distachyon*, Br: *Brassica rapa*, Ca: *Camellina sativa*, Cr: *Capsella rubella*, Et: *Eutrema salsugineum*, Gr: *Gossypium raimondii*, Mt: *Medicago truncatula*, Os: *Oryza sativa*, Po: *Populus*, TC: *Theobroma cacao*, Th: *Tarenaya Hassleriana*, Zm: *Zea mays*.

AD BD	Empty vector	LEC1	lec1 D <sub>86</sub> K
LEC2R <sub>168</sub> STOP	-	+++	++
<i>lec2.m12 R<sub>168</sub>STOP</i>	-	++	+
empty vector	-	-	-

**Table 1: The B2-domain of LEC2 and D86 residue of LEC1 are required for the interaction in yeast.**

The interactions between LEC1 with LEC2 have been investigated in yeast using two hybrid system, with LEC2 truncated (LEC2R<sub>168</sub>-STOP) and mutated (*lec2.m12-B3* = T<sub>105</sub>G A<sub>108</sub>T R<sub>109</sub>N, R<sub>168</sub>-STOP) proteins and wild type or mutated (*lec1D<sub>86</sub>K*) LEC1. Summary of the interactions is presented in this table. Significant (+, ++, or +++) and no detection (-) of interaction are indicated (the results of yeast growth are presented in sup data 2)

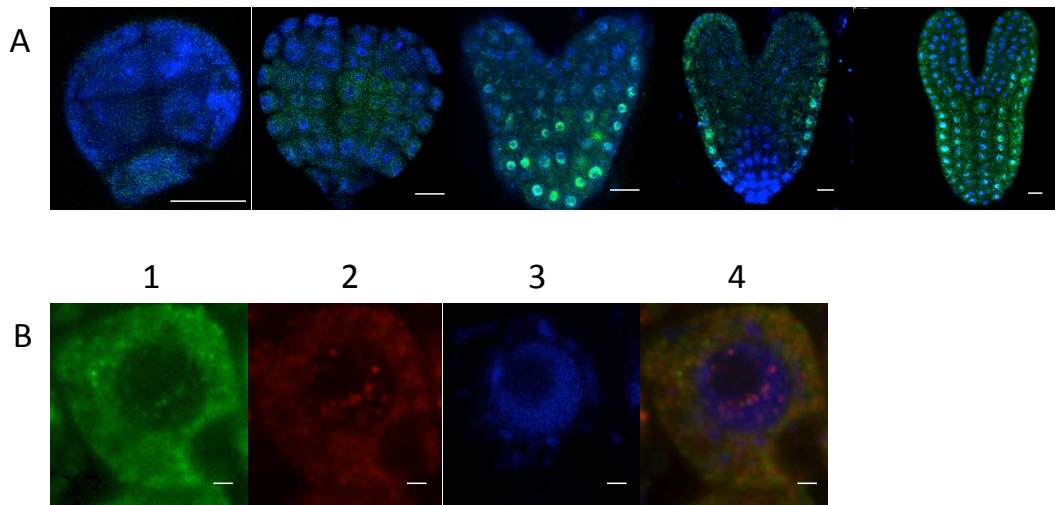


**Figure 5 The B2-domain is required for the induction of LEC2 activity by LEC1**

Transient expression assays were carried out using moss protoplasts transformed with LEC2, *lec2-m1* (T<sub>105</sub>G, K<sub>106</sub>S, A<sub>108</sub>T, R<sub>109</sub>N, I<sub>110</sub>C, K<sub>113</sub>S), FUS3 or LEC1 plasmids, alone or in combination, and the *ProOLE1*<sub>-257</sub>:GFP reporter construct. GFP activities presented are the means of at least three replicates (see material and methods for analyses).

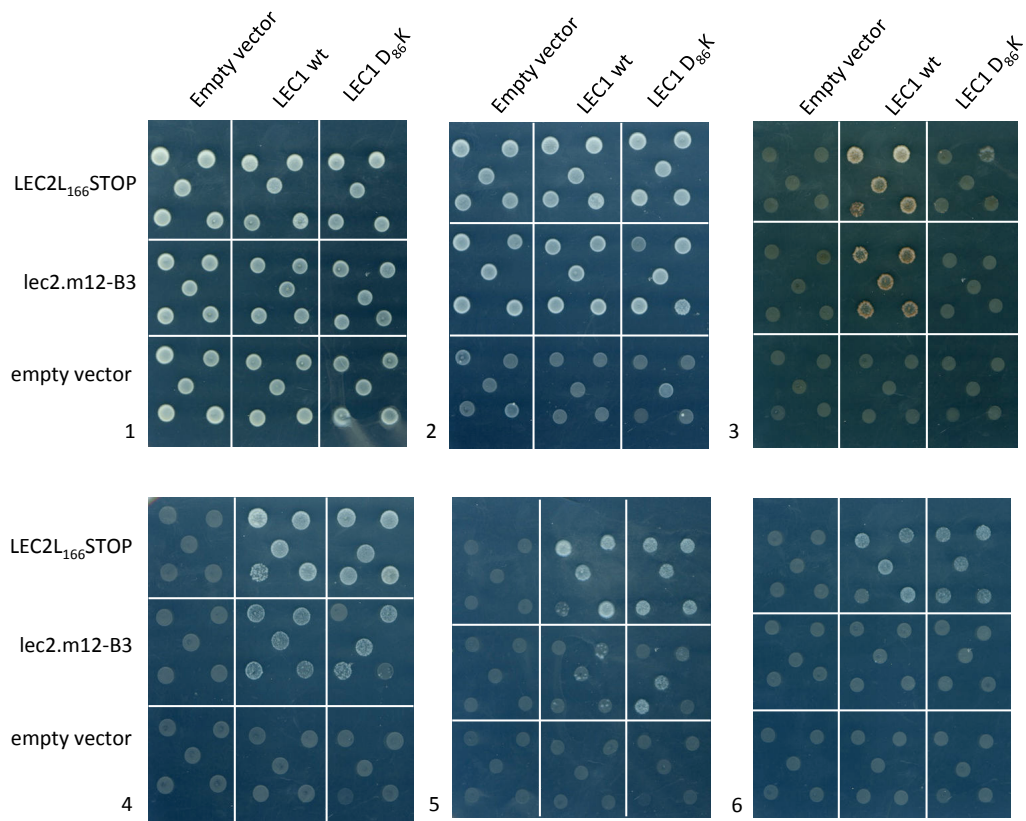
AU : fluorescence arbitrary units

**Supplementary figures:**



**Sup figure 1: LEC2 accumulation in developing embryos.**

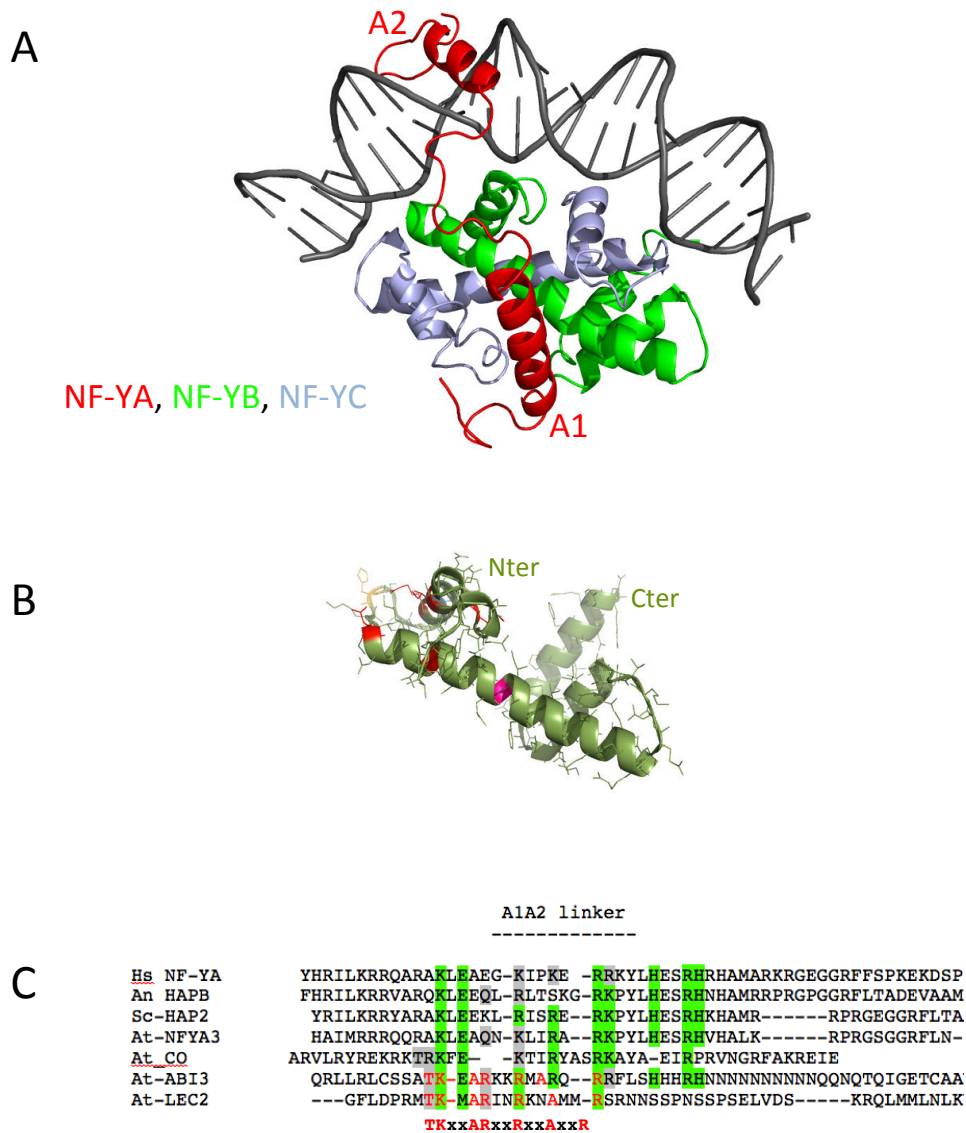
**A-** Second independent T-DNA line carrying pLEC2 (-1kb)::GFP-LEC2 during the seed development. Nuclei were stained with DAPI. Bar = 10  $\mu$ m. **B-** Co-immunolocalization of GFP:LEC2 and native LEC2 in embryo cells using a specific anti-GFP monoclonal primary antibody and a rat anti-LEC2 polyclonal antibody, with a mouse anti-GFP -alexa 488 (green) and a rat anti-LEC2 alexa 647 (red) secondary antibody respectively. Nuclei were stained with DAPI (blue). Bar = 1  $\mu$ m. Columns: (1), Alexa 488; (2), Alexa 647; (3), DAPI; (4), merged



**Sup. Figure 2: LEC1 and LEC2 interaction in yeast.**

Yeast two hybrid interaction test with the truncated LEC2, wild type sequence or mutated, and LEC1 wild type or with D86K mutation on 1: SD-WL, 2: SD-WLH, 3: SD-WLHA, 4: SD-WLH + 3AT 10mM, 5: SD-WLH +3AT 15mM, and 6: SD-WLH +3AT 25mM. Five different clones are tested for each interaction.

SD: synthetic dextrose-minimal medium (Clontech, Mountain view, USA) with dropout supplements, W: Tryptophan, L: Leucine, A: Adenine, 3AT: 3-aminotriazole.

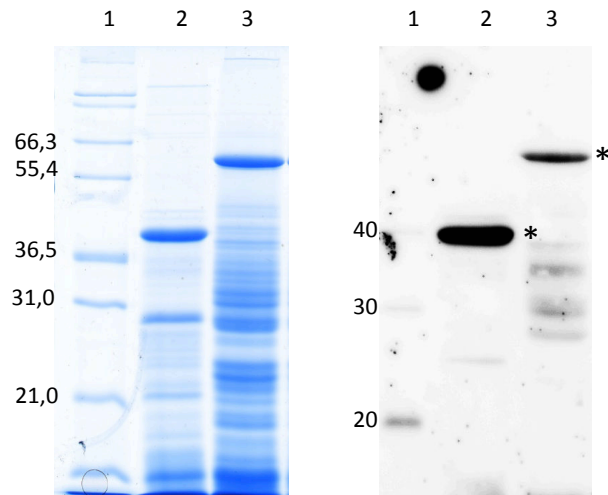


### Sup. Figure 3: Structure of proteins

**A-** Structure of the NF-YA/B/C complex (pdb 4awl). **B-** Structure of LEC1 Like (pdb 5g49). In colour are highlighted the amino acid residues important for *lec1* complementation (from (Lee *et al.* 2003)). **C-** Alignment of the A1A2 linker with the B2-domain. Identical residues are shaded in green. The conserved sequence in the B2 domain is highlighted in red.

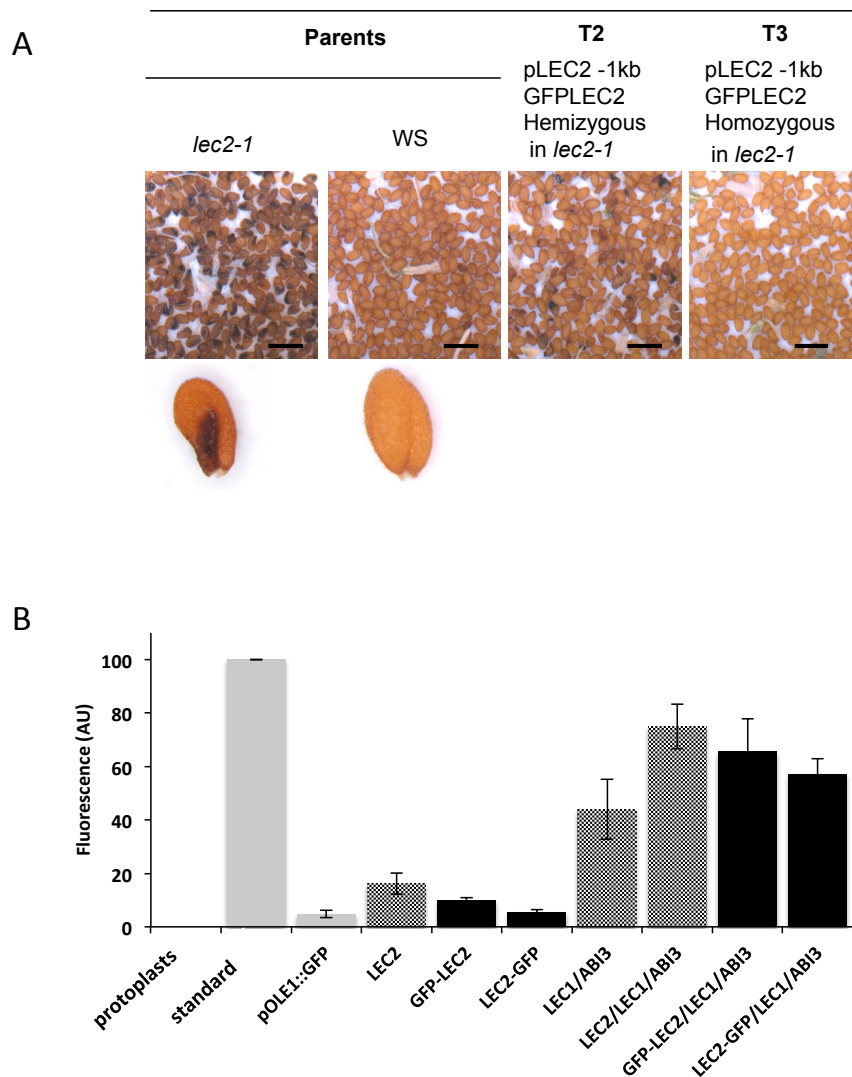
Hs: *Homo sapiens*, An: *Aspergillus nidulans*, Sc: *Saccharomyces cerevisiae*, At: *Arabidopsis thaliana*





**Sup figure 4: Production on His-tagged recombinant LEC1 and LEC2 proteins:**

Representative coomassie blue staining (left panel) and western blot (right panel) with an antibody raised against the HIS tag (right) of LEC1 and LEC2 after purification (1) Magic Mark standard, (2) LEC1 protein extract, (3) LEC2 protein extract. The asterix indicate the bands with the expected size.



**Sup figure 5: The chimeric LEC2:GFP is active *in planta* and in moss protoplast**

**A-** Phenotypic reversion of the *lec2-1* phenotype by transformation with *pLEC2-1kb::GFP:LEC2* chimeric gene . Bar = 1 mm. **B-** Activity of LEC2 tagged in N or C terminus with GFP in moss protoplasts in the presence / absence of LEC1 and ABI3.

**Sup Table 1: Oligonucleotides used in this study.**

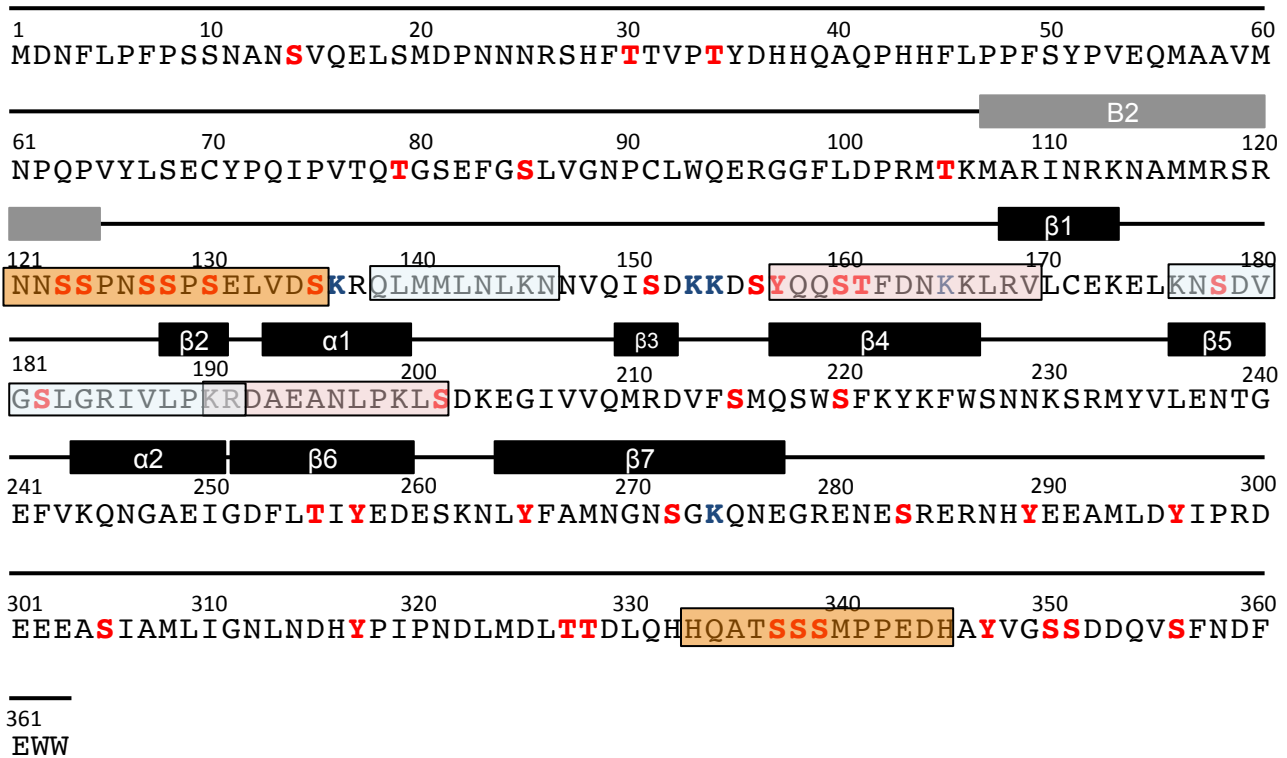
Oligo name	5'-3' sequence
Mutated LEC2 for	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGGATAAATTCTCCCTT
Mutated LEC2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACCACCACTCGAAATCG
pGWB2 PLEC2 -1KB_for	CCCAAGCTTGGGTGTTCTGATTC
pGWB2 PLEC2 -1KB_rev	GCTCTAGAGCTTTTCCCGGAGAG
pGWB2 PLEC2 -1KB_check_for	CCCAAGCTTGGGTGTTCTGATTC
pGWB2 PLEC2 -1KB_check_rev	CGCTCGAGCAAAAAAAGAGAGGTCG
2H LEC2 mutation for	GGTTTTCTTGATCCGCGTATGGGGAAGATGACAAACATCAACAGGAAAAACGCC
2H LEC2 mutation rev	GGCGTTTTTCCTGTTGATGTTTGTTCATCTTCCCATACGCGGATCAAGAAAACC
Short LEC2 for 2H for	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGGATAAATTCTTACCCTTTCC
Short LEC2 for 2H rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAAGCTTCTTGTTATCAAATG

A

TK-AR--R--A--R

At	LEC2	GGFLDPRM <b>TKMAR</b> IN <b>RKNAM</b> MRSRN
At	ABI3	LLRLCSSA <b>TKEARKK</b> R <b>M-ARQ</b> RRFLS
At	FUS3	LSASVPLLGVN <b>WKKR</b> M- <b>PRQ</b> RRSSS

B



**Figure 8 : Séquence et structure de la protéine LEC2.**

(A) Alignement des domaines B2 des AFL: (en rouge) les résidus conservés entre LEC2 et ABI3, (en bleu) tous les résidus conservés aux 3 AFL-B3, (en gras souligné) la séquence NLS d'ABI3. At: *Arabidopsis thaliana*. (B) Séquence en acides aminés de LEC2, structure secondaire de LEC2, le domaine B2 (en gris), le domaine B3 (en noir), site putatif de NLS (en rouge clair) <http://mleg.cse.sc.edu/seqNLS/>, sites putatifs de NES (en bleu clair) <http://prodata.swmed.edu/LocNES/LocNES.php>. Sites prédits ubiquitinylés (en bleu), simple avec un score moyen, en gras les sites fortement prédits, (<http://www.ubpred.org>). Sites prédits phosphorylés (en rouge), <http://www.cbs.dtu.dk/services/NetPhos/>. Domaine putatif PEST prédit par PESTFIND algorithm, en orange (<http://emboss.bioinformatics.nl/cgi-bin/emboss/epestfind>).



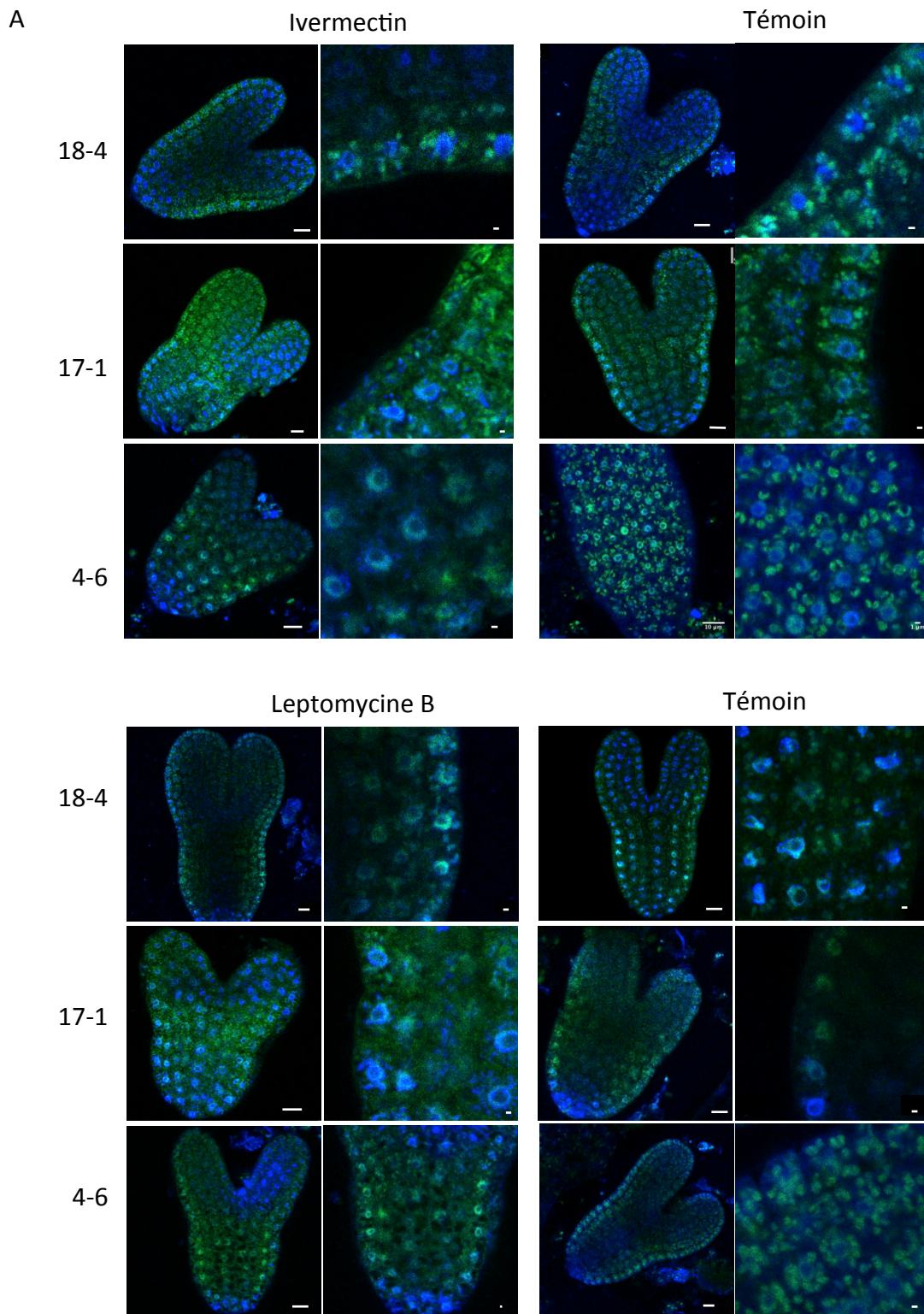
### **II.3.3 Résultats complémentaires**

Les facteurs de transcription LEC1 et LEC2 interagissent, sur un même promoteur, avec LEC1 qui augmente l'affinité (Kd) de LEC2 pour sa cible. Les premières observations ont permis de localiser, au cours du développement de la graine, LEC1 et LEC2 partiellement au noyau avec une localisation péri-nucléolaire (le reste de l'accumulation étant cytoplasmique). L'accumulation de la protéine et les données d'expression de l'ARN messager semblent correspondre aux mêmes stades de développement, du stade globulaire au stade torpille, (Chapitre I.2.2, figure 1 p. 20, Chapitre II.3) Les deux protéines interagissent par l'intermédiaire du domaine B2 de LEC2 et par D86 (au moins) chez LEC1. Chez ABI3, le domaine B2 est nécessaire pour la fonction activatrice, l'interaction des protéines et la localisation de la protéine au noyau. ABI3 contient dans son domaine B2 une séquence de localisation nucléaire (NLS) ARKKR (Marella and Quatrano 2007; Zhang *et al.* 2005). LEC2 a un domaine B2 différent de celui d'ABI3 et ne contient plus cette séquence NLS (Figure 8). La mobilité intracellulaire de LEC2 pourrait donc être différente de celle d'ABI3. La localisation subcellulaire de LEC1 et LEC2, montre un signal nucléaire mais aussi avec un bruit de fond cytoplasmique, qui laisserait penser à une mobilité des protéines qui pourrait être impliquée dans la régulation de l'activité des facteurs de transcription par les voies de l'import et l'export nucléaire. Des données préliminaires semblent montrer que LEC1 doit former un dimère avec un NF-YC pour être relocalisé au noyau (Hackenberg *et al.* 2012), mais aucune étude n'a été réalisée pour LEC2.

#### **II.3.3.1 L'import et l'export nucléaire**

Les procédés d'import et d'export nucléaire permettent le transfert rapide de protéines et sont très conservés. Ces systèmes sont notamment retrouvés chez la levure, l'Homme et *Arabidopsis thaliana* (Merkle 2001). Néanmoins, différentes protéines peuvent intervenir (Merkle 2001; Merkle 2011).

L'import nucléaire est caractérisé par la présence d'une séquence de localisation nucléaire (NLS). Les séquences NLS classiques sont constituées d'acides aminés basiques, composés principalement de lysine (K) et structurées en 1 ou en 2 séquences séparées par une courte séquence d'acides aminés acides. Elles sont affines pour les séquences Armadillo (ARM) de l'IMPORTINE  $\alpha$ , qui sont de courtes séquences répétées qui confèrent une structure en hélice et permettent l'interaction de l'IMPORTINE  $\alpha$  avec le NLS. (Merkle 2011). Les IMPORTINES  $\alpha$  interagissent avec les IMPORTINES  $\beta$  par la séquence « Importin  $\beta$ -Binding » (IBB) située en N terminal de la protéine  $\alpha$ . C'est par l'intermédiaire de l'IMPORTINE  $\beta$  que le complexe protéine NLS - IMPORTINE  $\alpha/\beta$  passe dans le noyau. Des séquences NLS moins caractéristiques se lient cependant directement à des membres de la famille des IMPORTINES  $\beta$ . Le passage au noyau des protéines s'effectue via les pores



**Figure 9: Localisation intracellulaire de GFP:LEC2**

Observation en microscopie confocale des 3 lignées *proLEC2(-1Kb):GFP:LEC2* (18-, 17-4 et 4-6), ; (A) après 3h d'incubation en présence d'ivermectin (300  $\mu$ M) ou de Glycérol/H<sub>2</sub>O 1:1, DMSO 0,01% pour témoin. (B) après 3h d'incubation en présence de leptomycine B (300 nM) ou de Glycérol/H<sub>2</sub>O 1:1 pour témoin.

nucléaires et le complexe de protéines NUCLEOPORINES qui interagissent avec les IMPORTINES  $\beta$ . L'affinité des NLS des protéines, l'import et l'export des protéines de transport sont médiés par les protéines G Ran et des échanges de GTP/GDP (Merkle 2001; Merkle 2011).

L'export nucléaire a un fonctionnement identique à l'import, une séquence d'export nucléaire (NES) est présente sur la séquence de la protéine cible. Cette séquence est riche en leucine et composée d'acides aminés hydrophobes. La NES est reconnue par des protéines de transport, les EXPORTINES qui interagissent avec les Ran GTP et par échange d'énergie vont exclure la protéine du noyau (Merkle 2011).

La régulation de l'export et de l'import peut être réalisée par l'intermédiaire de (i) modifications post-traductionnelles, notamment la phosphorylation des protéines, (ii) leurs conformations, (iii) les interactions avec des protéines tierces, (iv) la rétention des protéines dans un compartiment (Merkle 2001).

### ***II.3.3.2 Données préliminaires sur le trafic de LEC2***

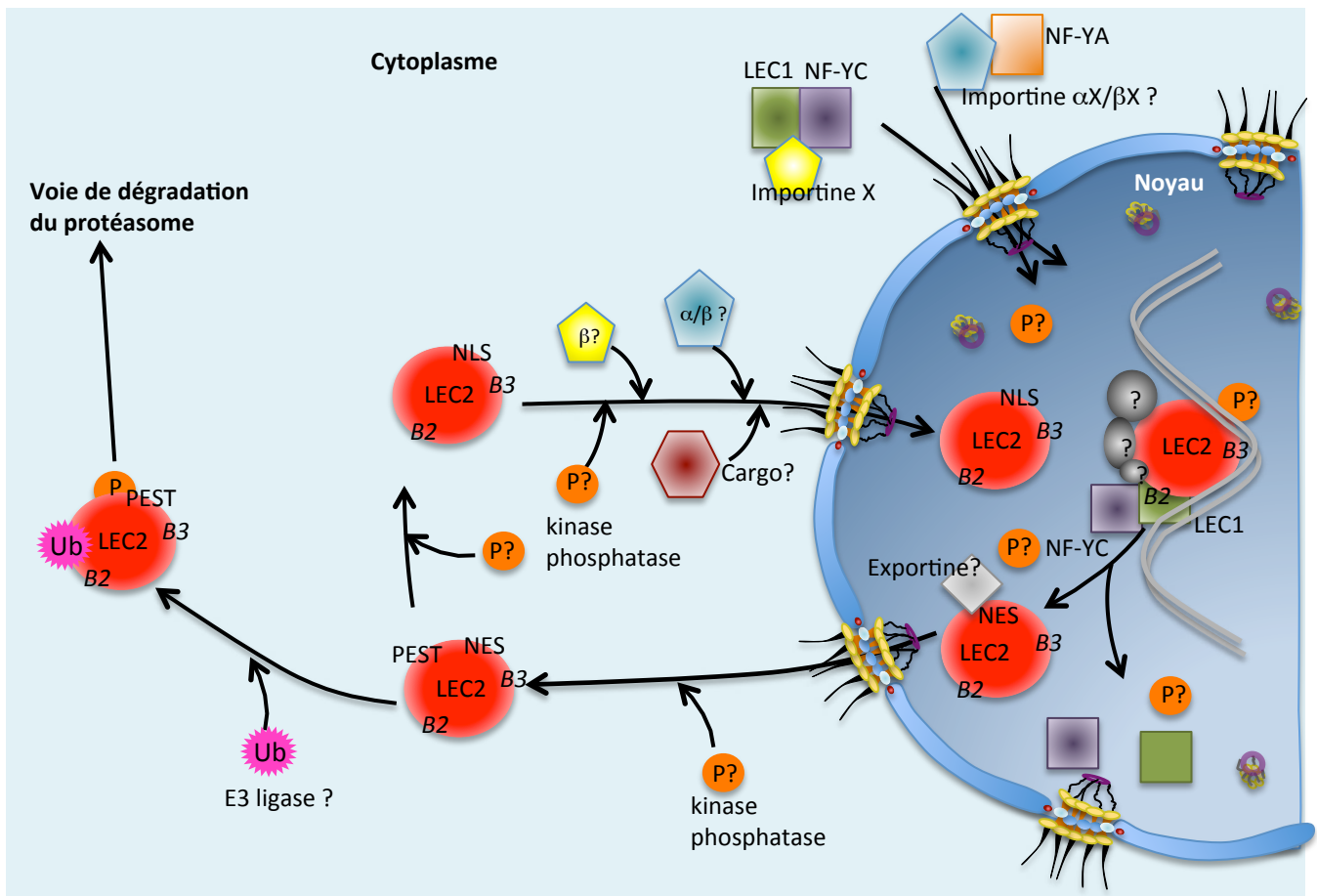
La recherche de présence de séquences spécifiques du transport nucléaire sur la séquence protéique de LEC2 montre une prédiction de 2 séquences NLS et de 2 séquences NES toutes situées dans le domaine B3 (figure 8).

Afin d'étudier le transport nucléaire de LEC2, les lignées GFP:LEC2 ont été observées en microscopie confocale après 3H d'incubation avec les drogues ivermectin et leptomycine B bloquant ainsi l'import et l'export nucléaire respectivement. Trois lignées indépendantes ont été observées, en présence ou non des drogues. Les résultats préliminaires ne montrent pas de rétention de LEC2 dans un des compartiments cellulaires pour l'import et pour l'export nucléaire, comparés au témoin (Figure 9).

### ***II.3.3.3 Discussion***

L'ivermectin n'affecte pas toutes les voies d'import, la drogue bloque spécifiquement l'import de protéines cargo IMPORTINE  $\alpha/\beta$  mais ne bloque pas le transport des l'IMPORTINES  $\beta$ . La constitution de la NLS de LEC2 ne permet pas de confirmer la voie d'import utilisée par LEC2 (Figure 10). Le trafic de LEC2 pourrait être réalisé par la voie de l'import  $\beta$  uniquement. La Leptomycine B agit sur les EXPORTINES CRM1 de l'homme ou XPO1 d'*Arabidopsis thaliana*. Il existe 6 EXPORTINES identifiées chez arabidopsis, dont 3 semblent s'apparenter à XPO1 (Merkle 2011). Toutes les EXPORTINES pourraient ne pas être bloquées par la leptomycine B. De même que pour l'import, l'export de LEC2 pourrait être effectué via des protéines non affectées par les drogues utilisées. La présence de la GFP dans le cytoplasme peut aussi être due à la dégradation de la protéine et au clivage de la GFP. Afin de comprendre le trafic de LEC2, il serait intéressant de tester différentes mutations des NES et NLS putatifs, permettant d'accumuler LEC2 dans les différents





**Figure 10: Résumé des hypothèses relatives aux transports et à la régulation de LEC2 dans la cellule.**

Les hypothèses proposées portent sur la mobilité cellulaire de la protéine LEC2, l'identification des sites de phosphorylation l'action des kinases et phosphatases pour la localisation au cytoplasme et au noyau, l'interaction avec l'une des protéines d'import et d'export. Toutes sont encore inconnues. La voie de dégradation proposée pour la régulation de l'activité de la protéine de LEC2 dans la cellule est basée sur la présence de sites de modifications post-traductionnelles (domaines PEST, sites prédits d'ubiquitination ou de phosphorylation). Le transport des NF-YA et du dimère NF-YB/C est basé sur le travail chez l'Homme d'après Kahle *et al* (2005), les importines  $\alpha/\beta$  et  $\beta$  ne sont pas identifiées chez les végétaux.

compartiments et d'identifier les protéines liées à LEC2 et responsables du transport. Enfin, le trafic des protéines est contrôlé par les modifications post-traductionnelles et notamment les phosphorylations. LEC2 possède de nombreux sites putatifs de phosphorylation qui, associés à des données de cristallographie, pourraient renseigner sur les différentes voies de régulations de l'activité et du trafic de LEC2 (Figure 8). Une étude menée sur la stabilité de FUS3 a permis d'identifier la présence d'un domaine PEST en C terminal de la protéine (Lu *et al.* 2010). Le domaine PEST riche en proline (P), acide glutamique (E), sérine (S), thréonine (T) est impliqué dans la régulation de l'activité de FUS3 par dégradation en lien avec le protéasome et, par conséquent, par la diminution du temps de demi-vie de la protéine (Rechsteiner and Rogers 1996; Lu *et al.* 2010; Xing *et al.* 2010; Park *et al.* 2014). Des domaines PEST sont prédits dans la séquence de LEC2 de part et d'autre du domaine B3 (Figure 8). L'activité de LEC2 et son temps de demi-vie pourraient également être régulés par ces domaines. Enfin, la prédiction de séquence NLS et NES n'exclut pas que LEC2 transite au noyau uniquement par l'intermédiaire d'une interaction avec une ou plusieurs autres protéines pour le moment inconnues.

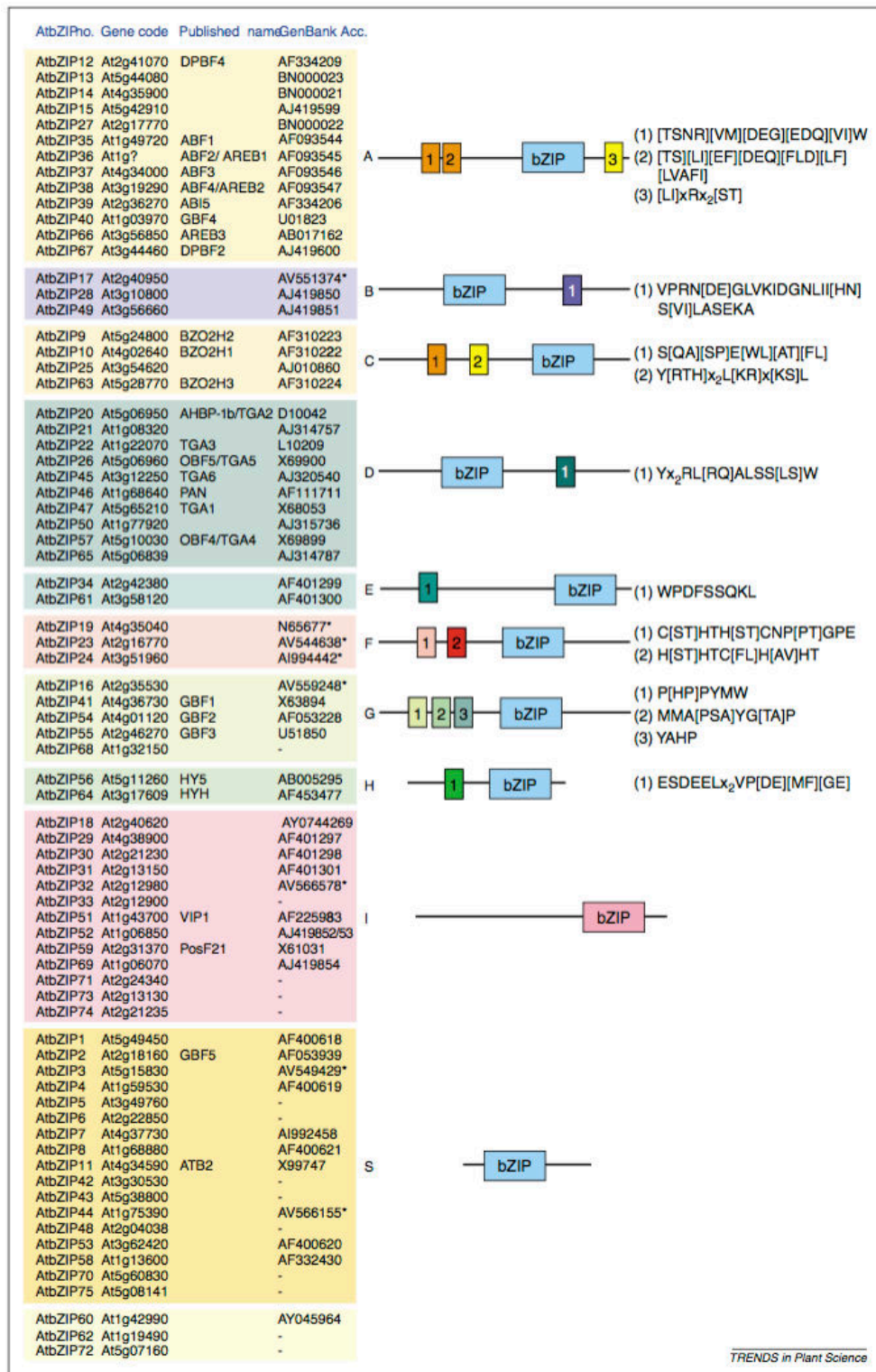


Figure 11 : Classement des 75 bZIPs identifiés chez *Arabidopsis thaliana* en fonction de leur similarité de séquence et de l'implication dans les processus biologiques (d'après Jakoby *et al* 2002).

## II.4 Des bZIPs font-ils partie du complexe LAFL ?

Nous avons montré que LEC1, LEC2, ABI3 forment un complexe sur le promoteur de l'*OLEOSINE1* et présentent une activité synergique sur ce promoteur (Baud *et al.* 2016). NF-YC2 semble intervenir également dans ce complexe. Les LAFL requièrent de nombreux partenaires lors de la mise en place des réserves protéiques, formant probablement ainsi des complexes sur les promoteurs des gènes des protéines de réserve (Baud *et al.* 2016; Kroj *et al.* 2003; Yamamoto *et al.* 2009), ou liés à d'autres fonctions. On peut en particulier citer les facteurs de transcription de type bZIPs, qui ont été décrits comme interagissant avec certains LAFL (Yamamoto *et al.* 2009; Mendes *et al.* 2013; Liu and Howell 2010; Nambara and Marion-Poll 2003; Lara *et al.* 2003; Alonso *et al.* 2009) et qui pourraient faire partie de tous, ou de certains, complexes protéiques liés aux LAFL.

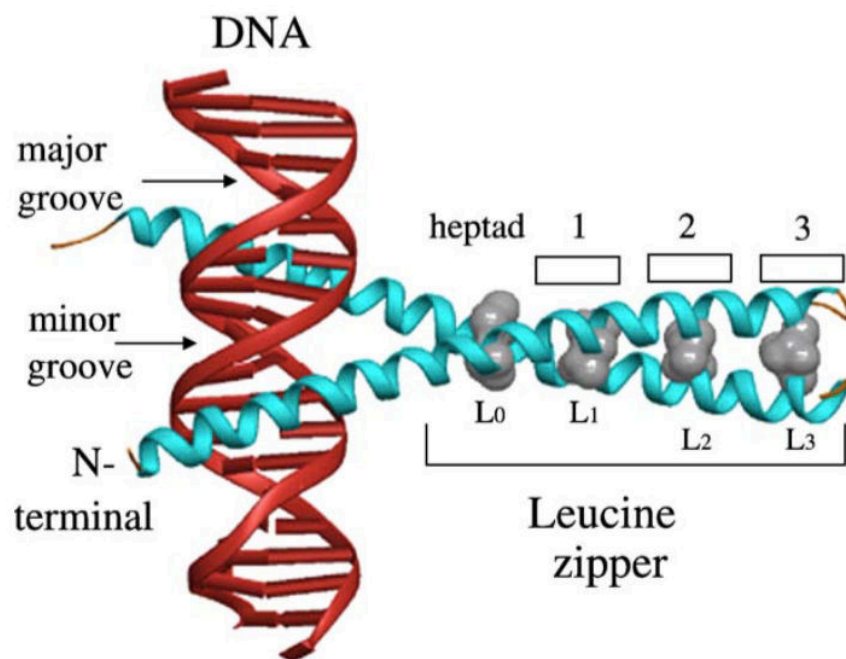
### II.4.1 La famille des facteurs de transcriptions bZIPs

Les bZIPs forment une famille de facteurs de transcription représentée chez tous les eucaryotes. Des analyses comparatives des génomes eucaryotes ont permis d'identifier 53, 21, 25 et 77 bZIPs chez *Homo sapiens*, *Drosophila*, *C. elegans* et *Arabidopsis thaliana* respectivement (Deppmann *et al.* 2004; Jakoby *et al.* 2002; Riechmann *et al.* 2000; Correa *et al.* 2008).

Des analyses du génome d'*Arabidopsis thaliana* ont permis d'identifier 77 bZIPs par une étude croisée de la similarité de séquences et de la phylogénie des protéines étudiées (Jakoby *et al.* 2002; Correa *et al.* 2008). Ces facteurs de transcription sont classés en 10 groupes A à I et un dernier groupe, S, selon les similarités de séquences de leur région basique et du leucine zipper, la présence de domaines conservés et leurs implications dans les procédés biologiques tels que : le développement et les stress (Jakoby *et al.* 2002) (Figure 11). Néanmoins, chez *Arabidopsis thaliana*, l'analyse du domaine basique, avec comme critère la présence systématique d'une asparagine dans ce domaine, n'a permis l'identification que de 67 bZIPs, les autres seraient des bZIPs rares (Deppmann *et al.* 2004).

Les bZIPs les plus étudiées chez les monocotylédones sont les protéines de la famille OPAQUE, notamment OPAQUE2 qui est impliquée dans la synthèse des protéines de réserve dans l'albumen (Hunter 2002). Chez les dicotylédones ABI5 est une protéine bZIP du groupe A qui interagit avec ABI3 et est impliquée dans la maturation tardive de la graine permettant l'acquisition de la tolérance à la dessiccation en lien avec l'acide abscissique (ABA) (Ooms *et al.* 1993; Righetti *et al.* 2015). ABI5 est impliquée avec ABI3 dans la synthèse des protéines LEA (Late Embryogenesis Abundant) au stade de la maturation tardive (Finkelstein and Lynch 2000; Lopez-Molina *et al.* 2002).

## B-ZIP motif



**Figure 12 : Structure d'un dimère de bZIPs avec une interaction avec l'ADN.**  
L représente une leucine du domaine leucine zipper (Vinson *et al* 2006).

## **II.4.2 Structure**

Les bZIPs agissent en homo ou hétérodimère au sein d'un même groupe ou d'un autre groupe. Ils se fixent sur les éléments cis de l'ADN, préférentiellement les boîtes « A » (TACGTA), « C » (GACGTC), « G » (CACGTG), avec pour séquence consensus CACGTGTC (Peng and Weselake 2011; Jakoby *et al.* 2002).

Les bZIPs sont des protéines structurées en hélice alpha et constituées en 2 parties : le domaine N terminal, qui est la région basique de la protéine, contient notamment la séquence NLS et le domaine de fixation à l'ADN ; la partie C terminale contient le leucine zipper impliqué dans la dimérisation, par la formation d'un coiled-coil entre deux hélices alpha. Le domaine leucine zipper est constitué d'heptades d'acides aminés où deux leucines vont être séparées par 6 acides aminés (Figure 12) (Jakoby *et al.* 2002; Vinson *et al.* 2006).

La composition en acides aminés des heptades du leucine zipper détermine l'homo ou l'hétérodimérisation des bZIPs. La structure du leucine zipper implique quatre acides aminés de l'heptade sur l'hélice alpha qui interagissent avec les acides aminés de même positionnement mais de charges opposées du second bZIP lors de la dimérisation (Vinson *et al.* 2006; Deppmann *et al.* 2004).

## **II.4.3 Régulation des bZIPs**

Les bZIPs sont régulées par de nombreuses voies de régulation et dépendants des dimères formés (pour revue : Schutze *et al.* 2008). La phosphorylation est une forme de régulation en particulier présente chez les bZIPs impliquées dans la signalisation par l'ABA (Lynch *et al.* 2012; Nakamura *et al.* 2001). La phosphorylation peut affecter l'interaction des protéines entre elles ou avec les macromolécules comme l'ADN. Par exemple, les sérines positionnées dans le domaine basique du bZIP63 réduisent l'interaction de la protéine avec l'ADN lorsqu'elles sont phosphorylées (Kirchler *et al.* 2010). L'action des dimères peut donc être régulée par des interactions entre un dimère de bZIPs et une autre protéine, ou encore par une régulation post-traductionnelle bloquant ainsi l'action du facteur de transcription (Schutze *et al.* 2008; Skubacz *et al.* 2016).

## **II.4.4 Les bZIPs identifiés ont une interaction avec certains AFL ou NF-Y**

L'ensemble des résultats est présenté sur le tableau 1.

### **II.4.4.1 Interaction des bZIPs avec les NF-Y**

bZIP28 fait partie du groupe B et est un capteur du réticulum endoplasmique (RE) et transducteur du signal en réponse au stress (Srivastava *et al.* 2014). L'accumulation de protéines déstructurées dans le réticulum endoplasmique entraîne un signal cellulaire de réponse au stress. bZIP28, ancrée dans la membrane du RE, est ainsi remobilisée à l'appareil de Golgi grâce à l'interaction avec la protéine Sar12 impliquée dans le transport

<b>bZIPs</b>	<b>Numéro AGI</b>	<b>Groupe</b>	<b>Interaction(s)</b>	<b>Élément cis de l'ADN</b>	<b>Dimère(s)</b>
AtbZIP67	At3G44460	A	NF-YC2 LEC1 L1L	G-box	?
AtbZIP28	At3G10800	B	NF-Y family (NF-YC2)	G-Box	Homodimer, Heterodimer: bZIP17, bZIP49, bZIP60 (B group)
AtbZIP10	At4G02640	C	ABI3	G-box	bZIP25, bZIP53
AtbZIP25	At3G54620	C	ABI3	G-box	bZIP10, bZIP53
AtbZIP53	At3G62420	S	ABI3	C and G box	bZIP10, bZIP53

**Tableau 1 : Tableau récapitulatif des données connues des bZIPS ayant une interaction avec un ou plusieurs LANL.**

(Liu and Howell 2010 ; Yamamoto et al. 2009; Mendes et al. 2013; Ehlert et al. 2006; Veerabagu et al. 2014; Llorca et al. 2015; Lara et al. 2003; Alonso et al. 2009).

des vésicules de COPII. Au niveau de l'appareil de Golgi, bZIP28 va subir une protéolyse par des enzymes protéolytiques (une sérine protéase, S1P et une métalloprotéase, S2P) libérant ainsi bZIP28 de son domaine transmembranaire et permettant à la protéine d'être localisée au noyau afin d'activer les gènes de réponse au stress (Liu *et al.* 2007; Srivastava *et al.* 2012). Il a été montré que bZIP28 peut former un complexe transcriptionnel avec des membres de la famille de NF-Y : NF-YA4 / NF-YB3 / NF-YC2 sur les promoteurs des gènes de réponse au stress contenant l'élément cis ERSE-I (RE stress réponse) (Liu and Howell 2010).

bZIP67 est une protéine du groupe A qui ont notamment une action en réponse à un stress biotique (Jakoby *et al.* 2002). En expression hétérologue, LEC1-LIKE (NF-YB6), ou LEC1 (NF-YB9) et NF-YC2 interagissent avec bZIP67 formant ainsi un trimère sur le promoteur cible. LEC1-like (L1L) interagit en particulier avec NF-YC2 et bZIP67 pour réguler l'expression des enzymes nécessaires à la synthèse des acides gras (Yamamoto *et al.* 2009). L'interaction entre LEC1, ou LEC1-like, avec bZIP67 et NF-YC2, semblerait aussi intervenir lors de la synthèse des protéines de réserve (Yamamoto *et al.* 2009; Mendes *et al.* 2013).

#### **II.4.4.2 Interaction des bZIPs avec un B3**

Les bZIPs du groupe C ont la particularité d'avoir un leucine zipper contenant plus de neuf heptades et sont impliquées dans l'expression des protéines de réserve chez *Arabidopsis thaliana*. Les bZIPs 10 et 25 sont des protéines du groupe C et ont une similarité de séquence et de composition de domaines avec OPAQUE2 (Jakoby *et al.* 2002; Lara *et al.* 2003). La famille des bZIPs OPAQUE est impliquée dans l'expression des protéines de réserve de l'albumen chez le maïs (Hunter 2002).

bZIP53 est une protéine bZIP du groupe S, qui regroupe des bZIPs ne contenant pas d'autres domaines connus et ayant une petite masse moléculaire (Jakoby *et al.* 2002). bZIP53 hétérodimérise avec les protéines d'autres groupes et notamment les bZIP10 et 25 du groupe C (Ehlert *et al.* 2006; Veerabagu *et al.* 2014; Llorca *et al.* 2015).

ABI3 est fortement impliquée dans la synthèse des protéines de réserve, les cruciférines, les oléosines et les LEA (Monke *et al.* 2012). Les gènes des protéines de réserve comme l'albumine 2S sont la cible du bZIP53 (Alonso *et al.* 2009). *ABI3*, *bZIP10* et 25 ont le même profil d'expression au cours du développement de la graine et l'expression des *bZIPs 10* et 25 en présence d'ABI3 augmente l'accumulation de l'albumine 2S (Lara *et al.* 2003). ABI3 formerait un complexe avec les bZIPs 10, 25, 53. Ce complexe agirait de façon synergique sur les promoteurs des gènes de protéines de réserve (Lara *et al.* 2003; Alonso *et al.* 2009). La cinétique de formation et la structure de ce complexe sont actuellement inconnues.



BD \ AD	bZIP10	bZIP25	bZIP28	bZIP53	bZIP67	pDEST22
<b>bZIP10</b>	+	+	+	+	-	-
<b>bZIP25</b>	-	-	-	+	-	-
<b>bZIP28</b>	Auto activation					
<b>bZIP53</b>	-	+	+	-	-	-
<b>bZIP67</b>	-	-	-	-	-	-
<b>pDEST32</b>	-	-	-	-	-	-

**Tableau 2 : Tableau récapitulatif des résultats de la formation de dimères de bZIPs en double hybride.**

(AD) Domaine d'Activation, (BD) Binding Domain de fixation, en gris les homodimères testés, en blanc les hétérodimères, (-) absence d'interaction, (+) présence d'interaction. pDEST 22 (AD) et pDEST 32 (BD) sont les vecteurs vides utilisés.

Les facteurs de transcription bZIPs 10/25/53/67 ont été décrits comme ayant un rôle dans l'expression des protéines de réserve, en interaction avec NF-YC2, LEC1 ou ABI3 (Lara *et al.* 2003; Alonso *et al.* 2009; Mendes *et al.* 2013; Yamamoto *et al.* 2009). *bZIP28* est exprimé au cours du développement de la graine, sans qu'il ait été démontré que la protéine intervienne dans l'expression des gènes de réserve. elle contient un domaine transmembranaire, est localisée au réticulum endoplasmique (RE) mais interagit avec NF-YC2 en réponse au stress du RE.

Les bZIPs étant capables d'homo ou d'hétérodimériser avec un membre du même groupe ou de groupes différents, j'ai réalisé des études systématiques de ces facteurs par double hybride et par des tests fonctionnels en expression transitoire chez la mousse afin d'identifier (i) les dimères formés entre les bZIPs 10/25/53/28/67, (ii) leurs actions sur le promoteur de l'*OLEOSINE1* et (iii) le rôle des bZIPs sur le complexe LEC1, ABI3, NF-YC2, LEC2 (LANL) intervenant dans l'expression de l'*OLEOSINE1*.

## **II.4.5 Résultats**

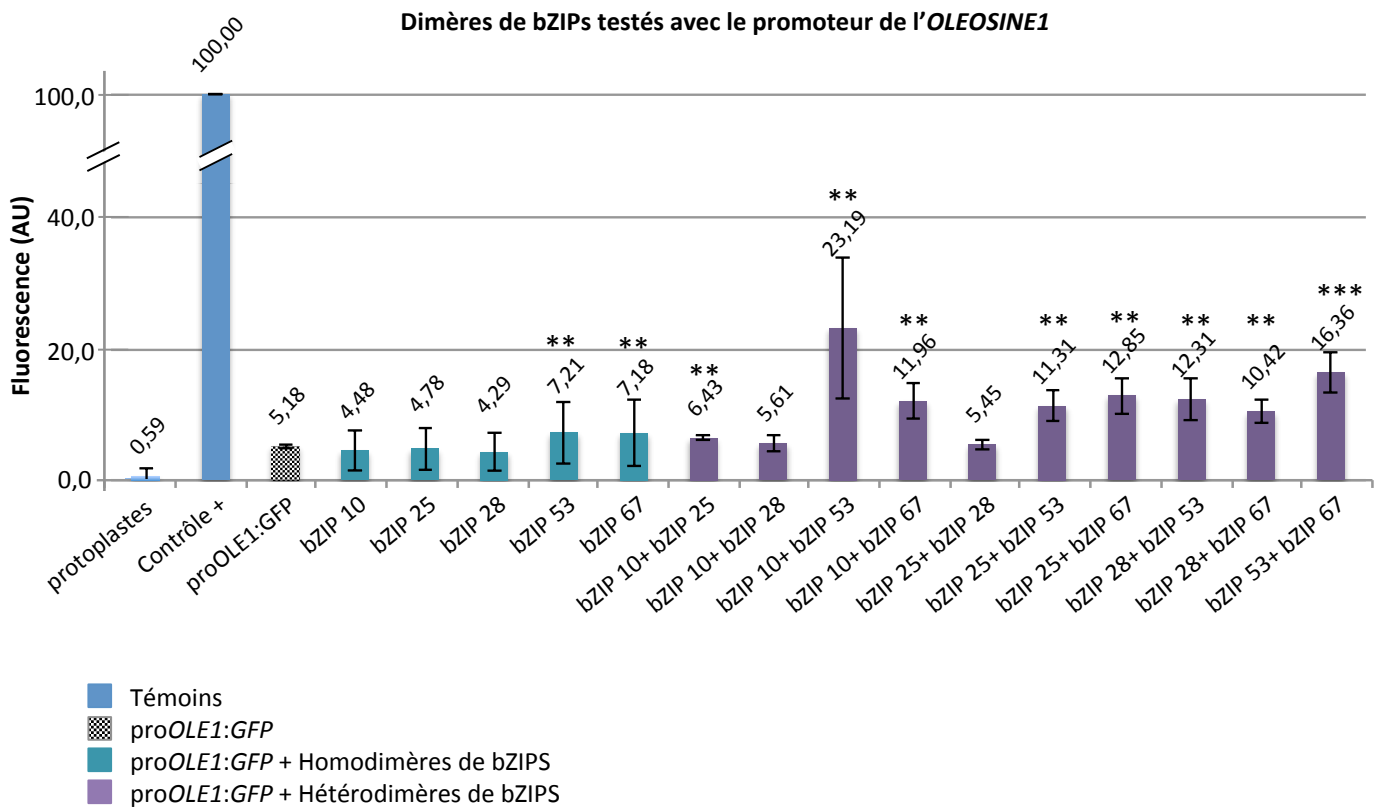
### **II.4.5.1 Dimérisation des bZIPs**

La formation de dimères de bZIPs entre certains groupes, a été testée dans la littérature par de nombreuses techniques comme le double hybride, ou la reconstitution bi-moléculaire de fluorescence (bIFC). Il a notamment été démontré que les bZIPs 10 et 25 du groupe C et le bZIP53 du groupe S sont capables d'interagir et de former des hétérodimères (Ehlert *et al.* 2006; Alonso *et al.* 2009). En outre, bZIP28 peut former des homo et hétérodimères avec les bZIPs du même groupe (Liu and Howell 2010).

Afin de tester la dimérisation au sein du groupe des protéines bZIPs sélectionnées, nous avons opté pour une analyse double hybride résumée dans le tableau 2. bZIP28 auto-active la transcription lorsqu'elle est associée au domaine de fixation à l'ADN (BD), les résultats ont donc été interprétés uniquement associé avec bZIP28 fusionnée au domaine d'activation (AD). Les autres bZIPs ont pu être étudiées avec les 2 domaines (AD et BD), en présence de 3-aminotriazole (3-AT) pour bloquer l'auto-activation.

Sur tous les bZIPs testées, seul bZIP10 forme un homodimère. Elle hétérodimérise avec bZIP25 et 53, de même que, réciproquement, bZIP 53 hétérodimérise avec bZIP10 et 25 et, enfin, bZIP 25 interagit avec bZIP10 et 53. Les deux bZIPs du groupe C interagissent avec bZIP53, appartenant au groupe S, ces résultats sont confirmés par les données publiées (Ehlert *et al.* 2006; Alonso *et al.* 2009).

La formation d'homodimère bZIP28 n'a pas pu être validée car bZIP28-BD auto-active, mais bZIP28-AD forme un dimère avec bZIP10 et 53. bZIP67 ne dimérise avec aucun des bZIPs testées.



**Figure 13: Activité des bZIPs avec le promoteur de l'OLEOSINE 1 par la méthode d'expression transitoire chez la mouche *Physcomitrella patens*.**

Les bZIPs sont testés en homodimères (bleu) ou hétérodimères (violet). Valeur des résultats du test de Student *t*-test, N=4 réplicats technique \* : < 0,05, \*\* : <0,01, \*\*\* :<0,001. (NS) : non significatif. Les résultats des bZIPs sont comparées à la valeur de proOLE1:GFP. Les témoins sont les protoplastes seuls et un témoin de transformation.

Les résultats de l'analyse systématique des bZIPs en double hybride sont validés par l'obtention des dimères déjà identifiés entre les bZIPs des groupes C et S. Seuls deux nouveaux hétérodimères ont été identifiés par cette méthode, bZIP28/10 et bZIP28/53.

De plus bZIP67 ne dimérise ni avec lui-même ni avec aucun des bZIPs testées. La littérature a démontré son action sur les promoteurs FAD et sur les promoteurs de protéines de réserve en présence des NF-Y mais sans tester la dimérisation (Mendes *et al.* 2013; Yamamoto *et al.* 2009). Le(s) partenaire(s) de bZIP67 reste(nt) donc encore inconnu(s).

#### **II.4.5.2 Action des bZIPs en dimère, seuls, sur le promoteur de l'OLEOSINE1**

Les bZIPs des groupes C et S (bZIP10/25/53) et le bZIP67 ont une activité transcriptionnelle connue sur des promoteurs des gènes de protéines de réserve (Lara *et al.* 2003; Alonso *et al.* 2009; Yamamoto *et al.* 2009). Le promoteur proOLE1 contient 3 boîtes « G » ayant pour séquence une première boîte imparfaite AACGTG et pour les 2 boîtes suivantes la séquence canonique CACGTG, ainsi que 2 boîtes RY positionnées entre les 3 boîtes « G ». Nous avons montré que ces boîtes sont reconnues par des bZIPs, en expression transitoire chez la mousse *Physcomitrella patens* (Figure 13).

Les bZIP10/25/28 en homodimères et les hétérodimères bZIP10/28 et 25/28 n'ont pas d'activité transcriptionnelle sur le promoteur proOLE1. Les homodimères bZIP53 et bZIP67 ainsi que les 8 autres hétérodimères ont une activité significative sur le promoteur. Le dimère bZIP10/25 a une activité plus faible que les hétérodimères testés, associées au bZIP53 ou bZIP67, les hétérodimères ont une plus grande activité transcriptionnelle sur proOLE1. La plupart d'entre eux ont une activité à effet additif qui est égale à la somme des activités des 2 homodimères de bZIPs.

