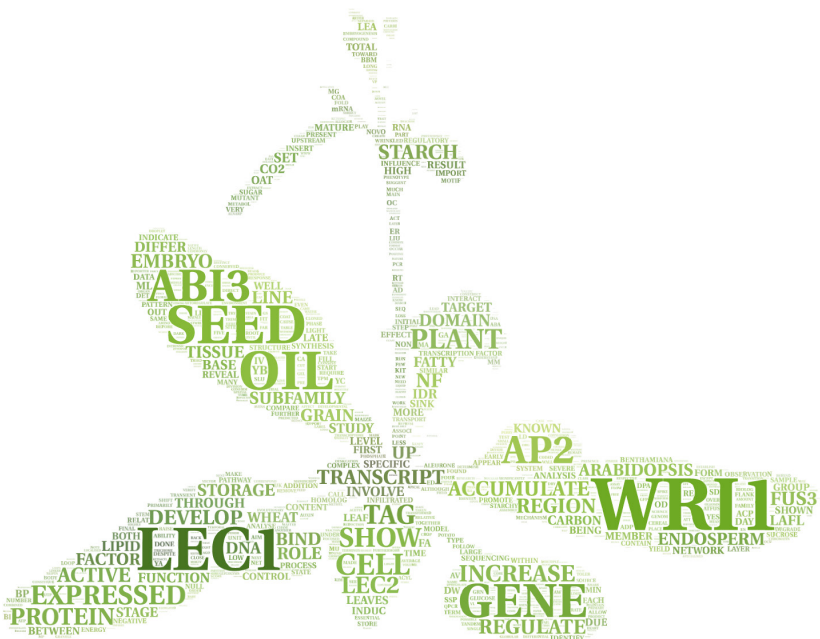
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2019:75
Faculty of Landscape Architecture, Horticulture and Crop Production Science

Dissecting the gene regulatory networks behind carbon allocation in plants

Ex situ studies on combinatorial and subdomains effects of seed transcription factors

Per Snell



Dissecting the gene regulatory networks behind carbon allocation in plants

Ex situ studies on combinatorial and subdomains effects of seed transcription factors

Per Snell

Faculty of Landscape Architecture, Horticulture and Crop Production Science
Department of Plant Breeding
Alnarp

Doctoral thesis
Swedish University of Agricultural Sciences
Alnarp 2019

Acta Universitatis agriculturae Sueciae
2019:75

Cover: Word cloud in the shape of *Arabidopsis thaliana* based on word usage data from this thesis.

ISSN 1652-6880
ISBN (print version) 978-91-7760-468-6
ISBN (electronic version) 978-91-7760-469-3

Print: SLU Service/Repro, Alnarp 2019

Dissecting the gene regulatory networks behind carbon allocation in plants. *Ex situ* studies on combinatorial and subdomains effects of seed transcription factors

Abstract

Plants store excess energy and metabolites derived from photosynthesis in a variety of storage compounds with starch, protein and triacyglycerol (oil, TAG) being the most common. As humans, we are completely dependent on these compounds either directly as food or indirectly, as materials for large number of goods or feed for animals. With an increased interest in sustainable production systems and the need to feed and clothe a growing population in a changing environment, the need for a better understanding of how plants allocate carbon for storage is increasingly important.

One key transcription factor involved in the accumulation of TAG into seeds is WRINKLED1 (WRI1). Using a transient gene expression system together with promoter-reporter gene constructs in *Nicotiana benthamiana* leaves, it was found that WRI1 is negatively regulating its own expression. This was further investigated by domain swapping between WRI1 homologs from diverse plant species to reveal that this mechanism was intrinsic to the tandem DNA binding AP2-domains of WRI1. Fluorescent electrophoretic mobility shift assay (fEMSA) was employed using purified WRI1 showing that WRI1 is unable to directly interact with its own upstream region.

During seed development, non-endospermic seeds accumulate large amounts of oil in the embryo during later stages of embryo development. This is known to be regulated by several master regulators commonly called the LAFL-network (LEAFY COTYLEDON1 (LEC1), ABSCISSIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3) and LEC2). Using transient expression in *N. benthamiana* leaves it was discovered that LEC1 significantly influences the transactivation ability of LEC2 and especially ABI3. Not only is LEC1 modulating the ability of ABI3 to induce *NbWRI1* expression, but also a large set of key genes involved in embryo morphogenesis. This suggests that LEC1 and ABI3 tightly collaborate to drive the embryo development from middle to late stage and the transition between morphogenesis to the acquisition of storage compounds.

With the purpose of investigating the differences in carbon partitioning between oat (*Avena sativa*) which is unique among the cereals in storing significant amount of oil in the endosperm with wheat (*Triticum aestivum*) that has basically no endosperm oil AsWRI1 was transformed into wheat. The resulting transformant lines showed a significantly altered seed phenotype with increased oil accumulation, disrupted endosperm development and reduced sink strength.

Keywords: Gene regulatory networks, AFL-subfamily, genetic engineering, gene regulation, biotechnology, autoregulation

Author's address: Per Snell, SLU, Department of Plant Breeding, P.O. Box 101, 230 53 Alnarp, Sweden

Utforskning av det regulatoriska gennätverket bakom kolallokering i växter. *Ex situ* studier kring kombinatoriska och subdomäneffekter av transkriptionsfaktorer från frö.

Abstract

Växter lagrar överskottsenergi och metaboliter från fotosyntesen i ett flertal olika former med stärkelse, protein och triacyglycerol (olja, TAG) som de vanligaste. Vi människor är helt beroende av denna inlagring, antingen direkt i form av mat eller foder, eller indirekt, som material för ett stort antal produkter. Med ett ökat intresse för hållbara produktionssystem och behovet av att föda och klä en växande befolkning i en allt mer föränderlig miljö, ökar stadigt behovet av att bättre förstå hur växter styr inlagringen av kol.

En viktig transkriptionsfaktor involverad i ansamlingen av TAG i frön är WRINKLED1 (WRI1). Genom att använde ett transient uttryck av transkriptionsfaktorer tillsammans med en promotordriven rapportgen i blad från *Nicotiana benthamiana*, konstaterades det att WRI1 reglerar negativt sitt eget uttryck. Detta undersöktes ytterligare genom att byta regioner mellan WRI1-homologer vilket avslöjade att denna mekanism var associerad med de DNA-bindande dubbla AP2-domänerna i WRI1. Fluorescent electrophoretic mobility shift assay (fEMSA) visade att WRI1 inte kan interagera direkt med sin egen uppströmsregion.

Under fröutveckling lagrar vissa frön stora mängder olja i embryot vilket sker sent i embryoutvecklinegn. Detta regleras av flera så kallade huvudregulatorer som tillsammans bildar LAFL-nätverket (LEAFY COTYLEDON1 (LEC1), ABSCISSIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3) och LEC2). Genom transient uttryck av dessa huvudregulatorer i blad från *N. benthamiana* upptäcktes att LEC1 signifikant påverkar transaktiveringsförmågan hos LEC2 och och i synnerhet hos ABI3. LEC1 modulerar inte bara ABI3s förmåga att inducera uttryck av *NbWRI1*, utan också en stor uppsättning nyckelgener involverade i embryomorfogenesen. Detta antyder att LEC1 och ABI3 tätt samarbetar för att driva stora delar av embryoutveckling och övergången mellan morfogenes till förvärv av lagringsmolekyler.

I syfte att undersöka skillnaderna i kollagring mellan spannmål som kan lagra olja i endospermet såsom havre (*Avena sativa*) och de som inte kunde göra det, såsom vete (*Triticum aestivum*) transformerades vete med *AsWRI1*. De resulterande linjerna uppvisade en signifikant förändrad fröfenotyp med ökad oljeackumulering, störd endospermutvecklingen och reducerad förmåga till kolinlagring.

Keywords: Regulatoriska gennätverk, AFL-subfamiljen, genmodifiering, genreglering, bioteknologi, självreglering

Author's address: Per Snell, SLU, Department of Plant Breeding, P.O. Box 101, 230 53 Alnarp, Sweden

“The sciences, each straining in its own direction, have hitherto harmed us little; but some day the piecing together of dissociated knowledge will open up such terrifying vistas of reality, and of our frightful position therein, that we shall either go mad from the revelation or flee from the light into the peace and safety of a new dark age.”

− Howard P. Lovecraft, The Call of Cthulhu, 1928

“How fleeting are all human passions compared with the massive continuity of ducks.”

− Dorothy L. Sayers, Gaudy Night, 1935

Contents

List of publications **11**

List of tables **13**

List of figures **15**

Abbreviations **19**

1 Introduction **21**

2 Background **23**
2.1 Regulation of gene expression 23
2.1.1 Levels of gene regulation 23
2.1.2 Transcription factors 25
2.1.3 Hit and run model of transcription 27
2.1.4 Gene regulatory networks 27
2.1.5 LEAFY COTYLEDON1 – A subunit in the Nuclear Factor Y complex 29
2.1.6 The AFL-subfamily 31
2.1.7 The AP2-subfamily 33
2.2 Plant storage compounds 36
2.2.1 Photosynthate partitioning 37
2.2.2 Starch 39
2.2.3 Oil 43
2.2.4 Protein 48
2.3 Seed development 49
2.3.1 Embryogenesis 50
2.3.2 Endosperm development 53
2.4 The LAFL-network 54
2.4.1 The internal regulatory system governing the LAFL-network 55
2.4.2 Temporal expression pattern of members of the LAFL-network during seed development 56
2.4.3 Spatial expression pattern of LAFL-network 57
2.4.4 Beyond the LAFL-network 58

3 Aims of this study 61
3.1 General aims 61
3.2 Specific aims 61

4 Results and discussion 63
4.1 WRINKLED1 in wheat (paper II) 63
4.1.1 WRINKLED1 leads to accumulation of oil when expressed in wheat endosperm 63
4.2 Physiological effects from overexpressing WRINKLED1 in wheat endosperm 64
4.2.1 WRINKLED1 influences wheat endosperm sink strength ambivalently 66
4.2.2 Field trials reveal further decrease in seed weight 68
4.3 Autoregulation of WRI1 expression (paper I) 70
4.3.1 WRI1 is suppressing its promoter activity through a negative feedback loop 70
4.3.2 Domain swapping reveals that the tandem AP2-domain is essential for WRI1 autoregulation 71
4.4 Understanding the structure of the AP2-subfamily of transcription factors (manuscript V) 72
4.4.1 The AP2-subfamily can be split into three groups 72
4.4.2 The regulatory specificity of the AP2-subfamily is intrinsic to the DNA-binding domain 73
4.5 LEC1 specifies the role of the AFL-subfamily during embryogenesis (manuscript III and IV) 74
4.5.1 Transcriptomic analysis reveals the unique roles of the AFL-subfamily 74
4.5.2 LEC1 and ABI3 activate key genes for embryo morphogenesis 75
4.5.3 LEC1 activates the expression of WRI1 through ABI3 77
4.5.4 ABI3 is able to increase oil accumulation through a non-WRI1 dependent pathway 78
4.5.5 The A-domain of ABI3 is essential for collaboration with LEC1 80

5 Conclusions 81

6 Future work 84

References 87

Popular science summary 99

Populärvetenskaplig sammanfattning **101**

Acknowledgements **103**

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Snell*, P., Grimberg, Å., Carlsson, A. S., Hofvander, P. (2019). WRINKLED1 Is Subject to Evolutionary Conserved Negative Autoregulation. *Frontiers in Plant Science*. 10, pp. 387.

II Grimberg*, Å., Wilkinson, M., Snell, P., De Vos, R. P., González-Thuillier, I., Tawfike, A., Ward J. L., Carlsson, A. S., Shewry, P., Hofvander, P. (2019). Transitions in wheat endosperm metabolism upon transcriptional induction of oil synthesis by oat endosperm WRINKLED1. (Submitted).

III Snell, P., Grimberg, Å., Hofvander, P. (2019). LEC1 regulate accumulation of seed storage compounds through interaction with ABI3. (Manuscript)

IV Snell, P., Grimberg, Å., Hofvander, P. (2019). Interaction between NF-Y and B3-family TFs differentially regulates key steps of embryo development. (Manuscript)

V Snell, P., Berg, H., Grimberg, Å., Carlsson, A, Hofvander, P. (2019). Dissection of members of the AP2-domain transcription factors reveal mechanisms behind target specificity. (Manuscript)

Papers I are reproduced with the permission of the publishers.

* Corresponding author.

The contribution of Per Snell to the papers included in this thesis was as follows:

I Participated in planning of the experiments, performed all experiments and participated in the analyses of the results. Wrote first draft of the manuscript.

II Participated in planning of the experiments, performed TAG analysis, RNA-seq analysis and actively participated in labelling experiments, histological preparations, dissection of seed tissues and growing of plant material. Wrote RNA-seq result section in the first draft of the manuscript.

III Participated in hypothesis development and planning of the experiments. Performed all experiments and participated in the analyses of the results. Wrote first draft of the manuscript.

IV Participated in hypothesis development and planned the experiments. Performed all experiments and participated in the analyses of the results. Wrote first draft of the manuscript.

V Participated in planning of the experiments, performed all experiments and participated in the analyses of the results. Wrote first draft of the manuscript.

List of tables

Table 1. Table over yield-related measurements on control and 12-insert transformed wheat lines expressing AsWRI1. Plants grown in climate chambers. Statistical analysis was conducted using Student's T-test. SD = Standard deviation. 67

Table 2. Table over yield-related measurements from field trials on control and 12-insert transformed wheat lines expressing AsWRI1. Statistical analysis was conducted using Student's T-test. SD = Standard deviation. 69

List of figures

Figure 1. The *hit and run* model of transcription. In the first step a pioneer transcription factor (Pi) binds its target sequence in the condensed and inaccessible chromatin. This leads to a localized opening allowing a first transcription factor (TF1) to bind together with a histone modifying factor (HMF) and further loosen the chromatin. TF1 is able to recruit TF2 which are able to initiate transcription through Polymerase II (Pol II). Once the chromatin is properly opened the TF1 and HMF dissociates to target the next gene. The chromatin is now fully accessible for TF2 allowing it to bind and initiate transcription without the help of TF1. Central region shows the temporal process where chromatin is the first to be influenced and remains so during the whole process while the transcriptome only is influenced at the later stages. Modified from Swift and Coruzzi (2017). 26

Figure 2. The basic components of a gene regulatory network. 28

Figure 3 A) General visualisation of the domains of Arabidopsis NF-YA, NF-YB, NF-YC. Modified from (Zhao *et al.*, 2016). B) Model of co-transportation of NF-YB and NF-YC where the mRNA is transported out of the nucleus (1) and translated in the cytosol (2). NF-YA, having a nuclear localisation signal, is being transported directly to the nucleus while NF-YB and NF-YC remain in the cytosol. NF-YB binds to the NF-YC (3), and the complex is then transported to the nucleus through the NF-YC nuclear localisation signal (4). In the nucleus, the full NF-Y complex can be assembled (5). 29

Figure 4. Structural representation of AtABI3, AtLEC2 and AtFUS3 that together make out the AFL-subfamily of transcription factors which are important for seed development. 32

Figure 5. Schematic drawing of Arabidopsis WRI1 containing a tandem AP2-domain and with intrinsically disordered regions (IDRs) towards the N-terminus (IDR1) and C-terminus (IDR2 and 3). 35

Figure 6. Simplified model over carbon allocation in developing seed tissue. With the help of cellular metabolism the carbon, in the form of sugars, can be shuffled into seed storage proteins, oil or starch. Figure modified from Baud *et al.* (2008). 38

Figure 7. Starch synthesis pathways for storage starch (A) and transient starch (B). Enzymes involved in the pathways are as follows: 1, Sucrose synthase; 2, UTP-glucose-1-phosphate uridylyltransferase; 3, Phosphoglycerate mutase (cytosolic); 4, Fructokinase; 5, Phosphoglucoisomerase; 6, Glucose-6-phosphate/Pi translocator; 7, ADP-glucose pyrophosphorylase (cytosolic); 8, ADP-glucose/ADP translocator; 9, Phosphoglycerate mutase; 10, ADP-glucose pyrophosphorylase; 11, Starch synthases (I-IV), Starch branching enzymes I-III), Isoamylases (I-III); 12, Granule-bound starch synthase. Metabolites are abbreviated as follows: UDP, Uridine diphosphate; Gluc, Glucose; F6P, Fructose-6-phosphate; G1P, Glucose-1-phosphate; ATP, Adenosine triphosphate; PPi, Pyrophosphate; UTP, Uridine triphosphate; ADP, Adenosine diphosphate. Transport pathways are noted by dashed arrows. Blue arrows indicate alternative pathway present in cereals. Figure modified from (Geigenberger, 2011). 42

Figure 8. Schematic overview of the *de novo* fatty acid synthesis taking place in the plastid and triacylglycerol assembly in the endosplasmic reticulum in plants. Enzymes involved in the pathways are as follows: 1, Pyruvate Dehydrogenase Complex; 2, acetyl-CoA carboxylase; 3, Malonyl-CoA:ACP Malonyltransferase; 4, Ketoacyl-ACP Synthase III; 5, Ketoacyl-ACP Reductase; 6, Hydroxyacyl-ACP Dehydrase; 7, Enoyl-ACP Reductase; 8, Ketoacyl-ACP Synthase I; 9, Ketoacyl-ACP Synthase II; 10, Stearoyl-ACP Desaturase; 11, Acyl-ACP Thioesterase A/B; 12, Glycerol-3-Phosphate Acyltransferase; 13, 1-Acylglycerol-3-Phosphate Acyltransferase; 14, Phosphatidate Phosphatase; 15, Acyl-CoA : Diacylglycerol Acyltransferase; 16, Phospholipid:Diacylglycerol Acyltransferase. Abbreviations: ER, Endoplasmic Reticulum; LD, Lipid Droplet. Transport pathways are noted by dashed arrows. Pathway and names after Li-Beisson *et al.* (2013). 46

Figure 9. Summary of the regulatory interplay observed in studies of the LAFL-network of transcription factors. 55

Figure 10. Expression of genes encoding the members of the LAFL-network of transcription factors during Arabidopsis seed development. X-axis denotes seed developmental stage spanning from mid-globular (3) to green-cotyledons embryo (10). Data from Schmid *et al.* (2005). 56
Figure 11. Interpreted expression pattern of A) LEC2, B) FUS3, C) ABI3 and D) LEC1 based on the promoter::GUS constructs reported by To *et al.* (2006b) and (Huang *et al.*, 2015). 57
Figure 12. Images from light microscopy showing wheat grains from control (A), the wrinkled phenotype of the AsWRI1-line with 12-inserts (B), the ear of the control (C) versus the 12-insert line (D) at 26 DPA. E-F shows a longitudinal section of the grain of control (E) and 12-insert line (F) at 26 DPA. Dark area in the middle of the grain in E is the crease, the dark area in F is a cavity filled with an aqueous liquid before seed dessication occurs. Scale bars are 10 mm (A, B), 20 mm (C, D) and 1 mm (E, F). Figure taken from manuscript II. 65
Figure 13. Light microscopy imaging of aleurone layer (Al), starchy endosperm (SE) and subaleurone layer (SAl) in control (A-B) and 12-insert line (C-D). Grain sections were stained with MAS staining proteins green and starch in dark purple. Scale bars 100 µm (A and C), 20 µm (B and D). Figure modified from manuscript II. 66
Figure 14. Unrooted phylogenetic tree based on sequence homology in the DNA-binding motif showing the three proposed groups of the 14 members of the AP2-subfamily transcription factors (left). Protein models are shown to scale to the right, where red boxes represent position of the tandem AP2-domains. 72
Figure 15. Chord diagram over significantly enriched (adjusted P-value <0.05) GO-terms for genes significantly affected by the co-expression of LEC1 with the individual members in the AFL-network of transcription factors. Size of chord is negatively correlated with P-value. Figure originally from manuscript IV. 76
Figure 16. Triacylglycerol (TAG) total fatty acid (FA) content from infiltrated leaves of N. benthamiana using Arabidopsis transcription factors. Bars represent the average from three biological replicates. Error bars = SD. Letters indicate significantly distinct groups based on Tukey´s test with a cutoff at $P \leq 0.05$. Dashed box represents the theoretical additive effect of LEC1+ABI3 on oil accumulation to the AtWRI1 infiltration. 79
Figure 17. Proposed model on the function of LAFL-network during seed development in Arabidopsis. Developmental stages of seed and embryo development in Arabidopsis with days after pollination

(DAP) marked (A). Gibberellins (GA) and abscisic acid (ABA) levels (B) and schematic gene expression profiles of the LAFL-network of transcription factors (C) in Arabidopsis seeds during development. Seed structures are not drawn to scale. ABA and GA profiles based on Okamoto et al. (2006) and Hu et al. (2018) respectively. Relative expression profiles of LAFL-network members is based on transcriptomic analysis made by Schmid et al. (2005) and Hofmann et al. (2019). D) Proposed model of how the LAFL-network regulate different stages of embryo development in collaboration with ABA and GA. During early embryogenesis with high levels of GA and low levels of ABA, LEC1 functions through the NF-Y complex as well as in a complex with LEC2 to regulate cell division, suspensor identity and suppressing late embryogenesis related genes targeted by LEC2. As ABI3 increases with increased ABA levels LEC1 cooperates with ABI3 to drive the transition from globular to heart stage. Towards late stages of embryo morphogenesis, LEC1 expression goes down allowing the AFL-network to fully function in the acquisition of desiccation tolerance and storage compound accumulation. 82

Abbreviations

ABA	Abscisic Acid
ABRE	ABA-Responsive Element
ACP	Acyl Carrier Protein
Al	Aleurone
ATP	Adenosine Triphosphate
bp	Base Pair
CoA	Coenzyme A
COAR	Co-Activator/Co-Repressor
CRE	Cis-Regulatory Element
DAG	Diacylglycerol
DNA	Deoxyribonucleic Acid
DPA	Days Post Anthesis
ER	Endoplasmic Reticulum
F6P	Fructose-6-Phosphate
FA	Fatty Acid
fEMSA	fluorescent Electromobility Shift Assay
FFA	Free Fatty Acid
G1P	Glucose-1-Phosphate
G3P	Glycerate 3-Phosphate
G6P	Glucose-6-Phosphate
GA	Gibberellic Acid
GFP	Green Fluorescent Protein
GO	Gene Ontology
GRN	Gene Regulatory Network
GUS	β-Galactosidase
H3K4	Histone 3, Lysine 4
H3K4me3	Histone 3, Lysine 27 Trimethylation
H3K9	Histone 3, Lysine 9

HFD	Histone Fold Domain
HMF	Histone-Modifying Factor
IDR	Intrinsically Disordered Region
LD	Lipid Droplet
LEA	Late Embryogenesis Abundant
LPA	2-Lysophosphatidic Acid
MAS	Masson's Trichrome Stain
mRNA	Messenger RNA
mRNA-seq	mRNA-Sequencing
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
PA	Phosphatidic Acid
PB/PBs	Protein Body/Protein Bodies
PC	Phosphatidylcholine
PPP	Pentose Phosphate Pathway
PSV	Protein Storage Vacuoles
RNA	Ribonucleic Acid
SAM	Shoot Apical Meristem
SCN	Stem Cell Niche
SD	Standard Deviation
SE	Starchy Endosperm
SAl	Subaleuronelayer
SSP	Seed Storage Protein
TAG	Triacylglycerol
TCA	Tricarboxylic Acid Cycle
TF	Transcription Factor
UDP	Uridine Diphosphate
UTP	Uridine Triphosphate
VSP	Vegetative Storage Proteins

1 Introduction

The invention of farming and the development of complex agricultural systems are the fundaments of human history. As history rose from pre-history, agriculture did not only coincide with the advent of writing and complex societal structures; it was the underlying reason for it. Still today, we can see that pivotal historical developments follow the development of agriculture. Moreover, until recently, agriculture was synonymous with the growing of plants, with animal husbandry serving the primary purpose of producing favourable conditions for crop production. Even today we are entirely reliant on crop production to feed the growing population of humans, either directly as food or through animal feed. In 2011, cereal grains supplied the world with 45 % of the consumed calories with wheat and rice making up over 80 % of that, which means that almost 37 % of the worlds consumed calories comes from two crops. In addition to this, almost half of the worlds consumed fats are from plant oils with oil palm, rapeseed, soybean and sunflower making up over 80 % of the produced oil crops (FAOSTAT, 2011). The common theme among all these crops is that we utilise their seeds or fruits. This makes the understanding of how seeds develop and accumulate storage compounds essential for both plant breeding and biotechnological applications. Especially when faced with a growing population and an increased interest in novel plant properties such as increased protein content.

In addition to this, plants are essentially sun-powered factories which can be utilised to produce specialised metabolites such as rare fatty acids which can be used to re-establish a plant-based chemical industry. Based on years of experience, we have come to understand that the establishment of such novel production systems does not come easy. It requires both an in-depth knowledge of the system being engineered as well as the processes surrounding it. The study of gene regulatory networks, as those focused on in this work, therefore becomes invaluable for future plant biotechnological applications.

2 Background

2.1 Regulation of gene expression

In the early days of genetics, the complexity of a species was thought to be a direct result of the number of genes coded for in their genome. More genes and larger genome equalled higher complexity. Today we know that the complexity of organisms mainly is a result of regulation of gene expression rather than the addition of large sets of novel genes. Proteins regulating the expression of individual genes are called transcription factors and switch on/off whole sets of genes for example conferring a developmental change. Transcription factors are in turn under strict control from other transcription factors. These complex networks are commonly known as gene regulatory networks which are one of the main reasons for eukaryotes being able to evolve complex, multicellular structures. However, gene regulatory networks are not only crucial for plant development. Plants, being sessile organisms, are extremely dependent on fast and precise regulation of gene expression to combat long and short term changes in their environment. As a result of this, gene regulatory networks are complex and prone to changes between species that have adapted to survival in different environments.

2.1.1 Levels of gene regulation

Gene regulation is an umbrella term that covers many aspects that influence processes that affects the production of messenger-RNA (mRNA), translation and function of the resulting protein. In brief, gene regulation can be divided into five major levels; (i) DNA-modifications, (ii) transcriptional regulation, (iii) post-transcriptional regulation, (iv) translational regulation and, (v) post-translational regulation.

At the first level, DNA-modifications, DNA can be physically modified to influence the expression of individual genes up to whole chromosomes. This modification can occur directly, through the methylation of DNA, or more indirect, through the modification of histones that control the packaging of the DNA. Although the exact nature of how these modifications influence gene expression is not fully understood, it is generally believed today that it is mainly through hindering or allowing transcription factors to interact with their target DNA-sequences physically. These modifications are usually tissue-specific, and some are known to be transgenerational inherited, laying the epigenetic background for the transcriptomic landscape (Quadrana & Colot, 2016; Mathieu *et al.*, 2007; Takeda & Paszkowski, 2006). The second level, transcriptional regulation, involves the direct regulation of the rate of which mRNA is being synthesised. This is controlled by DNA-binding proteins (transcription factors) that interact with regulatory elements of the gene to either increase (activate), or decrease (repress), mRNA transcriptional rate. Transcriptional regulation is commonly regarded as the most crucial step for controlling gene expression and protein levels.

After transcription, the pre-mRNA is being processed before being transported out of the nuclei for translation. During this processing, the mRNA undergoes modifications to get increased stability (capping and polyadenylation), splicing to yield the final coding sequence and RNA editing. mRNA can also be targeted for degradation by micro-RNA in a process called RNA-induced silencing. In not broken down in this process, the mature mRNA can then be utilised for translation into a protein by the ribosomes. Gene expression at this step is primarily regulated at the initiation step of translation, which is influenced by the ribosomal binding site of the mRNA, mRNA binding proteins and mRNA secondary structures. Except for the importance of the ribosomal binding sequence, we still know very little about how translational efficiency of mRNA into protein is modulated. The final level of gene regulation is traditionally regarded to be the post-translational modifications of proteins at which proteins can be covalently modified to influence their activity. This can e.g. be through phosphorylation, which can influence activity or ubiquitination that is associated with stability. In addition to this, many proteins require other proteins and cofactors to function correctly, which is usually not regarded as a level of gene regulation due to their, often, transient nature.

