Investigating the role of the VAL1 transcription factor in *Arabidopsis thaliana* embryo development

Andrew Paul Schneider

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Plant Pathology, Physiology, and Weed Science

Eva Collakova, Chair
Ruth Grene, Co-Chair
John G. Jelesko
Song Li

August 25, 2015
Blacksburg, VA

Keywords: Arabidopsis, seeds, transcription factor, epigenetic

Investigating the role of the VAL1 transcription factor in *Arabidopsis thaliana* embryo development

Andrew Paul Schneider

ABSTRACT

Developing oilseeds accumulate oils and seed storage proteins synthesized by the pathways of primary metabolism. Seed development and metabolism are positively regulated at the transcriptional level through the transcription factors belonging to the LAFL regulatory network. The *VAL* genes encode repressors of the seed maturation program in germinating seeds, but they are also expressed during early stages of seed maturation. VAL1 was identified through a reverse genetics approach as a regulator of seed metabolism, as *val1* mutant seeds accumulated elevated levels of storage proteins compared to the wild type. Two *VAL1* splice variants were identified, yielding the canonical protein and a truncated protein lacking the plant-homeodomain-like domain important for epigenetic repression. Transcriptomics analysis also revealed that VAL1 is a global epigenetic and transcriptional repressor in developing embryos, though none of the transcripts encoding the LAFL network regulators, including FUSCA3, were affected in *val1* embryos. However, VAL1 action is connected specifically to FUSCA3 as 38% of transcripts belonging to the FUSCA3 regulon, but not to other regulons, were largely de-repressed in the absence of VAL1. Based on our model, FUSCA3 activates expression of *VAL1* to repress transcription of seed maturation genes without interfering with expression of the core LAFL regulators.

Dedication

To Elizabeth Saris Schneider

Acknowledgements

This work would not have been possible without the support and encouragement of many individuals. First, I'd like to thank my advisor, Eva Collakova. To say she has been the best advisor I could have hoped for would be an understatement. When I was looking for a new lab, she contacted me out of the blue and even though I knew nothing about plants, she took a leap of faith and offered me a place in her lab. I think that paid off for both of us. I can't say enough about how her style of teaching was right for me. The lab was always a place I wanted to be and I never had to be worried about bad results or if I didn't know what something meant.

I'd like to thank my co-advisor, Ruth Grene for all the advice and great talks, often times not about science, that we've shared. She has a passion for learning that embodies what I think science is all about. She was an excellent counter-part to Eva.

I'd also like to thank my committee members, Drs. Jelesko, Li and Pilot, for all the encouragement and advice they gave me over the years. I learned a lot from each interaction with them, even in passing, about all things related to conducting experiments, analyzing data and knowing what's important and what is not.

I'd also like to thank Delasa Aghamirzaie, who's forgotten more about computational biology than I'll ever know. I'd like to thank Bhadra Gunesekera for all her support. I'd also like to thank past members of the Collakova lab, Yihui Fang and Daljit Singh. Their patience and assistance made joining the lab an easy process.

I'd like to thank all my friends from Latham Hall, Julien Besnard, Yu Shi, the Feds, John Herlihy, A.J. Prussin, Steve Rigoulot, and others who passed along encouragement, vital tips, tricks, and good times. I'd also like to thank Juin Yen, who not only bought me lunch two-times a week for a whole semester, but has been a good friend and someone I've approached for help

with problems many times. I'd also like to thank friends from outside the Latham Community whom I've known for many years: Jared Geissinger, Jake Kocher, Benjamin Webb, and Hardik Zatakia. Being around them made all this easy.

I'd like to thank The Undergrads: Aalok Parekh, Brittany Lassiter, and the awesomely-named Joseph Powers, for actually sticking around and doing what I said with minimal complaining, and for proving that the best way to learn something is to teach it to others. I particularly want to acknowledge Aalok, whose dedication to an incredibly tedious project approached super-human levels. Our working relationship developed into a friendship that continues to this day.

From my past life, I'd like to thank Dr. Wes Black, for teaching me a lot of the basic lab skills that I used throughout, for being a good friend in a tough time and beginning my education into the wonderful world of beer. I'd like to thank Dr. Florian Schubot for all of his advice and assistance during that time. I'd like to thank Sarah Riggs for possessing an inhuman level of cheerfulness, and also for being a complete nerd. I'd like to thank Regina Wallace for being one of the nicest people I've ever met. I'd also like to thank Dr. Zhaomin Yang, for making me realize what my limits are. The value of a lesson like that cannot be overstated.

Several people from my early education helped foster my interest in science: my 5th grade teacher, Sean Duffy and my 9th grade biology teacher Jean Packheiser. Also, Dr. Nicolas Proust, who was once my organic chemistry TA, who taught me the power of asking the right question. I'd also like to thank Marketta Wright, who was instrumental in making me believe in myself and kept me from simply giving up in high school.

When I was between labs, there was a period in which I had nothing to do. Dr. Birgit Scharf brought me into her lab and gave me a small project to do, and passed along to people she

knew that I needed a new lab, which is how I ended up in Eva's lab. Her kindness ultimately kept me from leaving science altogether, and I would not be here without her.

I'd like to thank Anne and Bill Hymen. I first met Anne when she gave my graduate recruitment orientation tour around campus. She and Bill have become very good friends of mine, and they were gracious enough to allow me to stay with them as I finished up the last month of my Ph.D. I'm blessed to have friends like them in my life.

I'd like to thank my parents, Paul and Giselle Schneider, for providing a home where my curiosity was never stifled and for never telling me I had limits on what I could do. I would never have made it without their encouragement and support. Also my siblings, for all the good times and non-serious discussions we'd have. I'd also like to thank my mother- and father-in law, James and Elizabeth Saris, for providing so much to my family as we went through all this together.

My wife, Ally Schneider, has meant more to me through this process than anyone or anything. She has stuck with me and encouraged me through all the tough times and helped me celebrate all the good times, even though that's usually not in my nature. We were apart for the first half of my doctorate, and since she joined me here, I've had the best years of my life. I love her with everything I have. I could never have done this without her. She gave me my beautiful daughter Lizzie, who completely changed my life.

Last, I'd like to thank my grandmother Mimi and my late grandfather Abo, who passed away shortly after I started in Eva's lab. They took the courageous steps to escape from Cuba and ensure their children and grandchildren wouldn't be raised under communism. Without their courage and conviction, I would never have been able to do this.

The last time I saw Abo, as he was getting into the car to leave, the final thing he said to me was “Dr. Andrew.” I can finally say “sí.”

Attributions

Several colleagues contributed to the projects, research, writing, and editing of the manuscript in this dissertation.

Chapter 2:

Eva Collakova, Ph.D. (Department of Plant Pathology, Physiology and Weed Science) is currently an Associate Professor at PPWS and is the corresponding author on the manuscript. She aided in project development, writing, and editing of the manuscripts.

Ruth Grene, Ph.D. (Department of Plant Pathology, Physiology and Weed Science) is a Professor at PPWS and is a co-author of the manuscript. She aided in project development, writing, and editing of the manuscripts.

Delasa Aghamirzaie is a co-author on the manuscript and is currently a Graduate Research Associate (Genetics, Bioinformatics, and Computational Biology Program). She contributed to the computational analyses.

Haitham Elmarakeby is a co-author on the manuscript and is currently a Graduate Research Associate (Department of Computer Science). He contributed to the computational analyses and writing of the manuscript.

Lenwood Heath, Ph.D. (Department of Computer Science) is a professor and co-author on the manuscript. He contributed to the computational analyses.

Abraham Koo, Ph.D. and Arati Nepal (Division of Biochemistry, Interdisciplinary Plant Group) are both at the University of Missouri. They provided sample analysis and contributed to the writing of the manuscript.

Table of Contents

Page

Appendix D

List of Tables

Page

List of Figures

Page

Chapter 2:
Potential targets of VAL1 repression in developing *Arabidopsis thaliana* embryos

Appendix A:
Supplemental figures

Chapter 1

Oilseed embryo development and potential regulatory functions of the transcription factor VAL1

A. Schneider[1]

[1]Department of Plant Pathology, Physiology and Weed Sciences, College of Agriculture and Life Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA.

1.1 Introduction

The colonization of land presented many challenges to plants, such as the lack of available water supply. Adapting to this challenge required, among other things, changes to reproductive strategies. Seed producing plants conquered land in part because the formation of seeds facilitates survival of plants in tough environmental conditions, such as drought, and allows for spreading to new environments that may be more suitable for vegetative growth (Bewley and Black, 1994; Bewley, 1997). Seeds have evolved extensive morphological and physiological traits that reflect the various strategies developed to overcome these challenges. Once released from the maternal plant, the seed will remain dormant until suitable growth conditions (which vary considerably depending on the species, ecotype, etc.) present themselves (Penfield et al., 2006; Footitt et al., 2011; Weitbrecht et al., 2011; Graeber et al., 2012; He et al., 2014).

Germination and early seedling development occur prior to the establishment of photosynthetic capability. Nutritive compounds (or seed storage compounds (SSCs)) stored within the seed tissues are catabolized into the chemical energy and structural components necessary until photoautotrophic processes are developed (Penfield et al., 2006; Theodoulou and Eastmond, 2012). Despite the aforementioned physiological and morphological variance, all seeds synthesize and store SSCs, which generally consist of a collection of oils, proteins, and carbohydrates (Baud et al., 2008). The proportions of these compounds vary depending on the species/ecotype/growth conditions and can account for up to 90% of the seed dry weight (Li et al., 2006; Baud et al., 2008). For example, monocotyledonous species such as maize primarily accumulate carbohydrates, while some dicotyledonous species such as soybean primarily accumulate proteins.

Seeds of the *Brassicaceae* family accumulate high levels of oils and proteins. For example, in *Brassica napus* seeds, oils contribute 50 – 60% to the dry weight (Li et al., 2006). High SSC levels in oilseeds such as *B. napus* are widely exploited for human and animal nutrition and industrial purposes. The existing uses and projected needs for increased food and fuel production have made oilseeds a target for metabolic engineering (Hills, 2004; Weselake et al., 2009). Understanding the complete processes (metabolic, regulatory, etc.) underlying SSC metabolism is critical if any attempt to engineer improved seed composition is to be attempted. Thale cress (*Arabidopsis thaliana*, Arabidopsis) is a small weed that is also a member of the *Brassicaceae* family. It is the primary model species for many aspects of plant biology and molecular biology, including seed development. Much of the molecular knowledge of dicotyledonous oilseed development is understood best in Arabidopsis due to the wealth of genomic and other molecular tools available. Despite the very small size of Arabidopsis seeds, a tremendous amount of physiological, morphological, and metabolic data have been obtained. Therefore, most of the results discussed below refer to work done on Arabidopsis.

1.2 The seed coat and endosperm support embryo development

The mature seed consists of three major structures: the seed coat, the endosperm, and the embryo. Each structure has specific roles during seed development and germination and is critical to produce a viable seed. For example, mutations that disrupt or prevent growth of the endosperm can be lethal to the seed as a whole (Lafon-Placette and Kohler, 2014). Each structure begins developing after fertilization (discussed below) and follows their own unique developmental programs (Lafon-Placette and Kohler, 2014). The overall development of the seed requires precise timing of developmental milestones, and there is likely communication

between the different structures. Ultimately the other structures serve to protect and facilitate growth of the embryo, which will develop into a vegetative plant. This section describes the development and roles of the seed coat and endosperm. The development of the embryo within the overall context of seed development is discussed in detail in Section 1.3.

1.2.1 The seed coat

The seed coat consists of several integumental layers derived from maternal ovular tissues (Debeaujon et al., 2000). During seed development the coat tissue forms the end of the symplastic connection to the maternal phloem and is the presumptive site of numerous sugar and amino acid transporters involved in endosperm/embryo development (Stadler et al., 2005; Zhang et al., 2007). These transporters are mostly unknown and release nutrients (principally glucose, sucrose, and glutamine) into the seed apoplast, which are subsequently imported into the endosperm (Patrick and Offler, 2001; Kim and Zambryski, 2005; Zhang et al., 2007). Recently, three members of the SWEET family of putative sucrose transporters were found to be involved in embryo development (Chen et al., 2015). Triple mutants of *SWEET 11*, *12* and *15* developed seeds that had delayed development and reductions in both starch and oil levels, suggesting that seed development largely relies on the supply of nutrients from the maternal organs. All three transporters were expressed in both the seed coat and endosperm of developing seeds. Other transporters involved in carbon and nitrogen export from the seed coat to the endosperm/embryo are predicted to exist, but have yet to be identified (de Jong et al., 1997; Tegeder, 2012).

Much more is known about the role of the seed coat in the maintenance of dormancy (Haughn and Chaudhury, 2005; Weitbrecht et al., 2011). As seed desiccation begins, the coat cells undergo programmed cell death, accumulate pigments, and begin to harden (Debeaujon et

al., 2000; Debeaujon et al., 2001; Nesi et al., 2001; Lepiniec et al., 2006). This produces a tough coat that protects the endosperm and embryo as the seed survives in the dormant state. The strength of the coat also has a role in maintaining dormancy by preventing the hypocotyl from breaking through (Bentsink and Koornneef, 2008). In addition, the various pigments that accumulate in the coat are also involved in dormancy. Mutants with reductions in coat pigmentation were found to have substantially reduced levels of dormancy (Debeaujon et al., 2000).

1.2.2 The endosperm

In Arabidopsis, the endosperm initially makes up the majority of the seed volume until it is consumed in the later stages of development. Endosperm development undergoes three phases: syncytial, cellular, and maturation phases. The syncytial phase begins with the double fertilization of the central cell, resulting in a triploid coenocyte surrounding a large central vacuole (Friedman, 2001; Baroux et al., 2002; Li and Berger, 2012; Lafon-Placette and Kohler, 2014). This multinucleate (~ 200 nuclei) cell initially occupies greater than 90% of the seed volume (Olsen, 2001; Berger et al., 2006). Cellularization is the next phase, and it is characterized by the *de novo* formation of cell membranes and walls around the nuclei and the loss of the central vacuole. This produces layers of differentiated endosperm cells with varying levels of thickness depending on the location within the seed (Li and Berger, 2012). In the final maturation phase, the endosperm volume declines, eventually occupying a single cell layer around the embryo called the aleurone layer (Olsen, 2001).

The endosperm is intimately involved in the regulation of embryo development, primarily through the establishment and maintenance of the source-sink relationship between the seed and

the maternal plant (Hill et al., 2003; Li and Berger, 2012; Lafon-Placette and Kohler, 2014). For example, in *B. napus*, the large central vacuole is the primary storage site for hexoses (Morley-Smith et al., 2008). As the central vacuole shrinks during cellularization, the amount of hexose in the seed declines with a concurrent increase in sucrose, which is taken in directly by the embryo (Kim et al., 2005; Morley-Smith et al., 2008). Mutations that prevent cellularization result in embryo arrest, most likely due to prevention of sucrose release (Baud et al., 2008; Morley-Smith et al., 2008; Pignocchi et al., 2009; Hehenberger et al., 2012). The endosperm also communicates with the developing embryo, presumably through the exchange of signaling molecules, although the nature and mechanism of this signaling is far from understood (Waters et al., 2013; Lafon-Placette and Kohler, 2014).

In the mature seed, the endosperm has a critical role in determining whether the seed will germinate or remain dormant (Yan et al., 2014). The strength of the micropylar endosperm tissues, which surround the embryo radicle, is the major determining factor governing radicle protrusion and the site of synthesis of many cell-wall loosening enzymes during germination (Dekkers et al., 2013). In addition, upon sensing water and/or low temperatures, the endosperm activates expression of abscisic acid (ABA) catabolic enzymes, breaking down endogenous ABA levels (Yamauchi et al., 2004; Okamoto et al., 2006). Enzymes involved in the biosynthesis of the germination stimulating gibberellins (GAs) are also upregulated in the endosperm in response to these environmental signals. Endosperm-synthesized GAs are detected by the embryo, leading to activation of germination specific processes such as expression of cell wall extending enzymes and the catabolism of seed storage compounds (Penfield et al., 2004; Penfield et al., 2006; Holdsworth et al., 2008a; Holdsworth et al., 2008b; Graeber et al., 2012; Finkelstein, 2013).

1.3 The embryo

Embryos grow into vegetative plants and are therefore the most critical part of the seed. Arabidopsis seeds take roughly 21 days to develop (under a 16-hour photoperiod), and progress through three basic stages, each discussed below (Baud et al., 2002; Fait et al., 2006; Jenik et al., 2007). The entire development process is measured based on the length of time from the fertilization event, which is classified as day 0 after pollination (DAP). The stages are relatively easy to identify visually, as the embryos undergo major changes in morphology, particularly during early seed development. As is the case for many seed species, specific points in seed development are often discussed in reference to embryo development (e.g., the endosperm cellularization during maturation refers to the stage of embryo development). It is easiest to discuss the phases of seed development in the context of the embryo, as it undergoes the most dynamic changes of the three major seed structures, including a series of cell divisions, pattern formation, rapid cell enlargement, synthesis and accumulation of SSCs, and finally, the accumulation of osmoprotectants and the acquisition of dormancy and desiccation tolerance.