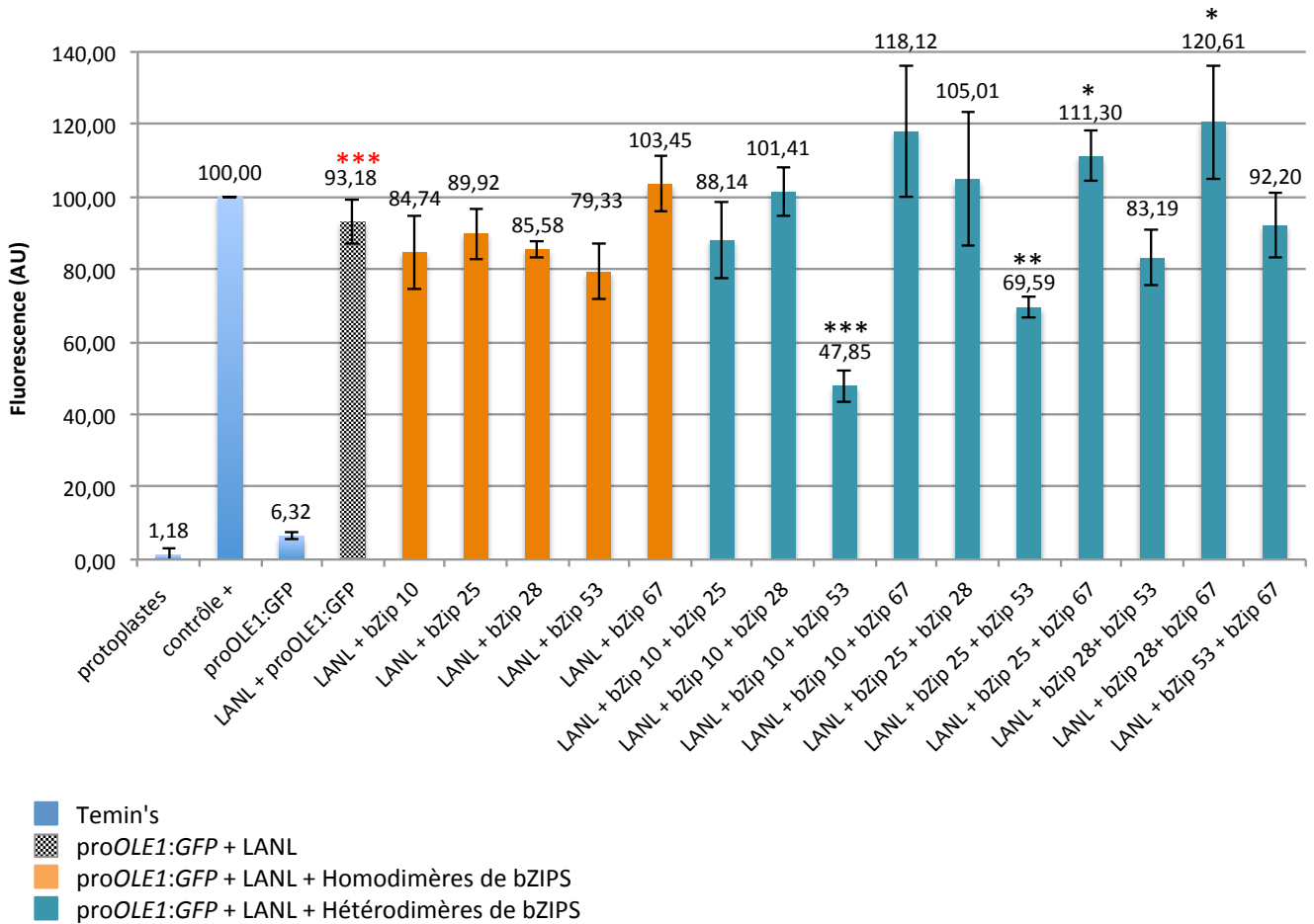
Le dimère bZIP10/53 semble avoir un effet synergique avec une activité supérieure à la somme des activités des homodimères bZIP10 et bZIP53 sur le promoteur. Cette activité a aussi été démontrée sur le promoteur de la protéine de réserve 2S2 (Alonso *et al.* 2009).

Les bZIPs 10, 25, 28 en homodimère n'ont pas d'effet. Mais associées en hétérodimères avec les bZIP53 ou 67, les protéines améliorent l'activité transcriptionnelle. bZIP53 et 67 semblent avoir une plus grande affinité pour les boîtes G de proOLE1 que les autres bZIPs testés. Toutefois l'hétérodimère bZIP53/67 semble avoir un effet additionnel et non synergique sur ce promoteur.

#### **II.4.5.3 Implication des bZIPs dans le complexe LEC1, ABI3, NF-YC2, LEC2, (LANL) sur le promoteur de proOLE1**

Les bZIPs des groupes C et S (bZIP10/25/53) ont été identifiées comme partenaires de ABI3, (Lara *et al.* 2003; Alonso *et al.* 2009). Les bZIP28 et 67 sont les partenaires des NF-Y : LEC1 et NF-YC2 (Yamamoto *et al.* 2009; Mendes *et al.* 2013). Afin de tester une possible

### Complexes protéiques testés avec le promoteur de l'*OLEOSINE1*



**Figure 14 : Activité des bZIPS avec le promoteur de l'*OLEOSINE1* en présence de LEC1, ABI3, NF-YC2, LEC2 (LANL) par la méthode d'expression transitoire chez la mousse *Physcomitrella patens*.**

Les bZIPS sont testés en homodimères (orange) ou hétérodimères (bleu). Valeur des résultats du test de Student *t*-test, N=4 réplicats technique \* : < 0,05, \*\* : <0,01, \*\*\* :<0,001. (NS) : non significatif. Les résultats obtenus avec le complexe est comparé à proOLE1:GFP seul (étoile rouge). Les valeurs des bZIPS sont comparées à la valeur d'obtenue avec le complexe ANLs seuls (étoiles noires). Les témoins sont les protoplastes seuls et un témoin de transformation.

activité de nos facteurs en complexe avec des bZIPs, celles-ci ont été testées par expression transitoire chez la mousse, avec le promoteur de l'*OLEOSINE1* en présence du complexe LANL, (Figure 14).

Les homodimères de bZIPs ne semblent pas avoir d'activité synergique ou additive avec les LANL sur le promoteur, seules bZIP53 et bZIP67 semblent avoir un effet mais non significatif. bZIP53 semble réduire l'activité du complexe LANL, alors que bZIP67 permettrait d'augmenter l'activité transcriptionnelle du complexe sur *proOLE1*. Les bZIPs 10/25/53 associées au bZIP28 en hétérodimère n'ont pas d'effet significatif sur l'activité transcriptionnelle du complexe, sur le promoteur *proOLE1*. Néanmoins bZIP10/28 et bZIP25/28 ont une tendance à avoir un effet synergique sur le promoteur. Les bZIPs 10/25/28 associées au bZIP67 en hétérodimérisation ont un effet positif sur l'activité transcriptionnelle du complexe LANL, contrairement au bZIP53 qui, lorsque la protéine dimérise avec les bZIP10 ou 25, réduit significativement l'activité transcriptionnelle du complexe. Les bZIP 53 et 67 ont une action opposée sur le complexe LANL, l'association de bZIP53/67 annule l'action de chacune, l'activité transcriptionnelle est identique à celle du complexe seul.

#### **II.4.6 Conclusion, discussion et perspectives**

(Tableau 3)

##### **II.4.6.1 Dimérisation des bZIPs et activités avec les LANL**

Les bZIPs seules ou en hétérodimères principalement, activent transcriptionnellement l'*OLEOSINE1*, néanmoins tous les dimères de bZIPs (autres que ceux testés dans cette étude) n'ont pas encore été identifiés.

Les bZIPs 53 et 67 ont une activité plus importante en homo ou hétérodimères que les autres. bZIP67 ne dimérise avec aucun des partenaires testés en double hybride, mais, en expression hétérologue, elle montre une action positive sur le complexe LANL, en homodimère ou hétérodimère associée aux bZIP10/25/28. bZIP67 semble associée au complexe LANL, sans identification formelle de(s) partenaire(s). Chez la mousse *Physcomitrella patens* 40 bZIPs ont été identifiées (Correa *et al.* 2008). bZIP67 pourrait donc interagir avec un bZIP endogène encore non identifiée.

bZIP67 joue un rôle dans l'expression des protéines de réserve et le gène a un profil d'expression similaire à celui de *LEC1-LIKE*, qui s'exprime de 6 à 18 DAP, plutôt que celui de *LEC1* de 3 à 14 DAP. bZIP67 agit sur l'expression des cruciférines avec *LEC1* et *LEC1-LIKE* en présence de NF-YC2 (Yamamoto *et al.* 2009, Chapitre I.2.2, figure 1 p. 20). *LEC1-Like* peut interagir avec d'autres bZIPs du même groupe telles que *EEL* et *ABI5*, avec d'autres membres des NF-YCs, pour activer l'expression des protéines de réserve et les

Expériences		bZIPs					
		bZIP10	bZIP25	bZIP28	bZIP53	bZIP67	
A	Interactions Protéines - protéines	bZIP10	+	+	+	+	-
		bZIP25	-	-	-	+	-
		bZIP28	Auto activation				
		bZIP53	-	+	+	-	-
		bZIP67	-	-	-	-	-
B	Activité des bZIPs avec le promoteur de l' <i>OLEOSINE1</i>	bZIP10	=	↗ **	=	↗↗ **	↗ **
		bZIP25		=	=	↗ **	↗ **
		bZIP28			=	↗ **	↗ **
		bZIP53				↗ **	↗ ***
		bZIP67					↗ **
C	Implication des bZIPs sur le complexe LANL avec <i>proOLE1</i>	bZIP10	=	=	↗ NS	↘↘ ***	↗↗
		bZIP25		=	↗ NS	↘↘ **	↗↗ *
		bZIP28			=	=	↗↗ *
		bZIP53				↘ NS	=
		bZIP67					↗ NS

**Tableau 3 : Résumé des résultats obtenus sur l'étude des bZIPs en relation avec les LANL, avec le promoteur de l'*OLEOSINE1*.**

(A) l'étude de la dimérisation des bZIPs par double hybride (-) absence d'interaction, (+) présence d'interaction. (B) l'étude de l'activité des dimères de bZIPs en homo ou hétéro-dimère sur le promoteur de l'*OLEOSINE1* par expression transitoire chez la mouche. (=) activité nulle sur *proOLE1:GFP*, (↗) activité transcriptionnelle du dimère avec *proOLE1:GFP*, (↗↗) activité transcriptionnelle plus importante du dimère sur *proOLE1:GFP*. (C) l'étude de l'activité des dimères de bZIPs en homo ou hétéro-dimère en présence du complexe ANLs L (LEC1), A (ABI3), N (NF-YC2), L (LEC2) sur le promoteur de l'*OLEOSINE1* par expression transitoire chez la mouche. (=) activité inchangée comparé au complexe LANL seul *proOLE1:GFP*, (↗) augmentation (↘) diminution de l'activité transcriptionnelle sur *proOLE1:GFP* en présence du complexe ANLs et du dimère comparé au complexe seul, (↗↗) augmentation (↘↘) diminution de l'activité transcriptionnelle plus importante sur *proOLE1:GFP* en présence du complexe ANLs et du dimère comparé au complexe seul. Valeur des résultats du test de Student *t*-test N=4 réplicats technique \* : < 0,05, \*\* : < 0,01, \*\*\* : < 0,001. (NS) : non significatif.

FAD (Yamamoto *et al.* 2009; Mendes *et al.* 2013). Les interactions entre bZIP67 et LEC1 ou LEC1-LIKE pourraient donc être médiées par NF-YC2.

Nous avons vu que les bZIPs peuvent interagir avec eux-même, des membres du même groupe et de groupes différents (Jakoby *et al.* 2002). Il est possible que bZIP67 puisse avoir différents partenaires au cours du développement de la graine. Une dimérisation différentielle de bZIP67 pourrait influencer, ou moduler, son activité au cours du développement par le changement de partenaire. Ainsi ABI5, qui est une protéine bZIP interagissant avec ABI3 pour le contrôle de l'expression de réserves tardives (Nakamura *et al.* 2001; Lopez-Molina *et al.* 2002), pourrait être le partenaire de bZIP67 pour l'expression de protéines de réserve lors de la maturation tardive.

ABI3 s'exprime durant tout le développement de la graine (Chapitre I.2.2, figure 1 p. 20). NF-YC2 semble intervenir dans le complexe entre ABI3 et LEC1 (Chapitre II.2.2, figure 6) et a une expression identique à ABI3. L'hypothèse de la formation d'un complexe formé par ABI3, LEC1-LIKE, bZIP67, ABI5 lors de l'expression des protéines de réserve au stade de la maturation tardive mériterait sans doute d'être étudiée.

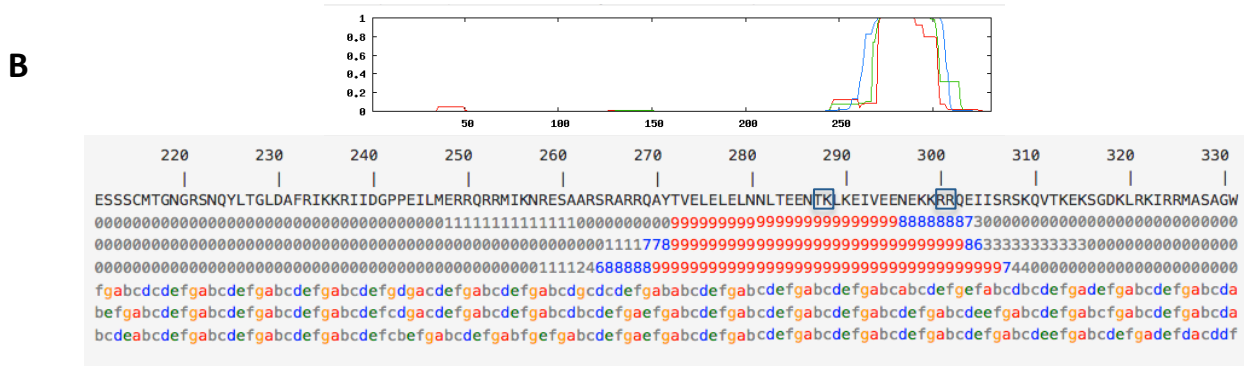
bZIP53 homodimérise, interagit avec les bZIPs 10, 25 et 28 et est également capable d'activer le promoteur de l'*OLEOSINE1*. bZIP53 en homodimère est moins active que lorsqu'elle est associée à une autre protéine bZIP (Figure 14) et l'activité du dimère bZIP10/53 sur le promoteur est plus importante que celles des autres dimères. Lorsqu'un dimère bZIP53/X est associé au complexe LEC1, ABI3, NF-YC2, LEC2, il semble avoir une activité antagoniste à bZIP67 et plus faible que son activité en homodimère. De même que l'homodimère, bZIP53, sur le promoteur est moins actif qu'un hétérodimère bZIP53/X (Figure 14). bZIP10/53 montre la plus forte baisse significative de l'activité du complexe comparée aux autres dimères.

Il a été montré qu'ABI3 interagit avec les dimères 10/53 ou 25/53 sur le promoteur des protéines de réserve et que cette interaction a un effet synergique sur leurs expressions (Lara *et al.* 2003). De la même façon, la dimérisation des bZIP10 et 25 avec bZIP53 augmente l'expression de l'*ALBUMINE 2S* dans des protoplastes transformés (Alonso *et al.* 2009). Le bZIP53 en homo ou hétérodimères avec les bZIP10 ou 25 pourrait donc être en compétition avec le complexe LANL. Les dimères 10/53, 25/53, pourraient titrer ABI3 et réduire l'efficacité transcriptionnelle du complexe LANL testé en réduisant la taille du complexe. La présence de ces dimères pourrait être un mode de contrôle de l'expression des protéines de réserve (Lara *et al.* 2003) ou un artefact de notre expérience, mettant en présence des complexes qui, normalement, ne sont pas présents en même temps dans la cellule.



**A**

At-NF-YA3	MRRRQQRAKLEAQN-KL-IRA-RKPYLHESRHHVHALK-----R
At-ABI3	LRLCSSATK-EARKKRM-ARQRRFLSHHHRHNNNNNNNNNNNQ
At-LEC2	GFLDPRMTK-MARINRKNAMMRSRNNSSPNSSPSELVDS----
At-bZIP67	283TEENTK-LKEIVEENEKKRRQEIIISRSKQV <sub>311</sub>
	306SKQVTK-EKSGDKLRKI-RRMASAGW <sub>331</sub>



**Figure 15 : Analyses de séquences du domaine B2.**  
 (A) Alignement des séquences correspondant au domaine B2 retrouvé chez les facteurs de transcription NF-YA, ABI3 et LEC2 et de la séquence du B2 putatif de bZIP67 d'*Arabidopsis thaliana*. (B) Analyse de la séquence de bZIP67 pour la recherche de prédiction de coiled-coil, situé dans le domaine B2 putatif. [https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_lupas.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_lupas.html) (Lupas *et al* 1991).

Le complexe entre les hétérodimères bZIPs 10/53 et 25/53 et ABI3 serait formé par l'interaction de bZIP10 et 25 avec ABI3, ABI3 n'interagissant pas avec l'homodimère bZIP53 (Alonso *et al.* 2009). Les dimères 28/53 et 53/53 ne baissent pas significativement l'activité du complexe sur *proOLE1*. bZIP 53 pourrait ne pas avoir pour partenaire uniquement ABI3 mais d'autres membres du complexe LANL, comme NF-YC2 qui interagit avec bZIP28 lors de l'expression des protéines en réponse à un stress (Liu and Howell 2010).

L'activité de dimères de bZIPs répond à un ratio optimal. En interaction avec ABI3, bZIP10/53 doivent être exprimées selon une stœchiométrie de 1/1 pour avoir une activité optimale de la transcription de la protéine de réserve (Alonso *et al.* 2009). La baisse d'activité du complexe LANL sur *proOLE1* pourrait être induite par la présence d'une stœchiométrie différente des bZIPs exprimées qui serait défavorable à l'activité synergique des bZIPs avec le complexe.

Enfin, les bZIPs 10 et 25 s'expriment tout au long du développement de la graine, contrairement à bZIP53 dont l'expression augmente au stade torpille (Lara *et al.* 2003; Weltmeier *et al.* 2009). L'activité transcriptionnelle des bZIP25/53 est moins importante que le dimère bZIP10/53 sur *proOLE1*, de même que sur le promoteur de l'*ALBUMINE 2S* (Alonso *et al.* 2009). bZIP53 n'est peut-être pas le partenaire principal de bZIP25, les bZIPs du groupe S dimérisant avec les bZIPs du groupe C (Ehlert *et al.* 2006), d'autres partenaires de bZIP25 pourraient être envisagés.

#### **II.4.6.2 Formation du complexe**

Les bZIPs 53 et 67 semblent avoir une action sur le promoteur de l'*OLEOSINE1*. Néanmoins ces résultats préliminaires ne sont pas suffisants pour conclure sur l'action répressive ou activatrice des bZIPs sur les facteurs de transcription individuellement.

bZIP67 interagit avec LEC1, LEC1-LIKE et NFYC2 sur les promoteurs de protéines de réserve (Yamamoto *et al.* 2009; Mendes *et al.* 2013). L'interaction moléculaire de bZIP67 n'a pas été étudiée et on ne peut pas affirmer que bZIP67 interagit individuellement avec un facteur ou avec les deux. L'analyse de la séquence de bZIP67 montre une micro-homologie de sa séquence sur la partie C terminale avec les facteurs de transcription AFL-B3 : ABI3 et LEC2. L'alignement révèle la présence d'une séquence similaire au domaine B2 retrouvée chez LEC2 et ABI3 (Figure 15) avec, en particulier, une prédiction forte de formation d'un domaine coiled-coil. Nous avons démontré que le domaine B2 est impliqué dans l'interaction entre LEC2 et LEC1 (Chapitre II.3). La séquence de bZIP67 identique au B2 est prédite, de la même manière que pour LEC2 et ABI3, comme pouvant former une structure coiled-coil (Figure 15). Nous pouvons supposer que bZIP67 interagit avec le NF-YB/C par l'intermédiaire de ce domaine. bZIP28 interagit également avec NF-YC2 (Liu and Howell

2010), mais aucune séquence similaire au domaine n'a été retrouvée dans la séquence de bZIP28. bZIP67 pourrait interagir directement avec LEC1 et LEC1-LIKE par l'intermédiaire de la microsimilarité au domaine B2. Le gène VP1, homologue de ABI3 chez le maïs, semble lui aussi se lier à OPAQUE2 par l'intermédiaire du domaine B2 (Hill *et al.* 1996). L'interaction d'ABI3 avec bZIP 10 et 25 pourrait être identique. L'importance de cette microsimilarité de bZIP67 pourra être testée par mutagenèse dirigée, dans le système d'expression transitoire de mousse.

Tout comme les *LAFL*, les *bZIPs* étudiés sont exprimés dans la graine. Néanmoins, l'expression des *bZIPs* peut varier au cours des différents stades du développement de l'embryon (embryogenèse, maturation, maturation tardive). Hormis *bZIP25* qui a une expression basse au cours du développement, l'expression des *bZIPs* 10/28/67/53 coïncide avec le profil d'expression d'*ABI3* et *LEC1-LIKE*. *LEC2* et *LEC1* s'expriment, eux, aux premiers stades de développement de l'embryon. *NF-YC2* est exprimé au même stade de développement que *LEC1* et au premier stade d'expression de *LEC1-LIKE* (Annexe 2 + Chapitre 1.2.2, figure 1 p. 20). Ce différentiel d'expression laisse supposer que le complexe responsable de l'expression des protéines impliquées dans le contrôle de l'accumulation des réserves lipidiques et protéiques évolue au cours du développement et de localisation des réserves. Ainsi d'autres *bZIPs* pourraient être impliquées dans l'expression de ces gènes à des stades de développement précis.

#### **II.4.6.3 Action des *bZIPs* sur le promoteur**

En système hétérologue de protoplastes de mousse, les *bZIPs* ont une activité transcriptionnelle sur le promoteur de l'*OLEOSINE1* bien que plus faible que le complexe LANL. Certains *bZIPs*, comme le *bZIP1*, ont différents modes d'action dans la régulation des gènes et les cibles de *bZIP1* ont été réparties en 3 classes (Para *et al.* 2014):

- classe I : les gènes qui contiennent une séquence de fixation du *bZIP*, mais ne sont pas régulés par le *bZIP* ;
- classe II : les gènes reconnus et régulés par le *bZIP* ;
- classe III : les cibles régulées par le *bZIP* mais dont l'interaction est difficile à démontrer expérimentalement. L'action des facteurs de transcription de classe III est communément appelée le « Hit and run » (Para *et al.* 2014). Pour répondre aux stimuli, l'action de ces facteurs induit un temps de réaction rapide (Li *et al.* 2015). Par exemple, *bZIP1*, qui interagit avec 10 et 25 (Ehlert *et al.* 2006), a par ailleurs une activité « hit and run », qui lui confère la possibilité d'initier la transcription des gènes et de maintenir l'expression par le recrutement d'autres facteurs de transcription, il joue ainsi le rôle de cofacteur (Li *et al.* 2015). L'activité des *bZIPs* testées montre que *proOLE1* est régulée positivement par les *bZIPs* seules. Cependant, les affinités des *bZIPs* pour *proOLE1* n'ont pas été mesurées ni individuellement

ni collectivement. Certaines bZIPs testées entreraient dans la classe II ou la classe III et pourraient aussi jouer le rôle de cofacteur dans le complexe de régulation qui nous intéresse ici.

#### **II.4.6.4 Conclusions et perspectives**

Les bZIPs que nous avons testées ont une activité transcriptionnelle sur *proOLE1*, moins forte que le complexe LANL seul. De même que les protéines testées n'ont pas toutes l'activité attendue sur *proOLE1* : bZIP67 a un effet synergique sur le complexe LANL, (avec une activité supérieure à la somme des activités des homodimères) alors que bZIP53 réduit l'activité transcriptionnelle du complexe ; d'autres n'ont pas d'effet, ni additif, ni synergique.

De nombreuses hypothèses ressortent des résultats préliminaires des dimères de bZIPs testées et laissent penser qu'ils peuvent avoir différentes activités telles que (i) un rôle de cofacteur, (ii) une interaction préférentielle qui empêche la formation du complexe, (iii) une interaction transitoire avec le promoteur, (iv) faire partie intégrante du complexe, (v) répondre à un ratio de bZIPs optimal pour former un dimère avec une activité plus forte.

Enfin les bZIPs testées ne sont pas forcément les « bons » dimères, *i.e.* présents *in vivo* à un stade de développement donné.

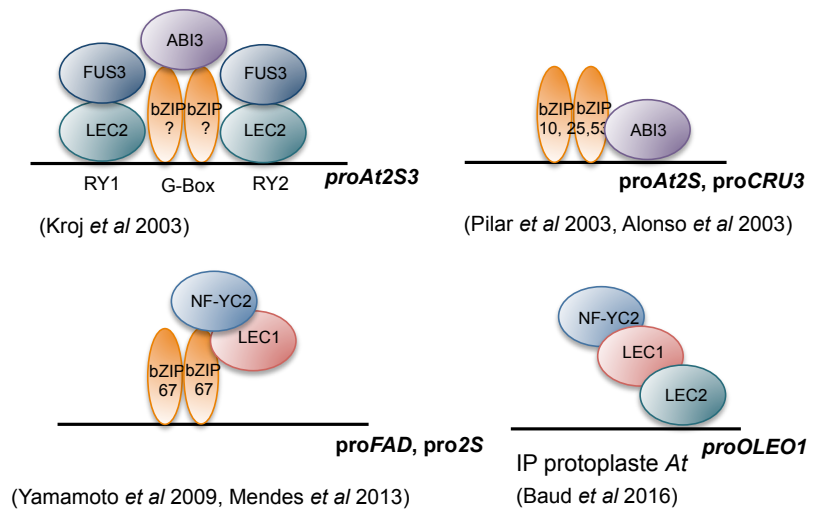
Afin de tester ces hypothèses, l'étude de l'interaction des dimères de bZIPs avec les facteurs de transcription LEC1, ABI3, NF-YC2, LEC2 (LANL), permettra de mieux identifier les partenaires LANL putatifs d'un dimère de bZIPs, en particulier pour bZIP67 et bZIP53. Une meilleure compréhension de la formation des dimères pourrait permettre de mieux caractériser la taille critique du complexe ou d'identifier le facteur inhibant l'interaction des hétérodimères bZIP53. Pour répondre à ces questions, l'étude de la stœchiométrie de bZIP10 avec bZIP53 sur *proOLE1* pourrait être menée par expression transitoire chez la mousse. Une étude similaire a notamment été utilisée sur le promoteur de l'*ALBUMINE 2S* par l'utilisation de bZIP10 muté, incapable de former un dimère (Alonso *et al.* 2009).

Il serait également envisageable d'identifier la classe des bZIPs testés par détermination du  $K_d$  (en thermophorèse) des dimères sur le promoteur *proOLE1* ce qui permettrait de déterminer l'activité propre de chacun des dimères et d'affiner leur action sur le promoteur.

Par ailleurs, une autre piste serait de rechercher de nouveaux dimères, notamment pour bZIP25, qui montre moins d'activité avec bZIP53 que le dimère bZIP10/53. En effet les bZIPs 10 et 25 interagissent en double hybride avec tous les membres du groupe S de bZIP53. Parmi les membres de ce groupe, les bZIPs 11 et 44 s'expriment aux premiers stades de développement de la graine (Ehlert *et al.* 2006; Weltmeier *et al.* 2009) et sont des candidats naturels pour jouer un rôle dans un complexe à ce stade du développement.

**A**

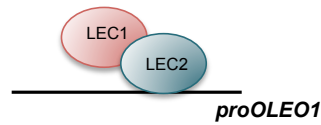
Littérature



**B**

Méthodes

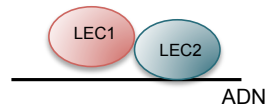
Thermophorèse



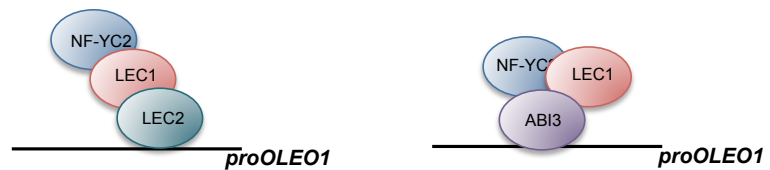
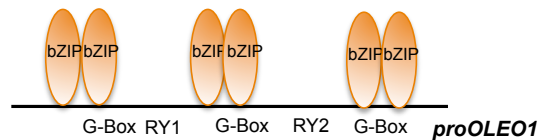
Double Hybride



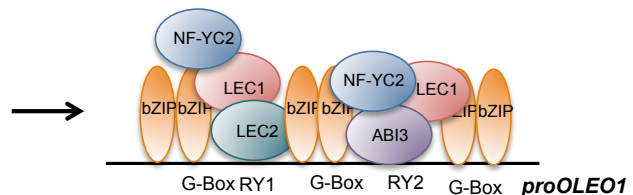
Immuno localisation  
*In planta*



Protoplaste de mousse



Constitution du  
complexe  
bZIPs +LANL  
individuellement  
??



**Figure 16 : Résumé des connaissances sur le complexe LAFL.**

(A) Résumé des connaissances issues de la littérature. (B) Résumé des résultats expérimentaux et hypothèses issus des travaux présentés.

À ce jour aucun dimère de bZIPs n'a été identifié comme ayant une interaction spécifique avec LEC2. Les bZIPs se fixent sur les boîtes « G ». Dans le promoteur de l'*OLEOSINE1*, les boîtes RY sont encadrées de boîtes « G » et l'activité des dimères de bZIPs ne montre pas toujours d'effet additif ou synergique. Les bZIPs pourraient avoir un rôle de stabilisation du complexe et favoriser une interaction indirecte des B3 sur le promoteur.

## **II.5 Conclusion, discussion et perspectives**

L'étude réalisée sur les complexes LANL et les résultats préliminaires de recherche de nouveaux partenaires ont permis de valider des interactions et d'en suggérer de nouvelles. L'interaction entre LEC2 et LEC1 a permis de mieux comprendre le fonctionnement de ce complexe et de découvrir un domaine nécessaire à sa formation (Figure 16).

LEC2 et LEC1 co-localisent en partie au noyau mais les protéines semblent aussi être localisées dans le cytoplasme. Les NF-YCs sont impliqués dans le transport et la localisation des NF-YBs (Hackenberg *et al.* 2012), LEC1, NF-YB9, pourraient être transportés au noyau par le NF-YC. Les protéines LEC2 étiquetées par la GFP montrent aussi une localisation cytoplasmique, en plus de leur localisation nucléaire. Deux hypothèses non exclusives sont donc possibles (i) LEC2 entrerait dans une voie (cytoplasmique) de dégradation une fois son rôle effectué, (ii) l'activité de LEC2 serait régulée finement par les voies de l'import et/ou l'export nucléaire (Figure 10). La séquence en acides aminés de LEC2 prédit 3 principaux sites d'ubiquitination : les lysines 153, 154, 273 sont fortement prédites pour être modifiées et permettraient de dégrader la protéine par la voie du protéasome. Les protéines ABI3 et FUS3 sont modifiées et régulées par AIP2, une E3 ligase (Duong *et al.* 2017; Zhang *et al.* 2005). L'accumulation de LEC2 pourrait être observée sur des lignées AIP2 déficientes. Par ailleurs, les acides aminés fortement prédits comme ubiquitinés pourraient être la cible de mutations ou, alternativement, l'utilisation de drogues bloquant le protéasome serait une première étape pour valider l'utilisation de cette voie de régulation et la dégradation de LEC2. L'activité de LEC2 pourrait être médiée au noyau par la voie de l'import et de l'export et par différents états de phosphorylations. Les EXPORTINES et IMPORTINES sont responsables du trafic nucléaire, les drogues testées responsables du blocage de certaines protéines de l'import et l'export n'ont pas d'effet sur la mobilité de LEC2 dans les conditions utilisées. Il est possible que LEC2 utilise d'autres voies de transport telles que les IMPORTINES  $\beta$  et d'autres EXPORTINES non touchées par la drogue. Enfin, certaines phosphorylations sont situées dans les NLS et NES, d'autres dans le domaine B3 et pourraient moduler l'accès à ces séquences. La modification de ces acides aminés par des mutations phosphomimétiques ou phosphodéficientes permettrait probablement de mieux comprendre la mobilité de LEC2.

LEC1 et LEC2 forment un complexe, les deux protéines interagissent entre elles par l'intermédiaire d'une séquence conservée du domaine B2, retrouvée chez *ABI3* et *LEC2* et absente dans la séquence de *FUS3*. Le domaine B2 semble avoir évolué à partir du gène d'*ABI3*. Le domaine B3 permet à *LEC2* de se fixer sur les boîtes RY de *proOLE1*, mais la protéine n'a pas une forte affinité pour sa cible. Lorsque *LEC1* forme un complexe avec *LEC2*, l'affinité pour l'ADN est plus importante que celle de *LEC2* seule. L'acide aminé D86 est nécessaire à l'activité de *LEC1* et est important dans l'interaction de *LEC1* avec *LEC2*. Une étude structurale des protéines permettrait sans doute de comprendre le rôle de cet acide aminé dans son interaction avec le B2, dans son impact sur le B3 et le rôle dans le changement d'affinité de *LEC2* pour l'ADN.

La littérature mentionne la capacité d'*ABI3* et de *LEC1* d'interagir avec d'autres facteurs de transcription (NF-YBs forment des dimères avec les NF-YC, *LEC1* interagit avec NF-YC2). *ABI3* et *LEC1* sont aussi capables d'interagir avec des bZIPs, les expériences menées ont permis de révéler la capacité des bZIPs à activer le promoteur *proOLE1* et d'être impliqués dans l'activité du complexe constitué de *LEC1*, *ABI3*, NF-YC2, *LEC2*. Les bZIPs 53 et 67 semblent, notamment, avoir des activités antagonistes sur le complexe. Ces résultats préliminaires seront affinés, afin de mieux comprendre le rôle de ces bZIPs sur le complexe et les interactions protéine-protéine des bZIPs pour l'un des facteurs de transcription. D'autres facteurs peuvent aussi être pris en compte notamment les bZIPs 11 et 44 du même groupe que bZIP53 et NF-YC3, dont le profil d'expression est similaire à celui de *LEC1* et qui interagit en double hybride avec *LEC1*.

Le complexe prédit, formé sur *proOLE1* n'a pas été reconstitué biochimiquement *in vitro*, néanmoins, l'existence d'une cinétique de formation du complexe est envisageable et des expériences de reconstitution du complexe prenant en compte cette problématique pourrait être envisagées. *LEC1* interagit avec *LEC2* sur le promoteur, mais aussi avec *ABI3*, *FUS3* et des bZIPs sur l'expression de gènes au cours de la maturation de la graine (Pelletier *et al.* 2017), la formation du complexe *LEC1*, *ABI3*, *LEC2* *in planta* est donc fortement probable.

### III Recherche de partenaires sans a priori

#### III.1 Introduction

La recherche de partenaires avec a priori a permis d'ouvrir des hypothèses de partenaires potentiels pour les LAFL, par exemple : LEC2 forme un complexe avec ABI3 et LEC1 sur le promoteur de l'*OLEOSINE1* et NF-YC2, LEC1 et LEC2 immunoprécipitent (Baud *et al.* 2016) mais aussi des partenaires liés à l'activité, le transport et la régulation. Cette recherche n'a pas permis d'identifier de nouveaux partenaires de LEC2. Il est cependant probable que d'autres protéines interagissent avec les LAFL et, en particulier, avec LEC2.

Dans cette seconde partie, nous nous sommes intéressés à la recherche de nouveaux partenaires, afin de connaître le réseau d'interaction de LEC2 qui permettrait d'appréhender toutes les activités de la protéine. Au démarrage de l'étude, n'ayant pas d'anticorps spécifiquement dirigés contre la protéine, la recherche sans a priori a été menée par purification de la protéine étiquetée selon les critères suivants :

- (i) pour identifier les partenaires impliqués dans les différentes activités de LEC2 (transport régulation et activité transcriptionnelle) la recherche de nouveaux partenaires liés à LEC2 a été réalisée *in vivo*,
- (ii) afin de couvrir un maximum d'identification de partenaires en spectrométrie de masse, l'étude a été réalisée sur les stades de développement précis de l'expression de LEC2, c'est-à-dire du stade globulaire au stade torpille (Chapitre I.2, Chapitre II.3),
- (iii) le choix du promoteur approprié est essentiel afin d'exprimer la protéine étiquetée dans un contexte embryonnaire normal (Stone *et al.* 2008; Feeney *et al.* 2013),
- (iv) la stratégie de capture de complexe a été réfléchi et choisie en fonction de l'étiquette et de sa technologie de purification, de sa position et de la fonctionnalité de la protéine étiquetée.

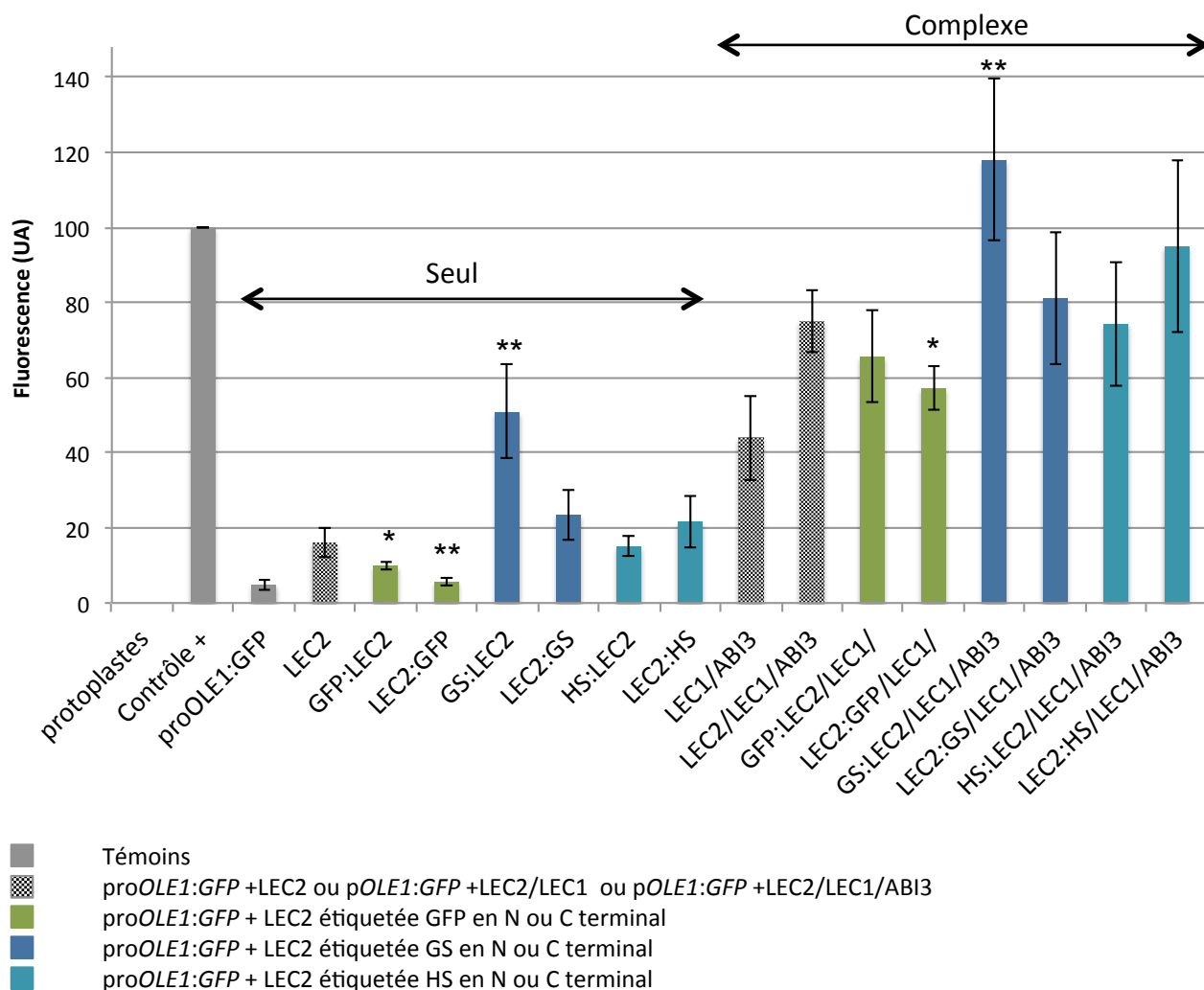
#### III.2 Recherche de partenaire

##### III.2.1 Fonctionnalité des LEC2 étiquetées

De nombreuses étiquettes avaient été utilisées auparavant pour des études *in vivo* sans succès, hormis le tag « myc » en immunoprécipitation avec des protoplastes d'*Arabidopsis* transformés (Baud *et al.* 2016). Pour permettre la capture de complexes protéiques et garder une protéine LEC2 fonctionnelle, le choix s'est porté sur 3 étiquettes de tailles différentes, de types de purification différents et testées en position N et C terminal de la protéine (Tableau 4). La GFP, est une grosse étiquette à double utilité : la localisation de la protéine *in vivo* et la disponibilité d'anticorps commerciaux très affins, intéressants pour la purification des complexes en immunoprécipitation. Le tag GS est communément utilisé pour la technique du TAP-TAG (Van Leene *et al.* 2008). Il est composé de 2 protéines G, d'un site de clivage de la protéase TEV et de la streptavidine qui a une forte affinité pour la biotine. Le tag HS est



### Activité de LEC2 étiquetée avec le promoteur de l'*OLEOSINE1*



**Figure 17 : Tests d'activité de LEC2 sur le promoteur de l'*OLEOSINE1*, en fonction de différentes étiquettes, par expression transitoire chez la mousse *Physcomitrella patens*.**

Étiquettes en N ou C terminal de différents poids moléculaire. GFP: 27 kDa, GS (protéine G\*2-TEV-streptII): 18 kDa, HS (Histidine \*6-StreptII): 3 kDa. Les activités des LEC2 étiquetées sont testées. Valeur des résultats du test de Student *t*-test N=4 réplicats technique \* : < 0,05, \*\* : < 0,01, \*\*\* : < 0,001. (NS) : non significatif. Les résultats des LEC2 étiquetés sont comparés aux valeurs de LEC2 sauvage seule ou en complexe.

une petite étiquette, pour s'assurer de la meilleure fonctionnalité de LEC2, qui est constituée de six histidines et de STREPII, sans site de clivage pour ne pas contaminer l'échantillon avec une protéase.

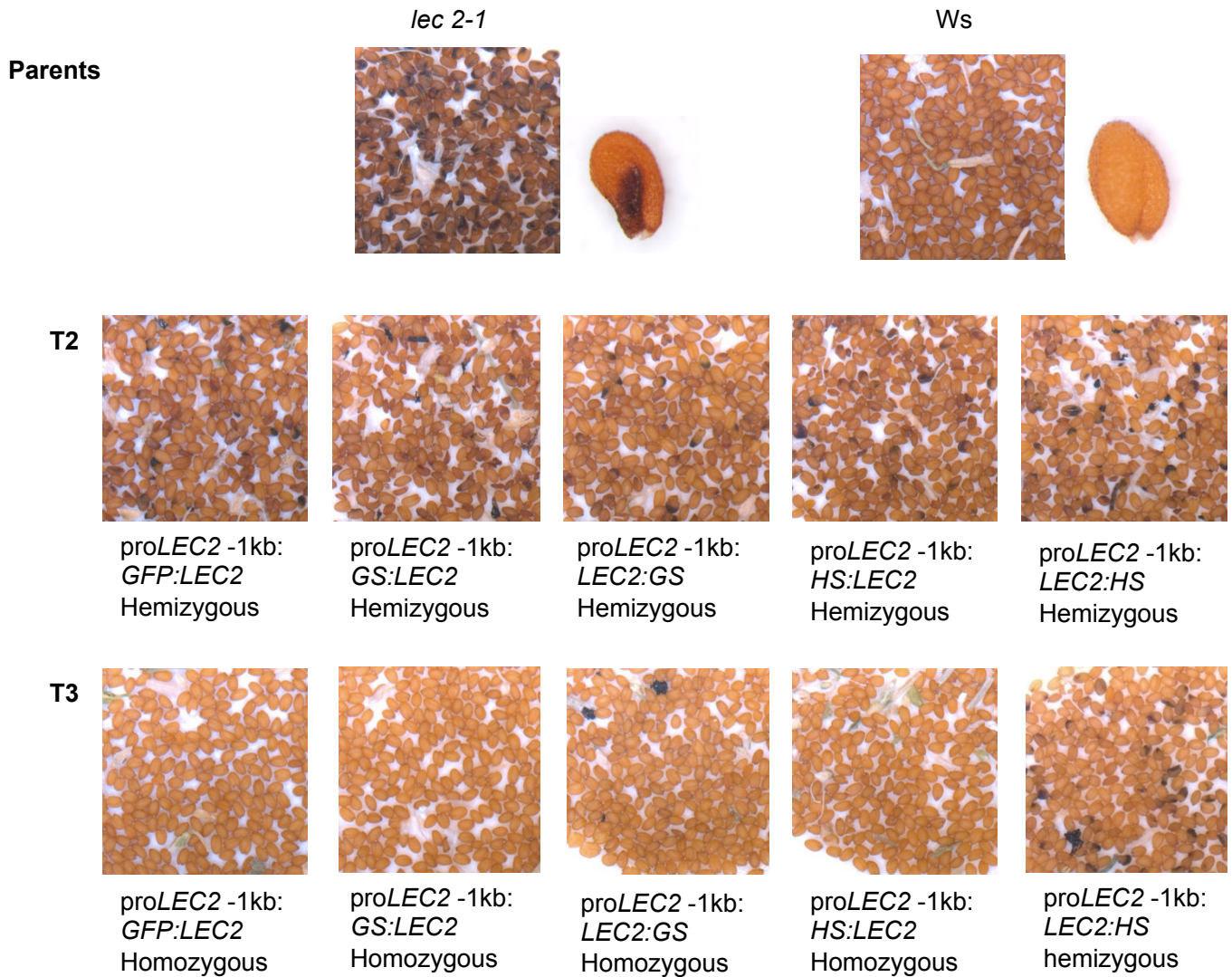
Étiquette	Purification	Masse moléculaire	Localisation	Détails
GFP	<u>Immunoprécipitation:</u> 1-Anti GFP	26 kDa	N / C Terminal	Gros Pas de LEC2GFP observée <i>in planta</i>
<b>GS:</b> Protéine G – Protéine G -TEV- SBP	<u>TAP-tag:</u> 1-Anti protéine G 2-Cleavage TEV protéase 3-Streptavidin (SBP)	18 kDa	N / C Terminal	Communément utilisée Grosse Contamination à la protéase TEV
His - StrepII (HS): His6-Gly3- STREPII-Gly7	<u>TAP-tag:</u> 1-résineNinta 2- <i>Strep</i> -Tactin®	3 kDa	N / C Terminal	Sans protéase, petite

**Tableau 4: récapitulatif des étiquettes et descriptif des étiquettes choisies pour la capture et la purification de complexe lié à LEC2.**

LEC2 est un gène dont l'expression est régulée transcriptionnellement (Berger *et al.* 2011). De plus, LEC2 induit l'embryogenèse somatique (Gaj *et al.* 2005; Stone *et al.* 2008) et sa surexpression dans les feuilles induit un changement d'état, les cellules des feuilles se redifférenciant en cellules embryonnaires avec l'apparition de structures spécifiques des embryons (corps lipidiques, lipides de réserve) jusqu'à, parfois, des embryons somatiques (Feeney *et al.* 2013; Guo *et al.* 2013). Les constructions, permettant l'expression de LEC2 chez la plante pour l'étude des fonctions de LEC2 dans l'embryon par l'identification des interactions protéiques et la formation de complexe lié à LEC2, ont été réalisées avec le promoteur endogène de LEC2 (-1Kb) et des étiquettes positionnées en N ou C terminal de la protéine. Pour cela la technologie du Gateway® double a été utilisée.

Afin de valider la fonctionnalité de LEC2 étiquetée en N et C terminal et ainsi sélectionner les étiquettes et leurs positions optimales permettant d'obtenir une protéine LEC2 fonctionnelle *in vivo*, l'activité des protéines étiquetées a été analysée en système hétérologue de mousse sur le promoteur de l'*OLEOSINE1*. LEC2 agit avec ABI3 et LEC1 sur ce même promoteur (Baud *et al.* 2016), les activités des différentes protéines LEC2 étiquetées ont été analysées seules et en complexe (Figure 17).

Les protéines étiquetées avec la GFP en N et C terminal et GS en N terminal, montrent une différence d'activité significative comparée avec celle de LEC2 sauvage seule. Les autres protéines étiquetées n'ont pas de différence d'activité significative à LEC2. Les protéines



**Figure 18 : Restauration phénotypique du mutant *lec2-1* par différentes constructions permettant d'étiqueter LEC2.**

Obtention des lignées homozygotes transformées avec proLEC2 -1kb: LEC2 étiquetées en fond *lec2-1*.

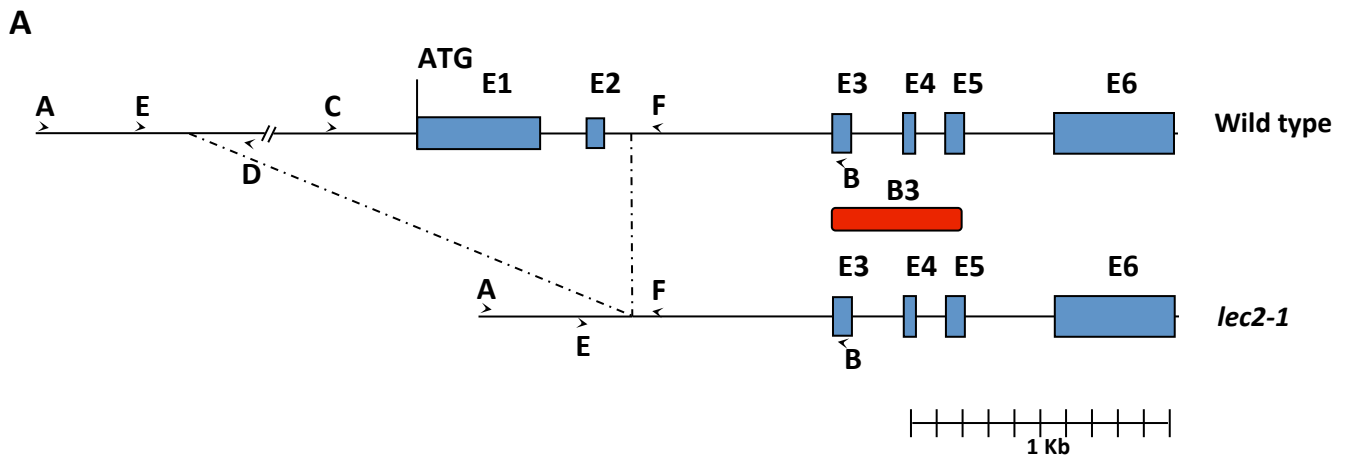
Les trois étiquettes sont: la GFP, le tag GS composé de 2 protéines G, d'un site de clivage de la protéase TEV et de la streptavidine, le tag HS constitué de six histidines et de STREPII, sans site de clivage.

LEC2 comportant les étiquettes ont été testées en complexe afin de s'assurer de la fonctionnalité de LEC2 avec les autres protéines en présence de l'étiquette. Concernant le tag GFP, comme pour LEC2 seule, la position en C-terminal réduit significativement l'activité du complexe, la différence entre le complexe à 3 protéines avec LEC2:GFP et le complexe constitué de ABI3 et LEC1 devient non significative, LEC2:GFP a peu d'activité et n'est pas assez fonctionnelle pour la formation d'un complexe. L'activité de GFP:LEC2 reste suffisante pour ne pas être significativement différente du sauvage en complexe. GS:LEC2 a une activité augmentée comparée à LEC2, l'étiquette semblant perturber l'activité de la protéine. LEC2:GS, HS:LEC2, LEC2:HS ne montrent pas de différences significatives comparé à LEC2 lors de la formation du complexe avec ABI3 et LEC1.

### III.2.2 Capture de complexe protéique lié à LEC2 *in vivo*

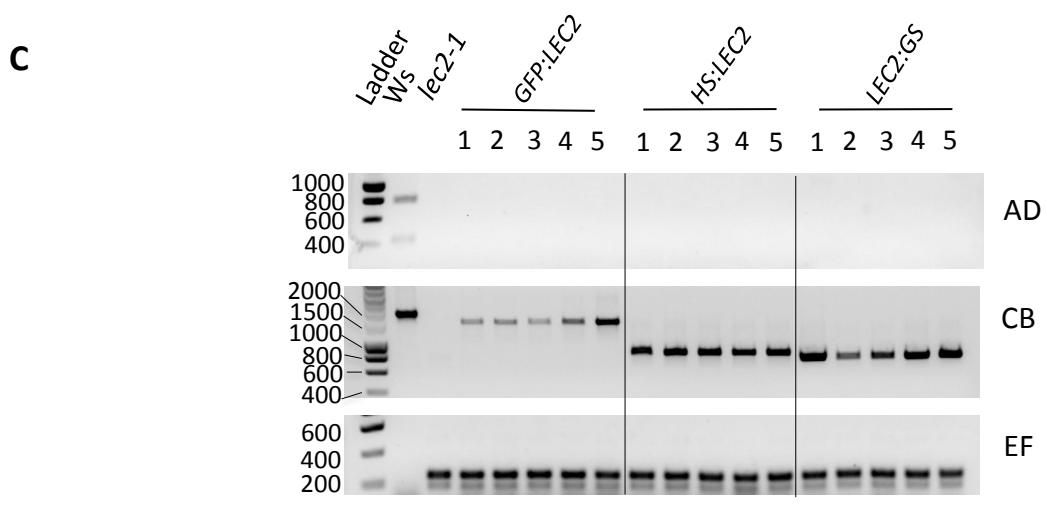
Afin de réaliser les expériences de capture de complexe protéique, les construits LEC2 étiquetés sous le contrôle du promoteur endogène de *LEC2* (-1Kb) ont été introduits dans des lignées *lec2-1* N3868 (Stone *et al.* 2001). La lignée mutante, en fond Ws, montre un fort phénotype par l'accumulation des anthocyanes dans les cotylédons et une pénétrance complète de la mutation (Figure 18). De par son manque de fonctionnalité, LEC2:GFP n'a pas été utilisée, les cinq autres LEC2 étiquetées ont été exprimées dans la lignée mutée. Après sélection, quatre des constructions montrent des lignées homozygotes qui complètent le mutant, pour trois des LEC2 étiquetées : GFP:LEC2, LEC2:GS et HS:LEC2, pour lesquelles 5 lignées indépendantes complétées ont été obtenues. Pour GS:LEC2, il n'a pas été possible d'obtenir suffisamment de lignées indépendantes, complétées, pour la suite des expériences. LEC2:HS ne complète pas en totalité la lignée mutée. Ces deux construits montrent une activité augmentée de LEC2 en expression transitoire qui est due à la présence de l'étiquette. *In vivo*, l'activité de LEC2 semble être aussi très régulée. Le génotypage des cinq lignées sélectionnées pour les construits GFP:LEC2, LEC2:GS et HS:LEC2, a permis de valider l'obtention des outils (version étiquetée de LEC2, fonctionnelle, homozygote, en fond *lec2*) permettant de réaliser les expériences de capture de complexe (Figure 19).

Afin de déterminer la quantité de matériel nécessaire pour réaliser l'étude, des recherches préliminaires d'identification de LEC2 étiquetées ont été réalisées en spectrométrie de masse à partir de matériel végétal en développement. Tout d'abord, avec un extrait brut de graines provenant de 400 siliques en développement, le protéome total révélant majoritairement des protéines du cytoplasme indique la faible teneur en protéine LEC2 dans l'extrait, cachée par les protéines plus abondantes. Ces résultats ont démontré la nécessité de purifier l'extrait. Une seconde recherche de LEC2 a été réalisée à partir de graines provenant de 400 siliques en développement dont les noyaux ont été extraits et le protéome



**B**

	Primers	DNA		
		Wild type (Ws) (DNA)	<i>lec2-1</i> (DNA)	pLEC2 (-1kB): Tagged LEC2 (cDNA)
mutant background	<b>AD</b>	800 pb	-	-
	<b>EF</b>	4191 pb	298 pb	298 pb
Insert	<b>CB</b>	1939 pb	-	<i>GFP:LEC2</i> : 1781 pb <i>HS:LEC2</i> : 1081 pb <i>LEC2:GS</i> : 1026 pb



**Figure 19 : Génotypage des transformants LEC2 - étiquetées.**

(A) schéma comparatif du locus LEC2 sauvage et mutant, avec la position des amorces utilisées. (B) Résultats attendus des fragments amplifiés par PCR en fonction des couples d'amorce testés. (C) Gels d'agarose, révélés par BET, récapitulatifs des amplifications par PCR en fonction des amorces testées sur les lignées LEC2 étiquetées sélectionnées.

a été analysé en spectrométrie de masse. L'objectif était de concentrer les recherches d'identification sur le noyau, supposé contenir la quantité majeure de la protéine et permettre d'éliminer le bruit de fond de protéines cytoplasmiques observées auparavant. Aucun peptide de LEC2 n'a été identifié dans l'extrait, l'analyse révélant des protéines majoritaires contenues dans le noyau. Une dernière étude a été effectuée sur une lignée GFP:LEC2, avec la même quantité de matériel initiale, où la protéine a été purifiée sur résine par immunoprécipitation. La totalité de l'éluat a été analysée en spectrométrie de masse. La liste des peptides de LEC2 obtenus de l'identification de la protéine recombinante (Annexe 3) a été utilisée afin de rechercher spécifiquement LEC2 lors de l'analyse. Aucun peptide spécifique de LEC2 n'a été identifié, probablement parce que la protéine est faiblement exprimée.

Bien que cette première recherche n'ait pas été fructueuse, le but était essentiellement de déterminer la quantité de matériel nécessaire par l'identification de LEC2 en spectrométrie de masse, la capture des complexes protéiques liés à LEC2 a quand même été réalisée sur 3 lignées de LEC2 étiquetée pour chacune des différentes étiquettes, associées à trois lignées témoins exprimant la GFP étiquetée avec le même tag. Les 3 purifications, différentes de celles des tests, laissaient espérer une possible identification de LEC2. Cependant pour éviter que cette absence d'identification ne soit liée spécifiquement à la purification de noyaux, les trois lignées ont été traitées selon deux méthodes différentes pour optimiser nos chances d'obtenir un résultat positif, mais qu'il faudra, bien entendu, reproduire. Pour chacune des lignées (LEC2 et témoin), les graines de 630 siliques en développement ont été prélevées. Pour 2 lignées (LEC2 et le témoin), les noyaux des extraits de graines ont été purifiés, alors que les troisièmes lignées (LEC2 et le témoin) ont été analysées à partir d'un extrait brut. La totalité des extraits ont ensuite été purifiés selon la composition des étiquettes.

Afin d'étudier le protéome et d'identifier les protéines spécifiquement extraites des lignées LEC2, les éluats des purifications ont été analysés en spectrométrie de masse, pour chacun de trois tags et des différentes répétitions. La liste de la totalité des protéines identifiées est présentée en annexe 4, Le tableau représente les résultats par groupe de protéines ayant les mêmes peptides identifiés, la protéine majoritaire du groupe apparaît en premier.

Cent vingt quatre groupes de protéines ont été identifiées en spectrométrie de masse pour la totalité des échantillons, localisés dans différents compartiments cellulaires : noyau, cytoplasme, chloroplaste, mitochondrie. Aucune protéine n'a été identifiée dans les échantillons comportant les tags GS et HS, aussi bien pour les témoins que pour les échantillons. La totalité des protéines a été identifiée à partir de la purification par immunoprécipitation de la GFP. Quel que soit l'échantillon (noyaux purifiés ou extraits bruts)

Group ID	Protein Description	Activity	Protein location	peptides identify	coverage	Expression
b51	AT1G20220.1   Alba DNA/RNA-binding protein	mRNA binding	cytosol, nucleus,	3	12	ubiquitously
b66	AT5G36230.1   ARM repeat superfamily protein	translation initiation factor activity, regulation of translational initiation	cytosol, membrane, nucleus	2	5	ubiquitously
b67	AT5G04530.1   Symbols: KCS19   3-ketoacyl-CoA synthase 19	fatty acid biosynthetic process, synthase family involved in the biosynthesis of VLCFA (very long chain fatty acids).	extracellular region,	2	5	ubiquitously Seed++
b77	AT5G42800.1   Symbols: DFR, TT3, M318   dihydroflavonol 4-reductase	in the biosynthesis of anthocyanins, coenzyme binding, flavanone 4-reductase activity	endoplasmic reticulum membrane	2	6	ubiquitously Seed ++
b79	AT5G11350.1   DNAse I-like superfamily protein	regulation of transcription, DNA-templated, transcription, DNA-templated, RNA binding, poly(A)-specific ribonuclease activity	cytoplasm, nucleus	2	3	ubiquitously
b96	AT2G28600.1   P-loop containing nucleoside triphosphate hydrolases superfamily protein	ATP binding, nucleic acid binding	nucleus	2	4	ubiquitously
c101	AT5G27720.1   Symbols: emb1644   Small nuclear ribonucleoprotein family protein	U6 snRNA binding, snRNP assembly, embryo development ending in seed dormancy	nucleus	1	10	ubiquitously
c104	AT3G42170.1   BED zinc finger ;hAT transposases family dimerisation domain	post-embryonic development, regulation of transcription from RNA polymerase II promoter, transcription, DNA-templated	cytosol, nucleus	1	2	ubiquitously
c109	AT1G70160.1   zinc finger MYND domain protein	putatif epigenetic activity, putatif homolog ASHR1 activity H3K36 methyltransferase activity	extracellular region	1	3	ubiquitously

**Tableau 5 : Résumé des identifications des protéines par spectrométrie de masse après purification des complexes protéiques par immuno-précipitation GFP**

Group ID : classement des protéines par groupe ayant les mêmes peptides identifiés. Protéine Description : description de la protéine identifiée du groupe. Protein Activity : Activité décrite de la protéine à partir de la base de donnée Tair. Location : localisation de la protéine dans la cellule décrite dans la base de données Tair (<http://www.arabidopsis.org>), seed ++: représente un pic d'expression dans la graine (Annexe 5). Peptide indentify : nombre de peptides identifiés. Coverage : pourcentage de couverture de la protéine à partir des peptides identifiés. Expression, profil d'expression des protéines au cours du développement de la graine.

tous les témoins ont des protéines identifiées dans la liste. La GFP n'a été identifiée que dans un seul échantillon, le témoin GFPGS 4-6, signifiant la faible abondance de la GFP.

Parmi les 124 groupes de protéines identifiées, seuls 9 sont représentés dans l'échantillon GFP :LEC2 17-1, dont la purification a été réalisée à partir de l'extrait brut (Annexe 4), aucune protéine n'est identifiée dans les deux autres échantillons GFP :LEC2 dont la purification a été réalisé à partir des noyaux. Toutes les identifications proviennent d'un seul type de purification, l'immunoprécipitation, mais celle-ci nécessitant moins de lavages, les protéines sont peut-être toutes des contaminants.

Parmi les neuf protéines identifiées (Tableau 5), les protéines snRNP (c101), BED zinc finger (c104) et à domaine MYND (c109), n'ont qu'un seul peptide identifié, ce qui est insuffisant pour les considérer comme de vrais positifs et valider dans la recherche de partenaires. KCS19 (b67) et TT3 (b77) ont une expression identique à celle de LEC2 (sur la base de l'accumulation des ARN des gènes considérés, annexe 5), néanmoins ces 2 protéines ne sont pas localisées au noyau et n'ont pas d'activité en lien direct celle de LEC2. Les autres protéines ont une expression ubiquitaire (Annexe 5). La Dnase I like (b79) peut probablement être considérée comme une contamination suite au traitement à la Dnase I de la purification des complexes. La protéine P-loop à domaine NTPase (b96) a une activité hydrolase, sa séquence n'est couverte qu'à 4%. La protéine à domaine ARM répété (b66) a une activité liée à l'initiation de la traduction, la séquence n'est couverte qu'à 5%. Les activités des protéines n'étant pas liées à une activité de dégradation, de transport et de transcription et le faible taux de recouvrement de séquences protéiques, ceci exclut ces protéines comme pouvant être des partenaires de LEC2. La protéine (b51) a activité de liaison ADN/ARN contenant le domaine Alba est annotée comme impliquée dans le métabolisme des ARNs. Néanmoins, le domaine Alba est aussi identifié comme impliqué dans l'architecture de la chromatine et la répression de la transcription. Ce domaine peut en particulier se dimériser et se lier à l'ADN chez l'Archae (Wardleworth *et al.* 2002). La protéine a été identifiée avec 3 peptides et recouverte à 12% de sa séquence. Malgré son activité de liaison à l'ADN, la protéine n'est identifiée qu'une seule fois au sein des différentes conditions testées, ce qui ne permet pas d'affirmer l'identification d'un partenaire potentiel de LEC2.

La recherche de nouveaux partenaires par les différentes méthodes reste donc largement insuffisamment concluante à ce stade.

### **III.3 Conclusion, discussion et perspectives**

Les expériences de capture de complexes afin d'identifier les partenaires liés à LEC2 ont été infructueuses ce qui, compte tenu de la lourdeur des expériences mises en place (clonage de 3 étiquettes, en N et C terminal, validation de l'activité, restauration phénotypique du



mutant *lec2*, extractions très lourdes de matériel végétal, purification des noyaux), reste très décevant. Quelques éléments pourraient expliquer que cette méthodologie n'ait pas donné les résultats attendus :

Les immunolocalisations de LEC2 au cours du développement révèlent une expression ciblée dans quelques cellules de l'embryon. Malgré la sélection des siliques en développement au stade où LEC2 s'exprime (3-6 jours après fécondation) et l'extraction des graines des siliques afin de diminuer le bruit de fond aspécifique, la quantité de matériel n'a probablement pas été suffisante pour identifier la protéine en spectrométrie de masse : la protéine est certainement diluée dans un extrait total, rendant le taux d'expression de LEC2 inférieur à la limite de détection de cette méthode d'identification. De plus, les protéines identifiées proviennent uniquement des purifications en immunoprécipitation, ces résultats pouvant être dus aux différences expérimentales entre les étiquettes de TAP-TAG et la GFP qui occasionnent moins de lavage. Les protéines identifiées sont, dans ce cas, plus facilement des contaminants que de vrais interactants.

Néanmoins, deux peptides de la GFP exprimée sous le contrôle du promoteur de LEC2 ont été identifiés en spectrométrie de masse sur un échantillon témoin GFP, ce qui pourrait indiquer un défaut de fixation des étiquettes sur la résine. Ce manque de fixation pourrait être dû (i) à la non accessibilité de l'étiquette dans sa conformation en fusion avec LEC2, même si LEC2 reste fonctionnelle lorsqu'elle contient une étiquette. L'accessibilité de l'étiquette ne peut pas être vérifiée que ce soit en N ou C terminal, des données de structure pourraient informer de cette possibilité, (ii) à la stabilité de la protéine étiquetée *in vivo* : la fonctionnalité de la protéine ne permet pas de déterminer le temps de demi-vie de LEC2 étiquetée *in vivo*. LEC2 très régulée, pourrait entrer très rapidement dans le processus de dégradation, limitant la quantité de protéine identifiable en spectrométrie de masse, même si la protéine devrait interagir avec des protéines liées au protéasome par exemple (pour mémoire, FUS3 a une demi-vie *in vitro* de 15 min (Lu *et al.* 2010). Dans ce cas une fixation préalable du matériel, dès la dissection, serait peut-être une possibilité pour stopper le processus.