2.1.2 Transcription factors

As mentioned earlier, the transcriptional rate of DNA into mRNA is the most critical stage at which gene regulation can occur due to its significant impact on mRNA and protein levels. Proteins involved in regulating this process are called transcription factors. Currently, more than 2000 transcription factors have been identified in Arabidopsis, which constitutes approximately 6 % of the total number of genes. This is comparable to other plants that have been analysed revealing that plants, in general, have four times as many transcription factors than other eukaryotes (Zhang *et al.*, 2011). This increased diversity of transcription factors among plants are likely a result of increased adaptiveness to selection pressure (Shiu *et al.*, 2005). In eukaryotes, they can be divided into four, although somewhat artificial, main classes; (i) general transcription factors, (ii) sequence-specific transcription factors, (iii) chromatin-related transcription factors and, (iv) transcriptional cofactors. General transcription factors, also known as basal transcription factors, are involved in transcription in a non gene-specific way such as the TRANSCRIPTION FACTOR II B which is a part of the Pol II enzyme complex. Chromatin-related transcription factors are proteins that regulate gene expression by the depositing or removal of histone modifications leading to changes in the condensation of the chromatin. Transcriptional cofactors constitute the middle ground between sequence-specific (see below) and general transcription factors. They act on a gene-specific basis but lack DNA-binding regions. Their role is varied but commonly described as parts of the mediator complex that relays information from sequence-specific transcription factors to the transcriptional complex (Allen & Taatjes, 2015).

Sequence-specific transcription factors

Sequence-specific transcription factors (from here onwards referred to as only transcription factors or TFs) are proteins that influence the expression of a specific set of genes through direct interaction with DNA target sequences called *cis-regulatory elements* (CREs) or *response element.* These target sequences are commonly located upstream of the translational start site, but they can also be situated downstream or even within the transcribed region. TFs are commonly grouped in families based on sequence homology within their DNA-binding domains (Riechmann *et al.*, 2000), and currently, there are around 60 families identified within the plant kingdom (Zhang *et al.*, 2011; Mitsuda & Ohme-Takagi, 2009). Compared to animals, plants show a much higher diversity in TFs with a large number of orthologues groups present (Zhang *et al.*, 2011). The structure of TFs has been described as modular in

nature, consisting of several clearly defined regions (Latchman, 1997; Frankel & Kim, 1991). However, more recent research has shown that many TFs contain regions important to their function devoid of any structural organisation at normal cellular conditions, so-called intrinsically disordered regions, which challenges this view (Ma *et al.*, 2015a; Pazos *et al.*, 2013). A special case of gene regulatory proteins are pioneer transcription factors, TFs that also have chromatin-modulating functions. These modulate chromatin states by altering histone modifications in and around a specific set of genes to allow non-pioneer TFs to get access to their target genes. They have been described as playing an essential role in changing the epigenetic landscape during development (Mayran & Drouin, 2018).

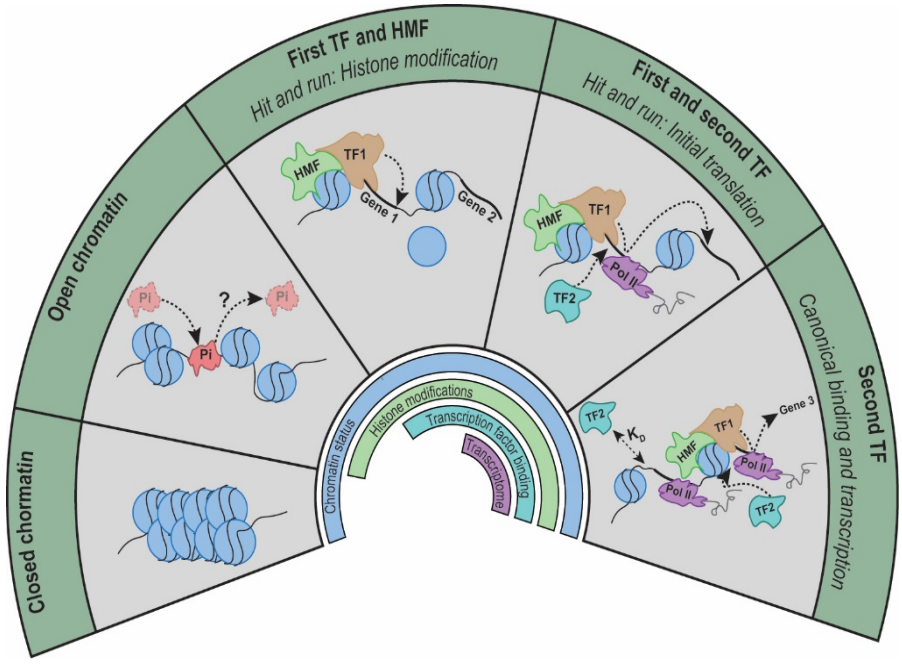


Figure 1. The *hit and run* model of transcription. In the first step a pioneer transcription factor (Pi) binds its target sequence in the condensed and inaccessible chromatin. This leads to a localized opening allowing a first transcription factor (TF1) to bind together with a histone modifying factor (HMF) and further loosen the chromatin. TF1 is able to recruit TF2 which are able to initiate transcription through Polymerase II (Pol II). Once the chromatin is properly opened the TF1 and HMF dissociates to target the next gene. The chromatin is now fully accessible for TF2 allowing it to bind and initiate transcription without the help of TF1. Central region shows the temporal process where chromatin is the first to be influenced and remains so during the whole process while the transcriptome only is influenced at the later stages. Modified from Swift and Coruzzi (2017).

2.1.3 Hit and run model of transcription

It is easy to make a mistake by viewing the genomic landscape as an exposed strand of DNA just waiting for TFs to bind and turn on the expression of their target genes. This is, of course, a too simplistic model and in reality, the DNA is in a constant shift between its open and available configuration where transcription occurs and different stages of inaccessibility. Development of new techniques and the combinatorial application of these have made it possible for us to increase our knowledge of how transcription is initiated within this complex and dynamic environment. One of these models is known as the *hit and run* model and describes how transcription can occur in regions of inaccessible DNA through the utilisation of specialised sets of TFs (Figure 1) (Lamke *et al.*, 2016; Para *et al.*, 2014). In this model, the closed chromatin is first targeted by a pioneer TF that can locate and bind to its CRE despite the condensed state of the chromatin. This leads to the localised opening of the chromatin structure around the CRE of the pioneer TF, allowing for a first set of TFs to bind. It is still unknown whether the pioneer TF remains bound to the DNA or disassociates to allow the following TF to bind, but in animals, it is known to remain (Yu *et al.*, 2011; Zaret & Carroll, 2011). This first TF recruits a histone-modifying factor (HMF) that further loosens the chromatin structure around the target gene (Lamke *et al.*, 2016). This temporary binding of the first TF allows it to recruit the second TF, which in turn can initiate transcription (Varala *et al.*, 2015). This constitutes the *run* part of the model. Once the first TF and HMF have disassociated, the chromatin is fully available, and the second TF can bind and initiate transcription independently.

2.1.4 Gene regulatory networks

In its purest form, a gene regulatory network (GRN) is a set of genes that interact with each other to produce a specific cellular response. Even with this straightforward definition, several conclusions can be drawn. A GRN must consist of at least two members with an, at least temporary, hierarchical co-dependency. Furthermore, the network must be able to sense the input of information and create output as a response. Since a system that only can be activated once is of minimal use, there must also be a regulatory feedback system in place. Although not always required, or even wanted, most GRN also requires tight regulatory control to avoid incorrect activation.

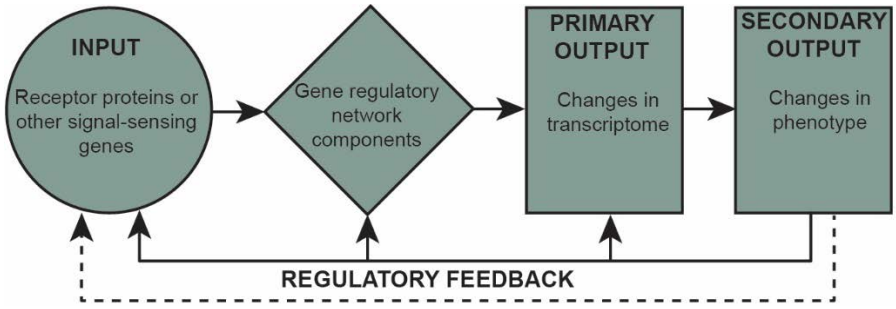


Figure 2. The basic components of a gene regulatory network.

With this information, we can also draw up a somewhat more complex regulatory chart where the output is divided into primary output (mRNA) and secondary output (protein and cellular response) (Figure 2). We also know that regulatory control is present at many different levels in GRN, creating a complex feedback system.

One major problem with working with GRNs today is how we define and limit them. GRNs are easy to define in theory, but any meaningful biological definition quickly becomes impossible to set unless the scope is limited. However, since no genes function in an empty void, it can be challenging, even impossible, to separate one GRN from another due to crosstalk and common factors. This becomes more and more obvious with the improvement of high-throughput techniques from which the available data of protein interactions can be retrieved.

The primary purpose of constructing GRNs is to be able to connect the genotype to the phenotype. Very few traits are determined by a single locus but a trait is rather the combined effect from thousands of genes. Within plant science, the ability to construct a bridge between the genome and a specific phenotype is especially valuable within plant breeding where the ability to, through GRN-based models, model how a certain genotype combined with a specific environmental interaction would influence the phenotype. Increased understanding of specific GRNs might also lead to the use of biomarkers within plant breeding.

2.1.5 LEAFY COTYLEDON1 – A subunit in the Nuclear Factor Y complex

LEAFY COTYLEDON1 (LEC1) is an extensively studied TF involved in many aspects of embryo development in plants. It is a member of the eukaryotic Nuclear Factor Y (NF-Y) TF complex, also known as CCAAT Binding Factor (CBF). The NF-Y complex consists of three subunits; A, B and C, together forming the core complex that can bind to the 5´-CCAAT-3´ DNA motif (Nardini *et al.*, 2013). While most eukaryotes genomes only encode single members of each subunits plants usually have several homologs with Arabidopsis having between 10-13 homologs of each subunit (Petroni *et al.*, 2012; Siefers *et al.*, 2009). LEC1 and the close homolog LEC1-like are both NF-YB subunits (NF-Y9 and NF-Y6 respectively).

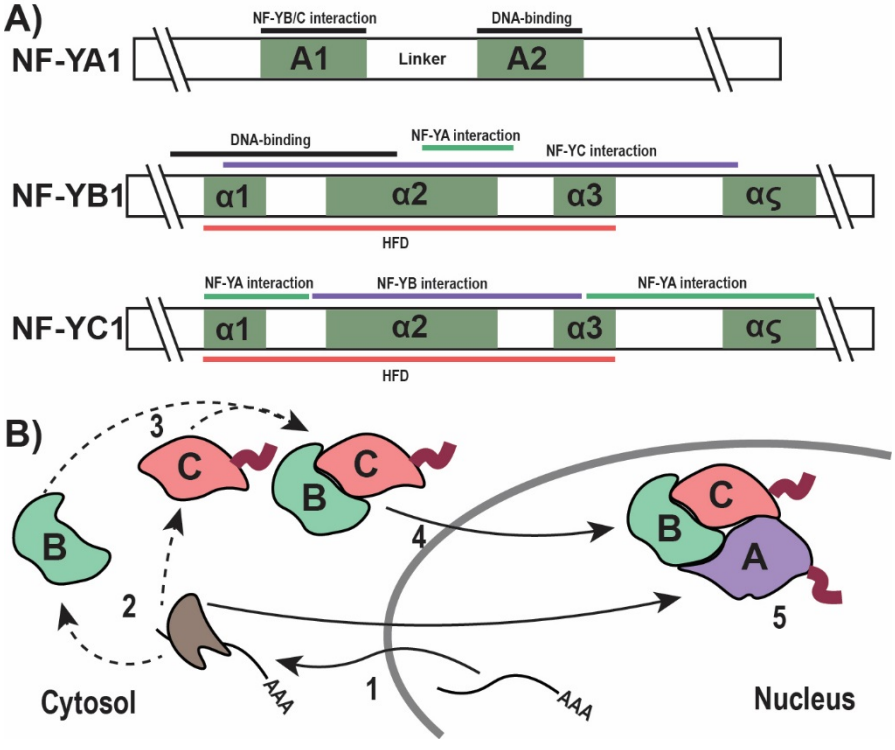


Figure 3 A) General visualisation of the domains of Arabidopsis NF-YA, NF-YB, NF-YC. Modified from (Zhao *et al.*, 2016). B) Model of co-transportation of NF-YB and NF-YC where the mRNA is transported out of the nucleus (1) and translated in the cytosol (2). NF-YA, having a nuclear localisation signal, is being transported directly to the nucleus while NF-YB and NF-YC remain in the cytosol. NF-YB binds to the NF-YC (3), and the complex is then transported to the nucleus through the NF-YC nuclear localisation signal (4). In the nucleus, the full NF-Y complex can be assembled (5).

NF-YA consists of two conserved α-helix domains where A1 is vital for interaction with NF-YB/C while the second one is the CCAAT-binding motif (Figure 1A) (Laloum *et al.*, 2013; Nardini *et al.*, 2013; Petroni *et al.*, 2012). The NF-YB and NF-YC both contain a histone fold domain (HFD), which is involved in protein-DNA and protein-protein interactions (Kahle *et al.*, 2005; Frontini *et al.*, 2004). The DNA-binding domain of NF-YB has been suggested to stabilise the binding to the CCAAT-motif by binding to non-specific flanking regions (Zhao *et al.*, 2016; Nardini *et al.*, 2013; Masaki *et al.*, 2005). NF-YA and NF-YC both have nuclear localisation signals, but only NF-YA is known to be persistently localised to the nucleus (Calvenzani *et al.*, 2012). The current model of how NF-YB and C are transported to the nucleus is that NF-YB piggy-backs on NF-YC which then can be transported through its nuclear localisation signal (Figure 3B) (Nardini *et al.*, 2013; Hackenberg *et al.*, 2012; Liu & Howell, 2010). In this model, the mRNAs are being transported to the cytosol for translation after which NF-YA is immediately localised to the nucleus. NF-YB and NF-YC then assemble their HFDs in a head-to-tail manner that most likely allows NF-YCs nuclear localisation signal to be exposed, leading to the dimer being transported past the nuclear membrane. In the nucleus, the NF-YA, which contain the CCAAT-binding domain, can interact with the dimer, thereby forming the complete trimeric complex (Figure 3A).

In Arabidopsis, LEC1 (NF-YB9) is primarily expressed during the early to middle stages of seed development (Schmid *et al.*, 2005). Loss-of-function mutants of LEC1 show defects during late embryogenesis with the most apparent phenotype being disturbed cotyledon development leading to enlarged cotyledons with trichome structures, i.e. phenotypes of mature leaves (Meinke *et al.*, 1994). The *lec1* mutant also have severe phenotypes in storage compound accumulation with a lowered number of oil bodies and disturbed spatial starch accumulation (Meinke *et al.*, 1994). LEC1 has also been shown to be essential for early embryo development with *lec1* mutants having defects in suspensor identity (Lotan *et al.*, 1998). Mature seeds of *lec1* mutants also show reduced seed dormancy and desiccation tolerance, ectopic accumulation of chlorophyll and anthocyanins (Parcy *et al.*, 1997; Meinke *et al.*, 1994; West *et al.*, 1994; Meinke, 1992). The *lec1* mutants also show a reduced response to abscisic acid (Parcy *et al.*, 1997; Meinke *et al.*, 1994). Ectopic expression of LEC1 causes vegetative cells to take on characteristics of embryonic structures causing somatic embryogenesis (Lotan *et al.*, 1998), accumulate oil (Mu *et al.*, 2008) and seed storage proteins (Kagaya *et al.*, 2005). Seed storage protein

accumulation by LEC1 is, however, less significant than those observed by *FUSCA3* and *ABSCISIC ACID INSENSITIVE3* (Kagaya *et al.*, 2005). Based on this, LEC1 has been suggested to be a master regulator of late embryogenesis involved in the maintenance of cell identity and the accumulation of seed storage compounds (Braybrook & Harada, 2008). Interestingly, LEC1 has been shown to regulate the expression of FLOWERING LOCUS C (FLC), a gene involved in the initiation of flowering, which could indicate that LEC1 has additional roles yet to be discovered (Tao *et al.*, 2017).

2.1.6 The AFL-subfamily

The B3-superfamily is a large family of TFs found in photosynthetic organisms (Yamasaki *et al.*, 2013), defined by their B3 DNA-binding domain, that in Arabidopsis has at least 93 members (Peng & Weselake, 2013). The first identified B3 TF was VIVIPAROUS1 (VP1) identified in maize by McCarty *et al.* (1991) to which an orthologue later was identified in Arabidopsis as the ABSCISIC ACID INSENSITIVE3 (ABI3) (Suzuki *et al.*, 2001; Giraudat *et al.*, 1992). Together with two other closely related B3 TFs identified in Arabidopsis, LEAFY COTYLEDON2 (LEC2) and FUSCA3 (FUS3), ABI3 forms the AFL-subfamily. It is believed that the AFL-subfamily originated in bryophytes from a proto-ABI3, indicating that the AFL-subfamily is specific to land living plants (Carbonero *et al.*, 2017). All so far studied higher plant species appear to maintain at least one copy of ABI3 and FUS3 but LEC2 appears to missing from all monocots (Devic & Roscoe, 2016). Also in several species of dicotyledons such as potato (*Solanum tuberosum* L.), tomato (*Solanum lycopersicum* L.) and grape (*Vitis vinifera* L.) no clear LEC2 homologs has been identified (Devic & Roscoe, 2016). The divergence within the AFL-subfamily has been suggested to be a result of specialization within the family (Li *et al.*, 2010).

Arabidopsis ABI3 is distinguished from LEC2 and FUS3 by the presence of four identified domains A, B1, B2 and B3 (Figure 4) (Suzuki *et al.*, 1997; Giraudat *et al.*, 1992). The B1-domain is involved in interacting with TFs from the bZIP-family while the B2-domain has been shown to be necessary for the activation of abscisic acid (ABA) responsive genes through the ABA-response element (ABRE) (Nakamura *et al.*, 2001; Ezcurra *et al.*, 2000; Bies-Etheve *et al.*, 1999; Hill *et al.*, 1996). The B2-domain of Arabidopsis ABI3 and LEC2 also contain a motif that has been shown to be responsible for direct interaction with LEC1 (Boulard *et al.*, 2018). The B3-domain is the DNA-binding domain which targets the RY-motif (5'-CATGCATG-3') although some

differentiation, or promiscuity, appear to be present among the AFL-subfamily (Nag *et al.*, 2005; Reidt *et al.*, 2000; Suzuki *et al.*, 1997). While LEC2 only have maintained the B2 and B3-domains, FUS3 also have an A-domain located on the opposite end (C-terminus) of ABI3 (Boulard *et al.*, 2018; Han *et al.*, 2017; Lu *et al.*, 2010). The A, B1 and B2 domains of ABI3 have been suggested to form a meta-domain, the co-activator/co-repressor (COAR) domain, involved in seed desiccation and dormancy (Suzuki *et al.*, 2014). Recently, a novel meta-domain of ABI3 was put forward by Boulard *et al.* (2018). This meta-domain, named SEEDEV, is composed of the TKxxARxxRxxAxxR motif found in the B2-domains of LEC2 and ABI3 and the B3-domain. This SEEDEV-domain was suggested to be involved in earlier developmental phases than those regulated by the COAR.

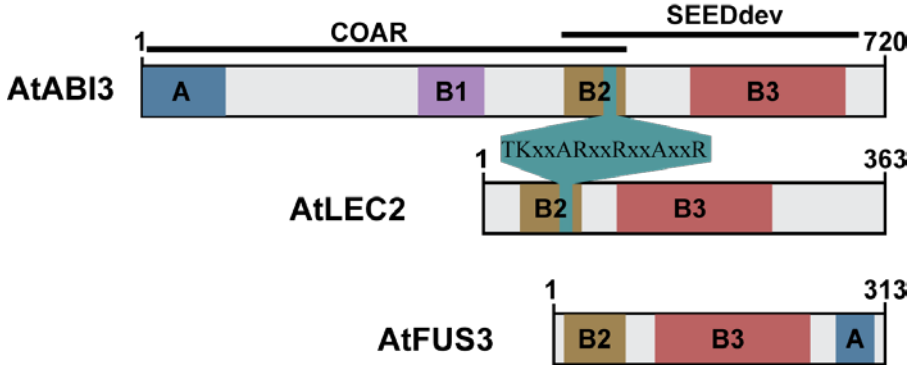


Figure 4. Structural representation of Arabidopsis ABI3, LEC2 and FUS3 that together make out the AFL-subfamily of transcription factors which are important for seed development.

In Arabidopsis, the AFL-subfamily is primarily expressed during seed development, with LEC2 being the earliest, followed by ABI3 and finally FUS3 (Schmid *et al.*, 2005). Much research has been done on the individual role of each of the three members of this subfamily, and it appears as if they have largely overlapping functionality. Mutant analysis of the AFL-subfamily has revealed that they all show severe phenotypes when it comes to seed maturation and seed filling. For example, all mutants show a severe defect in the accumulation of seed storage proteins (Kroj *et al.*, 2003; Parcy *et al.*, 1997; Nambara *et al.*, 1995; Keith *et al.*, 1994; Meinke *et al.*, 1994; Parcy *et al.*, 1994; West *et al.*, 1994; Meinke, 1992). However, distinct phenotypes separating the different mutations have also been identified. *abi3* show a severe inability to suppress chlorophyll accumulation during seed maturation, while *lec2* has a similar but more tissue-specific phenotype (Kroj *et al.*, 2003;

Nambara *et al.*, 1995). Related to this is the disturbed accumulation of anthocyanins in the cotyledons observed in *fus3* and *lec2* mutants but not in *abi3* (Stone *et al.*, 2001; Parcy *et al.*, 1997; Keith *et al.*, 1994; Meinke *et al.*, 1994). However, the anthocyanin accumulation in *lec2* is confined to specific regions and/or tissues just as the chlorophyll phenotype (Meinke, 1992). One of the classical phenotypes of *lec2* is the presence of trichomes on the cotyledons which are also shared by *fus3* but not by *abi3* (Keith *et al.*, 1994; Meinke *et al.*, 1994). Only *abi3* show a reduced response to abscisic acid while *lec2* is the only one to produce desiccation tolerant seeds (Parcy *et al.*, 1997; Keith *et al.*, 1994; Meinke *et al.*, 1994).

Members of the AFL-subfamily have also been studied from a regulatory perspective using transcriptomics and genome-wide chromatin immunoprecipitation techniques to understand their exact role better. For ABI3, a set of 98 direct targets have been identified through a combination of microarray and chromatin immunoprecipitation confirming that ABI3 is directly regulating genes involved in embryo maturation (Monke *et al.*, 2012). A similar approach has been used to study FUS3, revealing that approximately one-fifth of the FUS3 targeted genes also were targets of ABI3 (Wang & Perry, 2013; Monke *et al.*, 2012). Interestingly, only 19 % of the genes identified to be bound by FUS3 were also differentially regulated by the overexpression of FUS3 indicating either the presence of additional regulatory systems or that the proper targeting of FUS3 requires additional co-factors (Wang & Perry, 2013). In a microarray-based experiment done on Arabidopsis overexpressing LEC2, it was shown that also LEC2 targets genes involved in embryo maturation and seed filling with the addition of genes involved in auxin response (Braybrook *et al.*, 2006).

All in all, the AFL-subfamily are master regulators of late seed development involved in many aspects of seed filling, embryo maturation, and seed desiccation tolerance. Mutant analysis has revealed that they have both overlapping and specific roles which are somewhat mirrored in the, separately conducted, studies aiming towards the identification of their regulon.

2.1.7 The AP2-subfamily

In Arabidopsis, the superfamily AP2/ERF consists of a total 145 to 147 identified transcription factors divided among three major subfamilies (Nakano *et al.*, 2006; Sakuma *et al.*, 2002). It is present in all known plant species with 135 identified in grapevine (Jaillon *et al.*, 2007), 200 in poplar (Zhuang *et al.*,

2008) and 163 in rice (Rashid *et al.*, 2012). The superfamily is named after the DNA-binding domain first identified in APETALA2 (AP2) which, paradoxically for being the archetype for the superfamily, contains two AP2-domains situated in tandem (Riechmann & Meyerowitz, 1998; Huala & Sussex, 1992; Drews *et al.*, 1991). In Arabidopsis, 14 genes, all with tandem AP2-domains, form the AP2-subfamily while the other two subfamilies only have a single AP2-domain but in addition also have one or more ethylene responsive factors (AP/ERF subfamily) or the DNA-binding B3-domain (RAV-subfamily) also found in the AFL-subfamily (Kim *et al.*, 2006; Nakano *et al.*, 2006; Kagaya *et al.*, 1999; Riechmann & Meyerowitz, 1998). In addition to these a few soloist genes have also been identified in several species. Despite being clear members of the AP2/ERF-superfamily they are too different to easily be placed within a single subfamily (Rashid *et al.*, 2012). In addition to these there are three to four soloists that only contain a single AP2-domain (Nakano *et al.*, 2006; Sakuma *et al.*, 2002). These are often grouped together as a subgroup in the AP2-subfamily despite them not fulfilling the original definition of this subfamily. The largest of the subfamilies within the AP2/ERF superfamily is the AP2/ERF with 121 or 122 members in Arabidopsis while the RAV-subfamily only has six members (Nakano *et al.*, 2006; Sakuma *et al.*, 2002). As previously mentioned, the proper AP2-subfamily (i.e. including only the members with tandem AP2 domains) has 14 members in Arabidopsis many of which play important roles in embryo and flower development. Among these is the AP2 which is known to be essential for floral patterning, floral meristem maintenance, ovule development and seed coat growth (Ripoll *et al.*, 2011; Jofuku *et al.*, 2005; Ohto *et al.*, 2005; Jofuku *et al.*, 1994; Bowman *et al.*, 1993). Another is BABY BOOM (BBM) which is involved in meristem maintenance, stem cell niche (SCN) patterning of root tissue and embryogenesis during seed development (Horstman *et al.*, 2017; Aoyama *et al.*, 2012; Galinha *et al.*, 2007; Boutilier *et al.*, 2002). BBM has recently been highlighted as an important candidate target for future plant biotechnological applications (Horstman *et al.*, 2017).

WRINKLED1 is essential for de novo fatty acid synthesis in seeds

Another member of the AP2-subfamily that has been highlighted as a promising target for plant biotechnological application is WRINKLED1 (WRI1). Named after the wrinkled seed phenotype of Arabidopsis *wri1* mutants WRI1 has been shown to be an essential regulator of oil accumulation in seeds (Cernac & Benning, 2004; Focks & Benning, 1998). The wrinkled seed phenotype is a result of the seed losing up to 70 % of its oil content (Focks & Benning, 1998). WRI1 has been shown to influence oil accumulation by regulating key steps of glycolysis and *de novo* fatty acid synthesis leading to increased carbon flow to the triacylglycerol assembly in the endoplasmic reticulum (To *et al.*, 2012; Maeo *et al.*, 2009; Cernac & Benning, 2004). In addition to this WRI1 also downregulates starch biosynthesis further redirecting carbon towards fatty acid synthesis (Grimberg *et al.*, 2015). WRI1 also downregulates genes involved in photosynthesis (Grimberg *et al.*, 2015). This shows that WRI1 is an essential transcription factor when it comes to reprogramming plant tissues towards becoming strong sinks, a process that in developing dicot seeds typically takes place during later stages of embryogenesis. This is also the time point when *WRI1* is expressed in Arabidopsis (Schmid *et al.*, 2005).

Figure 5. Schematic drawing of Arabidopsis WRI1 containing a tandem AP2-domain and with intrinsically disordered regions (IDRs) towards the N-terminus (IDR1) and C-terminus (IDR2 and 3).

Arabidopsis WRI1 consists of a tandem AP2-domain located centrally in the protein which is flanked by one intrinsically disordered region (IDR) located towards the N-terminus and two IDRs located towards the C-terminus (Figure 5). IDRs are regions in the amino acid sequence with low three-dimensional order at physiological conditions and are common to find in eukaryotic transcription factors (Tsafou *et al.*, 2018). Their exact function is largely unknown but they have been shown to function e.g. as flexible linkers for binding, form ordered complexes upon binding, function in protein-protein interactions and act as phosphorylation sites (Staby *et al.*, 2017; Bu & Callaway, 2011; Fuxreiter *et al.*, 2007; Iakoucheva *et al.*, 2004). In the third IDR of WRI1 there is a PEST-motif which has been shown to be involved in protein stability (Ma *et al.*, 2015a). Deletion of the PEST-motif resulted in increased oil accumulation through increased stability of WRI1 (Ma *et al.*,

2015a). In the first AP2-domain a 14-3-3 binding domain has been identified in Arabidopsis WRI1 and co-expression of WRI1 and 14-3-3 proteins increase the stability of WRI1, potentially by protecting it from degradation of E3-ligases (Ma *et al.*, 2016). WRI1 is known to bind to the AW-box ([CnTnG](n)7[CG]) which has been found in the promoter regions of many glycolysis and fatty acid synthesis related genes (Maeo *et al.*, 2009). *WRI1* has been identified as a regulatory target of several members of the LAFL-network such as LEC2 and FUS3 (Wang & Perry, 2013; Baud *et al.*, 2007).