1.3.1 Fertilization and embryogenesis

Embryos develop from zygotes, which are produced through the single fertilization of egg cells (Dumas and Rogowsky, 2008). Zygotes establish polarity and elongate during the first phase of seed development, called embryogenesis (0 – 7 DAP) (Goldberg et al., 1994; Jurgens, 2001; Jenik et al., 2007; Capron et al., 2009; Lau et al., 2012). Polarity is established by a series of cell divisions: the first is asymmetrical and forms an apical and a basal cell that develop into the embryo and suspensor, respectively (Jurgens, 2001; Jenik et al., 2007; Lau et al., 2012; Radoeva and Weijers, 2014). The suspensor serves as a conduit for nutrients delivered from the

endosperm, and is necessary in establishing the sink strength of the embryo during embryogenesis (Morley-Smith et al., 2008; Kawashima and Goldberg, 2010). Once embryogenesis is complete, the suspensor degrades. The embryo proper develops from the apical cell through a series of cell divisions, which first establish radial symmetry followed by bilateral symmetry (Lau et al., 2012; Radoeva and Weijers, 2014). Once bilateral symmetry has been established, tissue differentiation begins (Raz et al., 2001; Baud et al., 2008). The differentiation process is identified through stages named after the visual shape of the embryo: the globular stage followed by the heart, and ending with the torpedo stage (Capron et al., 2009). Once the embryo resembles a torpedo, cell differentiation is largely complete and the embryo consists of the primordial root tissue connected to the two cotyledons via the hypocotyl.

1.3.2 Maturation

The maturation phase of seed development (7 – 16 DAP, beginning with torpedo-shaped embryos) can be broken down to three sub-phases: early, middle, and late (Baud et al., 2002; Baud et al., 2008). The overall maturation phase involves cell enlargement and the synthesis of substantial amounts of SSCs from maternally derived nutrients (Raz et al., 2001; Baud et al., 2002; Baud and Graham, 2006; Baud and Lepiniec, 2009, 2010). The early maturation sub-phase (7 – 10 DAP) begins as the cells of torpedo-shaped embryos expand to accommodate the substantial increase in stored biomass, causing the cotyledons to lengthen and bend. During this period, photosynthetic pigmentation increases, as embryos require photosynthetic capability to produce enough energy and reductant for SSC synthesis and re-fix CO_2 lost in catabolic reactions of central carbon and nitrogen metabolism (Ruuska et al., 2004; Schwender et al., 2004; Fait et al., 2006). During the middle sub-phase (10 – 13 DAP), embryos take on an up-turned-U shape

to accommodate the increasing amounts of SSCs, and embryos occupy a greater volume of the seed. During the late maturation sub-phase (14 – 16 DAP), embryos reach the maximal size and begin transitioning to the desiccation phase.

1.3.3 Desiccation

The final phase of seed development consists of the acquisition of desiccation tolerance and dormancy (17 – 21 DAP), which involves a dramatic reduction in water content (dry seed contains less than 10% water) concurrent with metabolic "quiescence" (Baud et al., 2008; Dekkers et al., 2015). This rapid loss of water is a necessary prerequisite for the active process of the acquisition of dormancy (roughly defined as the ability to survive in an inactive state for extended periods of time) (Verdier et al., 2013). Desiccation introduces various stresses (e.g., increases in reactive oxygen species levels, cell shrinkage, and macromolecular degradation) that would be lethal if embryos could not tolerate them (Kranner and Birtic, 2005; Dekkers et al., 2015). As water loss occurs, cells lose volume, leaving behind free space filled with osmoprotectants, such as late-embryo abundant proteins (LEAs), in addition to non-reducing sucrose and raffinose family oligosaccharides that prevent cellular collapse (Hoekstra et al., 2001; Baud et al., 2002; Buitink and Leprince, 2008; Angelovici et al., 2010; Leprince and Buitink, 2010). In addition, SSCs are stored in specialized vacuoles (discussed below) that help to prevent cellular collapse (Leprince et al., 1998). Metabolic processes slow considerably, and to cope with the buildup of reactive oxygen species, antioxidant activity increases and photosynthetic capability is lost (Pammenter and Berjak, 1999). By the end of desiccation, the seed has no chlorophyll and is dormant and ready for dispersal.

1.4 Seed storage compounds (SSCs)

The dormant seed is metabolically quiescent as both a prerequisite for dormancy and as a consequence of desiccation. This desiccated, quiescent state requires photosynthetic capability to be abolished (as a ROS-reducing measure) and therefore, germinating seedlings lack the photoautotrophic capability of vegetative plants. To cope with the lack of a readily available energy supply, seeds evolved methods to synthesize and store large quantities of highly reduced carbon sources and proteins for immediate catabolism as soon as the seed initiates germination (Penfield et al., 2004; Penfield et al., 2006). Once germination begins, catabolic processes are initiated, utilizing stored enzymes and proteases to break down the stored oils, carbohydrates, and proteins to small metabolites for incorporation into central carbon and nitrogen metabolism and protein synthesis. If photosynthetic ability is not established through seedling greening, storage reserves will be exhausted, resulting in growth arrest.

1.4.1 Carbon and nitrogen transporters

Developing seeds are predominately heterotrophic and require the import of maternally derived carbon and nitrogen precursors (Schwender and Ohlrogge, 2002; Schwender et al., 2006). Imported sugars, primarily glucose and sucrose, are used as carbon skeletons for anabolic processes, such as fatty acid synthesis, in addition to catabolic processes (Baud and Graham, 2006). As mentioned above, transport of nutrients within seeds is not well understood (Zhang et al., 2007; Lemoine et al., 2013). Maternally derived nutrients are symplastically delivered to the seed coat in a form of post-phloem transport (Stadler et al., 2005). From there, nutrients are exported into the seed apoplast through mostly unknown mechanisms and then imported into the endosperm/embryo (Chen et al., 2015).

To date, only a handful of importers have been identified in Arabidopsis seeds. The putative sucrose transporter *SUCROSE-PROTON SYMPORTER 5* (*SUC5*) is highly expressed in the endosperm of developing seeds (late embryogenesis to early maturation) and *suc5* mutant seeds demonstrated a substantial reduction in oil levels (Baud et al., 2005). Although originally believed to transport sucrose into the endosperm from the seed apoplast, studies of *suc5* mutants combined with mutants defective in biotin synthesis demonstrated that SUC5 transports biotin, an important cofactor in fatty acid synthesis (Nikolau et al., 2003), into the endosperm and not sucrose, and deficiencies in biotin transport result in seed oil level reductions (Pommerrenig et al., 2013). Transporters specific to the embryo have not been reported, as the SUC5 transporter is found only in the endosperm. In addition, the three SWEET transporters (discussed in section 1.2.1) involved in seed development are all found in the seed coat/endosperm (Chen et al., 2015).

Organic nitrogen is supplied to the embryo as amino acids, with glutamine being the predominant amino acid found in the endosperm and embryo (Tegeder, 2014). Amino acids are delivered to developing seeds through the activity of two amino acid transporters: AMINO ACID PERMEASE 1 and 8 (AAP1 and AAP8). Developing seeds of *aap1* mutants were found to have elevated levels of free amino acids and analysis of the seed coat/endosperm separated from the embryo indicated that the amino acids accumulated in the endosperm and were not imported into the embryo (Sanders et al., 2009). Storage protein levels were reduced in mutant seeds, while oil levels were not changed. In addition, increased levels of protein storage vacuoles were found in the endosperm relative to the embryo of the *aap1* mutant seeds, further implicating AAP1 as a key importer of amino acids to the embryo. The function of AAP8 in embryos has not been demonstrated, however embryos carrying an insertion in this gene demonstrated aberrant development and reductions in the amino acid content of siliques (Schmidt et al., 2007). Several

other amino acid transporters are suspected to be involved in nitrogen transport into seeds, but their roles have yet to be fully determined (Okumoto and Pilot, 2011; Ladwig et al., 2012).

1.4.2 Seed storage oils

Arabidopsis seeds generate and store oils mainly as triacylglycerols, which consist of three fatty acid chains of varying length and desaturation esterified to a glycerol backbone (Stymne and Stobart, 1987; Baud et al., 2008; Li-Beisson et al., 2013). Seed oils are a rich source of highly reduced carbon and are broken down during germination into glycerol and fatty acids (Quettier and Eastmond, 2009; Weitbrecht et al., 2011; Theodoulou and Eastmond, 2012; Li-Beisson et al., 2013). Specialized lipases and glycerol kinases are activated during imbibition and mutants of these enzymes have reduced, but not abolished, germination ability (Eastmond, 2004; Penfield et al., 2006; Kelly et al., 2011). Once released, fatty acids are broken down through β-oxidation in the peroxisome, producing molecules of acetyl-CoA that are metabolized to succinate and oxaloacetate through the glyoylate cycle in the cytosol (Penfield et al., 2006; Graham, 2008).

Before assembly into oils, fatty acids are first generated as long acyl chains in the plastid (Li-Beisson et al., 2013). Carbon dioxide and acetyl-CoA molecules are converted to malonyl-CoA in the first committed step of fatty acid synthesis, after which an acyl carrier protein (ACP) replaces the CoA (Konishi et al., 1996; Li-Beisson et al., 2013). This nascent fatty acid chain is extended by two carbons through condensation with a molecule of acetyl-CoA, and subsequent two-carbon extensions are through addition of malonyl-ACP molecules (Brown et al., 2006). The primary products formed by the plastid reactions are the saturated fatty acid chains 16:0-ACP and 18:0-ACP, and monounsaturated 18:1-ACP (Miquel and Browse, 1995; Pidkowich et al.,

2007). Prior to export, the ACP subunits are removed by thioesterases, releasing fatty acid chains that are subsequently activated by acyl-CoAs (Kjellberg et al., 2000; Bates et al., 2007). Acyl-CoA molecules are exported to the endoplasmic reticulum (ER) where they can be modified by elongation and/or by the introduction of double bonds at different positions (Somerville and Browse, 1996; Benning et al., 2006; Benning, 2009; Li-Beisson et al., 2013; Hurlock et al., 2014). Elongation of 18:0 or 18:1 fatty acids occurs at the ER membrane through the sequential addition of malonyl-ACP derived from cytosolic citrate (Von Wettsetein-Knowles, 1982; Millar and Kunst, 1997; Allen et al., 2015). The longest fatty acids typically found in Arabidopsis seed oils are 24 carbons long, although 20 carbon chains are more prevalent (O'Neill et al., 2003; Li et al., 2006). In addition to elongation, fatty acids are also subjected to desaturation (up to three double bonds), with 18:2 and 18:3 being the most common polyunsaturated fatty acid in Arabidopsis ecotype Columbia-0 (Col-0) (O'Neill et al., 2003; Li et al., 2006).

Fatty acids are assembled into oils through several possible pathways, the simplest being that described by Kennedy (Kennedy, 1961). In this pathway, glycerol-3-phosphate derived from dihydroxyacetone phosphate is imported into the ER from the cytosol (Baud and Graham, 2006). Acyl-CoA molecules are sequentially attached to the glycerol-3-phosphate, first at the *sn*-1 position, followed by the *sn*-2 position, and finally at the *sn*-3 position (Li-Beisson et al., 2013). Once formed, oils are stored in small vesicles derived from the ER membrane called oil bodies (Robenek et al., 2004). The oil body membrane is a phospholipid monolayer, which allows the hydrophobic membrane lipid tails to associate with the oil molecules in the lumen (Yatsu and Jacks, 1972).The exact mechanism of oil body formation in seeds has not been determined, although one possibility is that oil molecules accumulate within the ER membrane, forming a rapidly enlarging oil body that eventually buds off into the cytosol (Murphy and Vance, 1999;

Baud et al., 2008; Hurlock et al., 2014). Embedded within the membrane are a class of proteins called oleosins, which, in addition to providing structural stability during desiccation, may assist in targeting oil to the site of oil body synthesis (Huang, 1992; Tzen and Huang, 1992; Tauchi-Sato et al., 2002; Jolivet et al., 2004; Siloto et al., 2006). The embryo contains approximately 80 – 90% of the total seed oil, and the majority of the oil is found in the cotyledons (Mansfield and Briarty, 1992; Penfield et al., 2004). The remainder of the oil is found in the endosperm and seed coat (Li et al., 2006).

1.4.3 Seed storage proteins

Arabidopsis seed storage proteins (SSPs) are encoded by two families of genes: the legumin-type 12S globulins (also called cruciferins) and the napin-type 2S albumins (also called arabins) (Krebbers et al., 1988; Sjodahl et al., 1991; van der Klei et al., 1993; Fujiwara et al., 2002; Li et al., 2007). SSPs are broken down during germination by specialized proteinases to provide amino acids for anabolic processes as well as energy production (Muntz et al., 2001; Angelovici et al., 2011; Weitbrecht et al., 2011). There are four known globulin genes (Pang et al., 1988; Gruis et al., 2002; Gruis et al., 2004) and five albumin genes (Krebbers et al., 1988; Guerche et al., 1990), and all are among the most highly expressed genes in developing Arabidopsis seeds (Schmid et al., 2005; Le et al., 2010). The cruciferin gene sequences lack similarity to each other, although they each encode propeptides that are proteolytically processed into six α subunits (between 25 and 35 kDa) and six β subunits (approximately 22 kDa). Similar to the cruciferins, each albumin gene encodes a propeptide that is proteolytically processed into a large (6.5 kDa) and a small subunit (2 kDa) stabilized by disulfide-linkages.

Expression of SSPs begins during the late embryogenesis period and quickly reaches very high levels (Schmid et al., 2005). Regulation of SSP expression is primarily at the transcriptional level through very similar *cis*-elements, which is discussed in more detail below (Kroj et al., 2003; Verdier and Thompson, 2008; Xi and Zheng, 2011). Once transcribed, the propeptides are synthesized on the rough ER and processed into mature SSPs through specialized proteases (Gruis et al., 2002; Gruis et al., 2004; Otegui et al., 2006). SSPs are transported to protein storage vacuoles in vesicles that bud off from the ER and accumulate in the cytosol close to the nucleus (Mansfield and Briarty, 1992; Muntz, 1998; Shimada et al., 2003). The majority of these vacuoles are found in the embryonic cotyledons and roots, and to a lesser extent in the endosperm (Sanders et al., 2009).

1.5 Regulators of seed development

Seed development is a highly regulated process requiring precisely timed phase transitions (Suzuki and McCarty, 2008; Jia et al., 2014; Lafon-Placette and Kohler, 2014; Radoeva and Weijers, 2014). Metabolic reactions occurring in the various cell types also require tight control to manage limited resources as seed metabolism is linked to development (Gutierrez et al., 2007). In contrast to metabolic events, the regulatory processes that specify the timing of transitions and other developmental milestones are less understood. For example, enzymes functioning in fatty acid metabolism are characterized, but not the factors governing when and in what tissue these enzymes become functional. For any given enzyme, several levels of regulation could be operational, from epigenetic and transcriptional to post-transcriptional, post-translational and allosteric (Thelen et al., 2000; Martin et al., 2006; Uhrig et al., 2008; Nodine and Bartel, 2010). The lack of a clear understanding of regulation extends from the metabolic

level to the whole organism/systems level. What is known about regulation at this level is that seed development is regulated, in part, by the combinatorial activity of hormones and gene expression regulators, which includes both transcription factors (TFs) and epigenetic/chromatin modifiers. To date, most of the experimental evidence has been performed on TFs, which has led to models of large transcriptional regulatory networks (Ruuska et al., 2002; Braybrook et al., 2006; Santos-Mendoza et al., 2008; Suzuki and McCarty, 2008; Swaminathan et al., 2008; Agarwal et al., 2011; Gao et al., 2012; Junker and Baumlein, 2012; Monke et al., 2012; Wang and Perry, 2013; Jia et al., 2014). Regulation by the hormones, chromatin modifiers and TFs inferred to date to regulate seed development will be discussed below.

1.5.1 Regulation of seed development by ABA

Progression through the phases of seed development is regulated principally through the action of the hormones abscisic acid (ABA) and gibberellic acid (GA). These two hormones have antagonistic actions: ABA promotes maturation and dormancy, while GA promotes germination (Holdsworth et al., 1999; Braybrook and Harada, 2008; Cutler et al., 2010; Finkelstein, 2013). The ratio of ABA to GA is critical for regulated seed development, as too much ABA relative to GA inhibits development, while too little enables precocious germination (Finkelstein et al., 2002; Nambara and Marion-Poll, 2005). ABA levels in seeds change during development, reaching high levels ("peaks") during two phases of seed development (Kanno et al., 2010). ABA is initially supplied maternally, and ABA levels peak during the transition from embryogenesis to maturation. This first peak enables seeds to undergo maturation, since eliminating the ABA stimulus by removing the embryo from the seed, or otherwise abolishing ABA sensitivity, results in germination (Raz et al., 2001; Brocard-Gifford et al., 2003; Vicente-

Carbajosa and Carbonero, 2005). ABA levels decline during maturation until increasing again during late maturation, forming the second peak. This ABA is likely generated within the seed and is required to induce dormancy (Kanno et al., 2010). The ABA/GA ratio is also a critical determinant of germination ability. ABA levels in dormant seeds are higher than those of GA, and these seeds remain dormant in the presence of germination stimulants, whereas GA levels are higher in non-dormant seeds (Graeber et al., 2012). In addition to ABA and GA, other hormones have roles in seed development, although their roles in maturation and dormancy are unclear. Auxins are primarily involved in establishing pattern formation and tissue differentiation during embryogenesis (Lau et al., 2012; Radoeva and Weijers, 2014). While the importance of ABA, GA and auxin are clear due to their profound effect on seed development, further work is needed to elucidate the likely and more subtle roles of the other major hormones (e.g., brassinosteroids, cytokinins) in seed development (Sun et al., 2010). For example, jasmonates and ethylene have somewhat antagonistic roles in germination, in that ethylene promotes while jasmonates inhibit germination (Linkies et al., 2009; Linkies and Leubner-Metzger, 2012). Evidence for their roles in seed development is not as well established.