La mécanique d'action de LEC2 *in vivo* pourrait aussi être de type « Hit and Run » déjà observée chez les bZIPs (Para *et al.* 2014). LEC2 a un Kd faible pour le promoteur de l'*OLEOSINE1* (Chapitre II.3), le mode d'action de LEC2 sur l'ADN et l'interaction de la protéine avec ses partenaires pourraient être rapides et les interactions labiles. Le faible taux d'expression, associé un mode d'action rapide, pourraient augmenter le ratio protéine libre / protéine liée à un complexe et diminuer les chances d'identifier les partenaires de LEC2. Afin d'identifier si LEC2 a une activité « Hit and run », il serait envisageable de mesurer l'affinité de la protéine avec son partenaire déjà identifié LEC1 et son autre partenaire putatif ABI3.

Pour cette hypothèse, la fixation des complexes avant purification permettrait peut-être de palier les difficultés rencontrées, mais un tel mode d'action rendrait très délicat une éventuelle purification de complexe.

La quantité de matériel reste probablement la principale limitation pour la réalisation de cette étude. Le faible taux d'expression de LEC2 (sous contrôle de son promoteur endogène) dans des stades de développement précoces avec peu de cellules exprimant la protéine, rend très difficile la capture de complexe. Pour réaliser cette étude, une méthode d'enrichissement de cellules embryonnaires exprimant LEC2 pourrait être réalisée, par exemple (i) par la sélection de cellule exprimant LEC2 provenant de calcs embryonnaires ou (ii) par l'utilisation d'un promoteur fort qui demandera une longue analyse afin d'éliminer les faux positifs. Les perspectives pour cette étude sont axées vers la compréhension de l'activité de LEC2 et sa régulation. La présence d'un domaine PEST, démontrée chez FUS3, pourrait avoir une probable implication sur le temps de demi-vie de la protéine et sa dégradation. La délétion de ce domaine chez FUS3 montre une accumulation de la protéine dans la cellule et la diminution de sa dégradation (Lu *et al.* 2010). La prédiction de la présence de deux domaines PEST dans la séquence protéique de LEC2, laisse à penser que la protéine pourrait être dégradée par un mécanisme similaire. La délétion de ces domaines permettrait d'accumuler la protéine et d'améliorer la capture de complexe lié à LEC2. De façon similaire, une meilleure connaissance des sites de régulations post-traductionnelles et la modification de ceux-ci permettraient peut-être d'obtenir une protéine plus stable dont l'accumulation pourrait être augmentée.

## IV Conclusion et discussion générale

Mon travail de thèse reposait sur l'identification, avec ou sans a priori, de partenaires de LEC2 au sein d'un complexe de régulation de la transcription. Un complexe putatif a été identifié entre les LAFL avec *proOLE1*, ce complexe protéique semble être composé de LEC1, ABI3 et LEC2 (Baud *et al.* 2016). LEC1 et LEC2 sont des facteurs de transcription maîtres dans le développement de la graine, car ils peuvent induire toute une série de processus impliqués dans la formation et le développement de la graine. Au cours de mon travail de doctorat, j'ai tenté de confirmer l'existence de ce complexe et d'en rechercher de nouveaux partenaires. Une interaction directe entre LEC1 et LEC2 a été révélée par différentes techniques *in vitro*, supportée *in vivo* par une co-localisation subcellulaire des protéines, au niveau du noyau et dans le cytoplasme (Chapitre II.3).

LEC1 et LEC2 sont des facteurs de transcription des familles NF-YB et B3 respectivement. La littérature mentionne principalement la formation d'un trimère NF-Y A/B/C pour permettre la fixation des NF-Ys sur l'ADN (Nardone *et al.* 2016). LEC1 diffère des autres NF-YB présents chez *Arabidopsis thaliana*, en effet seuls 2 NF-YBs, LEC1 et LEC1-Like, ont un changement d'acide aminé en position 86 qui confère à ces deux protéines la capacité d'interagir avec des B3 (Baud *et al.* 2016; Lee *et al.* 2003). Néanmoins, les 2 NF-YBs ont une conformation identique aux autres membres du groupe et semblent former les mêmes interactions avec les autres NF-Y (A et C) pour former le trimère (Gnesutta *et al.* 2016). LEC2, quant à lui, est un facteur de transcription à domaine B3 capable de se fixer directement sur les éléments cis de l'ADN. Il a une affinité faible sur sa cible qui augmente lors de la formation du dimère LEC1/LEC2.

### IV.1 La formation du complexe LAL

Même si l'expression simultanée de toutes les protéines augmente drastiquement l'activité transcriptionnelle (Baud *et al.* 2016), la formation du complexe LEC1, ABI3, LEC2 n'a pas été totalement démontrée. Seule la formation d'un dimère LEC1/LEC2 et leur interaction sur l'ADN ont été démontrées (double hybride, thermophorèse et gel shift). Cependant, LEC1 et ABI3 semblent réguler des gènes en commun (Pelletier *et al.* 2017). Les tentatives d'études de formation du trimère sur l'ADN n'ont pas abouti, cela étant peut-être dû aux différences d'affinité des protéines pour la sonde utilisée, la stœchiométrie des protéines et la cinétique de formation du complexe. Afin d'étudier la formation de complexe *in vitro*, la valeur d'affinité des protéines B3 pour la sonde devrait être déterminée, pour ainsi pouvoir tester la formation avec différentes stœchiométries et différentes combinaisons d'apparition. L'hypothèse d'un complexe plus gros que celui identifié et la nécessité de la présence de partenaire(s) encore non identifié(s) peut-être aussi envisagée.

## IV.2 La recherche de nouveaux partenaires des AFL

Il a été démontré que les dimères NFY B/C peuvent interagir avec d'autres facteurs de transcription (Masiero *et al.* 2002; Ben-Naim *et al.* 2006; Gnesutta *et al.* 2017). Les LECs jouent un rôle important dans l'expression de gènes lors de la maturation, notamment LEC1 par l'interaction avec d'autres facteurs de transcription tels que ABI3 ou des bZIPs (Pelletier *et al.* 2017). Concernant LEC2, hormis LEC1, aucun facteur de transcription formant un complexe n'a été identifié.

La localisation périnucléolaire de LEC1 et LEC2 a permis de confirmer une interaction probable au noyau entre les deux protéines. Celles-ci ne sont pas en totalité co-localisées et ont certainement d'autres partenaires et/ou d'autres activités indépendantes du complexe LEC1/LEC2. La littérature montre que LEC1 interagit avec NF-YC2 et bZIP67 sur des promoteurs de gènes de protéines de réserve, de même pour ABI3 avec les bZIPs 10/25/53 (Alonso *et al.* 2009; Yamamoto *et al.* 2009). ABI3, LEC1 et LEC2 formant un complexe sur *proOLE1*, les bZIPs et NF-YC2 pourraient être impliqués dans ce complexe. Les études menées ont démontré que NF-YC2 interagit avec LEC1 et pourrait avoir un effet répresseur de l'activité transcriptionnelle lorsqu'il est associé à des NF-YBs « non LEC1 », mais qu'il ne montre pas d'interaction avec LEC2. De même, les bZIPs testées ont des effets positifs seules sur le promoteur. Les hétérodimères de bZIP67/ bZIPX et 53/ bZIPX ont, eux, des effets activateurs ou répresseurs respectivement. Une étude individualisée de chaque LANL avec les dimères de bZIPs, pour identifier les interactions directes ou les implications indirectes sur le complexe et l'effet de ces facteurs de transcription sur *proOLE1*, permettrait d'affiner les connaissances de la formation du complexe. De plus, d'autres bZIPs ainsi que NF-YC pourraient être impliqués, comme les bZIPs 11 et 44 ainsi que NF-YC3 : ces facteurs ont des profils d'expression similaires à celui de LEC1. NF-YC3 interagit avec le NF-YB9 en double hybride et bZIP11 et 44 sont du même groupe que bZIP53. Dans le cadre de la recherche de nouveaux partenaires, ces facteurs pourraient être des acteurs du complexe LANL qu'il faudrait étudier.

Enfin, l'activité synergique des facteurs ABI3, LEC1 et LEC2 sur *proOLE1*, a été identifiée en système hétérologue de mousse et l'hypothèse de la présence d'une protéine endogène de *Physcomitrella patens* impliquée dans le complexe doit être envisagée. La formation d'un complexe composé de ABI3, LEC1, LEC2 pourrait en effet nécessiter la présence d'autres partenaires pour être formé. La recherche globale de partenaire(s) en système hétérologue pourrait être une autre approche.

Une étude de recherche de partenaire(s) sans a priori a été réalisée afin d'identifier des partenaires liés à LEC2 dans les différentes fonctions cellulaires (activité, mobilité, régulation post-traductionnelle). Pour cela une étude complète de la fonctionnalité des protéines

étiquetées a été réalisée et a permis d'obtenir des lignées exprimant des protéines fonctionnelles étiquetées qui complètent le mutant *lec2-1*. Malgré 3 expériences réalisées en parallèle, avec trois étiquettes et technologies différentes, l'étude n'a pas permis d'identifier de candidat. Néanmoins, cette étude révèle la difficulté d'identifier les partenaires de la protéine, des hypothèses d'accessibilité des tags, de régulation de la stabilité de la protéine et de son temps de demi-vie peuvent être émises, mais aussi sur l'activité propre de LEC2. À ce jour, aucun partenaire de LEC2 autre que LEC1 n'a été identifié. LEC2 a une expression faible mais est pourtant un facteur de transcription essentiel à la formation de l'embryon. LEC2 pourrait avoir une activité « Hit and Run » sur les séquences cibles, cette hypothèse peut être envisagée car la protéine seule a une faible affinité pour le promoteur de l'*OLEOSINE1*. Dans ce cas, son activité pourrait être très rapide et l'interaction avec ces partenaires potentiels très labile, rendant la capture de complexe plus difficile. L'accumulation des connaissances sur LEC2 par l'étude de la protéine *in vitro* permettrait de mieux caractériser son activité, les zones d'interaction protéine - protéine, les modifications post-traductionnelles régulant son activité (phosphorylation, ubiquitine ou domaine de régulation PEST), les séquences d'import et d'export nucléaire. Ces connaissances permettraient également de mieux maîtriser son expression et peut-être de concentrer la protéine dans un compartiment cellulaire afin d'identifier les partenaires en fonction de son activité ou sa régulation. Enfin, nous avons montré que des expériences de double hybride avec la partie N terminale de LEC2 sont possibles (Chapitre II.3), alors que la protéine entière auto-active fortement. Il serait sans doute intéressant de réaliser un crible double hybride avec cette protéine tronquée pour tester si de potentiels partenaires pourraient être identifiés, en mettant de côté les problèmes éventuels de stabilité *in planta*.

#### **IV.3 Localisation, expression et régulation de LEC1 et LEC2**

Les quantifications des ARN messagers des deux gènes *LEC1* et *LEC2* ont démontré qu'ils n'ont pas le même niveau d'expression (Chapitre I.2) mais qu'ils s'expriment aux mêmes stades de développement de l'embryogenèse à la maturation (Chapitre I.2.2, figure 1 p. 20). Notre étude en immunolocalisation confirme l'expression des protéines aux mêmes stades de développement, leur localisation au noyau et au cytoplasme et leur co-localisation partielle (Chapitre II.3).

Cette localisation pose la question du contrôle de la mobilité des protéines dans la cellule. L'action de drogues connues pour bloquer le transport nucléaire, inhibant l'action de certaines IMPORTINES et EXPORTINES, n'a pas eu d'effet sur la mobilité de LEC2. LEC2 pourrait utiliser des voies différentes pour l'import, par exemple la voie de l'IMPORTINE  $\beta$  et d'autres voies pour l'export nucléaire comme par exemple un co-transport avec d'autres protéines nucléaires membres, ou pas, du complexe de régulation. Une étude plus ciblée sur

les NLS et NES de LEC2 (délétion et/ou mutagenèse dirigée) permettrait d'étudier plus précisément le trafic de LEC2 en observant l'accumulation de la protéine dans un compartiment cellulaire donné.

La régulation du trafic nucléaire passe par des structures définies, par exemple l'IMPORTINE $\alpha$ , dont l'état de phosphorylation permet l'accessibilité des séquences ARM structurée en hélice, qui pourront interagir avec la séquence NLS en fonction de sa structure. LEC2 a de nombreux sites putatifs de phosphorylation, notamment dans les séquences putatives de NLS et NES et certaines situées dans le domaine B3 dont deux tyrosines (Y257, Y265) qui semblent importantes pour la structure du B3. Ces acides aminés pourraient réguler l'accessibilité des séquences NLS et NES. L'état de phosphorylation de LEC2 pourrait donc influencer son activité d'interaction protéine-protéine, ou protéine-ADN, mais aussi jouer un rôle dans la mobilité intracellulaire de la protéine. L'identification et l'étude des sites de phosphorylation permettraient de faciliter l'étude de l'activité, de son trafic nucléaire et aider à identifier les protéines impliquées dans le transport. L'expression de LEC2 est fortement régulée (Chapitre I.2.2, figure 2 p. 21, Berger *et al.* 2011), mais on ne sait si c'est le cas de la régulation post-traductionnelle. Les immunolocalisations de LEC2 et LEC1 montrent que les protéines sont également présentes à des stades plus tardifs du développement de l'embryon. La localisation révèle aussi un bruit de fond cytoplasmique qui pourrait être lié à une régulation des protéines orientées dans des voies de dégradation. La séquence de LEC1 présente des sites d'ubiquitination (Nardini *et al.* 2013), les autres facteurs de transcription B3 (ie ABI3 et FUS3), sont régulés par AIP2 une E3 ligase (Zhang *et al.* 2005; Duong *et al.* 2017) et seraient orientés dans des voies de dégradation liées au protéasome. L'identification des modifications post-traductionnelles permettrait d'étudier la régulation de LEC2 et le temps de demi-vie de la protéine. LEC2 contient plusieurs sites d'ubiquitination dont 3 à fort score de probabilité (K153, K154, K273). Le motif PEST riche en proline (P), acide glutamique (E), sérine (S), thréonine (T) est impliqué dans la dégradation des protéines en lien avec le protéasome (Rogers *et al.* 1986; Rechsteiner and Rogers 1996; Xing *et al.* 2010; Park *et al.* 2014). Une étude réalisée sur FUS3 a révélé la présence du domaine PEST dans la partie C terminale de la protéine et son implication dans la dégradation et la diminution du temps de demi-vie de la protéine (Lu *et al.* 2010). Deux domaines PEST contenant des sérines phosphorylables sont prédits en N et C terminal du domaine B3 de LEC2. La protéine pourrait donc être inactivée par la voie de dégradation du protéasome, par l'intermédiaire d'un domaine PEST, ou subir une inactivation partielle et une accumulation selon des états de phosphorylation spécifiques. La caractérisation des sites de modifications post-traductionnelles, les interactions avec des kinases et des phosphatases, des E3 ligases et l'observation de l'accumulation des protéines suite à l'inactivation du

protéasome après modification des domaines PEST, sont donc une piste importante pour mieux comprendre la régulation l'activité de LEC2.

#### **IV.4 Structure des protéines**

LEC1 interagit avec LEC2 par l'intermédiaire du domaine B2 présent également chez ABI3. L'évolution des 2 gènes *ABI3* et *LEC2* suggère que *LEC2* a évolué à partir d'*ABI3*. Le domaine B2 de LEC2 diffère cependant de celui d'*ABI3*, notamment par sa taille plus petite et l'absence de la séquence NLS. FUS3 possède une petite partie de domaine B2 sans similarité avec celui d'*ABI3* et n'a pas d'activité transcriptionnelle augmentée lorsqu'il est exprimé en présence de LEC1, contrairement à LEC2 et ABI3. Comme pour LEC2, LEC1 pourrait donc être impliqué dans l'interaction avec ABI3 via le domaine le B2. Les changements d'affinité d'*ABI3* pour l'ADN n'ont pas été étudiés. Ces affinités, pourraient être mesurées afin de tester si LEC1 a le même rôle d'aide à l'efficacité transcriptionnelle du B3 par une amélioration de la fixation d'*ABI3* sur l'ADN et de mesurer les différences éventuelles d'affinité de LEC1 avec LEC2 ou ABI3.

Une séquence conservée du B2 entre LEC2 et ABI3 est retrouvée chez les NF-YA, au niveau de la boucle non structurée de connexion des hélices  $\alpha 1$  et  $\alpha 2$ . Une hypothèse d'action du B2 serait que le domaine mime la boucle du NF-YA. Néanmoins, il est à noter que LEC2 est une protéine différente, composée d'un domaine B3 se liant directement à l'ADN sans nécessité de former un complexe. Chez les facteurs de transcription à domaines B3, seule la structure du domaine B3 de RAV1 a été résolue (Yamasaki *et al.* 2004) Néanmoins, le domaine B2 identifié chez LEC2 est essentiel et nécessaire à l'interaction avec LEC1 et montre une prédiction de formation de domaine coiled coil sans que sa structure soit connue. Les NF-YB « non LEC1 » ne montrent pas d'effet sur l'activité transcriptionnelle de LEC2 et l'acide aminé D86 de LEC1 est essentiel pour lui conférer une affinité plus grande pour une interaction avec l'ADN. Un changement de conformation ou des interactions ADN-protéine plus importantes, pourraient être la cause de ce changement d'affinité. La résolution de la structure de la protéine entière de LEC2 permettrait de mieux étudier l'interaction du domaine B3 avec NF-YB9, déjà modélisée. La résolution de la structure du complexe LEC1/LEC2/ADN permettrait de comprendre l'implication de LEC1 dans cette interaction B3-ADN, le rôle de D86 dans les interactions protéine-protéine et protéine-ADN et l'interaction du B3 à l'ADN en présence de LEC1. La détermination de la structure de LEC2 permettrait également d'élucider les changements conformationnels liés aux modifications post-traductionnelles notamment les phosphorylations, sur la structure du B3 et l'accessibilité des séquences NLS et NES permettant la mobilité de LEC2 dans la cellule. Compte tenu de la difficulté à produire LEC2 de façon soluble en grande quantité chez la bactérie, ces analyses devront passer par (i) d'autres moyens de production de la

protéine (levures ou cellules d'insectes par exemple), (ii) par une optimisation de la production (taille plus courte, optimisation des codons).

#### **IV.5 Les séquences promotrices**

La formation d'un complexe B3/NF-Y semble avoir une importance dans l'expression des cibles, plus particulièrement sur *proOLE1* et faiblement sur *proOLE3*, le complexe ne fonctionnant pas sur *proOLE2* (Baud *et al.* 2016). L'action des B3 pourrait être différente en fonction de la structure du promoteur : de sa composition en éléments cis, les espacements entre les boîtes, l'affinité des facteurs pour la cible, la formation du complexe protéique de composition variée, où l'action de LEC1 serait plus ou moins requise. D'autres promoteurs, qui seraient les cibles de LEC1 et des B3, pourraient être étudiés afin d'affiner l'interaction et le rôle du complexe. La recherche de complexes sans a priori aurait permis de développer ces études, la formation d'un complexe en fonction du stade ou d'un tissu ou d'une cible. Les données de Chip on chip disponibles pourront être une première source d'information pour permettre de mieux définir la structure des promoteurs des gènes régulés par les AFL.

#### **IV.6 Conclusion**

Les études préliminaires permettant de définir la formation du complexe LAFL, ont révélé la complexité de l'action des B3. Plusieurs remarques peuvent être émises sur le fonctionnement de ces facteurs de transcription.

- (i) LEC1 et LEC2 forment un dimère via le domaine B2 de LEC2 et D86 de LEC1 qui augmente l'affinité de LEC2 sur sa cible. Par essence, les facteurs de transcription ABI3 et LEC2 se fixent à l'ADN par l'intermédiaire du domaine B3, néanmoins, j'ai montré que LEC2 seul n'a pas une grande affinité pour sa cible. Les domaines B3 et particulièrement LEC2 nécessiteraient l'intervention d'autres facteurs de transcription pour augmenter leurs affinités pour leur cible et permettre une meilleure activité transcriptionnelle.
- (ii) Des différences existent entre ABI3, LEC2 et FUS3 qui n'ont pas le même domaine B2. FUS3 n'a pas la même activité que les autres B3 en présence de LEC1, une différenciation du B2 déterminerait l'activité du facteur de transcription.
- (iii) La localisation de LEC2 a permis de révéler une mobilité de la protéine dans la cellule, mais aucune des drogues testées, la leptomycine B et l'invermectin, n'agit sur cette mobilité. LEC2 n'utiliserait pas les voies classiques de l'import et l'export.
- (iv) La recherche de nouveaux partenaires met en évidence l'action non exhaustive d'autres facteurs de transcription, notamment plusieurs bZIPs et le NF-YC2. Les premiers résultats ouvrent la voie de l'étude de la régulation de l'action des LAFL et de la formation d'un complexe plus gros que celui déjà identifié.



- (v) Ces résultats suggèrent que des données structurales sont essentielles pour comprendre les régulations, les interactions entre protéines et des protéines sur l'ADN.

L'action *in planta* de LEC2 reste à ce jour mal caractérisée : LEC2 intervient dans de nombreuses voies du développement de l'embryon, mais les données d'expression restreignent son action à des stades de développement précoces avec un taux d'expression faible. LEC2 joue un rôle dans l'accumulation des réserves avec ABI3 et LEC1, il pourrait jouer un rôle d'initiateur de la transcription dans la synthèse des gènes de réserves. La surexpression de LEC2 induit l'embryogenèse somatique et contrôle quelques gènes hormonaux ou d'autres facteurs de transcription importants, bien que LEC2 semble fortement régulé transcriptionnellement et post-traductionnellement (To *et al.* 2006; Berger *et al.* 2011). LEC2 peut interagir directement sur l'ADN en complexe avec LEC1 et forme un domaine Histone like motif avec NF-YC. Dans l'hypothèse d'une implication de LEC2 dans des complexes protéiques variés, son activité pourrait être intégrée dans différentes voies d'expression génique.

Les premiers résultats biochimiques orientés sur l'étude des protéines dans le cadre de ce travail confirment la diversité et la complexité d'action des LAFL. Toutes les hypothèses formées dans cette étude sont aussi des perspectives d'études à développer pour comprendre le fonctionnement de ces facteurs de transcription maîtres du développement de la graine.

## V Matériels et Méthodes

### V.1 Matériels

#### V.1.1 Le matériel végétal

Les travaux ont été réalisés sur les écotypes d'*Arabidopsis thaliana* Columbia (Col-0) et Wassilewskija (Ws) de la Station de Génétique et d'Amélioration des Plantes, INRA Versailles.

Les mutants *leafy cotyledon* utilisés sont :

*lec2-1* N3868, obtenu dans le fond génétique Ws-0, est un mutant d'insertion ADN-T, qui a occasionné une délétion en amont et aval du codon d'initiation: dans le promoteur de *LEC2* à -3048pb et dans la séquence de *LEC2* à 895 pb (Stone *et al* 2001), obtenu par David Meinke de l'université de l'Oklahoma (Meinke *et al* 1994).

*lec1-10*, N631219 et *lec1-11* N500450 obtenu dans le fond Col-0 sont des mutants d'insertion ADN-T. Les insertions sont toutes les deux positionnées dans l'intron du gène. Ces 2 lignées perte de fonction proviennent de l'European Arabidopsis Stock Centre (NASC).

#### V.1.2 Les bactéries

Les bactéries utilisées pour ces travaux sont :

*Escherichia Coli* souche Db3.1 : F- *gyrA462 endA1 Δ(sr1-recA) mcrB mrr hsdS20(rB-, mB-) supE44 ara-14 galK2 lacY1 proA2 rpsL20(SmR) xyl-5 λ- leu mtl1* (Thermo Fisher Scientific). Cette souche de bactéries électrocompétentes et insensibles au CcdB, est utilisée pour amplifier les vecteurs Gateway donneur ou de destination non recombinés contenant la séquence codante de la protéine CcdB. Cette protéine est cytotoxique pour les cellules et inhibe l'action de l'ADN gyrase. Ces bactéries électrocompétentes transformées seront résistantes au chloramphénicol pour la sélection des vecteurs non recombinés et à l'antibiotique du vecteur utilisé.

*Escherichia Coli* souche DH10B : F- *mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara leu) 7697 galU galK rpsL nupG λ-* (Thermo Fisher Scientific). Ce sont des bactéries électrocompétentes utilisées pour amplifier le produit de recombinaison Gateway. La sensibilité de cette souche à la protéine CcdB, permet une sélection positive des bactéries transformées avec un produit de recombinaison uniquement résistant à l'antibiotique du vecteur utilisé.

*Escherichia Coli* souche Mach1™ : F- *Φ80lacZΔM15 ΔlacX74 hsdR(rK-, mK+) ΔrecA1398 endA1 tonA* de Thermo Fisher Scientific.

Ce sont des bactéries chimiquement compétentes, utilisées ici essentiellement pour les recombinaisons doubles. Cette souche a la particularité d'avoir un temps de doublement plus rapide que les souches communément utilisées ce qui la rend plus efficace pour les recombinaisons doubles dont les événements sont plus difficiles à sélectionner.

*Escherichia coli* souche DE3 Rosetta [DE3, F2 ompT hsdS (r 2 m2) gal dcm (DE3) BBB pRARE (CamR)] (Novagen).

Ce sont des bactéries dérivées de la souche BL21 ; celles-ci contiennent les codons rares non présents chez les procaryotes. Elles améliorent ainsi l'expression de protéines eucaryotes.

*Agrobacterium tumefaciens* souche C58C1pMP90 58C1 (pMP90) (Koncz et al. 1986). Il s'agit de la souche utilisée pour transformer les plantes par agroinfiltration. Cette souche contient le gène de résistance à la rifampicine, et l'ADN-T du plasmide Ti a été substitué par le gène conférant la résistance à la gentamycine.

### V.1.3 Les levures

*Saccharomyces cerevisiae* souche AH109 (MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2 :: GAL1<sub>UAS</sub>-GAL1<sub>TATA</sub>-HIS3, GAL2<sub>UAS</sub>-GAL2<sub>TATA</sub>-ADE2, URA3 :: MEL1<sub>UAS</sub>-MEL1<sub>TATA</sub>-lacZ) James et al., 1996; A. Holtz, unpublished.

Cette souche disponible chez Clontech est utilisée pour tester les interactions par la technique de double hybride.

### V.1.4 Les vecteurs

(Annexe pour cartes)

**pDONR207** est un vecteur donneur Gateway (Invitrogen). Il contient les cassettes de recombinaison (attP1 - attP2), ainsi que le gène de résistance au chloramphénicol et la séquence CcdB. Ce vecteur permet de réaliser des réactions de recombinaison Gateway simples. Il peut être sélectionné grâce à sa résistance à la gentamycine

Le vecteur **pDONR221** est aussi un vecteur donneur de la réaction Gateway (Invitrogen), de même que pDONR207, il contient le CcdB et le gène de résistance au chloramphénicol. Toutefois, il est constitué des cassettes de recombinaison double (attP1 - attP5) ou (attP5R - attP2), qui permettent de fusionner deux séquences d'intérêt lors de la réaction LR double. pDONR221 recombiné est sélectionné grâce à sa résistance à la kanamycine.

**pETtrx-1a** est un vecteur qui permet l'expression des protéines étiquetées en N terminal de 6 histidines et d'une étiquette thiorédoxine qui permet d'améliorer la solubilisation des protéines recombinantes. L'étiquette peut être éliminée par la présence d'un site de clivage

TEV (tobacco etch virus protease) situé en C terminal de la thiorédoxine. Ce vecteur est composé de la séquence du gène *LACI* répresseur de l'opéron lactose qui permet une expression des protéines uniquement en présence d'IPTG. Le vecteur est résistant à la kanamycine (Baud *et al* 2012).

**pETG10A et pETG20A** sont des vecteurs d'expression Gateway pour *E. coli*. Après recombinaison, la protéine est étiquetée en N terminal respectivement d'un tag histidine et un tag thiorédoxine A avec 6 histidines. Ils possèdent la séquence du gène *LACI*, répresseur de l'opéron lactose. Les protéines seront exprimées par induction à l'IPTG. Les deux vecteurs sont sélectionnés grâce à la résistance à l'ampicilline (A. Geerlof).

**pDEST17** (Invitrogen) est un vecteur de destination Gateway pour *E. coli*. Après recombinaison, la protéine est étiquetée avec 6 histidines en N terminal. Ce vecteur permet l'expression des protéines par induction à l'IPTG, il contient le gène de résistance à l'ampicilline.

Tous les vecteurs utilisés pour l'expression des protéines sont dotés d'une étiquette histidine nécessaire pour une purification d'affinité grâce à la résine d'agarose d'acide de nickel-nitrilotriacétique (Ni-NTA) (Qiagen). De plus, l'étiquette permet de vérifier la présence de la protéine par western blot à l'aide d'un anticorps dirigé contre les histidines (anti histidine monoclonal de chez ThermoFisher Scientific étiqueté HRP).

Le vecteur **pGWB1** est un vecteur binaire contenant les cassettes de recombinaison Gateway attR1 – attR2 et a une double résistance à la kanamycine et l'hygromycine (Nakagawa *et al* 2007). Le vecteur pGWB1 ne contient ni promoteur ni séquence codante et permet l'insertion des séquences d'intérêt sous le contrôle du promoteur de son choix.

**pGWB2** est un vecteur binaire qui possède le promoteur p35S CaMV et les cassettes de recombinaison Gateway attR1 - attR2 et a une double résistance à la kanamycine et l'hygromycine (Nakagawa *et al* 2007). Il a été modifié pour remplacer le promoteur p35SCaMV par le promoteur endogène de *LEC2* aux sites *HindIII* et *XbaI*.

**pBS Tpp-A** ou vecteur « A » est le vecteur d'expression des gènes d'intérêt par expression transitoire chez la mousse *Physcomitrella patens*. Ce vecteur, basé sur le squelette pBluescript, a été modifié pour contenir les cassettes de recombinaison Gateway, comprenant les sites attR1 et attR2, le gène de résistance au chloramphénicol et la séquence de CCDB (Curtis & Grossniklaus, 2003). Construit à partir du vecteur pCOR104-CaMVter (Proust *et al.* 2011) le vecteur permet l'expression des gènes d'intérêts sous le contrôle du promoteur *ACTINE* de riz et est résistant à l'ampicilline (Baud *et al* 2012).

**pBS Tpp-B** ou vecteur « B » est un vecteur construit à partir du vecteur pBluescriptIIISK(+). Il contient les cassettes de recombinaison attR1 et attR2, la résistance au chloramphénicol et

la séquence codant de CcdB. Ce vecteur permet l'étude de séquence promotrice dans les expériences d'expression transitoire chez la mouche. La séquence promotrice d'intérêt insérée par réaction Gateway LR contrôle l'expression de la GFP dont la séquence est située en aval du promoteur inséré. Le vecteur recombiné est sélectionné grâce à sa résistance à l'ampicilline (Baud *et al* 2012).

Les vecteurs **pDEST22** et **pDEST32** sont des vecteurs utilisés pour la technique du double hybride.

**pDEST22** est un vecteur Gateway résistant à l'ampicilline et permet l'autotrophie pour le tryptophane lors de la sélection des levures. La recombinaison LR du vecteur pDEST22 permet de positionner le domaine d'activation « AD » en N terminal de la séquence du gène d'intérêt (ThermoFisher Scientific).

**pDEST32** est un vecteur Gateway résistant à l'ampicilline et permet l'autotrophie pour la leucine lors de la sélection des levures. La recombinaison LR du vecteur pDEST32 permet de positionner le domaine liaison « BD » en N terminal de la séquence du gène d'intérêt (ThermoFisher Scientific).

## **V.2 Méthodes**

### **V.2.1 Biologie moléculaire**

#### **V.2.1.1 Réaction de recombinaison Gateway BP simple**

La réaction Gateway de BP clonase a été simplifiée et permet l'insertion de fragment dans le vecteur donneur pDONR207 avec le plus petit volume possible (5 µL).

1 µL de chaque élément de la réaction (le produit PCR, le vecteur donneur, le tampon, l'enzyme BP clonase, l'eau) sont mélangés puis la réaction est incubée à 28°C pendant 3h. Des bactéries DH10b sont transformées par électroporation avec le produit de la réaction. Les clones positifs sont vérifiés par PCR avec des amorces spécifiques et séquencés après amplification et purification des plasmides avec le kit « nucleospin plasmid » de Macherey Nagel.

#### **V.2.1.2 Réaction de recombinaison Gateway BP pour recombinaison LR double**

La réaction Gateway de BP a été réalisée selon le protocole commercial ThermoFisher. Brièvement, sont ajoutés : 5 µL de produit PCR purifié sur colonne avec le kit PCR clean UP (Macherey-Nagel), 1 µL de vecteur Donneur pDONR221 P1-P5R ou P5-P2 complété à 8 µL avec le tampon TE (10 mM tris, 1mM EDTA pH 8,0), 2µL de BP Clonase<sup>®</sup> II. La réaction est incubée 2h à 25°C. Des bactéries DH10b sont transformées par électroporation avec le produit de la réaction.

Les clones sont vérifiés par PCR selon les amorces spécifiques et séquencés avec les amorces M13 sens / anti-sens situées en amont et aval des séquences de recombinaison attB1 et attB2. (Annexe 7 pour les séquences de amorces utilisées).

#### **V.2.1.3 Réaction de recombinaison Gateway BP inversée**

La réaction de BP clonase inversée permet d'insérer des fragments recombinés LR dans un vecteur de destination dans un vecteur donneur. Ainsi les éléments associés par la réaction LR Gateway double ont été insérés dans le vecteur pDONR207 attB1 attB2, les fragments associés en réaction BP double restent fusionnés.

La réaction est obtenue avec 1 µL de chaque élément (le vecteur de destination, le vecteur donneur, l'eau, le tampon, l'enzyme BP clonase), incubés à 28°C pendant 3h. Les bactéries DH10b sont électroporées pour sélectionner les clones positifs.

#### **V.2.1.4 Réaction de recombinaison Gateway LR simple**

De même que pour la réaction de BP, la réaction Gateway LR a été simplifiée et permet l'insertion de fragment dans le vecteur de destination avec le plus petit volume possible (5µl).

1 µL de chaque élément de la réaction (le vecteur donneur, le vecteur de destination, le tampon, l'enzyme LR clonase, l'eau) sont mélangés puis, incubés à 28°C pendant 3h, puis des bactéries DH10b sont transformées avec le produit de la réaction. Les clones positifs sont vérifiés par PCR avec des amorces spécifiques (annexe 7), séquencés après amplification et les plasmides purifiés avec le kit « nucleospin plasmid » de Macherey Nagel.

#### **V.2.1.5 Réaction de recombinaison Gateway LR double**

Afin d'insérer les étiquettes en 5' et 3' de la séquence codante de *LEC2* dans les vecteurs d'expression (attR1, attR2), une réaction de Gateway double est réalisée avec les vecteurs pDONR221 recombinés.

La réaction LR est réalisée avec différentes quantités de vecteurs donneur pDONR221 recombinés, en fonction de la taille de l'insert, soit 160ng de *LEC2* associé à 120ng pour les étiquettes (GFP ou GS), 1 µL de vecteur de destination, complétée à 8 µL de TE pH 8,0 puis 2 µL de mix d'enzyme LR Clonase<sup>®</sup> II Plus. La réaction est incubée à 25°C sur la nuit dans un bain à sec thermostaté. Des bactéries One Shot<sup>®</sup> Mach1<sup>™</sup> T1<sup>®</sup> Thermocompétentes sont transformées par chocs thermiques (30 minutes dans la glace, puis 30 secondes à 42°C et 2 min dans la glace) avec tout le produit de la réaction.

Pour la construction des témoins, la GFP est étiquetée avec GS en positions N et C terminales, la réaction de Gateway double est réalisée avec les mêmes quantités de vecteurs donneurs recombinés soit 120ng. Le reste de la réaction est identique.

### **V.2.1.6 Réaction TOPO**

Cette réaction permet l'insertion d'un fragment à bout franc (issu d'une PCR) dans un vecteur grâce à un vecteur déjà linéarisé et portant une topo isomérase à ses extrémités

Selon le kit commercial « topo cloning » de Thermofisher scientific, à 4 µL de produit PCR purifié sur colonne avec le kit PCR clean UP Macherey-Nagel et élué dans l'eau sont ajoutés 1 µL vecteur TOPO et 1 µL de tampon. La réaction est incubée 5 min à température ambiante. Les bactéries DH10b sont transformées avec le produit de la réaction pour sélectionner les clones positifs.

### **V.2.1.7 Extraction d'ADN génomique**

Les feuilles de jeunes plantules sont prélevées puis directement congelées dans l'azote liquide et stockées à - 80°C avant extraction. Les échantillons sont broyés à froid dans l'azote liquide. Les broyats sont repris dans 400µL de tampon d'extraction (200 mM tris pH 7,5, 150 mM NaCl, 25 mM EDTA, 0,5% SDS) puis mélangés vigoureusement pendant 5 secondes. Après une centrifugation de 1 minute à 13000rpm à température ambiante, 300 µL de surnageant sont prélevés et mélangés à de l'isopropanol 100% [V/V]. Après une incubation de 2 minutes à température ambiante, les acides nucléiques sont précipités par centrifugation pendant 5 minutes à 13000rpm à température ambiante. Le surnageant est éliminé et le culot d'ADN est séché à l'air puis repris dans 50 µL d'eau.

## **V.2.2 Biologie végétale**

### **V.2.2.1 Stérilisation et la culture du matériel végétal**

La lignée mutante *lec2-1* homozygote est sensible à la dessiccation et nécessite une stérilisation des siliques immatures. Après stérilisation des siliques immatures non sèches, les graines sont récupérées et semées. La méthode de repiquage a été uniformisée à toutes les lignées mutantes *lecs* homozygotes.

Les graines et les siliques sont stérilisées selon le même protocole de stérilisation. Une solution mère de Bayrochlore est réalisée à partir d'une pastille dissoute dans 40 mL d'eau stérile. La solution de stérilisation est réalisée avec 10% de solution mère dans l'éthanol 96%.

Le matériel végétal est stérilisé 7 minutes avec 1 mL de solution de stérilisation, puis lavé avec 1 ml d'éthanol absolu. Afin d'éliminer le reste d'éthanol contenu dans le tube, le matériel végétal est séché sur papier whatman stérile en conditions stériles (hotte à flux laminaire) puis les graines sont semées sur le milieu Murashige et Skoog et placées en culture sous conditions contrôlées, (jours longs : 16h jour, 8h nuit à 25°C) jusqu'à obtention des plantules. Celles-ci sont ensuite transférées en serre en jours longs : 16h jour à 19°C, 8h nuit 16°C.

### **V.2.2.2 Transformation des plantes d'*Arabidopsis thaliana* par *Agrobacterium tumefaciens***

La souche de bactéries d'*Agrobacterium tumefaciens* C58C1pMP90 a été transformée par électroporation et sélectionnée sur le milieu LB comprenant les antibiotiques rifampicine gentamycine et l'antibiotique porté par le plasmide porteur de l'ADN-T (en général kanamycine) à 50 µg/mL.

Une préculture en milieu LB liquide de 15 ml avec antibiotiques est réalisée à partir d'une colonie isolée et incubée 24h à 28°C, puis 500 mL de milieu LB avec antibiotiques sont inoculés avec la préculture et incubés jusqu'à saturation à 28°C sous agitation. Les bactéries sont ensuite culottées par centrifugation pendant 15 minutes, 6500 g, à température ambiante, puis resuspendues dans 250 ml de tampon d'infiltration (5% saccharose, 0,04% SILWET). Les plantes des lignées Col0 et *lec2-1* ont été transformées par agro-infiltration des boutons floraux (Bechtold *et al.* 1993).

### **V.2.2.3 Sélection des transformants**

Les transformants primaires sont sélectionnés sur du milieu Murashige et Skoog contenant de l'hygromycine (50 µg.mL<sup>-1</sup>). Les lignées homozygotes étudiées sont sélectionnées par résistance à l'hygromycine et génotypage dans le promoteur, l'exon ou l'intron de *LEC2* (Annexe 7). Pour chaque construction, plusieurs lignées transgéniques indépendantes sont analysées.

## **V.2.3 Analyse standard des protéines**

### **V.2.3.1 Gel SDS-PAGE**

Les protéines purifiées sont analysées sur gel SDS-PAGE selon Laemmli 1970. Le gel SDS-PAGE (gel de concentration à 6% d'acrylamide, gel de séparation à 12% d'acrylamide) est chargé avec l'échantillon repris dans du tampon Laemmli 1X final (50 mM tris HCl pH 6,8, 2%SDS, 10% glycérol, 0,1% BBP, 25 mM DTT). L'électrophorèse est conduite à 110V constant dans un tampon de migration 1X (25 mM tris, 192 mM glycine, 0,05% SDS). La masse moléculaire des protéines est estimée grâce au marqueur de poids moléculaire MARK 12™ d'Invitrogen. Le gel est ensuite coloré à l'aide d'une solution de coloration au bleu de coomassie R250 (0,05% bleu de coomassie R250, 50% éthanol, 10% acide acétique), révélé avec la solution de décoloration (40% éthanol, 10% acide acétique) et conservé dans l'eau.

### **V.2.3.2 Western blot**

Les protéines séparées par SDS-PAGE sont transférées sur une membrane PVDF (GE Healthcare), avec un tampon de transfert (25 mM tris, 190 mM glycine, 0,02% SDS, 20% éthanol) à l'aide d'un système de transfert Transblot SD Semi sec Biorad. Le transfert est



conduit à 15V constant pendant 1h. La membrane est ensuite saturée avec un tampon de saturation TBS-T-lait (20 mM tris, 150 mM NaCl pH 7,40, 5% tween20, 5% lait), 1h à température ambiante, puis incubée avec l'anticorps primaire dilué dans le TBS-T-lait, 3h à température ambiante. Après 3 lavages de 10 minutes de la membrane dans le TBS-T, celle-ci est incubée avec l'anticorps secondaire dilué dans le TBS-T-lait. Des lavages de 10 minutes successifs dans le TBS-T puis dans le TBS, finalisent les étapes d'incubations. La membrane est révélée par chimiluminescence selon les instructions du kit de révélation chimiluminescent (Thermo Scientific) avec le système d'imagerie Fujifilm LAS-4000. Un marqueur de poids moléculaire des protéines MagicMark™ XP Western d'Invitrogen a été utilisé pour visualiser les bandes standard.

## **V.2.4 Constructions lignées LEC2 étiquetés**

### **V.2.4.1 Construction de pGWB2 proLEC2-1Kb**

Afin d'exprimer chez la plante les protéines étiquetées sous le contrôle du promoteur endogène, le vecteur pGWB2 a été modifié. Le promoteur p35S présent initialement dans le vecteur a été remplacé par le promoteur de *LEC2* d'une longueur de -1 kb (à partir du codon d'initiation de la traduction de *LEC2*) par digestion / ligation enzymatique.

Les sites des enzymes de restriction *HindIII* et *XbaI* ont été placés aux extrémités 5' et 3' du fragment du promoteur *LEC2* amplifié par PCR (annexe 7) puis inséré dans le vecteur TOPO Blunt (Invitrogen). Le produit PCR et le vecteur pGWB2 ont été digérés successivement par les 2 enzymes puis purifiés sur gel avec le Kit Wizard SV clean UP system de Promega.

La ligation est réalisée selon le ratio produit PCR / vecteur de 9/1 avec 800 unités de T4 ligase (Biolabs) à 16°C, sur la nuit. Des bactéries DB3.1 ont été transformées avec le produit de ligation. Les colonies obtenues ont été vérifiées par PCR avec les amorces situées dans la séquence du promoteur de *LEC2* (annexe 7).

### **V.2.4.2 Constructions de LEC2 et GFP témoin étiquetés HS en N et C terminal**

Dans le cas de *LEC2* étiquetée avec HS en N terminal et SH en C terminal et des témoins associés HS-GFP et GFP-SH, l'étiquette HS inférieure à 200pb en taille ne pouvait pas être recombinée seule dans un vecteur donneur. Les séquences ont été synthétisées à façon par Geneart de la société ThermoFisher Scientific et encadrées par les cassettes de recombinaison attB1 et attB2 (annexe 6) permettant ainsi d'obtenir les différents vecteurs par recombinaison Gateway simple dans le vecteur donneur pDONR207.

Dans le cas de *LEC2*, la séquence du promoteur de *LEC2* d'une longueur de -550 pb a été inclus dans le produit de synthèse (Annexe 6). *HS:LEC2* et *LEC2:SH* sans promoteur dans pDONR207 ont été obtenus après PCR et recombinaison BP simple, puis recombinés LR dans le vecteur pBS Tpp-A contenant le promoteur ACTINE de riz utilisé pour l'expression

transitoire chez la mousse. Le produit de synthèse a permis d'insérer directement le fragment dans le vecteur pGWB1 pour transformer les plantes.

De même, les témoins GFP, étiquetés HS en N et C terminal (Annexe 6), sont recombinaison BP dans le vecteur pDONR207 à partir du produit de synthèse, puis insérés dans le vecteur pBS Tpp-A par recombinaison LR.

#### **V.2.4.3 Constructions de LEC2 et témoin étiquetés GFP et GS en N et C terminal**

Les séquences de la GFP, des étiquettes GS, SG et de LEC2 ont été amplifiées par PCR avec la polymérase Phusion (ThermoFisher Scientific) à activité de relecture. Les oligonucléotides ont été dessinés afin de positionner en 5' et 3' les cassettes de recombinaison doubles selon les couples suivants : (attB1 attB5r), (attB5, attB2). Cette configuration permet le positionnement des étiquettes GFP et GS en N terminal et C terminal de LEC2.

Dans le cas de LEC2 étiquetée avec la GFP, GS en N et C terminal, ainsi que les témoins GFP étiquetés GS en N et C terminal, les constructions ont été obtenues par réaction BP et LR par Gateway double, dans le vecteur A. Après une réaction de BP inversée permettant d'obtenir l'ensemble des constructions LEC2 étiquetés dans le vecteur donneur pDONR207, une réaction LR simple a permis de recombinaison le vecteur pGWB2 proLEC2 (-1kb) pour l'expression des LEC2 étiquetées chez la plante.

#### **V.2.5 Constructions des vecteurs d'expressions pour l'étude des bZIPs**

Les séquences codantes des bZIPs ont été synthétisées par Genart de la société ThermoFisher Scientific, encadrées par les cassettes de recombinaison simple (attB1, attB2).

Les séquences sont insérées dans le vecteur donneur pDONR 207 et les clones positifs vérifiés par PCR et séquençage (annexe 7). Les vecteurs de destination pour les expériences d'expression transitoire chez la mousse et de double hybride sont recombinaison par une réaction LR simple.

#### **V.2.6 Expression et purification des protéines**

##### **V.2.6.1 Étude des interactions par EMSA**

Annexe 1A : article Baud *et al* 2016.

Les séquences codantes de LEC1 et LEC2 sont insérées dans le vecteur d'expression pETtrx-1a. Les ADNc d'ABI3 et de NF-YC2, insérés dans le vecteur pDONR207, sont directement recombinaison LR dans pDest17. Les protéines sont produites dans *E. coli* Rosetta et induites avec 0,5 mM d'isopropyl-bD-thiogalactopyranoside (IPTG) respectivement à 28°C pendant 3 h, 17°C pendant 3 h et 17°C pendant une nuit. Les culots bactériens sont soumis à une sonication dans un tampon de lyse (20 mM Tris, 500 mM NaCl, 50 mM NaH<sub>2</sub>P0<sub>4</sub>, à

10% glycérol [v / v] et 5 mM imidazole, pH 8). Les lysats sont centrifugés à 21000 g pendant 45 min à 4°C. Les protéines solubles sont purifiées sur la résine d'agarose d'acide de nickel-nitrilotriacétique (Ni-NTA) (Qiagen) dans le tampon de lyse avec des concentrations croissantes d'imidazole (5-70 mM pour le lavage et 300 mM pour l'élution). Les protéines purifiées sont ensuite dialysées dans un tampon (0,05 M NaH<sub>2</sub>PO<sub>4</sub>, 0,15 M NaCl, 20% glycérol [v / v], pH 7,2) à 4°C, concentrées sur une cellule Amicon 10000 MWCO (Millipore) et quantifiées par la méthode de Bradford (Bradford, 1976). Après purification, les protéines sont stockées à -20°C.

#### **V.2.6.2 Production des anticorps spécifiques**

Les anticorps ont été produits à partir de corps d'inclusions de protéines recombinantes exprimées chez *E. coli*. Les bactéries (DE3) Rosetta sont transformées avec le vecteur pETG-10A contenant les séquences codantes de *LEC1* et *LEC2* et sont induites avec de l'IPTG à 0,5 mmol.L<sup>-1</sup> à 37°C pendant 3h. Les bactéries ont été récupérées par centrifugation 15 min, 5000g à 4°C, le culot est repris dans un tampon de lyse PBS 1X - 1% triton (137 mM NaCl, 2,7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1,8 mM KH<sub>2</sub>PO<sub>4</sub>, 1% triton). La solution bactérienne est lysée au sonicateur à 60% de puissance sur 1 minute par étapes de 10 secondes de sonication et de 10 secondes de repos dans la glace. Après centrifugation du lysat 45 min, 15000g à 4°C, le culot contenant les corps d'inclusions est lavé successivement avec 3 tampons (i) PBS, 2% triton, (ii) PBS 1X, 1 mol.L<sup>-1</sup> NaCl, 1 mol.L<sup>-1</sup> urée (iii) PBS 1X.

Les corps d'inclusion récupérés sont repris dans 1 mL de PBS 1X, puis dosés par la méthode de Bradford avec une gamme de BSA pour standard (Bradford, 1976).

Les corps d'inclusions sont analysés sur gel SDS-PAGE et western blot. La bande majoritaire est découpée et identifiée en spectrométrie de masse LC/MS-MS après digestion trypsique sur la plateforme Pappso, permettant ainsi de valider la présence majoritaire des protéines d'intérêt couvertes à plus de 50%.

#### **V.2.6.3 Étude des interactions par thermophorèse**

*LEC2* et *LEC1* sont produits à partir du vecteur pET<sub>trx</sub>-1a (Baud *et al* 2016). Les bactéries d'*E. coli* DE3 Rosetta transformées sont induites avec 0,5 mmol.L<sup>-1</sup> d'IPTG à 37°C pendant 3h. La purification est adaptée du protocole de Choudhury *et al* (2009). Le culot bactérien récupéré par centrifugation pendant 15min 5000g à 4°C est repris dans le tampon de solubilisation des corps d'inclusions (50 mM tris HCL pH8, 300 mM NaCl, 8 M urée, 5 mM imidazole, cocktail d'inhibiteurs de protéases complete Ultra mini EDTA free Roche). Après lyse des bactéries au sonicateur (10 fois 30 secondes à 60% de puissance), le lysat est incubé sur la nuit à 4°C sous agitation lente puis centrifugé 45 min 20000g à 4°C. Les protéines étiquetées histidines récupérées dans le surnageant sont incubées 2h avec la

résine d'agarose d'acide de nickel-nitrilotriacétique (Ni-NTA) (Qiagen), préalablement équilibrée avec le tampon de solubilisation. La purification est réalisée à 4°C, la résine est lavée par 3 volumes de colonne de tampon de lavage contenant des concentrations croissantes d'imidazole (50 mM tris-HCL pH8, 300 mM NaCl, 8M urée, 20 mM - 50 mM imidazole). La protéine est éluée avec 5 volumes de colonne du tampon de lavage contenant 250 mM d'imidazole.

L'éluat est immédiatement dialysé avec 3 bains de tampon PBS 1X, puis concentré avec des unités de filtration Amicon Ultra 10000 MWCO (Millipore).

## **V.2.7 Fabrication des outils spécifique de LEC1 et LEC2**

### ***V.2.7.1 Production des anticorps***

La production d'anticorps a été effectuée sur le plateau technique de production d'anticorps de l'unité Biopolymères Interaction Assemblages (BIA) du département CEPIA à l'INRA de Nantes. Les expérimentations animales ont été réalisées dans les installations de l'INRA, autorisées par le Service vétérinaire local (autorisation n° 44502). Deux lapins « New Zealand » ont été immunisés pour LEC1 et huit rats « Wistar » ont été immunisés pour LEC2 par injection intrapéritonéale répétée avec 300 µg de LEC1 et 100 µg de LEC2 émulsifié dans l'adjuvant de Freund (complet pour l'injection initiale, incomplet pour les autres). Les immunisations ont été répétées quatre fois à des intervalles de 2 semaines. Une semaine après la dernière injection, les animaux ont été saignés par ponction cardiaque sous anesthésie générale. Les sérums ont été préparés par centrifugation (2000g - 20 min) et conservés à -20°C.

### ***V.2.7.2 Purification des anticorps***

Pour augmenter leurs spécificités, les anticorps obtenus sont purifiés par immuno-affinité. Une quantité de 25 µg de protéines recombinantes nécessaires à l'immunisation sont déposées sur le gel SDS-PAGE 8%. Après transfert des protéines sur membrane PVDF et la coloration de celle-ci au rouge ponceau, les bandes colorées correspondant à la protéine recombinante sont découpées et saturées avec le tampon de saturation TBS-T-lait (20 mM tris, 150 mM NaCl, pH 7,4, 0,5% tween, 5% lait) 1h à 4°C. Le sérum est ensuite incubé avec les membranes à 4°C sur la nuit.

Après 4 lavages des membranes de 15 minutes avec du tampon TBS-T (20 mM tris, 150 mM NaCl, pH 7,4, 0,5% tween), les anticorps fixés sont élués plusieurs fois avec le tampon d'éluion (500 mM glycine, 500 mM NaCl, 0,1% tween 20, 1% BSA, pH 3,0). Les éluats sont immédiatement neutralisés avec une solution 1mol.L<sup>-1</sup> tris base.

L'anticorps purifié est concentré sur des unités de filtration Amicon Ultra 10000 MWCO (Millipore), puis stocké à -20°C.

## V.2.8 Observation et localisation

### V.2.8.1 Immunolocalisation et la co-immunolocalisation sur matériel fixé

#### V.2.8.1.1 Préparation des échantillons

Les siliques en développement sont prélevées et coupées à une longueur d'environ 0,5 cm, puis incubées 2h à 4°C dans la solution de fixation : (MSTB + formaldéhyde [50 mM PIPES, 5 mM MgSO<sub>4</sub>, 5 mM EGTA pH 6,9, 4% formaldéhyde]), puis lavées 30 minutes avec le tampon MTSB 0,5X (25 mM PIPES, 2,5 mM MgSO<sub>4</sub>, 2,5 mM EGTA pH 6,9) et 15 minutes dans du PBS 1X (137 mM NaCl, 2,7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1,8 mM KH<sub>2</sub>PO<sub>4</sub>).

Les échantillons sont ensuite déshydratés à l'aide de bains successifs de 30 minutes d'éthanol allant de 30 à 97% dans du PBS 1X et colorés 10 minutes avec du bleu de toluidine 0,01% dans l'éthanol à 40°C.

L'inclusion des échantillons est réalisée en 3 étapes à 40°C, (i) un bain de 10 minutes dans l'éthanol 97%, (ii) un bain d'éthanol 97% et de cire avec une interface entre les deux solutions sur la nuit, (iii) une fois les échantillons imbibés de cire, une dernière incubation de 2 heures dans la cire. Les siliques incluses sont ensuite positionnées dans des moules avec des pincettes chauffantes.

#### V.2.8.1.2 Coupe et immunolocalisation

Les échantillons inclus dans la cire sont coupés à l'aide d'un microtome automatique à rétraction LEICA sur une épaisseur de 8 µm et les bandes de cire ainsi obtenues sont positionnées sur des lames traitées à la polysine, séchées à l'air et protégées de la poussière. Afin d'éliminer la cire et de réhydrater les échantillons, les lames sont plongées dans des bains successifs de solutions d'éthanol décroissant allant de 100% à 30% dans du PBS 1X pendant 15 puis 10 minutes, puis traitées au citrate de sodium 1 mol.L<sup>-1</sup> pH 6,0 jusqu'à ébullition au micro-onde et, enfin, rincées 2 fois au PBS 1X.

Avant l'incubation avec l'anticorps primaire, les lames sont placées 1h dans une solution de saturation (1% BSA, 1X PBS), puis incubées sur la nuit à 4°C dans une atmosphère humide avec l'anticorps primaire spécifique purifié (Anti-LEC1 produit chez le lapin, Anti-LEC2 produit chez le rat), dilué au 1/500<sup>ème</sup> dans la solution de saturation.

Après 4 lavages successifs de 3 minutes dans du PBST (1X PBS, 0,1% Tween 20) ; les lames sont incubées à 37°C pendant 3h dans une atmosphère humide avec le second anticorps marqué d'un Alexa dilué au 1/100<sup>ème</sup> dans la solution de saturation. Les anticorps secondaires utilisés sont des : Anti-rat Alexa 488, Anti-lapin Alexa 488, Anti-lapin Alexa 647, Anti-souris Alexa 488 produit chez la chèvre par (Life Technologies).

Les lames sont lavées 2 fois 10 minutes dans une solution de PBS 1X, montées en présence d'un milieu de montage pour fluorescence (vectashiel H1000 de Vector Laboratories-INC complété avec 1 µg/mL de DAPI), puis scellées avec du vernis. Les lames sont stockées à 4°C jusqu'à l'observation en microscopie confocale.

#### **V.2.8.2 Localisation de GFP:LEC2 in vivo**

Les siliques en développement des lignées *GFP:LEC2* sont prélevées à des stades équivalents allant de 3 à 8 DAP. Les embryons sont extraits des graines sur des lames en présence d'eau glycérolée (50% glycérol, 0,3 µg/mL DAPI). La localisation de la GFP est observée en microscopie confocale.

#### **V.2.8.3 Observation confocale de GFP:LEC2 in vivo, étude du trafic de la protéine**

Pour étudier la mobilité de la protéine LEC2, 3 lignées indépendantes (*GFP:LEC2* 18-4, 17-1 et 4-6) ont été observées en microscopie confocale en présence de drogue bloquant l'import (ivermectin) ou l'export nucléaire (leptomycine B).

Les embryons en développement aux stades cœur et torpille ont été extraits des graines sur des lames en présence de (i) 50% glycérol, 300 mM de leptomycine B, 0,3 µg/mL DAPI, (ii) 0,3% DMSO, 300 µM ivermectin, 0,3 µg/mL DAPI et pour les témoins d'observations en présence de (i) eau glycérol 50% glycérol, 0,3 µm DAPI, (ii) eau à 0,3% DMSO, 0,3 µg/mL DAPI et incubés 5h à température ambiante avant l'observation en microscopie confocale selon les caractéristiques d'observation *in vivo* décrites dans le tableau 6.

#### **V.2.8.4 Observations confocales**

La localisation des protéines est observée à l'aide d'un microscope laser confocal LEICA TCS SP8 (Leica Microsystems) selon les caractéristiques suivantes (tableau) :

Signaux Fluorescent	Laser / Diode	Longueur d'onde excitation	Longueurs d'onde d'émission
<b>Observation sur échantillons fixés</b>			
DAPI	Diode 405	λ 405 nm	λ 415 - 459 nm
Alexa 488	Laser Argon	λ 488 nm	λ 503 - 527 nm
Alexa 647	Diode 633	λ 647 nm	λ 659 - 675nm
<b>Observation <i>in vivo</i></b>			
DAPI	Diode 405	λ 405 nm	λ 426 - 464 nm
GFP	Laser Argon	λ 488 nm	λ 490 - 522 nm
Auto fluorescence	Laser Argon	λ 488 nm	λ 660 - 748 nm

**Tableau 6 : Caractéristiques des observations des signaux de fluorescence en microscopie confocale selon la molécule fluorescente.**

## V.2.9 Études des interactions

### V.2.9.1 Interactions ADN / protéine

#### V.2.9.1.1 Interaction ADN / protéine par EMSA

Annexe 1A : article Baud *et al* 2016.

La sonde utilisée comprend 3 boîtes G (souligné) et 3 domaines RYL (gras): 5'-CAAACGTGTCTTTGAATAGACTCCTCTCGTTAACAC**ATGCAGCGGCTGCATGGT**GACGC CATTAACAGTGGCCTACAATTGCATG-ATGTCTCCATTGACCAGTACT-3'. Elle a été obtenue par traitement à la chaleur 5 minutes à 95°C des deux oligonucléotides de 80pb présentant un chevauchement de 48 pb, biotinylés en 5', synthétisés par la société Eurofins (5'-ggCAAACGTGTCTTTGAATAGACTCCTCTCGTTAACACATGCAGCGGCTG C ATGGTGACGCCATTAA-CACGTGGCCTACA-3' et 5'-ggAGTCACGTGTCAATGGAGACAT CATGCAATTGTAGGCCACGTGTTAATGGCGTACCATGCAGCCGCTGCATGTGTTA-3'). Les oligonucléotides sont dilués en quantité équimolaire et chauffés pendant 5 min à 95 °C. Les séquences double brin ainsi obtenues sont complétées par l'enzyme Klenow (fragment de Klenow, Thermo Scientific) pendant 1h à température ambiante afin d'obtenir une sonde double brin de 108 pb, biotinylée, à une concentration de 20 fmol.mL<sup>-1</sup>. Les réactions de liaison ont été réalisées avec le kit EMSA Shift Light Chemiluminescent (Thermo Scientific) et effectuées dans un tampon de liaison (Tris 10 mM, KCl 150 mM, dithiothréitol 1 mM, Nonidet P-40 à 0,05% [v / v] et 0,5 mg de poly [dl-dC], pH 7,5). 1µg pour LEC2, 1 à 3 µg pour LEC1 et 2 µg pour ABI3, de protéines ont été incubés (5 minutes sur la glace) dans du tampon de liaison avant d'ajouter 20 fmol de sonde biotinylée dans un volume réactionnel final de 20µL. Les réactions de liaison ont ensuite été incubées à température ambiante pendant 20 minutes dans l'obscurité. Les réactions de compétition ont été réalisées avec une augmentation de 25 fois de la sonde non marquée. Le gel d'électrophorèse (5% d'acrylamide dans 0,25% [v / v] de glycérol et 0,5X de Tris-borate / EDTA) a été soumis à une prémigration de 60 minutes à 120 V et les réactions de liaison ont été chargées sur le gel avec 5 µL de tampon de charge pour une migration de 6 h à 120 V et à 4°C. Le gel a été transféré sur une membrane d'ADN GeneScreen plus (Perkin-Elmer Life Sciences) pendant 1h à 25 V dans un tampon 0,5X Tris-borate / EDTA en utilisant un système de cellules de transfert Transblot SD Semi sec (Bio-Rad). L'ADN a été fixé sur la membrane avec un transilluminateur UV pendant 60 s à 254 nm avec la fonction auto-crosslink. Les bandes ont été révélées par chimioluminescence selon les instructions du kit de détection de l'acide nucléique chimiluminescent (ThermoFisher Scientific) en utilisant le système d'imagerie Fujifilm Las 4000.

#### V.2.9.1.2 Interaction ADN / protéine par Thermophorèse

La sonde double brin de 108pb du promoteur pro*OLE1* contient 3 boîtes G (souligné) et 3 domaines RYL (en gras) 5'-CAAACCGTGTCTTTGAATAGACTCCTCTCGTTAACACATGCAGCGGCT**TGCATGGT**GACGCCATTAACACGTGGCCTACAATTGCATG-ATGTCTCCATTGACACGTGACT-3'. La sonde a été préparée à partir d'oligonucléotides de 80 pb qui se chevauchent sur 48 pb, synthétisés par la société Sigma et marqués en 5' avec la cyanine 5 (5'-ggCAAACGTGTCTTTGAATAGACTCCTCTCGTTAACACATGCAGCGGCTGCATGGTGACGCCATTAACACGTGGCCTACA-3' et 5'-ggAGTCACGTGTCAATGGAGACATCATGCAATTGTAGGCCACGTGTTAATGGCGTCACCATGCAGCCGCTGCATGTGTTA-3').

Un mélange équimolaire de chaque oligonucléotide a été chauffé 5 min à 95°C, puis a été traité par la Klenow selon les recommandations du fournisseur (Biolabs). La sonde a été purifiée sur une colonne avec le kit MinElute reaction Cleanup (Qiagen). L'existence d'une seule population de sonde a été vérifiée sur un gel (5% d'acrylamide dans 0,25% [v / v] de glycérol et 0,5X de Tris-borate / EDTA, migration de 60 minutes à 120 V) et révélée par fluorescence de la cyanine 5 grâce à un lecteur de fluorescence Typhoon FLA9500 selon la méthode Cyanine 5 avec les paramètres suivants : laser à 635 nm, filtre LPR et 100 µm de résolution (Thermofisher).

L'interaction ADN / protéine a été mesurée par thermophorèse sur un Monolith NT.115 (NanoTemper). La protéine LEC2 recombinante contenue dans un tampon PBS 1X, 0,2 mM n-Dodecyl β-D-maltoside (DDM), concentrée à 650 µmol.L<sup>-1</sup> est diluée en cascade 1:1, 16 fois dans du PBS 1X + 0,2 mM DDM. La sonde marquée à la cyanine est diluée à une concentration finale de 20 nmol.L<sup>-1</sup> dans le tampon de liaison (Tris 10 mM, KCl 150 mM, dithiothréitol 1 mM, Nonidet P-40 à 0,05% [v / v] et 0,5 mg de poly [dl-dC], pH 7,5) et distribuée dans chaque tube contenant la protéine. La réaction est ensuite incubée 20 minutes à température ambiante dans l'obscurité, puis chargée dans des capillaires standards. Les capillaires sont scannés avec une excitation dans le rouge avec une puissance de LED power à 20%. Les mesures de thermophorèse de la sonde sont réalisées à une puissance de laser de 40% à 25°C constant, sur des intervalles de temps de 30 secondes (ON/OFF). Le logiciel de Nanotemper MO.Control V1.4.3 a été utilisé pour l'analyse des données.

#### V.2.9.1.3 Interaction ADN / protéine par expression transitoire chez *Physcomitrella patens*

L'étude de l'interaction des facteurs de transcription sur une séquence promotrice par expression transitoire chez la mousse *Physcomitrella patens* a été réalisée comme décrit par Thévenin *et al* (2012). Brièvement, les séquences codantes des facteurs de transcription étudiés sont recombinées par une réaction Gateway simple dans le vecteur pBS TPP-A (ou vecteur A) qui permet leurs expressions sous le contrôle du promoteur ACTINE. La



séquence du promoteur Oléosine 1 de -250pb, contenant 3 domaines RY et 3 boîtes G, est recombinée par réaction gateway simple dans le vecteur pBS Tpp-B (ou vecteur B), contrôlant l'expression de la GFP. Une interaction des facteurs de transcription sur la séquence promotrice induit l'activation du promoteur et l'expression de la GFP. L'expression de celle-ci est mesurée par cytométrie de flux comme décrit par Thévenin *et al* (2012).