2.2 Plant storage compounds

Plants are primarily photoautotrophs meaning that they can convert electromagnetic energy from photons to chemical energy. For most plants, this means that light energy is captured through the light-dependent reaction and the energy is temporarily stored as adenosine triphosphate (ATP) and the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH). In the next stage, the energy stored in ATP and NADPH is used for the synthesis of more stable energy-dense molecules in a process called the light-independent process. This is done by a series of reactions involving the fixation of CO_2 by reducing the carbon. The final product is glycerate 3-phosphate (G3P), a stable three carbon molecule allowing the previously fixated light energy to be bound in the reduced carbon skeleton. Together, these two stages are known as photosynthesis which is the primary source of biologically available carbon. The resulting moieties of G3P are then either used in the maintenance of the photosynthetic tissue or used for synthesising larger carbohydrates for either redistribution or storage. Simpler carbohydrates, mainly sugars, which stem directly from photosynthesis are generally called photosynthates.

Photosynthates can directly be used for storage, which some plants such as the sugar beet (*Beta vulgaris* ssp. v*ulgaris* L.) utilise by storing sucrose. Although it is a relatively inefficient storage medium mainly due to high oxidation state of the carbons in sucrose, it is “cheap” and easy to both synthesise, metabolise and transport for the plant. Apart from its low energy density, it is also readily available to pathogens and predators, and in high amounts, it influences osmotic potential as well as the chemical equilibrium of, e.g. glycolysis. To counter this, plants have developed more efficient ways of storing carbon by specialised storage compounds as well as spatial sequestration. When it comes to storage compounds from photosynthates plants mainly use starch, oil, and protein. Whereas both starch and protein can

be used for both short- and long-term storage, they have somewhat different roles. Starch is a polymer of simple sugar moieties and must go through additional metabolic remodelling to be utilised. Proteins, on the other hand, are composed of amino acids that can enter the metabolic pathways much further downstream. Energywise, the most efficient storage form of carbon in plants is oil, where the carbon reaches a much lower oxidation state. However, oil is both “expensive” to synthesise and require an extensive, specialised machinery to be utilised in the plant.

2.2.1 Photosynthate partitioning

A major challenge faced by plants is the redistribution of energy across the whole organism. Vascular plants primarily do this through a process called translocation, where soluble photosynthates are passively transported through the phloem. Two critical aspects of this are *phloem loading* and *unloading.* During phloem loading, photosynthates are initially transformed into non-reducing sugars, primarily sucrose, and transported to the sieve cell. This transport can be utterly passive via the symplast or at least partially active via the apoplast depending on species. From the sieve cell, the sugars can be loaded into the phloem and from there transported by an osmotic pressure gradient. The unloading occurs in a similar way where sugars are symplastically transported to the companion cell. From there, it can continue through a symplastic pathway, which is common in most tissues. The alternative is the use of an apoplastic pathway where sugars are unloaded into the apoplast and transported into the target cell by, e.g. proton-coupled sucrose carriers (Carpaneto *et al.*, 2005). Sucrose can also be converted to hexoses by invertases localised to the cell wall and then imported by hexose transporters (Sturm & Tang, 1999).

The transport direction of sugars is dependent on whether the tissue is a *source* or *sink* tissue. A source tissue, as the name implies, is a tissue where photosynthates are being produced in a higher rate than is used up, making it a net exporter of sugars. The opposite, a tissue where photosynthates are being consumed at a higher rate than it can produce, is, therefore, a sink tissue. Source tissues can be divided into primary and secondary source tissues where the primary are photosynthetic tissues actively producing sugars from photosynthesis while secondary are storage organs where sugars are being remobilized from storage compounds/organs where little or no photosynthesis occurs. A leaf, for example, is a primary source while the root of sugar beet during bolting is a secondary source. Sinks, on the other hand, are often

divided between weak and strong sinks. Weak sinks are more or less all non-storage organs such as meristems and other developing structures that require relatively large amounts of energy. Strong sinks are primarily seeds, storage roots, and tubers that require a large number of carbon moieties for storage purpose. *Sink strength* is dependent on two factors; (i) sink size and (ii) sink activity as follows:

$$Sink\ strength = sink\ size\ x\ sink\ activity$$

Where *sink size* is the absolute volume of the tissue and *sink activity* is the ability of the tissue to remove imported sugars. Phloem unloading usually differs between weak and strong sinks where strong sink tissues commonly utilise apoplastic transport.

Source-sink dynamics and carbon allocation

It is easy to think about the relationship between sources and sinks as fixed where leaves always are sources and roots acts as sinks. In reality, all tissues fluctuate between being a source and a sink during different parts of their life cycle. Leaves start out as sinks during their development, and at the end of the leaf development, they transition into a source. Similarly, seeds start as strong sinks only to function as source tissue during germination. Our current

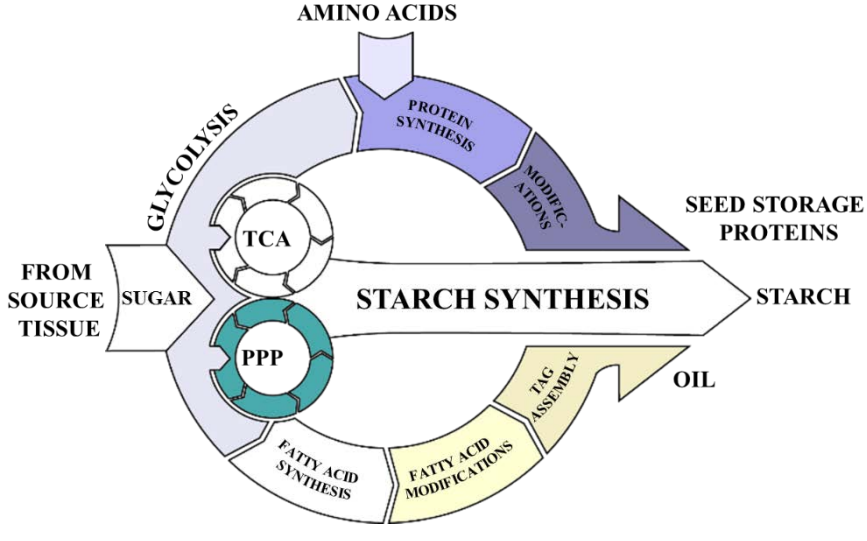


Figure 6. Simplified model over carbon allocation in developing seed tissue. With the help of cellular metabolism the carbon, in the form of sugars, can be shuffled into seed storage proteins, oil or starch. Figure modified from Baud *et al.* (2008).

knowledge of this transition is mainly based on mechanistic models of carbon partitioning between different parts of the plant. However, the process of how a tissue acquires its identity as a sink tissue has very little to do with this partition. Instead, it's based on the carbon allocation within the tissue. Carbon allocation is, although the term is sometimes used interchangeably with carbon partitioning, the process of diverting carbon flows in the cells metabolic pathways. This carbon allocation is the main factor behind *sink activity*, which is based on the ability to assimilate carbon moieties per biomass unit. In developing seeds, there are three main storage compounds to which carbon is being allocated, protein, starch and oil (Figure 6).

2.2.2 Starch

Starch is a homopolysaccharide consisting of a large number of glucose molecules linked together by glycosidic bonds. Two types of starch can be found in plants; amylopectin and amylose. Amylopectin consists of long chains with (6->100) glucosyl groups linked together by α-1,4-bonds. Each chain is linked to another chain with an α-1,6-bonds creating a large, branched structure. In contrast, amylose is a much less branched, almost linear, helical molecule. Starch is organised in granules consisting of ca. 75 % amylopectin and 25 % amylose, although the exact composition varies from species to species (Buleon *et al.*, 1998). It is generally believed that amylopectin is the older of the two starches with a direct evolutionary relationship with the glycogen found in animals (Ball & Morell, 2003). This is also supported by the fact that amylose is amorphous and cannot form granules on its own (Zeeman *et al.*, 2010). It has even been suggested that amylose synthesis requires the presence of amylopectin granules (Denyer *et al.*, 2001). The formation of starch granules is a complex and far from the fully understood process. For a comprehensive review, see Perez and Bertoft (2010). What is well known, however, is that granule size distribution and shape is species specific. Their size ranges from submicron to 100 µm in diameter. Starch granules are built up in layers of crystalline shells covering semi-amorphous layers creating a tree ring-like structure.

Starch as a dynamic carbon depot

In addition to the two chemically differentiated types, starch can also be divided into two types based on their biological role; (i) storage starch and, (ii) transitory starch (Streb & Zeeman, 2012). While storage starch is synthesized in sink tissues (such as seeds), transitory starch is synthesised in photosynthetic

source tissue during times of high photosynthesis activity (such as leaves during light hours) only to be degraded at a later timepoint.

As previously mentioned, storage starch is primarily synthesized in nonphotosynthetic sink tissues with the purpose of long term storage. Examples of specialised non-seed storage organs that accumulate starch can be found all over the plant kingdom, e.g. potato tubers and cassava roots. Storage starch is also transiently being accumulated in non-specialized tissue such as root cells but differs from transitory starch based on precursor origin and localisation. Starch synthesised for storage is primarily synthesized from hexoses such as glucose-6-phosphate and fructose-6-phosphate being imported over the cell membrane by transporters. Storage starch is synthesized in specialised organelles called amyloplasts where, in general, several starch granules are deposited in each amyloplast.

While storage starch is synthesised by specific cells, in specialised organelles with the purpose of long-term storage, transitory starch is produced in the chloroplast by photosynthetic cells for short-term storage of sugar moieties. This primarily occurs in the mesophyll cells that also make up the bulk cell volume of the leaves. While storage starch originates from hexoses transported across the cell membrane, the synthesis of transitory starch utilises fructose-6-phosphate originating from the light-independent reaction as the precursor. However, mutant analyses have revealed that there might be additional pathways able to supply the chloroplastic starch synthesis with substrates (Vitha et al., 2000; Streb et al., 2009; Tsai et al., 2009). Transitory starch is accumulated during times of high photosynthetic activity (e.g. day) at a nearly linear rate only to be degraded during periods with low photosynthetic activity (e.g. night) also in a near-linear manner. Despite this pattern, it should be noted that there is a delay in the synthesis and degradation of starch, indicating that the regulatory system is not purely based on photoresponse. The biological purpose of this cyclic process is to remove photosynthates efficiently during the day to allow photosynthesis to occur unhindered. The stored carbon is then remobilised, primarily in the form of non-reducing sugars to support metabolism with energy during periods of low photosynthesis. It is also likely that this system helps protect the metabolites utilised in the light-independent reaction from being depleted due to excessive transport of trioses.

Starch biosynthesis

While the formation of starch granules remain obscure, the biochemical process of synthesising the large homopolymers of glucose is well mapped out

and understood (Figure 7). In non-photosynthetic tissues, the first step of starch synthesis is the degradation of sucrose by sucrose synthase into one fructose and one UDP-glucose molecule (Figure 7A). From the UDP-glucose a phosphate group is removed to yield a glucose-1-phosphate (G1P) by the action of UTP-glucose-1-phosphate uridylyltransferase. Following this is the displacement of the phosphate group from the 1st to the 6th carbon by a cytosolic phosphoglycerate mutase. This G6P can then be transported into the amyloplast where a plastidial phosphoglycerate mutase remodels it back to G1P which then is charged using ATP to form adenosine diphosphate-glucose (ADP-Glucose) by an ADP-glucose pyrophosphorylase. The ADP-glucose can then be used for synthesising the glucose polymers. The fructose formed when sucrose is degraded in the first step can enter the glycolysis and be phosphorylated by a fructokinase forming fructose-6-phosphate (F6P). The F6P is then transformed to G6P by a phosphoglucoisomerase from which it can be transported into the amyloplast. The transportation into the amyloplast with the subsequent transformation into ADP-glucose constitute the first dedicated steps in this model of starch synthesis. In cereals, an alternative cytosolic pathway exists where G1P can be transformed into ADP-glucose by a cytosolic ADP-glucose pyrophosphorylase and then transported into the amyloplast by a specific ADP-glucose/ADP translocator. This specific pathway is dominating over the plastidic formation of ADP-glucose in cereals, leading to an earlier dedicated step in starch biosynthesis taking place already in the cytosol (Beckles *et al.*, 2001).

For transitory starch, the biosynthetic pathway mimics the steps of the storage starch after the transport of G6P into the amyloplast (Figure 7B). The main difference is that transitory starch is completely synthesized inside the chloroplast and that the G6P used as a substrate is supplied by the conversion of F6P by a plastidial phosphoglucoisomerase. Transient starch is also stored in the chloroplast and not in specialised storage organelles.

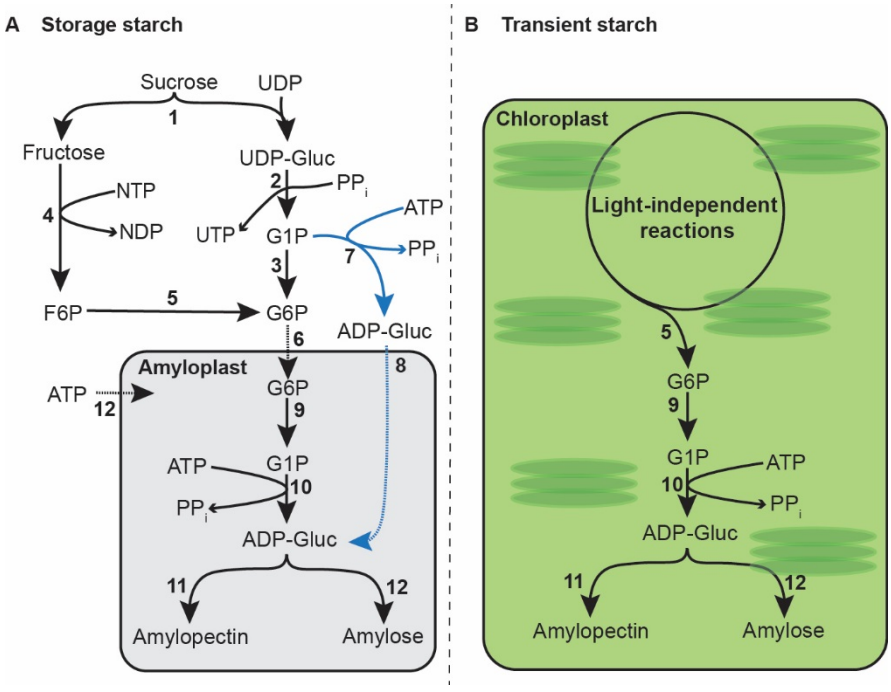


Figure 7. Starch synthesis pathways for storage starch (A) and transient starch (B). Enzymes involved in the pathways are as follows: 1, Sucrose synthase; 2, UTP-glucose-1-phosphate uridylyltransferase; 3, Phosphoglycerate mutase (cytosolic); 4, Fructokinase; 5, Phosphoglucoisomerase; 6, Glucose-6-phosphate/Pi translocator; 7, ADP-glucose pyrophosphorylase (cytosolic); 8, ADP-glucose/ADP translocator; 9, Phosphoglycerate mutase; 10, ADP-glucose pyrophosphorylase; 11, Starch synthases (I-IV), Starch branching enzymes I-III), Isoamylases (I-III); 12, Granule-bound starch synthase. Metabolites are abbreviated as follows: UDP, Uridine diphosphate; Gluc, Glucose; F6P, Fructose-6-phosphate; G1P, Glucose-1-phosphate; ATP, Adenosine triphosphate; PPi, Pyrophosphate; UTP, Uridine triphosphate; ADP, Adenosine diphosphate. Transport pathways are noted by dashed arrows. Blue arrows indicate alternative pathway present in cereals. Figure modified from (Geigenberger, 2011).

Starch degradation in endospermic tissue

Starch degradation involves the process of reverting the complex starch molecules back into simple and soluble sugars that can be further used in a multitude of metabolic pathways. The pathways utilised for starch degradation differs significantly between storage and transitory starch as well as between different types of storage tissues. Remobilisation of carbon deposits from the endosperm into the growing seedling is an important process during germination in endospermic seeds. In cereals, this process is initiated by a pulse of gibberellins diffusing from the embryo into the living aleurone layer surrounding the non-living starchy endosperm. This results in the synthesis and

release of α-amylase from the aleurone. α-amylase is able to cleave internal α-1,4-bonds in the polyglucose chains resulting in a release of shorter glucose polymers. These polyglucose chains are continuously targeted by α-amylases and β-amylases until the disaccharide maltose remains. Maltose is further broken down to glucose by maltases. The glucose can then be transported into the embryo for further utilisation on downstream processes. Branched starch chains are handled by a set of specialised enzymes called limit dextrinases that attacks the α-1,6-bonds of the branches creating linear polysaccharides that can be targeted by α-amylases.

2.2.3 Oil

In addition to starch, many plants also store carbon in an even more reduced form as triacylglycerols (TAGs), also called oil. TAGs consist of three fatty acids, usually 14-24 carbon long attached to a glycerol backbone. TAG is present, at low levels, in most plant tissues but are only accumulated in large amounts in specialised storage tissue such as seeds. Despite this, an elevated amount of TAG can be detected in several specialised tissues such as pollen as well as leaves during senescence (Kaup *et al.*, 2002). Much like starch, plants are generally not able to transport TAG, meaning that the precursors first must be transported to the target tissue, and there be utilized for TAG-synthesis. As previously mentioned, TAG is one of the most energy dense storage compounds available to plants. This has led to many plant species more or less altogether abandoning starch as an energy reserve in seeds and only accumulate oils and other fats. These seeds are commonly called oil-seeds and include many members of the Brassicaceae such as oil-seed rape and Arabidopsis. However, also, e.g. potato, oil palm, and flax are to be regarded as oil-seeds. In oil-seeds, the TAG is generally stored in the embryo and in some cases the aleurone. Storage within the endosperm of seeds is rare but occurs in, e.g. oat and castor bean. There also appears to be a trend towards oil-seeds in evolutionary younger plant families. This, together with our understanding of starch being the storage compound of an early type of seeds (Floyd & Friedman, 2001), indicates that TAG as the primary carbon storage in seeds is relatively recent in plant evolution.

Originally TAG was only thought to be a compound for the efficient storage of energy. However, an increasing amount of research shows that TAG is involved in processes as diverse as pollen tube penetration (Wolters-Arts *et al.*, 1998), gametophyte development (Shockey *et al.*, 2016) and biotic defence (Shimada & Hara-Nishimura, 2015). By increasing TAG synthesis plants also

protect cells against increased levels of free fatty acids (FFAs), which at high levels are toxic, that are released during, e.g. senescence by letting TAG sequester them until they can be broken down (Yang *et al.*, 2015).

De novo fatty acid synthesis

TAG is the end product of two separate pathways, *de novo* fatty acid synthesis and TAG-assembly. In plants, fatty acids (FAs) are synthesised in the chloroplast using acyl-CoA as the primary substrate while the adding of the FAs to the glycerol to form TAG takes place in the endoplasmic reticulum (Figure 8).

The first step of the *de novo* FA synthesis is the conversion of pyruvate to acetyl-CoA through the addition of a coenzyme A (CoA) and the subsequent removal of one carbon. This reaction is the first committed step in FA synthesis and is catalysed by a large heteromeric protein complex called the plastidial pyruvate dehydrogenase complex (PDHC). Experiments have shown that the amount of acetyl-CoA in the chloroplast is limited and that almost all (77 %) the cells available CoA is present in the chloroplast (Post-Beittenmiller *et al.*, 1992). This indicates that PDHC and the regeneration of acetyl-CoA from pyruvate have the potential for being a severe rate-limiting step in FA synthesis. Pyruvate is primarily produced by glycolysis in the cytosol and transported to the chloroplast, but also the chloroplast is able to generate pyruvate from more complex carbohydrates through similar processes. In the next step, the acetyl-CoA (two carbons in the acetyl group) can either be used directly as a substrate together with malonyl-ACP (three carbons in the malonyl group) by ketoacyl-ACP-synthase III to form the first elongation step. This yields a four carbon long 3-ketoacyl-ACP that in the following steps can be elongated by the addition of two more carbons from malonyl-ACP in a condensation reaction catalyzed by ketoacyl-ACP-synthase I. Acetyl-CoA not used for seeding the initialization of FA synthesis is converted to malonyl-CoA by acetyl-CoA carboxylase which in turn is transformed to malonyl-ACP by a malonyl-CoA acyl carrier protein malonyltransferase. The four carbon long 3-ketoacyl-ACP is reduced to a hydroxyacyl-ACP by a ketoacyl-ACP reductase. Following this is the dehydration step where hydroxyacyl-ACP dehydrase catalyses the formation an enoyl-ACP before the final enzyme, an enoyl-ACP reductase, reduces the enoyl-ACP to an acyl-ACP which then is ready for the next step of elongation. These four steps, condensation, reduction, dehydration and reduction is repeated six times to yield a final product of 16:0-ACP. To

produce 18:0-ACP a final round of the cycle is run with ketoacyl-ACP-synthase III catalysing the condensation. The 18:0-ACP can then be desaturated by stearoyl-ACP desaturase to form 18:1-ACP. The final plastidial step of FA synthesis is the removal of ACP from the 16:0-ACP, 18:0-ACP and 18:1-ACP to create FFA. This is catalyzed by fatty acyl thioesterase A and B. Before the FFA can be transported from the plastid they are bound to a CoA by a large group of enzymes called long-chain acyl-CoA synthetases.

Assembly of triacylglycerols

After the FA synthesis, which takes place in the plastid, the next step is the formation of TAG. Initially, the newly formed FAs are transported into the endoplasmic reticulum (ER) where it forms an acyl-CoA pool. The linear process in which three FA is attached to the glycerol-3-phosphate (G3P) is known as the Kennedy pathway. The first step is the acetylation of the *sn-1* position of the G3P by the enzyme glycerol-3-phosphate acyltransferase (GPAT) forming 2-lysophosphatidic acid (LPA). In the next step, the *sn-2* position is acetylated by a 2-lysophosphatidic acid acyltransferase (LPAAT) and the formed phosphatidic acid (PA) subsequently dephosphorylated by a phosphatidate phosphatase to form diacylglycerol (DAG). DAG does not only constitute a precursor for the final step in TAG-assembly but also function as a substrate for membrane lipid synthesis. Currently, three main pathways have been identified in the conversion of DAG to TAG. The first pathway is the acetylation of the *sn-3* position by a diacylglycerol acyltransferase (DGAT) using an acyl-CoA as an acyl donor. An alternative pathway is the acylation of DAG using phosphatidylcholine as the acyl-donor, which is catalysed by a phospholipid:diacylglycerol acyltransferase (PDAT). The third and final pathway is proposed to use another DAG moiety as the acyl-donor by a diacylglycerol:diacylglycerol transacylase although candidate genes have yet to be identified.

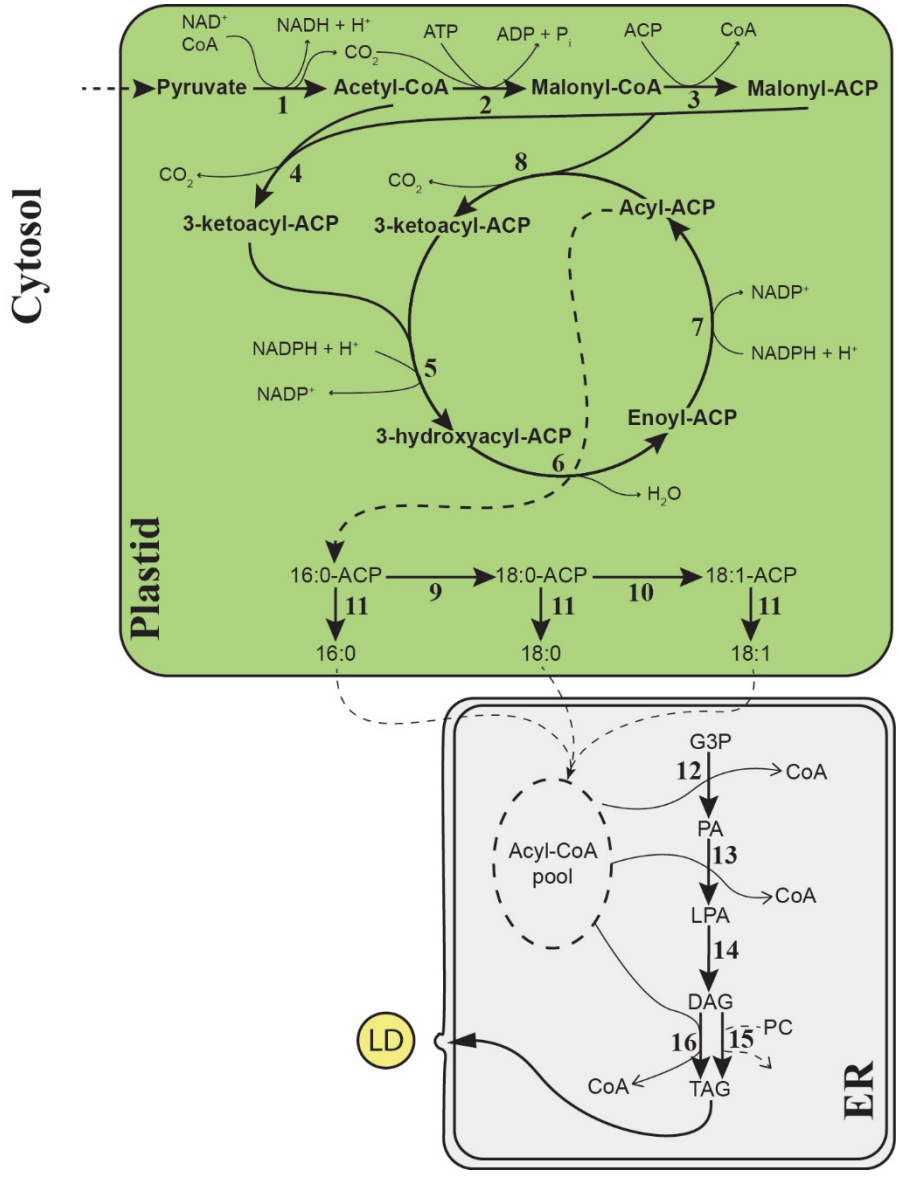


Figure 8. Schematic overview of the *de novo* fatty acid synthesis taking place in the plastid and triacylglyceral assembly in the endosplasmatic reticulum in plants. Enzymes involved in the pathways are as follows: 1, Pyruvate Dehydrogenase Complex; 2, acetyl-CoA carboxylase; 3, Malonyl-CoA:ACP Malonyltransferase; 4, Ketoacyl-ACP Synthase III; 5, Ketoacyl-ACP Reductase; 6, Hydroxyacyl-ACP Dehydrase; 7, Enoyl-ACP Reductase; 8, Ketoacyl-ACP Synthase I; 9, Ketoacyl-ACP Synthase II; 10, Stearoyl-ACP Desaturase; 11, Acyl-ACP Thioesterase A/B; 12, Glycerol-3-Phosphate Acyltransferase; 13, 1-Acylglycerol-3-Phosphate Acyltransferase; 14, Phosphatidate Phosphatase; 15, Acyl-CoA : Diacylglycerol Acyltransferase; 16, Phospholipid:Diacylglycerol Acyltransferase. Abbreviations: ER, Endoplasmic Reticulum; LD, Lipid Droplet. Transport pathways are noted by dashed arrows. Pathway and names after Li-Beisson *et al.* (2013).

Triacyglycerol storage and lipid droplets

During the TAG-synthesis in seeds newly formed TAG, being a neutral lipid is accumulated between the lipid bilayers of the ER membrane, eventually forms the swelling of the ER membrane (Figure 8). The exact mechanism and dynamics involved in this process are mostly unknown. We currently do not know if the accumulation of TAG inside the ER membrane is the cause for oleosins to aggregate in the cytosolic side of the ER membrane monolayer or if it is this aggregation of oleosins that specifies the accumulation of TAG. It could also be that both the aggregation of TAG and oleosins coincides, either dependent or independent of the other. Oleosins are small proteins that contain a long hydrophobic hairpin structure which is inserted into the outer monolayer of the lipid droplet (LD). Genes encoding oleosins are present in green algae, moss and vascular plants with an increasing number of copies and expression level (Huang, 2018). They are translated through signal-recognition particle-guided mRNAs on the cytosolic side of the ER membrane (Beaudoin *et al.*, 2000; Thoyts *et al.*, 1995; Qu *et al.*, 1986). We know that a reduction in oleosins to TAG ratio leads to larger oil bodies showing that oleosins are essential for the formation of the LD (Ting *et al.*, 1996). It is still unclear whether the mature LDs are formed at the ER or if smaller proto-LDs are merged. Many other proteins are associated with LDs in seeds, but their function is generally poorly understood. One of these groups are the SEIPINS, which is known to increase TAG accumulation when overexpressed in plants (Cai *et al.*, 2015).