1.5.2 Chromatin modifications in the seed

Epigenetic regulation consists of various mechanisms by which chromatin structure and, subsequently, gene expression, are controlled by modifications to DNA (such as methylation) or by post-translational modifications to histone proteins (Chen and Dent, 2014). Histone modifications change the structure of chromatin, which affects DNA-histone binding, the removal/replacement of histones or nucleosomes, or the sliding action of nucleosomes (Narlikar et al., 2013). These modifications alter the chromatin state, which has profound effects on gene

expression. Four types of histone proteins (H2A, H2B, H3 and H4) make up the core octamer around which DNA is wound. These histones are subject to several kinds of post-translational modifications (e.g., acetylation, ubiquitination, di- and tri-methylation), which affect the chromatin state by altering the tightness of DNA binding or serving as binding sites for other chromatin modifying enzymes (Chen and Dent, 2014). Many specific modifications to histones have been reported that generally indicate whether a particular gene is expressed. For example, genes for which the histone protein 3 is trimethylated at lysine residue 27 of (H3K27me3) are repressed, whereas H3K4me3 is associated with a gene that is being actively transcribed.

Chromatin remodeling enzymes are ubiquitous in eukaryotes. Chromodomain, helicase/ATPase and DNA-binding domain (CHD) proteins are a large family of chromatin modifiers with several subfamilies identified in plants (Murawska and Brehm, 2011). *PICKLE* (*PKL*) encodes a CHD3-like protein known to repress seed maturation regulators in developing seedlings by promoting H3K27me3, probably through interaction with polycomb group repressive complex 2 (PRC2) methyltransferases (Ogas et al., 1999; Aichinger et al., 2009). Aside from the data obtained with seedlings, very little is known about the epigenetic regulation of seed development itself. Chromatin compaction increases dramatically during the transition from embryogenesis to maturation reportedly by the activity of ABSISIC ACID INSENSITIVE 3 (ABI3) (van Zanten et al., 2011). This suggests that certain parts of the genome are "closed off" for expression during the maturation phase. Recently, the CHD1 protein CHR5 was reported to directly activate the expression of seed maturation regulators (Shen et al., 2015). Expression of the major seed maturation regulators LEAFY COTYLEDON 1 (LEC1), ABI3 and FUSCA 3 (FUS3) were reduced in the *chr5* mutant embryos, leading to reductions in SSP levels. However, the patterns of expression were unchanged and *chr5* seeds developed normally, suggesting that

CHR5 modulates gene expression as part of a larger activation mechanism. Furthermore, although direct binding of CHR5 to the promoters of the *ABI3* and *FUS3* genes was observed, no alterations in histone methylation were found in the *chr5* mutant. CHR5 may be involved in generating an active chromatin state in developing seeds, although much work remains to fully elucidate how this active state is maintained and regulated.

1.5.3 Global transcriptional regulators

The seed consists of three unique structures, each with their own developmental program, and development of a viable seed requires communication among them (Sun et al., 2010; Lafon-Placette and Kohler, 2014). The complete nature of these programs and the communication mechanism(s) are unknown and due to this complexity, it is often easiest to consider the seed as a whole unit. For example, the major known transcriptional regulators governing seed development (discussed below) are known to function in the embryo, while their activity in the endosperm is largely unknown. Furthermore, their transcript levels vary both spatially and temporally, indicating the activity of their protein products changes accordingly (To et al., 2006). Because of this, and especially because of our incomplete knowledge of specific target genes of these regulators in different tissues, the discussion of global regulators is presented in the context of *seed* development, not merely embryo development.

Known seed maturation transcriptional regulators form a complex network characterized by extensive mutual interactions (Gao et al., 2012; Roscoe et al., 2015). A group of global regulators are thought to form the "head" of this network and directly and indirectly regulate many of the same downstream genes. Recently, chromatin immunoprecipitation on DNA microarray (ChIP-chip) and genetic studies have identified putative regulons for each of these

global regulators, which include other TFs that function downstream of them (Braybrook et al., 2006; Junker et al., 2012; Monke et al., 2012; Wang and Perry, 2013). The four global regulators are FUS3, ABI3, LEC1 and LEC 2. FUS3, ABI3, and LEC2 are members of the plant-specific B3 DNA-binding domain TF family, while LEC1 contains a domain similar to the HAP3 subunit of CCAAT-box binding proteins (Giraudat et al., 1992; Meinke et al., 1994; Lotan et al., 1998; Luerssen et al., 1998; Stone et al., 2001; Lee et al., 2003; Gazzarrini et al., 2004; Holdsworth et al., 2008a). These four TFs were each initially discovered due to the profound effects on various aspects of seed development observed in mutant genotypes. In addition to substantial reductions in SSC levels, other alterations in seed-specific processes such as loss of dormancy (resulting in precocious germination) and loss of desiccation tolerance were observed in specific mutant alleles (Nambara et al., 1992; Keith et al., 1994; Meinke et al., 1994; West et al., 1994; Vicient et al., 2000).

The timing of expression for the global regulators provides some evidence of their likely functions when compared with their mutant phenotypes (Schmid et al., 2005; To et al., 2006; Le et al., 2010). *LEC1* expression begins during embryogenesis, and *LEC2* expression begins prior to the transition to maturation. Expression of *FUS3* and *ABI3* begin at the same time, and reach maximal levels during the middle-to-late maturation stages in Arabidopsis. *LEC1* expression peaks during embryogenesis and is likely an initiator of this developmental phase (Harada, 2001; Junker and Baumlein, 2012). Phenotypic analyses of *LEC1* null mutants suggest that *LEC1* specifies cotyledon identity, as these mutants form trichomes on cotyledons (Meinke, 1992; West et al., 1994). The presence of trichomes is typical for the epidermal cells of true leaves and their presence on embryonic cotyledons suggests possible cell identity changes. Consistent with a role in initiating embryogenesis, ectopic *LEC1* expression in seedlings results in the development of

embryo-like structures from vegetative tissues that accumulate SSCs (Lotan et al., 1998; Casson and Lindsey, 2006; Junker and Baumlein, 2012). Recently, it was determined from a ChIP-chip study that the LEC1 regulon includes the other global regulators and is also enriched in genes involved in cell elongation and auxin biosynthesis, particularly *YUCCA10* (Junker et al., 2012). Much like *LEC1*, null mutants of *LEC2* develop cotyledons with leafy characteristics (Meinke et al., 1994). Like the other global regulators, the *lec2* mutant displays reductions in SSC levels, although unlike the others, *lec2* is desiccation tolerant (Meinke et al., 1994; Stone et al., 2001; Braybrook and Harada, 2008). LEC1 activates *LEC2* expression, and transcript levels peak slightly later in seed development than *LEC1*, prior to the transition to seed maturation and LEC2 is probably involved in influencing this transition (Schmid et al., 2005; To et al., 2006). Like LEC1, LEC2 is capable of inducing formation of embryonic structures in vegetative tissues when ectopically expressed (Stone et al., 2008; Wojcikowska et al., 2013). ChIP-chip studies demonstrated that LEC2 targets both *ABI3* and *FUS3* in addition to several auxin biosynthesis genes, principally *YUCCA2* and *4* (Braybrook et al., 2006).

FUS3 expression is activated by LEC1 and LEC2 during embryogenesis, and through a combination of autoregulation and activation by ABI3, it reaches high levels during maturation (Kagaya et al., 2005a; Schmid et al., 2005; To et al., 2006; Monke et al., 2012; Wang and Perry, 2013). Similar to the LECs, *FUS3* null mutants form leafy cotyledons (Keith et al., 1994; Meinke et al., 1994) and show reductions in SSC levels (Parcy et al., 1997; Vicient et al., 2000; Kroj et al., 2003). The *fus3* mutant seeds are desiccation intolerant and capable of precocious germination, depending on the mutant allele (Keith et al., 1994; Tiedemann et al., 2008). The rapid increase in *FUS3* expression during maturation is consistent with its role as a central regulator of seed maturation, in which it integrates hormonal signals (primarily ABA) received

by the seed into transcriptional regulation (Gazzarrini et al., 2004; Kagaya et al., 2005b; Yamamoto et al., 2010). FUS3 is probably not a direct regulator of ABA signaling and *fus3* mutants do not have reductions in ABA sensitivity (Parcy et al., 1997; Kagaya et al., 2005b). In addition to ABA signal integration, FUS3 promotes maturation by repressing genes encoding GA-synthesizing enzymes (Curaba et al., 2004). Despite this demonstrated repression ability, FUS3 primarily functions as an activator and many embryogenesis- and seed-specific TF genes are included in the recently identified FUS3 regulon, including *FUS3* itself and *ABI3* (Wang and Perry, 2013).

ABI3 is directly involved in ABA signaling and mutants do not respond to exogenously applied ABA, resulting in desiccation intolerance, vivipary, and to reductions in SSC levels, depending on the severity of the mutant allele (Parcy et al., 1994; Parcy et al., 1997; Kroj et al., 2003). In addition to activation by the other three global regulators, ABI3 is also subjected to autoregulation, maintaining high expression levels when the two LECs are no longer expressed (To et al., 2006; Monke et al., 2012). As mentioned above, *FUS3* and *ABI3* have very similar expression patterns, although *ABI3* expression peaks later in seed development (at approximately 16 DAP) (Schmid et al., 2005; Le et al., 2010). In addition, the FUS3 and ABI3 regulons overlap to some extent, in that while they both regulate oil and SSP synthesis, FUS3 regulates oil synthesis genes more so than ABI3, while ABI3 regulates SSP gene expression to a greater extent than FUS3 (Roscoe et al., 2015). The ABI3 regulon includes many unique targets, including oleosins and numerous metabolic genes (Monke et al., 2012).

1.5.4 Additional TFs involved in regulating seed development

The global regulators discussed above are the major components of a much larger transcriptional network. A full description of all the known targets, TFs or otherwise, (Agarwal et al., 2011; Peng and Weselake, 2011; Gao et al., 2012; Jia et al., 2014) is outside the scope of this introduction. What follows are two brief examples of how the global regulators activate seed metabolism genes. Seed oil metabolism requires expression of many central carbon metabolism genes. WRINKLED1 (WRI1) is an AP2 family TF that is activated by both LEC1 and LEC2 which, in turn, activates expression of several glycolytic genes (Cernac and Benning, 2004; Baud et al., 2007; Junker and Baumlein, 2012; Junker et al., 2012). Null mutants have reduced expression of these genes, leading to substantial reductions in seed oil levels (Focks and Benning, 1998; Baud and Lepiniec, 2009). As another example, the basic leucine zipper TF *bZIP67* was recently found to regulate expression of the *FATTY ACID DESATURASE3* (*FAD3*) gene, which encodes an enzyme involved in generating 18:3 fatty acids (Mendes et al., 2013). bZIP67 forms a complex with the NUCLEAR FACTOR-YC2 (NF-YC2) and LEC1-LIKE (L1L) proteins to collectively activate *FAD3*. The loss of bZIP67 function results in the suppression of the *FAD3* gene, leading to reductions in the levels of 18:3 FA with no commensurate effects on total fatty acid levels. This transcriptional regulatory complex fits into the larger network through the activity of FUS3 and ABI3, which collectively activate expression of these three TFs.

1.6 Regulation of the seed maturation TF network by VAL transcriptional repressors

The TF network regulating seed development must be kept under tight control. Improper timing of expression can have substantial effects on downstream regulators and can disrupt the

timing of developmental transitions (Jia et al., 2014). In addition, maintaining expression in the proper physiological context is also essential. As mentioned above, ectopic expression of the global TF regulators can result in the formation of embryo-like proliferations in vegetative tissues. To date, several repressors of the seed maturation program, such as PKL and ARABIDOPSIS 6b-INTERACTING PROTEIN 1-Like 1 (ASIL1) that facilitate the transition from seed to seedling have been identified (Ogas et al., 1999; Gao et al., 2009; Gao et al., 2011). However none of these repressors have known functions in developing seeds. Since the timing of phase transitions is critical to proper seed development, it seems plausible that repressors would be functional during different stages of development to reduce the expression of "unneeded" genes. After all, major regulators such as LEC1 are repressed following the transition to seed maturation. Whether this is due to the activity of TFs that are primarily activators but can also function as repressors or TFs that are primarily repressors is unknown. One potential repressor of seed maturation genes is HIGH-LEVEL EXPRESSION OF SUGAR INDUCIBLE GENE 2/VIVIPAROUS1/ABI3-LIKE 1 (HSI2/VAL1, referred to as VAL1) and is the primary focus of this work.

1.6.1 Roles of VAL1 and VAL2 in Seedling Establishment

Several studies have demonstrated the importance of VAL1 and its close homolog VAL2 in repressing the seed maturation program during germination and seedling establishment (Suzuki et al., 2007; Tsukagoshi et al., 2007; Jia et al., 2013). The *VAL* genes are functionally redundant: while single knockouts have no obvious phenotype, knockouts in both genes are seedling lethal (Suzuki et al., 2007; Tsukagoshi et al., 2007). Both genes encode proteins of similar length and have the same four functional domains: a plant homeodomain-like (PHD-L)

region at the amino terminus, a B3 domain in the middle, a cysteine and tryptophan residue-containing (CW) domain slightly towards the carboxyl-terminus, and an ERF-associated amphiphilic repression (EAR) motif at the carboxyl-terminus (Suzuki et al., 2007; Jia et al., 2014). There is a third homolog, VAL3, that lacks the PHD-L domain, but it has not been functionally investigated to the extent of VAL1 and VAL2 (Jia et al., 2014).

VAL1 was originally identified through a forward genetics approach, as part of a screen of mutagenized leaves expressing both luciferase and β-glucuronidase under the control of a sugar responsive promoter (Tsukagoshi et al., 2005). These reporter genes were activated in the absence of sugar in the *val1* mutant background. Map-based cloning methods identified the mutation as a nonsense mutation in what is now known as the CW domain, which resulted in a truncated protein lacking the EAR motif. An insertion mutant in the seventh exon of the gene resulted in even greater reporter activity, whereas overexpression of *VAL1* resulted in loss of reporter gene activity. Subsequent work on VAL1 and VAL2 identified the critical regulatory activity of both TFs, as double knockout seeds germinated and developed shriveled seedlings that arrest after 7 – 9 days (Tsukagoshi et al., 2007). When grown in the presence of sucrose, *val1/val2* seedlings developed swollen hypocotyls and calli filled with SSCs. Analysis of global TF regulator (*LEC1*, *FUS3*, etc.) expression demonstrated a link between the concentration of exogenous sucrose and gene expression in the double mutant seedlings. In addition, increased sucrose concentration resulted in an increased *VAL1* and *VAL2* expression in the wild type seedlings. Levels of ABA in seven-day-old seedlings were found to increase in response to elevated sucrose exposure compared to the wild type seedlings. These results suggest that the exogenous sucrose triggered an increase in ABA levels, which, in turn, would result in a reversion to the embryonic state through activation of the seed maturation program. In the wild

type seedlings, the *VAL* genes would repress this program and allow germination to continue, but when both *VAL* genes are absent, the maturation program persists and results in seedling lethality. A microarray-based gene expression study in the *val1*/*val2* double mutant seedlings revealed over 800 genes that were greater than four fold upregulated, including three major seed maturation regulators (LEC1, LEC1-LIKE, FUS3, and ABI3) and several of their probable targets (Suzuki et al., 2007). These results clearly link the VAL function to repression of the seed maturation program in seedlings.

1.6.2 VAL1 functional domains and possible mechanisms of repression

Based on the analysis of known domains, VAL1 appears to have two potential mechanisms of targeting genes for repression. One mechanism is through a direct binding to DNA through the B3 domain, and the other is through binding to histone proteins. Direct binding to the RY/Sph motif has been reported for numerous other B3 domain-containing proteins (Suzuki et al., 1997; Reidt et al., 2000; Monke et al., 2004; Braybrook et al., 2006; Swaminathan et al., 2008). The presence of this domain in the VAL1 protein suggests that at least some of the genes regulated by this TF are targeted the same way. Analysis of the microarray data reported by Suzuki et al. (2007) demonstrated that 54% of the genes upregulated in the *val1*/*val2* seedlings have the RY/Sph motif in their promoters. The remainder of the upregulated genes could be targets of VAL1 through the domains discussed below. Despite the wealth of evidence regarding direct B3 domain binding to DNA, evidence for VAL binding *in planta* has not been demonstrated for either protein. Furthermore, neither the VAL1 and nor the VAL2 full-length protein were shown to bind to the DNA through a yeast one-hybrid approach but binding was demonstrated for FUS3 (Guerriero et al., 2009). However, when the B3 domain of VAL2 was

swapped with the same domain from FUS3, binding occurred, whereas the same approach with VAL1 failed. Therefore, attempts to demonstrate the binding of VAL to promoters have been unsuccessful, suggesting that VAL may require the formation of heterogeneous TF complexes for its action, as has been proposed for ABI3 (Nakamura et al., 2001), LEC1 (Yamamoto et al., 2009), and L1L (Mendes et al., 2013).