### **V.2.9.2 Interactions protéine / protéine**

#### V.2.9.2.1 Interaction protéine / protéine par double hybride

Les ADNcs de *LEC1* et *LEC1D<sub>86</sub>K* contenus dans pDONR207 (Baud *et al* 2016) ont été directement recombinés LR dans les vecteurs pDEST22 et pDEST32 (Life Technologies). *LEC2* a été muté dans le domaine B2 sur les acides aminés T105G A108T R109N par mutagenèse dirigée selon le protocole commercial Agilent quick change, avec les oligonucléotides suivant 5'-GGTTTTCTTGATCCGCGTATGGGGAAGATG ACAAACA TCAACAGGAAAAACGCC-3' et 5'-GGCGTTTTTCCTGTTGATGTTTGTTCATCTTCCCCATA CGCGGATCAAGAAAACC-3'. *LEC2* sauvage et *LEC2* T<sub>105</sub>G, A<sub>108</sub>T, R<sub>109</sub>N délétés du domaine B3 ont été amplifiés par PCR avec les couples d'oligonucléotides suivant respectivement : 5' - GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGGATAACTTCTTA CCCTTTCC - 3', 5'- GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAAGCTTCTTGTTA TCAAATG - 3', 5' - GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGGATAACTTCCTC CCTTTCC - 3', 5' - GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGAGTTTCTTGTTA TCGAAGG - 3', vérifiés par séquençage, introduits par recombinaison BP dans le vecteur pDONR207, puis introduits par recombinaison LR dans les vecteurs de double hybride pDEST22 et pDEST32. La souche de levure AH109 (Clontech) a été cultivée sur le milieu YPDA (Clontech) à 28°C. Une culture de 50 mL atteignant un DO<sub>600</sub> entre 1 et 2 a été recueillie, lavée avec de l'eau stérile, culottée par centrifugation à vitesse maximale pendant 15 secondes et remise en suspension dans 500 µL d'acétate de lithium 0,1 M. Cinq cents ng de chaque couple de plasmides ont été mélangés avec 10 µL d'ADN porteur (Finnzymes, Fisher Scientific), 50 µL de solution de cellules de levure et 36 µL d'acétate de lithium à 40% [v / v] de polyethylene glycol 4000. Après une incubation de 30 min à 28°C, le mélange de transformation a été soumis à un choc thermique de 42°C pendant 25 minutes. Après centrifugation à vitesse maximale pendant 15 secondes à température ambiante, le culot a été remis en suspension dans 200 µL d'eau stérile et étalé sur de la dextrose synthétique-milieu Trp-Leu (Clontech). Pour analyser les interactions, les cellules de levure co-transformées ont été transférées dans 100 µL de milieu YPDA, incubées à 28°C sur la nuit, puis testées sur un milieu sélectif par un dépôt de 5 µL. Les colonies croissant sur un milieu dépourvu de histidine (His) ou de histidine et d'Alanine (Ala) et en présence de diverses concentrations de 3-aminotriazole, ont été considérées comme des interactions positives.

## V.2.9.2.2 Interaction protéine / protéine par TAP TAG et Immunoprécipitation

### V.2.9.2.2.1 Préparation de l'échantillon

Les graines de 630 siliques sont prélevées manuellement à la loupe binoculaire et stockées dans l'azote liquide le temps du prélèvement puis conservées à -80°C jusqu'à l'extraction des protéines.

#### V.2.9.2.2.1.1 Extraction des protéines totales

Les graines sont broyées dans l'azote liquide à l'aide d'un mortier et d'un pilon, le broyat est repris dans 3 mL de tampon d'extraction des complexes (selon l'étiquette choisie), incubé 30 min à 4°C sous agitation douce, puis filtré sur un tamis cellulaire en nylon de 40 µm (22363547, Fisher Scientific) à 4°C. L'échantillon filtré est centrifugé à 13000 g, 10 min à 4°C, le surnageant est conservé pour la purification de complexe. Le culot est digéré à la DNase I pour fragmenter l'ADN (voir 2.9.2.2.1.3).

#### V.2.9.2.2.1.2 Purification des noyaux (*Nature Protocols, Vol 6 n°1, 2011*)

Les graines sont broyées dans l'azote liquide à l'aide d'un mortier et d'un pilon. Le broyat est repris dans 10 mL de tampon d'extraction de noyaux froid (20 mM MOPS pH7, 40 mM NaCl, 90 mM KCl, 2 mM EDTA pH7, 0,5 mM EGTA pH7, cocktail d'inhibiteurs de protéases [complete ultra mini EDTA free, Roche]), puis filtré sur un tamis cellulaire en nylon de 40 µm (22363547, Fisher Scientific) à 4°C. L'échantillon filtré est centrifugé à 1000g pendant 10 min à 4°C. Le culot est repris dans 1 ml de tampon d'extraction et transféré dans un tube de 1 mL, incubé 3 minutes dans la glace puis centrifugé à 1000 g, pendant 5 min à 4°C. Le culot correspondant aux noyaux purifiés est re-suspendu dans le tampon d'extraction des complexes (selon l'étiquette choisie) et incubé 30 minutes à 4°C sous agitation légère. Après une centrifugation de 10 min à 1500 g, le surnageant est récupéré et utilisé pour la suite de la purification de complexe, le culot cellulaire est digéré à la DNase I.

#### V.2.9.2.2.1.3 Digestion à la DNase I

Les culots cellulaires sont re-suspendus dans 1 mL de tampon DNase I (100 mM Tris pH7,5, 25 mM MgCl<sub>2</sub>, 5mM CaCl<sub>2</sub>) et incubés avec 10 unités de DNase I (EN0521, ThermoScientific) 1h à 15°C sous agitation légère. Après une centrifugation à 1500g pendant 10 minutes à 4°C, le surnageant est ajouté au premier surnageant.

Les surnageants issus de la purification et de la digestion à la DNase I sont regroupés pour les étapes de purification de complexe.

#### V.2.9.2.2.2 Purification en tandem par affinité (TAP-tag)

##### V.2.9.2.2.2.1 Etiquette 2 protéines G-streptII (GS) (Jonas et al 2016)

Le broyat de graines est repris dans le tampon d'extraction (25 mM Tris pH7,6, 15 mM MgCl<sub>2</sub>, 5 mM EGTA, 300 mM NaCl, 0,1% NP40, 1 mM DTT, Cocktail d'inhibiteurs de protéases [complete ultra mini EDTA free, Roche]).

Avant l'incubation de l'extrait avec la résine IgG Sepharose 6 fast Flow (GE Healthcare), l'échantillon est dilué pour atteindre une concentration de 150 mM de NaCl, puis dosé par la méthode de Bradford. Chaque couple d'échantillons LEC2 (étiqueté et témoin) sont en quantité de protéines équivalente pour la suite de la purification.

Vingt cinq µL de résine sont équilibrés dans le tampon de d'extraction (25 mM Tris pH7,6, 15 mM MgCl<sub>2</sub>, 5 mM EGTA, 150 mM NaCl, 0,1% NP40, 1 mM DTT, Cocktail d'inhibiteur de protéases [Roch complete ultra mini EDTA free, Roche]), puis incubés avec l'extrait protéique sous agitation douce pendant 1h à 4°C. La résine est lavée avec 100 volumes de tampon de lavage (10 mM Tris HCl pH8, 150 mM NaCl, 5 mM EDTA, 0,1% NP40, 5% Glycérol), puis avec 50 volumes de tampon TEV (10 mM Tris HCl pH8, 150 mM NaCl, 0,1% NP40, 0,5 mM EDTA, 5% glycérol, Cocktail d'inhibiteurs de protéases [complete ultra mini EDTA free, Roche]).

La résine est incubée avec 100 µL de tampon de TEV et 2x20 unités d'enzyme TEV (AcTEV, Thermo Fischer Scientific) 1h à 16°C. L'enzyme est ajouté en 2 fois à 30 minutes d'intervalle. L'éluat est récupéré, la résine est lavée avec 100µL de tampon TEV et mélangée à l'éluat initial. L'extrait protéique est purifié une seconde fois avec 25 µL de billes magnétiques Magstrep « type3 » XT beads (iba) équilibrées avec le tampon TEV. Le mélange est incubé 1h à 4°C sous agitation douce. Les billes sont lavées avec 100 volumes de tampon TEV. Les complexes purifiés sont élués avec 30 µL de tampon Laemmli chaud, les billes sont incubées 5 min à 95°C.

##### V.2.9.2.2.2.2 Etiquette 6 Histidines- streptII (HS)

Les graines broyées sont reprises dans le tampon d'extraction (20 mM Tris pH8, 300 mM NaCl, 5 mM EDTA pH7, 0,5% NP40, 10 mM imidazole, Cocktail d'inhibiteurs de protéases (Roche EDTA free)).

Après récupération du surnageant et avant l'incubation de l'extrait avec la résine d'agarose Ni-NTA (Qiagen), l'échantillon est dilué pour atteindre une concentration de 150 mM de NaCl. L'imidazole ne permettant pas le dosage par la méthode de Bradford, la totalité des extraits protéiques sont utilisés pour la suite de la purification des complexes.

Vingt cinq µL de résine Ni-NTA sont équilibrés dans le tampon d'interaction (20 mM Tris pH8, 300 mM NaCl, 5 mM EDTA pH7, 0,5% NP40, 10 mM imidazole, Cocktail d'inhibiteurs

de protéases [Roche EDTA free]), puis incubés avec l'extrait protéique sous agitation douce pendant 1h à 4°C. La résine est culottée par centrifugation à 2500g pendant 2 min à 4°C et lavée 3 fois avec 1 mL de tampon de lavage (20 mM Tris pH8, 300 mM NaCl, 5mM EDTA pH7, 0,5% NP40, 20 mM imidazole, Cocktail d'inhibiteurs de protéases [complete ultra mini EDTA free, Roche]). Les protéines fixées à la résine sont éluées avec le tampon d'éluion (20 mM Tris pH8, 150 mM NaCl, 5 mM EDTA pH7, 0,5% NP40, 200 mM imidazole). Les protéines éluées sont ensuite purifiées grâce à l'étiquette Strept II. Vingt cinq µL de billes magnétiques Magstrep « type3 » XT beads (iba) sont équilibrées dans le tampon d'éluion dilué au ½, afin d'obtenir une concentration d'imidazole de 100mM compatible avec la purification avec la Streptavidine. Les billes sont incubées avec les protéines purifiées, diluées au ½, 1h à 4°C sous agitation douce, puis lavées 3 fois avec 1 mL de tampon de lavage (10 mM Tris HCl pH8, 150 mM NaCl, 0,1% NP40, 0,5 mM EDTA, 5% glycérol, Cocktail d'inhibiteurs de protéases [complete ultra mini EDTA free, Roche]), Les protéines sont éluées avec 30 µL de tampon Laemmli 1X chaud et chauffées 5 min à 95°C.

#### *V.2.9.2.2.3 Immunoprécipitation*

##### *V.2.9.2.2.3.1 Etiquette GFP*

Les graines broyées sont reprises dans le tampon d'extraction (20 mM Tris pH7,5, 300 mM NaCl, 5 mM EDTA pH7, 1% NP40, Cocktail d'inhibiteurs de protéases [complete ultra mini EDTA free, Roche]).

Avant l'incubation de l'extrait avec la résine GFP, l'échantillon est dilué pour atteindre une concentration à 150 mM de NaCl, puis dosé par la méthode de Bradford.

Vingt cinq µL de résine GFP-Trap-M (Chromotek) sont équilibrés dans le tampon de lavage (20 mM Tris pH7,5, 150 mM NaCl, 5 mM EDTA pH7, 1% NP40, Cocktail d'inhibiteurs de protéases [complete ultra mini EDTA free, Roche]), puis incubés avec les extraits protéiques sous agitation douce à 4°C pendant 1 heure, chaque couple d'échantillons LEC2 étiquetée et témoin est en quantité de protéines équivalente. La résine est lavée 3 fois avec le tampon de lavage, après avoir culotté la résine par centrifugation à 2500 g pendant 2 min à 4°C. Les protéines fixées à la résine sont éluées avec le tampon Laemmli 1X chaud, les billes d'agarose sont chauffées avec 30 µL de tampon à 95°C pendant 5 minutes.

##### *V.2.9.2.2.4 Identification des partenaires de LEC2 par spectrométrie de masse*

Migration courte des échantillons purifiés

Les échantillons repris dans 30 µL de Laemmli 1X final sont déposés sur un gel NuPAGE bis-Tris 12% (NP0341, Invitrogen). Le gel est soumis à une migration à 200V constant avec un tampon de migration MOPS (NP0001, Invitrogen) pendant 4 minutes, puis est coloré au bleu de Coomassie R250 (0,125% Brillant blue R250, 50% éthanol 95%, 5% acide acétique).

Une fois le gel décoloré avec la solution de décoloration (40% éthanol 95%, 10% acide acétique), les bandes de gel concentrées sont découpées, conservées dans l'eau et envoyées pour analyse en spectrométrie de masse.

Les analyses en spectrométrie de masse ont été réalisées sur la plateforme de l'Institut Jacques Monod. Les bandes de gel ont été réduites avec 10 mM de dithiothréitol (DTT), alkylées avec 55 mM d'iodoacétamide (IAA) et incubées avec 20  $\mu\text{L}$  de  $\text{NH}_4\text{HCO}_3$  à 25 mmol  $\text{L}^{-1}$  de contenant de la trypsine de séquence séquentielle (12,5  $\mu\text{g mL}^{-1}$ , Promega) toute la nuit à 37°C. Les peptides résultants ont été extraits séquentiellement avec 30% d'acétonitrile, 0,1% d'acide formique et 70% d'acétonitrile, 0,1% d'acide formique. Les analyses peptidiques ont été effectuées par un Orbitrap Fusion (ThermoFisher Scientific, San Jose, CA) couplé à un système de chromatographie Easy-nLC Proxeon 1000 (Thermo Fisher Scientific, San Jose, CA). La séparation chromatographique des peptides a été effectuée avec les paramètres suivants: Acclaim Pepmap100 pré-colonne (5mm, 300  $\mu\text{m}$  id, C18, 5  $\mu\text{m}$ , 100 Å) et Acclaim PepMap-RSLC Proxeon colonne (50 cm, 75  $\mu\text{m}$  id, C18, 2  $\mu\text{m}$ , 100 Å), 300 nl / min de débit, gradient en hausse de 90% de solvant A (2% d'acétonitrile, 0,1% d'acide formique) à 40% de solvant B (100% d'acétonitrile, 0,1% d'acide formique) en 100 minutes puis une augmentation à 80% de solvant B en 5 minutes. Les peptides ont été analysés dans le mode de balayage par ions Orbitrap en mode ionogramme complet à une résolution de 120000 (à  $m/z$  200) et d'une gamme de masse de  $m/z$  350-1550. Des fragments ont été obtenus avec une activation de la dissociation collisionnelle à haute énergie (HCD) avec une énergie collisionnelle de 30%, une largeur d'isolation quadripolaire de 1,6 Da et ont été analysés dans le LTQ dans un deuxième événement de balayage. Les données MS / MS ont été acquises dans le piège à ions linéaire en mode haute vitesse, avec une exclusion dynamique de 50 secondes et une durée de répétition de 60 secondes. Les temps d'accumulation d'ions maximum ont été réglés sur 250 ms pour l'acquisition de MS et 60 ms pour l'acquisition de MS / MS en mode de parallélisation. Les données ont été traitées avec le logiciel Proteome Discoverer 2.1 (ThermoFisher Scientific) couplé à un serveur de recherche Mascot interne (Matrix Science, version 2.5.1). La tolérance de masse des ions fragments a été fixée à 6 ppm pour les ions précurseurs et 0,5 Da pour les fragments. Les modifications suivantes ont été utilisées dans les paramètres variables: oxydation (Met) phosphorylation (Ser / Thr / Tyr), carbamidométhylation (Cys), acétylation (terme N). Le nombre maximum de clivages oubliés était limité à deux pour la trypsine protéase. Les données MS / MS ont été recherchées contre la base de données SwissProt (02/2017) avec la taxonomie *Arabidopsis thaliana*. Les valeurs Q des peptides ont été calculées à l'aide de l'algorithme du percolateur et un filtre de 1% a été appliqué en tant que seuil de taux de découvertes fausses (FDR).

### V.2.9.2.3 Identification des protéines en spectrométrie de masse

#### V.2.9.2.3.1 Migration courte en gel SDS-PAGE

Les échantillons, repris dans 30 µL de Laemmli 1X final, sont déposés sur un gel SDS-PAGE 10%. Le gel est soumis à une migration à 100V constant jusqu'à atteindre 3 mm dans le gel de résolution, puis il est coloré au bleu de Coomassie R250 (0,125% Brillant blue R250, 50% éthanol 95%, 5% acide acétique). Une fois le gel décoloré (40% éthanol 95%, 10% acide acétique), les bandes de gel concentrées sont découpées et décolorées avec des bains de 8 minutes dans 25 mM NH<sub>4</sub>HCO<sub>3</sub>, puis dans 100% ACN, répétés jusqu'à décoloration complète.

#### V.2.9.2.3.2 Digestion trypsique sur gel

Celle-ci est réalisée sur des gels SDS-PAGE en courte migration ou en séparation SDS-PAGE 1D selon le protocole adapté de Jolivet *et al* (2004). Les morceaux de gels ont été lavés pendant 5 min avec de l'eau, déshydratés pendant 15 minutes avec de l'ACN et séchés par centrifugation sous vide. Les protéines sont réduites dans une solution de 10 mM dithiothréitol (DTT), 0,1 M NH<sub>4</sub>HCO<sub>3</sub> à 56°C pendant 30 min et alkylées avec une solution de 55 mM iodoacétamide, 0,1 M NH<sub>4</sub>HCO<sub>3</sub> à température ambiante pendant 20 minutes dans le noir. Les morceaux de gels ont été lavés dans le tampon 0,1 M NH<sub>4</sub>HCO<sub>3</sub>, déshydratés avec de l'ACN et séchés sous vide. Les protéines ont été digérées avec une solution de 25 mM NH<sub>4</sub>HCO<sub>3</sub>, 5 mM CaCl<sub>2</sub>, 10 ng.L<sup>-1</sup> trypsine (Roche Diagnostics) à 37°C sur la nuit. Les peptides résultants ont été extraits avec 5% d'acide formique (v / v), ACN / H<sub>2</sub>O (50/50, v / v) et ACN. Les extraits combinés ont été séchés.

#### V.2.9.2.3.3 Dessalage des protéines digérées

Une étape de dessalage du produit digéré est nécessaire avant d'effectuer les analyses en spectrométrie de masse, elle a été réalisée par la plateforme Pappso du Moulon (Gif-sur-Yvette). L'échantillon est dilué dans un tampon de dilution (0,06% acide acétique, 3% acétonitrile) pour un volume final de 525 µL. La colonne Strata™XL 100 µm polymère en phase inverse (Phenomenex) est reconditionnée avec de l'acétonitrile 100% puis équilibrée 3 fois avec 500µL de tampon de dilution. 500 µL d'échantillon sont chargés sur la colonne, puis la colonne est lavée 4 fois avec 500 µL de tampon de dilution.

Les peptides sont élués 2 fois avec 300 µL le tampon d'éluion (0,06% acide acétique, 40% acétonitrile). L'éluat est complètement évaporé au speed-vac sous vide puis stockés à -20°C et est repris dans 20 µL d'acétonitrile pour l'analyse en spectrométrie de masse.

#### V.2.9.2.3.4 Analyse en spectrométrie de masse

Celle-ci a été réalisée sur la plateforme de protéomique Pappso du Moulon (Gif-sur-Yvette). La chromatographie liquide a été effectuée sur un système NanoLC Ultra (Eksigent). Un échantillon de 4  $\mu\text{L}$  a été chargé à 7,5  $\mu\text{L} / \text{min}$  sur une cartouche pré-colonne (phase stationnaire: C18, particules de 5  $\mu\text{m}$ , colonne: 100  $\mu\text{m}$  i.d. 2 cm de longueur, NanoSeparations) et dessalé avec 0,1% d'acide formique dans l'eau. Après 3 min, la cartouche pré-colonne a été connectée à la colonne C18 séparée (phase stationnaire BIOSPHERE C18, particules de 3  $\mu\text{m}$ , colonne 75  $\mu\text{m}$  i.d. 150 mm de longueur NanoSeparations). La séparation des peptides a été obtenue en utilisant un gradient linéaire de 5 à 30% de solvant B (0,1% d'acide formique dans l'acétonitrile), avec le solvant A (0,1% d'acide formique dans l'eau) pendant 28 min à 300 nL / min et incluant une étape de régénération à 100% de B et une étape d'équilibrage à 100% de A, d'une durée de 45 minutes. Les peptides élués ont été analysés en ligne avec un spectromètre de masse Q-Exactive (ThermoFisher Scientific) à l'aide d'une interface nanoélectrospray (sonde capillaire non revêtue, pointe de 10  $\mu\text{m}$  i.d; longueur de 12 cm, nouvel objectif). Les ions peptidiques ont été analysés à l'aide de Xcalibur 2.3 avec les étapes d'acquisition dépendant des données suivantes: (1) analyse MS complète (rapport masse / charge ( $m / z$ ) 400 à 1400, mode centroïde, résolution 70 000, cible AGC réglée à 3.106) et (2) MS / MS (état de charge précurseur: 2 à 4, mode centroïde, résolution 17 500). L'étape 2 a été répétée pour les 8 ions majeurs détectés à l'étape 1. L'exclusion dynamique a été fixée à 40 s. Les fichiers de données brutes Xcalibur ont été transformés en format open source mzXML à l'aide du logiciel ProteoWizard version 3.0.

- Alonso R, Onate-Sanchez L, Weltmeier F, Ehlert A, Diaz I, Dietrich K, Vicente-Carbajosa J, Droge-Laser W (2009) A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. *Plant Cell* 21 (6):1747-1761. doi:10.1105/tpc.108.062968
- Baud S, Boutin J, Miquel M, Lepiniec L, Rochat C (2002) An integrated overview of seed development in Arabidopsis thaliana ecotype WS. *Plant physiology and biochemistry* 400
- Baud S, Dubreucq B, Miquel M, Rochat C, Lepiniec L (2008) Storage reserve accumulation in Arabidopsis: metabolic and developmental control of seed filling. *Arabidopsis Book* 6:e0113. doi:10.1199/tab.0113
- Baud S, Kelemen Z, Thevenin J, Boulard C, Blanchet S, To A, Payre M, Berger N, Effroy-Cuzzi D, Franco-Zorrilla JM, Godoy M, Solano R, Thevenon E, Parcy F, Lepiniec L, Dubreucq B (2016) Deciphering the Molecular Mechanisms Underpinning the Transcriptional Control of Gene Expression by Master Transcriptional Regulators in Arabidopsis Seed. *Plant Physiol* 171 (2):1099-1112. doi:10.1104/pp.16.00034
- Baud S, Lepiniec L (2009) Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis. *Plant Physiol Biochem* 47 (6):448-455. doi:10.1016/j.plaphy.2008.12.006
- Bäumlein H, Miséra S, Luerssen H, Kölle K, Horstmann C, Wobus U, Müller AJ (1994) the FUS3 gene of Arabidopsis thaliana is a regulator of gene expression during late embryogenesis *Plant J*
- Ben-Naim O, Eshed R, Parnis A, Teper-Bamnolker P, Shalit A, Coupland G, Samach A, Lifschitz E (2006) The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J* 46 (3):462-476. doi:10.1111/j.1365-313X.2006.02706.x
- Berger N, Dubreucq B, Roudier F, Dubos C, Lepiniec L (2011) Transcriptional regulation of Arabidopsis LEAFY COTYLEDON2 involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *Plant Cell* 23 (11):4065-4078. doi:10.1105/tpc.111.087866
- Bies-Etheve N, da Silva Conceicao A, Giraudat J, Koornneef M, Leon-Kloosterziel K, Valon C, Delseny M (1999) Importance of the B2 domain of the Arabidopsis ABI3 protein for Em and 2S albumin gene regulation. *Plant Mol Biol* 40 (6):1045-1054.
- Braybrook SA, Stone SL, Park S, Bui AQ, Le BH, Fischer RL, Goldberg RB, Harada JJ (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc Natl Acad Sci U S A* 103 (9):3468-3473. doi:10.1073/pnas.0511331103
- Calvenzani V, Testoni B, Gusmaroli G, Lorenzo M, Gnesutta N, Petroni K, Mantovani R, Tonelli C (2012) Interactions and CCAAT-binding of Arabidopsis thaliana NF-Y subunits. *PLoS One* 7 (8):e42902. doi:10.1371/journal.pone.0042902
- Carbonero P, Iglesias-Fernandez R, Vicente-Carbajosa J (2016) The AFL subfamily of B3 transcription factors: evolution and function in angiosperm seeds. *J Exp Bot.* doi:10.1093/jxb/erw458
- Carles C, Bies-Etheve N, Lorette Aspart L, Leon-Kloosterziel K, Koornneef M, Echeverria M, Michel Delseny M (2002) Regulation of Arabidopsis thaliana Em genes: role of ABI5. *The Plant Journal* 30 (3):373-383



- Choudhury D, Roy S, Chakrabarti C, Biswas S, Dattagupta JK (2009) Production and recovery of recombinant propapain with high yield. *Phytochemistry* 70 (4):465-472. doi:10.1016/j.phytochem.2009.02.001
- Correa LG, Riano-Pachon DM, Schrago CG, dos Santos RV, Mueller-Roeber B, Vincentz M (2008) The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. *PLoS One* 3 (8):e2944. doi:10.1371/journal.pone.0002944
- Curaba J, Moritz T, Blervaque R, Parcy F, Raz V, Herzog M, Vachon G (2004) AtGA3ox2, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in Arabidopsis. *Plant Physiol* 136 (3):3660-3669. doi:10.1104/pp.104.047266
- Debeaujon I, Nesi N, Perez P, Devic M, Grandjean O, Caboche M, Lepiniec L (2003) Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. *Plant Cell* 15 (11):2514-2531. doi:10.1105/tpc.014043
- Deppmann CD, Acharya A, Rishi V, Wobbles B, Smeekens S, Taparowsky EJ, Vinson C (2004) Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: a comparison to Homo sapiens B-ZIP motifs. *Nucleic Acids Res* 32 (11):3435-3445. doi:10.1093/nar/gkh653
- Devic M, Roscoe T (2016) Seed maturation: Simplification of control networks in plants. *Plant Sci* 252:335-346. doi:10.1016/j.plantsci.2016.08.012
- Dolfini D, Gatta R, Mantovani R (2012) NF-Y and the transcriptional activation of CCAAT promoters. *Crit Rev Biochem Mol Biol* 47 (1):29-49. doi:10.3109/10409238.2011.628970
- Dumas C, Rogowsky P (2008) Fertilization and early seed formation. *Comptes rendus biologies* 331 (10):715-725. doi:10.1016/j.crv.2008.07.013
- Duong S, Vonapartis E, Li CY, Patel S, Gazzarrini S (2017) The E3 ligase ABI3-INTERACTING PROTEIN2 negatively regulates FUSCA3 and plays a role in cotyledon development in Arabidopsis thaliana. *J Exp Bot* 68 (7):1555-1567. doi:10.1093/jxb/erx046
- Ehlert A, Weltmeier F, Wang X, Mayer CS, Smeekens S, Vicente-Carbajosa J, Droge-Laser W (2006) Two-hybrid protein-protein interaction analysis in Arabidopsis protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. *Plant J* 46 (5):890-900. doi:10.1111/j.1365-313X.2006.02731.x
- Fatihi A, Boulard C, Bouyer D, Baud S, Dubreucq B, Lepiniec L (2016) Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds. *Plant Sci* 250:198-204. doi:10.1016/j.plantsci.2016.06.013
- Feeney M, Frigerio L, Cui Y, Menassa R (2013) Following vegetative to embryonic cellular changes in leaves of Arabidopsis overexpressing LEAFY COTYLEDON2. *Plant Physiol* 162 (4):1881-1896. doi:10.1104/pp.113.220996
- Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. *New Phytol* 171 (3):501-523. doi:10.1111/j.1469-8137.2006.01787.x
- Finkelstein R, Reeves W, Ariizumi T, Steber C (2008) Molecular aspects of seed dormancy. *Annu Rev Plant Biol* 59:387-415. doi:10.1146/annurev.arplant.59.032607.092740
- Finkelstein RR, Gampala SS, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. *Plant Cell* 14 Suppl:S15-45

- Finkelstein RR, Lynch TJ (2000) The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. *Plant Cell* 12 (4):599-609
- Gaj M, Zhang S, Harada J, Lemaux P (2005) Leafy cotyledon genes are essential for induction of somatic embryogenesis of Arabidopsis *Planta* 222 (6):977-988
- Gazzarrini S, Tsuchiya Y, Lumba S, Okamoto M, McCourt P (2004) The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones gibberellin and abscisic acid. *Dev Cell* 7 (3):373-385. doi:10.1016/j.devcel.2004.06.017
- Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* 4 (10):1251-1261.
- Gnesutta N, Kumimoto RW, Swain S, Chiara M, Siriwardana C, Horner DS, Holt BF, 3rd, Mantovani R (2017) CONSTANS Imparts DNA Sequence Specificity to the Histone Fold NF-YB/NF-YC Dimer. *Plant Cell* 29 (6):1516-1532. doi:10.1105/tpc.16.00864
- Gnesutta N, Saad D, Chaves-Sanjuan A, Mantovani R, Nardini M (2016) Crystal Structure of the Arabidopsis thaliana L1L/NF-YC3 Histone-fold Dimer Reveals Specificities of the LEC1 Family of NF-Y Subunits in Plants. *Mol Plant*. doi:10.1016/j.molp.2016.11.006
- Gong W, Shen YP, Ma LG, Pan Y, Du YL, Wang DH, Yang JY, Hu LD, Liu XF, Dong CX, Ma L, Chen YH, Yang XY, Gao Y, Zhu D, Tan X, Mu JY, Zhang DB, Liu YL, Dinesh-Kumar SP, Li Y, Wang XP, Gu HY, Qu LJ, Bai SN, Lu YT, Li JY, Zhao JD, Zuo J, Huang H, Deng XW, Zhu YX (2004) Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. *Plant Physiol* 135 (2):773-782. doi:10.1104/pp.104.042176
- Guillon F, Larre C, Petipas F, Berger A, Moussawi J, Rogniaux H, Santoni A, Saulnier L, Jamme F, Miquel M, Lepiniec L, Dubreucq B (2012) A comprehensive overview of grain development in *Brachypodium distachyon* variety Bd21. *J Exp Bot* 63 (2):739-755. doi:10.1093/jxb/err298
- Guo F, Liu C, Xia H, Bi Y, Zhao C, Zhao S, Hou L, Li F, Wang X (2013) Induced expression of AtLEC1 and AtLEC2 differentially promotes somatic embryogenesis in transgenic tobacco plants. *PLoS One* 8 (8):e71714. doi:10.1371/journal.pone.0071714
- Gusmaroli G, Tonelli C, Mantovani R (2001) Regulation of the CCAAT-Binding NF-Y subunits in Arabidopsis thaliana. *Gene* 264 (2):173-185. doi:S0378111901003237 [pii]
- Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C (2007) Combined networks regulating seed maturation. *Trends in plant science* 12 (7):294-300. doi:10.1016/j.tplants.2007.06.003
- Hackenberg D, Wu Y, Voigt A, Adams R, Schramm P, Grimm B (2012) Studies on differential nuclear translocation mechanism and assembly of the three subunits of the Arabidopsis thaliana transcription factor NF-Y. *Mol Plant* 5 (4):876-888. doi:10.1093/mp/ssr107
- Han JD, Li X, Jiang CK, Wong GKS, Rothfels CJ, Rao GY (2017) Evolutionary Analysis of the LAFL Genes Involved in the Land Plant Seed Maturation Program. *FrontPlant Sci*. doi:doi:10.3389/fpls;2017.00439
- Hill A, Nantel A, Rock CD, Quatrano RS (1996) A Conserved Domain of the viviparous-1 Gene Product Enhances the DNA Binding Activity of the bZIP Protein EmBP-1 and Other Transcription Factors. *J Biol Chem* 271 (16):3366-3374
- Hunter BG (2002) Maize Opaque Endosperm Mutations Create Extensive Changes in Patterns of Gene Expression. *The Plant Cell Online* 14 (10):2591-2612. doi:10.1105/tpc.003905

- Jakoby M, Weisshaar B, Droge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F (2002) bZIP transcription factors in Arabidopsis. *Trends in plant science* 7 (3):106-111
- Jenik PD, Gillmor CS, Lukowitz W (2007) Embryonic patterning in Arabidopsis thaliana. *Annu Rev Cell Dev Biol* 23:207-236. doi:10.1146/annurev.cellbio.22.011105.102609
- Jia H, McCarty DR, Suzuki M (2013) Distinct roles of LAFL network genes in promoting the embryonic seedling fate in the absence of VAL repression. *Plant Physiol* 163 (3):1293-1305. doi:10.1104/pp.113.220988
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473 (7345):97-100. doi:10.1038/nature09916
- Jolivet P, Acevedo F, Boulard C, d'Andrea S, Faure JD, Kohli A, Nesi N, Valot B, Chardot T (2013) Crop seed oil bodies: from challenges in protein identification to an emerging picture of the oil body proteome. *Proteomics* 13 (12-13):1836-1849. doi:10.1002/pmic.201200431
- Kahle J, Baake M, Doenecke D, Albig W (2005) Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin beta and importin 13. *Molecular and cellular biology* 25 (13):5339-5354. doi:10.1128/MCB.25.13.5339-5354.2005
- Kersten B, Agrawal GK, Durek P, Neigenfind J, Schulze W, Walther D, Rakwal R (2009) Plant phosphoproteomics: an update. *Proteomics* 9 (4):964-988. doi:10.1002/pmic.200800548
- Khandelwal A, Cho SH, Marella H, Sakata Y, Perroud PF, Pan A, Quatrano RS (2010) Role of ABA and ABI3 in desiccation tolerance. *Science* 327 (5965):546. doi:10.1126/science.1183672
- Kirchler T, Briesemeister S, Singer M, Schutze K, Keinath M, Kohlbacher O, Vicente-Carbajosa J, Teige M, Harter K, Chaban C (2010) The role of phosphorylatable serine residues in the DNA-binding domain of Arabidopsis bZIP transcription factors. *European journal of cell biology* 89 (2-3):175-183. doi:10.1016/j.ejcb.2009.11.023
- Kirkbride RC, Fischer RL, Harada JJ (2013) LEAFY COTYLEDON1, a key regulator of seed development, is expressed in vegetative and sexual propagules of Selaginella moellendorffii. *PLoS One* 8 (6):e67971. doi:10.1371/journal.pone.0067971
- Kohler C, Hennig L (2010) Regulation of cell identity by plant Polycomb and trithorax group proteins. *Current opinion in genetics & development* 20 (5):541-547. doi:10.1016/j.gde.2010.04.015
- Kroj T, Savino G, Valon C, Giraudat J, Parcy F (2003) Regulation of storage protein gene expression in Arabidopsis. *Development* 130 (24):6065-6073. doi:10.1242/dev.00814
- Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* 15 (1):5-18
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227 (259):680-685

- Lara P, Onate-Sanchez L, Abraham Z, Ferrandiz C, Diaz I, Carbonero P, Vicente-Carbajosa J (2003) Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. *J Biol Chem* 278 (23):21003-21011. doi:10.1074/jbc.M210538200
- Laux T, Wurschum T, Breuninger H (2004) Genetic regulation of embryonic pattern formation. *Plant Cell* 16 Suppl:S190-202. doi:10.1105/tpc.016014
- Lee H, Fischer RL, Goldberg RB, Harada JJ (2003) Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci U S A* 100 (4):2152-2156
- Leon-Kloosterziel KM, Keijzer CJ, Koornneef M (1994) A Seed Shape Mutant of Arabidopsis That Is Affected in Integument Development. *Plant Cell* 6 (3):385-392. doi:10.1105/tpc.6.3.385
- Li Y, Jin KE, Zhu ZHU, Yang JI (2010) Stepwise Origin and Functional Diversification of the Afl Subfamily B3 Genes during Land Plant Evolution. *Journal of bioinformatics and computational biology* 08 (supp01):33-45. doi:10.1142/s0219720010005129
- Li Y, Varala K, Coruzzi GM (2015) From milliseconds to lifetimes: tracking the dynamic behavior of transcription factors in gene networks. *Trends Genet* 31 (9):509-515. doi:10.1016/j.tig.2015.05.005
- Liu JX, Howell SH (2010) bZIP28 and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in Arabidopsis. *Plant Cell* 22 (3):782-796. doi:10.1105/tpc.109.072173
- Liu JX, Srivastava R, Che P, Howell SH (2007) An endoplasmic reticulum stress response in Arabidopsis is mediated by proteolytic processing and nuclear relocation of a membrane-associated transcription factor, bZIP28. *Plant Cell* 19 (12):4111-4119. doi:10.1105/tpc.106.050021
- Llorca CM, Berendzen KW, Malik WA, Mahn S, Piepho HP, Zentgraf U (2015) The Elucidation of the Interactome of 16 Arabidopsis bZIP Factors Reveals Three Independent Functional Networks. *PLoS One* 10 (10):e0139884. doi:10.1371/journal.pone.0139884
- Lopez-Molina L, Mongrand S, McLachlin DT, Chait BT, Chua NH (2002) ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. *Plant J* 32 (3):317-328
- Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93 (7):1195-1205
- Lu QS, Paz JD, Pathmanathan A, Chiu RS, Tsai AY, 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 J* 64 (1):100-113. doi:10.1111/j.1365-313X.2010.04307.x
- Luan S (2003) Protein phosphatases in plants. *Annu Rev Plant Biol* 54:63-92. doi:10.1146/annurev.arplant.54.031902.134743
- Ludwikow A (2015) Targeting proteins for proteasomal degradation-a new function of Arabidopsis ABI1 protein phosphatase 2C. *Front Plant Sci* 6:310. doi:10.3389/fpls.2015.00310
- Luerksen H, Kirik V, Herrmann P, Misera S (1998) FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. *Plant J* 15 (6):755-764.

- Lynch T, Erickson BJ, Finkelstein RR (2012) Direct interactions of ABA-insensitive(ABI)-clade protein phosphatase(PP)2Cs with calcium-dependent protein kinases and ABA response element-binding bZIPs may contribute to turning off ABA response. *Plant Mol Biol* 80 (6):647-658. doi:10.1007/s11103-012-9973-3
- Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene* 239 (1):15-27
- Marella HH, Quatrano RS (2007) The B2 domain of VIVIPAROUS1 is bi-functional and regulates nuclear localization and transactivation. *Planta* 225 (4):863-872. doi:10.1007/s00425-006-0398-6
- Masiero S, Imbriano C, Ravasio F, Favaro R, Pelucchi N, Gorla MS, Mantovani R, Colombo L, Kater MM (2002) Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. *J Biol Chem* 277 (29):26429-26435. doi:10.1074/jbc.M202546200
- Meinke DW, Franzmann LH, Nickle TC, Yeung EC (1994) Leafy Cotyledon Mutants of Arabidopsis. *Plant Cell* 6 (8):1049-1064
- Mendes A, Kelly AA, van Erp H, Shaw E, Powers SJ, Kurup S, Eastmond PJ (2013) bZIP67 regulates the omega-3 fatty acid content of Arabidopsis seed oil by activating fatty acid desaturase3. *Plant Cell* 25 (8):3104-3116. doi:10.1105/tpc.113.116343
- Merkle T (2001) Nuclear import and export of proteins in plants: a tool for the regulation of signalling. *Planta* 213 (4):499-517. doi:10.1007/s004250100621
- Merkle T (2011) Nucleo-cytoplasmic transport of proteins and RNA in plants. *Plant Cell Rep* 30 (2):153-176. doi:10.1007/s00299-010-0928-3
- Monke G, Seifert M, Keilwagen J, Mohr M, Grosse I, Hahnel U, Junker A, Weisshaar B, Conrad U, Baumlein H, Altschmied L (2012) Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. *Nucleic Acids Res* 40 (17):8240-8254. doi:10.1093/nar/gks594
- Nakamura S, Lynch TJ, RR. F (2001) Physical interactions between ABA response loci of Arabidopsis. *Plant J* 26 (6):627-635
- Nambara E, Marion-Poll A (2003) ABA action and interactions in seeds. *Trends in plant science* 8 (5):213-217. doi:10.1016/s1360-1385(03)00060-8
- Nardini M, Gnesutta N, Donati G, Gatta R, Forni C, Fossati A, Vorrhein C, Moras D, Romier C, Bolognesi M, Mantovani R (2013) Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell* 152 (1-2):132-143. doi:10.1016/j.cell.2012.11.047
- Nardone V, Chaves-Sanjuan A, Nardini M (2016) Structural determinants for NF-Y/DNA interaction at the CCAAT box. *Biochim Biophys Acta*. doi:10.1016/j.bbagr.2016.09.006
- North H, Baud S, Debeaujon I, Dubos C, Dubreucq B, Grappin P, Jullien M, Lepiniec L, Marion-Poll A, Miquel M, Rajjou L, Routaboul JM, Caboche M (2010) Arabidopsis seed secrets unravelled after a decade of genetic and omics-driven research. *Plant J* 61 (6):971-981. doi:10.1111/j.1365-313X.2009.04095.x
- Ooms J, Leon-Kloosterziel KM, Bartels D, Koornneef M, Karssen CM (1993) Acquisition of Desiccation Tolerance and Longevity in Seeds of Arabidopsis thaliana (A Comparative Study Using Abscisic Acid-Insensitive abi3 Mutants). *Plant Physiol* 102 (4):1185-1191.

- Orozco-Arroyo G, Paolo D, Ezquer I, Colombo L (2015) Networks controlling seed size in *Arabidopsis*. *Plant Reprod* 28 (1):17-32. doi:10.1007/s00497-015-0255-5
- Para A, Li Y, Marshall-Colon A, Varala K, Francoeur NJ, Moran TM, Edwards MB, Hackley C, Bargmann BO, Birnbaum KD, McCombie WR, Krouk G, Coruzzi GM (2014) Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in *Arabidopsis*. *Proc Natl Acad Sci U S A* 111 (28):10371-10376. doi:10.1073/pnas.1404657111
- Park KC, Jeong J, Kim KI (2014) Regulation of mlkappaBNS stability through PEST-mediated degradation by proteasome. *Biochemical and biophysical research communications* 443 (4):1291-1295. doi:10.1016/j.bbrc.2013.12.140
- Pelletier JM, Kwong RW, Park S, Le BH, Baden R, Cagliari A, Hashimoto M, Munoz MD, Fischer RL, Goldberg RB, Harada JJ (2017) LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1707957114
- Peng FY, Weselake RJ (2011) Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in *Arabidopsis*. *BMC Genomics* 12:286. doi:10.1186/1471-2164-12-286
- Pimenta Lange MJ, Lange T (2006) Gibberellin biosynthesis and the regulation of plant development. *Plant Biol (Stuttg)* 8 (3):281-290. doi:10.1055/s-2006-923882
- Rechsteiner M, Rogers S (1996) PEST sequences and regulation by proteolysis. *Trends Biochem Sci* 21 (7):267
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290 (5499):2105-2110
- Righetti K, Vu JL, Pelletier S, Vu BL, Glaab E, Lalanne D, Pasha A, Patel RV, Provar NJ, Verdier J, Leprince O, Buitink J (2015) Inference of Longevity-Related Genes from a Robust Coexpression Network of Seed Maturation Identifies Regulators Linking Seed Storability to Biotic Defense-Related Pathways. *Plant Cell* 27 (10):2692-2708. doi:10.1105/tpc.15.00632
- Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 234 (4774):364
- Romanel EA, Schrago CG, Counago RM, Russo CA, Alves-Ferreira M (2009) Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. *PLoS One* 4 (6):e5791. doi:10.1371/journal.pone.0005791
- Roscoe TT, Guilleminot J, Bessoule JJ, Berger F, Devic M (2015) Complementation of Seed Maturation Phenotypes by Ectopic Expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in *Arabidopsis*. *Plant Cell Physiol* 56 (6):1215-1228. doi:10.1093/pcp/pcv049
- Sabelli PA, Larkins BA (2009) The development of endosperm in grasses. *Plant Physiol* 149 (1):14-26. doi:10.1104/pp.108.129437
- Santos Mendoza M, Dubreucq B, Miquel M, Caboche M, Lepiniec L (2005) LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in *Arabidopsis* leaves. *FEBS Lett* 579 (21):4666-4670. doi:10.1016/j.febslet.2005.07.037

- Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L (2008) Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *Plant J* 54 (4):608-620. doi:10.1111/j.1365-313X.2008.03461.x
- Schneitz K, Hülkamp M, Pruitt RE (1994) Wild-type ovule development in *Arabidopsis thaliana*: a light microscope study of cleared whole-mount tissue. *The Plant Journal* 7 (5):731-749
- Schutze K, Harter K, Chaban C (2008) Post-translational regulation of plant bZIP factors. *Trends Plant Sci* 13 (5):247-255. doi:10.1016/j.tplants.2008.03.002
- Shen Y, Devic M, Lepiniec L, Zhou DX (2015) Chromodomain, Helicase and DNA-binding CHD1 protein, CHR5, are involved in establishing active chromatin state of seed maturation genes. *Plant Biotechnol J*. doi:10.1111/pbi.12315
- Skubacz A, Daszkowska-Golec A, Szarejko I (2016) The Role and Regulation of ABI5 (ABA-Insensitive 5) in Plant Development, Abiotic Stress Responses and Phytohormone Crosstalk. *Front Plant Sci* 7:1884. doi:10.3389/fpls.2016.01884
- Sorensen MB (2002) Cellularisation in the endosperm of *Arabidopsis thaliana* is coupled to mitosis and shares multiple components with cytokinesis. *Development* 129 (24):5567-5576. doi:10.1242/dev.00152
- Srivastava R, Chen Y, Deng Y, Brandizzi F, Howell SH (2012) Elements proximal to and within the transmembrane domain mediate the organelle-to-organelle movement of bZIP28 under ER stress conditions. *Plant J* 70 (6):1033-1042. doi:10.1111/j.1365-313X.2012.04943.x
- Srivastava R, Deng Y, Howell SH (2014) Stress sensing in plants by an ER stress sensor/transducer, bZIP28. *Front Plant Sci* 5:59. doi:10.3389/fpls.2014.00059
- Stone SL, Braybrook SA, Paula SL, Kwong LW, Meuser J, Pelletier J, Hsieh TF, Fischer RL, Goldberg RB, Harada JJ (2008) *Arabidopsis* LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc Natl Acad Sci U S A* 105 (8):3151-3156. doi:10.1073/pnas.0712364105
- Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc Natl Acad Sci U S A* 98 (20):11806-11811. doi:10.1073/pnas.201413498
- Suzuki M, Wu S, Li Q, McCarty DR (2014) Distinct functions of COAR and B3 domains of maize VP1 in induction of ectopic gene expression and plant developmental phenotypes in *Arabidopsis*. *Plant Mol Biol* 85 (1-2):179-191. doi:10.1007/s11103-014-0177-x
- Swaminathan K, Peterson K, Jack T (2008) The plant B3 superfamily. *Trends in plant science* 13 (12):647-655. doi:10.1016/j.tplants.2008.09.006
- Ten Hove CA, Lu KJ, Weijers D (2015) Building a plant: cell fate specification in the early *Arabidopsis* embryo. *Development* 142 (3):420-430. doi:10.1242/dev.111500
- Thevenin J, Dubos C, Xu W, Le Gourrierc J, Kelemen Z, Charlot F, Nogue F, Lepiniec L, Dubreucq B (2012) A new system for fast and quantitative analysis of heterologous gene expression in plants. *New Phytol* 193 (2):504-512. doi:10.1111/j.1469-8137.2011.03936.x
- To A, Valon C, Savino G, Guilleminot J, Devic M, Giraudat J, Parcy F (2006) A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. *Plant Cell* 18 (7):1642-1651. doi:10.1105/tpc.105.039925

- Tunnacliffe A, Wise MJ (2007) The continuing conundrum of the LEA proteins. *Die Naturwissenschaften* 94 (10):791-812. doi:10.1007/s00114-007-0254-y
- van Dop M, Liao CY, Weijers D (2015) Control of oriented cell division in the Arabidopsis embryo. *Curr Opin Plant Biol* 23:25-30. doi:10.1016/j.pbi.2014.10.004
- Van Leene J, Witters E, Inze D, De Jaeger G (2008) Boosting tandem affinity purification of plant protein complexes. *Trends Plant Sci* 13 (10):517-520. doi:10.1016/j.tplants.2008.08.002
- Veerabagu M, Kirchler T, Elgass K, Stadelhofer B, Stahl M, Harter K, Mira-Rodado V, Chaban C (2014) The interaction of the Arabidopsis response regulator ARR18 with bZIP63 mediates the regulation of PROLINE DEHYDROGENASE expression. *Mol Plant* 7 (10):1560-1577. doi:10.1093/mp/ssu074
- Vicente-Carbajosa J, Carbonero P (2005) Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int J Dev Biol* 49 (5-6):645-651. doi:10.1387/ijdb.052046jc
- Vinson C, Acharya A, Taparowsky EJ (2006) Deciphering B-ZIP transcription factor interactions in vitro and in vivo. *Biochim Biophys Acta* 1759 (1-2):4-12. doi:10.1016/j.bbaexp.2005.12.005
- Waltner JK, Peterson FC, Lytle BL, Volkman BF (2005) Structure of the B3 domain from Arabidopsis thaliana protein At1g16640. *Protein Sci* 14 (9):2478-2483. doi:10.1110/ps.051606305
- Wardleworth B, Russell R, Bell S, Taylor G, White M (2002) Structure of Alba: an archaeal chromatin protein modulated by acetylation. *EMBO J* 21 (17):4654-4662
- Wehmeyer N, Hernandez LD, Finkelstein RR, Vierling E (1996) Synthesis of Small Heat-Shock Proteins Is Part of the Developmental Program of Late Seed Maturation. *Plant Physiol* 112 (2):747-757
- Weltmeier F, Rahmani F, Ehlert A, Dietrich K, Schutze K, Wang X, Chaban C, Hanson J, Teige M, Harter K, Vicente-Carbajosa J, Smeekens S, Droge-Laser W (2009) Expression patterns within the Arabidopsis C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. *Plant Mol Biol* 69 (1-2):107-119. doi:10.1007/s11103-008-9410-9
- Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierc J, Samach A, Coupland G (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. *Plant Cell* 18 (11):2971-2984. doi:10.1105/tpc.106.043299
- Xing H, Hong Y, Sarge KD (2010) PEST sequences mediate heat shock factor 2 turnover by interacting with the Cul3 subunit of the Cul3-RING ubiquitin ligase. *Cell stress & chaperones* 15 (3):301-308. doi:10.1007/s12192-009-0144-7
- Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T (2009) Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J* 58 (5):843-856. doi:TPJ3817 [pii] 10.1111/j.1365-313X.2009.03817.x
- Yamamoto A, Kagaya Y, Usui H, Hobo T, Takeda S, Hattori T (2010) Diverse roles and mechanisms of gene regulation by the Arabidopsis seed maturation master regulator FUS3 revealed by microarray analysis. *Plant Cell Physiol* 51 (12):2031-2046. doi:10.1093/pcp/pcq162



- Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Osanai T, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2004) Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1. *Plant Cell* 16 (12):3448-3459. doi:10.1105/tpc.104.026112
- Yotsui I, Saruhashi M, Kawato T, Taji T, Hayashi T, Quatrano RS, Sakata Y (2013) ABSCISIC ACID INSENSITIVE3 regulates abscisic acid-responsive gene expression with the nuclear factor Y complex through the ACTT-core element in *Physcomitrella patens*. *New Phytol* 199 (1):101-109. doi:10.1111/nph.12251
- Zhang F, Han M, Lv Q, Bao F, He Y (2015) Identification and expression profile analysis of NUCLEAR FACTOR-Y families in *Physcomitrella patens*. *Front Plant Sci* 6:642. doi:10.3389/fpls.2015.00642
- Zhang X, Garreton V, Chua N-H (2005) The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation. *Genes Dev* 19 (1):1532-1543

## **VI Annexes**

### **Annexe 1A : Article de résultats avec ses données supplémentaires**

Baud S, Kelemen Z, Thevenin J, Boulard C, Blanchet S, To A, Payre M, Berger N, Effroy-Cuzzi D, Franco-Zorrilla JM, Godoy M, Solano R, Thevenon E, Parcy F, Lepiniec L, Dubreucq B (2016) **Deciphering the Molecular Mechanisms Underpinning the Transcriptional Control of Gene Expression by Master Transcriptional Regulators in Arabidopsis Seed**. *Plant Physiol* 171 (2):1099-1112.

# Deciphering the Molecular Mechanisms Underpinning the Transcriptional Control of Gene Expression by Master Transcriptional Regulators in Arabidopsis Seed<sup>1</sup>

Sébastien Baud<sup>2</sup>, Zsolt Kelemen<sup>2</sup>, Johanne Thévenin, Céline Boulard, Sandrine Blanchet, Alexandra To, Manon Payre, Nathalie Berger, Delphine Effroy-Cuzzi, Jose Manuel Franco-Zorrilla, Marta Godoy, Roberto Solano, Emmanuel Thevenon, François Parcy, Loïc Lepiniec, and Bertrand Dubreucq\*

Institut Jean-Pierre Bourgin, Institut National de la Recherche Agronomique, AgroParisTech, Centre National de la Recherche Scientifique, Université Paris-Saclay, RD10, F-78026 Versailles cedex, France (S.Ba., Z.K., J.T., C.B., A.T., M.P., N.B., D.E.-C., L.L., B.D.); Laboratoire de Physiologie Cellulaire et Végétale, Université Grenoble Alpes, Centre National de la Recherche Scientifique Unité Mixte de Recherche 5168, Commissariat à l'Énergie Atomique/DRF/BIG, Institut National de la Recherche Agronomique Unité Mixte de Recherche 1417, 38054 Grenoble, France (S.Bl., E.T., F.P.); and Genomics Unit (J.M.F.-Z., M.G.) and Plant Molecular Genetics Department (R.S.), Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, Campus Universidad Autónoma, 28049 Madrid, Spain

ORCID IDs: 0000-0002-5507-327X (S.Ba.); 0000-0002-5293-5460 (J.T.); 0000-0002-0204-4045 (A.T.); 0000-0002-5552-0076 (M.G.); 0000-0003-2191-500X (F.P.); 0000-0002-5845-3323 (L.L.); 0000-0003-2661-5479 (B.D.).

In Arabidopsis (*Arabidopsis thaliana*), transcriptional control of seed maturation involves three related regulators with a B3 domain, namely LEAFY COTYLEDON2 (LEC2), ABSCISIC ACID INSENSITIVE3 (ABI3), and FUSCA3 (ABI3/FUS3/LEC2 [AFLs]). Although genetic analyses have demonstrated partially overlapping functions of these regulators, the underlying molecular mechanisms remained elusive. The results presented here confirmed that the three proteins bind RY DNA elements (with a 5'-CATG-3' core sequence) but with different specificities for flanking nucleotides. In planta as in the moss *Physcomitrella patens* protoplasts, the presence of RY-like (RYL) elements is necessary but not sufficient for the regulation of the *OLEOSIN1* (*OLE1*) promoter by the B3 AFLs. G box-like domains, located in the vicinity of the RYL elements, also are required for proper activation of the promoter, suggesting that several proteins are involved. Consistent with this idea, LEC2 and ABI3 showed synergistic effects on the activation of the *OLE1* promoter. What is more, LEC1 (a homolog of the NF-YB subunit of the CCAAT-binding complex) further enhanced the activation of this target promoter in the presence of LEC2 and ABI3. Finally, recombinant LEC1 and LEC2 proteins produced in Arabidopsis protoplasts could form a ternary complex with NF-YC2 in vitro, providing a molecular explanation for their functional interactions. Taken together, these results allow us to propose a molecular model for the transcriptional regulation of seed genes by the L-AFL proteins, based on the formation of regulatory multiprotein complexes between NF-YBs, which carry a specific aspartate-55 residue, and B3 transcription factors.

<sup>1</sup> This work was supported by ANR projects Plant TF Code (grant no. ANR-07-BLAN-0211), CERES (grant no. ANR-BLAN-1238), and ERA-CAPS ABCEEDS and by Labex Saclay Plant Sciences (grant no. ANR-10-LABX-0040-SPS to the Institut Jean-Pierre Bourgin).

<sup>2</sup> These authors contributed equally to the article.

\* Address correspondence to bertrand.dubreucq@versailles.inra.fr.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors ([www.plantphysiol.org](http://www.plantphysiol.org)) is: Bertrand Dubreucq ([bertrand.dubreucq@versailles.inra.fr](mailto:bertrand.dubreucq@versailles.inra.fr)).

S.Ba., Z.K., J.T., and C.B. performed most of the experiments; S.Bl., J.M.F.-Z., and M.G. performed research on SELEX and DNA-binding affinities; F.P. and R.S. supervised the PCR SELEX and DNA-binding affinity experiments; D.E.-C. provided technical assistance to S.Ba.; M.P. provided technical assistance to J.T.; A.T. and N.B. provided technical assistance to Z.K.; E.T. provided technical assistance to F.P.; L.L. and B.D. conceived the project (original screening and research plans) and wrote the article with contributions of all the authors.

[www.plantphysiol.org/cgi/doi/10.1104/pp.16.00034](http://www.plantphysiol.org/cgi/doi/10.1104/pp.16.00034)

Genetic and molecular analyses have delineated a complex network of transcriptional regulators controlling gene expression programs essential to accomplish seed maturation in Arabidopsis (*Arabidopsis thaliana*; for review, see Kagaya et al., 2005b; Braybrook et al., 2006; To et al., 2006; Braybrook and Harada, 2008; Santos-Mendoza et al., 2008; Suzuki and McCarty, 2008; Zhang and Ogas, 2009; Berger et al., 2011; Sreenivasulu and Wobus, 2013). This network comprises activators and repressors regulating phase transitions between embryogenesis, seed maturation, and vegetative development. Chromatin modifications also are involved, which repress the expression of the activators in late-maturing and germinating seeds as well as in vegetative organs (Suzuki and McCarty, 2008; Zhang and Ogas, 2009; Berger et al., 2011).

Among the key transcriptional activators of the maturation phase are three founding members of the B3 domain family, namely ABSCISIC ACID INSENSITIVE3

(ABI3), FUSCA3 (FUS3), and LEAFY COTYLEDON2 (LEC2; Giraudat et al., 1992; Luerssen et al., 1998; Stone et al., 2001). These AFL (ABI3/FUS3/LEC2) members act in concert with proteins homologous to the NF-YB subunit of the CCAAT box-binding protein called LEC1 and LEC1-LIKE (Lotan et al., 1998; Kwong et al., 2003; Cagliari et al., 2014; Hilioti et al., 2014). These genes will be named L-AFL (as proposed by Jia et al. [2013]). Additionally, basic leucine zippers (bZIPs) such as bZIP53 and bZIP67 (Alonso et al., 2009; Mendes et al., 2013), the MADS domain transcription factor AGAMOUS-LIKE15 (Thakare et al., 2008; Zheng et al., 2009), the MYB transcription factors MYB115 and MYB118 (Wang et al., 2009; Barthole et al., 2014), the WD repeat motif-containing EMB2757/TANMEI protein (Yamagishi et al., 2005), and the homeobox GLABRA2 (Shen et al., 2006) also were described as regulators of the maturation program.

LEC2, FUS3, and ABI3 share a B3 DNA-binding domain and play a major role in the transcriptional control of seed maturation (for review, see Koornneef et al., 1984; Giraudat et al., 1992; Meinke, 1992; Meinke et al., 1994; Parcy et al., 1997; Lotan et al., 1998; Luerssen et al., 1998; Harada, 2001). Complementary approaches have demonstrated the regulatory action exerted by these AFLs on embryo development, differentiation of zygotic tissues, accumulation of storage compounds (i.e. seed storage proteins [SSPs] and triacylglycerols [TAGs]), and acquisition of desiccation tolerance. Mutations affecting the AFL genes lead to partially overlapping and pleiotropic embryo phenotypes. Mutant embryos display abnormal suspensors, precocious cell cycle activation, growth of apical and root meristems, and cotyledons similar to young leaves accumulating high levels of chlorophyll and/or anthocyanins and lower amounts of storage compounds. Mutant seeds are less tolerant to desiccation and/or display precocious germination.

Several maturation-induced genes encode SSPs (e.g. *At2S1-4* and *CRA1*) and proteins involved in the storage of TAGs (e.g. *OLEOSIN1* [*OLE1*] and *HSD1*; Kagaya et al., 2005a; Santos Mendoza et al., 2005; Braybrook et al., 2006; Baud et al., 2009; Wang and Perry, 2013; Barthole et al., 2014). The B3 AFLs directly activate the expression of these genes through RY elements present in their promoter sequences (Fujiwara and Beachy, 1994; Ezcurra et al., 2000; Reidt et al., 2000; Reinders et al., 2002; Mönke et al., 2004). In agreement with this regulation, the core RY motif (5'-CATG-3') was shown to be necessary for the correct expression of several other seed-specific genes in Arabidopsis (Bäumlein et al., 1986, 1992; Ståhlberg et al., 1993; Conceição and Krebbers, 1994; Ellerström et al., 1996). This motif also is found in the promoter sequence of legume and monocot SSP genes (Dickinson et al., 1988; Bobb et al., 1997; Suzuki et al., 1997).

If the B3 DNA-binding domains of the AFL transcription factors display some variability in terms of DNA recognition (Golovenko et al., 2014), both FUS3 and ABI3 interact with RY motifs in vitro (Reidt et al., 2000; Mönke et al., 2004), and in vivo analyses have allowed defining a consensus sequence recognized by the two transcription factors (Mönke et al., 2004, 2012;

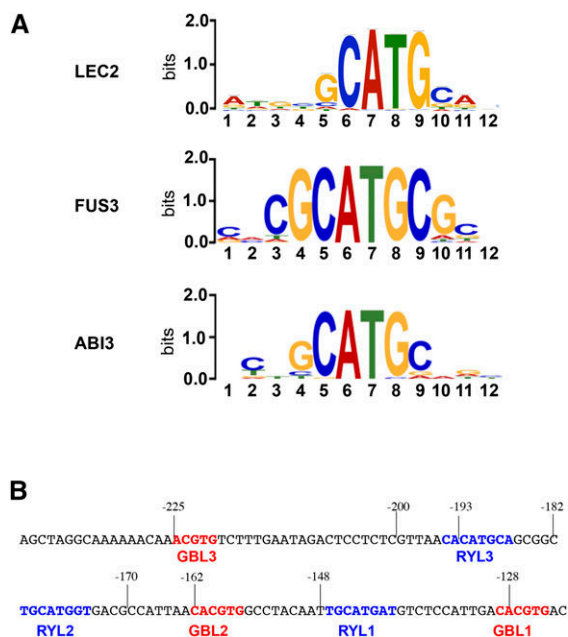
Wang and Perry, 2013). These results are consistent with the partially redundant phenotypes revealed both by loss- and gain-of-function analyses (To et al., 2006). In addition, yeast one-hybrid experiments and gel-shift assays have established that LEC2 and FUS3 can bind sequences containing RY elements separated by a G box (Kroj et al., 2003; Braybrook et al., 2006). Therefore, it was proposed that the B3 AFLs may act in concert with other transcription factors (e.g. bZIPs) interacting with these G boxes to confer proper expression to their common target genes during seed maturation (Sakata et al., 1997; Ezcurra et al., 1999; Hobo et al., 1999; Wobus and Weber, 1999; Kurup et al., 2000; Nakamura et al., 2001; Bensmihen et al., 2002; Brocard-Gifford et al., 2003; Lara et al., 2003; Zahn et al., 2005; Zhang et al., 2005; Nakashima et al., 2006).

Due to partial functional redundancy and the overlapping expression patterns of the B3 AFL genes (Roscoe et al., 2015), it is difficult to infer the specific in planta function of each AFL from genetic analyses. In the same line, ectopic overexpression of one of these proteins may trigger aspecific B3 effects, as observed previously with various transcriptional regulators (Xu et al., 2014), leading to possible misinterpretation. In order to decipher the specific role of each B3 AFL, we first conducted comprehensive analyses of the DNA-binding specificity of these proteins in vitro. In parallel, a functional analysis of the promoter of a known target of these AFLs, *OLE1* (At4g25140), was carried out in planta. Interactions of the AFLs with the putative cis-regulatory elements identified in this promoter were then investigated both in planta and in the moss *Physcomitrella patens* protoplasts. Last, the molecular interactions between the B3 AFLs and other transcriptional regulators like LEC1 were analyzed in moss protoplasts and in vitro. The results obtained showed that a regulatory complex comprising LEC2, ABI3, LEC1, and NF-YC2 is responsible for the synergistic effects of the L-AFL regulators. These results provide new insights into our understanding of the molecular mechanisms underpinning the transcriptional control of maturation-related genes in Arabidopsis.

## RESULTS

### Comprehensive Characterization of the DNA-Binding Sites of LEC2, ABI3, and FUS3 in Vitro

The DNA-binding sequences of the three AFL regulators were investigated in vitro using either high-throughput selection of ligand by exponential enrichment (SELEX) or protein-binding microarrays (Godoy et al., 2011; Franco-Zorrilla et al., 2014). The results confirmed that the three factors bind relatively similar elements made of a core 4-bp sequence (5'-CATG-3') flanked by nucleotides whose relative importance differs depending on the protein considered (Fig. 1; Supplemental Fig. S1A). FUS3 has a strong requirement for the two positions flanking the core sequence (5'-GCATGC-3'), whereas the specificities



**Figure 1.** Characterization of the B3 AFL-binding sites. A, Logos representing the DNA-binding specificities of ABI3, FUS3, and LEC2 as identified *in vitro*. Sequences presented for FUS3 and ABI3 correspond to the top-scoring 8-mers obtained for these proteins in protein-binding microarray assays; their corresponding position weight matrices (PWMs) were used for logo representations. The motif presented for LEC2 was derived from high-throughput SELEX experiments using MEME (see “Materials and Methods”). Scores are in arbitrary units (bits). B, Positions of the different cis-regulatory elements identified in the *OLE1* promoter. The RYL elements identified using PWMs obtained in A are highlighted in blue, whereas GBL elements identified using PLACE (<http://www.dna.affrc.go.jp/PLACE/index.html>) are highlighted in red. Numbers placed above the DNA sequence indicate nucleotide positions from the translational start site.

of ABI3 and LEC2 are more relaxed, since these proteins only require the 4-bp core sequence. These analyses provided us with models (PWMs) for the DNA-binding specificities of the B3 AFLs that could be used to identify putative binding sites in the regulatory regions of their target genes. *OLE1*, which encodes an oleosin participating in TAG storage in maturing embryos, is a known target of the B3 AFLs (Braybrook et al., 2006; Mönke et al., 2012). Scanning of a 257-bp *OLE1* promoter with these models resulted in the identification of three RY-like (RYL) elements (Fig. 1B; Supplemental Fig. S1B). The same sites were identified for ABI3, LEC2, and FUS3. In order to investigate the functions of these elements *in vivo*, a functional analysis of the *OLE1* promoter was carried out.