Triacyglycerol degradation

TAG degradation in seedlings takes place in specialised organelles called peroxisomes. At the initiation of TAG degradation peroxisomes are formed adjacent to the LDs. Between these, a tubular extension is formed, probably to allow transport of substrates to the peroxisome. A SUGAR-DEPENDENT LIPASE1 (SDP1) has been attributed to this process (Thazar-Poulot et al., 2015; Eastmond, 2006). This interaction has been shown to be negatively regulated by the presence of sucrose, revealing a direct connection between TAG breakdown and gluconeogenesis (Cui et al., 2016). It is currently unknown where TAG-degradation occurs, but it is believed to primarily occur in, or in association with, the LDs. The free fatty acids are then transported to the peroxisomes for the further breakdown in the β-oxidation from which acetyl-CoA is channelled to the glyoxylate cycle. The glyoxylate cycle is essentially a modified citric acid cycle from which succinate is transported out of the peroxisome to the citric acid cycle. In the citric acid cycle, substrates

used in gluconeogenesis can be formed or the succinate be completely oxidized to extract energy. It can be noted that in the endosperm of cereal grains, which is a dead tissue at maturity and therefore lacks functional peroxisomes, β-oxidation of fatty acids from TAG cannot occur. However, lipases were shown to be secreted from the scutellum into the endosperm of barley (Jensen & Heltved, 1982) which could release fatty acids from TAG. Furthermore, analytical and microscopy studies of germinating oat grains showed that lipids were taken up by the scutellum (Leonova et al., 2010) This shows that oil stored in the cereal endosperm is not a dead-end product but can be used as energy source during germination.

2.2.4 Protein

The third type of storage compound in seeds is the protein that differs from starch and protein since it is primarily not carbon storage but rather storage for amino acids and therefore nitrogen. This means that in order to utilise stored proteins, the cells do not need to degrade it fully but can instead salvage the amino acids reducing the strain on metabolic pathways. However, if needed, the cells can completely degrade the protein and recycle the nitrogen. Proteins are primarily stored in the embryo although an assortment of proteins can be found in the endosperm in seeds of many species. We still do not know whether these are to be regarded as true storage proteins or play essential roles in the organisation of the endosperm. Storage proteins are commonly grouped into two distinct groups; (i) vegetative storage proteins (VSP) and, (ii) seed storage proteins (SSP). VSPs accumulate exclusively in vegetative tissue such as leaves and stems while SSPs only accumulate in seeds. Underground tubers form a middle ground and, depending on species, are known to accumulate both SSPs and VSPs. The term storage protein is problematic as it originally applied to all proteins that are accumulating in the maturing seed, regardless of their function. This has led to some confusion as, e.g. oleosins were, and sometimes still are, regarded as SSPs as they accumulate in seeds. However, their role is not primarily to act as storage molecules but rather to help organise lipid bodies, making them not true SSPs. Although, future research might continue to describe additional functionality to true SSPs increasing the complexity of how they should be termed. In the context of this text, SSPs are defined as proteins accumulating in the embryo after cell division has been completed and with the primary known function of acting as energy and nutrient reserves for the seedling.

Seed storage proteins

SSPs are synthesised in the embryonic tissue of the seed during the maturation phase of embryo development. Even though seed protein content is an important agronomical trait of crops, as compared to the detailed genetic research on starch and oil relatively little research has been focused on processes governing their accumulation and remobilisation. Instead, many different SSPs have been identified in several species and were usually named after the species they were identified in such as legumin and vicilin which are found in legumes. More recent research has shown that although the number of copies varies between species, they can be grouped in distinct groups. In Arabidopsis, SSPs is made up of cruciferines and napins (12S globulins and 2S albumins respectively). While the napins are coded for by five different genes (*at2S1-at2S5*), only three cruciferines are known (*CRA1*, *CRB*, and *CRC*). In addition to these, Arabidopsis also carries genes coding for vicilin (a 7S globulin) although it is not accumulated in seeds. In potato (*Solanum tuberosum* L.) tubers, one of the primary storage protein is patatin (Shewry, 2003). Patatin also has lipase activity targeting membrane lipids. Arabidopsis also have genes coding for patatin-like proteins, which exclusively is expressed in roots and required for lateral root development (Rietz *et al.*, 2010). In Arabidopsis, SSP synthesis takes place at the rough ER, and immediately after translation, they are stored in protein storage organelles. In plants, there are two types of storage units; protein bodies (PBs) and protein storage vacuoles (PSVs). Different species utilise these storage forms in different ways. PBs are primarily formed from the rough ER and store a mixture of water-soluble albumins and water-insoluble prolamins (Choi *et al.*, 2000; Shewry *et al.*, 1995). PSVs from the smooth ER and mainly accumulate albumins in dicots (Choi *et al.*, 2000).

2.3 Seed development

Seeds generally consist of three main parts. The embryo, endosperm and the seed coat. While the seed coat develops from the maternal tissue surrounding the ovule, the embryo and endosperm develop from the fertilisation events of the egg cell and the central cell, respectively. In gymnosperms, during the pollination process, two separate fertilisation events coincide. It is believed that this double fertilisation initially resulted in two identical diploid embryos although it is not known why or what relation these two embryos shared. With evolutionary history, one of these embryos were maintained to form the proper embryo while the other fertilised cell underwent specialisation to form the endosperm (Friedman, 1998).

2.3.1 Embryogenesis

Embryogenesis is the course in which a single embryonic stem cell, through a series of highly regulated processes, develops into an embryo. In many animals, the embryo constitutes a miniaturised version of the fully matured body plan, but in plants, it consists of a less complex structure. In plants, several different types of embryogenesis occur, the two most commonly known are somatic embryogenesis and androgenesis. Furthermore, *de novo* organ development (e.g. root development from cuttings) shares many common aspects of embryogenesis. In this part, only embryogenesis related to seed development will be discussed. Plant embryogenesis can be divided into several more or less distinctly separated stages depending on the required resolution. Here, the following stages will be used; (i) pre-globular, (ii) globular, (iii) heart, (iv) torpedo and, (v) mature. The pre-globular stage starts with the fertilisation of the egg cell after which the zygote elongates, and an asymmetric cell division takes place, leading to a smaller apical and a larger basal cell. The basal cell continues to divide horizontally and forms the suspensor structure, which plays two important roles during embryogenesis. The first is to act as an umbilical cord for the developing embryo allowing the plant to supply the embryo with nutrients and other metabolites. The second is to act as physical support and pushing the embryo into the developing endosperm. The apical cell undergoes three rounds of cell division to produce a 16-cell embryo, this is the start of the globular stage. At this stage, the protoderm, that will become the epidermis, is present as the outermost cell layers.

Apical-basal patterning

As soon as the globular stage is entered the process of polarisation occurs. In this process, the embryo will establish an apical-basal patterning which will be maintained throughout the plants life. This is achieved through the local activation of GRN governing cell differentiation, thereby changing the cellular characteristics of cells in specific regions of the embryo. One major question in the study of plant embryogenesis was whether this process was due to inherited genetic programming or due to position and signalling. We know today that the absolute majority of this process is a direct result of local signalling (Torres-Ruiz & Jurgens, 1994). During the globular stages, auxin is being actively transported towards the suspensor structure by auxin transport proteins belonging to the PINFORMED-family (PINs) (reviewed by Jenik *et al.*

(2007)). This asymmetric auxin localisation is maintained throughout the embryo development and influences the stability of several auxin response factors such as MONOPTEROS and BODENLOS, which are required for proper axialization. The TF CUP SHAPED COTYLEDON3 is one of the earliest factors marking what will become the apical section of the embryo already at the start of the globular stage (Hibara *et al.*, 2006) while PLETHORA1 is expressed in what will become the basal cells (Aida *et al.*, 2004).

The transition between the globular stage and heart stage

In the middle of the globular stage, after the apical-basal patterning has taken place, the shoot apical meristem is established through a set of TFs where WUSCHEL (WUS) is the most important regulator. WUS forces the cells in the region to maintain their stem cell identity hindering other TFs from influencing them towards specific cell fates. WUS is counteracted by CLAVATA3, essentially a differentiation-promoting peptide, which is expressed in the periphery of the shoot apical meristem (SAM) thereby maintaining the size of the SAM. The establishment of both apical meristems marks the start of the transition between the globular stage to the heart stage. During this phase, the embryo leaves the radial symmetry and establishes bilateral symmetry by initiating the establishment of the cotyledons also yielding adaxial/abaxial differentiation. During this process, new GRN is activated by additional auxin focal points being established in the two areas where the cotyledons will be formed activating expression of the AP2-family TF DORNROSCHEN and the MYB-family TF ASYMMETRIC LEAVES1. At the same time, the expression of CUP SHAPED COTYLEDON1/2 marks the border between the two cotyledons. Cotyledon differentiation is determined by the TF KANADI, which is counteracted by a set of HD-ZIP III TFs expressed in the SAM, causing the abaxial/adaxial differentiation.

Embryo maturation and establishment of energy reserves

At the end of the heart stage, all organogenesis is completed, and the embryo tissue is fully formed. The next stage, the torpedo stage, marks the start of cell elongation and the filling of the embryo with energy reserves, primarily oil and protein. This also marks the turning point of starch accumulation in the endosperm of Arabidopsis which from here will be degraded to supply the metabolic processes with energy and carbon (Baud *et al.*, 2002a; Norton & Harris, 1975). Early in this process, chlorophyll starts to accumulate, and proplastids are remodelled to become chloroplasts, which also facilitates the

accumulation of fatty acids (Ruuska *et al.*, 2004; Mansfield & Briarty, 1991). The functionality and synthesis pathways of the storage compounds are discussed in chapter 2.2. How the transition between the morphogenesis phase to the embryo maturation phase is regulated is currently unknown, but it is known that sugar and hormone balances are crucial for this transition to occur.

During embryo development, there is a steady rise of ABA, which is known to early during embryogenesis promote embryo growth by counteracting GA. However, at the initiation of embryo maturation, there is a spike in ABA, resulting in the inhibition of embryo growth (Yang & Feng, 2015; Nambara & Marion-Poll, 2005). FUS3 has been suggested to play an essential role in this shift by repressing genes involved in gibberellin (GA) synthesis and upregulating ABA synthesis (Gazzarrini *et al.*, 2004; Nambara *et al.*, 2000). Initially, ABA is supplied from a maternal source but if embryos early in the maturation phase from Arabidopsis are excised from the developing pod, they are capable of completing maturation and germinate (Frey *et al.*, 2004; Raz *et al.*, 2001). This shows that embryonic ABA is in control of the maturation phase possibly suggesting that maternal ABA regulates morphogenesis while endogenous ABA marks the transition to maturation phase. Metabolic status has also been shown to be necessary for the transition to the maturation phase. High glucose to sucrose ratio has been shown to be correlated with morphogenesis while the opposite is present at the maturation phase (Ohto *et al.*, 2005; Baud *et al.*, 2002b; Weber *et al.*, 1997). This correlation might on the other hand, not be causational (Tomlinson *et al.*, 2004).

Late embryogenesis abundant proteins accumulate during embryo maturation
During the maturation phase, embryos accumulate large amounts of proteins clumped together under the common name of late embryogenesis abundant (LEA) proteins. LEA proteins are found in all land plants and consist of small hydrophilic, intrinsically disordered proteins that based on sequence homology, can be divided into a minimum of seven distinct groups (Hundertmark & Hincha, 2008; Garay-Arroyo *et al.*, 2000). In Arabidopsis, 51 genes coding for LEA proteins have been described with LEA group 4 and dehydrins being the most abundant groups (Hundertmark & Hincha, 2008). LEA proteins have been shown to, through their moldable structure, protect the structure and activity of other proteins during dehydration and freezing (Reyes *et al.*, 2005; Sanchez-Ballesta *et al.*, 2004; Bravo *et al.*, 2003; Hara *et al.*, 2001; Lin & Thomashow, 1992). In addition to this Arabidopsis dehydrins have been shown to be able to bind more water during dehydration than other proteins (Bokor *et al.*, 2005). This strongly points to LEA proteins being important for the

desiccation tolerance of the seed, but very little research has been focused on trying to distinguish differential functions of the different classes of LEA proteins.

2.3.2 Endosperm development

Three main classes of endosperm exist; the cellular, nuclear and helobial. These classes are based on the presence and type of cell walls formed. In the cellular endosperm, cell walls are maintained through all cell divisions while the nuclear endosperm has a phase of cell division without the formation of cell walls. The helobial endosperm is essentially a mix between the two other types with half of the endosperm being cellular in type while the other half is nuclear. The nuclear endosperm is the most common and will, therefore, be the foundation for the following description of endosperm development.

After the fertilisation of the central cell by the male gamete, the now triploid primary endosperm enters the syncytial stage in which the central cell undergoes several cell divisions (in Arabidopsis eight divisions) forming a large, cell wall free coenocyte (Olsen, 2001). At around 2-4 days after pollination, the coenocyte halts its cell division and enters the cellularization stage, which is tightly connected with the cell differentiation stage. During these two stages, cell walls are formed, separating the previous liquid endosperm into compartments, and several synchronous cell divisions take place expanding the cellular endosperm. At the same time, the endosperm is differentiated into four distinctly different types; the embryo surrounding region, transfer cells, the starchy endosperm, and the aleurone layer. The embryo surrounding region was initially described in maize as an embryo adjacent endosperm partition in which members of the embryo surrounding region (Esr) gene family was expressed (Opsahl-Ferstad *et al.*, 1997). In maize, it is known to be involved in biotic defence (Balandin *et al.*, 2005; Magnard *et al.*, 2000; Opsahl-Ferstad *et al.*, 1997), signaling required for embryo development (Leduc *et al.*, 1996; Mol *et al.*, 1995) and the formation of the embryonic cavern in which the embryo develops (Clark & Sheridan, 1991). Similar structures have been observed in Arabidopsis and wheat but might be absent in other species. Transfer cells are specialised endosperm cells involved in the import of photosynthates from the maternal vascular tissue to the endosperm. In maize, the transfer cell structure is three cell layers thick with transfer cell identity decreasing with distance from the vascular tissue (Schel *et al.*, 1984). The most peripheral part of the endosperm forms the aleurone layer which also forms the epidermis of the endosperm with the exception of the

transfer cells. Size and organisation of the aleurone cells differ between species with wheat having one single layer while barley has two-four with an average of three (Evers & Millar, 2002; Jones, 1969; Buttrose, 1963b). The aleurone cells contain a large number of aleurone bodies which are protein storage vacuoles surrounded by lipid droplets (Swanson *et al.*, 1998; Morrison *et al.*, 1975; Buttrose, 1963a). To survive the seed dormancy aleurone cells undergo a developmental program in which desiccation tolerance is established. The maize ABI3 homolog, VIVIPAROUS1 (VP1), has been shown to be essential for this process to take place in the aleurone layer (Kao *et al.*, 1996; Hoecker *et al.*, 1995; Robichaud & Sussex, 1986). The determination of the aleurone cell fate is still mostly unknown. The currently most favoured hypothesis is that the aleurone cells are positionally established as the outermost layer of the endosperm through directed signalling that is maintained throughout the endosperm development (Olsen, 2001; Olsen *et al.*, 1998). The starchy endosperm forms the bulk of the endosperm in endosperm-rich species such as the cereals, and primarily consists of starch synthesised and stored in amyloplasts. For more details regarding the starch synthesis and degradation, see chapter 2.2.2. The starchy endosperm also contains storage proteins but generally no oil with the exception of a few species such as oat (Shewry *et al.*, 1995; Youngs *et al.*, 1977). Also, the starchy endosperm in wheat contains a small number of lipid droplets (Hargin & Morrison, 1980a).

2.4 The LAFL-network

Gene regulatory networks, as previously explained, are essential for the transition into and maintenance of different phases in the life cycle of plants. One such network involved in the process of seed maturation is the LAFL-network consisting of **L**EC1, **A**BI3, **F**US3, and **L**EC2. Although much work has gone into trying to elucidate the internal structure of the LAFL-network it has proven to be very complicated yielding complex pictures which are difficult to pinpoint and sometimes even giving contradictory results. One of the, if not the primary, reason for this is that studies of the network mainly is carried out by either mutation analysis in the seed environment or ectopic expression outside the seed environment. This can be problematic since phenotypes stemming from mutations can be hidden or influenced by compensating regulatory systems and depending on the ectopic expression system used the regulatory background varies. With the members of the LAFL-network introduced in chapter 2.3, this chapter will try to focus on what is known about the regulation of this network and the regulatory structure of it.

2.4.1 The internal regulatory system governing the LAFL-network

Members of the LAFL-network are known to regulate the expression of other members as well as through autoregulation. When ectopically expressed in young Arabidopsis leaves LEC1 was shown to upregulate the expression of ABI3 and FUS3 (Kagaya *et al.*, 2005). In a study of Kroj *et al.* (2003) LEC2 was suggested to directly regulate FUS3 but not ABI3 in cotyledons. In 2006, To *et al.* (2006) used a combination of reporter gene constructs and mutants to reveal that the major role of LEC2 during embryo development was to regulate the expression of ABI3 and FUS3 and that both FUS3 and ABI3 activate their expression. In the same study, LEC1 was shown to positively regulate the expression of ABI3 and LEC2 in the cotyledons and that ABI3 and FUS3 positively regulated each other. In young primary leaves and cotyledons, LEC2 has been identified as an activator of LEC1 as well as ABI3 and FUS3 expression (Stone *et al.*, 2008). FUS3 is also known to bind to the promoter region of LEC1, but no activation of LEC1 by FUS3 has so far been reported (Wang & Perry, 2013). This information is summarised in Figure 9, which clearly shows that the regulatory control asserted within the LAFL-network is complex and not suited to be viewed as either hierarchical or linear. Since such a model, as presented in Figure 9 does not allow for the individual activation of any member of the LAFL-network without large parts of the system being activated. Therefore, additional regulatory mechanisms must be involved in the regulation of the LAFL-network.

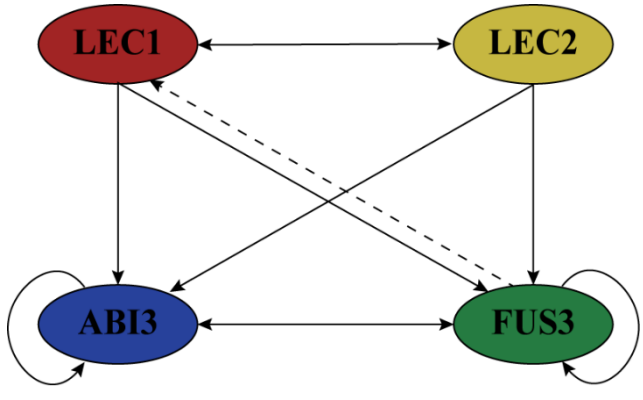


Figure 9. Summary of the regulatory interplay observed in studies of the LAFL-network of transcription factors.

2.4.2 Temporal expression pattern of members of the LAFL-network during seed development

By studying the data from Schmid *et al.* (2005) which is based on microarray analysis of developing seeds in Arabidopsis, we can start to draw up a rough, but interesting, picture of the temporal expression pattern of the LAFL-network (Figure 10). This data reveals that while LEC1 and LEC2 share almost identical temporal expression pattern ABI3 and FUS3 both have individual patterns. LEC1 and LEC2 reach their highest expression before or at the early globular stage. ABI3 reaches its highest expression level around the end of embryo morphogenesis and the start of seed filling but is present from at least middle/late globular stage. FUS3, on the other hand, lags behind ABI3 and reaches its peak right around when the most intensive seed filling takes place.

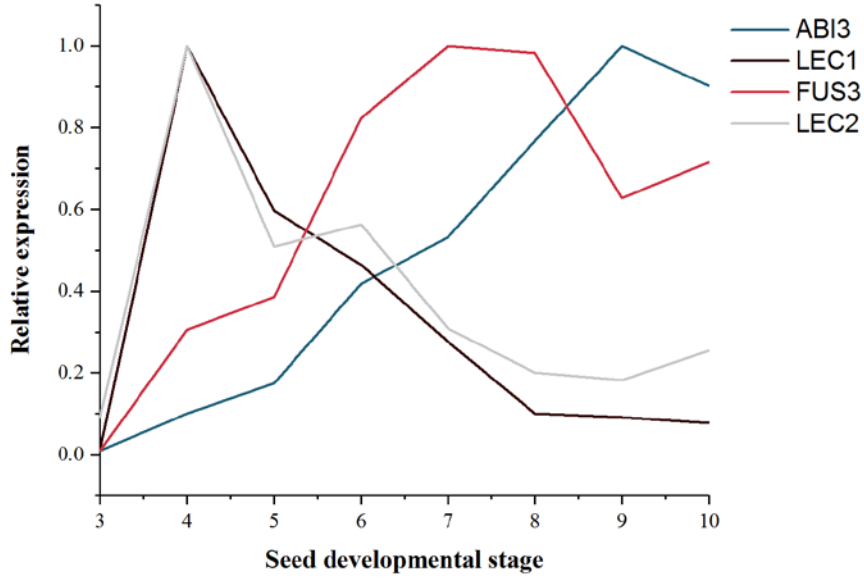


Figure 10. Expression of genes encoding the members of the LAFL-network of transcription factors during Arabidopsis seed development. X-axis denotes seed developmental stage spanning from mid-globular (3) to green-cotyledons embryo (10). Data from Schmid *et al.* (2005).

2.4.3 Spatial expression pattern of LAFL-network

Correctly mapping spatial expression of genes is difficult and tend to be error-prone. The most common approach is to fuse a piece of the promoter of a gene to a reporter gene such as GUS or GFP and visualise the expression of the reporter gene. Not only is the reporter protein a different protein in size and behaviour than the target protein, meaning that it will completely disregard directed transport and requirement for activation such as phosphorylation. It will also not be targeted by epigenetic or post-transcriptional regulation of gene expression meaning that a promoter that would otherwise be silent or silenced upon expression can be expressed. These modes of gene regulation are known to be very important in developmental processes to allow transcription and/or create gradients of proteins across tissues. Finally, GUS expression and the method for developing it is difficult to use for *in situ* quantification. None the less, alternative methods are expensive and time-consuming, reducing their usability. In a study by To *et al.* (2006b) promotors of FUS3, LEC2 and ABI3 were fused to a GUS reporter gene and transformed into Arabidopsis. The GUS expression pattern was analysed in mature embryos (10 days after pollination) right at the start of seed filling (Figure 11).

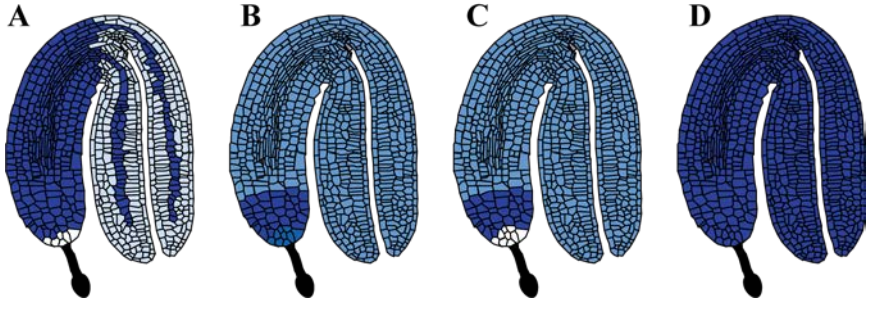


Figure 11. Interpreted expression pattern of A) LEC2, B) FUS3, C) ABI3 and D) LEC1 based on the promoter::GUS constructs reported by To *et al.* (2006b) and (Huang *et al.*, 2015).

These results indicate that *FUS3* is expressed throughout the mature embryo with a local maximum around the root meristem. LEC2 is primarily expressed in the hypocotyl and around the vascular tissue but not in the root tip. The expression of ABI3 is similar to that of LEC2 with the exception of the root tip, which lacks ABI3 expression. Similar studies using the LEC1 has revealed that LEC1 appear to be expressed in the whole embryo (Huang *et al.*, 2015).

2.4.4 Beyond the LAFL-network

In addition to the internal regulatory control observed within the LAFL-network, its members are also targets of external regulatory mechanisms often forming layers of complex, regulatory feedback loops. An example of this is the MADS-family transcription factor AGAMOUS-LIKE15 (AGL15) which is a direct target of LEC2 and FUS3. AGL15 in turn activates the AFL-subfamily creating a positive feedback loop. At the same time, AGL15 is subject to negative autoregulation silencing its own expression and thereby breaking the loop with the LAFL-network (Chen *et al.*, 2018b; Jia *et al.*, 2014; Wang & Perry, 2013; Zheng *et al.*, 2009a; Braybrook *et al.*, 2006; Zhu & Perry, 2005). The MYB-family transcription factors MYB115 and MYB118 downregulate LEC2 in the endosperm while at the same time upregulating LEC1 (Wang *et al.*, 2009; Zhang *et al.*, 2009). Also the AP2-subfamily member BABY BOOM has been shown to regulate all LAFL-network members in Arabidopsis seeds and seedlings (Horstman *et al.*, 2017). However, neither *agl15*, *myb115*, *myb118* nor *bbm* mutants show any of the phenotypes related to the loss of the LAFL-network strongly indicating that also the regulatory system for the LAFL-network is redundant in nature (Chen *et al.*, 2018b; Troncoso-Ponce *et al.*, 2016). The only repressors identified so far are the HIGH-LEVEL EXPRESSION OF SUGAR-INDUCIBLE GENE2 (HSI2), and HIGH-LEVEL EXPRESSION OF SUGAR-INDUCIBLE GENE2 -like1 (HSL1), which negatively regulate members of the LAFL-network after germination. However, it is still unclear whether this repression is direct or indirect through their confirmed repression of AGL15 (Chen *et al.*, 2018b; Chhun *et al.*, 2016; Schneider *et al.*, 2016; Jia *et al.*, 2014; Jia *et al.*, 2013; Veerappan *et al.*, 2012; Suzuki *et al.*, 2007; Tsukagoshi *et al.*, 2007; Tsukagoshi *et al.*, 2005). Also HSI2 and HSL1 are activated by FUS3, again creating a regulatory feedback loop (Zheng *et al.*, 2009b). Furthermore, PHAVOLUTA and PHABULOSA, HD-ZIP III transcriptions factors involved in establishing radial symmetry during embryo morphogenesis, have been shown to positively regulate *LEC2* expression (Tang *et al.*, 2012).

Epigenetic regulation of the LAFL-network

In addition to the more traditional transcriptional regulation of direct-acting transcription factors, members of the LAFL-network are also targets of chromatin-based regulation which play important roles during seed development. Chromatin-based regulation occur through the modifications of histones causing the chromatin to change between heterochromatin, a densely packed chromatin not allowing translation, and chromatin, an open chromatin where transcription can occur easily. Different histone modifications are

associated with different chromatin states with trimethylation of lysine four on histone 3 (H3K4me3) generally indicating an transcriptionally active state while H3K27me3 is indicative of silent genes (Sims et al., 2003). Chromatin-based regulation was previously considered to slow and cumbersome, only involved in long-term regulatory control, but more recent research has shown that e.g. H3K27me3 and histone ubiquitination by polycomb repressive complex (PRC) 1 and 2 confers fast response and allows for fine-tuning of gene repression (Mozgova & Hennig, 2015).

Among the LAFL-network several chromatin-based modifications are recognized to influence their expression. ABI3 expression is upregulated in mutants defective in H3K9 methylation and H3K4 demethylation (Zhao et al., 2015; Zheng et al., 2012; Liu et al., 2007). In mutants of PICKLE (PKL) and PICKLE-RELATED2 (PKR2), which both have reduced H3K27me3 levels, the LAFL-network remains expressed in post-embryogenesis tissues indicating that H3K27me3 plays an important role to silence the LAFL-network (Jing et al., 2013; Aichinger et al., 2011; Zhang et al., 2008; Rider et al., 2003). Acting antagonistically to PKL and PKR5 is the CHR5 which also has been associated in regulating members of the LAFL-network (Shen et al., 2015). Furthermore, PKL has been shown to function in gibberellic acid signalling by interacting with DELLA-proteins which have been shown to be important for the function of LEC1 during late embryogenesis (Hu et al., 2018; Zhang et al., 2014). In addition to this, CURLY LEAF (CLF) and FERTILIZATION-INDEPENDENT ENDOSPERM (FIE), both subunits in the PRC2-complex, mutants have been shown to have increased and ectopic expression of FUS3 and LEC2 as well as ectopic embryonic structures further supporting the idea that H3K27me3 is essential for the epigenetic regulation of the LAFL-network (Mozgova et al., 2017; Liu et al., 2016; Bouyer et al., 2011).