VAL1 contains a domain very similar to the plant homeodomain (PHD) and a CW domain, which are both implicated in recognizing active gene expression H3K4me3 marks (Bienz, 2006; He et al., 2010; Hoppmann et al., 2011; Sanchez and Zhou, 2011; Mouriz et al., 2015). Proteins containing PHD domains are found in all eukaryotes, where they mediate binding to chromatin through conserved zinc finger motifs and may coordinate interactions of histone-modifying enzymes that bring about repression of gene expression (Lee et al., 2009). The PHD-like (PHD-L) domain of VAL1 consists of three potential zinc fingers, as opposed to two in the canonical protein models (Suzuki et al., 2007; Sanchez and Zhou, 2011; Mouriz et al., 2015). Disruption of the PHD-L domain through a missense mutation led to moderately increased expression of several seed-specific genes in mutant seedlings, including *AGAMOUS-LIKE 15* (*AGL15*), *FUS3* and several SSP genes (Veerappan et al., 2012). These genes are likely regulated by VAL1 alone, as the expression level was unchanged when this missense mutant was crossed with a *val2* mutant. ChIP analysis indicated that the global regulator genes are enriched with H3K27me3 marks in the wild type seedlings and ABI3 and LEC1 showed reductions in H3K27me3 in the *val1* knockout mutant seedlings (Veerappan et al., 2014). These results suggest that VAL1 targets several seed-specific genes through the PHD-L domain and coordinates repression through an unidentified histone methyltransferase.

VAL1 also contains a domain consisting of conserved cysteine and tryptophan (CW) residues that form a zinc finger, resembling the PHD domain (He et al., 2010). The role that this domain has in VAL1 is less well understood than the PHD-L domain. Hoppmann et al. (2011) demonstrated that the VAL1 CW domain binds directly to H3K4me3. The CW domain of VAL2 was shown to interact directly with HISTONE DEACETYLASE 19 (HDA19), an epigenetic regulator, while no such interaction was detected for VAL1 (Zhou et al., 2013).

The VAL1 carboxyl-terminus contains a short ethylene-responsive element binding factor-associated amphiphilic repression (EAR) motif (Tsukagoshi et al., 2005). The VAL1 EAR motif is reported to recruit histone modifiers in a similar fashion. Disruption of the VAL1 EAR motif through a nonsense mutation resulted in loss of regulation of a sugar responsive reporter construct, as the reporter was active in the absence of sucrose (Tsukagoshi et al., 2005). Furthermore, deletion of the EAR motif resulted in elevated expression of the reporter construct in protoplasts. The EAR motif is ubiquitous amongst repressors of gene expression and many are reported to interact with co-repressors (Kagale and Rozwadowski, 2011). The model that has emerged is one in which a DNA binding protein with an EAR motif coordinates co-repressors such as histone deacetylases (HDACs) to the gene for epigenetic repression by removal of acetyl groups. One example of repression in this fashion is the interaction of ETHYLENE REPONSE FACTOR 7 (ERF7) with SWI-INDEPENDENT 3 (SIN3), which is a member of a conserved family of scaffolding proteins (Song et al., 2005; Grzenda et al., 2009). SIN3 in turn interacts with the HDAC HDA19 to repress gene expression.

VAL1 was reported to interact with SIN3-LIKE 1 (SNL1) in a yeast-two-hybrid screen, although these authors did not follow up on this interaction (Bowen et al., 2010). In this study, repression of a reporter construct required HDAC activity in SNL1 transfected protoplasts,

suggesting that SNL1 interacts with HDAC(s) to repress target genes. Furthermore, SNL1 was found to interact directly with HDA19 through both yeast-two-hybrid and bi-molecular fluorescence complementation, and the loss of function SNL1 mutants had increased histone acetylation (Wang et al., 2013). Interestingly this is the same HDAC, with which the VAL2 is reported to interact through the CW domain. These results suggest that VAL1 targets histone modifiers through the interaction with SNL1. However, the interaction between VAL1 and SNL1 still remains to be demonstrated *in planta*.

The available evidence suggests a model in which VAL1 targets histone modifiers to the target genes through two possible binding mechanisms. In one mechanism, genes are targeted through VAL1 binding to the RY/Sph motif in the promoter regions through the VAL1 B3 domain. In the other, VAL1 targets genes for repression through binding to active chromatin (H3K4me3) through either the PHD-L or the CW domain. In either case, VAL1 recruits histone modifiers through the EAR motif to the gene for repression through the removal of H3K4me3 and the addition of H3K27me3. Although evidence for VAL1 interacting with HDA19 through SNL1 is strong, *VAL1* repression through histone deacetylation has not been demonstrated.

In addition to the VAL1 involvement in histone trimethylation reported by Veerappan et al. (2014), VAL1 and 2 are also thought to be involved in ubiquitination of histone protein 2 subunit A (H2Aub), a repressive histone mark catalyzed by polycomb group repressive complex group 1 (PRC1) proteins (Yang et al., 2013). Triple mutant seedlings of the PRC1 genes BMI1 A/B/C have the same phenotype as the *val1*/*val2* double mutant, and VAL1 was found to bind with the BMI1 proteins through co-immunoprecipitation, although the binding domain is unknown. In addition, H2Aub and H3K27me3 levels in the promoters of seed maturation genes were reduced in the *val1*/*val2* double mutant and BMI1 triple mutants. Taken together, these

results suggest a model in which the VAL proteins target genes for repression either through binding to the RY/Sph motif through the B3 domain, or through recognition of H3K4me3 through the PHD-L domain. Then, VAL forms a complex with PRC1 proteins to target genes for repression by H2Aub followed by recruitment of a PRC2 class methyltransferase for H3K27me3 addition. Whether VAL1 functions in this manner in seeds remains to be determined.

References

Agarwal, P., Kapoor, S., and Tyagi, A.K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. BioEssays : news and reviews in molecular, cellular and developmental biology **33,** 189-202.

Aichinger, E., Villar, C.B., Farrona, S., Reyes, J.C., Hennig, L., and Kohler, C. (2009). CHD3 proteins and polycomb group proteins antagonistically determine cell identity in Arabidopsis. PLoS genetics **5,** e1000605.

Allen, D.K., Bates, P.D., and Tjellstrom, H. (2015). Tracking the metabolic pulse of plant lipid production with isotopic labeling and flux analyses: Past, present and future. Progress in lipid research.

Angelovici, R., Galili, G., Fernie, A.R., and Fait, A. (2010). Seed desiccation: a bridge between maturation and germination. Trends in plant science **15,** 211-218.

Angelovici, R., Fait, A., Fernie, A.R., and Galili, G. (2011). A seed high-lysine trait is negatively associated with the TCA cycle and slows down Arabidopsis seed germination. The New phytologist **189,** 148-159.

Baroux, C., Spillane, C., and Grossniklaus, U. (2002). Evolutionary origins of the endosperm in flowering plants. Genome biology **3,** reviews1026.

Bates, P.D., Ohlrogge, J.B., and Pollard, M. (2007). Incorporation of newly synthesized fatty acids into cytosolic glycerolipids in pea leaves occurs via acyl editing. The Journal of biological chemistry **282,** 31206-31216.

Baud, S., and Graham, I.A. (2006). A spatiotemporal analysis of enzymatic activities associated with carbon metabolism in wild-type and mutant embryos of Arabidopsis using in situ histochemistry. The Plant journal : for cell and molecular biology **46,** 155-169.

Baud, S., and Lepiniec, L. (2009). Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis. Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale **47,** 448-455.

Baud, S., and Lepiniec, L. (2010). Physiological and developmental regulation of seed oil production. Progress in lipid research **49,** 235-249.

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

Baud, S., Dubreucq, B., Miquel, M., Rochat, C., and Lepiniec, L. (2008). Storage reserve accumulation in Arabidopsis: metabolic and developmental control of seed filling. The Arabidopsis book / American Society of Plant Biologists **6,** e0113.

Baud, S., Mendoza, M.S., To, A., Harscoet, E., Lepiniec, L., and Dubreucq, B. (2007). WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. The Plant journal : for cell and molecular biology **50,** 825-838.

Baud, S., Wuilleme, S., Lemoine, R., Kronenberger, J., Caboche, M., Lepiniec, L., and Rochat, C. (2005). The AtSUC5 sucrose transporter specifically expressed in the endosperm is involved in early seed development in Arabidopsis. The Plant journal : for cell and molecular biology **43,** 824-836.

Benning, C. (2009). Mechanisms of lipid transport involved in organelle biogenesis in plant cells. Annu Rev Cell Dev Biol **25,** 71-91.

Benning, C., Xu, C., and Awai, K. (2006). Non-vesicular and vesicular lipid trafficking involving plastids. Current opinion in plant biology **9,** 241-247.

Bentsink, L., and Koornneef, M. (2008). Seed dormancy and germination. The Arabidopsis book / American Society of Plant Biologists **6,** e0119.

Berger, F., Grini, P.E., and Schnittger, A. (2006). Endosperm: an integrator of seed growth and development. Current opinion in plant biology **9,** 664-670.

Bewley, J.D. (1997). Seed Germination and Dormancy. Plant Cell **9,** 1055-1066.

Bewley, J.D., and Black, M. (1994). Seeds: Physiology and Development. (New York: Plenum).

Bienz, M. (2006). The PHD finger, a nuclear protein-interaction domain. Trends in biochemical sciences **31,** 35-40.

Bowen, A.J., Gonzalez, D., Mullins, J.G., Bhatt, A.M., Martinez, A., and Conlan, R.S. (2010). PAH-domain-specific interactions of the Arabidopsis transcription coregulator SIN3-LIKE1 (SNL1) with telomere-binding protein 1 and ALWAYS EARLY2 Myb-DNA binding factors. Journal of molecular biology **395,** 937-949.

Braybrook, S.A., and Harada, J.J. (2008). LECs go crazy in embryo development. Trends in plant science **13,** 624-630.

Braybrook, S.A., Stone, S.L., Park, S., Bui, A.Q., Le, B.H., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (2006). Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo

maturation and somatic embryogenesis. Proceedings of the National Academy of Sciences of the United States of America **103,** 3468-3473.

Brocard-Gifford, I.M., Lynch, T.J., and Finkelstein, R.R. (2003). Regulatory networks in seeds integrating developmental, abscisic acid, sugar, and light signaling. Plant physiology **131,** 78-92.

Brown, A.P., Affleck, V., Fawcett, T., and Slabas, A.R. (2006). Tandem affinity purification tagging of fatty acid biosynthetic enzymes in Synechocystis sp. PCC6803 and Arabidopsis thaliana. Journal of experimental botany **57,** 1563-1571.

Buitink, J., and Leprince, O. (2008). Intracellular glasses and seed survival in the dry state. C R Biol **331,** 788-795.

Capron, A., Chatfield, S., Provart, N., and Berleth, T. (2009). Embryogenesis: pattern formation from a single cell. The Arabidopsis book / American Society of Plant Biologists **7,** e0126.

Casson, S.A., and Lindsey, K. (2006). The turnip mutant of Arabidopsis reveals that LEAFY COTYLEDON1 expression mediates the effects of auxin and sugars to promote embryonic cell identity. Plant Physiol **142,** 526-541.

Cernac, A., and Benning, C. (2004). WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. The Plant journal : for cell and molecular biology **40,** 575-585.

Chen, L.Q., Lin, I.W., Qu, X.Q., Sosso, D., McFarlane, H.E., Londono, A., Samuels, A.L., and Frommer, W.B. (2015). A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the Arabidopsis embryo. Plant Cell **27,** 607-619.

Chen, T., and Dent, S.Y. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. Nature reviews. Genetics **15,** 93-106.

Curaba, J., Moritz, T., Blervaque, R., Parcy, F., Raz, V., Herzog, M., and Vachon, G. (2004). AtGA3ox2, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in Arabidopsis. Plant Physiol **136,** 3660-3669.

Cutler, S.R., Rodriguez, P.L., Finkelstein, R.R., and Abrams, S.R. (2010). Abscisic acid: emergence of a core signaling network. Annual review of plant biology **61,** 651-679.

de Jong, A., Koerselman-Kooij, J.W., Schuurmans, J., and Borstlap, A.C. (1997). The mechanism of amino Acid efflux from seed coats of developing pea seeds as revealed by uptake experiments. Plant Physiol **114,** 731-736.

Debeaujon, I., Leon-Kloosterziel, K.M., and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. Plant Physiol **122,** 403-414.

Debeaujon, I., Peeters, A.J., Leon-Kloosterziel, K.M., and Koornneef, M. (2001). The TRANSPARENT TESTA12 gene of Arabidopsis encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. Plant Cell **13,** 853-871.

Dekkers, B.J., Costa, M.C., Maia, J., Bentsink, L., Ligterink, W., and Hilhorst, H.W. (2015). Acquisition and loss of desiccation tolerance in seeds: from experimental model to biological relevance. Planta **241,** 563-577.

Dekkers, B.J., Pearce, S., van Bolderen-Veldkamp, R.P., Marshall, A., Widera, P., Gilbert, J., Drost, H.G., Bassel, G.W., Muller, K., King, J.R., Wood, A.T., Grosse, I., Quint, M., Krasnogor, N., Leubner-Metzger, G., Holdsworth, M.J., and Bentsink, L. (2013). Transcriptional dynamics of two seed compartments with opposing roles in Arabidopsis seed germination. Plant Physiol **163,** 205-215.

Dumas, C., and Rogowsky, P. (2008). Fertilization and early seed formation. Comptes rendus biologies **331,** 715-725.

Eastmond, P.J. (2004). Glycerol-insensitive Arabidopsis mutants: gli1 seedlings lack glycerol kinase, accumulate glycerol and are more resistant to abiotic stress. The Plant journal : for cell and molecular biology **37,** 617-625.

Fait, A., Angelovici, R., Less, H., Ohad, I., Urbanczyk-Wochniak, E., Fernie, A.R., and Galili, G. (2006). Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant Physiol **142,** 839-854.

Finkelstein, R. (2013). Abscisic Acid synthesis and response. The Arabidopsis book / American Society of Plant Biologists **11,** e0166.

Finkelstein, R.R., Gampala, S.S., and Rock, C.D. (2002). Abscisic acid signaling in seeds and seedlings. Plant Cell **14 Suppl,** S15-45.

Focks, N., and Benning, C. (1998). wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. Plant Physiol **118,** 91-101.

Footitt, S., Douterelo-Soler, I., Clay, H., and Finch-Savage, W.E. (2011). Dormancy cycling in Arabidopsis seeds is controlled by seasonally distinct hormone-signaling pathways. Proceedings of the National Academy of Sciences of the United States of America **108,** 20236-20241.

Friedman, W.E. (2001). Developmental and evolutionary hypotheses for the origin of double fertilization and endosperm. Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie **324,** 559-567.

Fujiwara, T., Nambara, E., Yamagishi, K., Goto, D.B., and Naito, S. (2002). Storage Proteins. The Arabidopsis book / American Society of Plant Biologists **1,** e0020.

Gao, M.J., Gropp, G., Wei, S., Hegedus, D.D., and Lydiate, D.J. (2012). Combinatorial Networks Regulating Seed Development and Seed Filling. In Embryogenesis, K.-I. Sato, ed (InTech), pp. 189-228.

Gao, M.J., Lydiate, D.J., Li, X., Lui, H., Gjetvaj, B., Hegedus, D.D., and Rozwadowski, K. (2009). Repression of seed maturation genes by a trihelix transcriptional repressor in Arabidopsis seedlings. Plant Cell **21,** 54-71.

Gao, M.J., Li, X., Lui, H., Gropp, G.M., Lydiate, D.D., Wei, S., and Hegedus, D.D. (2011). ASIL1 is required for proper timing of seed filling in Arabidopsis. Plant signaling & behavior **6,** 1886-1888.

Gazzarrini, S., Tsuchiya, Y., Lumba, S., Okamoto, M., and McCourt, P. (2004). The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones gibberellin and abscisic acid. Developmental cell **7,** 373-385.

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F., and Goodman, H.M. (1992). Isolation of the Arabidopsis ABI3 gene by positional cloning. The Plant cell **4,** 1251-1261.

Goldberg, R.B., de Paiva, G., and Yadegari, R. (1994). Plant embryogenesis: zygote to seed. Science **266,** 605-614.

Graeber, K., Nakabayashi, K., Miatton, E., Leubner-Metzger, G., and Soppe, W.J. (2012). Molecular mechanisms of seed dormancy. Plant, cell & environment **35,** 1769-1786.

Graham, I.A. (2008). Seed storage oil mobilization. Annual review of plant biology **59,** 115-142.

Gruis, D., Schulze, J., and Jung, R. (2004). Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases. Plant Cell **16,** 270-290.

Gruis, D.F., Selinger, D.A., Curran, J.M., and Jung, R. (2002). Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases. Plant Cell **14,** 2863-2882.