#### Functional Dissection of the *OLE1* Promoter in *Planta*

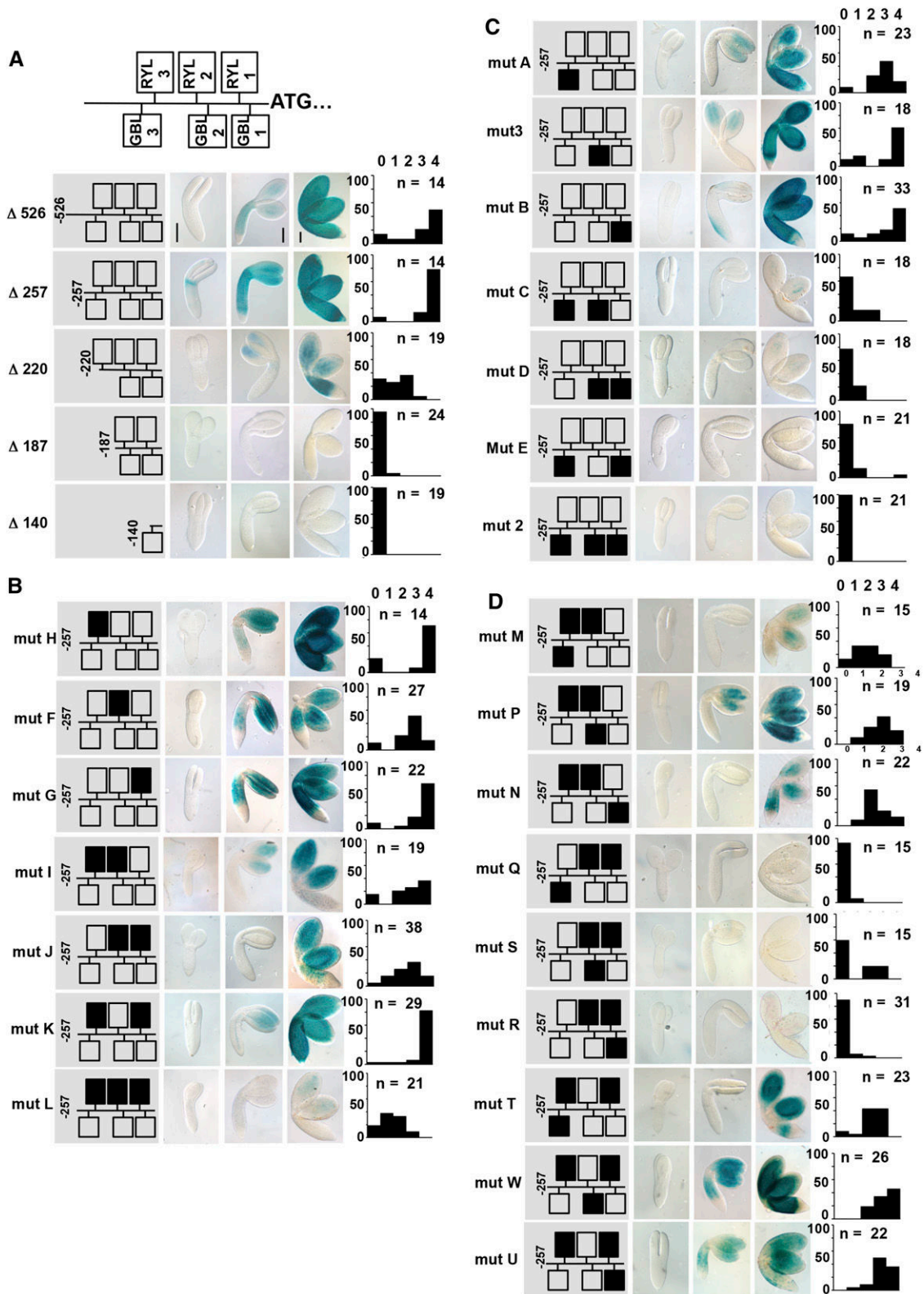
A deletion series of the *OLE1* promoter was fused translationally to the *uidA* reporter gene and stably introduced into plants. For each construct, at least 14 (and up to 38) independent transformants were obtained and analyzed for GUS activity (Fig. 2A). Both 526- and

257-bp fragments yielded strong activation of the reporter gene in maturing embryos. Deletion to position  $-220$  clearly decreased GUS activity. Further deletions to positions  $-187$  and  $-140$  resulted in a complete loss of activity. In light of these observations, the minimal 257-bp *OLE1* promoter fragment was analyzed further. This promoter region contained the three RYL motifs, named RYL1 (5'-TGCATGAT-3'), RYL2 (5'-TGCATGGT-3'), and RYL3 (5'-CACATGCA-3'), that were identified previously as putative binding sites of the B3 AFLs (Fig. 1B). These elements share a common 5'-CATG-3' sequence that fits well the consensus binding site determined for LEC2, ABI3, and FUS3 (Fig. 1A). This sequence is present in the four top-scoring motifs bound by FUS3 and ABI3 (Supplemental Fig. S1A). Interestingly, this promoter also contained three G box-like elements (GBL), two of those (GBL1 and GBL2) fitting perfectly the known consensus (5'-CACGTG-3') and the third one (GBL3) harboring a different nucleotide composition (5'-AACGTG-3'; Fig. 1B).

Different mutated versions of the *OLE1* minimal promoter (Supplemental Fig. S2) cloned upstream of the *uidA* reporter gene were stably introduced into *Arabidopsis* (Fig. 2, B and C). Mutation of one (mut H, F, and G) or two (mut I, J, and K) RY elements did not strongly affect the activity of the promoter, even though mutation of the RYL2 element led to slight and reproducible reduction of this activity (mut F, I, and J). Mutations of the three RYL elements (mut L) almost completely abolished the activity of the promoter. These results demonstrated a partial functional redundancy between the three elements in planta and confirmed the requirement for at least one RYL cis-regulatory element within the promoter of *OLE1* for its activation in maturing embryos. Similar analyses for the GBL elements (mut A, 3, B, C, D, E, and 2) demonstrated that at least two GBLs are required for proper promoter activity. These results confirmed the importance of the GBL elements, which also display partial functional redundancy. Interestingly, combining the mutations of two RYL and one GBL elements yielded contrasting results ranging from no activation (mut Q, R, and S) to slightly decreased activation in comparison with the native promoter (mut M, N, P, T, U, and W), showing that not all combinations of RYL and GBL elements are equivalent in planta (Fig. 2D). Taken together, these results demonstrated both the partial redundancy and functional interactions of the RYL and GBL elements for the transcriptional activation of *OLE1* in *Arabidopsis* embryos.

#### Synergistic Effect of the L-AFLs on the Activation of the *OLE1* Promoter

The ability of each transcriptional regulator to activate the *OLE1* promoter was investigated in the moss *P. patens* following a procedure established by Thévenin et al. (2012). The regulators were transiently expressed in the presence of a reporter GFP placed under the control



**Figure 2.** Functional dissection of the *OLE1* promoter in planta. A, At top is a schematic representation of the RYL and GBL elements identified within the *OLE1* promoter sequence. A series of 5' deletions was generated, and translational fusions to the *uidA* gene were prepared. The corresponding transgenic embryos were assayed for GUS activity at the torpedo, bent-cotyledon, and maturing stages. The length of the promoter tested is indicated at left. B and C, Mutations of the RYL (B) and GBL (C) elements



of a 257-bp-long *OLE1* promoter that contained all regulatory elements (Fig. 3 and supplemental Fig S3.). The results obtained showed that LEC2, ABI3, and FUS3 are able to induce the activity of this promoter. The strongest activation was observed with ABI3 and the weakest with FUS3. When expressed alone, LEC1 had no effect on the activity of the *OLE1* promoter. Combinations of transcription factors were then assayed. LEC2 and ABI3 synergistically activated the expression of the reporter gene. LEC1 further enhanced this transcriptional activity, whereas FUS3 did not (FUS3, therefore, was excluded from the next experiments).

The specific role of an amino acid residue (Asp-55) important for LEC1 function (Lee et al., 2003) and present in LEC1 but not in the homologous protein NF-YB7 (Asp being changed to Lys) was then investigated. LEC1, but neither NF-YB7 nor a mutated version of LEC1 (LEC1m D55K), could enhance the activity of the transcriptional complex, demonstrating the importance of residue Asp-55 in this activation (Fig. 4).

The synergistic effect observed between LEC1, LEC2, and ABI3 could result from an enhanced transcriptional activity of ABI3 or LEC2 or from changes in their DNA-binding properties within the complex. LEC1 also could affect promoter activity by providing, directly or indirectly, additional activation domains to the complex. To test the first hypothesis, ABI3 and LEC2 were fused to VP16, a strong transcriptional activation domain (Sadowski et al., 1988; Wilde et al., 1994; Parcy et al., 1998). Fusion proteins were expressed alone or in combination with LEC1 to test whether LEC1 still increased the transcriptional activity of complexes containing transcription factors fused to VP16 (Supplemental Fig. S4). VP16-LEC2 exhibited a stronger transcriptional activity than LEC2, demonstrating that the VP16 domain was fully active (Supplemental Fig. S4A). Yet, the addition of LEC1 further enhanced the activity of the VP16-LEC2-containing complex, suggesting that enhancement of the complex activity by LEC1 did not result from an increase in the transcriptional activity of LEC2 itself. A negative effect of the VP16 domain on ABI3 activity (lower activation of *ProOLE* expression, either alone or in combination with LEC1 or LEC2) prevented us from drawing any conclusion concerning ABI3 (Supplemental Fig. S4B).

#### Functional Interactions of the L-AFLs with the cis-Regulatory Elements of Oleosin Promoters

To study the interactions between the L-AFL complex and the cis-regulatory elements of its target promoters,

we first performed activation assays in moss protoplasts using the promoters of two oleosin genes closely related to *OLE1*, namely *OLE2* and *OLE3*. The promoter sequences of *OLE1* and *OLE3* both display repeats of RYL and GBL domains close to the translational start codon (Supplemental Fig. S5). On the contrary, RYL and GBL elements are scattered over a longer distance in *ProOLE2*. *ProOLE2* was weakly induced by the L-AFL complex, whereas a strong and synergistic induction of *ProOLE3* by the L-AFLs was observed, similar to the one reported previously for *ProOLE1*. These results showed that the relative positions of RYL and GBL elements within a promoter sequence dramatically affect the ability of the L-AFL complex to activate this promoter.

To further elucidate the interactions existing between the regulators under study and the cis-regulatory elements characterized, activation assays were then performed using mutagenized versions of the *OLE1* promoter (Supplemental Fig. S2). The activation of mutagenized promoter versions by LEC2, ABI3, or a combination of the three regulators LEC1, LEC2, and ABI3 was tested (Fig. 5). Mutation of a single RYL element (mut H, F, or G) had limited impact on the activation by ABI3 or by the combination of the three proteins, whereas activation by LEC2 alone was strongly decreased. Mutation of two or three RYL elements (mut I, J, K, and L) dramatically affected the ability of the transcriptional regulators, alone or in combination, to activate the promoter. Mutations affecting the GBL1 or GBL3 element (mut A and B) did not affect the induction of the reporter gene by the regulators tested, whereas mutation of the GBL2 element (mut 3) affected its induction by the complex (Fig. 5). Consistent with the results observed in planta, deletion of the three GBL domains dramatically affected promoter activity (mut 2). Taken together, these results confirmed the functional interactions between the L-AFL complex and the RYL and GBL elements under study in the *OLE1* promoter sequence. They also established the strong requirement of LEC2 for RYL elements to activate a promoter.

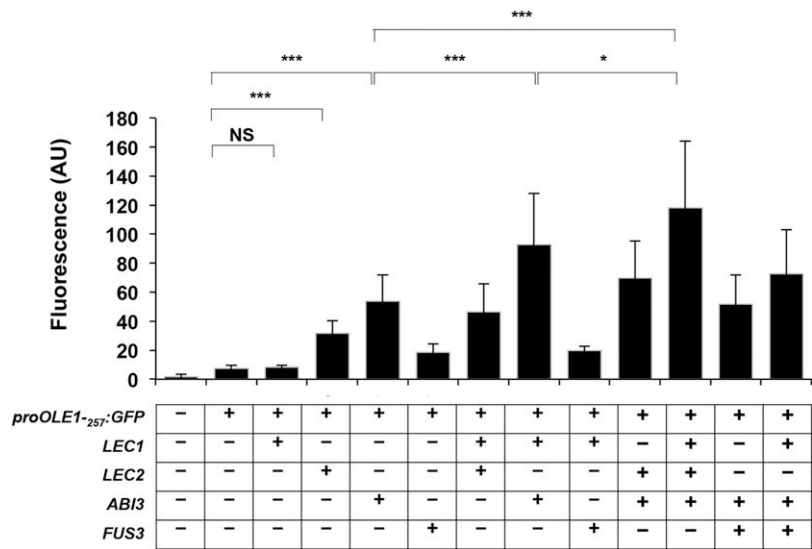
#### Molecular Interactions between the L-AFL Regulators and the *OLE1* Promoter

The ability of L-AFL proteins to interact physically with the *OLE1* promoter was tested in vitro by electrophoretic mobility-shift assay (EMSA; Supplemental Fig. S6). Both LEC2 and ABI3 proteins were able to

#### Figure 2. (Continued.)

were generated in the context of the 257-bp *OLE1* promoter, and translational fusions to the *uidA* gene were prepared. The corresponding transgenic embryos were assayed for GUS activity. D, Combinations of mutations affecting GBL and RY elements were generated in the context of the 257-bp *OLE1* promoter, and translational fusions to the *uidA* gene were prepared. The corresponding transgenic embryos were assayed for GUS activity. From left to right are the name of the mutagenized promoter under study, a schematic representation of the promoter with mutagenized elements in black, representative photographs of stained embryos at three different stages of the maturation process, and a bar graph showing the partition of staining intensities (from 0 = colorless to 4 = intense staining) among a population of *n* independent transformants.

**Figure 3.** Synergistic effects of LEC2, LEC1, and ABI3 on the activation of *ProOLE1* in moss protoplasts. Transient expression assays were carried out in moss protoplasts. GFP activity was measured in protoplasts transfected with the *ProOLE1*<sub>-257</sub>:GFP reporter and constructs allowing the expression of B3 AFL (LEC2, ABI3, or FUS3) or LEC1, alone or in combination. Means of at least three replicates ± sd are presented. AU, Arbitrary units; NS, nonsignificant. Asterisks indicate significant differences. \**P* value 5%, \*\*\**P* value 0.1%. Statistical analysis is provided in Supplemental Table S3.



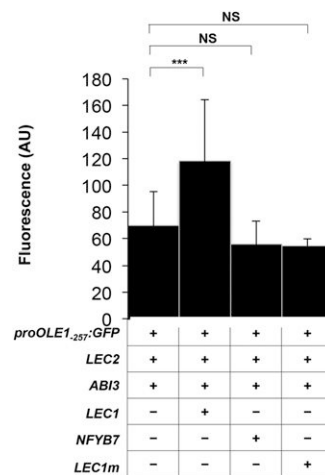
interact with the *ProOLE1* double-stranded probe (Supplemental Fig. S6, left). Surprisingly, LEC1 also was able to bind *ProOLE1* in vitro, although LEC1 has never been reported to bind DNA before. A supershift (Supplemental Fig. S6, arrows) was detected when LEC1 and LEC2, or LEC1, LEC2, and ABI3, were tested in combination, but not when LEC1 and ABI3, or LEC2 and ABI3, were mixed together. The intensity of this supershift appeared to be well correlated with LEC1 abundance (Supplemental Fig. S6, right). Competition experiments with wild-type nonlabeled probe demonstrated the specificity of these interactions.

**Molecular Interactions between the L-AFL Regulators**

Interaction studies between the four regulators were first investigated using a yeast two-hybrid approach (Supplemental Fig. S8). Due to a strong autoactivation of the reporter genes by LEC2 and ABI3 fused to the GAL4 DNA-binding domain, only a limited subset of interactions could be tested. No interaction was detected (Supplemental Fig. S8A) except the homodimerization of LEC1 (Supplemental Fig. S8B). This homodimerization was not affected by the D55K mutation.

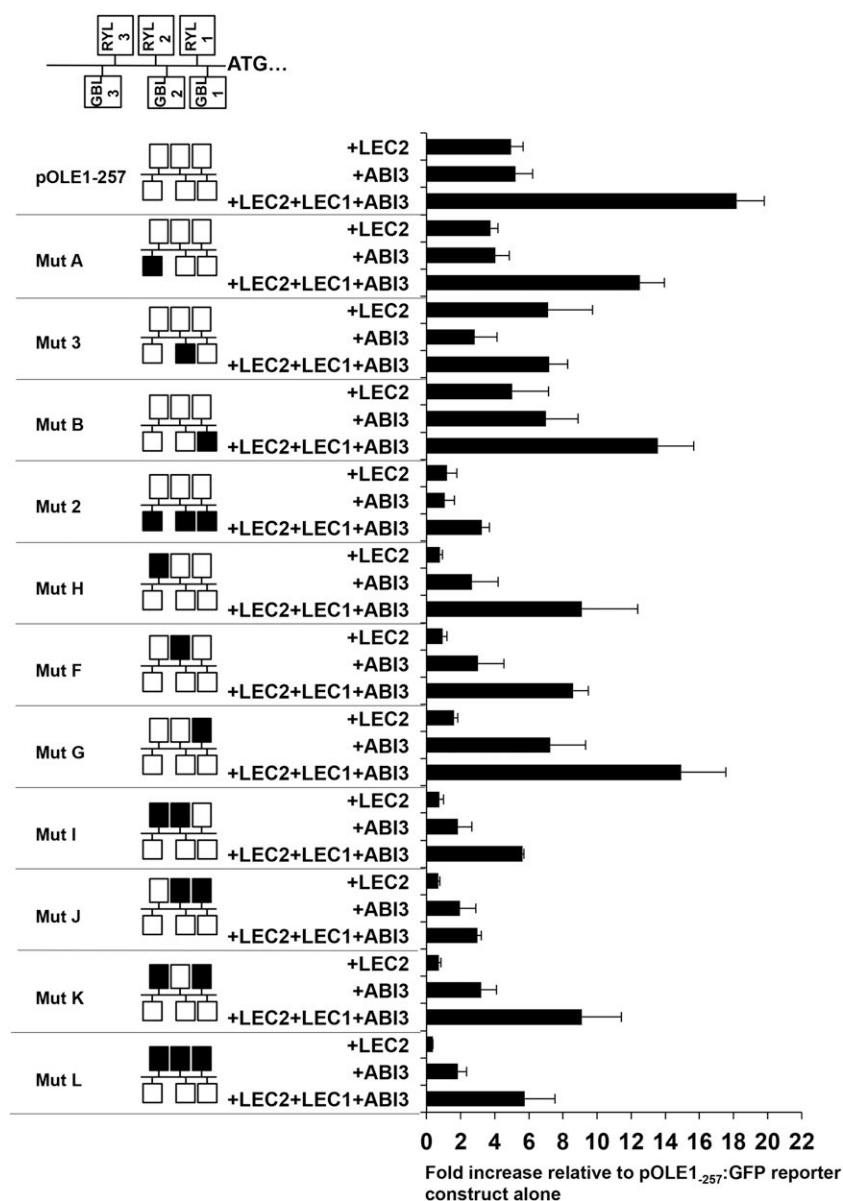
To test other interactions between the regulators, pull-down experiments were performed with recombinant hemagglutinin (HA)- or Myc-tagged proteins coexpressed in Arabidopsis protoplasts in the presence of a plasmid carrying the *OLE1* promoter. Proteins were immunoprecipitated using anti-HA antibodies and protein G-agarose beads. The immunoprecipitated proteins were then analyzed by western blot using both HA antibodies (as a control for immunoprecipitation) and anti-Myc antibodies to discover coprecipitated partners (Fig. 6; Supplemental Fig. S7). No interaction was detected using LEC1, LEC2, or ABI3 (Supplemental Fig. S8). Because NF-YC, a member of the NF-Y complex, was shown previously to interact with LEC1 (Hackenberg et al., 2012),

it was subsequently included in the experiments. Pairwise interactions were thus observed between NF-YC2 and LEC1 (or LEC1m D55K; Supplemental Fig. S6, A–C) but not with LEC2 (Supplemental Fig. S6, C and D). Then, NF-YC2, LEC2, and LEC1 (or LEC1m) were coexpressed simultaneously in the presence of the *OLE1* promoter. A strong band and a faint band corresponding to Myc-LEC1 (or LEC1m) and Myc-LEC2, respectively, were reproducibly detected in the immunoprecipitated fraction (Fig. 6B, bottom, columns 1 and 2, respectively). This result demonstrated that the three proteins can form a ternary



**Figure 4.** Enhanced activation of *ProOLE1* by NF-YB proteins is specific to the LEC1 type. Transient expression assays were carried out in moss protoplasts. GFP activity was measured in protoplasts transfected with the *ProOLE1*<sub>-257</sub>:GFP reporter and constructs allowing the expression of LEC2, ABI3, and LEC2 and ABI3 together with different NF-YB proteins, namely LEC1, NF-YB7 (At2g13570), and LEC1m (LEC1 D55K). Means of at least three replicates ± sd are presented. AU, Arbitrary units; NS, nonsignificant. Asterisks indicate a significant difference. \*\*\**P* value = 0.001. Statistical analysis is provided in Supplemental Table S3.





**Figure 5.** Importance of the RYL and GBL cis-regulatory elements in the activation of *proOLE1* by the L-AFLs in moss protoplasts. At top is a schematic representation of the RYL (RYL1–RYL3) and GBL (GBL1–GBL3) elements identified in the minimal *OLE1* promoter. At bottom are results from transient expression assays carried out in moss protoplasts with different mutagenized versions of the *OLE1* promoter. GFP activity was measured in protoplasts transfected with the *ProOLE1:GFP* reporter and constructs allowing the expression of LEC2, ABI3, or a combination of LEC2, LEC1, and ABI3. From left to right are the name of the mutagenized promoter under study, a schematic representation of the promoter with mutagenized elements in black, and a bar graph showing fluorescence intensities in moss protoplasts expressed as a ratio of the intensity measured with and without expression of the L-AFLs. Statistical analysis is provided in Supplemental Table S3.

complex with the *OLE1* promoter. Moreover, the detection of LEC2 was affected when replacing LEC1 by LEC1m (Fig. 6).

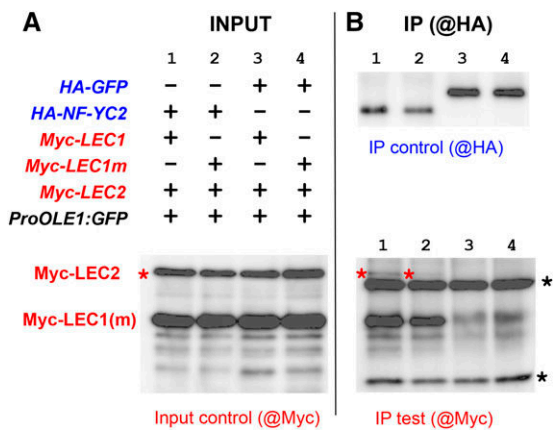
## DISCUSSION

Genetic analyses have unraveled functional interactions occurring in *Arabidopsis* seeds between the L-AFLs and the transcriptional regulation they exert, being mediated through RYL and GBL cis-regulatory elements present in their target promoters. Molecular analyses then established that the three B3 AFL regulators can bind RYL elements *in vitro* or in *Saccharomyces cerevisiae*. More recently, interactions of ABI3 and FUS3 with RYL and GBL domains were demonstrated in planta by chromatin immunoprecipitation (ChIP) approaches

(Wang and Perry, 2013), although the binding to cis-regulatory elements reported in these experiments could be direct or indirect. Here, we bring new insights into the functional and molecular interactions between B3 AFLs and LEC1 (NF-YB9) on one side and between these regulatory proteins and the different RYL and GBL cis-regulatory elements present in their target promoters on the other side.

### RY and G Boxes Are Involved in the Regulation of *OLE1* Expression in Maturing Embryos

We have conducted comprehensive *in vitro* analyses of DNA sequences bound by the B3 AFLs LEC2, ABI3, and FUS3 and showed that they recognize a similar 6-bp consensus sequence (5'-gCATGc-3'), which encompasses the core RY motif 5'-CATG-3'. Specific binding of FUS3



**Figure 6.** Interaction of LEC1, LEC2, and NF-YC2 in pull-down experiments. Coimmunoprecipitation experiments were performed using native or mutated LEC1 (LEC1m D55K) as a bait and LEC2 as a prey in the presence (or absence) of NF-YC2 and of a *proOLE1:GFP* construct. HA- and Myc-tagged proteins were coexpressed in Arabidopsis protoplasts and unraveled by western blot using anti-Myc and anti-HA antibodies, respectively. INPUT, 2% of the full protein extract; IP, immunoprecipitated sample. Asterisks indicate heavy (black) and light (red) chains of the antibody used for immunoprecipitation.

to the RY element has been shown by band-shift assays (Reidt et al., 2000) or biacore (Mönke et al., 2004), and the 5'-CATGCA-3' motif was found overrepresented within sequences bound by FUS3 (Wang and Perry, 2013). In our data, FUS3 exhibits a strong affinity for a longer consensus, namely 5'-cGCATGCg-3', that fits well the ChIP data of Wang and Perry (2013). Overall, the core RY sequence is always identified for the three B3 domain-containing factors with few differences in the 5' and 3' ends of the core sequence. Scanning of the *OLE1* minimal 257-bp promoter sequence with the PWMs led to the identification of three peaks, matching with RYL domains. Interestingly, a comparison of the promoter structures of the related *OLE1*, *OLE2*, and *OLE3* genes, all induced in maturing seeds and regulated by ABI3 (Mönke et al., 2012), revealed contrasting organizations of their RYL and GBL motifs. The promoters of *OLE1* and *OLE3* displayed similar arrangements of GBL and RYL elements close to the translational start site and exhibited a similar synergistic induction by the L-AFLs in moss protoplasts. On the contrary, this synergistic activation was not observed with the *OLE2* promoter, where GBL and RYL motifs are scattered over a long distance. These results and others (Che et al., 2009; Wang and Perry, 2013) emphasize the importance of the organization of RYL and GBL elements for the efficient activation of maturation-related genes by the L-AFLs. Further analyses with mutagenized versions of *ProOLE1* carried out in vivo (both in moss protoplasts and in planta) confirmed the importance of RYL and GBL elements for mediating the induction by L-AFLs. At least one RYL and two GBL domains are required for proper *ProOLE1* activity in planta. Nevertheless, it is still unknown whether the B3 AFLs can directly bind a GBL element. Although a direct

interaction cannot be excluded, neither our in vitro experiments nor recent ChIP analyses carried out with FUS3 (Mönke et al., 2004; Wang and Perry, 2013) support this idea. Rather, one can hypothesize that additional factors such as bZIPs are necessary for direct binding of the GBL elements. In agreement with this, the involvement of bZIP proteins in the accumulation of reserves during seed maturation is well established (for review, see Kroj et al., 2003; Vicente-Carbajosa and Carbonero, 2005). What is more, interactions between bZIPs and ABI3 were reported (Lara et al., 2003; Alonso et al., 2009), and some bZIPs (e.g. bZIP67 and bZIP28) already have been demonstrated to interact with LEC1 to activate seed-specific promoters (Yamamoto et al., 2009; Liu and Howell, 2010). Consistent with this hypothesis, bZIP homologs have been described in *P. patens* (Yotsui et al., 2013), which may explain why the L-AFL system requires GBL elements for a stronger activation of the *ProOLE1* promoter.

### Specific Roles of the B3 AFL Regulators

B3 proteins are classified into five subfamilies (Auxin Response Factor, Reproductive Meristem, High Level of Suc Inducible, Related to ABI3/VP1, and ABI3; Yamasaki et al., 2004), each subgroup exhibiting distinct DNA-binding characteristics (Boer et al., 2014; Franco-Zorrilla et al., 2014). They share a conserved B3 DNA-binding domain and can contain additional domains (e.g. B1 and B2) involved in protein-protein interactions (Romanel et al., 2009). The B3 AFLs, namely ABI3, LEC2, and FUS3, belong to the same subclass of ABI3-related transcription factors. Genetic and molecular analyses have demonstrated that these B3 AFL proteins have partially overlapping though nonidentical functions in planta (To et al., 2006; Mönke et al., 2012; Wang and Perry, 2013). Our results show that the three transcription factors can activate the *OLE1* promoter at similar levels when expressed alone in moss protoplasts. Nevertheless, only ABI3 and LEC2 display clear synergistic effects. The differential ability of these actors to cooperate and/or to be part of common transcriptional complexes may explain why their functions are not identical. Different DNA-binding properties or protein-protein interaction capacities may constitute the molecular bases explaining these different behaviors. Nonoverlapping expression patterns of these transcriptional regulators also may play a part in specifying their functions in developing embryos.

### LEC1 Acts Synergistically with LEC2 and ABI3

LEC1 (NF-YB9) is a member of the NF-YB protein family, which encodes components of an NF-Y regulatory complex (comprising NF-YA, NF-YB, and NF-YC subunits) involved in the recognition of the CCAAT box in yeast (Ronchi et al., 1995; Mantovani, 1999; Motta et al., 1999; Romier et al., 2003). With regard to

the large number of *NF-Y* genes present in plant genomes (in comparison with mammals and fungi), other roles have been postulated for these proteins in plants (Gusmaroli et al., 2001, 2002; Petroni et al., 2012). Members of the NF-YB family have thus been shown to work through mechanisms independent from the CCAAT box and to interact with proteins other than the NF-Ys (Masiero et al., 2002; Wenkel et al., 2006; Liu and Howell, 2010; Li et al., 2011). Similarly, LEC1 interacts functionally with ABRE-binding bZIP67 and bZIP28 to directly activate seed-specific genes (Yamamoto et al., 2009; Liu and Howell, 2010; Mendes et al., 2013).

In a somehow similar manner, this study establishes that LEC1, although unable to activate the *OLE1* promoter alone, enhances the transcriptional activity of LEC2 and ABI3. Interestingly, NF-YB7, a more distantly related protein used as a negative control, and a mutated version of LEC1 (D55K) did not. These results emphasized the importance of the amino acid Asp-55 in LEC1 for exerting a synergistic effect with LEC2 and ABI3. Consistent with these results is the role played by the Asp-55 residue in LEC1 function in planta (Lee et al., 2003).

To further elucidate the role of LEC1 in the synergistic activation of *ProOLE1* by the L-AFLs, we tested the effect of LEC1 acting in concert with B3 AFL proteins fused to the strong VP16 activation domain. Consistent with previous negative results obtained with VP16-ABI3 in yeast one-hybrid experiments (Kroj et al., 2003), the lack of activity of the chimerical protein prevented us from drawing any conclusion concerning ABI3. On the contrary, VP16-LEC2 was highly functional, and its activity was still further enhanced by the addition of LEC1, suggesting that the role of LEC1 in the complex is not to stimulate the transcriptional activity of individual B3 AFLs. Instead, LEC1 may favor the stability of the multiprotein complex, enhance the DNA-binding efficiency of the B3 AFLs, or provide additional activation domains to the complex.

### LEC1 and LEC2 Can Form a Ternary Complex in the Presence of NF-YC2

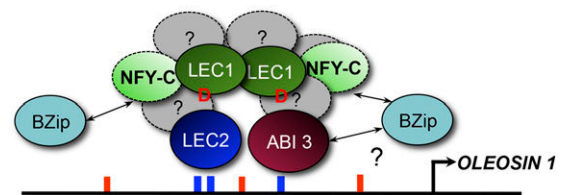
The tagged recombinant LEC1 and LEC2 proteins produced in *Arabidopsis* protoplasts can be coimmunoprecipitated when coexpressed with NF-YC2. Interestingly, the D55K mutation of LEC1 strongly impacted the coimmunoprecipitation of LEC2. This result is consistent with the lack of activity of the mutated LEC1m (D55K) and of NF-YB7 (Lys-55) in moss protoplasts. Then, homodimerization of LEC1 was demonstrated in yeast two-hybrid experiments. Taken together, these results provide some molecular bases for the key role played by LEC1 in structuring a complex (comprising NF-Y proteins and B3 AFLs) that is essential for the control of seed maturation (Lee et al., 2003). The Asp-55 residue found in LEC1 is of particular importance for the recruitment of LEC2 in the presence of NF-YC2. Interestingly, this amino acid is specific to

LEC1 and L1L, which are evolutionarily closely related and functionally redundant proteins (Kwong et al., 2003; Xie et al., 2008; Laloum et al., 2013).

### A Model for the Transcriptional Activation of Maturation-Related Genes by the L-AFLs

Taken together, the results of this study provide some important new insights into the molecular mechanisms underlying the partially overlapping functions of the B3 AFLs and the specific role of LEC1 in maturing seeds. They allow us to propose a model for the transcriptional activation of seed genes by LEC1, LEC2, ABI3, and NF-YC2 proteins through the formation of regulatory complexes (Fig. 7). In this model, LEC2 and ABI3 can bind RYL elements directly and independently. LEC1, which is able to form homodimers, also participates in the complex, and the amino acid Asp-55 of the protein is essential for LEC1 function through its synergistic action with LEC2 and ABI3. LEC1, as an NF-YB subunit, is expected neither to activate DNA transcription alone nor to directly bind DNA. In this respect, the shifted band obtained with LEC1 in gel-shift assays is surprising. LEC1 and L1L differ from the other NF-YB subunits and exhibit specific amino acid residues, some of which are essential for their activity. It might be that some of these residues may confer DNA-binding properties to LEC1. This aspect should be investigated further by performing a thorough structure-function analysis of the protein. Regardless of these DNA-binding capabilities, LEC1 remains unable to activate transcription by itself.

NF-YC2, which is also part of the complex, interacts with LEC1 (Calvenzani et al., 2012) and is essential for the formation of LEC1/LEC2/NF-YC2 ternary complexes in the presence of a target promoter in pull-down assays (Fig. 6). This role of NF-YC2 is consistent with its previously described involvement in the interactions between LEC1 and bZIP67 (Yamamoto et al., 2009). To explain the enhanced transcriptional activity of LEC2 and ABI3 by LEC1 observed in transient activation



**Figure 7.** Model for the transcriptional regulation of *OLE1* expression by the L-AFL proteins. In this model, LEC2 and ABI3 transcription factors directly and independently bind DNA and have a synergistic effect, together with LEC1, on the activation of the *ProOLE1* promoter. This may be due to their cooperating in a multiprotein complex, as shown by pull-down assays that unraveled the occurrence of LEC1/LEC2/NF-YC2 ternary complexes. bZIPs may play a role in this complex, although they were not tested in this study. D, Asp involved in the activity of LEC1; blue bars, RYL element; red bars, GBL element.

assays carried out in moss in the absence of AtNF-YC2, one has to consider that, whereas moss does not possess NF-YBs of the LEC1 type (containing the Asp-55 residue instead of a Lys; Kirkbride et al., 2013; Cagliari et al., 2014), NF-YA and NF-YC genes are present in moss and may substitute for their Arabidopsis counterparts. No direct evidence of direct or indirect binding between LEC1 and ABI3 could be obtained so far, despite the enhanced ABI3 activity observed in the presence of LEC1 in moss. Complementary analyses will be required to elucidate the molecular bases of this functional interaction.

A potential limit to the use of *P. patens* for performing activation studies may reside in the presence of endogenous transcription factors, such as B3s or bZIPs, that may be able to bind Arabidopsis promoter sequences and thus interfere with the candidates under study. Thirty-three B3 genes can be found in *P. patens*, seven of which belong to the ABI3-related or HIS families (Romanel et al., 2009). A gene homologous to *ABI3* was even identified that partially complements the Arabidopsis *abi3-6* mutant (Marella et al., 2006). PpABI3 was shown to interact cooperatively with PpNF-YC1 through the ACTT core element, which is different from the ATGC core of RY elements and the ACGT core of G boxes (Yotsui et al., 2013). As for bZIPs, the transcripts of which have been identified in *P. patens* (Richardt et al., 2010), it is difficult to speculate on their ability to interact with the system under study only on the basis of sequence similarities. Even if the possibility of PpB3s or PpbZIPs binding to *ProOLE* promoters cannot be ruled out (see above), the very low expression level of the reporter gene when the reporter construct was introduced alone in moss strongly suggests that the effect of endogenous *P. patens* factors alone is extremely limited.

In conclusion, this study reveals that the molecular mechanisms underpinning the transcriptional activation of maturation-related genes in seeds are much more complex than previously thought. They involve a set of protein partners assembled in a multiprotein complex, the structure of which may vary depending on the target promoter and the cell type considered. Aside from this modular organization, specific post-translational modifications of the components also may contribute to regulate the activity of the complexes and remain to be investigated.

## MATERIALS AND METHODS

### Plant Material and Growth Conditions

Arabidopsis (*Arabidopsis thaliana*) seeds of the accession Columbia were obtained from the Plant Observatory, Biological Resource Center (Institut National de la Recherche Agronomique). Plants were cultured as described previously (Baud et al., 2002). Material used for RNA extraction was frozen in liquid nitrogen immediately after harvest and then stored at  $-80^{\circ}\text{C}$ .

### Constructs for Plant and Moss Transformation

All the primers used in this work are described in Supplemental Table S1. For deletions of the *OLE1* promoter (*ProOLE1*), genomic DNA was PCR amplified

with Phusion High-Fidelity DNA Polymerase (Finnzymes) using primers carrying B1 and B2 Gateway recombination sites (Life Technologies). PCR products were cloned by BP recombination (according to the manufacturer's recommendations) into the entry vector pDONR207 and then LR recombined into the destination vector pBS Tpp-B for protoplast expression (Thévenin et al., 2012) or pBI101-R1R2-GUS for plant transformation. Mutations of the *ProOLE1* promoter cloned in pDONR207 were performed with the QuickChange Site Directed Mutagenesis kit (Agilent) according to the manufacturer's instructions. The mutations realized were as follows: GBL1\*, ACGTG→AGTTG; GBL2\*, ACGTG→AGTTG; GBL3\*, ACGTG→TTGTG; RYL1\*, TGCATGAT→TACAACAT; RYL2\*, TGCATGGT→TACAACGT; and RYL3\*, CACATGCA→CAGTTGTA (for oligonucleotide sequences, see Supplemental Table S2).

AFL transcription factors were PCR amplified from complementary DNAs (cDNAs) synthesized from RNAs extracted from developing seeds and BP recombined into pDONR207 (for oligonucleotide sequences, see Supplemental Table S1).

### Expression in Plant and in Moss Protoplasts

For plant transformation, binary vectors were electroporated into *Agrobacterium tumefaciens* strain C58C1 (pMP90) and used for agroinfiltration of Arabidopsis inflorescences (Bechtold et al., 1993). Primary transformants were selected on Murashige and Skoog medium containing kanamycin ( $50\text{ mg L}^{-1}$ ) and then transferred to soil for further characterization. For each construct, several independent transgenic lines were analyzed (at least 14 and up to 38; Fig. 2). Transient expression in the moss *Physcomitrella patens* protoplasts was performed as described (Thévenin et al., 2012). Briefly, transcription factors were recombined into a specific vector (pBS Tpp-A), allowing their expression in moss protoplasts. The promoters were recombined in a second vector (pBS Tpp-B) upstream of a *GFP* reporter gene. Clones were selected with carbenicillin ( $100\text{ }\mu\text{g mL}^{-1}$ ). Moss was cultured on solid standard PpNH4 medium (Trouiller et al., 2006).

### Cytometry Analysis

The activation studies were based on the quantification of the fluorescent signal in cotransformed protoplasts using a flow cytometer (Partec). Protoplasts were analyzed as described by Thévenin et al. (2012). The instrument was calibrated with calibration green beads (Sysmex-Partec; 05-4006). Transformed protoplasts in suspension were filtrated with a  $50\text{-}\mu\text{m}$  cell strainer to allow the analysis of 20,000 to 250,000 events. All the graphs are presented as means of at least three replicates  $\pm$  sd. The R stats package (<https://www.r-project.org/>) version 3.2.1 was used for statistical analysis (R Core Team, 2015).

### Microscopy and Image Analysis

Microscopic observations of seeds and embryos were carried out as described previously (Baud et al., 2007). For histochemical detection of GUS activity, tissues were incubated in  $0.1\text{ M}$  phosphate buffer, pH 7.2, containing  $2\text{ mM}$  5-bromo-3-indolyl- $\beta$ -D-glucuronide (Duchefa),  $0.1\%$  (v/v) Triton X-100:water,  $2\text{ mM}$  each potassium ferrocyanide and potassium ferricyanide, and  $10\text{ mM}$   $\text{Na}_2\text{-EDTA}$ . A vacuum was applied for 1 h before incubating for 3 h at  $37^{\circ}\text{C}$  in the dark. After incubation, the staining solution was removed and samples were cleared by sequential changes of 75% and 95% (v/v) ethanol.

### EMSA

The probe used comprised three GBL (underlined) and three RYL (boldface) domains: 5'-CAAACGTGCTTTGAATAGACTCCTCTCGTTAACACATG-CAGCGGCTGCATGGTGACGCCATTAACACGTGGCCTACAATTGCATG-ATGTCCTCATTGACACGTGACT-3'. It was obtained by annealing two 80-bp-long oligonucleotides exhibiting a 48-bp overlap that were 5' biotinylated and synthesized by the Eurofins Company (5'-ggCAAACGTGCTTTGAATA-GACTCCTCTCGTTAACACATGCAGCGGCTGCATGGTGACGCCATTAACACGTGGCCTACA-3' and 5'-ggAGTCACGTGTCATGGAGACATCATGCAA-TTGTAGGCCACGTGTTAATGGCGTCACCATGCAGCCGCTGCATGTGTTA-3'). The oligonucleotides were diluted and annealed for 5 min at  $95^{\circ}\text{C}$  in a PCR thermal block. The sequences were then double stranded by the action of Klenow enzyme (Klenow fragment; Thermo Scientific) for 1 h at room temperature in order to obtain a 108-bp double-stranded biotinylated probe at a concentration of  $20\text{ fmol }\mu\text{L}^{-1}$ . Binding reactions were performed with the

Shift Light Chemiluminescent EMSA kit (Thermo Scientific) and carried out with a binding buffer (10 mM Tris, 150 mM KCl, 1 mM dithiothreitol, 0.05% [v/v] Nonidet P-40, and 0.5  $\mu$ g of poly [dI-dC], pH 7.5). One to 6  $\mu$ g of proteins was incubated for 5 min on ice in binding buffer before adding 20 fmol of biotinylated probe. The binding reactions were then incubated at room temperature for 20 min in the dark. The competition reactions were performed with a 25-fold increase of unlabeled probe. The electrophoresis gel (5% acrylamide in 0.25% [v/v] glycerol and 0.5 $\times$  Tris-borate/EDTA) was subjected to a 60-min prerun at 120 V, and the binding reactions were loaded onto the gel with 5  $\mu$ L of loading buffer for 6 h of migration at 120 V and 4°C. The gel was transferred to a GeneScreen plus DNA membrane (Perkin-Elmer Life Sciences) for 1 h at 25 V with 0.5 $\times$  Tris-borate/EDTA using a Transblot SD Semi Dry transfer cell system (Bio-Rad). DNA was cross-linked on the membrane with a transilluminator for 60 s at 254 nm with the autocrosslink function. The bands were revealed by chemiluminescence according to instructions of the Chemiluminescent Nucleic Acid Detection module kit (Thermo Scientific) using the Fujifilm Las 4000 imaging system.

### Expression and Purification of Recombinant Proteins

*LEC1* and *LEC2* coding sequences were PCR amplified from AFLs in pDONR207 using primers carrying 5' *NcoI* and 3' *SalI* restriction sites (for primer sequences, see Supplemental Table S1). PCR products were ligated in the expression vector pETtrx-1a (Millipore) at the *NcoI/XhoI* sites with an N-terminal His-thioredoxin tag inserted in the oligonucleotide sequence. *ABI3* in pDONR207 was directly LR recombined into direpDest17 (Life Technologies) with an N-terminal His tag inserted in the oligonucleotide sequence. Proteins were produced in *Escherichia coli* Rosetta [DE3, F<sup>-</sup> *ompT hsdS<sub>B</sub>(r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) gal dcm* (DE3) pRARE (Cam<sup>R</sup>)] and induced with 0.5 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) at 28°C for 3 h, 17°C for 3 h, and 17°C overnight. Bacterial pellets were sonicated in lysis buffer (20 mM Tris, 500 mM NaCl, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 10% [v/v] glycerol, and 5 mM imidazole, pH 8). Lysates were centrifuged at 21,000g for 45 min at 4°C. Soluble proteins were purified on nickel-nitrilotriacetic acid agarose (Ni-NTA) resin (Bio-Rad) in lysis buffer with increasing concentrations of imidazole (5–70 mM for washing and 300 mM for elution). Purified proteins were then dialyzed in buffer (0.05 M phosphate, pH 7.2, 0.15 M NaCl, and 20% [v/v] glycerol) at 4°C, concentrated on Amicon cell 10,000 MWCO (Millipore), and finally quantified by the Bradford method (Bradford, 1976). After purification, proteins were stored at –20°C before subsequent analyses.

### Coimmunoprecipitation Experiments

The *HindIII/Stul* fragments of the pGWB15 and pGWB18 vectors (Nakagawa et al., 2007), containing promoter 35S/3 $\times$  HA epitope/Gateway recombination cassette/terminator and promoter 35S/4 $\times$  Myc epitope/Gateway recombination cassette/terminator sequence, respectively, were cloned into *HindIII/SmaI* sites of the pBluescript II KS+ vector (Stratagene) and used as destination vectors (pBS 35SHA-R1R2 and pBS 35MYC-R1R2) for coimmunoprecipitation experiments. *LEC2*, *LEC1*, and *ABI3* in pDONR207 were recombined directly into these destination vectors. NF-YC2 coding sequence was PCR amplified using primers NF-YC2 F GW and NF-YC2 R GW (for primer sequences, see Supplemental Table S1) from an Arabidopsis Columbia cDNA library and recombined directly into pDONR207 using BP Clonase, according to the manufacturer's instructions. The cDNA was then further transferred into the destination vector using LR Clonase.

Cell suspension culture of Arabidopsis ecotype Columbia, subcultured weekly in JPL medium (Takahashi et al., 2001), was used for the preparation of protoplasts and transformations. Protoplasts were isolated and transfected by the polyethylene glycol-mediated method (Kiegerl et al., 2000). After a 16-h incubation, cells were harvested by centrifugation, and proteins were extracted in an extraction buffer containing 25 mM Tris HCl, pH 7.7, 10 mM MgCl<sub>2</sub>, 15 mM EGTA, 75 mM NaCl, 15 mM  $\beta$ -glycerophosphate, 0.1% (v/v) Tween 20, 10% (v/v) glycerol, 1 mM dithiothreitol, 1 mM NaF, 0.5 mM NaVO<sub>3</sub>, 0.5 mM phenylmethylsulfonyl fluoride, and protease inhibitors (Complete; Roche). From 500  $\mu$ g of total protein extract of transfected Arabidopsis protoplasts, HA-tagged proteins were immunoprecipitated with 500 ng of 3F10 anti-HA monoclonal antibody (Roche) and washed three times with the extraction buffer and once with 20 mM Tris-HCl, pH 7.5. Proteins bound to the beads were eluted with 50  $\mu$ L of SDS loading buffer. Coimmunoprecipitation of the Myc-tagged proteins was revealed by SDS-PAGE followed by western blot using the 9E10 anti-Myc monoclonal antibody (Roche Life Sciences) and ECL detection (Amersham). The efficiency of the immunoprecipitation was tested using HA-probe (Y-11) polyclonal antibody (Santa Cruz Biotech).

### Yeast Experiments

*LEC2*, *LEC1*, and *ABI3* in pDONR207 were directly LR recombined into pDEST22 and pDEST32 vectors (Life Technologies) and stably cotransformed into yeast. The yeast AH109 strain (Clontech) was grown on YPDA plates (Clontech) at 28°C. A 2-d-old culture was collected and resuspended in 1 mL of sterile water. The yeast was pelleted by centrifugation (at maximum speed for 15 s) and resuspended in 1 mL of 0.1 M lithium acetate in 1 $\times$  Tris-EDTA solution. Five milligrams of each plasmid construct was mixed with 5  $\mu$ L of carrier DNA (Finnzymes, Fisher Scientific), 0.1 mL of yeast cell solution, and 0.6 mL of 40% [v/v] polyethylene glycol 4000/0.1 M lithium acetate/1 $\times$  Tris-EDTA solution. After a 30-min incubation at 28°C, the transformation mixture was heat shocked at 42°C for 25 min. After centrifugation (at maximum speed for 15 s), the pellet was resuspended in 250  $\mu$ L of sterile water and plated onto synthetic dextrose-Trp-Leu medium (Clontech). To analyze interactions, cotransformed yeast cells were transferred to 100  $\mu$ L of sterile water. Growth of a 5- $\mu$ L yeast solution was tested on selective medium. Colonies growing on medium lacking His or both His and Ala, and in the presence of various concentrations of 3-aminotriazole, were considered as positive interactions.

### Systematic Evolution of Ligands by Exponential Enrichment Experiments (SELEX) for the LEC2 Protein

*LEC2* coding sequence was cloned in pET-trx 1a. *LEC2* protein was expressed using *E. coli* strain Rosetta Blu (DE3). After induction by 0.5 mM IPTG, cells were grown overnight at 17°C under shaking. For cell lysis, the pellet of 1 L of culture was sonicated in 35 mL of loading buffer (500 mM NaCl, 20 mM Tris-HCl, pH 8, 5 mM imidazole, 5% [v/v] glycerol, and 5 mM tris(2-carboxyethyl)phosphine (TCEP) with one protease inhibitor cocktail tablet (Complete EDTA-free; Roche) and centrifuged for 40 min at 16,000g. The clear supernatant was incubated for 1 h with 1 mL of Ni-NTA resin (Qiagen). The resin was transferred into a column, washed with 15 mL of loading buffer and then 15 mL of washing buffer (loading buffer with 20 mM imidazole), and eluted with 10 mL of elution buffer (loading buffer with 300 mM imidazole). The protein concentration of each fraction was estimated using the Bradford assay (Bradford, 1976).

For SELEX, a random library was first synthesized using the 73-mer 5'-TGGAGAAGAGGAGATCTAGC(N)<sub>30</sub>CTGTTCCTCTTCGATCCGG-3' as a template with a fluorescent TAMRA-labeled forward primer (SElex-F) and a nonlabeled reverse primer (SElex-R), as described previously (Moyroud et al., 2011).

For each selection cycle, 1  $\mu$ M protein was mixed with 10 nM fluorescent double-stranded DNA in 225  $\mu$ L of SELEX buffer (25 mM HEPES, pH 8, 1 mg mL<sup>-1</sup> bovine serum albumin, 75 mM KCl, 10% glycerol, and 5 mM tris(2-carboxyethyl)phosphine [TCEP]) with 60  $\mu$ g mL<sup>-1</sup> fish sperm DNA added. After a 15-min incubation under shaking at 4°C, 25- $\mu$ L Ni-NTA magnetic beads (Qiagen), previously equilibrated in SELEX buffer, were added to the reaction mix to immobilize DNA/protein complexes via the His tag of the protein. After 30 min of incubation under shaking at 4°C, the reaction mix was placed on a tube magnet for 1 min, and the supernatant was removed from the separated beads to eliminate the unbound DNA. Six washes were performed subsequently at 4°C, each of them consisting of first adding 50  $\mu$ L of SELEX buffer and then 250  $\mu$ L of SELEX buffer containing 20  $\mu$ g mL<sup>-1</sup> fish sperm DNA. Between the two buffer additions, an aliquot was taken to perform PCR (see below). Each wash was followed by a 2-min incubation and 1 min on the tube magnet to discard the supernatant, and finally, the magnetic beads were resuspended in 50  $\mu$ L of SELEX buffer. Selected 73-mers were amplified by PCR as described previously (Moyroud et al., 2011) using 1  $\mu$ L of the magnetic beads solution as a template. PCR products were quantified as described previously (Moyroud et al., 2011), and the selection cycle was repeated two times, each time using the newly synthesized fluorescent DNA as a library. The 73-mer libraries obtained after three cycles of selection were used subsequently for barcoding and high-throughput sequencing as described previously (Chahtane et al., 2013). Sequences were finally analyzed with MEME software (<http://meme.nbrnet/meme/cgi-bin/meme.cgi>) to produce a PWM.

### Protein-Binding Microarrays on ABI3-B3 and FUS3

The *FUS3* coding sequence was PCR amplified using Phusion High Fidelity DNA Polymerase (Thermo Scientific) and the oSB110R/oETH132 primers. The amplified PCR product was subcloned into Zero Blunt (Life Technologies) and further transferred into pETM-40 (Dümmeler et al., 2005) at *NcoI-XhoI* restriction sites to yield pETH221 expression vector, allowing the



expression of FUS3 fused to a maltose-binding protein tag followed by a tobacco etch virus cleavage site.

The ABI3 B3 domain coding sequence was PCR amplified using Phusion High Fidelity DNA Polymerase (Thermo Scientific) and the oSB109F/oSB109R primers. The amplified PCR product was subcloned into Zero Blunt (Life Technologies) and further transferred into pMAL-C5x (New England Biolabs) at *NdeI*-*NotI* restriction sites to yield pETH214. This expression vector allows the expression of the ABI3 B3 domain fused to a maltose-binding protein tag followed by a Factor-Xa cleavage site.

These two proteins were expressed using *E. coli* strain Rosetta 2 (DE3). After induction by 0.4 mM IPTG, cells were grown overnight at 17°C, and soluble protein extracts were used for protein-binding microarray studies as described previously (Godoy et al., 2011).

## Supplemental Data

The following supplemental materials are available.

**Supplemental Figure S1.** E-score distribution of top four 6mers recognized by FUS3 and ABI3.

**Supplemental Figure S2.** Schematic representation of the mutagenized versions of the *OLE1* promoter used in this study.

**Supplemental Figure S3.** Activation of the native or mutated *OLE1* promoter versions by LEC2 in moss protoplasts.

**Supplemental Figure S4.** Complementary analyses related to activation studies in moss protoplasts.

**Supplemental Figure S5.** Test of synergistic effect of LEC2, LEC1, and ABI3 on the activation of several oleosin promoters in moss protoplasts.

**Supplemental Figure S6.** In vitro binding of the L-AFLs to the *OLE1* promoter.

**Supplemental Figure S7.** Complementary results to pull-down experiments.

**Supplemental Figure S8.** Study of protein-protein interactions in yeast.

**Supplemental Table S1.** Primers and genes used in this study.

**Supplemental Table S2.** List of primers used for site-directed mutagenesis of *ProOLE1*.

**Supplemental Table S3.** Statistical analyses.

## ACKNOWLEDGMENTS

We thank the Observatoire du Végétal for plant culture, access to the imaging facility, and assistance.

Received January 22, 2016; accepted April 7, 2016; published April 12, 2016.

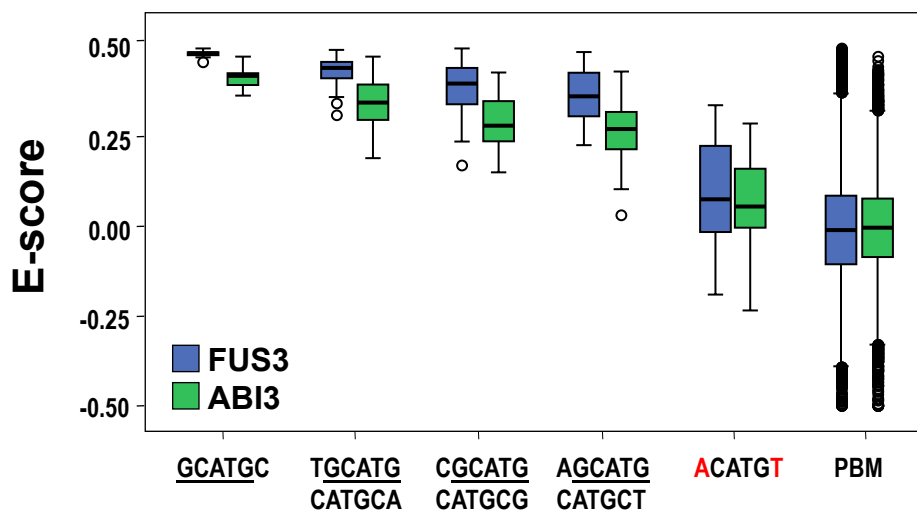
## LITERATURE CITED

- Alonso R, Oñate-Sánchez L, Weltmeier F, Ehlert A, Diaz I, Dietrich K, Vicente-Carbajosa J, Dröge-Laser W (2009) A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of *Arabidopsis* seed maturation gene expression based on heterodimerization and protein complex formation. *Plant Cell* **21**: 1747–1761
- Barthole G, To A, Marchive C, Brunaud V, Soubigou-Taconnat L, Berger N, Dubreucq B, Lepiniec L, Baud S (2014) MYB118 represses endosperm maturation in seeds of *Arabidopsis*. *Plant Cell* **26**: 3519–3537
- Baud S, Boutin JP, Miquel M, Lepiniec L, Rochat C (2002) An integrated overview of seed development in *Arabidopsis thaliana* ecotype Ws. *Plant Physiol Biochem* **40**: 151–160
- Baud S, Dichow NR, Kelemen Z, d'Andréa S, To A, Berger N, Canonge M, Kronenberger J, Viterbo D, Dubreucq B, et al (2009) Regulation of HSD1 in seeds of *Arabidopsis thaliana*. *Plant Cell Physiol* **50**: 1463–1478
- Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B (2007) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*. *Plant J* **50**: 825–838
- Bäumlein H, Nagy I, Villarroel R, Inzé D, Wobus U (1992) 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* **2**: 233–239
- Bäumlein H, Pustell J, Wobus U, Case ST, Kafatos FC (1986) The 3' ends of two genes in the Balbiani ring c locus of *Chironomus thummi*. *J Mol Evol* **24**: 72–82
- Bechtold N, Ellis J, Pelletier G (1993) In planta *Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *CR Acad Sci Paris* **316**: 1194–1199
- Bensmihen S, Ripka S, Lambert G, Jublot D, Pautot V, Granier F, Giraudat J, Parcy F (2002) The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. *Plant Cell* **14**: 1391–1403
- Berger N, Dubreucq B, Roudier F, Dubos C, Lepiniec L (2011) Transcriptional regulation of *Arabidopsis* LEAFY COTYLEDON2 involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *Plant Cell* **23**: 4065–4078
- Bobb AJ, Chern MS, Bustos MM (1997) Conserved RY-repeats mediate transactivation of seed-specific promoters by the developmental regulator PvALF. *Nucleic Acids Res* **25**: 641–647
- Boer DR, Freire-Rios A, van den Berg WA, Saaki T, Manfield IW, Kepinski S, López-Vidriero I, Franco-Zorrilla JM, de Vries SC, Solano R, et al (2014) Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell* **156**: 577–589
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248–254
- Braybrook SA, Harada JJ (2008) LECs go crazy in embryo development. *Trends Plant Sci* **13**: 624–630
- Braybrook SA, Stone SL, Park S, Bui AQ, Le BH, Fischer RL, Goldberg RB, Harada JJ (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc Natl Acad Sci USA* **103**: 3468–3473
- Brocard-Gifford IM, Lynch TJ, Finkelstein RR (2003) Regulatory networks in seeds integrating developmental, abscisic acid, sugar, and light signaling. *Plant Physiol* **131**: 78–92
- Cagliari A, Turchetto-Zolet AC, Korbes AP, Maraschin FdS, Margis R, Margis-Pinheiro M (2014) New insights on the evolution of Leafy cotyledon1 (LEC1) type genes in vascular plants. *Genomics* **103**: 380–387
- Calvenzani V, Testoni B, Gusmaroli G, Lorenzo M, Gnesutta N, Petroni K, Mantovani R, Tonelli C (2012) Interactions and CCAAT-binding of *Arabidopsis thaliana* NF-Y subunits. *PLoS ONE* **7**: e42902
- Chahtane H, Vachon G, Le Masson M, Thévenon E, Périgon S, Mihajlovic N, Kalinina A, Michard R, Moyroud E, Monniaux M, et al (2013) A variant of LEAFY reveals its capacity to stimulate meristem development by inducing RAX1. *Plant J* **74**: 678–689
- Che N, Yang Y, Li Y, Wang L, Huang P, Gao Y, An C (2009) Efficient LEC2 activation of OLEOSIN expression requires two neighboring RY elements on its promoter. *Sci China C Life Sci* **52**: 854–863
- Conceição AS, Krebbers E (1994) A cotyledon regulatory region is responsible for the different spatial expression patterns of *Arabidopsis* 2S albumin genes. *Plant J* **5**: 493–505
- Dickinson CD, Evans RP, Nielsen NC (1988) RY repeats are conserved in the 5'-flanking regions of legume seed-protein genes. *Nucleic Acids Res* **16**: 371
- Dümmeler A, Lawrence AM, de Marco A (2005) Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb Cell Fact* **4**: 34
- Ellerström M, Ståhlberg K, Ezcurra I, Rask L (1996) Functional dissection of a napin gene promoter: identification of promoter elements required for embryo and endosperm-specific transcription. *Plant Mol Biol* **32**: 1019–1027
- Ezcurra I, Ellerström M, Wycliffe P, Ståhlberg K, Rask L (1999) Interaction between composite elements in the napA promoter: both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol Biol* **40**: 699–709
- Ezcurra I, Wycliffe P, Nehlin L, Ellerström M, Rask L (2000) Transactivation of the *Brassica napus* napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different cis-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. *Plant J* **24**: 57–66
- Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R (2014) DNA-binding specificities of plant transcription factors