In conclusion, the LAFL-network is subject to complex internal and external regulatory mechanisms and several cases of regulatory feedback loops have been discovered. This suggest that the LAFL-network is able to, during certain circumstances, self-regulate to achieve a stable, or increasing, level of expression throughout the network. If the LAFL-network requires a threshold level to function, this could be a strategy for the plant to ensure that this threshold is reached. The idea of LAFL-function requiring a threshold level has been explored before by Santos-Mendoza *et al.* (2008), Devic and Roscoe (2016) and Lepiniec *et al.* (2018). In addition to this, the LAFL-network is also subject to chromatin-based regulation adding additional, and somewhat

difficult to interpret the importance of, regulatory levels on top of an already complex regulatory system.

3 Aims of this study

3.1 General aims

The general aim of this study was to further elucidate the transcriptional and gene regulatory network surrounding the accumulation of plant storage compounds in, as well as outside, the traditional seed environment. The work has mainly been focused towards the regulation of oil accumulation using the transcription factor WRINKLED1 to study how the master regulators LEAFY COTYLEDON1, LEAFY COTYLEDON2, ABSCISIC ACID INSENSITIVE3, and FUSCA3 interact to increase oil accumulation. An additional aim was to increase our understanding of how modification of carbon flow towards oil accumulation in non-oil accumulating tissues influences the metabolic and physiological environment. The aim was furthermore to evaluate the potential of using transcription factors to achieve this.

3.2 Specific aims

- How is seed storage compound, especially oil, accumulation controlled by the LAFL-network of transcription factors? (Manuscript III and IV).
- How does WRI1 influence the carbon flow between different storage compounds when expressed in wheat endosperm? (Manuscript II).
- How is the structure and function of the members in the AP2-subfamily of transcription factors related to each other? (Manuscript V).

- Are there novel mechanisms in the regulation of oil accumulation, either WRI1-dependent or WRI1-independent, to be described? (Paper I, manuscript II and IV).

4 Results and discussion

4.1 WRINKLED1 in wheat (paper II)

Several studies have shown that increased expression of WRI1 in already oil-accumulating dicotyledonous seeds leads to increased oil buildup through the increased redirection of carbon flow into *de novo* fatty acid synthesis (Chen *et al.*, 2018a; Ivarson *et al.*, 2017; An & Suh, 2015; Shen *et al.*, 2010; Cernac & Benning, 2004). Among cereals, oat (*Avena sativa* L.) is unique in accumulating relatively large amount of oil in its endosperm. While wheat only stores minuscule amounts of oil in the endosperm, oat can amass up to 17 % (Liu, 2011; Banas *et al.*, 2007; Peterson & Wood, 1997; Hargin & Morrison, 1980b). Oat has therefore previously been used as a model organism to study the partitioning of carbon between lipids and starch taking place in the endosperm (Hayden *et al.*, 2011; Ekman *et al.*, 2009). With the purpose of better understanding the differences in carbon allocation in the endosperm between one cereal able to accumulate oil (oat) and one who does not accumulate oil (wheat) we transformed oat endosperm *WRI1* (*AsWRI1*) into wheat behind a starchy endosperm specific promoter.

4.1.1 WRINKLED1 leads to accumulation of oil when expressed in wheat endosperm

Analysis of the transformed lines revealed a more than 9-fold increase in oil (from 0.6 % to 6.4 % dry of weight) accumulation in the endosperm together with seed coat parts of mature grains of the line carrying 12-inserts of *AsWRI1*. Additional lines with fewer gene inserts also showed increased oil accumulation, but the 12-insert line was by far the most efficient in amassing

oil. Lipid analysis during three different time points of development (10, 18 and 26 days post-anthesis (DPA)) revealed that oil is being accumulated in a linear pattern during the whole developmental process. Meanwhile, the mRNA-seq analysis revealed that the *AsWRI1* transcript levels were the highest at 10 DPA (around 5000 TPM) and then steadily decreased to around 100 TPM at 26 DPA. Despite this, the target genes of AsWRI1 did not appear to follow a similar expression pattern suggesting that; (i) WRI1 does not function in a dose-dependent way, (ii) levels of WRI1 is heavily influenced by post-transcriptional regulation or, (iii) the function of WRI1 is dependent on post-translational modifications. The oil content in mature grains also corresponded negatively with the starch content which in the 12-insert line was reduced from 62 % to 22 % of dry weight. The increase in oil accumulation and loss of starch is likely due to WRI1 upregulating central parts of the glycolysis and *de novo* fatty acid synthesis while downregulating starch synthesis as confirmed by mRNA-seq analysis. This fits well with what has been reported regarding the function of oat WRI1 and WRI1 in general (Grimberg *et al.*, 2015; Cernac & Benning, 2004). A small upregulation of genes involved starch degradation can also be observed in the mRNA-seq data potentially indicating that WRI1 also regulates parts of the remobilisation of starch observed in oilseeds. More surprising was the observation that the amount of sucrose increased from 1.5 % to 10 %.

4.2 Physiological effects from overexpressing WRINKLED1 in wheat endosperm

WRI1 is named after the wrinkled seed phenotype observed in the loss-of-function mutants in Arabidopsis (Focks & Benning, 1998). It is, therefore, a quite ironic twist that the overexpression of oat endosperm WRI1 in wheat endosperm also causes a wrinkled phenotype in the mature grain (Figure 12A-B). This is most likely a direct effect of the swollen phenotype that the lines transformed with AsWRI1 exhibit during seed development, which is also the primary cause for the thicker ears observed (Figure 12C-D). This swollen phenotype is caused by the accumulation of a sugar-rich liquid leading to the formation of a cavity as it disappears during seed desiccation, causing the seed coat to collapse inwards (Figure 12E-F). Furthermore, the 12-insert line also displays an intriguing phenotype related to aleurone layers. As previously mentioned, wheat, as most cereals, only has a single aleurone layer. In the 12-insert line, multiple layers of aleurone cells could be observed along the periphery of the starchy endosperm (Figure 13A-D). They also exhibit an oblong shape as opposed to the regular cuboid shape observed for most

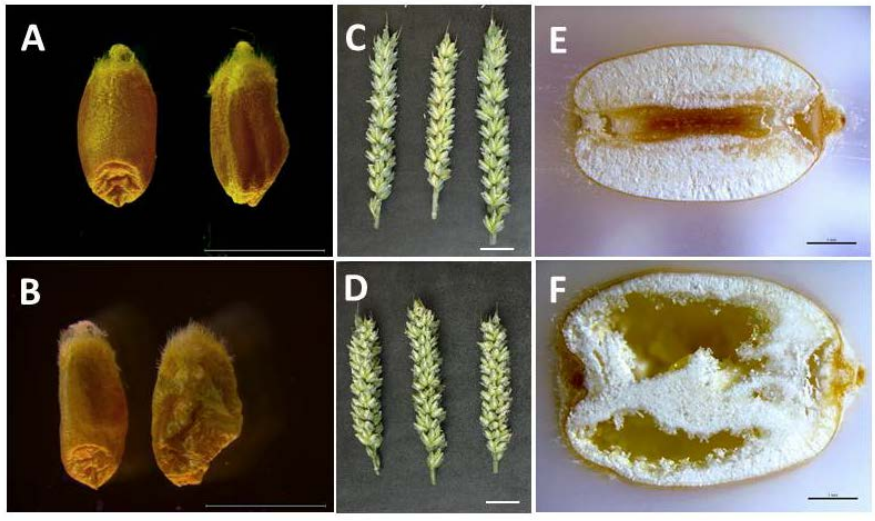


Figure 12. Images from light microscopy showing wheat grains from control (A), the wrinkled phenotype of the *AsWRI1*-line with 12-inserts (B), the ear of the control (C) versus the 12-insert line (D) at 26 DPA. E-F shows a longitudinal section of the grain of control (E) and 12-insert line (F) at 26 DPA. Dark area in the middle of the grain in E is the crease, the dark area in F is a cavity filled with an aqueous liquid before seed dessication occurs. Scale bars are 10 mm (A, B), 20 mm (C, D) and 1 mm (E, F). Figure taken from manuscript II.

aleurone cells. As the high-molecular-weight glutenin promoter used to control the expression of *AsWRI1* should not be active in the aleurone cells (Lamacchia *et al.*, 2001) the effect on the differentiation of aleurone cells should be as a result from the WRI1 expression in the starchy endosperm. Initial differentiation of the aleurone cell identity is known to occur as early as 6-8 DPA (Morrison *et al.*, 1975) while the promoter used in our study first is active 10-12 DPA (Lamacchia *et al.*, 2001). Auxin is known to be involved in aleurone establishment and has also been associated with WRI1 (Kong *et al.*, 2017). However, no transcripts associated with auxin synthesis or transport was identified as differentially expressed in the endosperm. As the ratio between different sugars have been shown to be important for other signalling purposes, it becomes the prime suspect for the observed multiplication of aleurone cells observed here although it is currently not possible to give a final answer to this question. A study covering the early part of grain development, i.e. spanning the time where cells are differentiated into aleurone cells, could yield more answers. Analysis of starch granule size revealed that the lines expressing AsWRI1 had generally smaller starch granules than the control with larger granules completely gone from the 12-insert line.

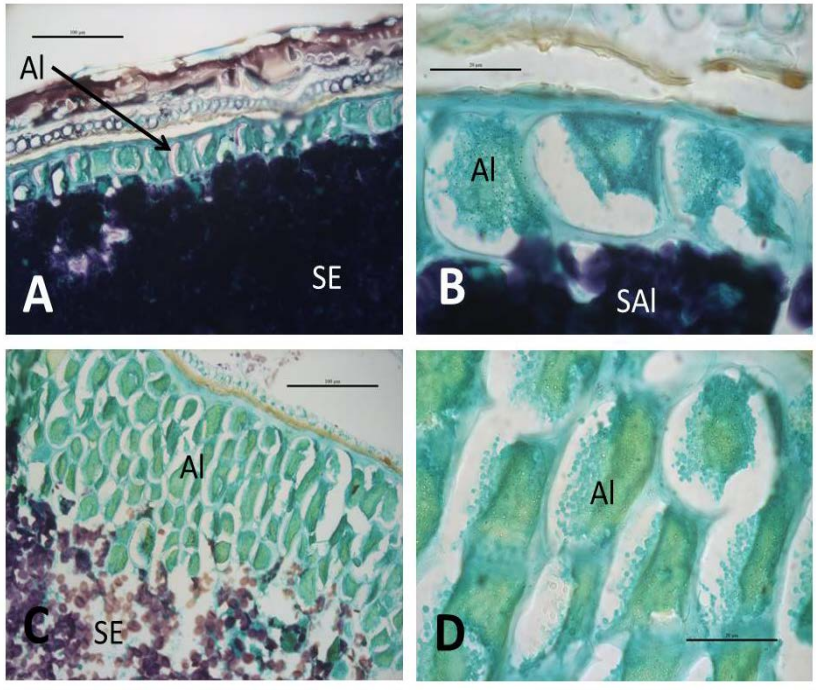


Figure 13. Light microscopy imaging of aleurone layer (Al), starchy endosperm (SE) and subaleurone layer (SAl) in control (A-B) and 12-insert line (C-D). Grain sections were stained with MAS staining proteins green and starch in dark purple. Scale bars 100 µm (A and C), 20 µm (B and D). Figure modified from manuscript II.

Despite these phenotypes, seeds maintained normal germination ability after desiccation but were less tolerant to sterilisation due to the fragile wrinkled seed coat from the collapsing of the central cavity.

4.2.1 WRINKLED1 influences wheat endosperm sink strength ambivalently

During seed development, the endosperm transitions from a weak to a strong sink with the initiation of starch accumulation. To better understand how endospermic expression of *AsWRI1* influenced the sink strength of the seed, a number of yield-related traits were measured on plants grown in climate chambers (Table 1). This revealed that the 12-insert line yield/plant (g) was 24 % lower than that of the control. Height, number of spikes, number of spikelets/spike and seeds/plant were unaffected, indicating strongly that the yield reduction is due to a reduction in seed size. Based on carbon equivalents calculated from sugar, oil and starch content of mature grains, the 12-insert line had a 45 % lower amount of carbon stored than the control. In an effort to

characterise the carbon flow dynamics, a pulse-chase experiment was carried out on developing ears. Similar experiments have been conducted on oat and wheat (Grimberg, 2014; Ekman *et al.*, 2008; Singh & Jenner, 1983). ^{14}C-labelled sucrose was fed to detached spikes, and the net accumulation of ^{14}C was measured up to 192 hours after the pulse. This revealed that there was no difference between the 12-insert line and the control in total sucrose net accumulation per grain, although the AsWRI1-line appeared to have a faster uptake as observed from the first sampled time point (0 hours), possibly indicating a stronger sink at this developmental stage. The most significant difference in ^{14}C accumulation was found in the oil fraction of the grains, where the 12-insert line accumulated 6-fold more than the control while the starch accumulation only was 50 % compared to the control. The water-soluble phase, containing, e.g. sucrose, of the 12-insert line contained 5-fold more ^{14}C than the control at 192 hours, indicating either the recycling of carbon into sucrose and other sugars or the inability of the developing seed to divert sucrose into starch synthesis.

Taken together, this shows that *AsWRI1* expressed in the wheat endosperm is able to create a stronger sink activity than what is present in the wild type. Despite this, sucrose can be observed being accumulated in high amounts inside the endosperm. This accumulation of sucrose is most likely the reason for the loss of sink strength as it inhibits the phloem unloading and the sink size. This sucrose can either be unutilized sucrose imported from the phloem or *de novo* synthesised sucrose stemming either from starch or fatty acid

Table 1. Table over yield-related measurements on control and 12-insert transformed wheat lines expressing *AsWRI1*. Plants grown in climate chambers. Statistical analysis was conducted using Student's T-test. SD = Standard deviation.

	Control		**12-insert line (AsWRI1)**		
	Average	**SD**	**Average**	**SD**	**P-value**
Height (cm)	84.1	8.88	80	3.26	0.071
Number of spikes	9	1.63	8.7	1.38	0.479
Spike length (mm)	107	11.43	90	4.02	<0.001
Number of spikelets/spike	23	0.97	23	0.61	0.823
Spike density (spikelets/mm)	0.22	0.02	0.26	0.01	<0.001
Seeds/plant	501.2	105.17	531.4	93.61	0.356
Yield/plant (g)	16.8	4.23	13.6	2.69	0.007
Seed weight (mg)	34	5.06	26	1.55	<0.001
Yield/spike (g)	1.9	0.37	1.6	0.27	0.006

degradation. Genes involved in starch degradation are upregulated in the mRNA-seq data but so are a few genes involved in β-oxidation. It is, therefore, possible that the increased inflow of fatty acids cannot be handled by the TAG-assembly enzymes causing a build-up of free fatty acids, which in high levels becomes toxic to cell membranes, which are shuttled to the peroxisome to be degraded and re-shuffled back into the central metabolism. Another potential explanation is that the ectopic TAG-accumulation stimulates TAG-breakdown, causing feedback of carbon into the glycolysis. However, as TAG-breakdown has been shown to be repressed by high levels of sucrose, it is less likely that this would be the case here. We know that WRI1 stimulates carbon flow into *de novo* fatty acid biosynthesis primarily in the lower part of glycolysis, i.e. far downstream of sucrose. Therefore, as degraded fatty acids are being fed back into the glycolysis as pyruvate, it would make more sense for these carbon moieties to be shuffled back to fatty acid synthesis rather than through gluconeogenesis back to sucrose. Starch breakdown, on the other hand, yields sucrose making this a more straightforward explanation to the source of the sucrose being accumulated.

4.2.2 Field trials reveal further decrease in seed weight

One of the primary goals of this study was to evaluate the potential use of WRI1 to, trough biotechnological methods, create novel oil accumulation in non-oil accumulating endosperm tissue. A similar study done on maize (*Zea mays)* in which maize *WRI1* was expressed in the maize endosperm failed to yield any substantial increase in oil (Shen *et al.*, 2010) indicating either that not all species are good targets for improving oil accumulation using WRI1 or that different homologs of WRI1 differs in their peripheral function. However, it is well known that results from tightly controlled lab experiments often translate poorly to real-life field trials. Therefore, with the purpose of evaluating the effects of field trials on traits previously measured in climate chamber experiments a small field trial of *AsWRI1*-wheat was conducted during the summer of 2019. The field trial was conducted at Borgeby, Lomma municipality, Sweden (DMS: 55°75'36.2"N, 13°05'36.9"E) by Hushållningssällskapet during the summer of 2019 with three plots (1.6 m x 2.5 m) for each of the evaluated lines (control and 12-insert line). Ten plants from each plot was harvested and threshed by hand to allow for more precise measurements.

Table 2. Table over yield-related measurements from field trials on control and 12-insert transformed wheat lines expressing *AsWRI1*. Statistical analysis was conducted using Student's T-test. SD = Standard deviation.

	Control		**12-insert line (AsWRI1)**		
	Average	**SD**	**Average**	**SD**	**P-value**
Height (cm)	56.3	6.62	45.8	2.51	<0.001
Number of spikes	1.2	0.48	1.4	0.61	0.170
Spike length (mm)	68.8	9.19	68.2	6.57	0.764
Number of spikelets/spike	15	1.29	16.1	1.51	0.004
Spike density (spikelets/mm)	0.22	0.02	0.24	0.02	0.004
Seeds/plant	35.7	20.58	47.1	20.71	0.039
Yield/plant (g)	1.15	0.55	0.85	0.34	0.014
Seed weight (mg)	34.2	6.51	18.5	3.75	<0.001
Yield/spike (g)	1.0	0.38	0.63	0.17	0.003

The field trial revealed that the 12-insert line generally behaved in a similar way in a field setting as it did in the Biotron when compared to the control line (Table 2). Total number of spikes, spike length and number of grains per plant didn't differ between the 12-insert line and the control in the field. This is similar to what was observed in the Biotron with the exception of spike length which was significantly reduced in the 12-insert line when grown in the Biotron but not in the field. This is likely due to differences in spikelet density which were more prominent in the Biotron than in the field. Interestingly, the trend towards shorter plant height observed in the Biotron (Table 2) became very noticeable as well as significant in the field setting. The reason for this is unknown but as the expression of *AsWRI1* is under control of a strict starchy endosperm promoter it is most likely a secondary effect of the changes in carbon flow and sink activity and not due to changes in the vegetative parts. One potential mechanism is that the increased sink activity arguably observed during the early stages of seed development outcompetes the weaker sink of the shoot apical meristem causing a reduced carbon flow to straw elongation. Despite that the total grain yield per plants was unaffected both in the Biotron and in the field, the total mass yield of the 12-insert line was reduced further in the field compared to the Biotron with a 24 % reduction in seed weight in Biotron compared to the 44 % reduction observed in the field.

4.3 Autoregulation of WRI1 expression (paper I)

As discussed earlier, and shown above, WRI1 is a prime target for biotechnological applications targeting oil accumulation in plant storage tissues. However, it is also evident that care needs to be taken when selecting promoters to use in these applications. During seed development in Arabidopsis, WRI1 has an unexpectedly small window of expression right at the start of seed filling despite known regulators of WRI1 expression being highly expressed well beyond that point. This shows that we are, despite WRI1 being a relatively well-studied gene, lacking critical aspects of the regulation of WRI1. The expression pattern of WRI1 fits that of genes which are under negative autoregulatory loops, which also has been proposed to be present for Arabidopsis WRI1 by Cernac and Benning (2004). With this background, we set up a transient expression system in *Nicotiana benthamiana* leaves where the activity of the Arabidopsis WRI1 promoter could be analysed using the GUS reporter gene. As expression of agroinfiltrated constructs has been shown to be well correlated with the OD_{600} of the *Agrobacterium tumefaciens* (at low OD_{600}) we could use this system to analyse the effect of WRI1 on the activity of the Arabidopsis WRI1 promoter.

4.3.1 WRI1 is suppressing its promoter activity through a negative feedback loop

In the initial experiments, we observed that the promoter of Arabidopsis WRI1 was active and able to initiate expression in *N. benthamiana* leaves. As the normal state of eukaryotic promoters is to be silent in the ground state, this was surprising. This could indicate that WRI1 is epigenetically silenced in vegetative tissue, perhaps through a similar mechanism as the LAFL-network. However, this active state of the promoter when expressed in leaves by agroinfiltration gave us a unique opportunity to study repressors of the WRI1 promoter. We were able to show that increased expression of Arabidopsis WRI1 led to decreased Arabidopsis WRI1 promoter activity, thereby demonstrating the presence of a repressive regulatory feedback loop. WRI1 homologs from different species are known to be well conserved and similar in function (Grimberg *et al.*, 2015). Based on this, we showed that homologs of WRI1 from oat, potato and nutsedge also were able to repress the activity of the Arabidopsis WRI1 promoter. Furthermore, these results showed that there was a significant difference between different homologs in the ability to affect the promoter activity with WRI1 from potato having the most severe effect and WRI1 from oat having the mildest effect. Truncating the Arabidopsis WRI1 promoter down to 250 bp showed no effect on the autoregulation, proving that

the cis-regulatory element involved is located in very close proximity to the translational start site.

4.3.2 Domain swapping reveals that the tandem AP2-domain is essential for WRI1 autoregulation

WRI1, being a member of the AP2-subfamily of transcription factors, can be divided into three main regions. Two central AP2-domains in tandem flanked by two more or less unstructured regions (called the N-terminus region and C-terminus region). Previous research has revealed that Arabidopsis WRI1 contains three predicted intrinsically disordered regions (IDRs) (Ma *et al.*, 2015b). Based on *in silico* sequence analyses using PONDR we identified that all WRI1 homologs had a very similar structure with conserved IDRs flanking the DNA-binding domain. To elucidate whether these flanking regions played a role in the negative autoregulation, we swapped the DNA-binding and the C-terminus domains between Arabidopsis and oat WRI1. To verify their functionality, they were agroinfiltrated into *N. benthamiana*, showing that they all were fully functional in inducing accumulation of oil. The lipid analysis also revealed that the difference in oil accumulation capacity observed between different species by Grimberg *et al.* (2015) appears to be heavily influenced by the DNA-binding region. This was also the case for the promoter activity assay where the permutations containing the DNA-binding region from oat were significantly worse at downregulating the promoter activity than those with the DNA-binding motif from Arabidopsis. Together, this is a proof of that the autoregulatory mechanism is profoundly relying on the sequence of the DNA-binding domain. With this proven to be the case, Arabidopsis WRI1 was purified from *E. coli*, and the binding capacity to the promoter region was evaluated using fluorescent electromobility shift assay (fEMSA) which showed that the autoregulatory mechanism observed is not due to direct binding of WRI1 to its own upstream region.

In conclusion, this data shows that WRI1 is under evolutionary conserved negative autoregulation. Although we cannot define precisely what kind of mechanism is behind it, we do know that it is not a direct interaction between WRI1 and its upstream region. Based on this, two likely hypotheses can be drawn. In the first, a downstream target of WRI1 is upregulated and in turn downregulates WRI1 expression. In the second hypotheses, WRI1 is able to interact with another transcription factor enabling it to bind the WRI1 promoter and silencing it. Despite this, the autoregulation observed here gives an explanatory model for the observed expression pattern of Arabidopsis WRI1

during seed development in which the relatively short-lived expression of WRI1 is a result of it downregulating its own expression.

4.4 Understanding the structure of the AP2-subfamily of transcription factors (manuscript V)

The previous observation that much of the regulatory specificity was bound to the DNA-binding domain of WRI1 and that it is possible, without serious negative consequences, to remodel these transcription factors raised a question on how the family to which WRI1 belongs to, the AP2-subfamily of the AP2/ERF-superfamily, is constructed. This subfamily contain many important developmental transcription factors such as APETALA2 (AP2) which is involved in floral patterning and floral meristem maintenance and BABY BOOM (BBM) which is known to be essential for embryogenesis and root patterning (Horstman *et al.*, 2017; Ripoll *et al.*, 2011; Galinha *et al.*, 2007; Jofuku *et al.*, 2005; Boutilier *et al.*, 2002; Drews *et al.*, 1991). Despite the importance of the members of this subfamily, relatively little work has been focused on trying to understand the mechanisms behind the specialisation of these transcription factors.

4.4.1 The AP2-subfamily can be split into three groups

Previous studies have been able to show, based primarily on sequence homology within the first AP2-domain, that the AP2-subfamily can be divided into two groups (Kim *et al.*, 2006). Based on additional sequence analysis as

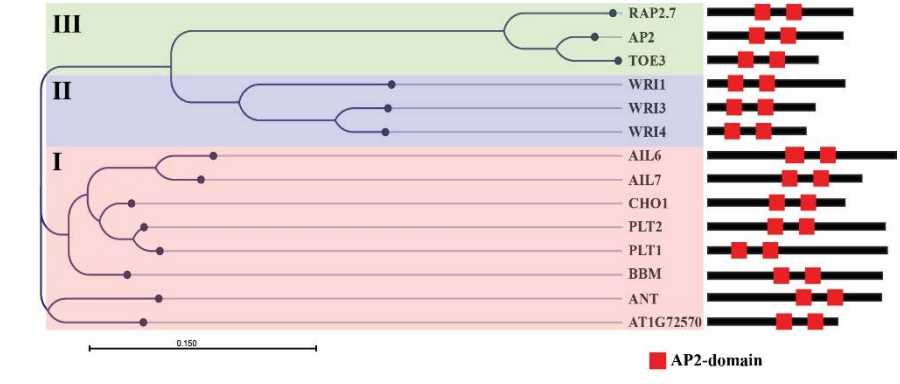


Figure 14. Unrooted phylogenetic tree based on sequence homology in the DNA-binding motif showing the three proposed groups of the 14 members of the AP2-subfamily transcription factors (left). Protein models are shown to scale to the right, where red boxes represent position of the tandem AP2-domains.

well as previously published functional data, we propose that the AP2-subfamily can be divided into three groups (Figure 14). Group I consist of eight members with BBM and AINTEGUMENTA (ANT) being the two most

prominent, while group II and III both have three members each. Structurally, group I tends to be longer with an average of 545 amino acids long compared to 365 and 416 of group II and III respectively. Group I also tends to have less distinct IDRs flanking the DNA-binding domain in addition to the tandem AP2-domains which have a tendency to be situated towards the C-terminus. Both group II and III have more easily identifiable flanking IDRs with the DNA-binding domain of group II and III being located towards the N-terminus and the centre, respectively. Group I appears to primarily be involved in general developmental processes such as root structuring, root and shoot meristem maintenance and cell proliferation. Meanwhile, group II and III appear to have more specialised roles with group II being involved in lipid biosynthesis and group III being involved in floral development and the transition to flowering. Furthermore, *in silico* analysis revealed that all AP2-subfamily members contain at least one predicted IDR flanking the DNA-binding region. Additionally, the 35 amino acid long linker between the two AP2-domains turn out to be exceptionally well conserved in length but not in sequence. However, there appears to be a strong tendency for the linker to have a high level of disorder, suggesting that it needs to be flexible to allow the two AP2-domains to bind.

4.4.2 The regulatory specificity of the AP2-subfamily is intrinsic to the DNA-binding domain

The C-terminus of Arabidopsis WRI1 has been shown to be important for the stability of the protein (Ma *et al.*, 2015b). Other than this, very little is known about the function of the flanking regions and the predicted IDRs. To evaluate this, Arabidopsis mutants *bbm-1*, *wri1-1* and *wri4-3* were analysed for seed oil content. This analysis confirmed previous studies showing that WRI1, but not WRI4, is essential for seed oil accumulation (To *et al.*, 2012) as well as that the *bbm-1* mutant has a small, but significant, decrease in seed oil content. Arabidopsis BBM, WRI1 and WRI4 were then ectopically expressed in *N. benthamiana* leaves showing that both WRI1 and WRI4, but not BBM, was able to induce oil accumulation in leaf tissue. Taken together, this strongly indicates that WRI1 and WRI4 induce oil accumulation through activation of *de novo* fatty acid synthesis while the reduction of oil in *bbm-1* is the result of disturbed embryo development. Following this, the DNA-binding region with

the tandem AP2 domains of BBM and WRI4 was isolated and used to replace the DNA-binding region of WRI1 creating two hybrid transcription factors (WRI1:BBM:WRI1 and WRI1:WRI4:WRI1). These as well as WRI1 were agroinfiltrated into *N. benthamiana* leaves, and the oil content analysed. This analysis showed that the WRI1 with the DNA-binding region from BBM failed to activate oil accumulation while WRI1 with the DNA-binding region from WRI4 had almost the same capacity as WRI1. This shows that the target specificity of these members of the AP2-subfamily is tightly linked to the DNA-binding specificity of the tandem AP2-domains and that the primary function of the flanking regions could be to regulate the activation level by interacting with general transcription factors.