Grzenda, A., Lomberk, G., Zhang, J.S., and Urrutia, R. (2009). Sin3: master scaffold and transcriptional corepressor. Biochimica et biophysica acta **1789,** 443-450.

Guerche, P., Tire, C., De Sa, F.G., De Clercq, A., Van Montagu, M., and Krebbers, E. (1990). Differential Expression of the Arabidopsis 2S Albumin Genes and the Effect of Increasing Gene Family Size. Plant Cell **2,** 469-478.

Guerriero, G., Martin, N., Golovko, A., Sundstrom, J.F., Rask, L., and Ezcurra, I. (2009). The RY/Sph element mediates transcriptional repression of maturation genes from late maturation to early seedling growth. The New phytologist **184,** 552-565.

Gutierrez, L., Van Wuytswinkel, O., Castelain, M., and Bellini, C. (2007). Combined networks regulating seed maturation. Trends in plant science **12,** 294-300.

Harada, J.J. (2001). Role of *Arabidopsis LEAFY COTYLEDON* genes in seed development. J Plant Physiol **158,** 405-409.

Haughn, G., and Chaudhury, A. (2005). Genetic analysis of seed coat development in Arabidopsis. Trends Plant Sci **10,** 472-477.

He, F., Umehara, T., Saito, K., Harada, T., Watanabe, S., Yabuki, T., Kigawa, T., Takahashi, M., Kuwasako, K., Tsuda, K., Matsuda, T., Aoki, M., Seki, E., Kobayashi, N., Guntert, P., Yokoyama, S., and Muto, Y. (2010). Structural insight into the zinc finger CW domain as a histone modification reader. Structure **18,** 1127-1139.

He, H., de Souza Vidigal, D., Snoek, L.B., Schnabel, S., Nijveen, H., Hilhorst, H., and Bentsink, L. (2014). Interaction between parental environment and genotype affects plant and seed performance in Arabidopsis. Journal of experimental botany **65,** 6603-6615.

Hehenberger, E., Kradolfer, D., and Kohler, C. (2012). Endosperm cellularization defines an important developmental transition for embryo development. Development **139,** 2031-2039.

Hill, L.M., Morley-Smith, E.R., and Rawsthorne, S. (2003). Metabolism of sugars in the endosperm of developing seeds of oilseed rape. Plant Physiol **131,** 228-236.

Hills, M.J. (2004). Control of storage-product synthesis in seeds. Current opinion in plant biology **7,** 302-308.

Hoekstra, F.A., Golovina, E.A., and Buitink, J. (2001). Mechanisms of plant desiccation tolerance. Trends Plant Sci **6,** 431-438.

Holdsworth, M., Kurup, S., and McKibbin, R. (1999). Molecular and genetic mechanisms regulating the transition from embryo development to germination. Trends in plant science **4,** 275-280.

Holdsworth, M.J., Bentsink, L., and Soppe, W.J. (2008a). Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. The New phytologist **179,** 33-54.

Holdsworth, M.J., Finch-Savage, W.E., Grappin, P., and Job, D. (2008b). Post-genomics dissection of seed dormancy and germination. Trends in plant science **13,** 7-13.

Hoppmann, V., Thorstensen, T., Kristiansen, P.E., Veiseth, S.V., Rahman, M.A., Finne, K., Aalen, R.B., and Aasland, R. (2011). The CW domain, a new histone recognition module in chromatin proteins. EMBO J **30,** 1939-1952.

Huang, A.H.C. (1992). Oil bodies and oleosins in seeds. Annu. Rev. Plant Physiol. **43,** 177-200.

Hurlock, A.K., Roston, R.L., Wang, K., and Benning, C. (2014). Lipid trafficking in plant cells. Traffic **15,** 915-932.

Jenik, P.D., Gillmor, C.S., and Lukowitz, W. (2007). Embryonic patterning in Arabidopsis thaliana. Annu Rev Cell Dev Biol **23,** 207-236.

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

Jia, H., Suzuki, M., and McCarty, D.R. (2014). Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. WIREs Dev Biol **3,** 135-145.

Jolivet, P., Roux, E., D'Andrea, S., Davanture, M., Negroni, L., Zivy, M., and Chardot, T. (2004). Protein composition of oil bodies in Arabidopsis thaliana ecotype WS. Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale **42,** 501-509.

Junker, A., and Baumlein, H. (2012). Multifunctionality of the LEC1 transcription factor during plant development. Plant signaling & behavior **7,** 1718-1720.

Junker, A., Monke, G., Rutten, T., Keilwagen, J., Seifert, M., Thi, T.M., Renou, J.P., Balzergue, S., Viehover, P., Hahnel, U., Ludwig-Muller, J., Altschmied, L., Conrad, U., Weisshaar, B., and Baumlein, H. (2012). Elongation-related functions of LEAFY COTYLEDON1 during the development of Arabidopsis thaliana. The Plant journal : for cell and molecular biology **71,** 427-442.

Jurgens, G. (2001). Apical-basal pattern formation in Arabidopsis embryogenesis. EMBO J **20,** 3609-3616.

Kagale, S., and Rozwadowski, K. (2011). EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression. Epigenetics : official journal of the DNA Methylation Society **6,** 141-146.

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

Kagaya, Y., Okuda, R., Ban, A., Toyoshima, R., Tsutsumida, K., Usui, H., Yamamoto, A., and Hattori, T. (2005b). Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in Arabidopsis. Plant & cell physiology **46,** 300-311.

Kanno, Y., Jikumaru, Y., Hanada, A., Nambara, E., Abrams, S.R., Kamiya, Y., and Seo, M. (2010). Comprehensive hormone profiling in developing Arabidopsis seeds: examination of the site of ABA biosynthesis, ABA transport and hormone interactions. Plant & cell physiology **51,** 1988-2001.

Kawashima, T., and Goldberg, R.B. (2010). The suspensor: not just suspending the embryo. Trends in plant science **15,** 23-30.

Keith, K., Kraml, M., Dengler, N.G., and McCourt, P. (1994). fusca3: A Heterochronic Mutation Affecting Late Embryo Development in Arabidopsis. Plant Cell **6,** 589-600.

Kelly, A.A., Quettier, A.L., Shaw, E., and Eastmond, P.J. (2011). Seed storage oil mobilization is important but not essential for germination or seedling establishment in Arabidopsis. Plant Physiol **157,** 866-875.

Kennedy, E.P. (1961). Biosynthesis of complex lipids. Federation proceedings **20,** 934-940.

Kim, I., and Zambryski, P.C. (2005). Cell-to-cell communication via plasmodesmata during Arabidopsis embryogenesis. Current opinion in plant biology **8,** 593-599.

Kim, I., Kobayashi, K., Cho, E., and Zambryski, P.C. (2005). Subdomains for transport via plasmodesmata corresponding to the apical-basal axis are established during Arabidopsis embryogenesis. Proceedings of the National Academy of Sciences of the United States of America **102,** 11945-11950.

Kjellberg, J.M., Trimborn, M., Andersson, M., and Sandelius, A.S. (2000). Acyl-CoA dependent acylation of phospholipids in the chloroplast envelope. Biochimica et biophysica acta **1485,** 100-110.

Konishi, T., Shinohara, K., Yamada, K., and Sasaki, Y. (1996). Acetyl-CoA carboxylase in higher plants: most plants other than gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. Plant & cell physiology **37,** 117-122.

Kranner, I., and Birtic, S. (2005). A modulating role for antioxidants in desiccation tolerance. Integrative and comparative biology **45,** 734-740.

Krebbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans, J., Van Damme, J., Segura, M., Gheysen, G., Van Montagu, M., and Vandekerckhove, J. (1988). Determination of the Processing Sites of an Arabidopsis 2S Albumin and Characterization of the Complete Gene Family. Plant Physiol **87,** 859-866.

Kroj, T., Savino, G., Valon, C., Giraudat, J., and Parcy, F. (2003). Regulation of storage protein gene expression in Arabidopsis. Development **130,** 6065-6073.

Ladwig, F., Stahl, M., Ludewig, U., Hirner, A.A., Hammes, U.Z., Stadler, R., Harter, K., and Koch, W. (2012). Siliques are Red1 from Arabidopsis acts as a bidirectional amino acid transporter that is crucial for the amino acid homeostasis of siliques. Plant physiology **158,** 1643-1655.

Lafon-Placette, C., and Kohler, C. (2014). Embryo and endosperm, partners in seed development. Current opinion in plant biology **17,** 64-69.

Lau, S., Slane, D., Herud, O., Kong, J., and Jurgens, G. (2012). Early embryogenesis in flowering plants: setting up the basic body pattern. Annual review of plant biology **63,** 483-506.

Le, B.H., Cheng, C., Bui, A.Q., Wagmaister, J.A., Henry, K.F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S., Drews, G.N., Fischer, R.L., Okamuro, J.K., Harada, J.J., and Goldberg, R.B. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proceedings of the National Academy of Sciences of the United States of America **107,** 8063-8070.

Lee, H., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (2003). Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. Proceedings of the National Academy of Sciences of the United States of America **100,** 2152-2156.

Lee, W.Y., Lee, D., Chung, W.I., and Kwon, C.S. (2009). Arabidopsis ING and Alfin1-like protein families localize to the nucleus and bind to H3K4me3/2 via plant homeodomain fingers. The Plant journal : for cell and molecular biology **58,** 511-524.

Lemoine, R., La Camera, S., Atanassova, R., Dedaldechamp, F., Allario, T., Pourtau, N., Bonnemain, J.L., Laloi, M., Coutos-Thevenot, P., Maurousset, L., Faucher, M., Girousse, C., Lemonnier, P., Parrilla, J., and Durand, M. (2013). Source-to-sink transport of sugar and regulation by environmental factors. Frontiers in plant science **4,** 272.

Lepiniec, L., Debeaujon, I., Routaboul, J.M., Baudry, A., Pourcel, L., Nesi, N., and Caboche, M. (2006). Genetics and biochemistry of seed flavonoids. Annual review of plant biology **57,** 405-430.

Leprince, O., and Buitink, J. (2010). Desiccation tolerance: From genomics to the field. Plant Science **179,** 554-564.

Leprince, O., van Aelst, A.C., Pritchard, H.W., and Murphy, D.J. (1998). Oleosins prevent oil-body coalescence during seed imbibition as suggested by a low-temperature scanning electron microscope study of dessication-tolerant and -sensitive oilseeds. Planta **204,** 109-119.

Li, J., and Berger, F. (2012). Endosperm: food for humankind and fodder for scientific discoveries. The New phytologist **195,** 290-305.

Li, Q., Wang, B.C., Xu, Y., and Zhu, Y.X. (2007). Systematic studies of 12S seed storage protein accumulation and degradation patterns during Arabidopsis seed maturation and early seedling germination stages. Journal of biochemistry and molecular biology **40,** 373-381.

Li, Y., Beisson, F., Pollard, M., and Ohlrogge, J. (2006). Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. Phytochemistry **67,** 904-915.

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., and Ohlrogge, J. (2013). Acyl-lipid metabolism. The Arabidopsis book / American Society of Plant Biologists **8,** e0133.

Linkies, A., and Leubner-Metzger, G. (2012). Beyond gibberellins and abscisic acid: how ethylene and jasmonates control seed germination. Plant cell reports **31,** 253-270.

Linkies, A., Muller, K., Morris, K., Tureckova, V., Wenk, M., Cadman, C.S., Corbineau, F., Strnad, M., Lynn, J.R., Finch-Savage, W.E., and Leubner-Metzger, G. (2009). Ethylene interacts with abscisic acid

to regulate endosperm rupture during germination: a comparative approach using Lepidium sativum and Arabidopsis thaliana. The Plant cell **21,** 3803-3822.

Lotan, T., Ohto, M., Yee, K.M., West, M.A., Lo, R., Kwong, R.W., Yamagishi, K., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (1998). Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell **93,** 1195-1205.

Luerssen, H., Kirik, V., Herrmann, P., and Misera, S. (1998). FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. The Plant journal : for cell and molecular biology **15,** 755-764.

Mansfield, S.G., and Briarty, L.G. (1992). Cotyledon Cell-Development in Arabidopsis-Thaliana during Reserve Deposition. Can J Bot **70,** 151-164.

Martin, R.C., Liu, P.P., and Nonogaki, H. (2006). MicroRNAs in seeds: Modified detection techniques and potential applications. Can J Bot **84,** 189-198.

Meinke, D.W. (1992). A Homoeotic Mutant of Arabidopsis thaliana with Leafy Cotyledons. Science **258,** 1647-1650.

Meinke, D.W., Franzmann, L.H., Nickle, T.C., and Yeung, E.C. (1994). Leafy Cotyledon Mutants of Arabidopsis. Plant Cell **6,** 1049-1064.

Mendes, A., Kelly, A.A., van Erp, H., Shaw, E., Powers, S.J., Kurup, S., and Eastmond, P.J. (2013). bZIP67 regulates the omega-3 fatty acid content of Arabidopsis seed oil by activating fatty acid desaturase3. Plant Cell **25,** 3104-3116.

Millar, A.A., and Kunst, L. (1997). Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. The Plant journal : for cell and molecular biology **12,** 121-131.

Miquel, M., and Browse, J. (1995). Lipid biosynthesis in developing seeds. In Seed Development and Germination, J. Kigel and G. Galili, eds (New York: Dekker), pp. 169-193.

Monke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H.P., Baumlein, H., and Conrad, U. (2004). Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. Planta **219,** 158-166.

Monke, G., Seifert, M., Keilwagen, J., Mohr, M., Grosse, I., Hahnel, U., Junker, A., Weisshaar, B., Conrad, U., Baumlein, H., and Altschmied, L. (2012). Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. Nucleic acids research **40,** 8240-8254.

Morley-Smith, E.R., Pike, M.J., Findlay, K., Kockenberger, W., Hill, L.M., Smith, A.M., and Rawsthorne, S. (2008). The transport of sugars to developing embryos is not via the bulk endosperm in oilseed rape seeds. Plant Physiol **147,** 2121-2130.

Mouriz, A., Lopez-Gonzalez, L., Jarillo, J.A., and Pineiro, M. (2015). PHDs govern plant development. Plant Signaling & Behavior**,** 10, e993253.

Muntz, K. (1998). Deposition of storage proteins. Plant Mol Biol **38,** 77-99.

Muntz, K., Belozersky, M.A., Dunaevsky, Y.E., Schlereth, A., and Tiedemann, J. (2001). Stored proteinases and the initiation of storage protein mobilization in seeds during germination and seedling growth. Journal of experimental botany **52,** 1741-1752.

Murawska, M., and Brehm, A. (2011). CHD chromatin remodelers and the transcription cycle. Transcription **2,** 244-253.

Murphy, D.J., and Vance, J. (1999). Mechanisms of lipid-body formation. Trends in biochemical sciences **24,** 109-115.

Nakamura, S., Lynch, T.J., and Finkelstein, R.R. (2001). Physical interactions between ABA response loci of Arabidopsis. Plant J **26,** 627-635.

Nambara, E., and Marion-Poll, A. (2005). Abscisic acid biosynthesis and catabolism. Annual review of plant biology **56,** 165-185.

Nambara, E., Naito, S., and Mccourt, P. (1992). A Mutant of Arabidopsis Which Is Defective in Seed Development and Storage Protein Accumulation Is a New Abi3 Allele. Plant Journal **2,** 435-441.

Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. Cell **154,** 490-503.

Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell **13,** 2099-2114.

Nikolau, B.J., Ohlrogge, J.B., and Wurtele, E.S. (2003). Plant biotin-containing carboxylases. Archives of biochemistry and biophysics **414,** 211-222.

Nodine, M.D., and Bartel, D.P. (2010). MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes & development **24,** 2678-2692.

O'Neill, C.M., Gill, S., Hobbs, D., Morgan, C., and Bancroft, I. (2003). Natural variation for seed oil composition in Arabidopsis thaliana. Phytochemistry **64,** 1077-1090.

Ogas, J., Kaufmann, S., Henderson, J., and Somerville, C. (1999). PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. Proc Natl Acad Sci U S A **96,** 13839-13844.

Okamoto, M., Kuwahara, A., Seo, M., Kushiro, T., Asami, T., Hirai, N., Kamiya, Y., Koshiba, T., and 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,** 97-107.

Okumoto, S., and Pilot, G. (2011). Amino acid export in plants: a missing link in nitrogen cycling. Molecular plant **4,** 453-463.

Olsen, O.A. (2001). ENDOSPERM DEVELOPMENT: Cellularization and Cell Fate Specification. Annu Rev Plant Physiol Plant Mol Biol **52,** 233-267.

Otegui, M.S., Herder, R., Schulze, J., Jung, R., and Staehelin, L.A. (2006). The proteolytic processing of seed storage proteins in Arabidopsis embryo cells starts in the multivesicular bodies. Plant Cell **18,** 2567-2581.

Pammenter, N.W., and Berjak, P. (1999). A review of recalcitrant seed physiology in relation to desiccation-tolerance mechanisms. Seed Science Research **9,** 13-37.

Pang, P.P., Pruitt, R.E., and Meyerowitz, E.M. (1988). Molecular cloning, genomic organization, expression and evolution of 12S seed strorage protein genes of *Arabidopsis thaliana*. Plant Mol Biol **11,** 805-820.

Parcy, F., Valon, C., Kohara, A., Misera, S., and Giraudat, J. (1997). The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. Plant Cell **9,** 1265-1277.

Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M., and 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,** 1567-1582.

Patrick, J.W., and Offler, C.E. (2001). Compartmentation of transport and transfer events in developing seeds. Journal of experimental botany **52,** 551-564.

Penfield, S., Pinfield-Wells, H.M., and Graham, I.A. (2006). Storage reserve mobilisation and seedling establishment in Arabidopsis. The Arabidopsis book / American Society of Plant Biologists **4,** e0100.

Penfield, S., Rylott, E.L., Gilday, A.D., Graham, S., Larson, T.R., and Graham, I.A. (2004). Reserve mobilization in the Arabidopsis endosperm fuels hypocotyl elongation in the dark, is independent of abscisic acid, and requires PHOSPHOENOLPYRUVATE CARBOXYKINASE1. Plant Cell **16,** 2705-2718.

Peng, F.Y., and Weselake, R.J. (2011). Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis. BMC genomics **12,** 286.

Pidkowich, M.S., Nguyen, H.T., Heilmann, I., Ischebeck, T., and Shanklin, J. (2007). Modulating seed beta-ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil. Proceedings of the National Academy of Sciences of the United States of America **104,** 4742-4747.

Pignocchi, C., Minns, G.E., Nesi, N., Koumproglou, R., Kitsios, G., Benning, C., Lloyd, C.W., Doonan, J.H., and Hills, M.J. (2009). ENDOSPERM DEFECTIVE1 Is a Novel Microtubule-Associated Protein Essential for Seed Development in Arabidopsis. Plant Cell **21,** 90-105.

Pommerrenig, B., Popko, J., Heilmann, M., Schulmeister, S., Dietel, K., Schmitt, B., Stadler, R., Feussner, I., and Sauer, N. (2013). SUCROSE TRANSPORTER 5 supplies Arabidopsis embryos with biotin and affects triacylglycerol accumulation. The Plant journal : for cell and molecular biology **73,** 392-404.

Quettier, A.L., and Eastmond, P.J. (2009). Storage oil hydrolysis during early seedling growth. Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale **47,** 485-490.

Radoeva, T., and Weijers, D. (2014). A roadmap to embryo identity in plants. Trends in plant science.

Raz, V., Bergervoet, J.H., and Koornneef, M. (2001). Sequential steps for developmental arrest in Arabidopsis seeds. Development **128,** 243-252.

Reidt, W., Wohlfarth, T., Ellerstrom, M., Czihal, A., Tewes, A., Ezcurra, I., Rask, L., and Baumlein, H. (2000). Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. The Plant journal : for cell and molecular biology **21,** 401-408.

Robenek, M.J., Severs, N.J., Schlattmann, K., Plenz, G., Zimmer, K.P., Troyer, D., and Robenek, H. (2004). Lipids partition caveolin-1 from ER membranes into lipid droplets: updating the model of lipid droplet

biogenesis. FASEB journal : official publication of the Federation of American Societies for Experimental Biology **18,** 866-868.

Roscoe, T.T., Guilleminot, J., Bessoule, J.J., Berger, F., and Devic, M. (2015). Complementation of Seed Maturation Phenotypes by Ectopic Expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in Arabidopsis. Plant & cell physiology **56,** 1215-1228.

Ruuska, S.A., Schwender, J., and Ohlrogge, J.B. (2004). The capacity of green oilseeds to utilize photosynthesis to drive biosynthetic processes. Plant Physiol **136,** 2700-2709.

Ruuska, S.A., Girke, T., Benning, C., and Ohlrogge, J.B. (2002). Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell **14,** 1191-1206.

Sanchez, R., and Zhou, M.M. (2011). The PHD finger: a versatile epigenome reader. Trends in biochemical sciences **36,** 364-372.

Sanders, A., Collier, R., Trethewy, A., Gould, G., Sieker, R., and Tegeder, M. (2009). AAP1 regulates import of amino acids into developing Arabidopsis embryos. The Plant journal : for cell and molecular biology **59,** 540-552.

Santos-Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M., and Lepiniec, L. (2008). Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. The Plant journal : for cell and molecular biology **54,** 608-620.

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

Schmidt, R., Stransky, H., and Koch, W. (2007). The amino acid permease AAP8 is important for early seed development in Arabidopsis thaliana. Planta **226,** 805-813.

Schwender, J., and Ohlrogge, J.B. (2002). Probing in vivo metabolism by stable isotope labeling of storage lipids and proteins in developing Brassica napus embryos. Plant Physiol **130,** 347-361.

Schwender, J., Shachar-Hill, Y., and Ohlrogge, J.B. (2006). Mitochondrial metabolism in developing embryos of Brassica napus. J Biol Chem **281,** 34040-34047.

Schwender, J., Goffman, F., Ohlrogge, J.B., and Shachar-Hill, Y. (2004). Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. Nature **432,** 779-782.

Shen, Y., Devic, M., Lepiniec, L., and 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,** 811-820.

Shimada, T., Fuji, K., Tamura, K., Kondo, M., Nishimura, M., and Hara-Nishimura, I. (2003). Vacuolar sorting receptor for seed storage proteins in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **100,** 16095-16100.

Siloto, R.M., Findlay, K., Lopez-Villalobos, A., Yeung, E.C., Nykiforuk, C.L., and Moloney, M.M. (2006). The accumulation of oleosins determines the size of seed oilbodies in Arabidopsis. Plant Cell **18,** 1961-1974.

Sjodahl, S., Rodin, J., and Rask, L. (1991). Characterization of the 12S globulin complex of Brassica napus. Evolutionary relationship to other 11-12S storage globulins. European journal of biochemistry / FEBS **196,** 617-621.

Somerville, C., and Browse, J. (1996). Dissecting desaturation: plants prove advantageous. Trends in cell biology **6,** 148-153.

Song, C.P., Agarwal, M., Ohta, M., Guo, Y., Halfter, U., Wang, P., and Zhu, J.K. (2005). Role of an Arabidopsis AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses. Plant Cell **17,** 2384-2396.

Stadler, R., Lauterbach, C., and Sauer, N. (2005). Cell-to-cell movement of green fluorescent protein reveals post-phloem transport in the outer integument and identifies symplastic domains in Arabidopsis seeds and embryos. Plant Physiol **139,** 701-712.

Stone, S.L., Kwong, L.W., Yee, K.M., Pelletier, J., Lepiniec, L., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. Proceedings of the National Academy of Sciences of the United States of America **98,** 11806-11811.

Stone, S.L., Braybrook, S.A., Paula, S.L., Kwong, L.W., Meuser, J., Pelletier, J., Hsieh, T.F., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (2008). Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. Proceedings of the National Academy of Sciences of the United States of America **105,** 3151-3156.

Stymne, S., and Stobart, A.K. (1987). Triacylglycerol biosynthesis. (New York: Academic Press).

Sun, X., Shantharaj, D., Kang, X., and Ni, M. (2010). Transcriptional and hormonal signaling control of Arabidopsis seed development. Current opinion in plant biology **13,** 611-620.

Suzuki, M., and McCarty, D.R. (2008). Functional symmetry of the B3 network controlling seed development. Current opinion in plant biology **11,** 548-553.

Suzuki, M., Kao, C.Y., and McCarty, D.R. (1997). The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. Plant Cell **9,** 799-807.

Suzuki, M., Wang, H.H., and McCarty, D.R. (2007). Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. Plant Physiol **143,** 902-911.

Swaminathan, K., Peterson, K., and Jack, T. (2008). The plant B3 superfamily. Trends in plant science **13,** 647-655.

Tauchi-Sato, K., Ozeki, S., Houjou, T., Taguchi, R., and Fujimoto, T. (2002). The surface of lipid droplets is a phospholipid monolayer with a unique Fatty Acid composition. The Journal of biological chemistry **277,** 44507-44512.

Tegeder, M. (2012). Transporters for amino acids in plant cells: some functions and many unknowns. Current opinion in plant biology **15,** 315-321.

Tegeder, M. (2014). Transporters involved in source to sink partitioning of amino acids and ureides: opportunities for crop improvement. Journal of experimental botany **65,** 1865-1878.

Thelen, J.J., Miernyk, J.A., and Randall, D.D. (2000). Pyruvate dehydrogenase kinase from Arabidopsis thaliana: a protein histidine kinase that phosphorylates serine residues. The Biochemical journal **349,** 195-201.

Theodoulou, F.L., and Eastmond, P.J. (2012). Seed storage oil catabolism: a story of give and take. Current opinion in plant biology **15,** 322-328.

Tiedemann, J., Rutten, T., Monke, G., Vorwieger, A., Rolletschek, H., Meissner, D., Milkowski, C., Petereck, S., Mock, H.P., Zank, T., and Baumlein, H. (2008). Dissection of a complex seed phenotype: novel insights of FUSCA3 regulated developmental processes. Developmental biology **317,** 1-12.

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

Tsukagoshi, H., Morikami, A., and 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,** 2543-2547.

Tsukagoshi, H., Saijo, T., Shibata, D., Morikami, A., and 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 Physiol **138,** 675-685.

Tzen, J.T., and Huang, A.H. (1992). Surface structure and properties of plant seed oil bodies. The Journal of cell biology **117,** 327-335.

Uhrig, R.G., O'Leary, B., Spang, H.E., MacDonald, J.A., She, Y.M., and Plaxton, W.C. (2008). Coimmunopurification of phosphorylated bacterial- and plant-type phosphoenolpyruvate carboxylases with the plastidial pyruvate dehydrogenase complex from developing castor oil seeds. Plant Physiol **146,** 1346-1357.

van der Klei, H., Van Damme, J., Casteels, P., and Krebbers, E. (1993). A fifth 2S albumin isoform is present in Arabidopsis thaliana. Plant Physiol **101,** 1415-1416.

van Zanten, M., Koini, M.A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., Koornneef, M., Fransz, P., and Soppe, W.J. (2011). Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. Proceedings of the National Academy of Sciences of the United States of America **108,** 20219-20224.

Veerappan, V., Chen, N., Reichert, A.I., and Allen, R.D. (2014). HSI2/VAL1 PHD-like domain promotes H3K27 trimethylation to repress the expression of seed maturation genes and complex transgenes in Arabidopsis seedlings. BMC plant biology **14,** 293.

Veerappan, V., Wang, J., Kang, M., Lee, J., Tang, Y., Jha, A.K., Shi, H., Palanivelu, R., and 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,** 1-17.

Verdier, J., and Thompson, R.D. (2008). Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. Plant & cell physiology **49,** 1263-1271.

Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I., Righetti, K., Bandyopadhyay, K., Leprince, O., Chatelain, E., Vu, B.L., Gouzy, J., Gamas, P., Udvardi, M.K., and Buitink, J. (2013). A regulatory

network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of Medicago truncatula seeds. Plant physiology **163,** 757-774.

Vicente-Carbajosa, J., and Carbonero, P. (2005). Seed maturation: developing an intrusive phase to accomplish a quiescent state. Int J Dev Biol **49,** 645-651.

Vicient, C.M., Bies-Etheve, N., and Delseny, M. (2000). Changes in gene expression in the leafy cotyledon1 (lec1) and fusca3 (fus3) mutants of Arabidopsis thaliana L. Journal of experimental botany **51,** 995-1003.

Von Wettsetein-Knowles, P.M. (1982). Elongase and epicuticular wax biosynthesis. Physiol. Vég **20,** 797-809.

Wang, F., and Perry, S.E. (2013). Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. Plant Physiol **161,** 1251-1264.

Wang, Z., Cao, H., Sun, Y., Li, X., Chen, F., Carles, A., Li, Y., Ding, M., Zhang, C., Deng, X., Soppe, W.J., and Liu, Y.X. (2013). Arabidopsis paired amphipathic helix proteins SNL1 and SNL2 redundantly regulate primary seed dormancy via abscisic acid-ethylene antagonism mediated by histone deacetylation. The Plant cell **25,** 149-166.

Waters, A., Creff, A., Goodrich, J., and Ingram, G. (2013). "What we've got here is failure to communicate": zou mutants and endosperm cell death in seed development. Plant signaling & behavior **8,** e24368.

Weitbrecht, K., Muller, K., and Leubner-Metzger, G. (2011). First off the mark: early seed germination. Journal of experimental botany **62,** 3289-3309.

Weselake, R.J., Taylor, D.C., Rahman, M.H., Shah, S., Laroche, A., McVetty, P.B., and Harwood, J.L. (2009). Increasing the flow of carbon into seed oil. Biotechnol Adv **27,** 866-878.

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

Wojcikowska, B., Jaskola, K., Gasiorek, P., Meus, M., Nowak, K., and Gaj, M.D. (2013). LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of Arabidopsis, via YUCCA-mediated auxin biosynthesis. Planta **238,** 425-440.

Xi, D., and Zheng, C. (2011). Transcriptional regulation of seed storage protein genes in Arabidopsis and cereals. Seed Science Research **21,** 247-254.

Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S., and Hattori, T. (2009). Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. The Plant journal : for cell and molecular biology **58,** 843-856.

Yamamoto, A., Kagaya, Y., Usui, H., Hobo, T., Takeda, S., and Hattori, T. (2010). Diverse roles and mechanisms of gene regulation by the Arabidopsis seed maturation master regulator FUS3 revealed by microarray analysis. Plant & cell physiology **51,** 2031-2046.

Yamauchi, Y., Ogawa, M., Kuwahara, A., Hanada, A., Kamiya, Y., and Yamaguchi, S. (2004). Activation of gibberellin biosynthesis and response pathways by low temperature during imbibition of Arabidopsis thaliana seeds. Plant Cell **16,** 367-378.

Yan, D., Duermeyer, L., Leoveanu, C., and Nambara, E. (2014). The functions of the endosperm during seed germination. Plant & cell physiology **55,** 1521-1533.

Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M. (2013). VAL- and AtBMI1-mediated H2Aub initiate the switch from embryonic to postgerminative growth in Arabidopsis. Current biology : CB **23,** 1324-1329.

Yatsu, L.Y., and Jacks, T.J. (1972). Spherosome membranes: half unit-membranes. Plant Physiol **49,** 937-943.

Zhang, W.-H., Zhou, Y., Dibley, K.E., Tyerman, S.D., Furbank, R.T., and Patrick, J.W. (2007). Nutrient loading of developing seeds. Functional Plant Biology **34,** 314-331.

Zhou, Y., Tan, B., Luo, M., Li, Y., Liu, C., Chen, C., Yu, C.W., Yang, S., Dong, S., Ruan, J., Yuan, L., Zhang, Z., Zhao, L., Li, C., Chen, H., Cui, Y., Wu, K., and Huang, S. (2013). HISTONE DEACETYLASE19 interacts with HSL1 and participates in the repression of seed maturation genes in Arabidopsis seedlings. The Plant cell **25,** 134-148.

Chapter 2

Potential targets of VAL1 repression in developing *Arabidopsis thaliana* embryos

A. Schneider[1]
D. Aghamirzaie[2]
H. Elmarakeby[3]
A. Nepal[4]
A. Koo[4]
L. Heath[3]
R. Grene[1]
E. Collakova[1]

[1]Department of Plant Pathology, Physiology and Weed Sciences, College of Agriculture and Life Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA;
[2]Genetics, Bioinformatics and Computational Biology Program, Virginia Tech, Blacksburg, VA, USA;
[3]Department of Computer Science, Virginia Tech, Blacksburg, VA, USA;
[4]Division of Biochemistry, Interdisciplinary Plant Group, University of Missouri, Columbia, MO, USA

Manuscript in preparation

2.1 Abstract

Developing oilseeds accumulate oils and seed storage proteins synthesized by the pathways of primary metabolism. Seed development and metabolism are positively regulated through transcription factors belonging to the LEC1, ABI3, FUS3, LEC2 (LAFL) regulatory network. The *VP1/ABI3-Like* (*VAL*) gene family encodes repressors of the seed maturation program in germinating seeds. They are also expressed during seed maturation, where their role is not yet understood. VAL1 was identified through a reverse genetics approach as a regulator of seed metabolism, since *val1* mutant seeds accumulated elevated levels of storage proteins compared to the wild type. Two *VAL1* splice variants were identified through RNA-Seq analysis: a canonical form and a truncated form lacking the plant-homeodomain-like domain associated with epigenetic repression. Transcriptomics analysis revealed evidence that VAL1 is a global epigenetic and transcriptional repressor in developing embryos. None of the transcripts encoding the core LAFL network transcription factors, including FUSCA3, were affected in *val1* embryos. However, 38% of transcripts belonging to the FUSCA3 regulon were de-repressed in the absence of VAL1. The LEC1 and LEC2 regulons responded to a lesser extent. Activation of VAL1 by FUSCA3 appears to result in repression of a subset of seed maturation genes downstream of core LAFL regulators.

2.2 Introduction

The evolution of seeds allowed plants to conquer land by developing a dormant state that facilitates survival in seasonally harsh environments. Seed storage compounds, primarily consisting of oils, proteins, and carbohydrates, are synthesized in metabolically active developing seeds to provide energy and structural components for germinating seedlings (Penfield et al., 2006; Baud et al., 2008). Seed development can be divided into three major phases (Meinke, 1995; Baud et al., 2008). First, in embryogenesis, cells divide and differentiate to form shoot and root meristems and the embryo axis and cotyledons. Second, the maturation phase is characterized primarily by biosynthesis and accumulation of seed storage compounds. Third, seed development ends with the acquisition of desiccation tolerance and dormancy.