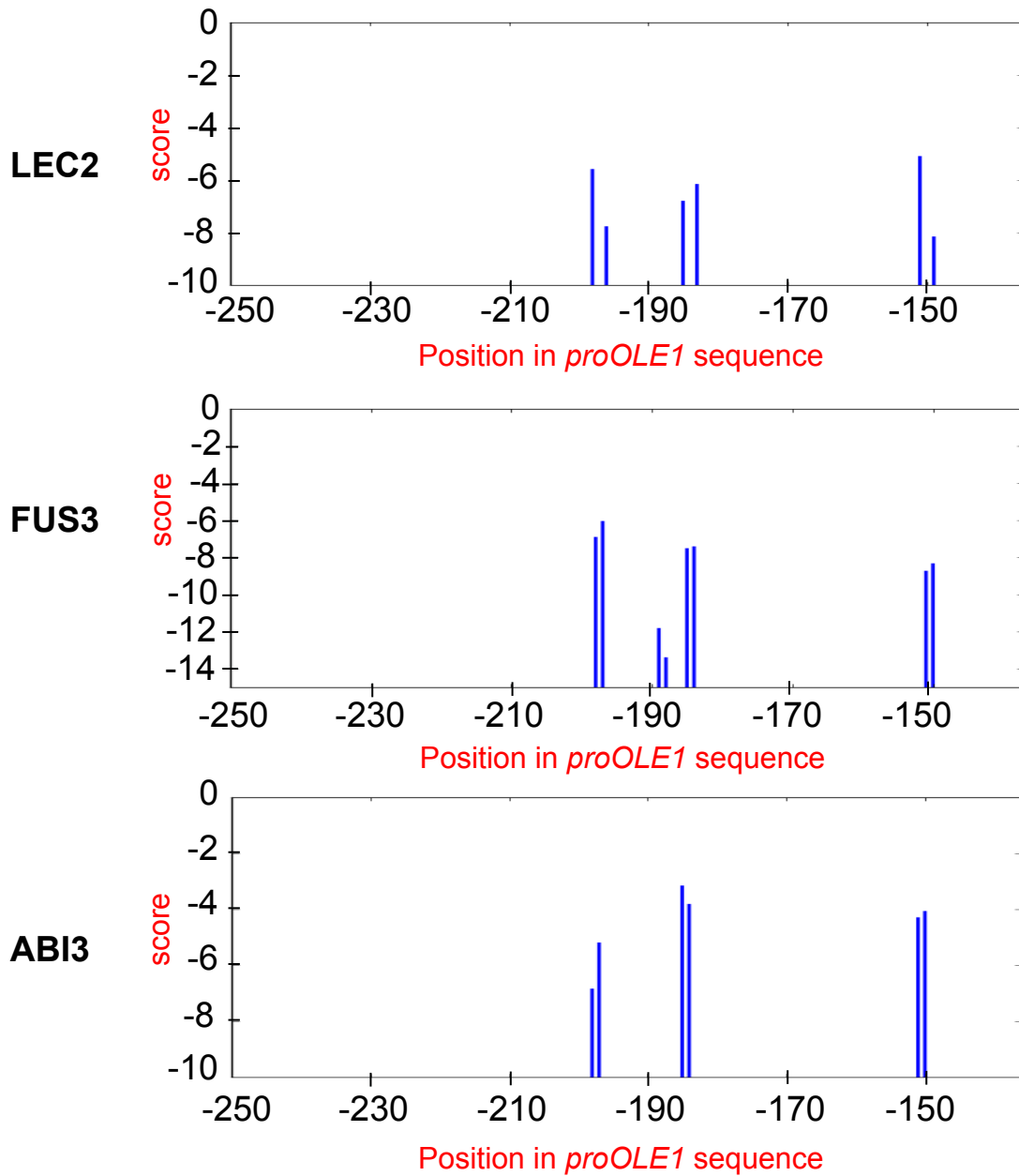
- and their potential to define target genes. *Proc Natl Acad Sci USA* **111**: 2367–2372
- Fujiwara T, Beachy RN** (1994) Tissue-specific and temporal regulation of a beta-conglycinin gene: roles of the RY repeat and other cis-acting elements. *Plant Mol Biol* **24**: 261–272
- Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM** (1992) Isolation of the *Arabidopsis* ABI3 gene by positional cloning. *Plant Cell* **4**: 1251–1261
- Godoy M, Franco-Zorrilla JM, Pérez-Pérez J, Oliveros JC, Lorenzo O, Solano R** (2011) Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. *Plant J* **66**: 700–711
- Golovenko D, Manakova E, Zakrys L, Zaremba M, Sasnauskas G, Grazulis S, Siksnys V** (2014) Structural insight into the specificity of the B3 DNA-binding domains provided by the co-crystal structure of the C-terminal fragment of BfII restriction enzyme. *Nucleic Acids Res* **42**: 4113–4122
- Gusmaroli G, Tonelli C, Mantovani R** (2001) Regulation of the CCAAT-binding NF-Y subunits in *Arabidopsis thaliana*. *Gene* **264**: 173–185
- Gusmaroli G, Tonelli C, Mantovani R** (2002) Regulation of novel members of the *Arabidopsis thaliana* CCAAT-binding nuclear factor Y subunits. *Gene* **283**: 41–48
- Hackenberg D, Wu Y, Voigt A, Adams R, Schramm P, Grimm B** (2012) Studies on differential nuclear translocation mechanism and assembly of the three subunits of the *Arabidopsis thaliana* transcription factor NF-Y. *Mol Plant* **5**: 876–888
- Harada JJ** (2001) Role of *Arabidopsis* LEAFY COTYLEDON genes in seed development. *J Plant Physiol* **158**: 405–409
- Hilioti Z, Ganopoulos I, Bossis I, Tsaftaris A** (2014) LEC1-LIKE paralog transcription factor: how to survive extinction and fit in NF-Y protein complex. *Gene* **543**: 220–233
- Hobo T, Asada M, Koyama Y, Hattori T** (1999) ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. *Plant J* **19**: 679–689
- Jia H, McCarty DR, Suzuki M** (2013) Distinct roles of LAFL network genes in promoting the embryonic seedling fate in the absence of VAL repression. *Plant Physiol* **163**: 1293–1305
- Kagaya Y, Okuda R, Ban A, Toyoshima R, Tsutsumida K, Usui H, Yamamoto A, Hattori T** (2005a) Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in *Arabidopsis*. *Plant Cell Physiol* **46**: 300–311
- Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, Hattori T** (2005b) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol* **46**: 399–406
- Kiegl S, Cardinale F, Siligan C, Gross A, Baudouin E, Liwosz A, Eklöf S, Till S, Bögre L, Hirt H, et al** (2000) SIMKK, a mitogen-activated protein kinase (MAPK) kinase, is a specific activator of the salt stress-induced MAPK, SIMK. *Plant Cell* **12**: 2247–2258
- Kirkbride RC, Fischer RL, Harada JJ** (2013) LEAFY COTYLEDON1, a key regulator of seed development, is expressed in vegetative and sexual propagules of *Selaginella moellendorffii*. *PLoS ONE* **8**: e67971
- Koornneef M, Reuling G, Karssen CM** (1984) The isolation and characterization of abscisic acid insensitive mutants of *Arabidopsis thaliana*. *Physiol Plant* **61**: 377–383
- Kroj T, Savino G, Valon C, Giraudat J, Parcy F** (2003) Regulation of storage protein gene expression in *Arabidopsis*. *Development* **130**: 6065–6073
- Kurup S, Jones HD, Holdsworth MJ** (2000) Interactions of the developmental regulator ABI3 with proteins identified from developing *Arabidopsis* seeds. *Plant J* **21**: 143–155
- Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ** (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* **15**: 5–18
- Laloum T, De Mita S, Gamas P, Baudin M, Niebel A** (2013) CCAAT-box binding transcription factors in plants: Y so many? *Trends Plant Sci* **18**: 157–166
- Lara P, Oñate-Sánchez L, Abraham Z, Ferrándiz C, Díaz I, Carbonero P, Vicente-Carbajosa J** (2003) Synergistic activation of seed storage protein gene expression in *Arabidopsis* by ABI3 and two bZIPs related to OPAQUE2. *J Biol Chem* **278**: 21003–21011
- Lee H, Fischer RL, Goldberg RB, Harada JJ** (2003) *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci USA* **100**: 2152–2156
- Li C, Distelfeld A, Comis A, Dubcovsky J** (2011) Wheat flowering repressor VRN2 and promoter CO2 compete for interactions with NUCLEAR FACTOR-Y complexes. *Plant J* **67**: 763–773
- Liu JX, Howell SH** (2010) bZIP28 and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in *Arabidopsis*. *Plant Cell* **22**: 782–796
- Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ** (1998) *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* **93**: 1195–1205
- Luerssen H, Kirik V, Herrmann P, Miséra S** (1998) FUSCA3 encodes a protein with a conserved VP1/ABI3-like B3 domain which is of functional importance for the regulation of seed maturation in *Arabidopsis thaliana*. *Plant J* **15**: 755–764
- Mantovani R** (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene* **239**: 15–27
- Marella HH, Sakata Y, Quatrano RS** (2006) Characterization and functional analysis of ABSCISIC ACID INSENSITIVE3-like genes from *Physcomitrella patens*. *Plant J* **46**: 1032–1044
- Masiero S, Imbriano C, Ravasio F, Favaro R, Pelucchi N, Gorla MS, Mantovani R, Colombo L, Kater MM** (2002) Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. *J Biol Chem* **277**: 26429–26435
- Meinke DW** (1992) A homeotic mutant of *Arabidopsis thaliana* with leafy cotyledons. *Science* **258**: 1647–1650
- Meinke DW, Franzmann LH, Nickle TC, Yeung EC** (1994) Leafy cotyledon mutants of *Arabidopsis*. *Plant Cell* **6**: 1049–1064
- Mendes A, Kelly AA, van Erp H, Shaw E, Powers SJ, Kurup S, Eastmond PJ** (2013) bZIP67 regulates the omega-3 fatty acid content of *Arabidopsis* seed oil by activating fatty acid desaturase3. *Plant Cell* **25**: 3104–3116
- Mönke G, Altschmied L, Tewes A, Reidt W, Mock HP, Bäumlein H, Conrad U** (2004) Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. *Planta* **219**: 158–166
- Mönke G, Seifert M, Keilwagen J, Mohr M, Grosse I, Hähnel U, Junker A, Weisshaar B, Conrad U, Bäumlein H, et al** (2012) Toward the identification and regulation of the *Arabidopsis thaliana* ABI3 regulon. *Nucleic Acids Res* **40**: 8240–8254
- Motta MC, Caretti G, Badaracco GF, Mantovani R** (1999) Interactions of the CCAAT-binding trimer NF-Y with nucleosomes. *J Biol Chem* **274**: 1326–1333
- Moyroud E, Minguet EG, Ott F, Yant L, Posé D, Monniaux M, Blanchet S, Bastien O, Thévenon E, Weigel D, et al** (2011) Prediction of regulatory interactions from genome sequences using a biophysical model for the *Arabidopsis* LEAFY transcription factor. *Plant Cell* **23**: 1293–1306
- Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T** (2007) Development of series of Gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* **104**: 34–41
- Nakamura S, Lynch TJ, Finkelstein RR** (2001) Physical interactions between ABA response loci of *Arabidopsis*. *Plant J* **26**: 627–635
- Nakashima K, Fujita Y, Katsura K, Maruyama K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K** (2006) Transcriptional regulation of ABI3- and ABA-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of *Arabidopsis*. *Plant Mol Biol* **60**: 51–68
- Parcy F, Nilsson O, Busch MA, Lee I, Weigel D** (1998) A genetic framework for floral patterning. *Nature* **395**: 561–566
- Parcy F, Valon C, Kohara A, Miséra S, Giraudat J** (1997) The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of *Arabidopsis* seed development. *Plant Cell* **9**: 1265–1277
- Petroni K, Kumimoto RW, Gnesutta N, Calvenzani V, Fornari M, Tonelli C, Holt BF III, Mantovani R** (2012) The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. *Plant Cell* **24**: 4777–4792
- R Core Team** (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna
- Reidt W, Wohlfarth T, Ellerström M, Czihal A, Tewes A, Ezcurra I, Rask L, Bäumlein 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

- Reinders A, Schulze W, Kühn C, Barker L, Schulz A, Ward JM, Frommer WB (2002) Protein-protein interactions between sucrose transporters of different affinities colocalized in the same enucleate sieve element. *Plant Cell* **14**: 1567–1577
- Richardt S, Timmerhaus G, Lang D, Qudeimat E, Corrêa LG, Reski R, Rensing SA, Frank W (2010) Microarray analysis of the moss *Physcomitrella patens* reveals evolutionarily conserved transcriptional regulation of salt stress and abscisic acid signalling. *Plant Mol Biol* **72**: 27–45
- Romanel EA, Schrago CG, Couñago RM, Russo CA, Alves-Ferreira M (2009) Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. *PLoS ONE* **4**: e5791
- Romier C, Cocchiarella F, Mantovani R, Moras D (2003) The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J Biol Chem* **278**: 1336–1345
- Ronchi A, Bellorini M, Mongelli N, Mantovani R (1995) CCAAT-box binding protein NF-Y (CBF, CP1) recognizes the minor groove and distorts DNA. *Nucleic Acids Res* **23**: 4565–4572
- Roscoe TT, Guillemot J, Bessoule JJ, Berger F, Devic M (2015) Complementation of seed maturation phenotypes by ectopic expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in *Arabidopsis*. *Plant Cell Physiol* **56**: 1215–1228
- Sadowski I, Ma J, Triezenberg S, Ptashne M (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**: 563–564
- Sakata Y, Chiba Y, Fukushima H, Matsubara N, Habu Y, Naito S, Ohno T (1997) The RY sequence is necessary but not sufficient for the transcription activation of a winged bean chymotrypsin inhibitor gene in developing seeds. *Plant Mol Biol* **34**: 191–197
- Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L (2008) Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *Plant J* **54**: 608–620
- Santos Mendoza M, Dubreucq B, Miquel M, Caboche M, Lepiniec L (2005) LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in *Arabidopsis* leaves. *FEBS Lett* **579**: 4666–4670
- Shen B, Sinkevicius KW, Selinger DA, Tarczynski MC (2006) The homeobox gene GLABRA2 affects seed oil content in *Arabidopsis*. *Plant Mol Biol* **60**: 377–387
- Sreenivasulu N, Wobus U (2013) Seed-development programs: a systems biology-based comparison between dicots and monocots. *Annu Rev Plant Biol* **64**: 189–217
- Stålberg K, Ellerström M, Josefsson LG, Rask L (1993) Deletion analysis of a 2S seed storage protein promoter of *Brassica napus* in transgenic tobacco. *Plant Mol Biol* **23**: 671–683
- Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc Natl Acad Sci USA* **98**: 11806–11811
- Suzuki M, Kao CY, McCarty DR (1997) The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. *Plant Cell* **9**: 799–807
- Suzuki M, McCarty DR (2008) Functional symmetry of the B3 network controlling seed development. *Curr Opin Plant Biol* **11**: 548–553
- Takahashi S, Katagiri T, Hirayama T, Yamaguchi-Shinozaki K, Shinozaki K (2001) Hyperosmotic stress induces a rapid and transient increase in inositol 1,4,5-trisphosphate independent of abscisic acid in *Arabidopsis* cell culture. *Plant Cell Physiol* **42**: 214–222
- Thakare D, Tang W, Hill K, Perry SE (2008) The MADS-domain transcriptional regulator AGAMOUS-LIKE15 promotes somatic embryo development in *Arabidopsis* and soybean. *Plant Physiol* **146**: 1663–1672
- Thévenin J, Dubos C, Xu W, Le Gourrierec J, Kelemen Z, Charlot F, Nogué F, Lepiniec L, Dubreucq B (2012) A new system for fast and quantitative analysis of heterologous gene expression in plants. *New Phytol* **193**: 504–512
- To A, Valon C, Savino G, Guillemot J, Devic M, Giraudat J, Parcy F (2006) A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. *Plant Cell* **18**: 1642–1651
- Trouiller B, Schaefer DG, Charlot F, Nogué F (2006) MSH2 is essential for the preservation of genome integrity and prevents homeologous recombination in the moss *Physcomitrella patens*. *Nucleic Acids Res* **34**: 232–242
- Vicente-Carbajosa J, Carbonero P (2005) Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int J Dev Biol* **49**: 645–651
- Wang F, Perry SE (2013) Identification of direct targets of FUSCA3, a key regulator of *Arabidopsis* seed development. *Plant Physiol* **161**: 1251–1264
- Wang X, Niu QW, Teng C, Li C, Mu J, Chua NH, Zuo J (2009) Overexpression of PGA37/MYB118 and MYB115 promotes vegetative-to-embryonic transition in *Arabidopsis*. *Cell Res* **19**: 224–235
- Wenck S, Turck F, Singer K, Gissot L, Le Gourrierec J, Samach A, Coupland G (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell* **18**: 2971–2984
- Wilde RJ, Cooke SE, Brammar WJ, Schuch W (1994) Control of gene expression in plant cells using a 434:VP16 chimeric protein. *Plant Mol Biol* **24**: 381–388
- Wobus U, Weber H (1999) Seed maturation: genetic programmes and control signals. *Curr Opin Plant Biol* **2**: 33–38
- Xie Z, Li X, Glover BJ, Bai S, Rao GY, Luo J, Yang J (2008) Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes. *Mol Biol Evol* **25**: 1581–1592
- Xu W, Lepiniec L, Dubos C (2014) New insights toward the transcriptional engineering of proanthocyanidin biosynthesis. *Plant Signal Behav* **9**: 9
- Yamagishi K, Nagata N, Yee KM, Braybrook SA, Pelletier J, Fujioka S, Yoshida S, Fischer RL, Goldberg RB, Harada JJ (2005) TANMEI/EMB2757 encodes a WD repeat protein required for embryo development in *Arabidopsis*. *Plant Physiol* **139**: 163–173
- Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T (2009) *Arabidopsis* NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J* **58**: 843–856
- Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, et al (2004) Solution structure of the B3 DNA binding domain of the *Arabidopsis* cold-responsive transcription factor RAV1. *Plant Cell* **16**: 3448–3459
- Yotsui I, Saruhashi M, Kawato T, Taji T, Hayashi T, Quatrano RS, Sakata Y (2013) ABSCISIC ACID INSENSITIVE3 regulates abscisic acid-responsive gene expression with the nuclear factor Y complex through the ACTT-core element in *Physcomitrella patens*. *New Phytol* **199**: 101–109
- Zahn LM, Leebens-Mack J, DePamphilis CW, Ma H, Theissen G (2005) To B or not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *J Hered* **96**: 225–240
- Zhang H, Ogas J (2009) An epigenetic perspective on developmental regulation of seed genes. *Mol Plant* **2**: 610–627
- Zhang W, Ruan J, Ho TD, You Y, Yu T, Quatrano RS (2005) Cis-regulatory element based targeted gene finding: genome-wide identification of ABA- and abiotic stress-responsive genes in *Arabidopsis thaliana*. *Bioinformatics* **21**: 3074–3081
- Zheng Y, Ren N, Wang H, Stromberg AJ, Perry SE (2009) Global identification of targets of the *Arabidopsis* MADS domain protein AGAMOUS-Like15. *Plant Cell* **21**: 2563–2577

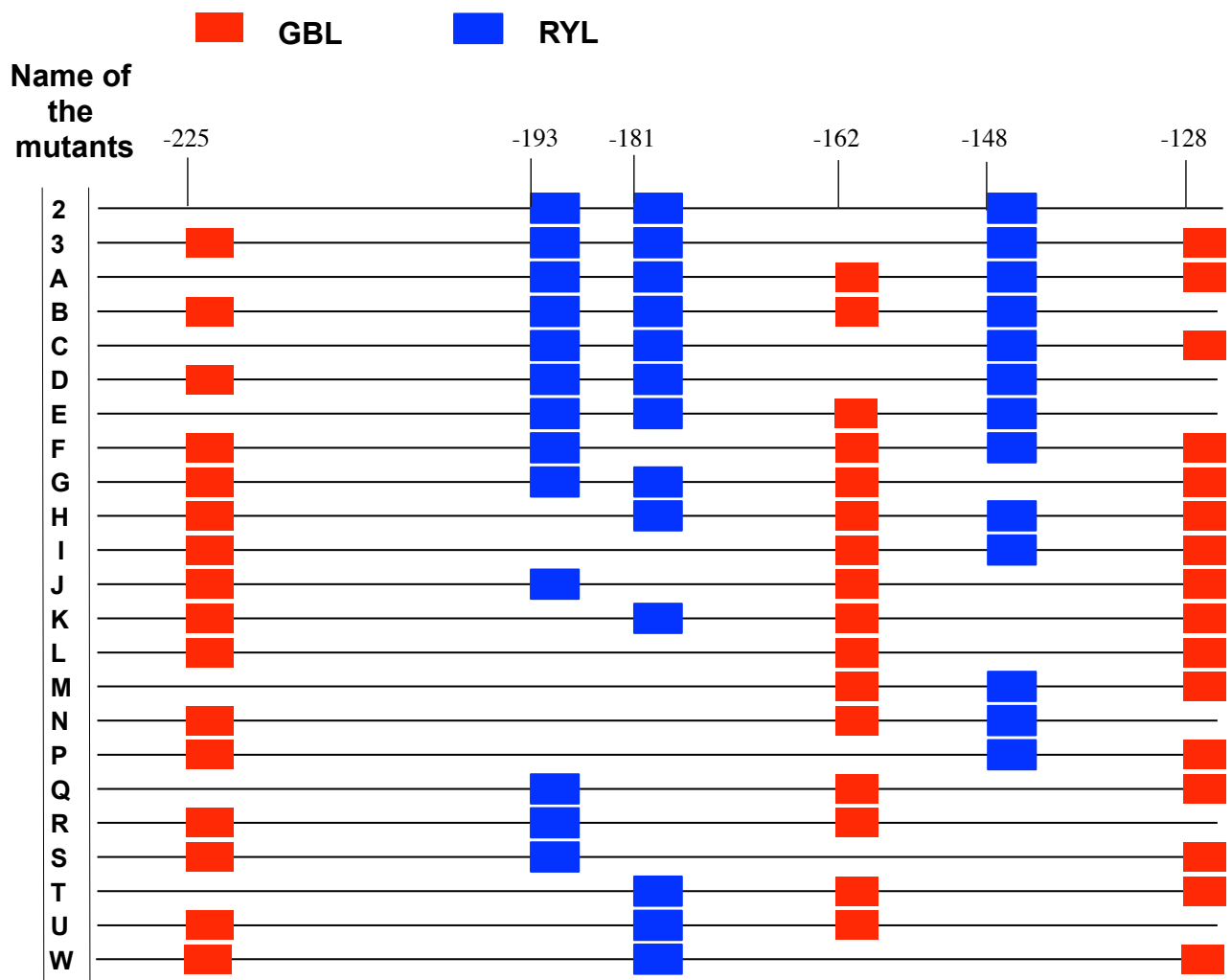




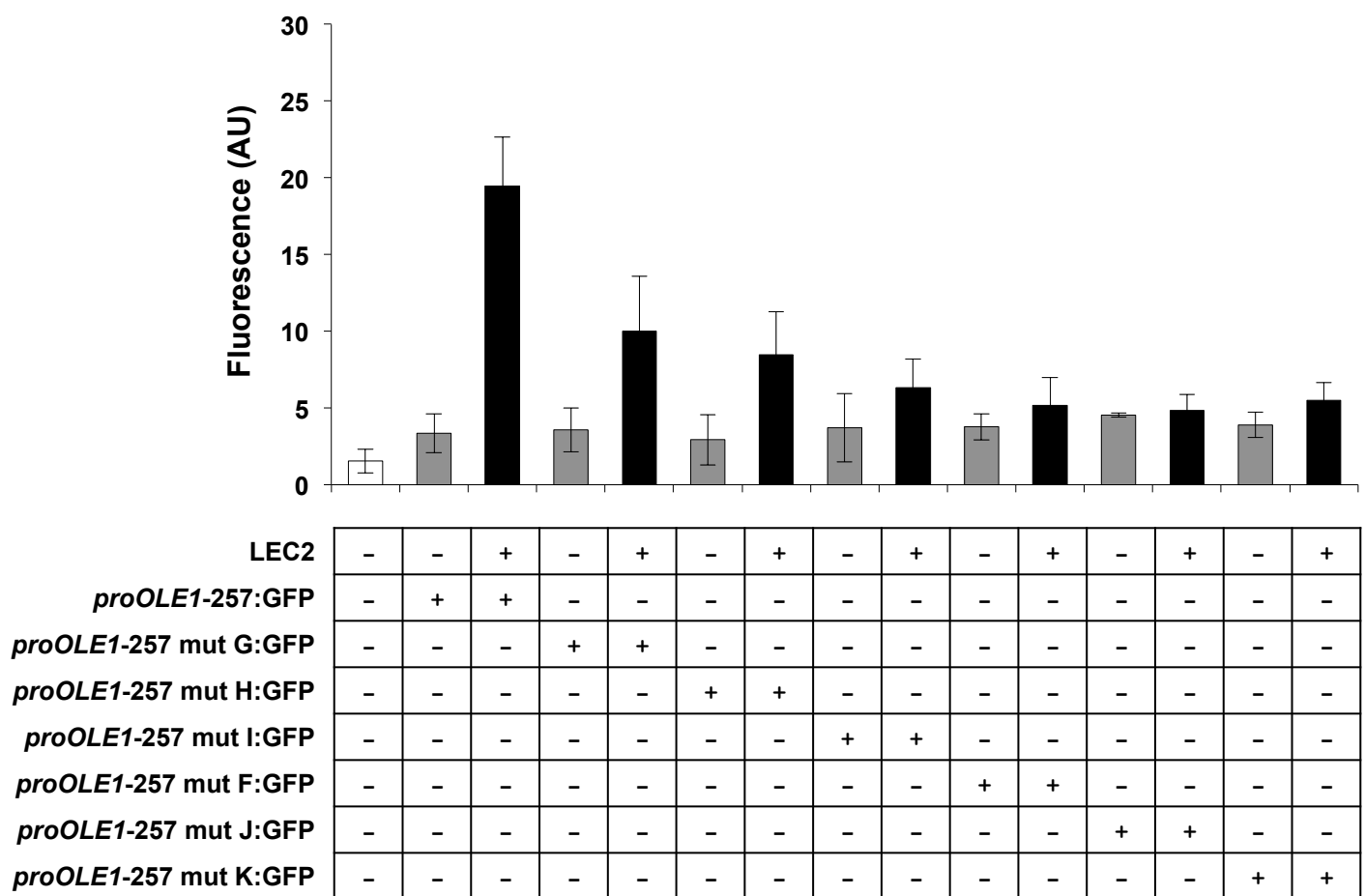
**Fig S1 A :E-score distributions of top 4, 6mer cis-elements recognized by FUS3 and ABI3**  
 The elements are shown in descending order in relation to their E-score. The motif *ACATGC* corresponds to a “mutant” version of the top-scoring *GCATGC* motif, reflecting a decrease in E-score. All the top scoring elements recognized by both proteins share the sequence *CATGC*, which could be considered as an RY-core. Note that the “canonical” RY 6mer previously described (*CATGCA*) appears in the second position for both proteins, with a clear decrease in affinity in relation to the top scoring 6mer (*GCATGC*). In general, both proteins recognize the same elements, although DNA-binding affinity was slightly lower for ABI3



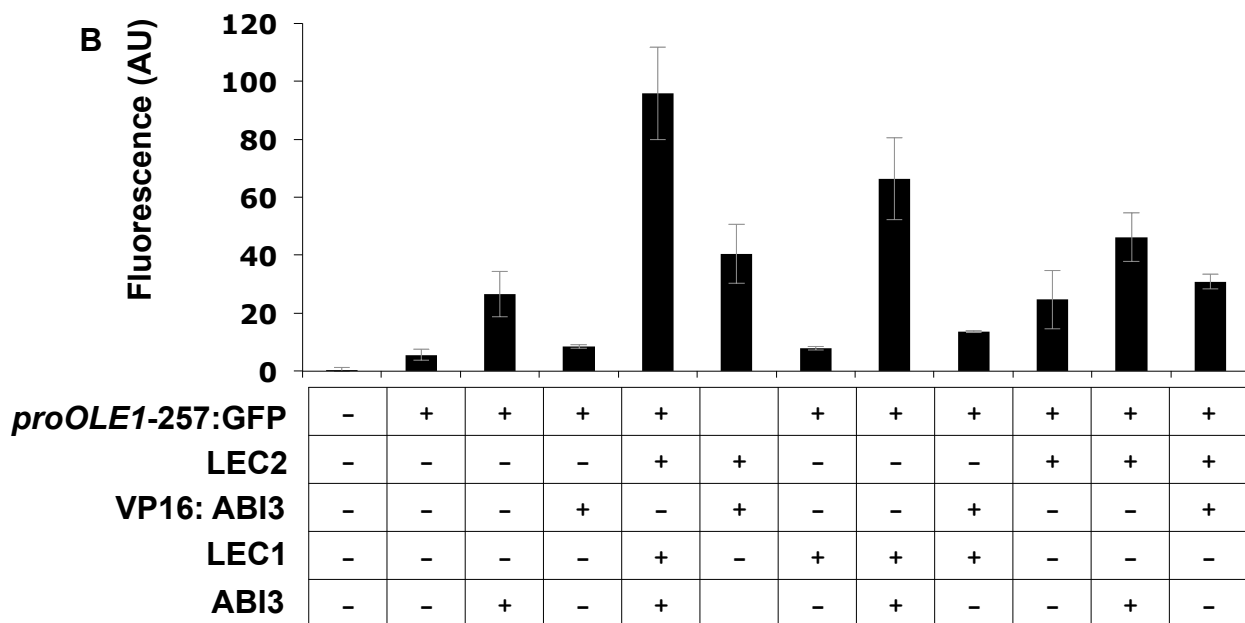
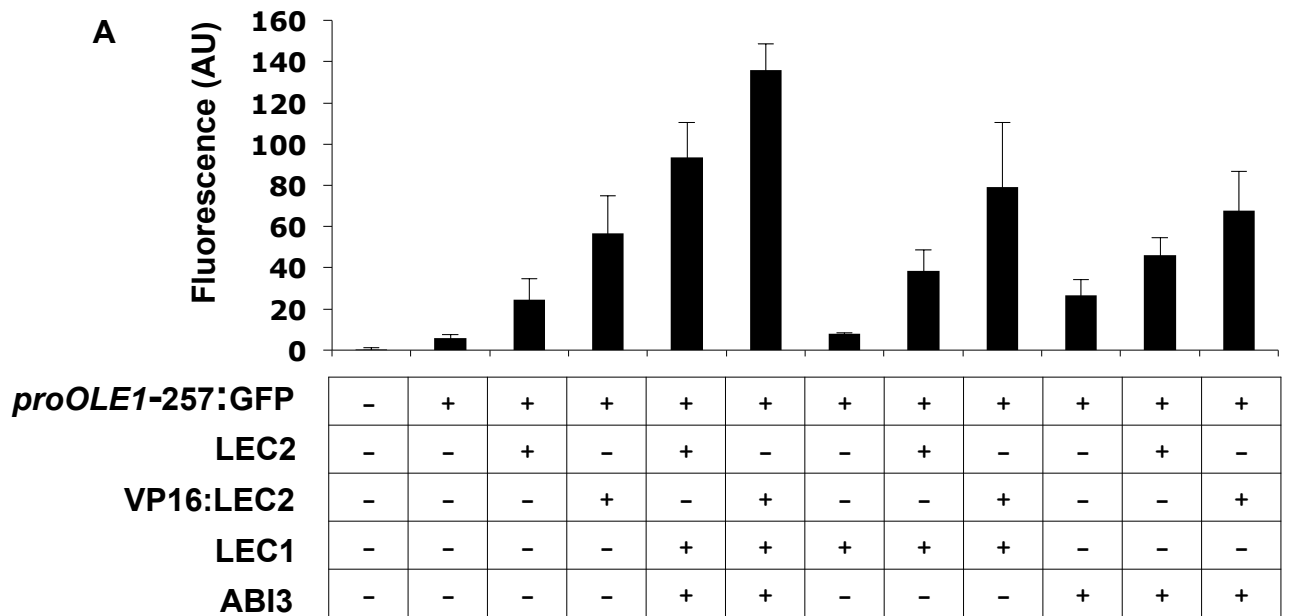
**Figure S1B:** Score plot of the *ProOLE1* sequence using matrix obtained for LEC2, FUS3 and ABI3. *ProOLE1* sequence was scanned using LEC2, FUS3 and ABI3 PWM. As these matrices are pseudo-palindromic, each binding site is systematically identified on both DNA strands with slightly different scores, explaining the presence of double bars in the plots. Numbers on the X axis indicate nucleotide number from translation start.



**Figure S2:** schematic representation of the mutations of the regulatory elements (mutated version of RYL and mutated version of GBL) found in *proOLE1*. Numbers on the top indicate nucleotide number from translation start.

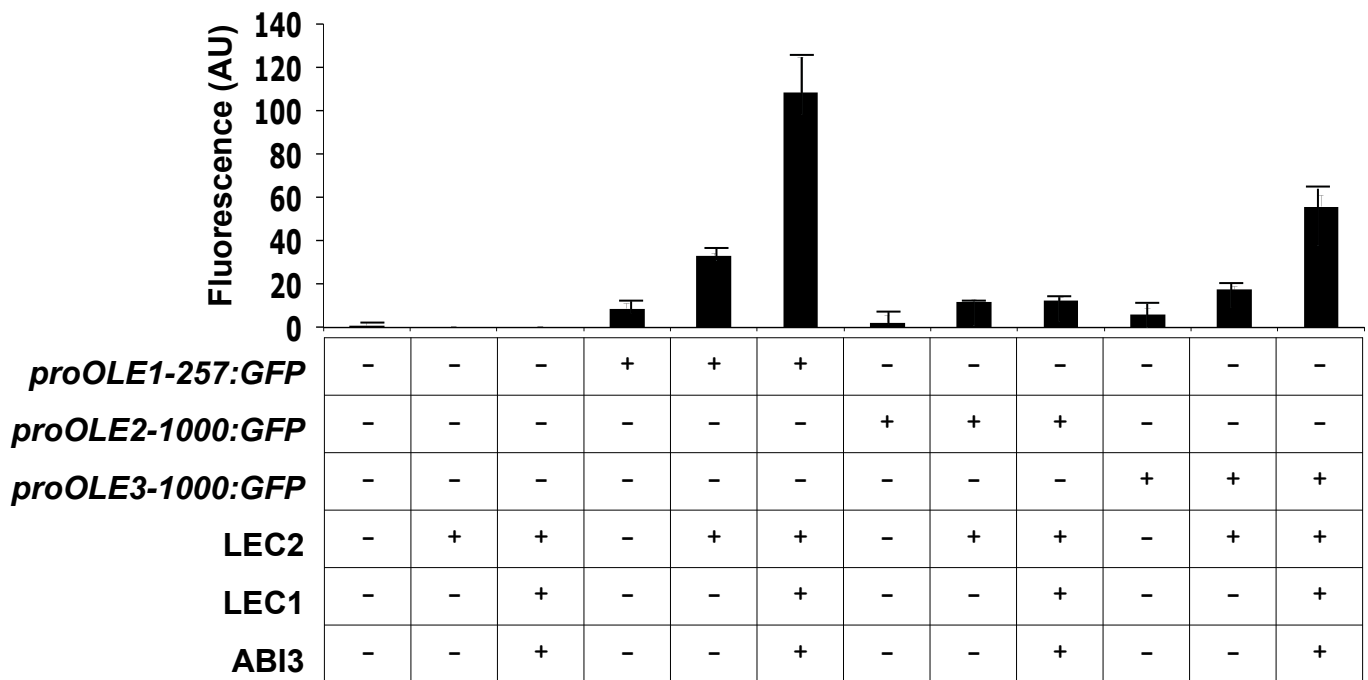


**Figure S3:** activation of the native or mutated *OLE1* promoter by LEC2 in moss protoplasts

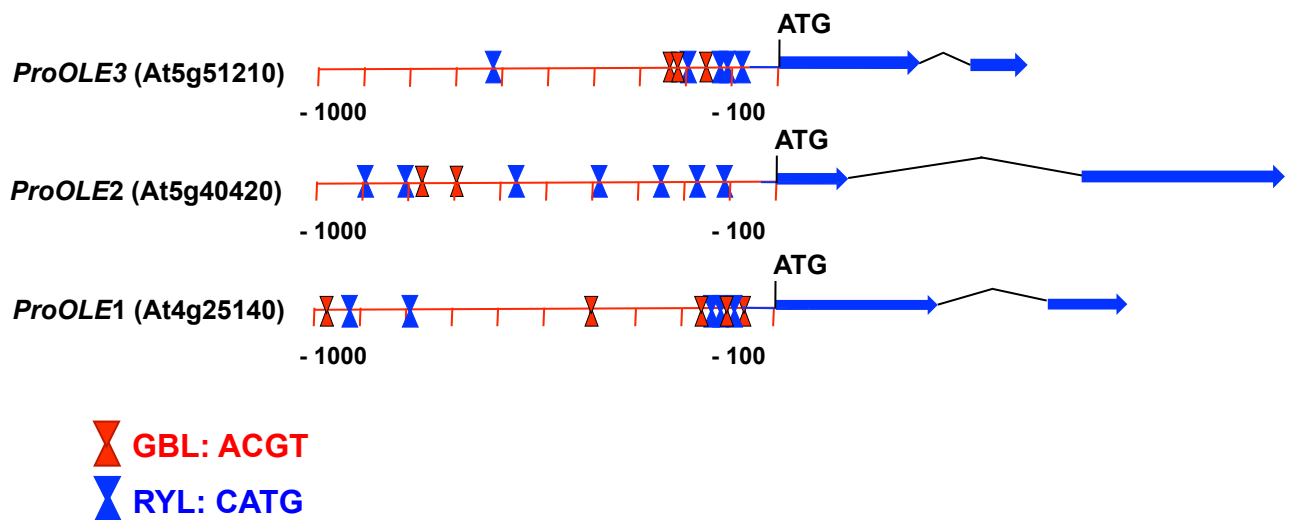


**Figure S4** : activation of *OLE1* promoter by LEC2 (A) or ABI3 (B) fused to a strong transcriptional activator (VP16) in moss protoplasts

**A**



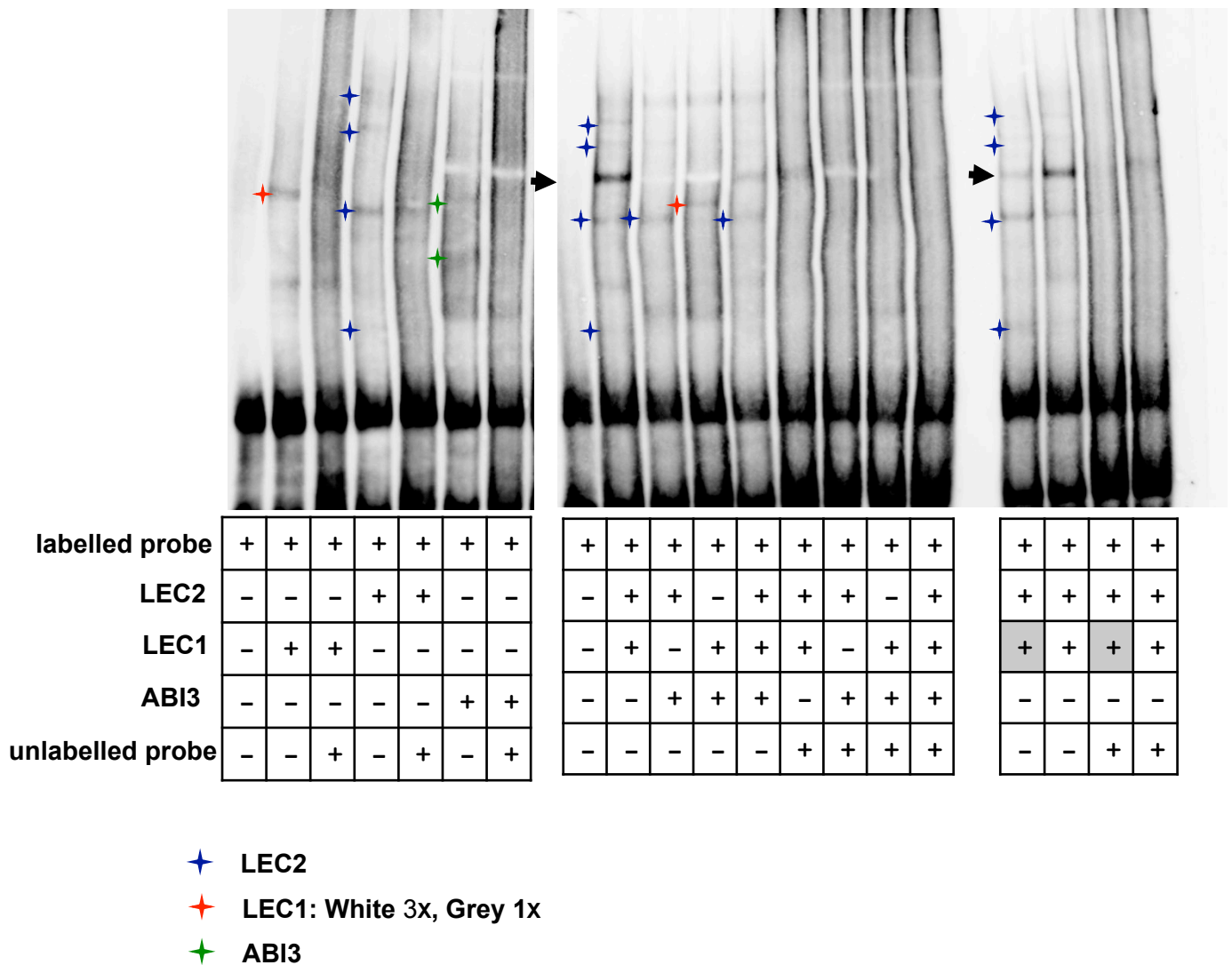
**B**



**Figure S5:** Effect of LEC2 and AFL complex on other oleosin promoters

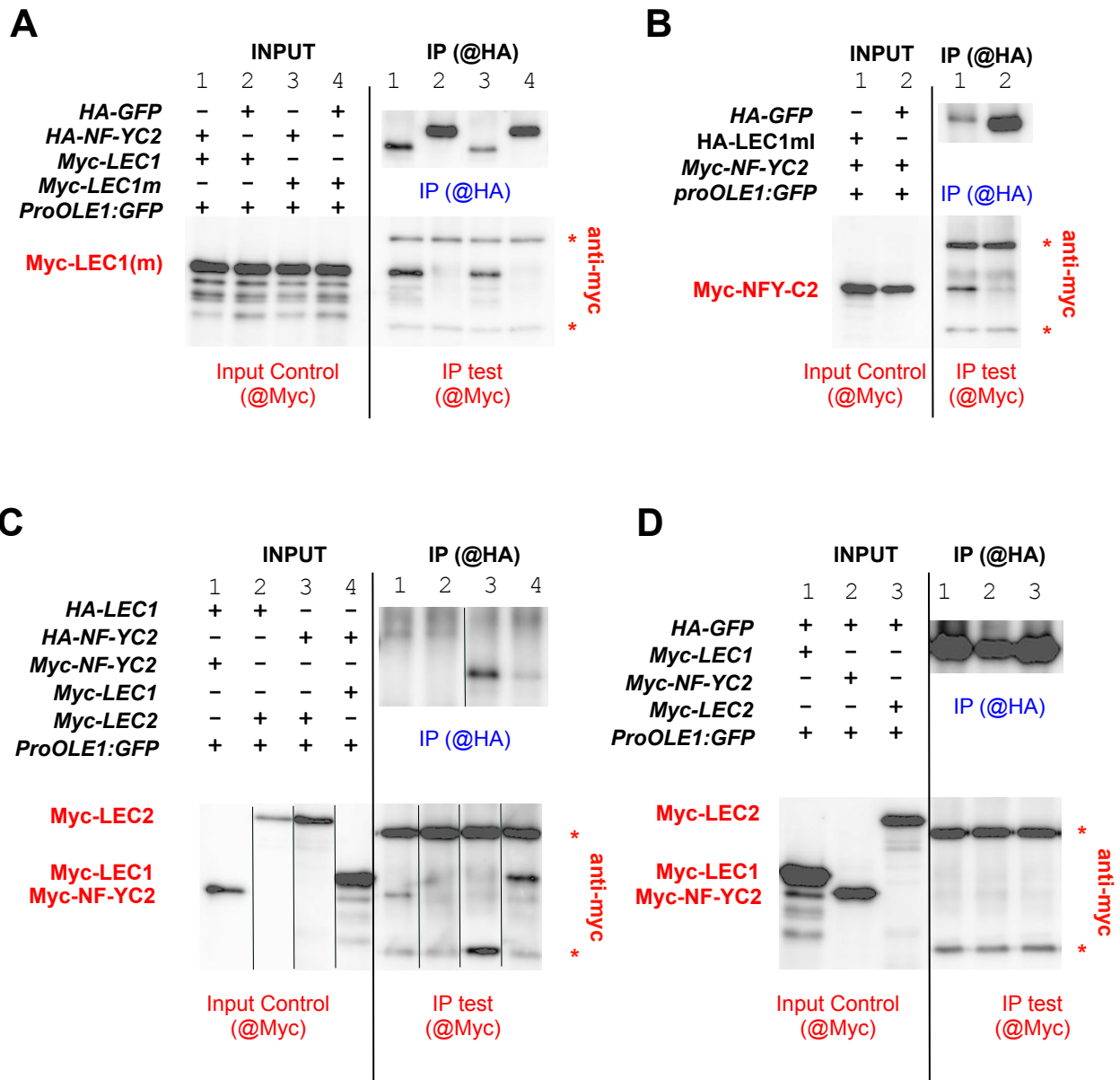
A- activation of promoter *OLE1*, *OLE2*, and *OLE3* by the LEC1, LEC2, ABI3 complex in moss protoplasts.

B- schematic representation of *proOLE1*(At4g25140), *proOLE2* (AT5G40420) and *proOLE3* (At5g51210) genomic regions. The RYL and the GBL core regions are highlighted in blue and red respectively.



**Fig S6: LEC1 and LEC2 bind to proOLE1.**

EMSA binding assay with LEC1, LEC2 and ABI3 alone or in combination with labelled 100bp wild-type proOLE1 probe. Competitions were performed with 25x more labelled probe. Three time more LEC1 were usually used for interaction studies, except on the right panel were less (1x) was used to check stoichiometry.



**Fig S7** : Interaction controls.  
in presence of *proOLE1:GFP*

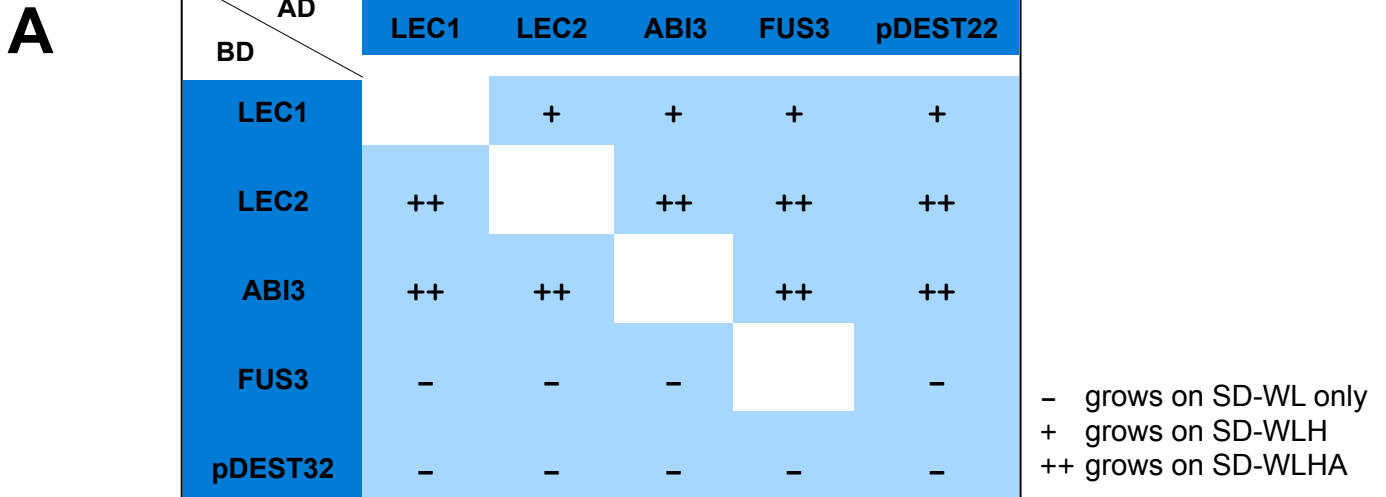
**A-B NF-YC2 interacts with the mutated LEC1 (Lec1mut).**

The HA- or Myc-tagged proteins have been co-expressed in Arabidopsis protoplasts. INPUT : 2 % of the full protein extract. IP : immunoprecipitated sample (\*) : heavy and light chains of the antibody used for the immunoprecipitation

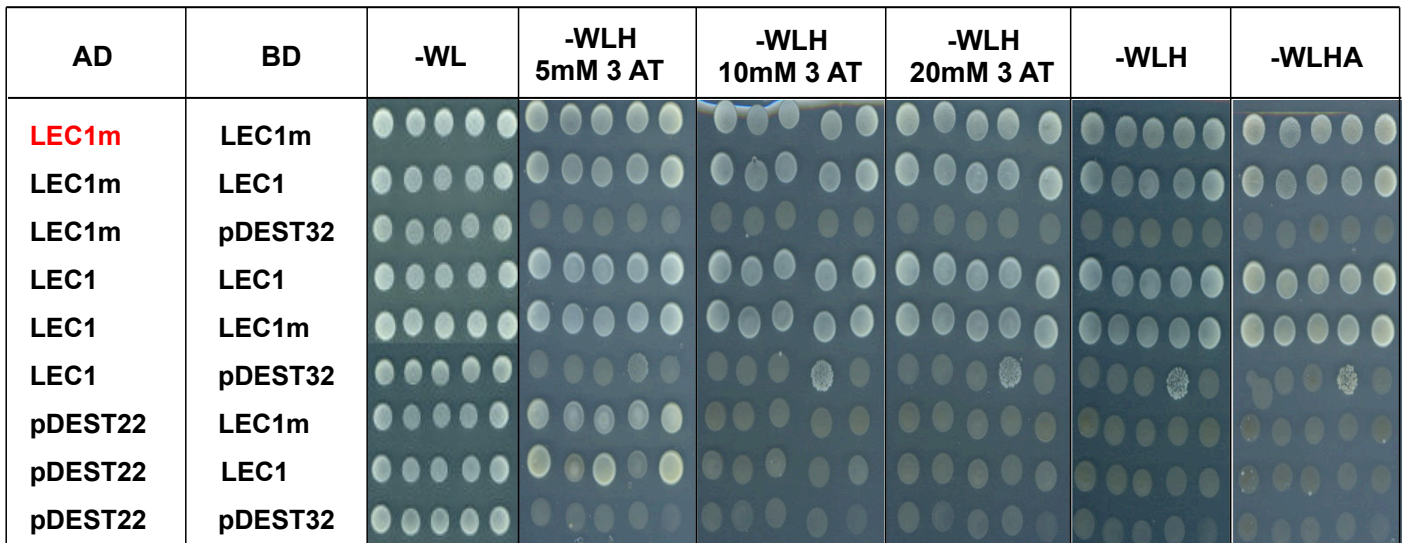
**C-D LEC1 and NF-YC2 interact with each other but do not interact with LEC2.**

The HA- or Myc-tagged proteins have been co-expressed in Arabidopsis protoplast. INPUT : 2 % of the full protein extract. IP : immunoprecipitated sample. (\*) : heavy and light chains of the antibody used for the immunoprecipitation.





**B**



FigS8: Interactions in Yeast

723

724 **Supplementary figures:**

725 **Figure S1A. E-score distribution of top four 6mers recognized by FUS3 and ABI3.**

726 The elements are shown in descending order in relation to their E-score. All top scoring elements  
727 recognized by both proteins share the sequence GCATG (underlined), which could be considered as a  
728 RY-core. The motif ACATGT corresponds to a “mutant” version of RY-core in which the nucleotide G  
729 at the 5' end has been replaced by an A, reflecting a decrease in E-score. Both proteins recognize  
730 similar elements, although DNA-binding affinity was slightly lower for ABI3 compared to FUS3.

731

732 **Figure S1B.** Score plots of the *proOLE1* sequence scanned with DNA-binding matrices described for  
733 LEC2, FUS3 and ABI3.

734 Since matrices correspond to pseudo-palindromic sequences, each binding site is systematically  
735 identified on both DNA strands with slightly different scores. This explains the presence of double bars  
736 in the plots.

737

738 **Figure S2.** Schematic representation of the mutagenized versions of the *OLE1* promoter used in this  
739 study.

740 Regulatory elements (RYL and GBL domains) present in the minimal *OLE1* promoter were  
741 mutagenized using site-directed mutagenesis. Mutagenized RYL (blue box) and GBL (red box) are  
742 represented. Figures indicate the position of these elements relative to the translational start site of  
743 *OLE1*.

744

745 **Figure S3.** Activation of the native or mutated *OLE1* promoter versions by LEC2 in moss protoplasts.

746 Means of at least 3 replicates +/- SD

747

748 **Figure S4.** Complementary analyses related to activation studies in moss protoplasts.

749 Transient activation of the *OLE1* promoter by LEC2 (A) or ABI3 (B) fused to a strong activator domain  
750 (VP16) were assayed in moss protoplasts. Means of at least 3 replicates +/- SD are presented.

751

752 **Figure S5. Test of synergistic effect of LEC2, LEC1, and ABI3 on the activation of several**  
753 **oleosin promoters in moss protoplasts.**

754 (A) Transient expression assays were carried out in moss protoplasts. GFP activity was measured in  
755 protoplasts transfected with *ProOLE1:GFP*, *ProOLE2:GFP*, or *ProOLE3:GFP* reporter and constructs  
756 allowing the expression of L-AFLs (LEC2, ABI3 or LEC1), alone or in combination. Means of at least 3  
757 replicates +/- SD are presented. AU, arbitrary units.

758 (B) Schematic representation of *proOLE1* (At4g25140), *proOLE2* (AT5G40420), and *proOLE3*  
759 (At5g51210) genomic regions. The RYL and GBL elements are highlighted in blue and red,  
760 respectively.

761

762 **Figure S6. *In vitro* binding of the L-AFLs to the *OLE1* promoter.**

763 Electrophoretic mobility shift assays (EMSA) with LEC1, LEC2, and ABI3 alone or in combination were  
764 carried out using a labelled 100-bp long probe corresponding region -228 to -121 (relative to the  
765 translational start site) of the *OLE1* promoter. Competition experiments were performed by adding  
766 non-labelled competitor DNA added in 25-fold molar excess. Shifted bands are indicated with stars,  
767 arrows indicate supershifts.

768

769 **Figure S7. Complementary results to pull-down experiments (Figure 6).**

770 (A-B) NF-YC2 interacts with LEC1 or its mutated version (LEC1m, D55K).

771 The HA- or Myc-tagged proteins have been co-expressed in *Arabidopsis* protoplasts. INPUT, 2 % of  
772 the full protein extract; IP, immunoprecipitated sample; \*, heavy and light chains of the antibody used  
773 for the immunoprecipitation.

774 (C-D) LEC1 and NF-YC2 interact with each other but do not interact with LEC2

775 The HA- or Myc-tagged proteins have been co-expressed in *Arabidopsis* protoplasts. INPUT, 2 % of  
776 the full protein extract; IP, immunoprecipitated sample; \*, heavy and light chains of the antibody used  
777 for the immunoprecipitation.

778

779

780 **Figure S8. Study of protein-protein interactions in yeast.**

781 (A) Summary of interaction studies carried out using a yeast two-hybrid system in strain AH109.  
782 Reciprocal interactions between LEC1, LEC2, ABI3, and FUS3 were assayed.  
783 AD, fusion to activation domain; BD, fusion to DNA-binding domain; W, tryptophan; L, leucine; H,  
784 histidine; A, alanine.  
785 (B) Study of homodimer formation between native or mutant LEC1 proteins. Yeast two hybrid  
786 experiments were performed using native (LEC1) or mutant (LEC1m D55K) proteins. W, tryptophan;  
787 L, leucine; H, Histidine; A, alanine; 3-AT, 3-amino -1,2,4-triazole.

788

789

790

791

792

793

794 **Supplementary table 1. Primers and genes used in this study.**

795

Primer name	5'→3' sequence	Gene	AGI
oSB110R	TATGCTCGAGCTAGTAGAAGTCATCGA	FUS3	At3g26790
oETH1132	GGCCATGGTTGATGAAAATGTGGAA	FUS3	At3g26790
oSB109F	TGCCATATGGCAGAGACTCAGTTACCT	ABI3	At3g24650
oSB109R	TGCGGCCGCTTGCCGCTCGGTTGTCT	ABI3	At3g24650
73 mer selex	TGGAGAAGAGGAGAGATCTAGC(N) <sub>30</sub> CTTGTTCTTCTTCGATTCCGG	selex	
Selex F	TGGAGAAGAGGAGAG ATCTAG	selex	
Selex R	CCGG AATCGAAGAAGAACA	selex	
LEC2 F GW	GGGGACAAGTTTGTACAAAAAGCAGGCTCGATGGATAACTTCTTACC	LEC2	At1g28300
LEC2 R GW	GGGGACCACTTTGTACAAGAAAGCTGGGTTCCACCACCACTCAAAG	LEC2	At1g28300
ABI3 F GW	GGGGACAAGTTTGTACAAAAAGCAGGCTCGATGAAAAGCTTGCATGTGGCG G	ABI3	At3g24650
ABI3 R GW	GGGGACCACTTTGTAAAGAAAGCTGGGT GGAAATGGTGTGGTTATAT	ABI3	At3g24650
FUS3 F GW	GGGGACAAGTTTGTACAAAAAGCAGGCTCGATGATGGTTGATGAAAATGTG G	FUS3	At3g26790
FUS3 R GW	GGGGACCACTTTGTAAAGAAAGCTGGGTCTAGTAGAAGTCATCGAGAG	FUS3	At3g26790
LEC1 F GW	GGGGACAAGTTTGTACAAAAAGCAGGCTCG ATGGAACGTGGAGCTCCCTTC	LEC1	At1g21970
LEC1 R GW	GGGGACCACTTTGTAAAGAAAGCTGGGTTCACTTATACTGACCATAATGG	LEC1	At1g21970
NF-YC2 F GW	GGGGACAAGTTTGTACAAAAAGCAGGCTTCATGGAGCAGTCAGAAGAGGGT C	NF-YC2	AT1G56170
NF-YC2 R GW	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAGACTCATCAGGGTGTG	NF-YC2	AT1G56170
LEC2-stop R GW	GGGGACCACTTTGTAAAGAAAGCTGGGT CCCACTCAAAGTCGTTAAA	LEC2	At1g28300
LEC1-stop R GW	GGGGACCACTTTGTAAAGAAAGCTGGGTCCTTATACTGACCATAATGG	LEC1	At1g21970
ABI3-stop R GW	GGGGACCACTTTGTAAAGAAAGCTGGGTCAATGGTGTGGTTATAT	ABI3	At3g24650
NF-YB 7 F GW	GGGGACAAGTTTGTACAAAAAGCAGGCTATGACTGAGGAGAGCCAGAAGA AG	NF-YB7	At2g13570

NF-YB 7 R GW	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACCAGTGAATTGAATCAATGT TATTATC	NF-YB7	At2g13570
L1L F GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGAACGTGGAGGCTTCCATGG C	L1L	At5g47670
L1L R GW	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGTACTTATGTTGTTGAGTCG G	L1L	At5g47670
pOleo R GW	GGGGACCACTTTGTAAAGAAAGCTGGGTCTTTTTTGTCTTTGTTACTAG	pOleo	At4g25140
pOleo Δ526 F GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGGAAGAGAGACGACG	pOleo	At4g25140
pOleo Δ257 F GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAACGCCACATTCTGAG	pOleo	At4g25140
pOleo Δ220 F GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTTTGAATAGACTCCTCTCGTTA AC	pOleo	At4g25140
pOleo Δ187 F GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTGCGGCTGCATGGTGACGCC	pOleo	At4g25140
pOleo Δ140 F GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTCTCCATTGACACGTGACTTC	pOleo	At4g25140

796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807

808 Supplementary table 2. List of primers used for site-directed mutagenesis of  
 809 *ProOLE1*.  
 810

Name of the construct	Mutations (*)	Construct used as a DNA template	Primers	5'→3' Sequence
mutA	RY3,2,1/G*,2,1	ProS3-257 (wild type) in pDONR207	mutA-F mutA-R	AGGCAAAAAACAATTGTGTCTTTGAA TTCAAAGACACAATTGTTTTTGCCT
mutB	RY3,2,1/G3,2,*	ProS3-257 (wild type) in pDONR207	mutB-F mutB-R	CATGATGTCTCCATTGACAGTTGACTTCTCGTCTCCTTTC GAAAGGAGACGAGAAGTCAACTGTCAATGGAGACATCATG
mutC	RY3,2,1/G*,*,1	mut3 (RY3,2,1/G3,*,1)	mutC-F mutC-R	AGGCAAAAAACAATTGTGTCTTTGAA TTCAAAGACACAATTGTTTTTGCCT
mutD	RY3,2,1/G3,*,*	mut3 (RY3,2,1/G3,*,1)	mutD-F mutD-R	CATGATGTCTCCATTGACAGTTGACTTCTCGTCTCCTTTC GAAAGGAGACGAGAAGTCAACTGTCAATGGAGACATCATG
mutE	RY3,2,1/G*,2,*	mutA (RY3,2,1/G*,2,1)	mutE-F mutE-R	CATGATGTCTCCATTGACAGTTGACTTCTCGTCTCCTTTC GAAAGGAGACGAGAAGTCAACTGTCAATGGAGACATCATG
mutF	RY3,*,1/G3,2,1	ProS3-257 (wild type) in pDONR207	mutF-F mutF-R	CGTTAACACATGCAGCGGCTACAACGTGACGCCATTAACACGTG CACGTGTTAATGGCGTCACGTTGTAGCCGCTGCATGTGTTAACG
mutG	RY3,2,*/G3,2,1	ProS3-257 (wild type) in pDONR207	mutG-F mutG-R	GGCCTACAATTACAACATGTCTCCATTG CAATGGAGACATGTTGTAATTGTAGGCC
mutH	RY*,2,1/G3,2,1	ProS3-257 (wild type) in pDONR207	mutH-F mutH-R	AGACTCCTCTCGTTAACAGTTGTAGCGGCTGCATGGTGACG CGTCACCATGCAGCCGCTACAACGTGTTAACGAGAGGAGTCT
mutI	RY*,*,1/G3,2,1	mutH (RY*,2,1/G3,2,1)	mutI-F mutI-R	CGTTAACAGTTGTAGCGGCTACAACGTGACGCCATTAACACGTG CACGTGTTAATGGCGTCACGTTGTAGCCGCTACAACGTGTTAACG
mutJ	RY3,*,*/G3,2,1	mutG (RY3,2,*/G3,2,1)	mutJ-F mutJ-R	CGTTAACACATGCAGCGGCTACAACGTGACGCCATTAACACGTG CACGTGTTAATGGCGTCACGTTGTAGCCGCTGCATGTGTTAACG
mutK	RY*,2,*/G3,2,1	mutH (RY*,2,1/G3,2,1)	mutK-F mutK-R	GGCCTACAATTACAACATGTCTCCATTG CAATGGAGACATGTTGTAATTGTAGGCC
mutL	RY*,*,*/G3,2,1	mutK (RY*,2,*/G3,2,1)	mutL-F mutL-R	CGTTAACAGTTGTAGCGGCTACAACGTGACGCCATTAACACGTG CACGTGTTAATGGCGTCACGTTGTAGCCGCTACAACGTGTTAACG
mutM	RY*,*,1/G*,2,1	mutI (RY*,*,1/G3,2,1)	mutM-F mutM-R	AGGCAAAAAACAATTGTGTCTTTGAA TTCAAAGACACAATTGTTTTTGCCT
mutN	RY*,*,1/G3,2,*	mutI (RY*,*,1/G3,2,1)	mutN -F mutN -R	CATGATGTCTCCATTGACAGTTGACTTCTCGTCTCCTTTC GAAAGGAGACGAGAAGTCAACTGTCAATGGAGACATCATG
mutP	RY*,*,1/G3,*,1	mutI (RY*,*,1/G3,2,1)	mutP-F mutP-R	CGTGACGCCATTAACAGTTGGCCTACAATTGCATG CATGCAATTGTAGGCCAAGTCAACTGTTAATGGCGTCACG
mutQ	RY3,*,*/G*,2,1	mutJ (RY3,*,*/G3,2,1)	mutQ-F mutQ-R	AGGCAAAAAACAATTGTGTCTTTGAA TTCAAAGACACAATTGTTTTTGCCT
mutR	RY3,*,*/G3,2,*	mutJ (RY3,*,*/G3,2,1)	mutR-F mutR-R	CAACATGTCTCCATTGACAGTTGACTTCTCGTCTCCTTTC GAAAGGAGACGAGAAGTCAACTGTCAATGGAGACATGTTG
mutS	RY3,*,*/G3,*,1	mutJ (RY3,*,*/G3,2,1)	mutS-F mutS-R	CGTGACGCCATTAACAGTTGGCCTACAATTACAAC GTTGTAATTGTAGGCCAAGTCAACTGTTAATGGCGTCACG
mutT	RY*,2,*/G*,2,1	mutK (RY*,2,*/G3,2,1)	mutT-F mutT-R	AGGCAAAAAACAATTGTGTCTTTGAA TTCAAAGACACAATTGTTTTTGCCT
mutU	RY*,2,*/G3,2,*	mutK (RY*,2,*/G3,2,1)	mutU-F mutU-R	CAACATGTCTCCATTGACAGTTGACTTCTCGTCTCCTTTC GAAAGGAGACGAGAAGTCAACTGTCAATGGAGACATGTTG
mutW	RY*,2,*/G3,*,1	mutK (RY*,2,*/G3,2,1)	mutW-F	GGTGACGCCATTAACAGTTGGCCTACAATTACAAC

			mutW-R	GTTGTAATTGTAGGCCAACTGTTAATGGCGTCACC
mut2	RY3,2,1/G*,**	mutD (RY3,2,1/G3,*,*)	mut2-F mut2-R	AGGCAAAAAACAATTGTGTCTTTGAA TTCAAAGACACAATTGTTTTGCCT
mut3	RY3,2,1/G3,*,1	ProS3-257 (wild type) in pDONR207	mut3-F mut3-R	GGTGACGCCATTAACAGTTGGCCTACAATTGCATG CATGCAATTGTAGGCCAACTGTTAATGGCGTCACC





A

	proOLE1:GFP	proOLE1:GFP + LEC1	proOLE1:GF P + LEC2	proOLE1:GFP + ABI3	proOLE1:GFP + FUS3	proOLE1:GFP + LEC1 + LEC2	proOLE1:GFP + ABI3 + LEC1	proOLE1:GFP + FUS3 + LEC1	proOLE1:GFP + ABI3 + LEC2	proOLE1:GFP + FUS3 + ABI3	proOLE1:GFP + LEC2 + ABI3 + LEC1
proOLE1:GFP + LEC1	0.2406										
proOLE1:GFP + LEC2	5.2e-13	1.9e-07									
proOLE1:GFP + ABI3	6.6e-12	4.2e-07	1.33e-4								
proOLE1:GFP + FUS3	4.8e-07	6.4e-05	3.819e-4	4.741e-08							
proOLE1:GFP + LEC1 + LEC2	3.21-11	1.2e-06	4.066e-3	0.6098	NR						
proOLE1:GFP + ABI3 + LEC1	6.6e-12	4.2e-07	1.515e-09	8.839e-06	NR	8.83e-05					
proOLE1:GFP + FU3 + LEC1	3.06e-08	1.0e-05	7.082e-05	3.647e-09	0.2636	5.41e-07	1.802e-07				
proOLE1:GFP + ABI3 + LEC2	1.3e-11	5.3e-07	8.532e-08	6.117e-3	NR	NR	0.0151	NR			
proOLE1:GFP + ABI3 + LEC1 + LEC2	8.2e-15	6.0e-08	2.299e-11	1.643e-07	NR	6.331e-07	0.02368	NR	6.243e-05		
proOLE1:GFP + FUS3 +ABI3	8.742e-07	5.017e-05	5.891 <sup>e</sup> -3	0.472	2.955e-4	NR	0.00783	NR	0.1375	NR	
proOLE1:GFP + FUS3 + ABI3 + LEC1	2.997e-06	9.486e-05	8.544e-05	6.115e-3	6e-4426	NR	0.9628	6.426e-4	NR	0.09447	8.03e-3

3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

B

	pS3 <sup>-257</sup> :GFP	pS3 <sup>-257</sup> :GFP + LEC2	pS3 <sup>-257</sup> :GFP + LEC2 + LEC1	pS3 <sup>-257</sup> :GFP + ABI3	pS3 <sup>-257</sup> :GFP + ABI3 + LEC1	pS3 <sup>-257</sup> :GFP + ABI3 + LEC2	pS3 <sup>-257</sup> :GFP + LEC2 + LEC1 + ABI3	pS3 <sup>-257</sup> :GFP + LEC2 + ABI3 + NFYB7
pS3 <sup>-257</sup> :GFP + LEC2	5.222e-13							
pS3 <sup>-257</sup> :GFP + LEC2 + LEC1	3.214e-11	4.066e-3						
pS3 <sup>-257</sup> :GFP + LEC2 + NFYB7	4.209e-3	0.195	0.3667					
pS3 <sup>-257</sup> :GFP + LEC2 + LEC1mut	0.004209	0.1003	0.01807					
pS3 <sup>-257</sup> :GFP + ABI3	6.654e-12	1.332e-4	0.6098					
pS3 <sup>-257</sup> :GFP + ABI3 + LEC1	6.654e-12	1.515e-09	8.83e-05	8.839e-06				
pS3 <sup>-257</sup> :GFP + ABI3 + NFYB7	4.209e-3	NR	NR	0.4407	0.02917			
pS3 <sup>-257</sup> :GFP + ABI3 + LEC1mut	0.004209	NR	NR	0.04274	5.092e-3			
pS3 <sup>-257</sup> :GFP + ABI3 + LEC2	1.336e-11	8.532e-08	NR	0.006117	0.01515			
pS3 <sup>-257</sup> :GFP + LEC2 + LEC1 + ABI3	8.232e-15	2.299e-11	6.331e-07	1.643e-07	0.02368	6.243e-05		
pS3 <sup>-257</sup> :GFP + LEC2 + ABI3 + NFYB7	1.821e-4	NR	NR	NR	NR	0.2421	8.974e-3	
pS3 <sup>-257</sup> :GFP + ABI3 + LEC2 + LEC1mut	4.209e-3	NR	NR	NR	NR	0.1337	0.01367	0.8607

4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17

# C-1

		pWt +LEC2	pWt +ABI3	pWt +LEC1 LEC2 ABI3	pmutA +LEC2	pmutA +ABI3	pmutA +LEC1 LEC2 ABI3	pmut3 +LEC2	pmut3 +ABI3	pmut3 +LEC1 LEC2 ABI3	pmutB +LEC2	pmutB +ABI3	pmutB +LEC1 LEC2 ABI3	pmut2 +LEC2	pmut2 +ABI3	pmut2 +LEC1 LEC2 ABI3	pmutH +LEC2	pmutH +ABI3	pmutH +LEC1 LEC2 ABI3	
pmut A	gain LEC2	0.12																		
pmut A	gain ABI3		0.28																	
pmut A	gain LEC2 LEC1 ABI3			0.014	0.005	0.002														
pmut 3	gain LEC2	0.25			0.15															
pmut 3	gain ABI3		0.11			0.25														
pmut 3	gain LEC2 LEC1 ABI3			0.001			0.008	0.97	0.012											
pmutB	gain LEC2	0.84			0.40			0.34												
pmutB	gain ABI3		0.16			0.09			0.039											
pmutB	gain LEC2 LEC1 ABI3			0.057			0.52			0.019	0.007	0.016								
pmut2	gain LEC2	0.003			0.004			0.05			0.079									
pmut2	gain ABI3		0.001			0.008			0.13			0.023								
pmut2	gain LEC2 LEC1 ABI3			0.002			0.004			0.014			0.010	0.010	0.007					
pmutH	gain LEC2	0.009			0.003			0.051			0.072			0.34						
pmutH	gain ABI3		0.13			0.26			0.89			0.037			0.21					
pmutH	gain LEC2 LEC1 ABI3			0.02			0.20			0.42			0.13			0.086	0.047	0.056		
pmutF	gain LEC2	0.007			0.002			0.052			0.077			0.55			0.35			
pmutF	gain ABI3		0.18			0.37			0.88			0.047			0.14					0.79
pmutF	gain LEC2 LEC1 ABI3			0.002			0.024			0.16			0.040			0.002				0.81
pmutG	gain LEC2	0.013			0.006			0.06			0.11			0.26			0.004			
pmutG	gain ABI3		0.75			0.45			0.18			0.31		0.06					0.17	
pmutG	gain LEC2 LEC1 ABI3			0.040			0.76			0.07			0.45			0.026				0.31
pmutI	gain LEC2	0.006			0.0014			0.049			0.070			0.30			0.82			
pmutI	gain ABI3		0.011			0.03			0.34			0.026			0.27				0.47	
pmutI	gain LEC2 LEC1 ABI3			0.005			0.014			0.13			0.022			0.010				0.20
pmutJ	gain LEC2	0.009			0.005			0.049			0.070			0.26			0.41			
pmutJ	gain ABI3		0.018			0.044			0.42			0.026			0.23				0.55	
pmutJ	gain LEC2 LEC1 ABI3			0.003			0.006			0.017			0.012			0.42				0.081
pmut K	gain LEC2	0.009			0.004			0.050			0.071			0.28			0.57			
pmut K	gain ABI3		0.08			0.29			0.70			0.052			0.035				0.63	
pmut K	gain LEC2 LEC1 ABI3			0.006			0.08			0.31			0.057			0.042				0.93
pmutL	gain LEC2	0.009			0.005			0.045			0.062			0.13			0.03			
pmutL	gain ABI3		0.002			0.022			0.32			0.033			0.16				0.45	
pmutL	gain LEC2 LEC1 ABI3			0.001			0.007			0.30			0.008		0.12					0.21

# C-2

	pmutF +LEC2	pmutF +ABI3	pmutF +LEC1 LEC2 ABI3	pmutG +LEC2	pmutG +ABI3	pmutG +LEC1 LEC2 ABI3	pmutI +LEC2	pmutI +ABI3	pmutI +LEC1 LEC2 ABI3	pmutJ +LEC2	pmutJ +ABI3	pmutJ +LEC1 LEC2 ABI3	pmutK +LEC2	pmutK +ABI3	pmutK +LEC1 LEC2 ABI3	pmutK +LEC2	pmutK +ABI3	
pmut A gain LEC2																		
pmut A gain ABI3																		
pmut A gain LEC2 LEC1 ABI3																		
pmut 3 gain LEC2																		
pmut 3 gain ABI3																		
pmut 3 gain LEC2 LEC1 ABI3																		
pmutB gain LEC2																		
pmutB gain ABI3																		
pmutB gain LEC2 LEC1 ABI3																		
pmut2 gain LEC2																		
pmut2 gain ABI3																		
pmut2 gain LEC2 LEC1 ABI3																		
pmutH gain LEC2																		
pmutH gain ABI3																		
pmutH gain LEC2 LEC1 ABI3																		
pmutF gain LEC2																		
pmutF gain ABI3																		
pmutF gain LEC2 LEC1 ABI3	0.002	0.009																
pmutG gain LEC2	0.013																	
pmutG gain ABI3		0.22																
pmutG gain LEC2 LEC1 ABI3			0.14	0.020	0.026													
pmutI gain LEC2	0.34			0.09														
pmutI gain ABI3		0.32			0.052													
pmutI gain LEC2 LEC1 ABI3			0.027				0.0003	0.015										
pmutJ gain LEC2	0.18			0.005			0.76											
pmutJ gain ABI3		0.38			0.10			0.84										
pmutJ gain LEC2 LEC1 ABI3			0.005			0.026			0.001	0.0014	0.19							
pmut K gain LEC2	0.21			0.003			0.87			0.79								
pmut K gain ABI3		0.86			0.23			0.13			0.18							
pmut K gain LEC2 LEC1 ABI3			0.84			0.20			0.12			0.041	0.023	0.031				
pmutL gain LEC2	0.048			0.007			0.11			0.021			0.036					
pmutL gain ABI3		0.31			0.10			1			0.82			0.10				
pmutL gain LEC2 LEC1 ABI3			0.08			0.033			0.92			0.11			0.12	0.034	0.052	

- 2
- 3 (A) Statistical analyses corresponding to the results presented in Figure 3.
- 4 (B) Statistical analyses corresponding to the results presented in Figure 4
- 5 (C) Statistical analyses corresponding to the results presented in Figure 5 (C1, to be continued with C2)

6 Red: significant differences at 5%

7 NR: not relevant.

8 For A and B, the normality of the data was checked with a Shapiro-Wilk test. As the samples were not all following a Normal law, a Kolmogorov-Smirnov test  
 9 was then performed on the relevant comparisons. All the tests are included in the R package

2 C: The normality of the data was checked with a Shapiro-Wilk test. Student t-test was performed on the relevant comparisons using the R stats Package  
3 (<https://www.r-project.org/>) version 3.2.