4.5 LEC1 specifies the role of the AFL-subfamily during embryogenesis (manuscript III and IV)

The LAFL-network, named after the LEC1, ABI3, FUS3 and LEC2 transcription factors, is a network of master regulators governing seed development during late embryogenesis. Their functions have been well studied and are known to primarily be centred around the accumulation of protein and oil, with more peripheral roles in cotyledon differentiation and embryo de-greening (Jia *et al.*, 2013). Even though the vast knowledge about their individual function, few studies have been conducted in a manner that makes it easy to compare and differentiate the roles of the individual members. This becomes problematic when they are known to have overlapping as well as unique functions. Furthermore, very few studies have looked at the combinatorial effect of the NF-YB transcription factor LEC1 and the AFL-subfamily. In this work, we utilise ectopic gene expression in *N. benthamiana* leaves to achieve a comparable data set of all AFL-members with and without co-expression of LEC1 with the purpose of elucidating the role of the interaction within the LAFL-network.

4.5.1 Transcriptomic analysis reveals the unique roles of the AFL-subfamily

Comparative transcriptome analysis of the individual AFL-subfamily members yielded a set of 1312 differentially expressed transcripts (FDR <0.05) common to all members of the subfamily. Approximately half of these were upregulated (log2 FC >1). These 1312 differentially expressed transcripts form the frame of a core regulome for the AFL-subfamily, although it is important to remember that this is in a *N. benthamiana* leaf environment. The core

regulome was analysed for enriched biological process GO-terms. This analysis revealed that the AFL-subfamily share two main roles, metabolic processes and development of the epidermis. The metabolic processes appeared to be mainly focused towards lipid and protein precursor biosynthesis. Interestingly, despite the lack of phenotypes pointing to this based on mutant studies, also ABI3 regulates genes involved in the development of the epidermis. In general, these findings fit well with what has already been identified as the primary role of these transcription factors showing not only that the experimental system is functional but also that the AFL-subfamily is able to function within the leaf environment.

To identify the distinctive role of the subfamily members, differentially expressed transcripts unique to each transcription factor was selected and analysed for GO-term enrichment. This revealed that all individual AFL-subfamily members regulated unique sets of genes involved in metabolism and signalling pathways. ABI3 also regulated a broad set of unique genes involved in abscisic acid (ABA) response, late embryo development and metabolism. LEC2 appeared to be more focused towards protein precursor synthesis while FUS3 largely mimics the role of ABI3.

4.5.2 LEC1 and ABI3 activate key genes for embryo morphogenesis

While the expression of Arabidopsis LEC1 by itself had a relatively minor influence on the transcriptome, it had a significant influence on the transcriptomes of the AFL-subfamily members when co-expressed with them. While the number of differentially expressed transcripts in the transcriptome of FUS3 was reduced by 75 % when co-expressed with LEC1, only a few transcripts (70) were identified as differentially expressed when the LEC1+FUS3 infiltrations were compared to the individual infiltrations of LEC1 and FUS3. Since no clear pattern among these could be identified, they were discarded as background noise suggesting that LEC1 does not specify the actions of FUS3 and that the observed effect on the FUS3 transcriptome could be due to other effects such as competitive binding. On the other hand, when co-expressed with LEC2 and ABI3, LEC1 significantly affected the expression of 347 and 1090 genes, respectively. The majority of these genes turned out to be *de novo* expressed when LEC1 was added, indicating that LEC1 collaborates with LEC2 and ABI3 to activate specific genes. Nevertheless, around one-third of the genes showed a reversed state of transcription, going from downregulated to upregulated or vice versa. This is interesting as it reveals a mechanism through which LEC1 can control specific functions of

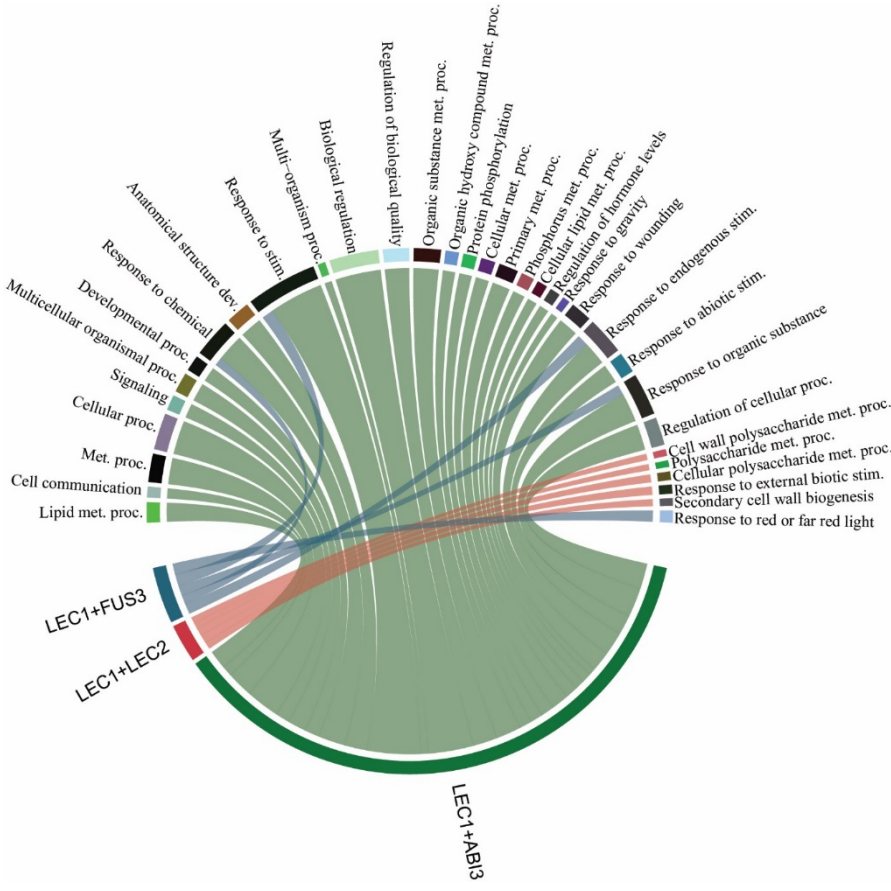


Figure 15. Chord diagram over significantly enriched (adjusted P-value <0.05) GO-terms for genes significantly affected by the co-expression of LEC1 with the individual members in the AFL-network of transcription factors. Size of chord is negatively correlated with P-value.

LEC2 and ABI3 during specific stages of embryo development. To better understand the biological role of this regulation, a GO-enrichment analysis was run on the identified sets of genes (Figure 15). This analysis revealed that LEC1 influences LEC2 and ABI3 in different ways. While LEC1 co-expressed with LEC2 targets genes involved in cell wall biosynthesis, LEC1 and ABI3 together target a broad set of developmental genes. Among these genes is a homolog to the Arabidopsis PEI1, a zinc finger transcription factor, essential for the transition between globular and heart stage which is only differentially expressed when LEC1 is co-expressed with ABI3. Another is a homolog to the Arabidopsis *ARGONAUTE10* (*AGO10*) which is involved in the bilateral patterning of the embryo through maintaining HD-ZIPIII transcription factor expression. Interestingly, LEC1+ABI3 also upregulated *NbFUS3* suggesting

that LEC1+ABI3 regulates developmental phases partially through FUS3. Our data also show that LEC2 and ABI3 have overlapping functions when it comes to activating late embryogenesis abundant (LEA) proteins. However, ABI3 appears to generally be more efficient in activating the expression of LEA genes. While the addition of LEC1 to ABI3 does not yield any apparent differences in the expression of LEA genes, the addition of LEC1 to LEC2 downregulates large parts of the group 1 and group 4 LEA genes. This suggests that LEC1 is essential for repressing certain aspects of the late-embryogenesis role of LEC2 during early embryogenesis. Surprisingly, FUS3 was unable to affect the expression of LEA genes, which indicates that FUS3 requires additional transcription factors to achieve this and that these are not present in the leaf environment. This suggests that FUS3 functions through at least two different mechanisms, one where it independently can induce the expression of developmental genes and one where it functions in a larger complex to activate seed storage genes. This could imply that FUS3 is involved in other aspects of plant development outside the seed environment and that this mechanism allows it to do so without activating genes specifically involved in late-embryogenesis.

4.5.3 LEC1 activates the expression of WRI1 through ABI3

The LAFL-network is known to regulate the accumulation of oil in the embryo during late embryogenesis. To better understand this process, leaves infiltrated with members of the LAFL-network were analysed for oil content showing that only ABI3 had a significant effect on total oil content. However, LEC2 and especially FUS3 both showed a tendency towards a small increase in increased oil content. Interestingly, LEC1, despite having no influence on oil accumulation itself, significantly boosted the ability of ABI3 to initiate oil accumulation. Previous studies have revealed that WRI1 is a target of LEC1, LEC2 and FUS3 thereby yielding a mechanism for how the LAFL-network regulates oil accumulation (Wang & Perry, 2013; Monke *et al.*, 2012; Yamamoto *et al.*, 2010; Baud *et al.*, 2007). Based on our transcriptome data, only ABI3 when co-expressed with LEC1 significantly influenced the expression of *NbWRI1* while LEC1+LEC2 and FUS3 showed a small but non-significant increase. To verify this data, the experiment was repeated, and RT-qPCR was used to analyse the expression of *NbWRI1* verifying the induction of *NbWRI1* by LEC1+ABI3 together, but not by ABI3 or LEC1 individually. Furthermore, it showed that FUS3 are in fact, able to significantly increase *NbWRI1* expression but not LEC2. To our surprise, pathway analysis of genes involved in glycolysis and *de novo* fatty acid synthesis showed that many of

the known targets of WRI1 were downregulated. With the purpose of explaining these results, a time study stretching over 14 days was conducted on leaves infiltrated with LEC1 and ABI3 which revealed that *NbWRI1* reaches its highest expression level already before 24 hours after the infiltration and stabilises between 3-6 days after infiltration. This was confirmed using two well-known target genes of WRI1, *NbKASI* and *NbBCCP2*, which both showed a similar although slightly lagging pattern of expression. This indicates that LEC1 is able to transiently activate WRI1 expression through ABI3. As we know that other targets of LEC1 and ABI3 have a much longer time-window of expression, this indicates that the leaf is able to either counteract the WRI1 expression or silence WRI1 through other means.

4.5.4 ABI3 is able to increase oil accumulation through a non-WRI1 dependent pathway

In this work we show that the ectopic expression of AtABI3 in *N. benthamiana* leaves is sufficient to significantly increase the accumulation of oil (Figure 16). ABI3 is also known to be important for fatty acid accumulation in Arabidopsis with the loss-of-function mutant having reduced total fatty acid content (Roscoe *et al.*, 2015). Furthermore, we show that the expression of *NbWRI1* is not influenced by AtABI3 by itself but requires the presence of AtLEC1. This raises the question of how ABI3 is able to induce oil accumulation in a WRI1-independent way. Since triacylglycerol is deposited in lipid droplets, one possible explanation is that ABI3 is primarily involved in earmarking these lipid droplets for storage by changing the composition of lipid droplet associated proteins. By conducting RNA-seq on infiltrated leaves we confirmed that ABI3 is able to induce a large set of lipid droplet associated proteins such as *SEIPIN1-like* genes. Arabidopsis SEIPIN1 is known to increase oil accumulation by influencing lipid droplet dynamics (Cai *et al.*, 2015) indicating that ABI3, through the activation of genes coding for a multitude of lipid droplet associated proteins, could play a similar role. This role appears to be positively influenced by co-infiltration with AtLEC1. This is further supported by the observation that the co-infiltration of AtABI3 and AtWRI1 is able to induce higher oil accumulation than would be expected by simple additive effects. Furthermore, when co-infiltrated with AtLEC1, AtABI3 is able to induce *NbWRI1* creating a synergistic effect where NbWRI1 increase *de novo* fatty acid biosynthesis while ABI3 (partially together with AtLEC1) increase the storage stability of the formed lipid droplets. These observations fit with the suggested increased β-oxidation observed in the wheat transformed with AsWRI1. Since the wheat endosperm is not programmed to

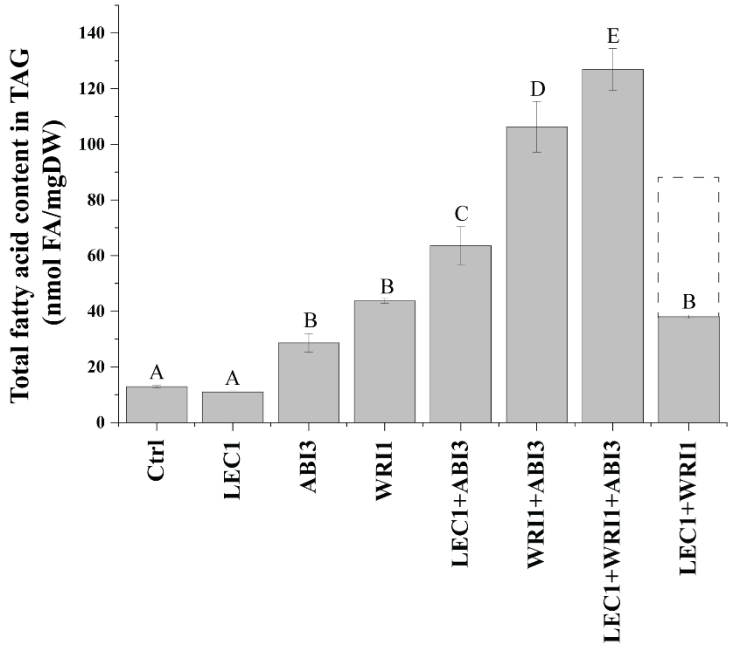


Figure 16. Triacylglycerol (TAG) total fatty acid (FA) content from infiltrated leaves of *N. benthamiana* using Arabidopsis transcription factors. Bars represent the average from three biological replicates. Error bars = SD. Letters indicate significantly distinct groups based on Tukey´s test with a cutoff at $P \leq 0.05$. Dashed box represents the theoretical additive effect of LEC1+ABI3 on oil accumulation to the *AtWRI1* infiltration.

accumulate oil many of the lipid droplet associated proteins is missing potentially resulting in active or passive recycling of newly formed triacylglycerol. It is not unthinkable that co-expression of LEC1 and ABI3 would facilitate in protecting the lipid droplets from degradation therefore increasing the oil accumulation potential of the endosperm. Finally, it is interesting to note that the loss-of-function WRI1 mutant identified in Arabidopsis has a 80% reduction of seed oil content (Focks & Benning, 1998). If we make the assumption that one of the main roles of LEC1 is to activate the expression of WRI1 through ABI3 we can compare the oil accumulation between the AtLEC1+AtABI3+AtWRI1 infiltration with that of only AtABI3 to see how much influence the removal of WRI1-expression has. In our system, the removal of WRI1 yields an 82% reduction of total oil accumulation which is remarkably similar to the 80% reduction observed in the original Arabidopsis mutant. This is of course only an intriguing observation based on a simplistic theoretical model of seed oil accumulation but forms an interesting

base for further investigation into the role of these three transcription factors during the transition to seed filling in Arabidopsis.

4.5.5 The A-domain of ABI3 is essential for collaboration with LEC1

ABI3 is known to consist of four conserved regions, the A, B1, B2 and B3 domains (McCarty *et al.*, 1991). In LEC2, which only contains the B2 and B3 domains, LEC1 has been shown to interact with LEC2 through the B2-domain (Boulard *et al.*, 2018). To test the importance of the different domains present in ABI3 on the observed effect of LEC1 truncated ABI3s was infiltrated into leaves, with and without LEC1, and the result analysed with RT-qPCR. This revealed that the A-domain of ABI3 is vital for the LEC1-dependent activation of the developmental genes *NbPEI1* and *NbYUC10*. The seed storage protein gene *NbCRA1* was also used as a non-developmental related marker gene showing that the B2 and B3 domains are enough to activate its expression, although at a significantly lower level than the native ABI3. Interestingly, the addition of LEC1 to these truncated ABI3 completely extinguished their ability to activate *NbCRA1* indicating that LEC1 either outcompetes the ability of ABI3 to bind to the target promoter or inhibits ABI3 ability to initiate transcription. One potential explanation could be that LEC1 interacts with the B2-domain, as observed in LEC2, but requires the A-domain to recruit the polymerase causing a non-functional complex to block the promoter. This observation also yields a clue to another potential function of LEC1 in collaboration with ABI3. As it is highly unlikely that all truncated ABI3 and LEC1 moieties are able to form a complex, there should be a basal activity of the *NbCRA1* promoter caused by the sporadic binding of truncated ABI3. The lack of this would then indicate that the hypothetical complex between LEC1 and ABI3 either is assembled on the promoter itself or that LEC1 influences the binding constant of ABI3 when binding to it making it much more stable in its interaction with the promoter.

5 Conclusions

In this work we have shown that (i) WRI1 is negatively regulating its own expression; (ii) LEC1 influences the regulomes of LEC2 and ABI3; (iii) LEC1 and ABI3 together regulate the expression of several key genes in embryo development, (v) LEC1 regulates WRI1 expression through ABI3 and (vi) ABI3 and WRI1 synergistically regulate oil accumulation through two different pathways. Based on these observations, we can build a model where LEC1 plays an essential role in regulating the functions of LEC2 and ABI3 during early embryogenesis and the transition to embryo maturation. In this model, LEC1, likely through the NF-Y complex, regulates genes involved in early morphogenesis such as the establishment and maintenance of the suspensor cell identities. At the same time, LEC1 represses the late-embryogenesis role of LEC2 while allowing it to function towards cell wall biosynthesis during the globular stage. At the transition between globular and heart stage, ABI3 has reached high enough levels to start impacting the development. LEC1 can then, through ABI3, initiate genes such a *PEI1*, *YUC10*, *AGO10*, *CLV2* and *SEU* which are essential for the establishment of the bilateral symmetry ending in the late heart stage. Some of these functions are likely a result of the activation of *FUS3*. At the same time, LEC1, again through ABI3, activates *WRI1* expression, thereby drastically changing the sink activity and kick-starting the oil accumulation. We know that the WRI1 expression during Arabidopsis seed development is transient in nature, similar to what we can observe by the activation of WRI1 in *N. benthamiana* leaves. This is possibly, perhaps even likely, to be due to the negative autoregulation of WRI1 indicating that a similar process does take place in the seed environment as well. As ABI3 levels rise even further and LEC1 levels decrease, ABI3 gradually becomes more and more focused on targeting the accumulation of seed storage compounds. Also, *LEC2* is gradually released from LEC1 repression, allowing it to fully function during seed maturation. As it has been

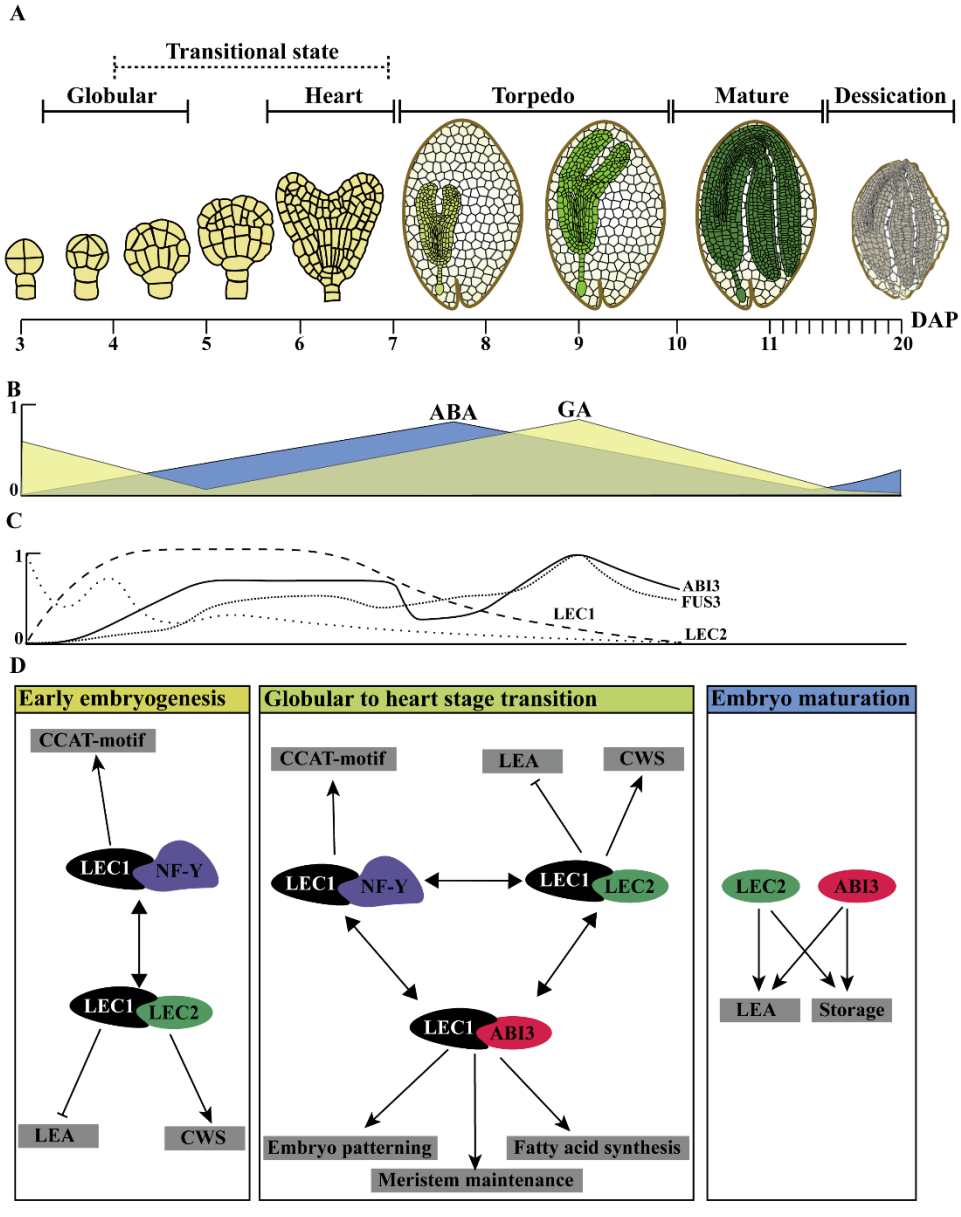


Figure 17. Proposed model on the function of LAFL-network during seed development in Arabidopsis. Developmental stages of seed and embryo development in Arabidopsis with days after pollination (DAP) marked (A). Gibberellins (GA) and abscisic acid (ABA) levels (B) and schematic gene expression profiles of the LAFL-network of transcription factors (C) in Arabidopsis seeds during development. Seed structures are not drawn to scale. ABA and GA profiles based on Okamoto *et al.* (2006) and Hu *et al.* (2018) respectively. Relative expression profiles of LAFL-network members is based on transcriptomic analysis made by Schmid *et al.* (2005) and Hofmann *et al.* (2019).

Figure 18 (cont.). D) Proposed model of how the LAFL-network regulate different stages of embryo development in collaboration with ABA and GA. During early embryogenesis with high levels of GA and low levels of ABA, LEC1 functions through the NF-Y complex as well as in a complex with LEC2 to regulate cell division, suspensor identity and suppressing late embryogenesis related genes targeted by LEC2. As ABI3 increases with increased ABA levels LEC1 cooperates with ABI3 to drive the transition from globular to heart stage. Towards late stages of embryo morphogenesis, LEC1 expression goes down allowing the AFL-network to fully function in the acquisition of desiccation tolerance and storage compound accumulation.

shown that both ABI3 and FUS3 are capable of positive autoregulation, it is likely that once this process has been started, it is difficult, if not impossible, to stop. However, at least the autoregulation of ABI3 appears to be dependent on the environment as it could not be detected in this work based on the leaf environment. This model does not only add novel functionality to the LAFL-network but also suggests a mechanism for the interaction of gibberellic acids and abscisic acid during seed and embryo development.

Furthermore, in this work we also have shown that WRI1 is an interesting target for plant biotechnological applications of cereals. An oat endosperm WRI1 was able to increase the sink activity and drastically increase the oil accumulation when expressed in wheat endosperm. A fascinating effect was also revealed on the specification process of the endosperm since the overexpression of WRI1 influences the establishment of both the starchy endosperm and the aleurone layer. It also reveals that increased oil accumulation is likely not the only aspect that differs between oil and non-oil accumulating cereals. It also shows that accurate spatial and temporal control of genes is essential when using transcription factors for biotechnological purposes. Finally, we show that the AP2-subfamily of transcription factors, to which WRI1 belongs to, can be divided into three distinct groups and that their specificity is primarily a result of sequence variation within their DNA-binding region.

6 Future work

This work raises many critical questions for future work. The model proposed above requires a lot of conformational and challenging work to verify and complement. Some questions to be answered are:

- Do LEC2 and ABI3 directly interact with LEC1 *in vivo*?
- What role does the B2-region of the AFL-subfamily members play in the cooperation with LEC1?
- Through which mechanism is ABI3 able to initiate the accumulation of oil without WRI1?
- How does the regulatory complex surrounding FUS3 look like during the transcriptional initiation of genes involved in storage compounds accumulation in seeds?
- What are the roles of the IDRs present in the AP2-subfamily?
- Through which mechanism does the autoregulation of WRI1 function?
- How can WRI1 overexpression influence aleurone identity of endosperm cells in cereal grains?
- Is it possible for us to better understand the LAFL-network by creating even higher-resolution spatial maps during different stages of development?

In addition to these questions, this work also points out several major problems left to solve. E.g. what is the precise role of LEC2 during embryogenesis? We know it to be an evolutionarily conserved transcription factor but it is becoming more and more evident that it is not present in the Solanaceae family. How does the Solanaceae family seed development differ from that of other plant families based on this? This work also raises the question of how internal sink dynamics work and how we can mimic gene regulatory networks present in other species and tissues to avoid adverse effects. It also shows the importance of proper promoter screening, and perhaps the development of

synthetic promoters, for the efficient use of transcription factors in biotechnology. Finally, this work also raises the critical question about the spatial and temporal differences of gene regulatory networks and that all studies must take into consideration in which tissue type they are conducted.

References

Aichinger, E., Villar, C.B.R., Di Mambro, R., Sabatini, S. & Kohler, C. (2011). The CHD3 Chromatin Remodeler PICKLE and Polycomb Group Proteins Antagonistically Regulate Meristem Activity in the Arabidopsis Root. *Plant Cell*, 23(3), pp. 1047-1060.

Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., Nussaume, L., Noh, Y.S., Amasino, R. & Scheres, B. (2004). The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. *Cell*, 119(1), pp. 109-20.

Allen, B.L. & Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology*, 16, p. 155.

An, D. & Suh, M.C. (2015). Overexpression of Arabidopsis WRI1 enhanced seed mass and storage oil content in Camelina sativa. *Plant Biotechnology Reports*, 9(3), pp. 137-148.

Aoyama, T., Hiwatashi, Y., Shigyo, M., Kofuji, R., Kubo, M., Ito, M. & Hasebe, M. (2012). AP2-type transcription factors determine stem cell identity in the moss Physcomitrella patens. *Development*, 139(17), pp. 3120-3129.

Balandin, M., Royo, J., Gomez, E., Muniz, L., Molina, A. & Hueros, G. (2005). A protective role for the embryo surrounding region of the maize endosperm, as evidenced by the characterisation of ZmESR-6, a defensin gene specifically expressed in this region. *Plant Molecular Biology*, 58(2), pp. 269-282.

Ball, S.G. & Morell, M.K. (2003). From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu Rev Plant Biol*, 54, pp. 207-33.

Banas, A., Debski, H., Banas, W., Heneen, W.K., Dahlqvist, A., Bafor, M., Gummeson, P.O., Marttila, S., Ekman, A., Carlsson, A.S. & Stymne, S. (2007). Lipids in grain tissues of oat (Avena sativa): differences in content, time of deposition, and fatty acid composition. *Journal of Experimental Botany*, 58(10), pp. 2463-2470.

Baud, S., Boutin, J.-P., Miquel, M., Lepiniec, L. & Rochat, C. (2002a). An integrated overview of seed development in Arabidopsis thaliana ecotype WS. *Plant Physiology and Biochemistry*, 40(2), pp. 151-160.

Baud, S., Boutin, J.P., Miquel, M., Lepiniec, L. & Rochat, C. (2002b). An integrated overview of seed development in Arabidopsis thaliana ecotype WS. *Plant Physiology and Biochemistry*, 40(2), pp. 151-160.