Seed storage compounds are synthesized through well-understood metabolic pathways, while the regulation of these pathways is less well understood (Baud et al., 2008; Jia et al., 2014). In the model oilseed species *Arabidopsis thaliana*, regulation has been primarily studied at the level of transcription. Several transcription factors (TFs) are known to positively and globally regulate distinct aspects of seed development (Jia et al., 2014). These global regulators include the CAAT-box family protein LEAFY COTYLEDON 1 (LEC1) and the B3-family proteins ABCISIC ACID INSENSTIVE 3 (ABI3), FUSCA 3 (FUS3), and LEC2, together making up part of the core LEC1, ABI3, FUS3, LEC2 (LAFL) regulatory network (Jia et al., 2013). LEC1 and 2 function primarily during embryogenesis, while FUS3 and ABI3 activate maturation-specific processes. Together, the core LAFL TFs form a complex transcriptional network regulating seed development, with many overlapping targets, including mutual and auto regulation, and the participation of other TFs, such as the LEC1 homolog LEC1-LIKE and the basic leucine zipper TF bZIP67 (Mendes et al., 2013).

Most of, but not all, the known seed development regulators function primarily as activators of gene expression, although several repressors have also been identified. Recently, TRANSPARENT TESTA GLABRA 1 (TTG1) was identified as a repressor of ABI3, LEC2 and several genes involved in seed storage compound synthesis in developing seeds (Chen et al., 2015). The function of the other known repressors has primarily been defined as the repression of the seed maturation program in germinating seeds and seedlings (Jia et al., 2014). The VIVAPOROUS1/ABI3-LIKE (VAL) TFs are repressors of the core LAFL network in developing seedlings, and are required for the transition from embryonic to the vegetative state. Single knockouts of *VAL1* or *VAL2* do not have major effects on seedling and vegetative development, but double knockouts produce seedlings that abort after 7 – 10 days following germination (Suzuki et al., 2007; Tsukagoshi et al., 2007). Seed storage proteins are substantially elevated in *val1*/*val2* seedlings, a phenotype related to the retention of the embryonic state. Microarray analysis showed that core LAFL network genes were massively upregulated in these seedlings (Suzuki et al., 2007). To date, the influence of VAL TFs on seed development itself has not been elucidated, despite its known expression in maturing seeds (Schmid et al., 2005).

VAL1 (At2g30470) contains several domains that are implicated in chromatin/DNA binding. Two of these domains, a domain very similar to the canonical plant homeodomain (PHD) zinc finger domain and a conserved zinc finger Cys- and Trp-containing (CW) domain, bind to histone proteins trimethylated at the fourth Lys residue of histone protein 3 (H3K4me3), which is indicative of active transcription (Hoppmann et al., 2011; Sanchez and Zhou, 2011). This PHD-like (PHD-L) domain contains three potential zinc fingers as opposed to two that are present in the canonical PHD domains (Suzuki et al., 2007; Sanchez and Zhou, 2011; Mouriz et al., 2015). While the VAL1 PHD-L domain has not been demonstrated to bind to chromatin, it is

involved in repressing several seed maturation genes in seedlings through the promotion of trimethylation of Lys residue 27 on histone protein 3 (H3K27me3) (Veerappan et al., 2012; Veerappan et al., 2014). Trimethylation of this residue is a repressive epigenetic mark catalyzed by POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) methyltransferases (Chanvivattana et al., 2004; Schubert et al., 2006). In contrast, the VAL1 CW domain has been demonstrated to bind directly to H3K4me3 (Hoppmann et al., 2011). VAL1 contains the well-known B3 DNA-binding domain, utilized by many plant-specific transcriptional activators, such as ABI3, to bind directly to the RY/Sph motif (CATGCA) in promoter regions of target genes (Monke et al., 2004), although direct DNA binding by VAL1 remains to be demonstrated (Tsukagoshi et al., 2005; Suzuki et al., 2007; Swaminathan et al., 2008; Guerriero et al., 2009). VAL1 also contains an ethylene-responsive element binding factor-associated amphiphilic repression (EAR) motif, which is necessary for repression of a reporter construct driven by a sugar responsive promoter (Tsukagoshi et al., 2005). EAR motifs are ubiquitous among repressors and likely function through recruitment of histone modifiers to genes for chromatin modifications (Kagale and Rozwadowski, 2011).

In addition to expression in seedlings, *VAL1* is expressed in the seed, with a peak during the transition from early to middle maturation (Schmid et al., 2005); however, its role has not been elucidated in developing seeds. In the work reported here, VAL1 was identified in a mutant screen aimed at finding TFs involved in the regulation of seed development and metabolism. The *val1* mutant accumulated elevated protein levels in dry seeds compared to the wild type. Potential functions of *VAL1* in developing embryos were further investigated in transcriptomics and metabolomics time-course experiments covering seed maturation and early desiccation phases of embryo development. Studies of Arabidopsis seed development have largely been

limited to whole seeds or siliques due to the small seed size (Le et al., 2010; Jia et al., 2013; Chen et al., 2015). To our knowledge, this study represents the first detailed investigation of the embryo transcriptome and metabolome in Arabidopsis and the first evidence for action of a global transcriptional repressor that functions during embryo development.

2.3 Results

2.3.1 VAL1 was identified in a high-throughput screen of TF mutants for alterations in seed storage compound levels and/or composition

A multi-step screening method was developed to identify TFs whose action influences the accumulation of seed storage compounds in Arabidopsis. First, a list of over 2,338 non-redundant entries for TFs was obtained by compiling TFs from The Database of Arabidopsis Transcription Factors, the University of Potsdam Plant Transcription Factor Database, and the RIKEN Arabidopsis Transcription Factor Database (Guo et al., 2005; Iida et al., 2005; Perez-Rodriguez et al., 2010). This list contained 1,136 TFs that were expressed in developing siliques and/or seeds based on the microarray data available at the time (Ruuska et al., 2002; Swarbreck et al., 2008). The TFs were prioritized such that a total of 251 TFs that showed the largest gene expression changes during seed development were selected. At least one corresponding Salk T-DNA insertion mutant line was identified for 184 of the 251 TFs and 156 homozygous mutants were obtained. These TF mutants were subsequently screened for alterations in seed storage compound accumulation. Several known TF mutants with associated alterations in seed storage compound levels were identified in a blind-design manner as a result of this screen (e.g., ABI3 and FUS3), indicating that the selection method could identify seed maturation regulators. Collectively, we identified 43 mutants that are potentially affected in seed storage compound levels.

Disruption of the *VAL1* gene in the SALK_088606C mutant was shown to affect seed metabolism through this high-throughput screen. VAL1 has a known role in seedling establishment (Suzuki et al., 2007; Tsukagoshi et al., 2007), and based on the results from the screen, it is also involved in regulating seed metabolism in Arabidopsis. Estimation of the dry

seed protein content through a fluorescent-based hydrophobic protein assay revealed that wild type and *val1* mutant seeds accumulated 80.6 ± 8.6 and 93.3 ± 11.28 µg mg^{-1}, respectively, of hydrophobic proteins per dry weight (DW) of seed, representing an approximate 12% increase in protein content in *val1* seeds (22 replicates for wild type, 16 for *val1*, *p*-value = 0.00083). In contrast, the amount and composition of seed storage proteins did not appear to be affected in the *val1* mutant when analyzed through SDS-PAGE (Figure A.1), but a marginal increase in protein, such as 12%, is difficult to capture by protein gels. The oil levels and fatty acid composition of *val1* seeds remained unchanged (Figure A.2).

2.3.2 The SALK_088606C mutant contains one T-DNA insertion

The SALK_088606C mutant contains an insertion in the seventh exon of the *VAL1* gene (Figure 1A), which encodes part of the B3 domain. This insertion mutant has been characterized previously in studies using seedlings and mature vegetative plants (Tsukagoshi et al., 2005; Suzuki et al., 2007; Tsukagoshi et al., 2007; Veerappan et al., 2012; Jia et al., 2013; Sharma et al., 2013; Veerappan et al., 2014), but not in seeds. In addition, the results of the T-DNA SEQ project (http://www.ncbi.nlm.nih.gov/nucgss/KO428765) indicated that the parental line SALK_088606 contains a second T-DNA insertion in the 5'-untranslated region of an unrelated gene (At5g63350). PCR-based genotyping on plants grown from seed directly obtained from the Arabidopsis Biological Resource Center (ABRC, The Ohio State University) confirmed that this insertion is not present in the SALK_088606C mutant, and the expression pattern of At5g63350 is unaltered in *val1* mutant embryos, suggesting that this mutant line is free of any additional insertions other than that present in the *VAL1* exon (Figure A.3).

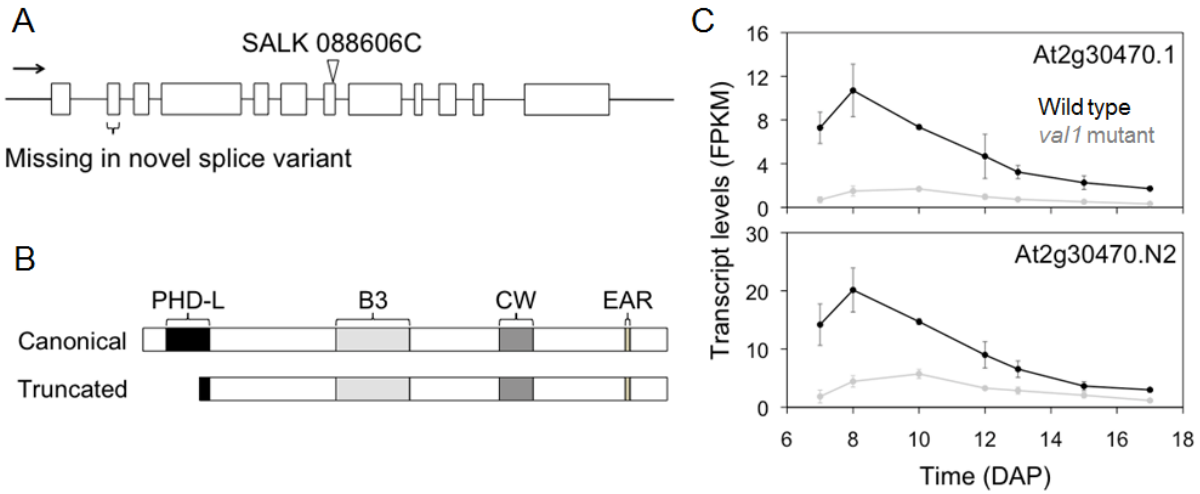

Figure 1. The *VAL1* gene and corresponding splice variants. A) The *VAL1* gene. The SALK_088606C mutant carries a T-DNA insertion in the seventh exon, which encodes part of the B3 domain. The novel splice variant is missing the entire second exon. qPCR confirmation of the truncated splice variant is found in Figure A.4D. B) Splicing of the *VAL1* premature RNA yields two splice variants encoding two different protein variants. The full-length protein contains a PHD-L domain, B3 DNA-binding domain, CW domain, and an EAR motif. The truncated splice variant is missing 77% of the PHD-L domain. C) Expression of both *VAL1* splice variants in developing Arabidopsis embryos. The full-length transcript is expressed nearly two fold less at each time point in the wild type than the truncated variant. Based on RNAseq data, the expression of both splice variants is reduced in the *val1* mutant.

2.3.3 Temporal aspects of Arabidopsis embryo development in wild type and *val1* mutant

The majority of seed biomass consists of storage compounds that are predominantly found in the embryo of dry seeds (Baud et al., 2002; Higashi et al., 2006; Baud et al., 2008). Increased protein levels in *val1* dry seeds suggested that VAL1 might have a role in regulating embryo development and metabolism. In order to elucidate the specific role(s) of VAL1 in this context, wild type and mutant embryos at the maturation and desiccation phases were investigated. This comparative time course (Figure 2) covered the early maturation (7 and 8 days after pollination (DAP)), middle maturation (10, 12 and 13 DAP), and late maturation/early desiccation (15 and 17 DAP) phases. The transcriptomes, metabolomes, and levels of selected

phytohormones were analyzed in the wild type and *val1* mutant embryos, free from contaminating seed coats and endosperm.

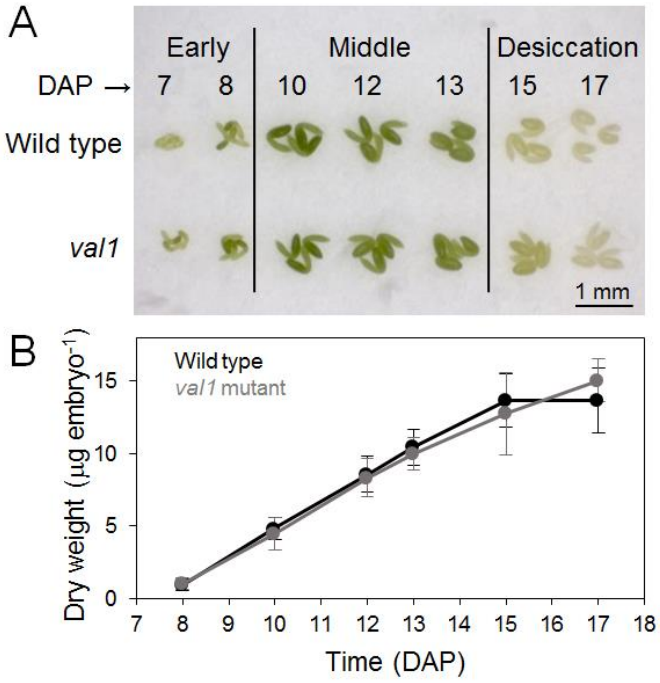

Figure 2. Developing wild type and *val1* mutant embryos. A. Embryos dissected from seeds of specified age from a representative wild type and *val1* mutant plant, respectively, are shown. Separation of embryo development into three major phases of seed maturation (early seed filling, middle seed filling, and desiccation) is shown by vertical lines. B. Biomass accumulation in developing wild type and *val1* mutant embryos. Averages ± SD of nine biological replicates for each line are shown. No differences in the appearance or DW of wild type and mutant embryos of comparable ages were observed.

2.3.4 The VAL1 gene encodes a novel splice variant missing the PHD-L domain

We performed high-throughput RNA sequencing (RNAseq) on total RNA extracted from the developing wild type and *val1* embryos. Analysis of the RNAseq data showed that a full-length transcript and also a novel splice variant missing the second exon of *VAL1* are co-expressed throughout the time course in both wild type and mutant embryos (Figure 1C). The novel splice variant produces a protein with an 85-amino acid truncation from the amino terminus, removing the majority (77%) of the PHD-L domain (Figure 1B). The novel splice variant is also more abundant (nearly two fold at each time point) than the full-length transcript

in wild type embryos. From the RNAseq data, it appears that low levels of these two different *VAL1* transcripts are still detected in the *val1* mutant throughout the time course, which can be explained by the way read mapping was performed as part of the RNAseq analysis. Nearly wild type levels of reads (Figure A.4A) were mapped to the 5' end of the VAL1 transcript in *val1*, as confirmed by qPCR (Figure A.4B), while very few reads were mapped downstream of the T-DNA insertion site. Determining expression levels of transcripts is based on averaging the mapped reads, which is consistent with the overall low levels of the two splice variants in *val1*. The reads that mapped to the downstream region could be derived from the full-length transcript containing the T-DNA insertion. Additionally, as the insertion site is very close to an intron-exon border, the insertion could be spliced out. However, when the region flanking the T-DNA insertion site was assayed through RT-PCR, no amplification was detected in the *val1* mutant, while the band of correct size was detected in the wild type (Figure A.4C), suggesting that the T-DNA insertion was retained. Despite the levels and proper splicing of the 5' end, the VAL1 transcript from *val1* embryos is probably rendered non-functional and would produce a severely truncated protein.

Table 1: Results from RNAseq and classification of transcripts. The total number of genes and transcripts are indicated, as are Cuffcompare classifications defined in Trapnell et al., (2010). Transcript classification: "=" are known transcripts, "j" are novel splice variants, "o" have partial overlap with known reference transcripts, "s" are antisense intron transcripts, "u" are intergenic transcripts, and "x" are antisense exon transcripts. *Includes chloroplast and mitochondrial genes.

	Genes*	**Transcripts**	**=**	**j**	**o**	**x**	**u**	**s**
Total detected	33,544	53,988	41,934	10,947	457	350	289	11
Upregulated in *val1*	4,030	5,320	4,283	903	49	44	40	1
Downregulated in *val1*	2,667	3,625	2,828	667	56	41	31	2

2.3.5 VAL1 regulates embryo development through FUS3 without a direct effect on the expression of core LAFL genes

The RNAseq data (summarized in Table 1) show that 5,320 transcripts (including all detectable splice variants encoded by a given gene) are upregulated during embryo maturation in *val1* embryos, in at least one time point, whereas 3,625 transcripts are downregulated. Of the upregulated transcripts, 261 (encoded by 204 genes) are upregulated (FDR-adjusted *p*-value < 0.05) at every time point in *val1* embryos. VAL1 can repress gene expression by promoting PRC2-catalyzed H3K27me3 repression in seedlings, among other possible mechanisms (Veerappan et al., 2014). To determine whether any of the 204 upregulated genes could be repressed through this type of chromatin modification, we compared our results with those genes carrying the H3K27me3 mark (Zhang et al., 2007). Fifty-five genes were identified (27%) that carry this epigenetic modification mark. In addition, as VAL1 contains a B3 domain, we searched for the presence of the RY/Sph motif in promoter regions (1000 base pairs upstream of the start codon) of the 204 upregulated genes using the MEME suite (Bailey et al., 2009) We found that 100 genes (49%) contain at least one RY/Sph motif in their promoters. Of the 55 genes upregulated at every time point, 37 (67.3%) contain at least one RY/Sph motif (Table 2). This group of genes represents candidate direct targets of VAL1 repression. Many of these genes encode regulatory proteins, including the auxin signaling TF AUXIN RESISTANT 3 (AXR3, Figure 3), the putative repressor PICKLE-RELATED 2 (PKR2, Figure 3) and the embryogenesis regulating TF WUSCHEL-RELATED HOMEOBOX 2 (WOX2, Figure 3) (Lie et al., 2012).