**Annexe 1B : Article de revue**

Fatihi A, Boulard C, Bouyer D, Baud S, Dubreucq B, Lepiniec L. (2016) **Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds.** Plant Sci. Sep;250:198-204.



ELSEVIER

Contents lists available at ScienceDirect

## Plant Science

journal homepage: [www.elsevier.com/locate/plantsci](http://www.elsevier.com/locate/plantsci)

Review article

# Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds



Abdelhak Fatihi<sup>a,\*</sup>, Céline Boulard<sup>a</sup>, Daniel Bouyer<sup>b</sup>, Sébastien Baud<sup>a</sup>,  
Bertrand Dubreucq<sup>a</sup>, Loïc Lepiniec<sup>a,\*</sup>

<sup>a</sup> IJPB, INRA, AgroParisTech, CNRS, Université Paris-Saclay, RD10, 78026, Versailles Cedex, France

<sup>b</sup> Institut de Biologie de l'ENS, CNRS UMR8197, Ecole Normale Supérieure, 46 rue d'Ulm, 75230, Paris cedex 05, France

## ARTICLE INFO

## Article history:

Received 23 April 2016

Received in revised form 16 June 2016

Accepted 18 June 2016

Available online 23 June 2016

## Keywords:

Seed development

Seed maturation

Transcriptional regulation

LAFL network

Oil

Seed storage proteins

## ABSTRACT

Increasing yield and quality of seed storage compounds in a sustainable way is a key challenge for our societies. Genome-wide analyses conducted in both monocot and dicot angiosperms emphasized drastic transcriptional switches that occur during seed development. In *Arabidopsis thaliana*, a reference species, genetic and molecular analyses have demonstrated the key role of LAFL (LEC1, ABI3, FUS3, and LEC2) transcription factors (TFs), in controlling gene expression programs essential to accomplish seed maturation and the accumulation of storage compounds. Here, we summarize recent progress obtained in the characterization of these LAFL proteins, their regulation, partners and target genes. Moreover, we illustrate how these evolutionary conserved TFs can be used to engineer new crops with altered seed compositions and point out the current limitations. Last, we discuss about the interest of investigating further the environmental and epigenetic regulation of this network for the coming years.

© 2016 Elsevier Ireland Ltd. All rights reserved.

## Contents

1. Introduction .....	199
2. Master regulators of seed development and maturation .....	199
2.1. LAFL genes .....	199
2.2. Epigenetic regulation of LALF genes expression .....	199
2.3. LAFL targets and partners .....	200
3. Engineering new crops with modified storage reserves .....	200
3.1. LAFL functions are conserved in angiosperms .....	200
3.2. Using LALF and WRI genes for engineering crops .....	202
4. Conclusions and prospects .....	202
Acknowledgments .....	202
References .....	202

\* Corresponding authors.

E-mail addresses: [Abdelhak.Fatihi@versailles.inra.fr](mailto:Abdelhak.Fatihi@versailles.inra.fr) (A. Fatihi),  
[Loic.Lepiniec@versailles.inra.fr](mailto:Loic.Lepiniec@versailles.inra.fr) (L. Lepiniec).

## 1. Introduction

Seed and grain productions are strategic for agriculture, food supply, economy and, more broadly, humankind. They provide half of the world's intake of dietary proteins, oils and starch and green renewable carbon resources (e.g. oil or starch) that represent an alternative to fossil hydrocarbon chains for chemical industries. Thus, there is a need to develop new strategies to modify the composition of oil, storage proteins or carbohydrate in crops for specific industrial uses. These issues are compounded by the dwindling of arable lands and the forecasted climate change, characterized by an increase in average temperature, already impacting negatively seed production [1]. In this context, increasing yield and quality of seed storage compounds represents a global key challenge not only for agriculture but, more broadly, for our societies and industries.

Seed development in both monocots and dicots has been well described [2]. Proper seed formation requires the coordinated development of diploid embryo, triploid endosperm, and maternal testa, as well as inter-compartmental signaling [3]. In exalbuminous species, such as *Arabidopsis*, embryo growth occurs at the expense of the endosperm, which progressively degenerates. Finally, the embryo, which is surrounded by a one-cell layer of remaining endosperm (i.e. the aleurone layer), fills the mature seed (Fig. 1). Contrastingly, mature cereal grains are mainly composed of a large endosperm, whereas the embryo remains small in size. During seed maturation, photo-assimilates are converted into storage compounds in the form of oil, storage proteins, or starch. The nature and location of these reserves in the mature seeds are highly variable depending on the species considered [2] (Fig. 1). Starch constitutes the main reserve in the endosperm of cereal grains [4,5]. Embryos can store mainly oil (e.g. rapeseed), proteins (e.g. soybean) or starch (e.g. pea) [6,7]. Some species (e.g. sugar beet or coffee) accumulate their reserves in a perisperm originating from the maternal nucellus [8].

*Arabidopsis* seed development and maturation have been extensively described [9]. The metabolic routes leading to the accumulation of the main storage compounds (i.e. oil or proteins), which determine seed quality, are well characterized too [7]. Genome wide transcriptional profiles have been established at different stages of seed development [2,10]. Besides an overall elevated transcriptional activity observed during the maturation phase, these analyses have revealed that the late embryonic phase is characterized by a fundamental shift in gene expression profiles, with many genes being specifically active during this phase. This has been confirmed in other plants including cereals such as maize and rice [11,12]. Taken together these results suggest that important transcriptional switches are required to trigger and orchestrate the maturation process and the accumulation of storage compounds, which define to a large extent seed quality. Consistent with this idea, a complex network of transcriptional regulators that controls seed maturation has been characterized and is presented below.

## 2. Master regulators of seed development and maturation

Several seed transcriptional regulators have been identified and characterized in *Arabidopsis* as well as in other plant species. They control different aspects of seed growth and development [13]. Most of them regulate testa or early endosperm and embryo development, and/or are involved in hormone signaling and cell division [14,15]. Here we focus on the members of the LAFL regulatory network that controls seed maturation, their protein partners and their target regulatory genes.

### 2.1. LAFL genes

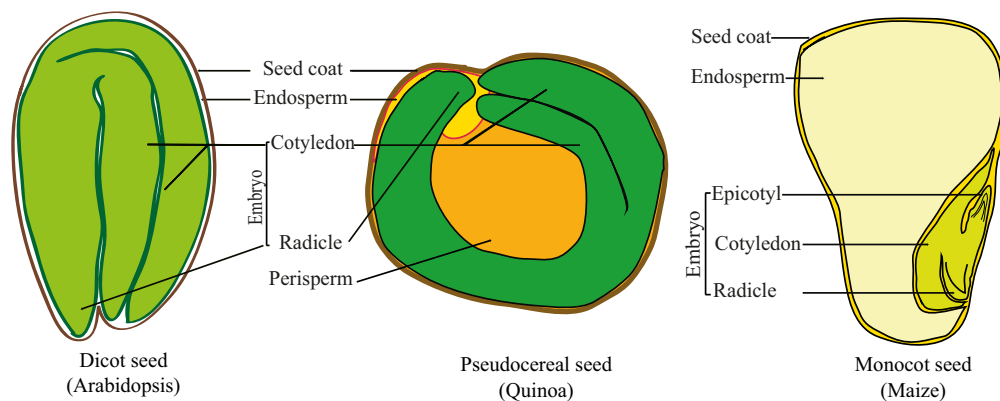
The *Arabidopsis* LAFL genes code for LEAFY COTYLEDON1 (LEC1), ABSICISIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3), and LEC2, respectively [16–19]. *Arabidopsis* LEC1, one of the first characterized regulators of seed development, is a member of the NF-YB protein family (HAP3 subunit of the CCAAT-box binding factors). ABI3, FUS3, and LEC2 belong to the conserved, plant specific, family of B3 domain transcription factors and thus are collectively named AFL-B3. LEC1 is expressed in developing seeds at early stages of embryogenesis from the globular to the bent-cotyledon stages whereas LEC2 is weakly and specifically expressed at early stages of embryogenesis, with highest mRNA levels detected between the pre-globular and bent-cotyledons stages (Fig. 2). By contrast, ABI3 and FUS3 accumulate during late stages of embryogenesis. Interestingly, LAFL transcripts are also detected in the endosperm, suggesting that additional regulators are required to specify the proper accumulation of different storage compounds in the endosperm and embryo (see Section 2.3).

Characterization of *laf1* single and multiple mutants has demonstrated that the LAFL proteins have partially overlapping and synergistic functions during embryogenesis and seed maturation [16–20]. The structural similarity between the AFL-B3 proteins and their resulting ability to bind similar DNA cis-elements (see Section 2.3) account for this functional redundancy. In addition, the expression of FUS3 and ABI3 is controlled by a network of partially redundant regulations that involves LEC1 and LEC2 and, more surprisingly, FUS3 and ABI3 themselves, through positive regulatory feedback loops (Fig. 3). Moreover, we have recently demonstrated that the AFL-B3 proteins can take part in multi-protein complexes, which include LEC1 or LEC1-like, to control the expression of their target genes [21]. Nevertheless, the different LAFL also have more specific functions [17,19][17,19 and references therein]. For instance, ABI3 is involved in chlorophyll degradation in late maturing seeds [22], but has only a limited impact on oil synthesis and accumulation, in comparison to FUS3 [19]. Conversely LEC1, LEC2, and FUS3, but not ABI3, control embryo identity, and loss of their function induces the formation of true leaves in embryos (instead of cotyledons) that display trichomes and accumulate anthocyanin (instead of storage compounds) [17]. Finally, ectopic expression of LEC1 and LEC2 in vegetative cells can induce somatic embryogenesis and *lec1* and *lec2* mutants have lower ability to undergo somatic embryogenesis [23,24 and references therein] suggesting an essential role of these genes in embryonic differentiation. This surprising ability of LEC genes to induce somatic embryogenesis provides a possible explanation for the establishment of strong and apparently redundant genetic and epigenetic mechanisms that repress their expression in vegetative tissues [25][25 and references therein].

### 2.2. Epigenetic regulation of LAFL genes expression

LAFL are repressed in vegetative tissues by Polycomb Repressive Complexes (PRC) 1 and 2 that are responsible for histone H2A monoubiquitination and trimethylation of histone H3 on lysine 27 (H3K27me3), respectively [26–30]. Consistent with these results, FUS3 and ABI3 are de-repressed in *curly leaf* (*clf*) mutant seeds (*CLF* encodes one of the methyltransferase of PRC2 complexes), leading to ectopic accumulation of oil and enhanced seed growth [31]. Moreover, LAFL genes are also marked by the activating H3K4me3 modification in seed, indicating that both, their inhibition as well as their activation, are tightly regulated by chromatin-based mechanisms [29,32]. Consistently, two chromatin remodelers of the CHD3 family (PICKLE and PRK2) are also involved in H3K27me3-dependent LAFL repression, although how these proteins affect PRC2 function remains a matter of debate [33–35]. Conversely, proper induction of ABI3 and FUS3 in seeds, requires another chro-





**Fig. 1.** Schematic representation of the main seed architectures. In mature seeds of Arabidopsis, the two cotyledons are the main storage organs and the embryo fills most of the seed cavity. A single-cell layer of remaining endosperm (the aleurone layer) surrounds the embryo. In quinoa a perispermic dicot seed, most of the reserves accumulate in the perisperm, a maternal tissue originating from the nucellus. In maize grain, the mature seed contains a small embryo surrounded by a large persistent triploid endosperm, which is the main storage organ.

matin remodeler of the CHD1 family (i.e. CHR5) for establishing an active chromatin state [36]. SCARECROW-LIKE15 (SCL15), a member of the plant specific GRAS family of transcription factors, binds to HISTONE DEACETYLASE19 (HDA19) leading to the repression of *ABI3* and *LEC1* [37]. Moreover, repression of the *LAF1* genes in seedlings is reinforced by a family of VIVIPAROUS1/ABI3-LIKE (VAL) transcription factors that also contain a B3 domain [25,38]. VAL1 and VAL2 are thought to repress *AFL* expression by modifying chromatin structure and to compete with the *AFL* proteins by binding to their target promoters.

Post-transcriptional gene silencing (PTGS) also contributes to the repression of *LEC2* and *FUS3* via two *miRNAs* (i.e. *miRNA* 156 and 166) [39–41]. The *miR166* inhibits *LEC2* in vegetative tissues by repressing *PHABULOSA* and *PHAVOLUTA* that encode two HD-ZIPs, the former being a direct activator of *LEC2* [41]. Similarly, by silencing *SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE 10* (*SPL10*) and *SPL11* expression during early embryogenesis, *miR156* represses *LEC2* and *FUS3*, thus preventing precocious onset of the maturation program during seed development [39,40]. Last, acting downstream of the *miRNA* pathway, the trihelix transcription factors ASIL1 and ASIL2, and the histone deacetylase HDA6/SIL1 also participate in the repression of *LEC2* and *FUS3* during early embryogenesis [40]. Nevertheless, the underlying molecular mechanisms remain to be ascertained.

### 2.3. *LAF1* targets and partners

Our current understanding of the regulatory interactions between the *LAF1* and their target genes, with regard to control of storage compound accumulation during seed development is summarized in Fig. 3. The three *AFL*-B3 proteins bind “RY” DNA motifs (including a 5′-CATG-3′ core sequence) both in vitro and in vivo, but with different specificities regarding the flanking nucleotides of the core sequence [19,21][19,21 and references therein]. The direct targets genes of the *AFL*-B3 proteins include genes encoding seed storage proteins (i.e. cruciferins and napins) and proteins involved in oil storage (e.g. oleosins), as well as transcriptional regulators such as AGAMOUS-LIKE15 (*AGL15*), MYB118, and WRINKLED1 (*WRI1*) [42–44]. *AGL15* encodes a MADS-domain transcription factor interacting with the *LAF1* network [42]. Indeed, *AGL15* controls *LEC2* expression through a putative positive feedback loop, limited by the fact that *AGL15* negatively regulates itself. Moreover, *agl15* mutant seeds do not display *lec2* phenotypes, suggesting that *AGL15* acts redundantly with paralogous MADS proteins to control the network. MYB118 is specifically expressed in the endosperm, where it represses both *LEC2* expression and reserve deposition

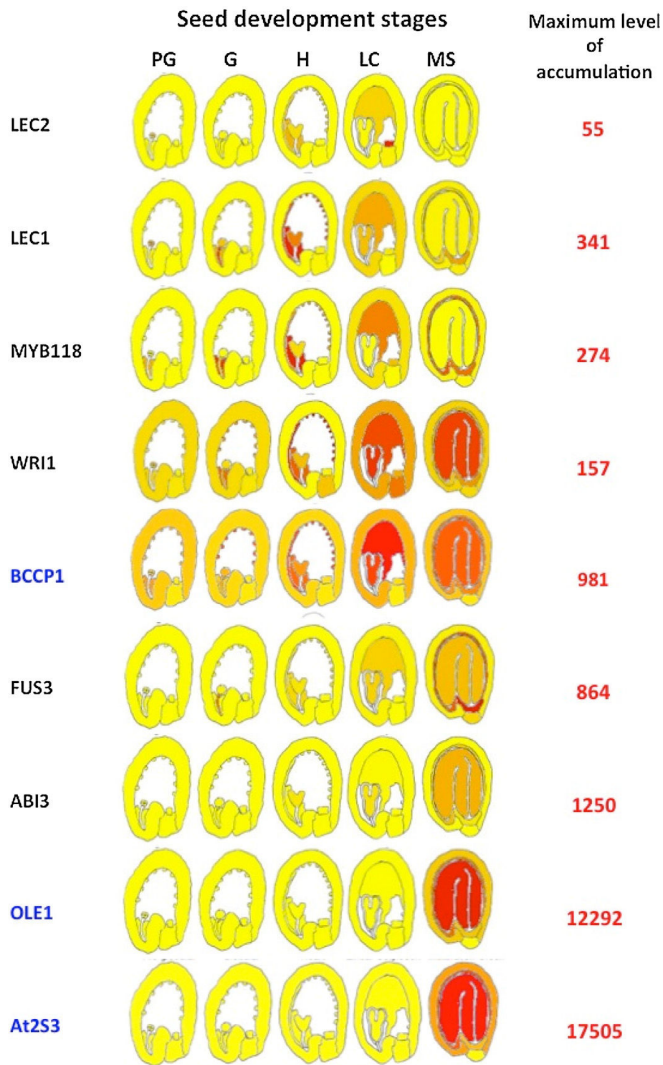
[43]. Thus, the specific expression of *MYB118* in the endosperm may result from the activity of endosperm-specific activators and/or embryo-specific repressors cooperating with *LEC2*. *WRI1* belongs to the AP2/EREB family of TFs and controls the expression of genes encoding enzymes of the late glycolysis and the fatty acid biosynthetic network [45]. The reduced seed oil content observed in *lec2* and *lec1* mutants is partly due to the down-regulation of *WRI1* [46]. This regulatory cascade allows fine-tuning the expression of fatty acid biosynthetic genes in seeds [7].

In addition to RY motifs, the presence of G-Box elements is required for proper activation of target promoters of the *LAF1*, *in planta* [21]. This result suggested that the *AFL*-B3 collaborate with bZIPs TFs that interact with these G-Box elements. Consistent with this hypothesis, some bZIPs (e.g. bZIP10, bZIP25 and bZIP53 or bZIP67 and bZIP28) interact with *ABI3* or *LEC1*, respectively, to activate seed specific promoters [4,21][for references see 4,21]. In addition, although bZIP53 does not interact directly with *ABI3* it can form ternary complexes with other bZIPs and *ABI3*, increasing the expression of target genes [47]. *LEC1* also activates gene expression through G-box elements, by interacting with PIF4 (a bHLH), during post-embryonic development [48,49]. More broadly, some bZIPs and other transcription factors such as DOF or MYB proteins also play key roles during cereal grains filling [4][4 and references inside]. Among them OPAQUE2 (O2), a maize kernel-specific transcription factor interact with a MADS factor (ZmMADS47) to regulate the expression of *ZEIN* genes, which encode the predominant storage proteins in maize [50]. HvVP1, the barley ortholog of *ABI3*, modulates the activity of bZIP, DOF and MYB transcription factors to control gene expression in both endosperm and embryo [4], as described below. Therefore, in addition to *LAF1*, transcription factors of the MYB, bZIP, MADS, DOF or AP2 families are also involved in the maturation regulatory network, as protein partners or direct target genes. They modulate and/or specify the function of the *LAF1* network, providing variations in the end-targets of the transcriptional cascade, thus, contributing to specify the accumulation of a variety of storage compounds (i.e. oil, starch or seed storage proteins) in various seed tissues (embryo or endosperm).

## 3. Engineering new crops with modified storage reserves

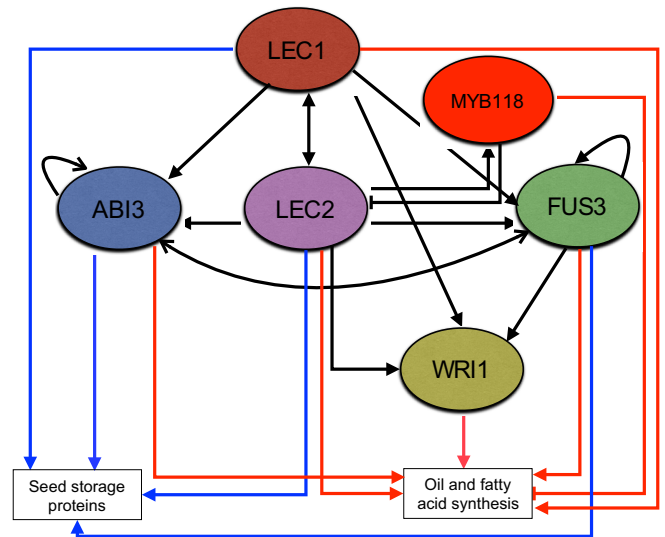
### 3.1. *LAF1* functions are conserved in angiosperms

*LAF1* orthologs are found in most angiosperms studied so far, including agronomically important monocots such as maize, rice, wheat, and barley [51–53]. Most of these genes are expressed in the seed and control the expression of both structural and regu-



**Fig. 2.** Expression pattern of the LAFL and some of their target genes. Transcriptomic data from Goldberg and Harada Labs [90] (<http://seedgenenetwork.net>) are displayed using the Arabidopsis eFP Browser [91] (<http://bar.utoronto.ca/efp>) and the « absolute » mode. The different stages of seed development are pre-globular embryo (PG), globular (G), heart (H), linear cotyledons (LC) and mature seed (MS). The maximum level of accumulation for each mRNA is indicated and corresponds to the red colour in the scheme. The yellow colour indicates that the mRNA is not detected.

latory genes involved in the accumulation of storage compounds. For instance, the loss of *FUS3* ortholog in *Brassica napus*, which is evolutionary closely related to Arabidopsis, results in phenotypes similar to those observed in Arabidopsis *fus3* mutants seeds, including the accumulation of anthocyanin, a marked decrease in seed oil content, and a concomitant increase in the amount of seed storage proteins [54]. Additionally, *Bn fus3* mutants are affected in the expression of other LAFL genes and display a reduction in seed yield by 12% due to the reduced number of siliques per plant. Interestingly, orthologous genes have also been characterized in monocots such as barley *HvFUS3* and *HvVP1* orthologs of *FUS3* and *ABI3*, respectively [4,55]. *HvFUS3* complements the Arabidopsis *fus3* mutants by restoring the proper biosynthesis of seed storage proteins and weak accumulation of anthocyanins. The gene is expressed in both, the developing barley endosperm and embryo, with higher expression level in the endosperm, and promotes the expression of seed storage proteins. More surprisingly, *HvVP1* acts as a negative regulator repressing *B-HORDEIN* expression during



**Fig. 3.** Schematic representation of the gene regulatory network controlling seed maturation. Arrows and blunted lines indicate activation and repression, respectively. Colors of the arrows stand for the regulation of transcription factors (black), seed storage proteins (blue) and genes involved in biosynthesis and storage of lipids (red).

early development of the endosperm and inhibiting the activation of alpha-amylase by *HvGAMYB* in the embryo [55]. Genetic, expression, and functional analyses of five maize AFL-B3 orthologs (*ZmAFL1* to 5) also support the conservation of *FUS3* and *ABI3* functions [56]. *ZmAFL4*, which is expressed in seed, may have a different function in the endosperm, being involved in carbon metabolism and starch accumulation [56]. The spatio-temporal expression pattern of *LEC1* homologs in rye (*ScLEC1*), maize (*ZmLEC1*), and radish also suggest that the role of *LEC1* during seed development is highly conserved in these crops [57–59]. Wheat LAFL homologs (i.e. *TaL1L*, *TaL2L* and *TaFUS3*) are also expressed during early stage of seed development, but functional analyses are required to confirm their orthologous function [60]. Homologs of *WRI1* have been also found in all species investigated such as *Brassica napus* (*BnWRI1*), maize (*ZmWRI1*), or oil palm (*EgWRI1*), and successfully complement Arabidopsis *wri1* mutants by restoring wild-type seed oil content and seed coat phenotypes [61–63]. Nevertheless, in contrast with *BnWRI1* and *ZmWRI1* that are mainly expressed in the developing seeds, *EgWRI1* is expressed in the mesocarp tissue of oil palm [62].

These results clearly establish that the global function of the LAFL network (i.e. to activate maturation pathways and inhibit germination) is conserved in angiosperms, but can take place in different tissues depending on seed architecture (e.g. cereals which accumulate starch in the endosperm or oleoproteaginous such as rapeseed which store most of their reserves in the embryo). Therefore, a comprehensive functional analysis of these orthologs remains an important challenge for deciphering possible variations in their functioning (protein partners and target genes) in different storage tissues of various crop seeds. The recent development of CRISPR/Cas9 genome editing system provides new rapid, efficient, versatile and powerful tools to tackle this challenge (i.e. genetic and functional analyses of gene families in complex plant genomes) and for the improvement of crops [64]. For instance, it has been successfully used for generating targeted and heritable mutations of two paralogous genes in barley or a specific member of a gene family in *Brassica oleracea* [65].

### 3.2. Using LALF and WRI genes for engineering crops

Consistent with their conservation and potential agronomic interest, the *LALF* and *WRI* genes have already been successfully used to promote oil accumulation not only in seeds of brassicaceae, but also in leaves, tubers and seeds in other crops [44,66–69]. For instance, overexpression of *AtLEC1*, and of its homologs in rapeseed (*BnLEC1*) and maize (*ZmLEC1*) induces an increase in seed oil content [59,70,71]. Overexpression of *ZmWRI1* promotes oil accumulation in transgenic maize grains by 30–46% of total oil, without any detrimental effect on agronomic traits such as early stand counts, flowering time, plant height, or seed yield [59]. This higher oil content is mainly due to a redirection of the carbon flux from starch biosynthesis to oil biosynthesis in the embryos of the transgenic grains. However, some detrimental effects on agronomic traits such as grain yield (total grain weight per plant) or seed vigor have been encountered [15,66,70,72]. Using promoters that confer specific expression patterns (i.e. triggering moderate, tissue- or time-specific ectopic expression) allows preventing these undesirable effects. For instance, the extension of the duration of *WRI1* expression during mid-phase of Arabidopsis seed development, using the *FUS3* promoter, significantly enhances seed oil content [73]. Moreover, the overexpression of *WRI* rescues feedback inhibition of fatty acid synthesis in hydroxylase-expressing Arabidopsis seeds, leading to higher accumulation of both, hydroxyl fatty acids and total oil content [74].

Last, aside from the LALF network, other regulators that control seed size and represent potentially interesting targets for obtaining larger seeds through biotech approaches [13,75,76]. Although there is no direct link between seed size and yield, there are a few reports of positive agronomic impacts of larger seeds. For instance, the overexpression of *OsSPL13*, which belongs to the SBP family of transcription factors, results in an increase in seed size, weight and yield in transgenic rice [77]. In wheat, the overexpression of *TaGS5-3A* (encoding a putative carboxypeptidase) in the endosperm induces an up regulation of genes associated with cell proliferation and a subsequent increase in grain size of transgenic lines [78].

## 4. Conclusions and prospects

The regulators of seed development and maturation that consist of the *LALF* genes and their associated regulatory network have been well studied in Arabidopsis. These master regulators are crucial for proper embryogenesis, accumulation of storage reserves and seed maturation. The conservation of this network among higher plants should facilitate knowledge translation to crop improvement by classical breeding or genetic engineering. While genetic engineering of crops with altered seed storage reserves using *LALF* and *WRI* genes is very promising, further investigations are required to confirm agronomical interest for major cereal crops and legumes such as soybean, pea, or alfalfa. It might reveal functional variations between known orthologous transcription factors, or uncover new actors, that may impact the mechanisms controlling seed quality in different plants. On this line, the variability in binding affinities of the AFL-B3 over conserved partners and/or targets genes readily offers the possibility of modulating carbon partitioning and fatty acid composition by biotech approaches. The availability of new, fast, and robust tools for genome editing should allow a better understanding of the LALF regulatory network in important crops, and could help speeding up the engineering of adapted crops with modified seed storage reserves to fulfill agronomic or industrial needs. This turns out to be critical in the light of environmental (biotic or abiotic) impact on seed production, maturation and quality [79,80].

For instance, heat stress negatively impacts seed development, yield and quality (e.g. accumulation of storage compounds, storability, and the ability to germinate under various conditions) have been observed for various crops, including rapeseed, wheat, barley, soybean or maize [81–83]. Over the past three decades, heat has caused a decrease of about 5% in the global maize and wheat production. Projections predict a drastic decline in the production of major cereals by 2020, including 9% for maize, 11% for rice, and 14% for wheat [83]. Similarly, biotic stress has also detrimental effects on the LALF network and, thus, seed weight and quality [84].

However, current knowledge about the specific impact of stress during seed development and about the mechanisms involved is limited [85–87]. A better understanding of the molecular and genetic regulation of seed development and maturation by environmental factors would provide new molecular tools for improving crops and agronomic practices [88]. More specifically, it has been demonstrated that the expression of the *LALF* genes is controlled by epigenetic mechanisms. This likely underpins the plasticity and the ability of plants to adapt swiftly to ever changing environmental conditions [89]. Thus, the response of the LALF network and its chromatin landscape to environmental stress, may contribute to the development of crops better adapted to climate change.

## Acknowledgments

We apologize to all our colleagues whose works and original articles could not be cited because of space limitations and the request to cite only the most recent references. A.F. is supported by Agreenskills fellowship (FP7-26719) and ABCEEDS program (ERA-CAPS, FP7 #291864). Some work presented in this review was supported by CERES project (ANR-10-BLAN-1238). We thank the editor and anonymous reviewers for numerous suggestions that strongly improve the manuscript, F. Roudier for discussions and our colleagues at IJPB E. Fiume, D. Grain, M. Miquel, J. Thévenin, and A. To for their support to different projects related to this work. The IJPB benefits from the support of the Labex Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

## References

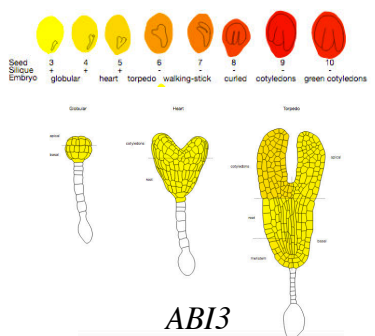
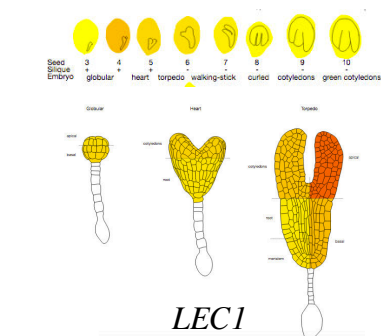
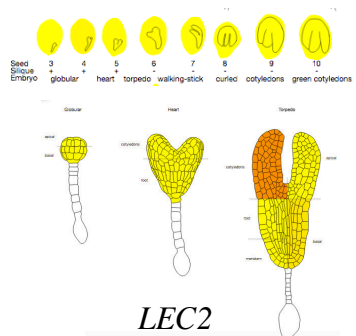
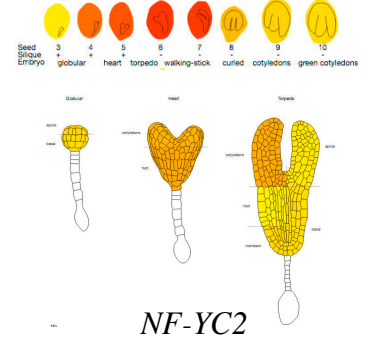
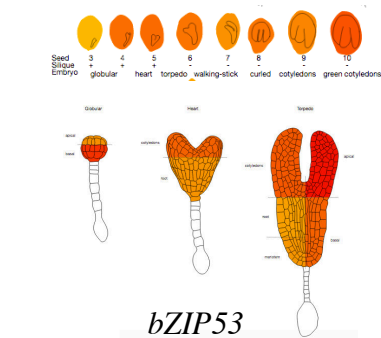
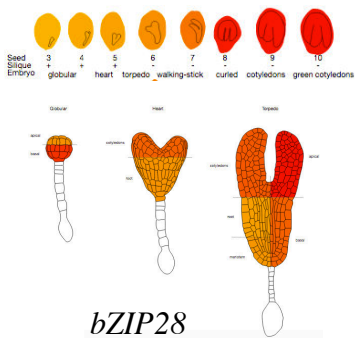
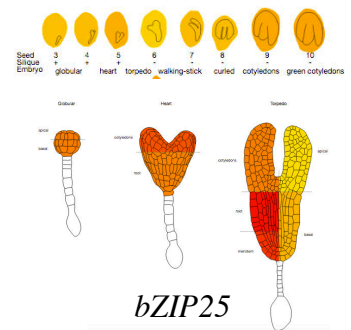
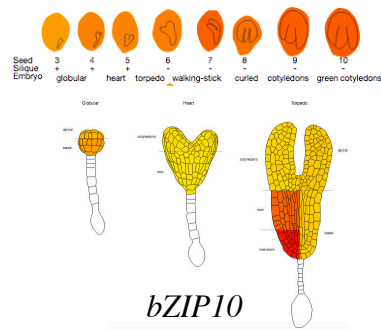
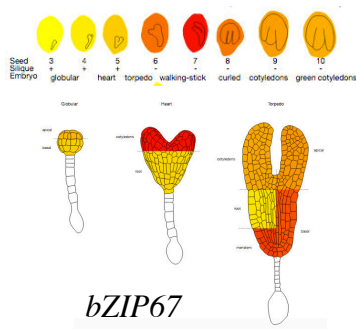
- [1] C. Lesk, P. Rowhani, N. Ramankutty, Influence of extreme weather disasters on global crop production, *Nature* 529 (2016) 84–87.
- [2] N. Sreenivasulu, U. Wobus, Seed-development programs: a systems biology-based comparison between dicots and monocots, *Annu. Rev. Plant Biol.* 64 (2013) 189–217.
- [3] M.K. Nowack, A. Ungru, K.N. Bjerkan, P.E. Grini, A. Schnittger, Reproductive cross-talk: seed development in flowering plants, *Biochem. Soc. Trans.* 38 (2010) 604–612.
- [4] Z. Abraham, et al., A developmental switch of gene expression in the barley seed mediated by HvVP1 (Viviparous1) and HvGAMYB interactions, *Plant Physiol.* 170 (2016) 2146–2158.
- [5] G. Barthole, L. Lepiniec, P.M. Rogowsky, S. Baud, Controlling lipid accumulation in cereal grains, *Plant Sci.* 185–186 (2012) 33–39.
- [6] H. Weber, L. Borisjuk, U. Wobus, Molecular physiology of legume seed development, *Annu. Rev. Plant Biol.* 56 (2005) 253–279.
- [7] S. Baud, L. Lepiniec, Physiological and developmental regulation of seed oil production, *Prog. Lipid Res.* 49 (2010) 235–249.
- [8] H.P. Burrieza, M.P. Lopez-Fernandez, S. Maldonado, Analogous reserve distribution and tissue characteristics in quinoa and grass seeds suggest convergent evolution, *Front. Plant Sci.* 5 (2014) 546.
- [9] H. North, et al., Arabidopsis seed secrets unravelled after a decade of genetic and omics-driven research, *Plant J.* 61 (2010) 971–981.
- [10] M.G. Becker, S.W. Hsu, J.J. Harada, M.F. Belmonte, Genomic dissection of the seed, *Front. Plant Sci.* 5 (2014) 464.
- [11] H. Xu, Y. Gao, J. Wang, Transcriptomic analysis of rice (*Oryza sativa*) developing embryos using the RNA-Seq technique, *PLoS One* 7 (2012) e30646.
- [12] J. Chen, et al., Dynamic transcriptome landscape of maize embryo and endosperm development, *Plant Physiol.* 166 (2014) 252–264.
- [13] G. Orozco-Arroyo, D. Paolo, I. Ezquer, L. Colombo, Networks controlling seed size in Arabidopsis, *Plant Reprod.* 28 (2015) 17–32.
- [14] P.E. Jameson, J. Song, Cytokinin: a key driver of seed yield, *J. Exp. Bot.* 67 (2016) 593–606.



- [15] A. Fatihi, A.M. Zbierzak, P. Dormann, Alterations in seed development gene expression affect size and oil content of Arabidopsis seeds, *Plant Physiol.* 163 (2013) 973–985.
- [16] S.A. Braybrook, J.J. Harada, LECs go crazy in embryo development, *Trends Plant Sci.* 13 (2008) 624–630.
- [17] M. Santos-Mendoza, et al., Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis, *Plant J.* 54 (2008) 608–620.
- [18] J. Verdier, R.D. Thompson, Transcriptional regulation of storage protein synthesis during dicotyledon seed filling, *Plant Cell Physiol.* 49 (2008) 1263–1271.
- [19] T.T. Roscoe, J. Guillemot, J.J. Bessoule, F. Berger, M. Devic, Complementation of seed maturation phenotypes by ectopic expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in Arabidopsis, *Plant Cell Physiol.* 56 (2015) 1215–1228.
- [20] A. Yamamoto, et al., Cell-by-cell developmental transition from embryo to post-germination phase revealed by heterochronic gene expression and ER-body formation in Arabidopsis leafy cotyledon mutants, *Plant Cell Physiol.* 55 (2014) 2112–2125.
- [21] S. Baud, et al., Deciphering the molecular mechanisms underpinning the transcriptional control of gene expression by L-AFL proteins in Arabidopsis seed, *Plant Physiol.* 171 (2016) 1099–1112.
- [22] F. Delmas, et al., ABI3 controls embryo degreening through Mendel's 1 locus, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 3888–3894.
- [23] A. Junker, H. Baumlein, Multifunctionality of the LEC1 transcription factor during plant development, *Plant Signal. Behav.* 7 (2012) 1718–1720.
- [24] B. Wojcikowska, et al., LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of Arabidopsis, via YUCCA-mediated auxin biosynthesis, *Planta* 238 (2013) 425–440.
- [25] H. Jia, M. Suzuki, D.R. McCarty, Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks, *Wiley Interdiscip. Rev. Dev. Biol.* 3 (2014) 135–145.
- [26] N. Berger, B. Dubreucq, F. Roudier, C. Dubos, L. Lepiniec, Transcriptional regulation of Arabidopsis LEAFY COTYLEDON2 involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27, *Plant Cell* 23 (2011) 4065–4078.
- [27] D. Bouyer, et al., Polycomb repressive complex 2 controls the embryo-to-seedling phase transition, *PLoS Genet.* 7 (2011) e1002014.
- [28] C. Yang, et al., VAL- and AtBMI1-mediated H2Aub initiate the switch from embryonic to postgerminative growth in Arabidopsis, *Curr. Biol.* 23 (2013) 1324–1329.
- [29] A.M. Molitor, Z. Bu, Y. Yu, W.H. Shen, Arabidopsis AL PHD-PRC1 complexes promote seed germination through H3K4me3-to-H3K27me3 chromatin state switch in repression of seed developmental genes, *PLoS Genet.* 10 (2014) e1004091.
- [30] M. Ikeuchi, et al., PRC2 represses dedifferentiation of mature somatic cells in Arabidopsis, *Nat. Plants* 1 (2015) 15089.
- [31] J. Liu, et al., CURLY LEAF regulates gene sets coordinating seed size and lipid biosynthesis in Arabidopsis, *Plant Physiol.* 171 (2016) 424–436.
- [32] K. Muller, D. Bouyer, A. Schnittger, A.R. Kermode, Evolutionarily conserved histone methylation dynamics during seed life-cycle transitions, *PLoS One* 7 (2012) e51532.
- [33] B. Carter, et al., Crosstalk between sporophyte and gametophyte generations is promoted by chd3 chromatin remodelers in Arabidopsis thaliana, *Genetics* (2016), <http://dx.doi.org/10.1534/genetics.1115.180141>.
- [34] E. Aichinger, C.B. Villar, R. Di Mambro, S. Sabatini, C. Kohler, The CHD3 chromatin remodeler PICKLE and polycomb group proteins antagonistically regulate meristem activity in the Arabidopsis root, *Plant Cell* 23 (2011) 1047–1060.
- [35] S.K. Han, M.F. Wu, S. Cui, D. Wagner, Roles and activities of chromatin remodeling ATPases in plants, *Plant J.* 83 (2015) 62–77.
- [36] Y. Shen, M. Devic, L. Lepiniec, D.X. Zhou, Chromodomain, Helicase and DNA-binding CHD1 protein, CHR5, are involved in establishing active chromatin state of seed maturation genes, *Plant Biotech. J.* 13 (2015) 811–820.
- [37] M.J. Gao, et al., SCARECROW-LIKE15 interacts with HISTONE DEACETYLASE19 and is essential for repressing the seed maturation programme, *Nat. Commun.* 6 (2015) 7243.
- [38] A. Schneider, et al., Potential targets of VIVIPAROUS1/ABI3-LIKE1 (VAL1) repression in developing Arabidopsis thaliana embryos, *Plant J.* 85 (2016) 305–319.
- [39] M.D. Nodine, D.P. Bartel, MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis, *Genes Dev.* 24 (2010) 2678–2692.
- [40] M.R. Willmann, A.J. Mehalick, R.L. Packer, P.D. Jenik, MicroRNAs regulate the timing of embryo maturation in Arabidopsis, *Plant Physiol.* 155 (2011) 1871–1884.
- [41] X. Tang, et al., MicroRNA-mediated repression of the seed maturation program during vegetative development in Arabidopsis, *PLoS Genet.* 8 (2012) e1003091.
- [42] Y. Zheng, N. Ren, H. Wang, A.J. Stromberg, S.E. Perry, Global identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like15, *Plant Cell* 21 (2009) 2563–2577.
- [43] G. Barthole, et al., MYB118 represses endosperm maturation in seeds of Arabidopsis, *Plant Cell* 26 (2014) 3519–3537.
- [44] C. Marchive, K. Nikovics, A. To, L. Lepiniec, S. Baud, Transcriptional regulation of fatty acid production in higher plants: molecular bases and biotechnological outcomes, *Eur. J. Lipid Sci. Technol.* 116 (2014) 1332–1343.
- [45] A. To, et al., WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in Arabidopsis, *Plant Cell* 24 (2012) 5007–5023.
- [46] S. Baud, et al., WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis, *Plant J.* 50 (2007) 825–838.
- [47] R. Alonso, et al., A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation, *Plant Cell* 21 (2009) 1747–1761.
- [48] A. Junker, et al., Elongation-related functions of LEAFY COTYLEDON1 during the development of Arabidopsis thaliana, *Plant J.* 71 (2012) 427–442.
- [49] M. Huang, Y. Hu, X. Liu, Y. Li, X. Hou, Arabidopsis leafy cotyledon1 mediates postembryonic development via interacting with phytochrome-interacting factor4, *Plant Cell* 27 (2015) 3099–3111.
- [50] Z. Qiao, et al., ZmMADS47 regulates zein gene transcription through interaction with opaque2, *PLoS Genet.* 12 (2016) e1005991.
- [51] A. Cagliari, et al., New insights on the evolution of Leafy cotyledon1 (LEC1) type genes in vascular plants, *Genomics* 103 (2014) 380–387.
- [52] F.Y. Peng, R.J. Weselake, Genome-wide identification and analysis of the B3 superfamily of transcription factors in Brassicaceae and major crop plants, *Theor. Appl. Genet.* 126 (2013) 1305–1319.
- [53] Y. Wang, et al., Systematic analysis of plant-specific B3 domain-containing proteins based on the genome resources of 11 sequenced species, *Mol. Biol. Rep.* 39 (2012) 6267–6282.
- [54] N. Elahi, R.W. Duncan, C. Stasolla, Decreased seed oil production in FUSCA3 Brassica napus mutant plants, *Plant Physiol. Biochem.* 96 (2015) 222–230.
- [55] M.A. Moreno-Risueno, et al., FUSCA3 from barley unveils a common transcriptional regulation of seed-specific genes between cereals and Arabidopsis, *Plant J.* (2008), <http://dx.doi.org/10.1111/j.1365-1313X.2007.03382.x>.
- [56] A. Grimault, et al., Role of B3 domain transcription factors of the AFL family in maize kernel filling, *Plant Sci.* 236 (2015) 116–125.
- [57] A. Gruszczynska, M. Rakoczy-Trojanowska, Expression analysis of somatic embryogenesis-related SERK, LEC1 VP1 and NiR orthologues in rye (*Secale cereale* L.), *J. Appl. Genet.* 52 (2011) 1–8.
- [58] L. Zhai, et al., Transcriptional identification and characterization of differentially expressed genes associated with embryogenesis in radish (*Raphanus sativus* L.), *Sci. Rep.* 6 (2016) 21652.
- [59] B. Shen, et al., Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize, *Plant Physiol.* 153 (2010) 980–987.
- [60] K. Rikiishi, M. Maekawa, Seed maturation regulators are related to the control of seed dormancy in wheat (*Triticum aestivum* L.), *PLoS One* 9 (2014) e107618.
- [61] B. Pouvreau, et al., Duplicate maize Wrinkled1 transcription factors activate target genes involved in seed oil biosynthesis, *Plant Physiol.* 156 (2011) 674–686.
- [62] W. Ma, et al., Wrinkled1, a ubiquitous regulator in oil accumulating tissues from Arabidopsis embryos to oil palm mesocarp, *PLoS One* 8 (2013) e68887.
- [63] X. Wang, et al., New insights into the genetic networks affecting seed fatty acid concentrations in Brassica napus, *BMC Plant Biol.* 15 (2015) 1–18.
- [64] M.K. Samanta, A. Dey, S. Gayen, CRISPR/Cas9: an advanced tool for editing plant genomes, *Transgenic Res.* (2016), <http://dx.doi.org/10.1007/s11248-11016-19953-11245>.
- [65] T. Lawrenson, et al., Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease, *Genome Biol.* 16 (2015) 258.
- [66] Q. Li, et al., WRINKLED1 accelerates flowering and regulates lipid homeostasis between oil accumulation and membrane lipid anabolism in Brassica napus, *Front. Plant Sci.* 6 (2015) 1015.
- [67] P. Hofvander, et al., Potato tuber expression of Arabidopsis WRINKLED1 increase triacylglycerol and membrane lipids while affecting central carbohydrate metabolism, *Plant Biotech. J.* (2016), <http://dx.doi.org/10.1111/pbi.12550>.
- [68] D. An, M.C. Suh, Overexpression of Arabidopsis WRI1 enhanced seed mass and storage oil content in Camelina sativa, *Plant Biotech. Rep.* 9 (2015) 137–148.
- [69] J. Zale, et al., Metabolic engineering of sugarcane to accumulate energy-dense triacylglycerols in vegetative biomass, *Plant Biotech. J.* 14 (2016) 661–669.
- [70] H. Tan, et al., Enhanced seed oil production in canola by conditional expression of Brassica napus LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds, *Plant Physiol.* 156 (2011) 1577–1588.
- [71] N. Elahi, R.W. Duncan, C. Stasolla, Modification of oil and glucosinolate content in canola seeds with altered expression of Brassica napus LEAFY COTYLEDON1, *Plant Physiol. Biochem.* 100 (2016) 52–63.
- [72] Y. Yang, et al., ectopic expression of WRINKLED1 affects fatty acid homeostasis in Brachypodium distachyon vegetative tissues, *Plant Physiol.* 169 (2015) 1836–1847.
- [73] M. Kanai, S. Mano, M. Kondo, M. Hayashi, M. Nishimura, Extension of oil biosynthesis during the mid-phase of seed development enhances oil content in Arabidopsis seeds, *Plant Biotech. J.* 14 (2016) 1241–1250.
- [74] N.D. Adhikari, P.D. Bates, J. Browse, WRINKLED1 rescues feedback inhibition of fatty acid synthesis in hydroxylase-expressing seeds, *Plant Physiol.* 171 (2016) 179–191.
- [75] Y. Zhang, et al., Transcription factors SOD7/NGAL2 and DPA4/NGAL3 act redundantly to regulate seed size by directly repressing KLU expression in Arabidopsis thaliana, *Plant Cell* 27 (2015) 620–632.

- [76] M. Noguero, et al., DASH transcription factor impacts *Medicago truncatula* seed size by its action on embryo morphogenesis and auxin homeostasis, *Plant J.* 81 (2015) 453–466.
- [77] L. Si, et al., OsSPL13 controls grain size in cultivated rice, *Nat. Genet.* 48 (2016) 447–456.
- [78] L. Ma, et al., TaGS5-3A, a grain size gene selected during wheat improvement for larger kernel and yield, *Plant Biotech. J.* 14 (2016) 1269–1280.
- [79] T. Obata, et al., Metabolite profiles of maize leaves in drought, heat, and combined stress field trials reveal the relationship between metabolism and grain yield, *Plant Physiol.* 169 (2015) 2665–2683.
- [80] S.D. Singer, J. Zou, R.J. Weselake, Abiotic factors influence plant storage lipid accumulation and composition, *Plant Sci.* 243 (2016) 1–9.
- [81] M.H. Siebers, et al., Heat waves imposed during early pod development in soybean (*Glycine max*) cause significant yield loss despite a rapid recovery from oxidative stress, *Glob. Chang. Biol.* 21 (2015) 3114–3125.
- [82] E. Mangelsen, et al., Transcriptome analysis of high-temperature stress in developing barley caryopses: early stress responses and effects on storage compound biosynthesis, *Mol. Plant* 4 (2011) 97–115.
- [83] C. Kole, et al., Application of genomics-assisted breeding for generation of climate resilient crops: progress and prospects, *Front. Plant Sci.* 6 (2015) 563.
- [84] E. Terrasson, et al., Identification of a molecular dialogue between developing seeds of *Medicago truncatula* and seedborne xanthomonads, *J. Exp. Bot.* 66 (2015) 3737–3752.
- [85] C. Chen, et al., Molecular characterization of rice endosperm development under heat stress identifies OsMADS87 as a determinant of seed size and thermal sensitivity, *Plant Physiol.* 171 (2016) 606–622.
- [86] J.L. Liao, et al., Transcriptome changes in rice (*Oryza sativa* L.) in response to high night temperature stress at the early milky stage, *BMC Genomics* 16 (2015) 18.
- [87] N. Sreenivasulu, et al., Designing climate-resilient rice with ideal grain quality suited for high-temperature stress, *J. Exp. Bot.* 66 (2015) 1737–1748.
- [88] C.E. Bitra, T. Gerats, Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops, *Front. Plant Sci.* 4 (2013) 273.
- [89] D. Grimanelli, F. Roudier, Epigenetics and development in plants: green light to convergent innovations, *Curr. Top. Dev. Biol.* 104 (2013) 189–222.
- [90] M.F. Belmonte, et al., Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 435–444.
- [91] D. Winter, et al., An Electronic Fluorescent Pictograph browser for exploring and analyzing large-scale biological data sets, *PLoS One* 2 (2007) e718.

**Annexe 2 :** Représentation de l'accumulation des ARNs messagers des bZIPs testés et de *NF-YC2*, *LEC1*, *LEC2*, *ABI3*, au cours du développement de la graine. (D'après Waese *et al* 2017) <http://bar.utoronto.ca/eplant>.



**Annexe 3** : Tableau récapitulatif des peptides identifiés en spectrométrie de masse de LEC2 recombinant

Peptides	charges	Modification	Masse théorique
GGFIDPR	2		761,394
DAEANIPK	2		857,437
NSDVGSIGR	2		904,449
EGIVVQMR	2		931,503
EGIVVQMR	2	M: +15.99491	947,498
QIMMINIK	2	M:+15.99491 - M:-17.02655	989,516
NNVQISDKK	2		1045,564
YKFWSNK	2		1086,537
RQIMMINIK	2		1146,649
RQIMMINIK	3	M:+15.99491	1162,644
RQIMMINIK	3	M:+15.99491	1178,639
EISMDPNNR	2	M:+15.99491	1205,522
EIKNSDVGSIGR	2		1274,67
SHFTTVPTYDH	2		1304,591
NIYFAMNGNSGK	2		1315,61
PTYDHHQAQPH	2		1330,593
NIYFAMNGNSGK	2	M:+15.99491	1331,605
DSYQQSTFDNK	2		1332,571
DVFSMQSWSFK	2		1361,62
ISDKEGIVVQMR	2		1374,741
DVFSMQSWSFK	2	M:+15.99491	1377,615
ISDKEGIVVQMR	2	M:+15.99491	1390,736
MYVIENTGEFVK	2		1429,704
SHFTTVPTYDHH	2		1441,65
MYVIENTGEFVK	2	M:+15.99491	1445,699
DSYQQSTFDNKK	2		1460,666
SVQEISMDPNNR	2		1503,686
SVQEISMDPNNR	2	M7:+15.99491	1519,681
KDSYQQSTFDNKK	2		1588,761
HFIPPFYYPVEQM	2		1591,762
NHYEEAMIDYIPR	2		1650,758
NNSSPNSSPSEIVDSK	2		1661,762
NHYEEAMIDYIPR	2	M:+15.99491	1666,753
ANSVQEISMDPNNR	2	M:+15.99491	1704,761
HFIPPFYYPVEQMAA	2	M:+15.99491	1749,831
NNSSPNSSPSEIVDSKR	2		1817,863
VLCEKELKNSDVGSLGR	3	M:+57.02146	1903,991
SRNNSSPNSSPSEIVDSK	2		1904,895
SNANSVQEISMDPNNR	2	M:+15.99491	1905,836
ERNHYEEAMLIDYIPR	2		1935,902
ERNHYEEAMIDYIPR	3	M:+15.99491	1951,897
SHFTTVPTYDHHQAQPH	2		2002,916
QNGAEIGDFITIEDESK	3		2028,94
SSNANSVQEISMDPNNR	2	M:+15.99491 - M:+42.01056	2034,879
SRNNSSPNSSPSEIVDSKR	3		2060,996
QAQPHHFIPPFYYPVEQM	3		2153,028
DVFSMQSWSFKYKFWSNK	3		2429,139
DVFSMQSWSFKYKFWSNK	3	M:+15.99491	2445,134
NNVQISDKKDSYQQSTFDNKK	3		2487,211
SHFTTVPTYDHHQAQPHHFIPPFYYPVEQM	3		3575,659

**Annexe 4 :** Tableau d'identification des protéines par spectrométrie de masse après purification des complexes protéiques par immunoprécipitation GFP :

Le tableau représente les résultats par groupe de protéines ayant les mêmes peptides identifiés, la protéine majoritaire du groupe apparait en premier.

Group ID : classement des protéines par groupe ayant les mêmes peptides identifiés.

Top Protein Description : description de la première protéine identifiée du groupe.

Protein location : localisation de la protéine dans la cellule décrite dans la base de données Tair (<http://www.arabidopsis.org>).

Number of proteins : nombre total de protéines du groupe.

GFP:GS: lignées témoins, les expériences ont été réalisées pour les lignées 10-2 et 4-6 à partir de noyaux purifiés, la lignée 17-4 à partir de l'extrait brut.

GFP:LEC2 : lignées étudiées, les expériences ont été réalisées pour les lignées 4-6 et 18-4 à partir de noyaux purifiés, la lignée 17-1 à partir de l'extrait brut.

En jaune les protéines présentent uniquement dans les échantillons, en rouge les résultats d'identification de la GFP.

Les critères de sélections des données, sont : Evalue peptidique maximale de 0,03, Evalue protéique maximale -4, nombre de peptide minimum par protéine de 1.