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, p. e0113.

Baud, S., Mendoza, M.S., 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 Journal*, 50(5), pp. 825-38.

Beaudoin, F., Wilkinson, B.M., Stirling, C.J. & Napier, J.A. (2000). In vivo targeting of a sunflower oil body protein in yeast secretory (sec) mutants. *Plant Journal*, 23(2), pp. 159-170.

Beckles, D.M., Smith, A.M. & ap Rees, T. (2001). A cytosolic ADP-glucose pyrophosphorylase is a feature of graminaceous endosperms, but not of other starch-storing organs. *Plant Physiology*, 125(2), pp. 818-827.

Bies-Etheve, N., Conceicao, A.D., 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 Molecular Biology*, 40(6), pp. 1045-1054.
Bokor, M., Csizmok, V., Kovacs, D., Banki, P., Friedrich, P., Tompa, P. & Tompa, K. (2005). NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. *Biophysical Journal*, 88(3), pp. 2030-2037.
Boulard, C., Thevenin, J., Tranquet, O., Laporte, V., Lepiniec, L. & Dubreucq, B. (2018). LEC1 (NF-YB9) directly interacts with LEC2 to control gene expression in seed. *Biochim Biophys Acta Gene Regul Mech*, 1861(5), pp. 443-450.
Boutilier, K., Offringa, R., Sharma, V.K., Kieft, H., Ouellet, T., Zhang, L.M., Hattori, J., Liu, C.M., van Lammeren, A.A.M., Miki, B.L.A., Custers, J.B.M. & Campagne, M.M.V. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell*, 14(8), pp. 1737-1749.
Bouyer, D., Roudier, F., Heese, M., Andersen, E.D., Gey, D., Nowack, M.K., Goodrich, J., Renou, J.P., Grini, P.E., Colot, V. & Schnittger, A. (2011). Polycomb Repressive Complex 2 Controls the Embryo-to-Seedling Phase Transition. *Plos Genetics*, 7(3).
Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitz, E.M. & Smyth, D.R. (1993). Control of Flower Development in Arabidopsis-Thaliana by Apetala1 and Interacting Genes. *Development*, 119(3), pp. 721-743.
Bravo, L.A., Gallardo, J., Navarrete, A., Olave, N., Martinez, J., Alberdi, M., Close, T.J. & Corcuera, L.J. (2003). Cryoprotective activity of a cold-induced dehydrin purified from barley. *Physiologia Plantarum*, 118(2), pp. 262-269.
Braybrook, S.A. & Harada, J.J. (2008). LECs go crazy in embryo development. *Trends in Plant Science*, 13(12), pp. 624-30.
Braybrook, S.A., Stone, S.L., Park, S., Bui, A.Q., Le, B.H., Fischer, R.L., Goldberg, R.B. & Harada, J.J. (2006). Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 103(9), pp. 3468-3473.
Bu, Z.M. & Callaway, D.J.E. (2011). Proteins Move! Protein Dynamics and Long-Range Allostery in Cell Signaling. *Advances in Protein Chemistry and Structural Biology: Protein Structure and Diseases, Vol 83*, 83, pp. 163-221.
Buleon, A., Colonna, P., Planchot, V. & Ball, S. (1998). Starch granules: structure and biosynthesis. *Int J Biol Macromol*, 23(2), pp. 85-112.
Buttrose, M.S. (1963a). Ultrastructure of Developing Aleurone Cells of Wheat Grain. *Australian Journal of Biological Sciences*, 16(4), pp. 768-&.
Buttrose, M.S. (1963b). Ultrastructure of Developing Wheat Endosperm. *Australian Journal of Biological Sciences*, 16(2), pp. 305-&.
Cai, Y.Q., Goodman, J.M., Pyc, M., Mullen, R.T., Dyer, J.M. & Chapman, K.D. (2015). Arabidopsis SEIPIN Proteins Modulate Triacylglycerol Accumulation and Influence Lipid Droplet Proliferation. *Plant Cell*, 27(9), pp. 2616-2636.
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), p. e42902.
Carbonero, P., Iglesias-Fernandez, R. & Vicente-Carbajosa, J. (2017). The AFL subfamily of B3 transcription factors: evolution and function in angiosperm seeds. *J Exp Bot*, 68(4), pp. 871-880.
Carpaneto, A., Geiger, D., Bamberg, E., Sauer, N., Fromm, J. & Hedrich, R. (2005). Phloem-localized, proton-coupled sucrose carrier ZmSUT1 mediates sucrose efflux under the control of the sucrose gradient and the proton motive force. *J Biol Chem*, 280(22), pp. 21437-43.
Cernac, A. & Benning, C. (2004). WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *Plant Journal*, 40(4), pp. 575-85.
Chen, L., Zheng, Y., Dong, Z., Meng, F., Sun, X., Fan, X., Zhang, Y., Wang, M. & Wang, S. (2018a). Soybean (Glycine max) WRINKLED1 transcription factor, GmWRI1a, positively regulates seed oil accumulation. *Mol Genet Genomics*, 293(2), pp. 401-415.
Chen, N.C., Veerappan, V., Abdelmageed, H., Kang, M. & Allen, R.D. (2018b). HSI2/VAL1 Silences AGL15 to Regulate the Developmental Transition from Seed Maturation to Vegetative Growth in Arabidopsis. *Plant Cell*, 30(3), pp. 600-619.
Chhun, T., Chong, S.Y., Park, B.S., Wong, E.C.C., Yin, J.L., Kim, M. & Chua, N.H. (2016). HSI2 Repressor Recruits MED13 and HDA6 to Down-Regulate Seed Maturation Gene Expression

Directly During Arabidopsis Early Seedling Growth. *Plant and Cell Physiology*, 57(8), pp. 1689-1706.
Choi, S.B., Wang, C.L., Muench, D.G., Ozawa, K., Franceschi, V.R., Wu, Y.J. & Okita, T.W. (2000). Messenger RNA targeting of rice seed storage proteins to specific ER subdomains. *Nature*, 407(6805), pp. 765-767.
Clark, J.K. & Sheridan, W.F. (1991). Isolation and Characterization of 51 Embryo-Specific Mutations of Maize. *Plant Cell*, 3(9), pp. 935-951.
Cui, S.K., Hayashi, Y., Otomo, M., Mano, S., Oikawa, K., Hayashi, M. & Nishimura, M. (2016). Sucrose Production Mediated by Lipid Metabolism Suppresses the Physical Interaction of Peroxisomes and Oil Bodies during Germination of Arabidopsis thaliana. *Journal of Biological Chemistry*, 291(38), pp. 19734-19745.
Denyer, K., Johnson, P., Zeeman, S. & Smith, A.M. (2001). The control of amylose synthesis. *Journal of Plant Physiology*, 158(4), pp. 479-487.
Devic, M. & Roscoe, T. (2016). Seed maturation: Simplification of control networks in plants. *Plant Sci*, 252, pp. 335-346.
Drews, G.N., Bowman, J.L. & Meyerowitz, E.M. (1991). Negative Regulation of the Arabidopsis Homeotic Gene Agamous by the Apetala2 Product. *Cell*, 65(6), pp. 991-1002.
Eastmond, P.J. (2006). SUGAR-DEPENDENT1 encodes a patatin domain triacylglycerol lipase that initiates storage oil breakdown in germinating Arabidopsis seeds. *Plant Cell*, 18(3), pp. 665-675.
Ekman, A., Hayden, D.M., Dehesh, K., Bulow, L. & Stymne, S. (2008). Carbon partitioning between oil and carbohydrates in developing oat (Avena sativa L.) seeds. *Journal of Experimental Botany*, 59(15), pp. 4247-4257.
Ekman, A., Hayden, D.M., Dehesh, K., Bulow, L. & Stymne, S. (2009). Carbon partitioning between oil and carbohydrates in developing oat (Avena sativa L.) seeds (vol 59, pg 4247, 2008). *Journal of Experimental Botany*, 60(1), pp. 350-350.
Evers, T. & Millar, S. (2002). Cereal Grain Structure and Development: Some Implications for Quality. *Journal of Cereal Science*, 36(3), pp. 261-284.
Ezcurra, I., Wycliffe, P., Nehlin, L., Ellerstrom, 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 Journal*, 24(1), pp. 57-66.
FAOSTAT (2011). Food and agriculture data. *Rome: Food and Agriculture Organization of the United Nations.*
Floyd, S.K. & Friedman, W.E. (2001). Developmental evolution of endosperm in basal angiosperms: evidence from Amborella (Amborellaceae), Nuphar (Nymphaeaceae), and Illicium (Illiciaceae). *Plant Systematics and Evolution*, 228(3), pp. 153-169.
Focks, N. & Benning, C. (1998). wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiology*, 118(1), pp. 91-101.
Frankel, A.D. & Kim, P.S. (1991). Modular structure of transcription factors: implications for gene regulation. *Cell*, 65(5), pp. 717-9.
Frey, A., Godin, B., Bonnet, M., Sotta, B. & Marion-Poll, A. (2004). Maternal synthesis of abscisic acid controls seed development and yield in Nicotiana plumbaginifolia. *Planta*, 218(6), pp. 958-64.
Friedman, W.E. (1998). The evolution of double fertilization and endosperm: an "historical" perspective. *Sexual Plant Reproduction*, 11(1), pp. 6-16.
Frontini, M., Imbriano, C., Manni, I. & Mantovani, R. (2004). Cell cycle regulation of NF-YC nuclear localization. *Cell Cycle*, 3(2), pp. 217-22.
Fuxreiter, M., Tompa, P. & Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. *Bioinformatics*, 23(8), pp. 950-956.
Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R. & Scheres, B. (2007). PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. *Nature*, 449(7165), pp. 1053-1057.
Garay-Arroyo, A., Colmenero-Flores, J.M., Garciarrubio, A. & Covarrubias, A.A. (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *Journal of Biological Chemistry*, 275(8), pp. 5668-5674.

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), pp. 373-85.
Geigenberger, P. (2011). Regulation of starch biosynthesis in response to a fluctuating environment. *Plant Physiology*, 155(4), pp. 1566-77.
Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F. & Goodman, H.M. (1992). Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell*, 4(10), pp. 1251-61.
Grimberg, A. (2014). Preferred carbon precursors for lipid labelling in the heterotrophic endosperm of developing oat (Avena sativa L.) grains. *Plant Physiology and Biochemistry*, 83, pp. 346-355.
Grimberg, A., Carlsson, A.S., Marttila, S., Bhalerao, R. & Hofvander, P. (2015). Transcriptional transitions in Nicotiana benthamiana leaves upon induction of oil synthesis by WRINKLED1 homologs from diverse species and tissues. *BMC Plant Biol*, 15, p. 192.
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. *Molecular Plant*, 5(4), pp. 876-88.
Han, J.D., Li, X., Jiang, C.K., Wong, G.K.S., Rothfels, C.J. & Rao, G.Y. (2017). Evolutionary Analysis of the LAFL Genes Involved in the Land Plant Seed Maturation Program. *Frontiers in Plant Science*, 8.
Hara, M., Terashima, S. & Kuboi, T. (2001). Characterization and cryoprotective activity of cold-responsive dehydrin from Citrus unshiu. *Journal of Plant Physiology*, 158(10), pp. 1333-1339.
Hargin, K.D. & Morrison, W.R. (1980a). The Distribution of Acyl Lipids in the Germ, Aleurone, Starch and Non-Starch Endosperm of 4 Wheat-Varieties. *Journal of the Science of Food and Agriculture*, 31(9), pp. 877-888.
Hargin, K.D. & Morrison, W.R. (1980b). The distribution of acyl lipids in the germ, aleurone, starch and non-starch endosperm of four wheat varieties. *Journal of the Science of Food and Agriculture*, 31(9), pp. 877-888.
Hayden, D.M., Rolletschek, H., Borisjuk, L., Corwin, J., Kliebenstein, D.J., Grimberg, A., Stymne, S. & Dehesh, K. (2011). Cofactome analyses reveal enhanced flux of carbon into oil for potential biofuel production. *Plant Journal*, 67(6), pp. 1018-1028.
Hibara, K., Karim, M.R., Takada, S., Taoka, K., Furutani, M., Aida, M. & Tasaka, M. (2006). Arabidopsis CUP-SHAPED COTYLEDON3 regulates postembryonic shoot meristem and organ boundary formation. *Plant Cell*, 18(11), pp. 2946-57.
Hill, A., Nantel, A., Rock, C.D. & Quatrano, R.S. (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. *Journal of Biological Chemistry*, 271(7), pp. 3366-3374.
Hoecker, U., Vasil, I.K. & McCarty, D.R. (1995). Integrated control of seed maturation and germination programs by activator and repressor functions of Viviparous-1 of maize. *Genes Dev*, 9(20), pp. 2459-69.
Hofmann, F., Schon, M.A. & Nodine, M.D. (2019). The embryonic transcriptome of Arabidopsis thaliana. *Plant Reprod*, 32(1), pp. 77-91.
Horstman, A., Li, M.F., Heidmann, I., Weemen, M., Chen, B.J., Muino, J.M., Angenent, G.C. & Boutilier, K. (2017). The BABY BOOM Transcription Factor Activates the LEC1-ABI3-FUS3-LEC2 Network to Induce Somatic Embryogenesis. *Plant Physiology*, 175(2), pp. 848-857.
Hu, Y.L., Zhou, L.M., Huang, M.K., He, X.M., Yang, Y.H., Liu, X., Li, Y.G. & Hou, X.L. (2018). Gibberellins play an essential role in late embryogenesis of Arabidopsis. *Nature Plants*, 4(5), pp. 289-298.
Huala, E. & Sussex, I.M. (1992). LEAFY Interacts with Floral Homeotic Genes to Regulate Arabidopsis Floral Development. *Plant Cell*, 4(8), pp. 901-913.
Huang, A.H.C. (2018). Plant Lipid Droplets and Their Associated Proteins: Potential for Rapid Advances. *Plant Physiology*, 176(3), pp. 1894-1918.
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), pp. 3099-111.
Hundertmark, M. & Hincha, D.K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. *BMC Genomics*, 9, p. 118.

Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z. & Dunker, A.K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research*, 32(3), pp. 1037-1049.

Ivarson, E., Leiva-Eriksson, N., Ahlman, A., Kanagarajan, S., Bulow, L. & Zhu, L.H. (2017). Effects of Overexpression of WRI1 and Hemoglobin Genes on the Seed Oil Content of Lepidium campestre. *Frontiers in Plant Science*, 7.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.F., Weissenbach, J., Quetier, F., Wincker, P. & Public, F.-I. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449(7161), pp. 463-U5.

Jenik, P.D., Gillmor, C.S. & Lukowitz, W. (2007). Embryonic patterning in Arabidopsis thaliana. *Annu Rev Cell Dev Biol*, 23, pp. 207-36.

Jensen, S.A. & Heltved, F. (1982). Visualization of enzyme activity in germinating cereal seeds using a lipase sensitive fluorochrome. *Carlsberg Research Communications*, 47(5), p. 297.

Jia, H.Y., McCarty, D.R. & Suzuki, M. (2013). Distinct Roles of LAFL Network Genes in Promoting the Embryonic Seedling Fate in the Absence of VAL Repression. *Plant Physiology*, 163(3), pp. 1293-1305.

Jia, H.Y., Suzuki, M. & McCarty, D.R. (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), pp. 135-145.

Jing, Y.J., Zhang, D., Wang, X., Tang, W.J., Wang, W.Q., Huai, J.L., Xu, G., Chen, D.Q., Li, Y.L. & Lin, R.C. (2013). Arabidopsis Chromatin Remodeling Factor PICKLE Interacts with Transcription Factor HY5 to Regulate Hypocotyl Cell Elongation. *Plant Cell*, 25(1), pp. 242-256.

Jofuku, K.D., Denboer, B.G.W., Vanmontagu, M. & Okamuro, J.K. (1994). Control of Arabidopsis Flower and Seed Development by the Homeotic Gene Apetala2. *Plant Cell*, 6(9), pp. 1211-1225.

Jofuku, K.D., Omidyar, P.K., Gee, Z. & Okamuro, J.K. (2005). Control of seed mass and seed yield by the floral homeotic gene APETALA2. *Proceedings of the National Academy of Sciences of the United States of America*, 102(8), pp. 3117-3122.

Jones, R.L. (1969). The fine structure of barley aleurone cells. *Planta*, 85(4), pp. 359-75.

Kagaya, Y., Ohmiya, K. & Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Research*, 27(2), pp. 470-478.

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A. & Hattori, T. (2005). LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant and Cell Physiology*, 46(3), pp. 399-406.

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. *Mol Cell Biol*, 25(13), pp. 5339-54.

Kao, C.Y., Cocciolone, S.M., Vasil, I.K. & McCarty, D.R. (1996). Localization and interaction of the cis-acting elements for abscisic acid, VIVIPAROUS1, and light activation of the C1 gene of maize. *Plant Cell*, 8(7), pp. 1171-1179.

Kaup, M.T., Froese, C.D. & Thompson, J.E. (2002). A role for diacylglycerol acyltransferase during leaf senescence. *Plant Physiology*, 129(4), pp. 1616-26.

Keith, K., Kraml, M., Dengler, N.G. & Mccourt, P. (1994). Fusca3 - a Heterochronic Mutation Affecting Late Embryo Development in Arabidopsis. *Plant Cell*, 6(5), pp. 589-600.

Kim, S., Soltis, P.S., Wall, K. & Soltis, D.E. (2006). Phylogeny and domain evolution in the APETALA2-like gene family. *Mol Biol Evol*, 23(1), pp. 107-20.

Kong, Q., Ma, W., Yang, H., Ma, G., Mantyla, J.J. & Benning, C. (2017). The Arabidopsis WRINKLED1 transcription factor affects auxin homeostasis in roots. *J Exp Bot*, 68(16), pp. 4627-4634.

Kroj, T., Savino, G., Valon, C., Giraudat, J. & Parcy, F. (2003). Regulation of storage protein gene expression in Arabidopsis. *Development*, 130(24), pp. 6065-6073.
Laloum, T., De Mita, S., Gamas, P., Baudin, M. & Niebel, A. (2013). CCAAT-box binding transcription factors in plants: Y so many? *Trends in Plant Science*, 18(3), pp. 157-66.
Lamacchia, C., Shewry, P.R., Di Fonzo, N., Forsyth, J.L., Harris, N., Lazzeri, P.A., Napier, J.A., Halford, N.G. & Barcelo, P. (2001). Endosperm-specific activity of a storage protein gene promoter in transgenic wheat seed. *Journal of Experimental Botany*, 52(355), pp. 243-250.
Lamke, J., Brzezinka, K., Altmann, S. & Baurle, I. (2016). A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *EMBO J*, 35(2), pp. 162-75.
Latchman, D.S. (1997). Transcription factors: an overview. *Int J Biochem Cell Biol*, 29(12), pp. 1305-12.
Leduc, N., MatthysRochon, E., Rougier, M., Mogensen, L., Holm, P., Magnard, J.L. & Dumas, C. (1996). Isolated maize zygotes mimic in vivo embryonic development and express microinjected genes when cultured in vitro. *Developmental Biology*, 177(1), pp. 190-203.
Leonova, S., Grimberg, A., Marttila, S., Stymne, S. & Carlsson, A.S. (2010). Mobilization of lipid reserves during germination of oat (Avena sativa L.), a cereal rich in endosperm oil. *J Exp Bot*, 61(11), pp. 3089-99.
Lepiniec, L., Devic, M., Roscoe, T.J., Bouyer, D., Zhou, D.X., Boulard, C., Baud, S. & Dubreucq, B. (2018). Molecular and epigenetic regulations and functions of the LAFL transcriptional regulators that control seed development. *Plant Reproduction*, 31(3), pp. 291-307.
Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M.X., Arondel, V., Bates, P.D., Baud, S., Bird, D., DeBono, A., Durrett, T.P., Franke, R.B., Graham, I.A., Katayama, K., Kelly, A.A., Larson, T., Markham, J.E., Miquel, M., Molina, I., Nishida, I., Rowland, O., Samuels, L., Schmid, K.M., Wada, H., Welti, R., Xu, C., Zallot, R. & Ohlrogge, J. (2013). Acyl-Lipid Metabolism. *The Arabidopsis Book*, 2013(11).
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, pp. 33-45.
Lin, C.T. & Thomashow, M.F. (1992). A Cold-Regulated Arabidopsis Gene Encodes a Polypeptide Having Potent Cryoprotective Activity. *Biochemical and Biophysical Research Communications*, 183(3), pp. 1103-1108.
Liu, J., Deng, S., Wang, H., Ye, J., Wu, H.-W., Sun, H.-X. & Chua, N.-H. (2016). CURLY LEAF Regulates Gene Sets Coordinating Seed Size and Lipid Biosynthesis. *Plant Physiology*, 171(1), pp. 424-436.
Liu, J.X. & Howell, S.H. (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), pp. 782-96.
Liu, K.S. (2011). Comparison of Lipid Content and Fatty Acid Composition and Their Distribution within Seeds of 5 Small Grain Species. *Journal of Food Science*, 76(2), pp. C334-C342.
Liu, Y.X., Koornneef, M. & Soppe, W.J.J. (2007). The absence of histone H2B monoubiquitination in the Arabidopsis hub1 (rdo4) mutant reveals a role for chromatin remodeling in seed dormancy. *Plant Cell*, 19(2), pp. 433-444.
Lotan, T., Ohto, M., Yee, K.M., West, M.A.L., Lo, R., Kwong, R.W., Yamagishi, K., Fischer, R.L., Goldberg, R.B. & Harada, J.J. (1998). Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell*, 93(7), pp. 1195-1205.
Lu, Q.S., dela Paz, J., Pathmanathan, A., Chiu, R.S., Tsai, A.Y.L. & Gazzarrini, S. (2010). The C-terminal domain of FUSCA3 negatively regulates mRNA and protein levels, and mediates sensitivity to the hormones abscisic acid and gibberellic acid in Arabidopsis. *Plant Journal*, 64(1), pp. 100-113.
Ma, W., Kong, Q., Grix, M., Mantyla, J.J., Yang, Y., Benning, C. & Ohlrogge, J.B. (2015a). Deletion of a C-terminal intrinsically disordered region of WRINKLED1 affects its stability and enhances oil accumulation in Arabidopsis. *Plant Journal*, 83(5), pp. 864-74.
Ma, W., Kong, Q., Grix, M., Mantyla, J.J., Yang, Y., Benning, C. & Ohlrogge, J.B. (2015b). Deletion of a C-terminal intrinsically disordered region of WRINKLED1 affects its stability and enhances oil accumulation in Arabidopsis. *Plant Journal*, 83(5), pp. 864-874.
Ma, W., Kong, Q., Mantyla, J.J., Yang, Y., Ohlrogge, J.B. & Benning, C. (2016). 14-3-3 protein mediates plant seed oil biosynthesis through interaction with AtWRI1. *Plant Journal*, 88(2), pp. 228-235.

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 Journal*, 60(3), pp. 476-87.
Magnard, J.L., Le Deunff, E., Domenech, J., Rogowsky, P.M., Testillano, P.S., Rougier, M., Risueno, M.C., Vergne, P. & Dumas, C. (2000). Genes normally expressed in the endosperm are expressed at early stages of microspore embryogenesis in maize. *Plant Molecular Biology*, 44(4), pp. 559-574.
Mansfield, S.G. & Briarty, L.G. (1991). Early embryogenesis in Arabidopsis thaliana. II. The developing embryo. *Canadian Journal of Botany*, 69(3), pp. 461-476.
Masaki, T., Mitsui, N., Tsukagoshi, H., Nishii, T., Morikami, A. & Nakamura, K. (2005). ACTIVATOR of Spomin::LUC1/WRINKLED1 of Arabidopsis thaliana transactivates sugar-inducible promoters. *Plant Cell Physiol*, 46(4), pp. 547-56.
Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C. & Paszkowski, J. (2007). Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell*, 130(5), pp. 851-62.
Mayran, A. & Drouin, J. (2018). Pioneer transcription factors shape the epigenetic landscape. *J Biol Chem*, 293(36), pp. 13795-13804.
McCarty, D.R., Hattori, T., Carson, C.B., Vasil, V., Lazar, M. & Vasil, I.K. (1991). The Viviparous-1 developmental gene of maize encodes a novel transcriptional activator. *Cell*, 66(5), pp. 895-905.
Meinke, D.W. (1992). A Homoeotic Mutant of Arabidopsis thaliana with Leafy Cotyledons. *Science*, 258(5088), pp. 1647-50.
Meinke, D.W., Franzmann, L.H., Nickle, T.C. & Yeung, E.C. (1994). Leafy Cotyledon Mutants of Arabidopsis. *Plant Cell*, 6(8), pp. 1049-1064.
Mitsuda, N. & Ohme-Takagi, M. (2009). Functional analysis of transcription factors in Arabidopsis. *Plant Cell Physiol*, 50(7), pp. 1232-48.
Mol, R., Matthys-Rochon, E. & Dumas, C. (1995). Embryogenesis and plant regeneration from maize zygotes by in vitro culture of fertilized embryo sacs. *Plant Cell Rep*, 14(12), pp. 743-7.
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 Research*, 40(17), pp. 8240-8254.
Morrison, I.N., Kuo, J. & O'Brien, T.P. (1975). Histochemistry and fine structure of developing wheat aleurone cells. *Planta*, 123(2), pp. 105-16.
Mozgova, I. & Hennig, L. (2015). The Polycomb Group Protein Regulatory Network. *Annual Review of Plant Biology, Vol 66*, 66, pp. 269-296.
Mozgova, I., Munoz-Viana, R. & Hennig, L. (2017). PRC2 Represses Hormone-Induced Somatic Embryogenesis in Vegetative Tissue of Arabidopsis thaliana. *Plos Genetics*, 13(1).
Mu, J.Y., Tan, H.L., Zheng, Q., Fu, F.Y., Liang, Y., Zhang, J.A., Yang, X.H., Wang, T., Chong, K., Wang, X.J. & Zuo, J.R. (2008). LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. *Plant Physiology*, 148(2), pp. 1042-1054.
Nag, R., Maity, M.K. & DasGupta, M. (2005). Dual DNA binding property of ABA insensitive 3 like factors targeted to promoters responsive to ABA and auxin. *Plant Molecular Biology*, 59(5), pp. 821-838.
Nakamura, S., Lynch, T.J. & Finkelstein, R.R. (2001). Physical interactions between ABA response loci of Arabidopsis. *Plant Journal*, 26(6), pp. 627-635.
Nakano, T., Suzuki, K., Fujimura, T. & Shinshi, H. (2006). Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant Physiology*, 140(2), pp. 411-432.
Nambara, E., Hayama, R., Tsuchiya, Y., Nishimura, M., Kawaide, H., Kamiya, Y. & Naito, S. (2000). The role of ABI3 and FUS3 loci in Arabidopsis thaliana on phase transition from late embryo development to germination. *Dev Biol*, 220(2), pp. 412-23.
Nambara, E., Keith, K., Mccourt, P. & Naito, S. (1995). A Regulatory Role for the Abi3 Gene in the Establishment of Embryo Maturation in Arabidopsis-Thaliana. *Development*, 121(3), pp. 629-636.
Nambara, E. & Marion-Poll, A. (2005). Abscisic acid biosynthesis and catabolism. *Annual Review of Plant Biology*, 56, pp. 165-185.
Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., Vonrhein, 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), pp. 132-43.
Norton, G. & Harris, J.F. (1975). Compositional Changes in Developing Rape Seed (Brassica-Napus L). *Planta*, 123(2), pp. 163-174.
Ohto, M., Fischer, R.L., Goldberg, R.B., Nakamura, K. & Harada, J.J. (2005). Control of seed mass by APETALA2. *Proceedings of the National Academy of Sciences of the United States of America*, 102(8), pp. 3123-3128.
Okamoto, M., Kuwahara, A., Seo, M., Kushiro, T., Asami, T., Hirai, N., Kamiya, Y., Koshiba, T. & Nambara, E. (2006). CYP707A1 and CYP707A2, which encode abscisic acid 8'-hydroxylases, are indispensable for proper control of seed dormancy and germination in Arabidopsis. *Plant Physiology*, 141(1), pp. 97-107.
Olsen, O.A. (2001). ENDOSPERM DEVELOPMENT: Cellularization and Cell Fate Specification. *Annu Rev Plant Physiol Plant Mol Biol*, 52, pp. 233-267.
Olsen, O.A., Lemmon, B. & Brown, R. (1998). A model for aleurone development. *Trends in Plant Science*, 3(5), pp. 168-169.
Opsahl-Ferstad, H.G., Le Deunff, E., Dumas, C. & Rogowsky, P.M. (1997). ZmEsr, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. *Plant Journal*, 12(1), pp. 235-46.
Para, A., Li, Y., Marshall-Colon, A., Varala, K., Francoeur, N.J., Moran, T.M., Edwards, M.B., Hackley, C., Bargmann, B.O., Birnbaum, K.D., McCombie, W.R., Krouk, G. & Coruzzi, G.M. (2014). Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in Arabidopsis. *Proc Natl Acad Sci U S A*, 111(28), pp. 10371-6.
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), pp. 1265-1277.
Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M. & Giraudat, J. (1994). Regulation of gene expression programs during Arabidopsis seed development: roles of the ABI3 locus and of endogenous abscisic acid. *Plant Cell*, 6(11), pp. 1567-82.
Pazos, F., Pietrosemoli, N., Garcia-Martin, J.A. & Solano, R. (2013). Protein intrinsic disorder in plants. *Front Plant Sci*, 4, p. 363.
Peng, F.Y. & Weselake, R.J. (2013). Genome-wide identification and analysis of the B3 superfamily of transcription factors in Brassicaceae and major crop plants. *Theor Appl Genet*, 126(5), pp. 1305-19.
Perez, S. & Bertoft, E. (2010). The molecular structures of starch components and their contribution to the architecture of starch granules: A comprehensive review. *Starch-Starke*, 62(8), pp. 389-420.
Peterson, D.M. & Wood, D.F. (1997). Composition and structure of high-oil oat. *Journal of Cereal Science*, 26(1), pp. 121-128.
Petroni, K., Kumimoto, R.W., Gnesutta, N., Calvenzani, V., Fornari, M., Tonelli, C., Holt, B.F., 3rd & Mantovani, R. (2012). The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. *Plant Cell*, 24(12), pp. 4777-92.
Post-Beittenmiller, D., Roughan, G. & Ohlrogge, J.B. (1992). Regulation of plant Fatty Acid biosynthesis : analysis of acyl-coenzyme a and acyl-acyl carrier protein substrate pools in spinach and pea chloroplasts. *Plant Physiology*, 100(2), pp. 923-30.
Qu, R., Wang, S.M., Lin, Y.H., Vance, V.B. & Huang, A.H.C. (1986). Characteristics and Biosynthesis of Membrane-Proteins of Lipid Bodies in the Scutella of Maize (Zea-Mays-L). *Biochemical Journal*, 235(1), pp. 57-65.
Quadrana, L. & Colot, V. (2016). Plant Transgenerational Epigenetics. *Annu Rev Genet*, 50, pp. 467-491.
Rashid, M., He, G.Y., Yang, G.X., Hussain, J. & Yan, X. (2012). AP2/ERF Transcription Factor in Rice: Genome-Wide Canvas and Syntenic Relationships between Monocots and Eudicots. *Evolutionary Bioinformatics*, 8, pp. 321-355.
Raz, V., Bergervoet, J.H. & Koornneef, M. (2001). Sequential steps for developmental arrest in Arabidopsis seeds. *Development*, 128(2), pp. 243-52.
Reidt, W., Wohlfarth, T., Ellerstrom, M., Czihal, A., Tewes, A., Ezcurra, I., Rask, L. & Baumlein, H. (2000). Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. *Plant Journal*, 21(5), pp. 401-408.

Reyes, J.L., Rodrigo, M.J., Colmenero-Flores, J.M., Gil, J.V., Garay-Arroyo, A., Campos, F., Salamini, F., Bartels, D. & Covarrubias, A.A. (2005). Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects in vitro. *Plant Cell and Environment*, 28(6), pp. 709-718.

Rider, S.D., Henderson, J.T., Jerome, R.E., Edenberg, H.J., Romero-Severson, J. & Ogas, J. (2003). Coordinate repression of regulators of embryonic identity by PICKLE during germination in Arabidopsis. *Plant Journal*, 35(1), pp. 33-43.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K. & Yu, G. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, 290(5499), pp. 2105-10.

Riechmann, J.L. & Meyerowitz, E.M. (1998). The AP2/EREBP family of plant transcription factors. *Biological Chemistry*, 379(6), pp. 633-646.

Rietz, S., Dermendjiev, G., Oppermann, E., Tafesse, F.G., Effendi, Y., Holk, A., Parker, J.E., Teige, M. & Scherer, G.F.E. (2010). Roles of Arabidopsis Patatin-Related Phospholipases A in Root Development Are Related to Auxin Responses and Phosphate Deficiency. *Molecular Plant*, 3(3), pp. 524-538.

Ripoll, J.J., Roeder, A.H., Ditta, G.S. & Yanofsky, M.F. (2011). A novel role for the floral homeotic gene APETALA2 during Arabidopsis fruit development. *Development*, 138(23), pp. 5167-76.

Robichaud, C. & Sussex, I.M. (1986). The Response of Viviparous-1 and Wild-Type Embryos of Zea-Mays to Culture in the Presence of Abscisic-Acid. *Journal of Plant Physiology*, 126(2-3), pp. 235-242.

Roscoe, T.T., Guilleminot, J., Bessoule, J.J., 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), pp. 1215-28.

Ruuska, S.A., Schwender, J. & Ohlrogge, J.B. (2004). The capacity of green oilseeds to utilize photosynthesis to drive biosynthetic processes. *Plant Physiology*, 136(1), pp. 2700-9.

Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K. & Yamaguchi-Shinozaki, K. (2002). DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochemical and Biophysical Research Communications*, 290(3), pp. 998-1009.

Sanchez-Ballesta, M.T., Rodrigo, M.J., LaFuente, M.T., Granell, A. & Zacarias, L. (2004). Dehydrin from citrus, which confers in vitro dehydration and freezing protection activity, is constitutive and highly expressed in the flavedo of fruit but responsive to cold and water stress in leaves. *Journal of Agricultural and Food Chemistry*, 52(7), pp. 1950-1957.

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 Journal*, 54(4), pp. 608-620.

Schel, J.H.N., Kieft, H. & Vanlammeren, A.A.M. (1984). Interactions between Embryo and Endosperm during Early Developmental Stages of Maize Caryopses (Zea-Mays). *Canadian Journal of Botany-Revue Canadienne De Botanique*, 62(12), pp. 2842-2853.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D. & Lohmann, J.U. (2005). A gene expression map of Arabidopsis thaliana development. *Nature Genetics*, 37(5), pp. 501-506.

Schneider, A., Aghamirzaie, D., Elmarakeby, H., Poudel, A.N., Koo, A.J., Heath, L.S., Grene, R. & Collakova, E. (2016). Potential targets of VIVIPAROUS1/ABI3-LIKE1 (VAL1) repression in developing Arabidopsis thaliana embryos. *Plant Journal*, 85(2), pp. 305-319.

Shen, B., Allen, W.B., Zheng, P., Li, C., Glassman, K., Ranch, J., Nubel, D. & Tarczynski, M.C. (2010). Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. *Plant Physiology*, 153(3), pp. 980-7.

Shen, Y., Devic, M., Lepiniec, L. & Zhou, D.X. (2015). Chromodomain, Helicase and DNA-binding CHD1 protein, CHR5, are involved in establishing active chromatin state of seed maturation genes. *Plant Biotechnology Journal*, 13(6), pp. 811-820.

Shewry, P.R. (2003). Tuber storage proteins. *Annals of Botany*, 91(7), pp. 755-769.

Shewry, P.R., Napier, J.A. & Tatham, A.S. (1995). Seed Storage Proteins - Structures and Biosynthesis. *Plant Cell*, 7(7), pp. 945-956.

Shimada, T.L. & Hara-Nishimura, I. (2015). Leaf oil bodies are subcellular factories producing antifungal oxylipins. *Current Opinion in Plant Biology*, 25, pp. 145-50.
Shiu, S.H., Shih, M.C. & Li, W.H. (2005). Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiology*, 139(1), pp. 18-26.
Shockey, J., Regmi, A., Cotton, K., Adhikari, N., Browse, J. & Bates, P.D. (2016). Identification of Arabidopsis GPAT9 (At5g60620) as an Essential Gene Involved in Triacylglycerol Biosynthesis. *Plant Physiology*, 170(1), pp. 163-79.
Siefers, N., Dang, K.K., Kumimoto, R.W., Bynum, W.E.t., Tayrose, G. & Holt, B.F., 3rd (2009). Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiology*, 149(2), pp. 625-41.
Sims, R.J., Nishioka, K. & Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. *Trends in Genetics*, 19(11), pp. 629-639.
Singh, B.K. & Jenner, C.F. (1983). Culture of Detached Ears of Wheat in Liquid Culture - Modification and Extension of the Method. *Australian Journal of Plant Physiology*, 10(2), pp. 227-236.
Staby, L., O'Shea, C., Willemoes, M., Theisen, F., Kragelund, B.B. & Skriver, K. (2017). Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochemical Journal*, 474, pp. 2509-2532.
Stone, S.L., Braybrook, S.A., Paula, S.L., Kwong, L.W., Meuser, J., Pelletier, J., Hsieh, T.F., Fischer, R.L., Goldberg, R.B. & Harada, J.J. (2008). Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 105(8), pp. 3151-3156.
Stone, S.L., Kwong, L.W., Yee, K.M., Pelletier, J., Lepiniec, L., Fischer, R.L., Goldberg, R.B. & Harada, J.J. (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20), pp. 11806-11811.
Streb, S. & Zeeman, S.C. (2012). Starch metabolism in Arabidopsis. *Arabidopsis Book*, 10, p. e0160.
Sturm, A. & Tang, G.Q. (1999). The sucrose-cleaving enzymes of plants are crucial for development, growth and carbon partitioning. *Trends in Plant Science*, 4(10), pp. 401-407.
Suzuki, M., Kao, C.Y., Cocciolone, S. & McCarty, D.R. (2001). Maize VP1 complements Arabidopsis abi3 and confers a novel ABA/auxin interaction in roots. *Plant Journal*, 28(4), pp. 409-18.
Suzuki, M., Kao, C.Y. & McCarty, D.R. (1997). The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. *Plant Cell*, 9(5), pp. 799-807.
Suzuki, M., Wang, H.H.Y. & McCarty, D.R. (2007). Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiology*, 143(2), pp. 902-911.
Suzuki, M., Wu, S., Li, Q. & McCarty, D.R. (2014). Distinct functions of COAR and B3 domains of maize VP1 in induction of ectopic gene expression and plant developmental phenotypes in Arabidopsis. *Plant Molecular Biology*, 85(1-2), pp. 179-91.
Swanson, S.J., Bethke, P.C. & Jones, R.L. (1998). Barley aleurone cells contain two types of vacuoles. Characterization Of lytic organelles by use of fluorescent probes. *Plant Cell*, 10(5), pp. 685-98.
Swift, J. & Coruzzi, G.M. (2017). A matter of time - How transient transcription factor interactions create dynamic gene regulatory networks. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms*, 1860(1), pp. 75-83.
Takeda, S. & Paszkowski, J. (2006). DNA methylation and epigenetic inheritance during plant gametogenesis. *Chromosoma*, 115(1), pp. 27-35.
Tang, X.R., Bian, S.M., Tang, M.J., Lu, Q., Li, S.B., Liu, X.G., Tian, G., Nguyen, V., Tsang, E.W.T., Wang, A.M., Rothstein, S.J., Chen, X.M. & Cui, Y.H. (2012). MicroRNA-Mediated Repression of the Seed Maturation Program during Vegetative Development in Arabidopsis. *Plos Genetics*, 8(11).
Tao, Z., Shen, L.S., Gu, X.F., Wang, Y.Z., Yu, H. & He, Y.H. (2017). Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. *Nature*, 551(7678), pp. 124-+.
Thazar-Poulot, N., Miquel, M., Fobis-Loisy, I. & Gaude, T. (2015). Peroxisome extensions deliver the Arabidopsis SDP1 lipase to oil bodies. *Proceedings of the National Academy of Sciences of the United States of America*, 112(13), pp. 4158-4163.
Thoyts, P.J.E., Millichip, M.I., Stobart, A.K., Griffiths, W.T., Shewry, P.R. & Napier, J.A. (1995). Expression and in-Vitro Targeting of a Sunflower Oleosin. *Plant Molecular Biology*, 29(2), pp. 403-410.

Ting, J.T.L., Lee, K.Y., Ratnayake, C., Platt, K.A., Balsamo, R.A. & Huang, A.H.C. (1996). Oleosin genes in maize kernels having diverse oil contents are constitutively expressed independent of oil contents - Size and shape of intracellular oil bodies are determined by the oleosins oils ratio. *Planta*, 199(1), pp. 158-165.

To, A., Joubes, J., Barthole, G., Lecureuil, A., Scagnelli, A., Jasinski, S., Lepiniec, L. & Baud, S. (2012). WRINKLED Transcription Factors Orchestrate Tissue-Specific Regulation of Fatty Acid Biosynthesis in Arabidopsis. *Plant Cell*, 24(12), pp. 5007-5023.

To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J. & Parcy, F. (2006a). A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell*, 18(7), pp. 1642-1651.

To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J. & Parcy, F. (2006b). A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell*, 18(7), pp. 1642-51.

Tomlinson, K.L., McHugh, S., Labbe, H., Grainger, J.L., James, L.E., Pomeroy, K.M., Mullin, J.W., Miller, S.S., Dennis, D.T. & Miki, B.L. (2004). Evidence that the hexose-to-sucrose ratio does not control the switch to storage product accumulation in oilseeds: analysis of tobacco seed development and effects of overexpressing apoplastic invertase. *J Exp Bot*, 55(406), pp. 2291-303.

Torres-Ruiz, R.A. & Jurgens, G. (1994). Mutations in the FASS gene uncouple pattern formation and morphogenesis in Arabidopsis development. *Development*, 120(10), pp. 2967-78.

Troncoso-Ponce, M.A., 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), pp. 2666-2682.

Tsafou, K., Tiwari, P.B., Forman-Kay, J.D., Metallo, S.J. & Toretsky, J.A. (2018). Targeting Intrinsically Disordered Transcription Factors: Changing the Paradigm. *J Mol Biol*, 430(16), pp. 2321-2341.

Tsukagoshi, H., Morikami, A. & Nakamura, K. (2007). Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. *Proceedings of the National Academy of Sciences of the United States of America*, 104(7), pp. 2543-2547.

Tsukagoshi, H., Saijo, T., Shibata, D., Morikami, A. & Nakamura, K. (2005). Analysis of a sugar response mutant of Arabidopsis identified a novel B3 domain protein that functions as an active transcriptional repressor. *Plant Physiology*, 138(2), pp. 675-685.

Wang, F. & Perry, S.E. (2013). Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. *Plant Physiology*, 161(3), pp. 1251-64.

Wang, X.C., Niu, Q.W., Teng, C., Li, C., Mu, J.Y., Chua, N.H. & Zuo, J.R. (2009). Overexpression of PGA37/MYB118 and MYB115 promotes vegetative-to-embryonic transition in Arabidopsis. *Cell Research*, 19(2), pp. 224-235.

Varala, K., Li, Y., Marshall-Colon, A., Para, A. & Coruzzi, G.M. (2015). "Hit-and-Run" leaves its mark: catalyst transcription factors and chromatin modification. *Bioessays*, 37(8), pp. 851-6.

Weber, H., Borisjuk, L. & Wobus, U. (1997). Sugar import and metabolism during seed development. *Trends in Plant Science*, 2(5), pp. 169-174.

Veerappan, V., Wang, J., Kang, M.Y., Lee, J., Tang, Y.H., Jha, A.K., Shi, H.Z., Palanivelu, R. & Allen, R.D. (2012). A novel HSI2 mutation in Arabidopsis affects the PHD-like domain and leads to derepression of seed-specific gene expression. *Planta*, 236(1), pp. 1-17.

West, M., Yee, K.M., Danao, J., Zimmerman, J.L., Fischer, R.L., Goldberg, R.B. & Harada, J.J. (1994). LEAFY COTYLEDON1 Is an Essential Regulator of Late Embryogenesis and Cotyledon Identity in Arabidopsis. *Plant Cell*, 6(12), pp. 1731-1745.

Wolters-Arts, M., Lush, W.M. & Mariani, C. (1998). Lipids are required for directional pollen-tube growth. *Nature*, 392(6678), pp. 818-821.

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 and Cell Physiology*, 51(12), pp. 2031-2046.

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), pp. 267-76.

Yang, F.W. & Feng, X.Q. (2015). Abscisic acid biosynthesis and catabolism and their regulation roles in fruit ripening. *Phyton-International Journal of Experimental Botany*, 84(2), pp. 444-453.

Yang, Y., Munz, J., Cass, C., Zienkiewicz, A., Kong, Q., Ma, W., Sanjaya, Sedbrook, J. & Benning, C. (2015). Ectopic Expression of WRINKLED1 Affects Fatty Acid Homeostasis in Brachypodium distachyon Vegetative Tissues. *Plant Physiology*, 169(3), pp. 1836-47.
Youngs, V.L., Puskulcu, M. & Smith, R.R. (1977). Oat Lipids .1. Composition and Distribution of Lipid Components in 2 Oat Cultivars. *Cereal Chemistry*, 54(4), pp. 803-812.
Yu, H.B., Johnson, R., Kunarso, G. & Stanton, L.W. (2011). Coassembly of REST and its cofactors at sites of gene repression in embryonic stem cells. *Genome Research*, 21(8), pp. 1284-1293.
Zaret, K.S. & Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes & Development*, 25(21), pp. 2227-2241.
Zeeman, S.C., Kossmann, J. & Smith, A.M. (2010). Starch: Its Metabolism, Evolution, and Biotechnological Modification in Plants. *Annual Review of Plant Biology, Vol 61*, 61, pp. 209-234.
Zhang, D., Jing, Y.J., Jiang, Z.M. & Lin, R.C. (2014). The Chromatin-Remodeling Factor PICKLE Integrates Brassinosteroid and Gibberellin Signaling during Skotomorphogenic Growth in Arabidopsis. *Plant Cell*, 26(6), pp. 2472-2485.
Zhang, H., Jin, J., Tang, L., Zhao, Y., Gu, X., Gao, G. & Luo, J. (2011). PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. *Nucleic Acids Research*, 39(Database issue), pp. D1114-7.
Zhang, H., Rider, S.D., Henderson, J.T., Fountain, M., Chuang, K., Kandachar, V., Simons, A., Edenberg, H.J., Romero-Severson, J., Muir, W.M. & Ogas, J. (2008). The CHD3 remodeler PICKLE promotes trimethylation of histone H3 lysine 27. *Journal of Biological Chemistry*, 283(33), pp. 22637-22648.
Zhang, Y.F., Cao, G.Y., Qu, L.J. & Gu, H.Y. (2009). Involvement of an R2R3-MYB transcription factor gene AtMYB118 in embryogenesis in Arabidopsis. *Plant Cell Reports*, 28(3), pp. 337-346.
Zhao, H., Wu, D., Kong, F., Lin, K., Zhang, H. & Li, G. (2016). The Arabidopsis thaliana Nuclear Factor Y Transcription Factors. *Front Plant Sci*, 7, p. 2045.
Zhao, M.L., Yang, S.G., Liu, X.C. & Wu, K.Q. (2015). Arabidopsis histone demethylases LDL1 and LDL2 control primary seed dormancy by regulating DELAY OF GERMINATION 1 and ABA signaling-related genes. *Frontiers in Plant Science*, 6.
Zheng, J., Chen, F.Y., Wang, Z., Cao, H., Li, X.Y., Deng, X., Soppe, W.J.J., Li, Y. & Liu, Y.X. (2012). A novel role for histone methyltransferase KYP/SUVH4 in the control of Arabidopsis primary seed dormancy. *New Phytologist*, 193(3), pp. 605-616.
Zheng, Y., Ren, N., Wang, H., Stromberg, A.J. & Perry, S.E. (2009a). Global identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like15. *Plant Cell*, 21(9), pp. 2563-77.
Zheng, Y.M., Ren, N., Wang, H., Stromberg, A.J. & Perry, S.E. (2009b). Global Identification of Targets of the Arabidopsis MADS Domain Protein AGAMOUS-Like15. *Plant Cell*, 21(9), pp. 2563-2577.
Zhu, C. & Perry, S.E. (2005). Control of expression and autoregulation of AGL15, a member of the MADS-box family. *Plant Journal*, 41(4), pp. 583-594.
Zhuang, J., Cai, B., Peng, R.H., Zhu, B., Jin, X.F., Xue, Y., Gao, F., Fu, X.Y., Tian, Y.S., Zhao, W., Qiao, Y.S., Zhang, Z., Xiong, A.S. & Yao, Q.H. (2008). Genome-wide analysis of the AP2/ERF gene family in Populus trichocarpa. *Biochemical and Biophysical Research Communications*, 371(3), pp. 468-474.

Popular science summary

Essentially, plants are self-replicating solar powered robots able to extract carbon dioxide from the air only extruding oxygen as a by-product. The carbon bound by this process is used by the plants to form a large number of different molecules such as proteins, starch and oil that can be used for growth or for energy storage. This makes plants excellent tools to use in the production of sustainable and environmentally friendly oils which can be used to substitute fossil oil in a number of applications. These include plastics, cosmetics, paints as well as other specialised applications. Many plants naturally produce oil in their seeds to act as a nutrient reserve for the seedling and much human effort has been put into trying to improve on these so called oil crops to increase their yield and oil quality. However, with an increased interest in plant-based oil there is also an increased interest in new and improved oil crops which can be grown in new areas and even utilised in novel ways.

Similar to robots being controlled by their coded programs, plants are controlled by complex networks of genes encoded by the DNA. With the advent of new techniques we are now able to not only study how plants are constructed but also write our own programs using the DNA-code as our programming language. This allows us to for example change where and how plants produce oil which can be a key to increase oil production without having to reduce food or feed production. However, to do this we need to better understand how this code is written and structured. In this thesis the regulatory programs controlling where and how carbon is to be stored during seed development was investigated. We show that this process is highly regulated by several developmental programs and that it is also able to turn itself off during certain circumstances. We are also able to show that it is possible to increase oil accumulation in wheat using parts of this complex programming allowing us to better understand how this can be utilised to help humanity transition towards a future bio-based economy.

Populärvetenskaplig sammanfattning

Ur ett bioteknologiskt perspektiv kan växter ses som självreplikerande soldrivna robotar som kan utvinna koldioxid från luften med syre som enda biprodukt. Det kol som binds används sedan av växterna för att bygga ett stort antal olika molekyler såsom proteiner, stärkelse och olja. Detta gör dem till utmärkta verktyg att använda för produktion av hållbara och miljövänliga oljor som kan ersätta fossil olja i ett antal tillämpningar. Dessa inkluderar plast, kosmetika, färger samt andra specialiserade användningsområden. Många växter producerar olja i sina frön som en näringsreserv för fröplantan och mycket arbete har lagts ned på att försöka förbättra dessa så kallade oljegrödor i syfte att öka deras avkastning och oljekvalité. Med med ett ökat intresse för växtbaserad olja finns det också ett ökat intresse för nya och förbättrade oljegrödor som kan odlas i nya områden och användas på nya sätt.

I likhet med robotar som styrs av kodade program styrs växter av komplexa nätverk av gener som kodas av DNA. Med tillkomsten av nya tekniker kan vi nu inte bara studera hur växters program är uppbyggda utan även skriva våra egna program med DNA-koden som programmeringsspråk. Detta ger oss en nyckel till att ändra var och hur växter producerar olja kan ge ökad oljeproduktionen utan att behöva minska mat- eller foderavkastningen. Men för att lyckas med detta måste vi först förstå hur denna kod är skriven och strukturerad. I denna avhandling undersöks hur programmen som styr var och hur kol ska lagras under fröutveckling är konstruerade. Denna process är under hård kontroll av flera utvecklingsprogram och att den också har möjlighet att stänga av sig själv under vissa omständigheter. Vi kan också visa att det är möjligt att öka oljemängden i vete med hjälp av delar av detta program. På så sätt kan vi bättre förstå hur bioteknik kan användas för att hjälpa mänskligheten övergång till en framtida biobaserad ekonomi.

Acknowledgements

Despite primarily being presented as the work of a single individual, a doctoral thesis is the result of the combined effort of a large group of researchers and technical personnel. Without me, this thesis would still be written, although probably a bit different in content and style. Therefore, this work is as much theirs as it is mine.

First and foremost, I want to extend my deepest gratitude to my main supervisor *Per Hofvander* for welcoming me into his lab and giving me his unwavering support. Your talent as a supervisor is only surpassed by your skill as a researcher and your ability to identify small details in huge sets of data and to put them into a theoretical context still baffles me after four years. It has been an honour to work for and with you and I sincerely hope that I will be able to continue our professional relationship in the future.

I am also extremely grateful to my co-supervisors *Åsa Grimberg* and *Anders Carlsson*. *Åsa*, your genuine enthusiasm for science and the application of our research has been and still is, a big source of inspiration for me. Your profound interest in bringing results from the lab into real-life use is one of the most important tasks a researcher can tackle, but it is also one of the hardest. I am however convinced that you are exactly the right person to take on this mission and I am glad to have had you as a supervisor during these years. *Anders*, I thank you for the support and for your involvement in much of the fundamental work upon which this work is built. A special thank to *Sten Stymne*. Not only for accepting me into your group or your time as my supervisor but also as the great source of inspiration and knowledge that you are. Not only for me but for so many other people around the world. Your work forms the stable foundation upon which it is now up to us in the younger generation of scientists to continue building a better society.

Thank also go to *Ida Lager* for being such great support and "hönsmamma" in the lab. Despite you always claiming the opposite you always had the answers for every question that I could come up with. And I think I've asked them all. You are an extremely talented researcher, a wonderful friend and an excellent teacher. It has been a privilege to share a lab with you, and I hope that I will be able to do it again in the future.

Furthermore, I want to extend my sincerest thanks to *Helen Lindgren*, *Helle Turesson* and *Niklas Olsson*. *Helen*, you are one of those people who rarely get the recognition that you deserve. Your tireless work, often uphill, not only to maintain but also to improve the lab, is vital for all research being carried out within the group. *Helle*, when I first started I couldn't stand still for more than a couple of seconds before you asked me if I needed any help with anything. Every time. And you still do. It is a great example of your character. Always helpful, caring and willing to support lost students develop new skills. You will always be a role model for me! And *Niklas*, I have really enjoyed your company and help in the lab. Without you, the everyday work would both be much more difficult and boring.

Thanks also go out to my office mates *Simon Jeppson*, *Magnus Carlsson* and *Joel Marklund* for your support and for the interesting times and discussions we have shared. *Simon*, we have shared crazy ideas, scientific projects and fun adventures as well as both easy and hard times. Your support has been invaluable to me, and I'm forever indebted to you for this. Thanks must also go to *Maja Brus-Szkalej* for your wonderful friendship and giving me the kick in the ass I so often need. Without your help this thesis would have been finished a year earlier. Never change! I also want to acknowledge the friendship and help I have received from *Kristin Konopatzki*, who never did anything I asked her to do but still succeeded perfectly.

To all my friends and colleagues at the Department of Plant Breeding, I send a heartfelt thank you for creating such a wonderful, warm and engaging workplace. Especially *Erik Alexanderson*, for your appreciated scrutiny of my project on several occasions as well as your valuable work with PlantLink. Also *Inger Åhman*, for your work on improving the doctoral education which has benefited me greatly.

But science also requires a democratic and open society to prosper. Therefore I also want to thank my brothers and sisters in arms in the 214th

reconnaissance company (214. Underrättelsekompaniet) for your contributions to our society. Especially *Roger Isaksson*, *Joakim Andersson*, *Johan Spennare*, *Per Elfving*, *Oskar Ehrning*, *Georg Stenström* and *Mikael Hagafors* for standing tall with me during times of difficulties.

I also want to acknowledge that this work would have been impossible without the financial aid from the Swedish Foundation for Strategic Research (Stiftelsen för Strategisk Forskning, SSF) and SLU. I also want to thank the Royal Physiographic Society in Lund (Kungliga Fysiografiska Sällskapet i Lund) for supporting me while presenting my research in an international setting.

Acta Universitatis Agriculturae Sueciae
Doctoral Thesis No. 2019:75

Three compounds — oil, starch and protein – form not only the basis for plant energy storage but also the foundation of agriculture and, in extension, human society. In this thesis, we investigate the gene regulatory networks involved in directing carbon flow towards oil accumulation in plants. Additionally, we explore how individual transcription factors as well as gene regulatory networks can be utilised to improve and expand the current set of commercially available oil crops.

Per Snell received his graduate education at the Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Alnarp. He received his MSc in Biology from Uppsala University, Uppsala.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

Online publication of thesis summary: https://pub.epsilon.slu.se

ISSN 1652-6880
ISBN (print version) 978-91-7760-468-6
ISBN (electronic version) 978-91-7760-469-3