Group ID	Top Protein Description	Protein location	Number of proteins	GFP:GS 4-6	GFP:GS 17-4	GFP:GS 10-2	GFP:LEC2 17-1	GFP:LEC2 4-6	GFP:LEC2 18-4
a1	AT1G27090.1   glycine-rich protein	chloroplast, cytosol	1	11	2	2	3	3	0
a2	AT5G62690.1   Symbols: TUB2   tubulin beta chain 2	Golgi apparatus, cytoplasm, membrane, microtubule	2	11	0	0	2	2	0
a2	AT5G23860.1   Symbols: TUB8   tubulin beta 8	Golgi apparatus, cytoplasm, membrane, microtubule	2	8	0	0	2	3	0
a2	AT5G44340.1   Symbols: TUB4   tubulin beta chain 4	Golgi apparatus, cell wall, chloroplast, cytoplasm, microtubule, membrane,	1	10	0	0	3	0	0
a3	AT2G21660.1   Symbols: ATGRP7, CCR2, GR-RBP7, GRP7   cold, circadian rhythm, and rna binding 2	chloroplast, cytoplasm, mitochondrion, nucleus, peroxisome	1	10	2	1	2	1	0
a4	AT5G37720.1   Symbols: ALY4   ALWAYS EARLY 4	nucleolus, nucleoplasm	2	6	1	1	2	0	0
a4	AT1G66260.1   RNA-binding (RRM/RBD/RNP motifs) family protein	nucleolus, nucleoplasm, nucleus	2	4	0	0	2	0	0
a5	AT2G42520.1   P-loop containing nucleoside triphosphate hydrolases superfamily protein	peroxisome	1	7	2	0	0	0	0
a5	AT3G58510.1   DEA(D/H)-box RNA helicase family protein	nucleolus, peroxisome, membrane	3	5	0	0	3	0	0
a6	AT2G18020.1   Symbols: EMB2296   Ribosomal protein L2 family	chloroplast, cytoplasm, nucleolus, plasma membrane	1	3	2	2	2	0	1
a7	AT4G20360.1   Symbols: ATRAB8D, ATRABE1B, RABE1b   RAB GTPase homolog E1B	apoplast, chloroplast, membrane, nucleolus, nucleus, plastid	1	5	0	0	5	0	0
a8	AT2G41840.1   Symbols:   Ribosomal protein S5 family protein	cytosol, membrane, nucleolus, plasmodesma	1	4	1	0	4	0	0
a8	AT1G58380.1   Symbols: XW6   Ribosomal protein S5 family protein	chloroplast, cytosolic, membrane, plasmodesma	4	1	0	0	3	0	0
a9	AT5G02530.1   RNA-binding (RRM/RBD/RNP motifs) family protein	nucleoplasm	2	8	0	0	0	0	1
b10	AT5G02870.1   Ribosomal protein L4/L1 family   chr5:657830-659526 FORWARD LENGTH=407	chloroplast, cytoplasm, membrane, nucleolus, plasma membrane, plasmodesma, vacuole	2	3	0	2	1	0	0
b10	AT3G09630.1   Ribosomal protein L4/L1 family	cell wall, chloroplast, cytoplasm, membrane, nucleolus, plasma membrane, plasmodesma, vacuole	2	5	0	0	1	0	0
b11	AT2G47610.1   Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	chloroplast, cytoplasm, nucleolus, plasmodesma, vacuolar membrane	2	4	0	2	2	0	0
b12	AT5G59950.1   RNA-binding (RRM/RBD/RNP motifs) family protein	nucleolus, nucleoplasm, nucleus	3	5	0	2	0	0	0
b12	AT5G59950.2   RNA-binding (RRM/RBD/RNP motifs) family protein	nucleolus, nucleoplasm, nucleus	1	4	0	1	0	0	0
b13	AT1G48920.1   Symbols: ATNUC-L1, PARL1, NUC-L1   nucleolin like 1	mitochondrion, nucleolus	1	4	2	0	1	0	0
b14	AT1G07920.1   GTP binding Elongation factor Tu family protein	chloroplast, cytoplasm, membrane, mitochondrion, nucleolus, plasmodesma, vacuolar membrane	6	7	0	0	0	0	0

Group ID	Top Protein Description	Protein location	Number of proteins	GFP:GS 4-6	GFP:GS 17-4	GFP:GS 10-2	GFP:LEC2 17-1	GFP:LEC2 4-6	GFP:LEC2 18-4
b15	AT3G05060.1   NOP56-like pre RNA processing ribonucleoprotein	membrane, nucleolus, nucleus	1	5	0	1	0	0	0
b15	AT5G27120.1   NOP56-like pre RNA processing ribonucleoprotein	nucleolus, nucleus, plasmodesma	1	2	0	0	0	0	0
b16	AT1G56110.1   Symbols: NOP56   homolog of nucleolar protein NOP56	cell wall, nucleolus, nucleus, plasmodesma	1	7	0	0	0	0	0
b17	AT3G08580.1   Symbols: AAC1   ADP/ATP carrier 1	cell wall, chloroplast, membrane, mitochondrion, nucleolus, vacuole	2	6	0	0	0	0	0
b17	AT5G13490.1   Symbols: AAC2   ADP/ATP carrier 2	chloroplast, mitochondrion, vacuolar membrane	2	4	0	0	0	0	0
b18	AT4G22770.1   AT hook motif DNA-binding family protein	nucleus	1	3	2	1	0	0	0
b19	AT2G31610.1   Ribosomal protein S3 family protein	chloroplast, cytoplasm, membrane, plasmodesma, vacuole	1	4	0	1	1	0	0
b20	AT2G07698.1   ATPase, F1 complex, alpha subunit protein	chloroplast, membrane, mitochondrion, nucleolus, plasma membrane, vacuole	2	4	2	0	0	0	0
b21	AT3G25520.1   Symbols: ATL5, PGY3, OLI5, RPL5A   ribosomal protein L5	chloroplast, cytoplasm, membrane, nucleolus, nucleus, plasma membrane, ribosome, vacuole	1	4	0	0	2	0	0
b22	AT2G44120.1   Ribosomal protein L30/L7 family protein	chloroplast, cytoplasm, cytosol, membrane, nucleolus, plasmodesma, vacuole	2	4	0	1	0	0	0
b22	AT2G01250.1   Ribosomal protein L30/L7 family protein	cell wall, chloroplast, cytoplasm, membrane, vacuole	1	3	0	0	0	0	0
b23	AT1G22780.1   Symbols: PFL, RPS18A, PFL1   Ribosomal protein S13/S18 family	cell wall, cytoplasm, plasma membrane, plasmodesma, vacuolar membrane	3	6	0	0	0	0	0
b24	AT5G01530.1   Symbols: LHCB4.1   light harvesting complex photosystem II	chloroplast, thylakoid	1	4	1	1	0	0	0
b25	AT1G62060.1   unknown protein	apoplast	1	4	0	0	1	0	0
b26	AT3G01540.1   Symbols: DRH1, ATDRH1   DEAD box RNA helicase 1	cytoplasm, nucleolus, nucleus	4	2	0	0	1	0	0
b26	AT3G06480.1   DEAD box RNA helicase family protein	integral component of membrane, nucleus, vacuole	1	3	0	0	0	0	0
b27	AT2G43030.1   Ribosomal protein L3 family protein	chloroplast, membrane, ribosome	1	0	1	2	2	0	0
b28	AT5G02500.1   Symbols: HSC70-1, HSP70-1, AT-HSC70-1, HSC70   heat shock cognate protein 70-1	Golgi apparatus, apoplast, cell wall, chloroplast, cytoplasm, membrane, nucleolus, nucleus, plasmodesma, vacuolar membrane	2	1	0	0	2	0	0
b28	AT5G02490.1   Heat shock protein 70 (Hsp 70) family protein	Golgi apparatus, cell wall, cytosol, nucleus, plasma membrane	1	2	0	0	1	0	0
b29	AT5G49460.1   Symbols: ACLB-2   ATP citrate lyase subunit B 2	cytoplasm, cytosol, plasma membrane	1	4	0	1	0	0	0

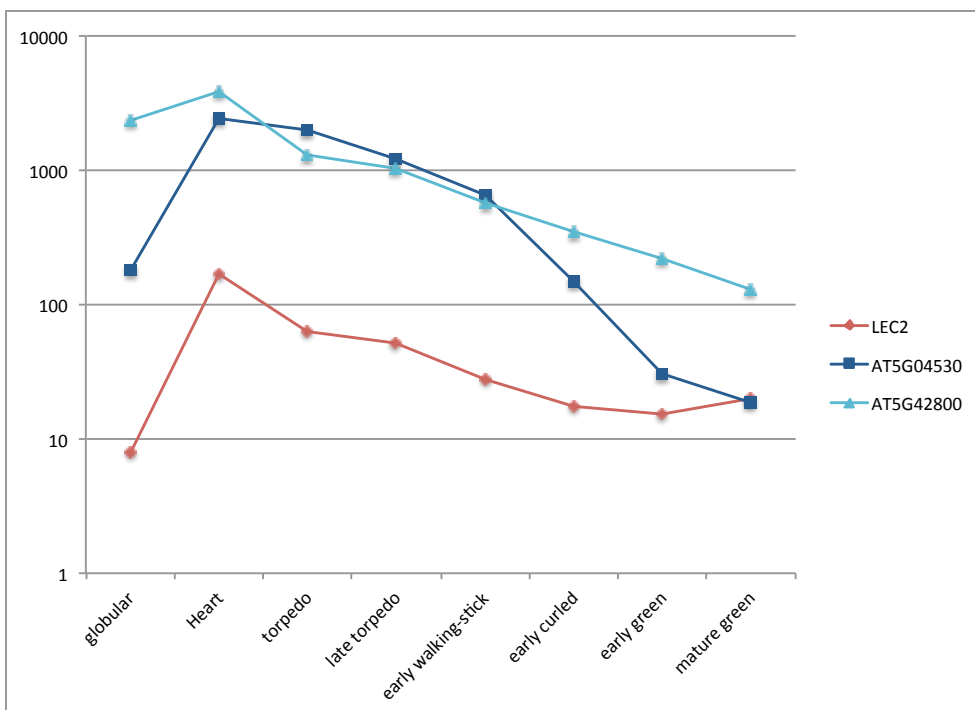
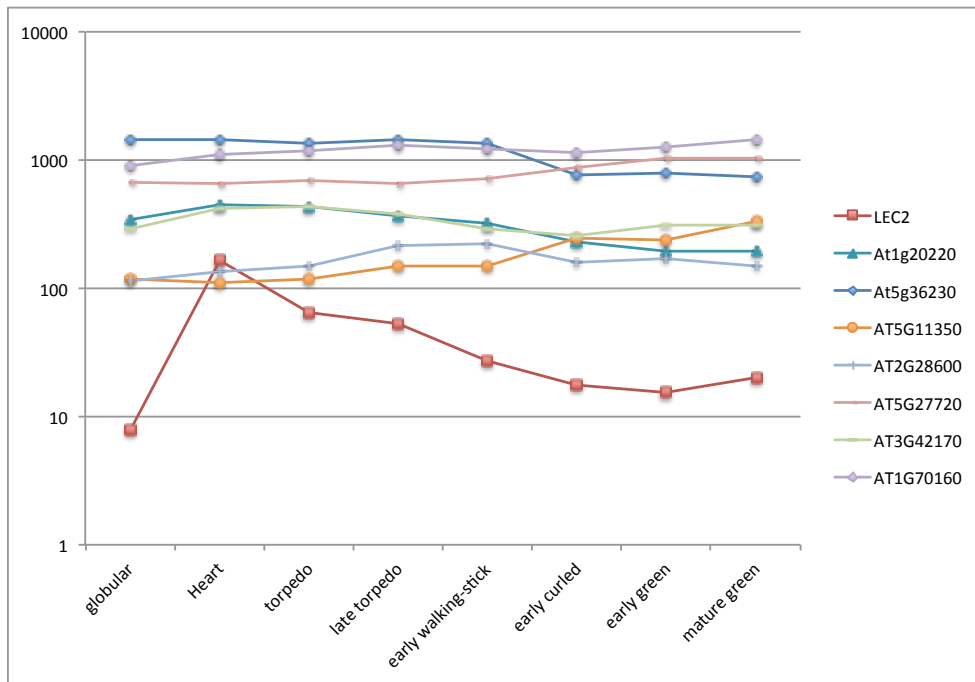
Group ID	Top Protein Description	Protein location	Number of proteins	GFP:GS 4-6	GFP:GS 17-4	GFP:GS 10-2	GFP:LEC2 17-1	GFP:LEC2 4-6	GFP:LEC2 18-4
b30	AT4G31700.1   Symbols: RPS6, RPS6A   ribosomal protein S6	chloroplast, cytosol, membrane, mitochondrion, nucleolus, nucleus, plasma membrane, plasmodesma	1	4	0	0	0	0	0
b31	AT4G31480.1   Coatomer, beta subunit	COPI vesicle coat, Golgi membrane, cytosol, nucleus, plasmodesma	3	3	1	0	0	0	0
b32	AT5G20290.1   Ribosomal protein S8e family protein	cell wall, chloroplast, cytoplasm, cytosol, membrane, nucleolus, plasma membrane, plasmodesma,	1	2	0	1	1	0	0
b33	AT3G44110.1   Symbols: ATJ3, ATJ   DNAJ homologue 3	cell wall, cytoplasm, cytosol, nucleolus, plasma membrane, plasmodesma	1	3	0	0	1	0	0
b34	AT3G05590.1   Symbols: RPL18   ribosomal protein L18	chloroplast, cytoplasm, plasma membrane, vacuole	2	2	0	0	2	0	0
b35	AT1G33120.1   Ribosomal protein L6 family	chloroplast, cytoplasm, membrane, nucleolus, plasmodesma, vacuole	2	3	0	0	1	0	0
b36	AT1G09760.1   Symbols: U2A'   U2 small nuclear ribonucleoprotein A	Cajal body, chloroplast, cytoplasm, cytosol, nucleolus, nucleoplasm, nucleus	1	0	1	2	1	0	0
b37	AT1G04270.1   Symbols: RPS15   cytosolic ribosomal protein S15	cytoplasm, cytosol, plasma membrane, plasmodesma, vacuolar membrane	3	3	0	0	1	0	0
b38	AT4G18100.1   Ribosomal protein L32e	cytoplasm, nucleolus, ribosome	1	0	0	1	2	0	0
b39	AT2G43970.1   RNA-binding protein	nucleus	2	1	0	0	2	0	0
b40	AT3G50370.1   unknown protein	cytoplasm	1	1	0	2	0	0	0
b41	AT5G23900.1   Ribosomal protein L13e family protein	cytoplasm, membrane	1	1	2	0	0	0	0
b42	AT2G36160.1   Ribosomal protein S11 family protein	chloroplast, cytoplasm, membrane, plasmodesma, vacuolar membrane	3	2	0	1	0	0	0
b43	AT3G49910.1   Translation protein SH3-like family protein	chloroplast, cytosol, membrane, nucleolus, nucleus, plasma membrane, vacuolar membrane	1	2	0	0	1	0	0
b44	AT1G53750.1   Symbols: RPT1A   regulatory particle triple-A 1A	cytosol, nucleus, plasma membrane	1	2	0	0	1	0	0
b45	AT3G49720.1   unknown protein	Golgi apparatus, chloroplast thylakoid membrane, endosome, membrane, vacuolar membrane	2	1	0	1	1	0	0
b46	AT1G74050.1   Ribosomal protein L6 family protein	cytoplasm, membrane, plasmodesma	2	3	0	0	0	0	0
b47	AT5G43960.1   Nuclear transport factor 2 (NTF2) family protein with RNA binding (RRM-RBD-RNP motifs) domain	intracellular	1	3	0	0	0	0	0
b48	ATCG00280.1   Symbols: PSBC   photosystem II reaction center protein C	chloroplast, membrane, plastoglobule, thylakoid	1	1	0	0	1	1	0
b49	AT1G29910.1   Symbols: CAB3, AB180, LHCB1.2   chlorophyll A/B binding protein 3	chloroplast, mitochondrion, plastoglobule, thylakoid	3	2	0	1	0	0	0

Group ID	Top Protein Description	Protein location	Number of proteins	GFP:GS 4-6	GFP:GS 17-4	GFP:GS 10-2	GFP:LEC2 17-1	GFP:LEC2 4-6	GFP:LEC2 18-4
b50	AT5G13020.1   Emsy N Terminus (ENT)/ plant Tudor-like domains-containing protein	nucleus	1	3	0	0	0	0	0
b51	AT1G20220.1   Alba DNA/RNA-binding protein	cytosol, nucleus, plant-type cell wall, plasmodesma	1	0	0	0	3	0	0
b52	AT2G20490.1   Symbols: NOP10, EDA27   nucleolar RNA-binding Nop10p family protein	Cajal body, nucleolus, nucleus	1	0	0	1	2	0	0
b53	AT4G17530.1   Symbols: RAB1C, ATRAB1C, ATRABD2C   RAB GTPase homolog 1C	Golgi apparatus, cytosol, endoplasmic reticulum membrane, membrane, plasmodesma, vacuole	1	3	0	0	0	0	0
b54	AT3G18035.1   Symbols: HON4   winged-helix DNA-binding transcription factor family protein	nucleosome, nucleus	1	2	0	0	0	1	0
b55	AT5G48650.1   Nuclear transport factor 2 (NTF2) family protein with RNA binding (RRM-RBD-RNP motifs) domain	cytoplasm, cytosol	1	2	0	0	1	0	0
b56	AT2G33620.1   AT hook motif DNA-binding family protein	cytosol, nucleus	4	3	0	0	0	0	0
b57	AT1G74470.1   Pyridine nucleotide-disulphide oxidoreductase family protein	chloroplast, membrane	1	2	0	0	1	0	0
b58	AT5G10220.1   Symbols: ANN6, ANNAT6   annexin 6	chloroplast	1	3	0	0	0	0	0
b59	AT3G53520.1   Symbols: UXS1, ATUXS1   UDP-glucuronic acid decarboxylase 1	Golgi apparatus, endosome, membrane	4	3	0	0	0	0	0
b60	AT1G15820.1   Symbols: LHCB6, CP24   light harvesting complex photosystem II subunit 6	chloroplast, membrane, plastoglobule, thylakoid	1	0	0	1	2	0	0
b61	AT1G02780.1   Symbols: emb2386   Ribosomal protein L19e family protein	cytoplasm, membrane, nucleolus, plasmodesma	1	3	0	0	0	0	0
b62	AT2G42740.1   Symbols: RPL16A   ribosomal protein large subunit 16A	chloroplast, cytoplasm, cytosol, cytosolic large ribosomal subunit, nucleolus, plasmodesma, vacuole	5	2	0	1	0	0	0
b63	AT2G17360.1   Ribosomal protein S4 (RPS4A) family protein	chloroplast, cytoplasm, plasma membrane, vacuolar membrane	4	3	0	0	0	0	0
b64	AT1G27650.2   Symbols: ATU2AF35A   U2 snRNP auxiliary factor small subunit, putative	nucleus	1	3	0	0	0	0	0
b65	AT4G16143.1   Symbols: IMPA-2   importin alpha isoform 2	cytoplasm, cytosol, nuclear pore, nucleolus, nucleoplasm, plasmodesma	2	1	0	0	2	0	0
b66	AT5G36230.1   ARM repeat superfamily protein	cytosol, membrane, nucleus	1	0	0	0	2	0	0
b67	AT5G04530.1   Symbols: KCS19   3-ketoacyl-CoA synthase 19	extracellular region, integral component of membrane	1	0	0	0	2	0	0
b68	AT5G08120.1   Symbols: MPB2C   movement protein binding protein 2C	microtubule cytoskeleton, nucleus	1	2	0	0	0	0	0
b69	AT5G08180.1   Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	nucleolus, nucleus	2	2	0	0	0	0	0

Group ID	Top Protein Description	Protein location	Number of proteins	GFP:GS 4-6	GFP:GS 17-4	GFP:GS 10-2	GFP:LEC2 17-1	GFP:LEC2 4-6	GFP:LEC2 18-4
b70	AT2G44440.1   Emsy N Terminus (ENT) domain-containing protein	nucleus	1	2	0	0	0	0	0
b71	AT1G72370.1   Symbols: P40, AP40, RP40, RPSAA   40s ribosomal protein SA	chloroplast, cytoplasm, membrane, nucleus, plasma membrane, plasmodesma	2	2	0	0	0	0	0
b72	AT1G12000.1   Phosphofructokinase family protein	cell wall, cytoplasm, membrane	1	0	1	0	1	0	0
b73	AT1G29030.1   Apoptosis inhibitory protein 5 (API5)	cytosol	1	2	0	0	0	0	0
b74	AT5G43310.1   COP1-interacting protein-related	membrane, nucleus	4	2	0	0	0	0	0
b75	ATCG00900.1   Symbols: RPS7.1, RPS7   Ribosomal protein S7p/S5e family protein	chloroplast	2	2	0	0	0	0	0
b76	GFP		1	2	0	0	0	0	0
b77	AT5G42800.1   Symbols: DFR, TT3, M318   dihydroflavonol 4-reductase	extrinsic component of endoplasmic reticulum membrane	1	0	0	0	2	0	0
b78	AT3G04840.1   Ribosomal protein S3Ae	chloroplast, cytoplasm, plasma membrane, plasmodesma	1	2	0	0	0	0	0
b79	AT5G11350.1   DNase I-like superfamily protein	cytoplasm, nucleus	1	0	0	0	2	0	0
b80	AT1G01610.1   Symbols: ATGPAT4, GPAT4   glycerol-3-phosphate acyltransferase 4	chloroplast, membrane	1	2	0	0	0	0	0
b81	AT3G20550.1   Symbols: DDL   SMAD/FHA domain-containing protein	chloroplast, cytosol, nucleus	1	2	0	0	0	0	0
b82	AT2G30490.1   Symbols: ATC4H, C4H, CYP73A5, REF3   cinnamate-4-hydroxylase	Golgi apparatus, endoplasmic reticulum, membrane, plasma membrane, plasmodesma, vacuolar membrane	1	2	0	0	0	0	0
b83	AT3G28500.1   60S acidic ribosomal protein family	cytoplasm, plasma membrane	1	1	0	0	0	0	1
b84	AT3G11400.1   Symbols: EIF3G1, ATEIF3G1   eukaryotic translation initiation factor 3G1	cytoplasm, plasma membrane	2	2	0	0	0	0	0
b85	AT1G48830.1   Ribosomal protein S7e family protein	cell wall, chloroplast, cytosol, plasma membrane	2	2	0	0	0	0	0
b86	AT1G10290.1   Symbols: ADL6, DRP2A   dynamin-like protein 6	Golgi apparatus, cytoplasm, membrane, microtubule, phragmoplast, plasma membrane, plasmodesma, vacuolar membrane	1	2	0	0	0	0	0
b87	AT3G09680.1   Ribosomal protein S12/S23 family protein	cytoplasm, nucleolus	2	2	0	0	0	0	0
b88	AT5G61780.1   Symbols: Tudor2, AtTudor2, TSN2   TUDOR-SN protein 2	cytoplasm, endoplasmic reticulum, nuclear envelope, perinuclear region of cytoplasm, plasmodesma	1	2	0	0	0	0	0
b89	AT2G34040.1   Apoptosis inhibitory protein 5 (API5)	cytosol, nucleus, plasmodesma	2	2	0	0	0	0	0

Group ID	Top Protein Description	Protein location	Number of proteins	GFP:GS 4-6	GFP:GS 17-4	GFP:GS 10-2	GFP:LEC2 17-1	GFP:LEC2 4-6	GFP:LEC2 18-4
b90	AT3G01090.1   Symbols: AKIN10, SNRK1.1, KIN10   SNF1 kinase homolog 10	cytoplasm, nucleus	3	2	0	0	0	0	0
b91	AT3G54110.1   Symbols: ATPUMP1, UCP, PUMP1, ATUCP1, UCP1   plant uncoupling mitochondrial protein 1	chloroplast, mitochondrion, plasmodesma, vacuolar membrane	1	2	0	0	0	0	0
b92	AT5G22650.1   Symbols: HD2B, HDT02, HDT2, ATHD2B, HDA4, HD2, ATHD2   histone deacetylase 2B	cell wall, cytosol, mitochondrion, nucleolus, nucleus, vacuolar membrane	2	2	0	0	0	0	0
b93	AT4G34555.1   Ribosomal protein S25 family protein	cytoplasm	3	2	0	0	0	0	0
b94	AT4G02840.1   Small nuclear ribonucleoprotein family protein	membrane, nuclear speck, nucleolus, nucleus	2	0	0	0	2	0	0
b95	AT4G16720.1   Ribosomal protein L23/L15e family protein	cytoplasm, plasma membrane, vacuolar membrane	2	2	0	0	0	0	0
b96	AT2G28600.1   P-loop containing nucleoside triphosphate hydrolases superfamily protein	nucleus	1	0	0	0	2	0	0
b97	AT2G46960.2   Symbols: CYP709B1   cytochrome P450, family 709, subfamily B, polypeptide 1	integral component of membrane	1	1	1	0	0	0	0
b98	AT4G10340.1   Symbols: LHCB5   light harvesting complex of photosystem II 5	chloroplast, membrane, thylakoid	1	1	0	0	1	0	0
b99	AT3G26650.1   Symbols: GAPA, GAPA-1   glyceraldehyde 3-phosphate dehydrogenase A subunit	apoplast, chloroplast, membrane	1	2	0	0	0	0	0
c100	AT1G10200.1   Symbols: WLIM1   GATA type zinc finger transcription factor family protein	cytoplasm, cytoskeleton, plasma membrane	1	1	0	0	0	0	0
c101	AT5G27720.1   Symbols: emb1644   Small nuclear ribonucleoprotein family protein	nucleus	1	0	0	0	1	0	0
c102	AT2G45140.1   Symbols: PVA12   plant VAP homolog 12	cytosol, endoplasmic reticulum, nucleus, plasma membrane	1	1	0	0	0	0	0
c103	AT3G60770.1   Ribosomal protein S13/S15	cell wall, chloroplast, cytosol, nucleolus	2	1	0	0	0	0	0
c104	AT3G42170.1   BED zinc finger ;hAT family dimerisation domain	cytosol, nucleus	1	0	0	0	1	0	0
c105	AT2G44200.1   CBF1-interacting co-repressor CIR, N-terminal;Pre-mRNA splicing factor	nucleus	1	1	0	0	0	0	0
c106	AT3G48410.1   alpha/beta-Hydrolases superfamily protein	endoplasmic reticulum, nucleus	1	1	0	0	0	0	0
c107	AT5G19760.1   Mitochondrial substrate carrier family protein	cell wall, chloroplast, mitochondrion, plasmodesma, vacuolar membrane	1	1	0	0	0	0	0
c108	AT4G39260.1   Symbols: CCR1, ATGRP8, GR-RBP8, GRP8   cold, circadian rhythm, and RNA binding 1	Golgi apparatus, cell wall, chloroplast, nucleolus, nucleus, peroxisome, plasma membrane, plasmodesma	4	1	0	0	0	0	0
c109	AT1G70160.1   zinc finger MYND domain protein	extracellular region	1	0	0	0	1	0	0
c110	AT3G58740.1   Symbols: CSY1   citrate synthase 1	peroxisome	1	1	0	0	0	0	0

**Annexe 5 :** Données d'expression des gènes identifiés d'après les résultats d'immunoprécipitation avec l'outil AtGenExpress visualisation (AVT) d'après Schmid *et al* 2005.









TCGGTCTCGAGCCCGGAGACAAGCTTATACTGTGGAACCTGGAGTTGGAATTGAACAACCTCACGGAAG  
AAAACACGAAGCTGAAGGAAATTGTGGAGGAAAATGAGAAGAAAAGAAGACAAGAGATAATAAGTAGA  
AGCAAACAAGTGAATAAGAGAAGAGCGGAGACAAATTGAGAAAAGATTTCGGAGGATGGCCAGTGCCGG  
GTGGTAA**GACCCAGCTTTCTTGTACAAAAGTGGTCCCC**

**B1 - *AtbZIP10* - B2 (1315pb)**

**GGGGACAAGTTTGTACAAAAAAGCAGGCTCG**ATGAACAGTATCTTCTCCATTGACGATTTCTCCGATC  
CTTTCTGGGAAACTCCTCCGATTCCCTCTCAATCCCGACTCTTCTAAGCCTGTTACGGCGGATGAAGTT  
AGCCAGAGTCAACCGGAATGGACTTTTCGAGATGTTTCTCGAAGAGATTTCTTCGTCGGCGGTGAGCTC  
TGAGCCACTTGGTAACAACAACAACGCGATCGTCGGTGTTCCTTCGGCGCAATCTCTTCTTCTGTTT  
CCGGACAGAATGATTTTCGAGGATGATAGTCGATTTTCGTGATCGCGATTTCGGGAAATTTGGATTGTGCT  
GCTCCCATGACGACGAAGACGGTGATTGTTGATTCCGATGATTATCGTCGTGTTCTTAAGAACAAGCT  
TGAGACTGAGTGCCTACTGTTGTTTCTTTCGGGTTGGGTCTGTGAAGCCTGAAGATTTCGACTAGTT  
CTCCAGAACTCAACTTCAACCAGTTCAATCCAGTCTCTTACTCAAGGTTCTTTGATGACCCAGGA  
GAACTTGGTGTACTTCTTCCCTTACCAGCTGAGGTGAAAAAACTGGTGTATCAATGAAGCAGGTTAC  
TAGTGGATCGTCGAGAGAATATTCTGATGACGAGGACCTTGATGAAGAGAATGAAACCACCGGTTCTCT  
TGAAGCCAGAGGACGTTAAAAAATCTAGAAGGATGCTGTCAAATCGTGAGTCAGCTAGGCGATCTAGA  
AGGAGAAAGCAGGAGCAAACAAGTGACCTCGAAACACAGGTTAATGATCTAAAAGGTGAGCATTCATC  
ACTTCTTAAACAACCTGAGCAACATGAATCACAAGTATGACGAGGCTGCTGTTGGCAATAGAATACTAA  
AGGCTGACATTGAGACATTAAGAGCTAAGGTGAAAATGGCGGAAGAAACCGTGAAGAGAGTAACAGGA  
ATGAATCCGATGCTTCTCGGAAGATCAAGTGGACATAACAACAACAACAGAATGCCAATAACTGGTAA  
CAACAGGATGGATTCTTCTAGCATTATTCCAGCTTATCAACCACACTCAAACCTAAACCATATGTCAA  
ACAAAACATCGGGATCCCAACCATTTCTACCTCCAAGACTCGGAAACAATTTTCGCTGCTCCTCCATCC  
CAAACAGCTCTCCCTTGCAGAGAATTAGAAATGGGCAAAATCACCATGTTACTCCAAGCGCAACCC  
GTATGGCTGGAATACCGAACCTCAGAACGATTGAGCATGGCCGAAAAAATGCGTGGACTGA**GACCCAG  
CTTTCTTGTACAAAAGTGGTCCCC**

**B1 - *AtbZIP25* - B2 (1273pb)**

**GGGGACAAGTTTGTACAAAAAAGCAGGCTCG**ATGCACATCGTGTTCCTCTGTGGATGATCTCACCGAGT  
CTTTCTGGCCTGTTCCCTGCTCCAGCTCCTTCTCCTGGATCTTCTTCTACACCTTCTCCAACCCAGAAC  
GTGGCAGATGGAATGACCAGATCTCAATCTGAGTGGGCTTTCCACAGGCTCATCAACGAGCTTTCTGG  
ATCTGATTCTTCTCCTACCACCAATACTATCGAGAGGTCACCTCCTCCAGTGCAGTCTTTGTCTAGAC  
TCGAAGAGACAGTGGATGAGACTGAGGATGTGGTGGAAATCCAGAAGCCTCAGAACCATAGAAGGCTC  
CCTGTTGATGATCAGGGAAAGAACAGAAACAGGGCTCCTTCTTCTGATCCTGTGGATTCTTCTGCTCC  
AGTGGTGGTTGATCCTAACCAGTACCATGCTATCCTCAAGTCTAAGCTCGAGTTGGCTTGTGCTGCTG  
TGGCTAGAAGAGTTGGAAGTGTGAAGCCAGAGGATTTCTTTCAGCTTCTGCTTCTAACCAGAAGCAGGCA  
CAGGGATCTATCGTGGCTCAAACCTCTCCTGGTGCATCTTCTGTGAGGTTCTCTCCTACTACCTCTAC  
CCAGAAAAAGCCTGATGTGCCAGCTAGGCAGACCTCTATCTCTTCTAGGGATGATTGAGACGATGATG  
ATCTCGATGGTGTGATGCTGATAACGGTGTGATCCTACCGATGTGAAGAGGGCTAGAAGGATGCTCTCTAAC  
AGAGAGTCTGCTAGAAGAAGTAGGAGAAGAAAGCAAGAGCAGATGAACGAGTTTCGATACCCAGGTTGG  
ACAGCTTAGGGCTGAGCATTCTACCCTTATCAACAGGCTCTCTGATATGAACCACAAGTACGATGCTG  
CTGCAGTGGATAACAGAATCCTCAGGGCTGATATCGAGACTCTCAGGACCAAGGTTAAGATGGCTGAG  
GAAACCGTGAAGAGGTTACCGGTGTTAACCTCTCCATTGGTCTAGGCCTAACATGGGAATCCCTTT  
CTCTAATACCCCTTCTGCTAGTTCTTCAATCCCTCCTAACTAACCACATCTTGAAGCCTGCTAACT  
CATCTACCAACACCTCTGCTGGACTCGCTCAAAACCAGAGATTGAGACTGCTAACTTCTTGCCTGAG  
CAGGTTAACAGGGAAGGTATGCAGAATCCTTTCGCTCCTGATTCTAACCTCTACGAGACTCTTCTCTCA  
CTGGAACCATAAGCACTAA**GACCCAGCTTTCTTGTACAAAAGTGGTCCCC**

B1 - *AtbZIP28* - B2 (2089pb)

GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGACGGAATCAACATCCGTGGTTGCTCCTCCGCCGG  
AGATACCTAATCTGAACCCTAGCATGTTTTCTGAGTCCGATTTGTTTTCTATTCCGCCGCTAGATCCT  
CTTTTCTATCTGATTCTGATCCGATTTCAATGGATGCGCCAATCTCCGATCTCGACTTCTTACTCGA  
CGATGAGAACGGAGATTTTCGCTGATTTTGATTTCTCGTTTTGATAATTCTGATGATTTCTTCGATTTCCG  
ATTTATCGGAGCCCGCGGTGGTGTATCCCTGAGGAGATCGGTAACAATCGTTCGAATTTGGACTCATCG  
GAAAACAGAAGCGGCGATGGAGGTTTAGAAGGAAGATCTGAGTCTGTTTCATTCACAGGTTTCATCTCA  
AGGCTCCAAGACTTTTGTGTCCGACACCGTTGACGCATCATCCTCCCCTGAATCAAGCAATCACCAGA  
AATCTTCTGTTAGCAAGAGGAAGAAGGAAAATGGAGACTCCAGTGGCGAATTAAGGAGCTGCAAGTAC  
CAAAAGTCCGATGATAAATCAGTCGCTACGAACAACGAAGGTGATGATGACGACGACAAGAGGAAGTT  
GATAAGGCAGATTAGGAACCGTGAAAGTGCTCAGCTTTTCGAGGTTGAGGAAGAAGCAACAAACTGAGG  
AGCTTGAAAGAAAAGTGAAGAGTATGAATGCTACCATTGCTGAATTGAATGGTAAGATTGCTTATGTT  
ATGGCTGAGAATGTCGCTTTAAGGCAACAAATGGCTGTTGCTTCTGGTGCTCCTCCTATGAATCCTTA  
TATGGCTGCCCGCCTTTACCGTATCAATGGATGCCGTATCCGCCGTATCCTGTTAGGGGATATGGAT  
CACAGACACCTTTGGTTCCCATTCCTAAGTTAAATCCTAAGCCTGTCTCGAGTTGTAGACCGAAGAAG  
GCAGAGAGTAAGAAGAATGAGGGTAAAAGTAAAGCTCAAGAAGGTTGCTAGTATTAGTTTTATTGGAAT  
TCTCTTCTTTGTCTTCTTGTGTTGGTACATTGGTTCCTTTTATGAATGTAAAATTTGGAGGAGAACGTG  
GAAGCTTTGGCGGTTTGTCTAAATATGATGGCCACCGGTATTACGATGAACATAAGGGGAGGGTTTTTA  
ATGGTCGGCGATGGTTCGATGTTAGAAGAAATAGTGGAAATTTCTGAAGGAAATATCCATTCAGTAG  
GATTAGTCATGGTGAGAGAGATAGTTGTGGAGGAGTAGATTATAATGCTCATCCGAAAGTAGAAGGAC  
GACCAAGTTCGTTGAGCAATGCCAGTGATCCTCTCTTTGCTTCTCTCTATGTCCCAAGAAACGATGGG  
CTTGTGAAGATCGATGGGAACCTTGATAATCACTCTGTTTTGGCGAGCGAGAAAGCAAGGGGTTTTAGG  
AAAGAAGAATATCACTGAAACAGTAAAACTAAAGAACCGGATTTGACCATTCTGGTGCATGTCTT  
CTGCATTAGCTGTTCCGGGGGTAAGAGGAAATGCAGCAATGCTTCCGCATTC AACAGCTCTCTCTTCT  
GAAGGGAAAAGACTTCACCAATGGTTTCATGAAGGTGGCTCAGGGCCACTAATGGATTACAGCATGTG  
CACCGAGGTTTTCCAGTTTGATATTGCTCCTGGTGCTATAGTCCCGTCATCAGTCTCCAGCATTTCTG  
CGGAGCATCTCCAAAATGTCACTACCCACGGCAAGAGAATGAAGAACAGGAGAATCCTTGAGGGACTT  
CCTGTTTCACTTGTGGCGTCAGAGCTCAATATCACCGGAACCCAGCCAAACAAAGACGCTCAAAATAA  
GACCTTTAATGGAAACACTAACAAACCCACATCATCCTCCATGGTTGTCTCAGTGTACTTGTGATC  
CAAGAGAGGTGCTTGACTCTGAAACCGACAGAGTGGTTCCCCCAAACCCAAAATCACTTTCCCGGATC  
TTTGTGGTGGTGTCTTGTGACAGTGTCAAGTACGTTACCTACTCATGCGTTCCTCCTCGATCGGGTCT  
CCATCTCGTAGCCACCTGAGACCCAGCTTTCTTGTACAAAGTGGTCCCC

## Annexe 7 : Tableaux des amorces

Construction de pGWB2 pLEC2_1kb	
HINDIII_pLEC2_1Kb_fwd	cccAAGCTTgggTGTCTGATTC
proLEC2_1Kb_XbaI_rev	gcTCTAGAgcTTTTCCCGGAGAG
Construction en Gateway multiple	
B1LEC2B5Rfwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGATAACTTCTTACCCTTTC
B1LEC2B5Rrev	GGGGACAAC TTTTGTATACAAAGTTGTCCACCACTCAAAGTCGTTAAAG
B1GFPB5Rfwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAGTAAAGGAGAAGA AACT
B1GFPB5Rrev	GGGGACAAC TTTTGTATACAAAGTTGTTTTGTATAGTTCATCCATGCCAT
B1GSB5Rfwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAATTCATGGGCACCCC
B1GSB5Rrev	GGGGACAAC TTTTGTATACAAAGTTGTTGAACCGCCTCCACCCGGCT
B5LEC2B2fwd	GGGGACAAC TTTTGTATACAAAAGTTGTAATGGATAACTTCTTACCCTTTC
B5LEC2B2rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACCACCACTCAAAGTCGT
B5GFPB2fwd	GGGGACAAC TTTTGTATACAAAAGTTGTAATGAGTAAAGGAGAAGA AACT
B5GFPB2rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTCTATTTGTATAGTTCATCCA
B5GSB2fwd	GGGGACAAC TTTTGTATACAAAAGTTGTAATGCAGAAGCTTATCTCCGAGG
B5GSB2rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTCTATTCAGTGACAGTGAAAG
Construction en Gateway simple	
B1HS-LEC2	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGCATCATCATCACCATCA
HS-LEC2B2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACCACCACTCGAAAATCGT
B1LEC2-SH	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGGATAACTTCTCCCTTTT
LEC2-SHB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGTGGTGATGGTGATGAT
Amorces pour la vérification par PCR et séquençage	
proLEC2-50 FWD	CGCTGGAGCAAAAAAAGAG
proLEC2 -250rev	GGGGACCACTTTGTACAAGAAAGCGGGACATATACATACACTA
LEC2 FWD	ATGGATAACTTCTTACCCTTTC
LEC2 REV	CCACCACTCAAAGTCGTTAAAG
GFP FWD	ATGAGTAAAGGAGAAGA AACT
GFP REV	CTATTTGTATAGTTCATCCA
GS_N FWD	ATGGAATTCATGGGCACCCC
GS_C REV	CTATTCAGTGACAGTGAAAG
HS_N FWD	ATGCATCATCATCATCATCATG
HS_C REV	TTAGTGGTGATGGTGATGATGTC
LEC2T-B3_F	ATGGATAACTTCTTACCCTTTC
bZIP67 FWD	ATGTCGGTTTTTCGAATCGG
bZIP67 REV	TTACCACCCGGCACTGG
bZIP53 FWD	ATGGGGTCGTTGCAAATGC
bZIP53 REV	GCTGATATGTTTGATTGCTGA
bZIP28 FWD	ATGACGGAATCAACATCCG
bZIP28 REV	TCAGGTGGCTACGAGATG
bZIP25 FWD	ATGCACATCGTGTCTCTG

bZIP25 REV	TTAGTGCTTATGGTTCCAG
bZIP10 FWD	ATGAACAGTATCTTCTCCATTG
bZIP10 REV	TCAGTCCACGCATTTTTTCG
<b>Amorces pour le géotypage</b>	
A : pro <i>LEC2</i> (-3587pb)	CGCACTTGATCAAAAGGCTTG
B : <i>LEC2</i> exon 645pb	TGAAAGACCAAGACTGCATA
C : pro <i>LEC2</i> (-500 pb)	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTAAAGGTACAAAGAT
D : pro <i>LEC2</i> (-2787 pb)	GTCTAAGTAATGCTCTCGTTGAG
E pro <i>LEC2</i> (-3246pb)	CCTCATCTAAAGATTTGAGC
F <i>LEC2</i> intron 945	GACGGAAGTCTAATCCTC
<b>Construction de <i>LEC2</i> muté</b>	
<i>LEC2</i> -TGATRN_FWD	GGTTTTCTTGATCCGCGTATGGGGAAGATGACAAACATCAACAGGAAAAACGCC
<i>LEC2</i> -TGATRN_rev	GGCGTTTTTCTGTTGATGTTTGTTCATCTTCCCATACGCGGATCAAGAAAAACCC
<b>Construction de <i>LEC2</i> tronqué pour les expériences de double hybride</b>	
B1 <i>LEC2</i> +B2-B3	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGGATAACTTCTTACCCTTTCC
B2 <i>LEC2</i> +B2-B3	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAAGCTTCTTGTTATCAAATG
B1 <i>LEC2</i> muté-B2-B3	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGGATAACTTCTTCCCTTTC
B2 <i>LEC2</i> muté-B2-B3	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGAGTTTCTTGTTATCGAAGG

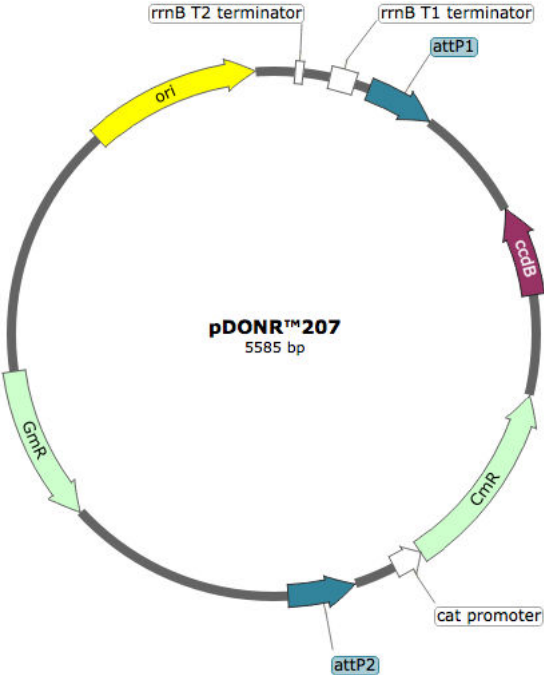
**Annexe 8 : Milieux pour les expériences de double hybride**

	Auto activation				Test double hybride		
	SD - leu	SD - leu - his	SD - tryp	SD - tryp - his	SD - tryp, - leu	SD - leu - tryp - his	SD - leu - tryp - his - ade
SD	8,01g	8,01g	8,01g	8,01g	8,01g	8,01g	8,01g
Aa (- tryp)			222mg				
Aa (- leu)	207mg						
Aa (- leu, - tryp)					192mg		
Aa (- leu - tryp, - his)		186mg		186mg		186mg	
Aa (- leu, - tryp, - his - ade)							180mg
ade	30mg	30mg	30mg	30mg	30mg	30mg	
tryp		30mg					
leu				30mg			
Agar (biomar)	6g	6g	6g	6g	6g	6g	6g

Milieu YPDA :

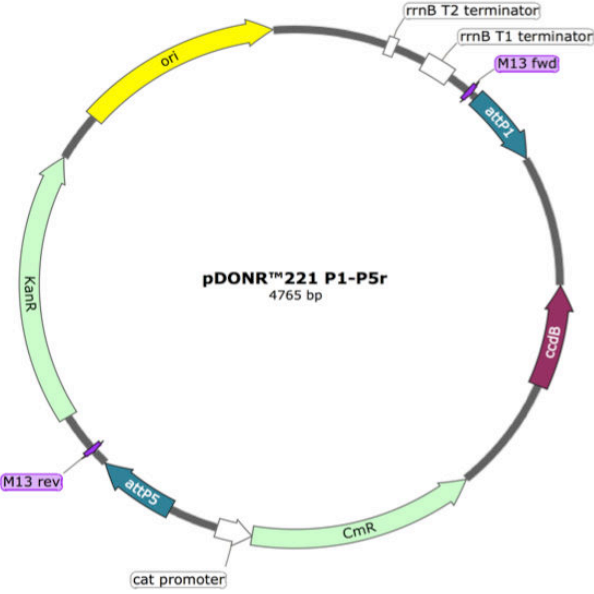
12,5g de milieu YPD  
30mg d'Adénine  
pour 250mL

**Annexe 9 : Vecteur Gateway simple**

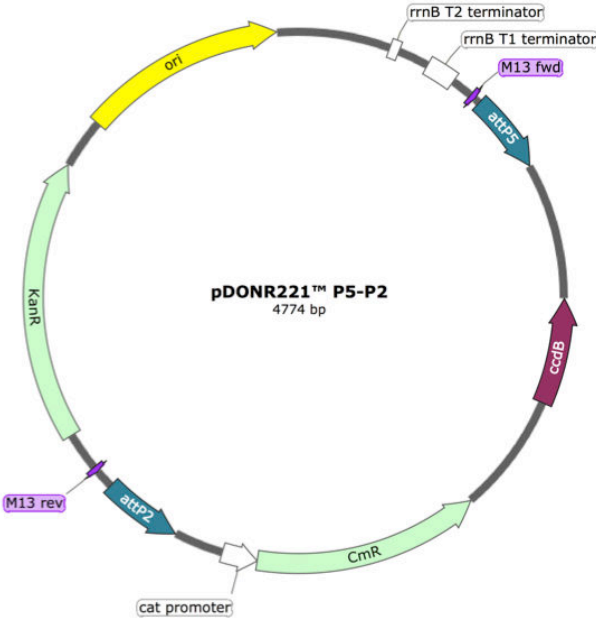


Carte du vecteur donneur pDONR207 (Invitrogen)

**Annexe 10 : Vecteurs Gateway double**



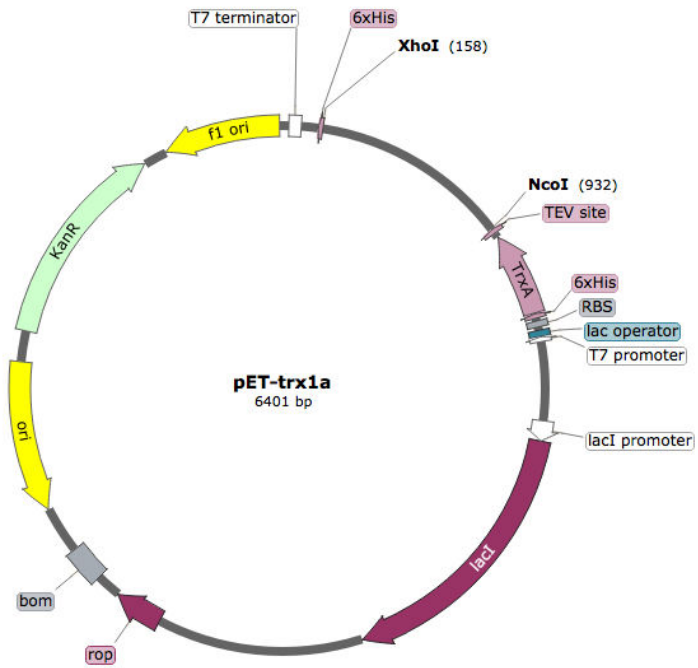
Carte du vecteur donneur pDONR221 P1-P5r (Invitrogen)



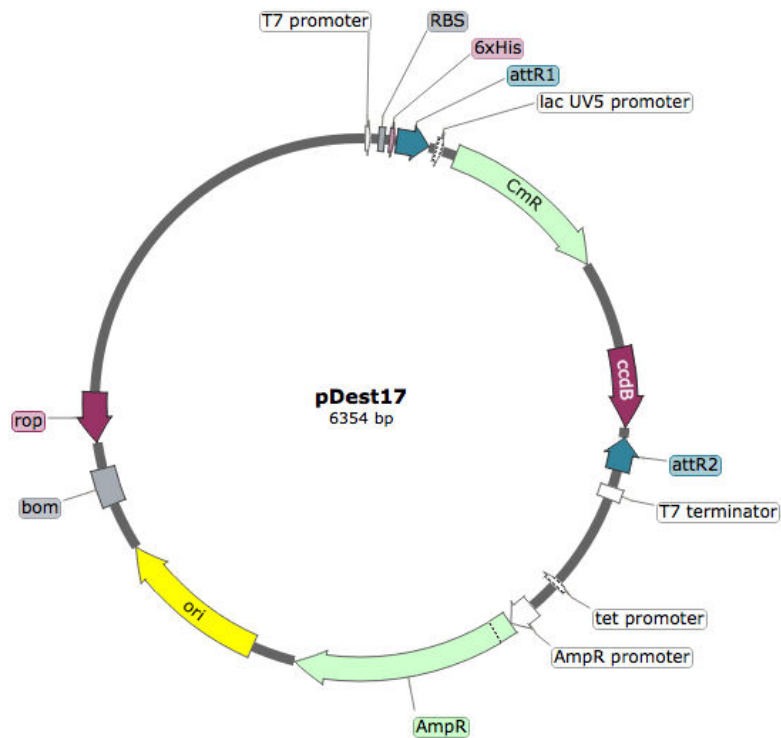
Carte du vecteur donneur pDONR221 P5-P2 (Invitrogen)



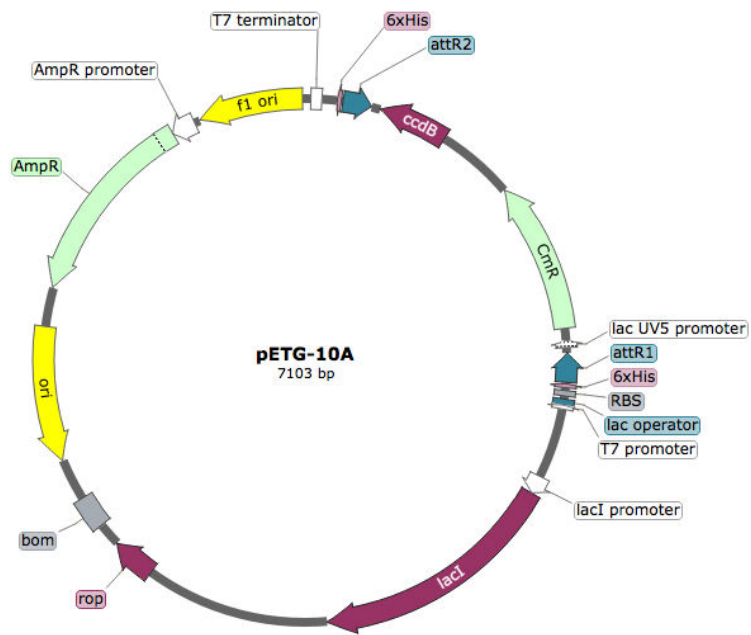
## Annexe 11 : Vecteurs d'expression des protéines



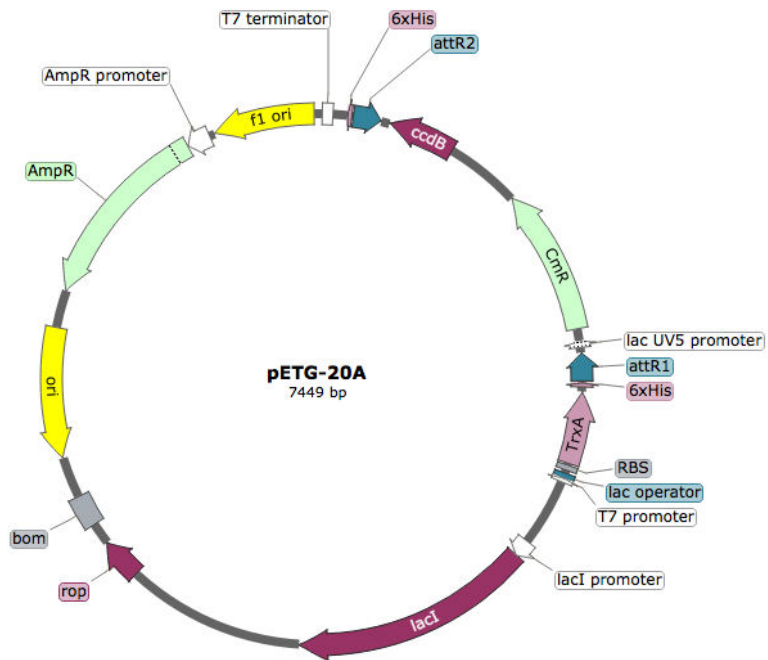
Carte du vecteur pETtrx-1a  
(Baud *et al* 2016)



Carte du vecteur pDEST17  
(Invitrogen)

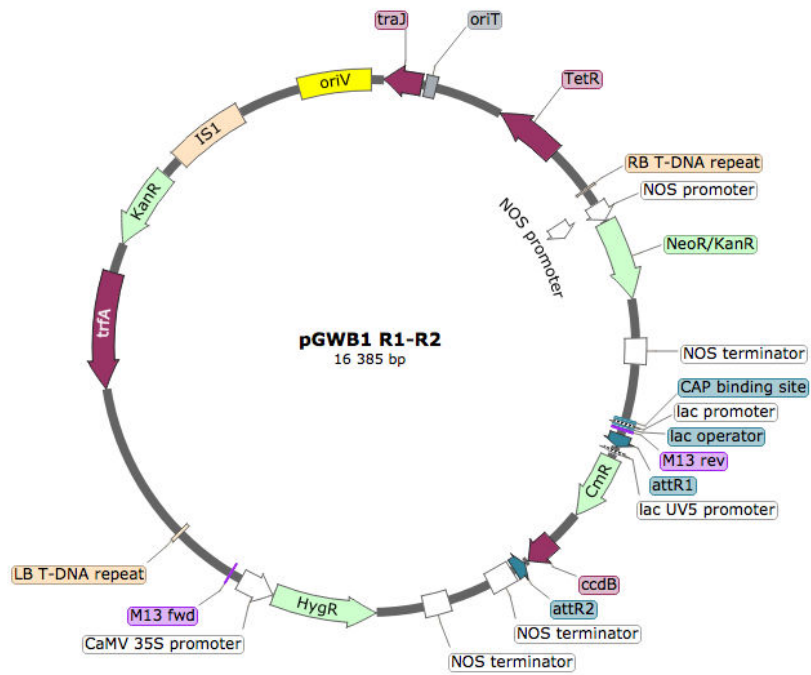


Carte du vecteur pETG-10A  
(EMBL, A. Geerlof)

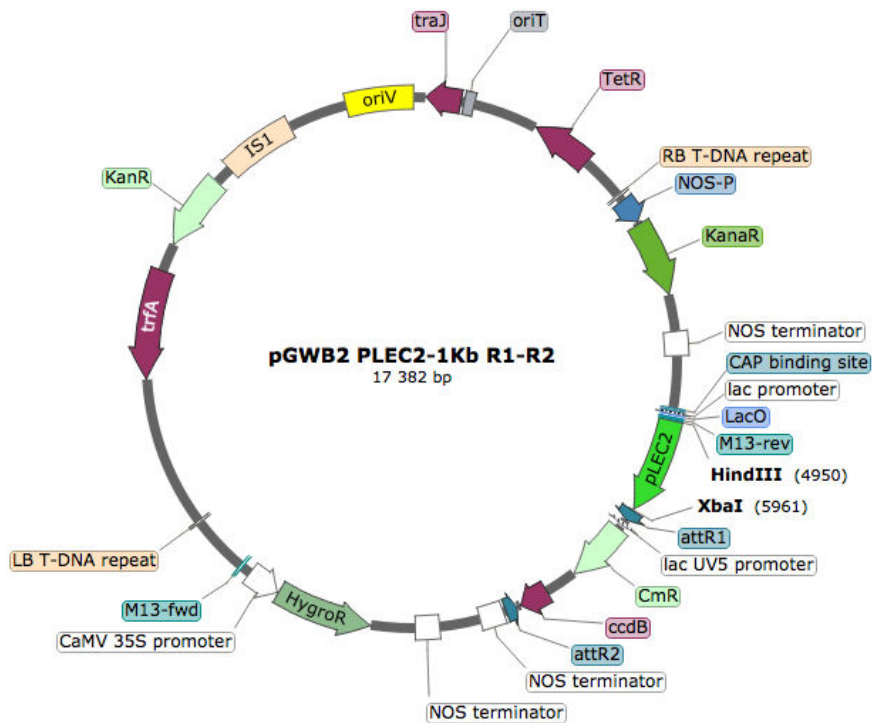


Carte du vecteur pETG-20A  
(EMBL, A. Geerlof)

## Annexe 12 : Vecteurs Binaire

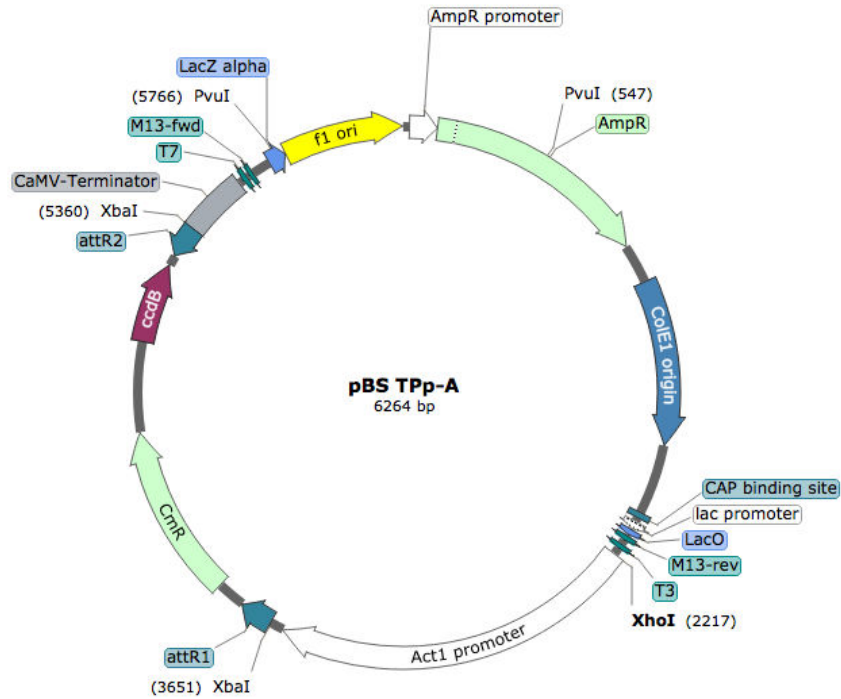


Carte du vecteur pGWB1  
(Nakagawa *et al* 2007)

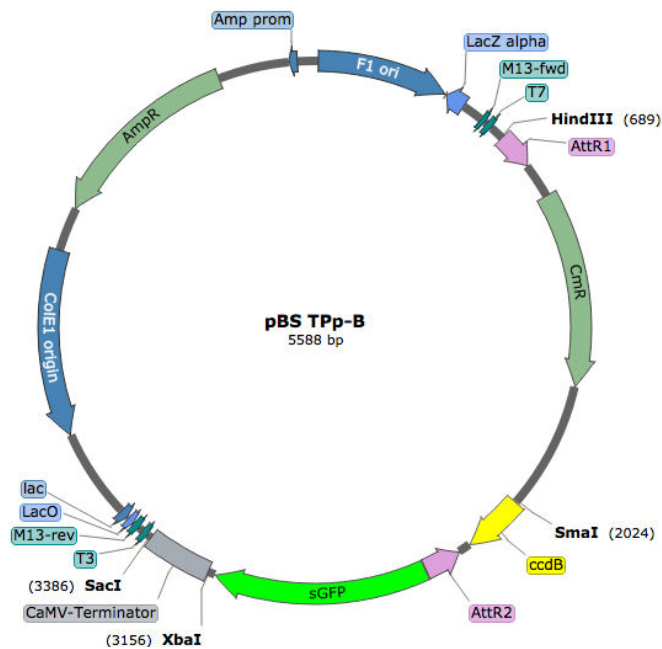


Carte du vecteur pGWB2 proLEC2-1Kb

**Annexe 13** : Vecteurs d'expression transitoire chez la mouche *Physcomitrella patens*

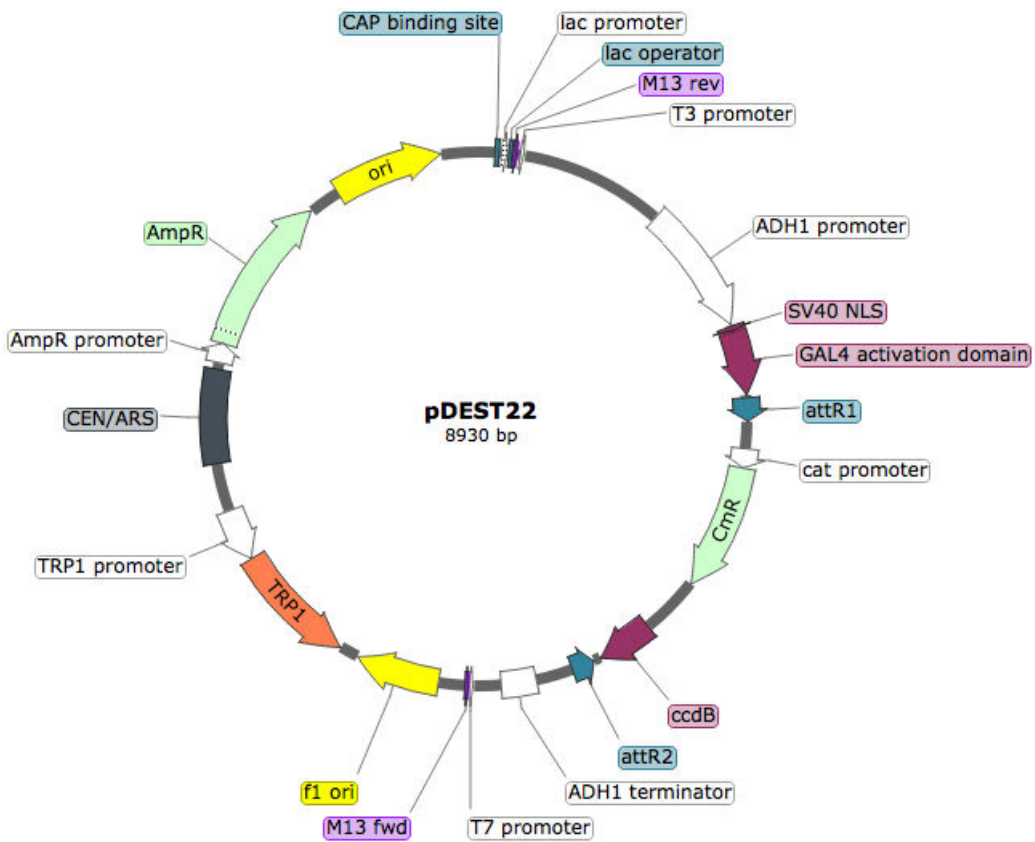


Carte du vecteur pBs Tpp-A ou vecteur A (Thévenin *et al* 2012)

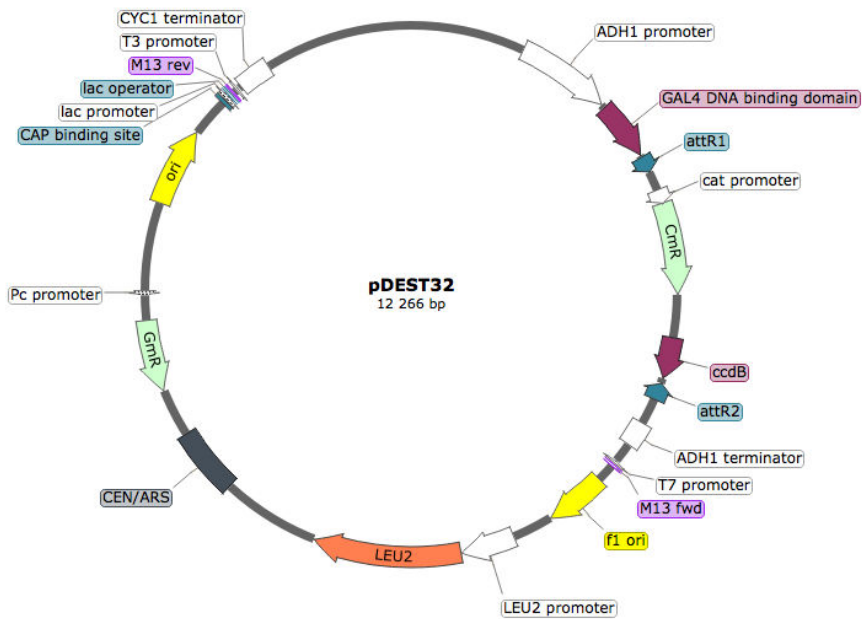


Carte du vecteur pBs Tpp-B ou vecteur B (Thévenin *et al* 2012)

**Annexe 14 : Vecteurs pour les expériences de double hybride**



Carte du vecteur pDEST22 (Invitrogen)



Carte du vecteur pDEST32 (Invitrogen)

**Titre :** Caractérisation biochimique du complexe AFL et identification des partenaires LEC2 lors du développement de la graine.

**Mots clés :** *Arabidopsis thaliana*, AFLs, domaine B3, NF-Y, interaction, LEC2, LEC1.

**Résumé :** Les spermatophytes majoritairement représentés par les angiospermes ou plantes à fleur, ont le trait fondamental de pouvoir se multiplier par la formation de graine. Les graines représentent un caractère essentiel pour la survie des plantes, leur dissémination et un usage important pour l'industrie, notamment agro-alimentaire. Lors de la formation de la graine, des transitions cruciales ont lieu, en particulier pour le développement, l'établissement des réserves, la maturation, l'acquisition de la tolérance à la dessiccation et la dormance. Ces étapes sont régulées par des facteurs de transcription et notamment les LAFL (LEC1, ABI3, FUS3, LEC2). Les LAFL sont composés de deux familles, les facteurs de transcription à domaine B3 (ABI3, FUS3, LEC2) et les NF-Y (LEC1), qui sont impliqués dans la régulation génique dans différentes étapes du développement de la graine. Des études suggèrent la possibilité de la formation d'un complexe composé de ABI3, LEC1 et LEC2

sur le promoteur de l'*OLEOSINE1* permettant d'activer plus fortement la transcription du gène. Néanmoins la caractérisation des protéines, leur action, régulation et leur capacité à interagir entre elles et former un complexe, restent encore méconnues. Dans un premier temps, je me suis attachée à caractériser la formation du complexe protéique composé d'ABI3, LEC1 et LEC2, plus particulièrement, l'interaction entre LEC1 et LEC2. La localisation subcellulaire des deux protéines par étiquetage *in planta*, ou grâce à des anticorps spécifiques, leur interaction dans un système double hybride et la mesure de l'effet de LEC1 sur LEC2 par thermophorèse ont permis de montrer que LEC2 et LEC1 peuvent former un dimère sur le promoteur de l'*OLEOSINE1* et que cette association change l'affinité de LEC2 pour son ADN cible. Dans une deuxième approche, j'ai cherché à identifier des partenaires avec (données bibliographiques) et sans (approche de TAP-tag) a priori pour compléter l'étude.

**Title :** Biochemical characterization of the AFL complex and identification of LEC2 partners during seed development.

**Keywords :** *Arabidopsis thaliana*, AFLs, domain B3, NF-Y, interaction, LEC2, LEC1

**Abstract :** The spermatophytes predominantly represented by angiosperms or flowering plants, have the fundamental ability to disseminate through seeds. Seeds are essential for the survival of plants, their dissemination and important resource for industry, especially food industry. During seed formation, critical transitions occur, in particular for development, establishment of reserves, maturation, acquisition of desiccation tolerance and dormancy. These steps are regulated by transcription factors and especially LAFL (LEC1, ABI3, FUS3, LEC2). LAFLS are composed of two families, B3 domain containing transcription factors (ABI3, FUS3, and LEC2) and NF-Y (LEC1) that are involved in gene regulation in different stages of seed development. Studies have suggested the possibility that a complex composed of ABI3, LEC1 and LEC2 is formed on the promoter of *OLEOSINE1*, allowing activating the transcription of the gene. Nevertheless, the

characterization of proteins, their action, regulation and their capacity to interact with each another and form a complex, are still unknown. In a first step, I focused on characterizing the formation of the protein complex composed of ABI3, LEC1 and LEC2, more particularly on the interaction between LEC1 and LEC2. The subcellular localization of the two proteins were followed by labelling *in planta*, or by means of specific antibodies, their interaction tested in a double hybrid system and the measurement of the effect of LEC1 on LEC2 validated by thermophoresis. All the results demonstrated that LEC2 and LEC1 can form a dimer on the promoter of *OLEOSINE1* and that this association change the affinity of LEC2 for its target DNA. In a second approach, I sought to identify partners with (bibliographic data) and without (TAP tagging approach) a priori to complete the study.