VAL1 regulates expression of the core LAFL genes in seedlings in conjunction with VAL2 (Suzuki et al., 2007). In contrast to germinating seedlings, expression of these genes is not affected by the *val1* mutation in developing embryos (Figure A.5). The LAFL TFs are part of a

larger transcriptional regulatory network and each TF has its own regulon with both unique and overlapping targets (Jia et al., 2014). VAL1 could regulate the LAFL network activity through repression of members of these regulons. Therefore, we compared our results with the regulons of LEC1 (Junker et al., 2012), LEC2 (Braybrook et al., 2006), ABI3 (Monke et al., 2012), and FUS3 (Wang and Perry, 2013). The LEC1 regulon overlapped with 104 transcripts (79 genes, 2% of the 5,320 transcripts) upregulated in at least one time point in *val1* embryos, 84 of these were unique to LEC1. Similarly, 295 transcripts (207 genes, 4% of the 5,320 transcripts) were also present in the LEC2 regulon, 240 were unique to LEC2 and seven were upregulated at all time points. Only 35 transcripts (37 genes, <1%) of the ABI3 regulon were upregulated in *val1* embryos at any time point, eight were unique to ABI3.

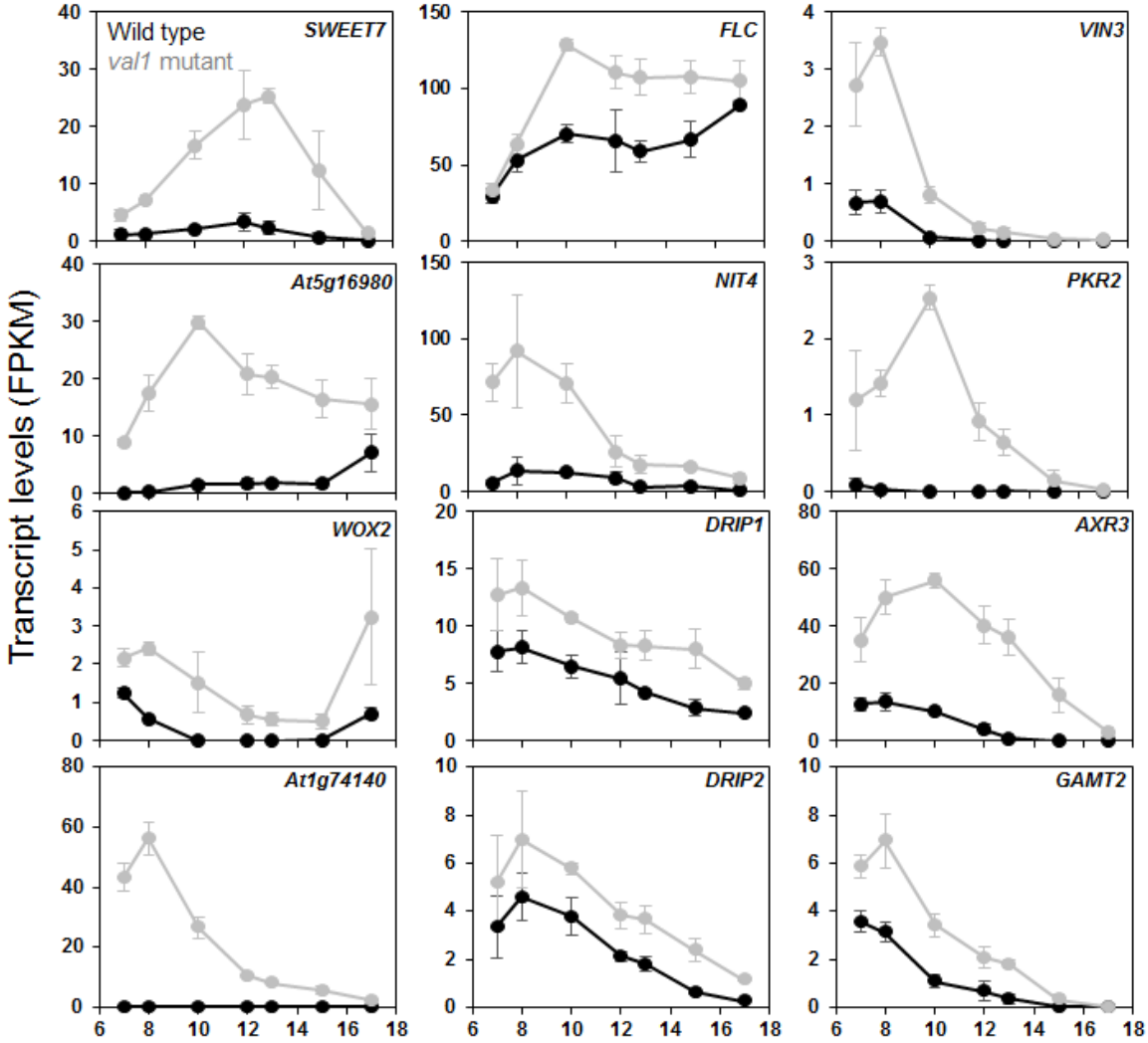

Figure 3. Selected transcripts upregulated in *val1*. All transcripts shown are known transcripts, not novel, and with the exception of *FLC*, are upregulated (FDR-adjusted *p* value < 0.05) at all time points. Validation by qPCR of selected transcripts is shown in Figure A.6.

Table 2. Selected genes upregulated at each time point in *val1* embryos. The PRC2 H3K27me3 mark and the RY motif were present in the upstream region of each gene.

General category	AGI	Name	Function	Transcript count	LAFL regulator
Auxin response	At4g32810	CCD8	Carotenoid cleavage dioxygenase 8	1	LEC2
	At1g66360	At1g66360	Calcium-dependent lipid-binding family	1	
Cell wall modification	At1g12070	At1g12070	Immunoglobulin E-set superfamily	1	
	At1g21310	RSH	Extensin 3	1	FUS3
Chlorophyll degradation	At1g19670	AtCLH1	Chlorophyllase 1	1	
Chromosome segregation	At3g58210	At3g58210	TRAF-like family protein	1	
Defense	At1g56520	At1g56520	Disease resistance protein (TIR-NBS-LRR class) family	2	
GA catabolism	At5g56300	GAMT2	GA methyltransferase 2	1	FUS3, LEC2
Glycolipid transporter	At1g21360	GLTP2	Glycolipid transfer protein 2	1	FUS3
Iron homeostasis	At1g01580	FRD1	Ferric reduction oxidase 2	3	FUS3
	At3g08040	AtFRD3	Ferric Reductase Defective 3	3	
Nitrogen detoxification	At5g22300	NIT4	Nitrilase 4	1	
Phospholipase	At2g26870	NPC2	Non-specific phospholipase C2	1	FUS3
Protein kinase	At1g23700	At1g23700	Protein kinase superfamily protein	1	
Putative F box protein	At5g44220	At5g44220	F-box family protein	1	
Steroid metabolism	At4g25850	ORP4B	OSBP (oxysterol binding protein)-related protein 4B	1	
Stress response	At4g33420	At4g33420	Peroxidase superfamily protein	1	FUS3
Terpene synthesis	At4g15370	BARS1	Baruol synthase 1	1	
Transcription factor	At3g12730	At3g12730	Homeodomain-like superfamily	2	FUS3
	At4g31900	PKR2	Pickle Related 2	1	
	At5g59340	WOX2	WUSCHEL related homeobox 2	1	FUS3
	At1g04250	AXR3	Auxin Resistant 3	1	FUS3
Transferase	At3g12470	At3g12470	Polynucleotidyl transferase, ribonuclease H-like superfamily	1	
Transporter	At1g10010	AAP8	Amino acid permease 8	1	
	At1g79520	At1g79520	Cation efflux family protein	1	
	At3g43270	At3g43270	Plant invertase/pectin methylesterase inhibitor superfamily	1	
	At4g10850	SWEET7	Nodulin MtN3 family protein	1	LEC2
Unknown	At1g11740	At1g11740	Ankyrin repeat family protein	1	FUS3
	At1g54530	At1g54530	Calcium-binding EF hand family	1	
	At1g77655	At1g77655	Unknown protein	1	FUS3
	At2g18460	LCV3	Like COV 3	1	
	At2g41360	At2g41360	Galactose oxidase/kelch repeat superfamily protein	1	
	At3g21030	At3g21030	Copia-like retrotransposon family	2	
	At3g57620	At3g57620	Glyoxal oxidase-related protein	1	
	At4g30460	At4g30460	Glycine-rich protein	1	
	At5g09210	At5g09210	GC-rich sequence DNA-binding factor-like protein	1	
	At5g57785	At5g57785	Unknown protein	1	FUS3

In contrast, 701 transcripts (467 genes, 13% of the 5320 transcripts) that were upregulated in *val1* mutant embryos in at least one time point are encoded by genes that belong to the FUS3 regulon (623 were unique). Forty-eight of the unique transcripts were upregulated and differentially expressed at all time points. Several of the transcripts upregulated in *val1* embryos that are part of the FUS3 regulon are also members of the LAFL network (Table 3), including the TFs FLOWERING LOCUS C (FLC, Figure 3) and CUP-SHAPED COTYLEDONS 1 (CUC1, Figure A.5), both targets of FUS3, and also genes involved in phytohormone synthesis and signaling. For example, GIBBERLLIC ACID METHYLTRANSFERASE 2 (GAMT2), an enzyme involved in GA catabolism, was upregulated across the entire time course (Figure 3), and AXR3 (a target of both FUS3 and LEC2) was profoundly upregulated in *val1* embryos. We also compared our list of genes that were upregulated in *val1* embryos at all time points with *val1* seedling microarray data (Suzuki et al., 2007) and found eight genes. PKR2, NITRILASE 4 (Figure 3), the putative sucrose transporter SWEET7 (Figure 3), At4g35650, At1g16705, At1g12070, and At1g74140 (Figure 3) are upregulated in both *val1* embryos and seedlings.

Table 3: Genes in the LAFL network involved in transcription and hormone responses upregulated in *val1* embryos. H3K27me3 indicates whether or not a given gene is known to carry H3K27me3 (Zhang et al., 2007). RY motif indicates if the canonical RY/Sph motif (CATGCA) was found in the promoter region 1000 bp upstream of the start codon. Y stands for yes, while N for no. Abbreviations: ABA – abscisic acid; GA – gibberellic acid; BR – brassinosteroid; CK – cytokinin.

General category	AGI	Gene name	H3K27me3	RY Motif	Function (TAIR)	Upregulated in *val1* embryos (DAP)	LAFL Regulator
Transcription regulation	At3g15170	CUC1	N	N	Apical meristem formation	7,10,12,13,15, 17	FUS3
	At5g10140	FLC	Y	N	Repressor of floral transition	10,12,13,15,1 7	FUS3
GA catabolism	At5g56300	GAMT2	Y	Y	Gibberellic acid methyltransferase	7,8,10,12,13, 15,17	FUS3 LEC2
	At1g47990	GA2OX4	Y	Y	Gibberellic acid oxidase	7,8,10	FUS3
GA synthesis	At4g21690	GA3OX3	Y	N	Gibberellic acid oxidase	7,8,10,12,13	FUS3 LEC2
	At4g02780	GA1	Y	Y	GA synthesis	8,10,12,13,15, 17	-
ABA signaling	At3g58120	bZIP61	N	Y	Basic leucine zipper TF	8,10,12,13,15	FUS3
Auxin catabolism	At5g55250	IAMT1	Y	Y	IAA carboxylmethyltransfe-rase	8,10,12,13,15, 17	FUS3
	At1g44350	ILL6	Y	N	IAA amino acid hydrolase	7,12,13,15	FUS3
Auxin signaling	At1g04250	AXR3	Y	Y	Auxin responsive TF	7,8,10,12,13, 15,17	FUS3 LEC2
	At3g62100	IAA30	Y	Y	Auxin responsive TF	10,12,13,15, 17	LEC2
BR synthesis	At4g36380	ROT3	N	Y	3-Epi-6-deoxocathasterone 23-monooxygenase	8,10,13,15,17	FUS3
CK synthesis	At1g25410	IPT6	Y	N	Adenylate isopentenyltransferase	10,12,13,15	FUS3
	At1g75450	CKX6	Y	N	Cytokinin dehydrogenase	13,15,17	FUS3

2.3.6 Stress responsive genes and transcriptional regulators are among the genes upregulated in developing *val1* mutant embryos

Some of the genes upregulated at all time points are involved in stress responses. Gene Ontology (GO) enriched terms for biological process among the upregulated genes included 37 genes (17%, *p*-value = 0.00925) are involved in stress response, 33 (16%, *p*-value = 0.026) in

response to abiotic or biotic stimulus, and 6 (3%, *p*-value = 0.00385) in transcription. The stress categories contained genes that respond to several abiotic stresses, including cold (e.g., COLD REGULATED 314 (At1g29395)), water-deprivation (e.g., At1g36060, a member of the DEHYDRATION-RESPONSIVE ELEMENT BINDING (DREB) protein A-6 family related to a known redox sensor (Shaikhali et al., 2008)), and salt (e.g., EARLY RESPONSE TO DEHYDRATION 3, At4g19120) stresses. The TF WOX2 was also present in this category. Genes that are responsive to oxidative stress (e.g., At5g16980, a Zn-dehydrogenase (Babiychuk et al., 1995)) were highly upregulated in the *val1* mutant throughout the time course (Figure 3). Several of the genes in this GO-enriched category are known to respond to multiple abiotic stresses (e.g., At4g20260, a plasma membrane associated cation binding protein 1). It is noteworthy that the patterns of expression of many of these stress-responsive transcripts were similar in wild type and *val1* embryos, with the difference between the two residing only in the higher extent of expression in the mutant than in the wild type throughout the time course (an example of this observation is DRIP 1 in Figure 3). Of the stress-responsive genes, eight carry the H3K27me3 mark, including three biotic stress response genes (At1g12200, At1g01580, and At1g19670) and the regulatory genes WOX2 and VERNALIZATION INSENSITIVE 3 (VIN3, Figure 3). In addition, the POLYCOMB REPRESSIVE COMPLEX 1 (PRC1) proteins DREB 2A INTERACTING PROTEIN 1 and 2 (DRIP1 and 2), which negatively regulate drought responses (Qin et al., 2008) and interact with VAL1 (Yang et al., 2013), are upregulated at each time point in *val1* embryos compared to the wild type (Figure 3).

WOX2 and VIN3 also appeared in the TF category. VIN3 is associated with PRC2-mediated H3K27me3 repression (Bond et al., 2009) and is required to repress FLOWERING LOCUS C (FLC, which is substantially elevated starting at 10 DAP) as part of the vernalization

response (Lee et al., 2015). At2g42720, which encodes a leucine rich repeat protein involved in chromatin silencing by small RNAs, was also present in the TF category. It is noteworthy that At1g16705, which encodes an acetyltransferase, is highly upregulated throughout the majority of the time course (Figure A.5) with the exception of the last time point at 17 DAP. In this case, there was little expression of this gene in the wild type embryos, as opposed to the behavior of the stress-associated genes discussed above, in which expression followed the same pattern, but was elevated in *val1*. Similarly, the PICKLE homolog PKR2, a putative chromatin remodeling repressor (Aichinger et al., 2009), was barely detectable in wild type embryos and was substantially upregulated in *val1* embryos (Figure 3).

2.3.7 Accumulation of raffinose family oligosaccharides is delayed in *val1* embryos

To investigate the effect of the *val1* mutation on the metabolome of the developing embryos, three independent metabonomics studies were conducted. A total of 51 compounds were detected, including sugars, sugar acids, sugar alcohols, sugar phosphates, phenolics, organic acids, organic amines, and amino acids, in addition to lipid-derived fatty acids, and hydrophobic proteins. A combination of gas chromatography coupled to mass spectrometry (GC-MS), GC coupled to flame ionization detection (GC-FID), ultra-performance liquid chromatography (UPLC), and the hydrophobic protein assay were employed. The phytohormones abscisic acid (ABA), auxin, 12-oxo-phytodienoic acid (OPDA), jasmonic acid (JA), and jasmonoyl-isoleucine (JA-Ile) were also analyzed through liquid chromatography-tandem MS (LC-MS/MS). Embryos were taken between 8 – 17 DAP for these analyses, to match the timing in the transcriptomics experiment, though it was not possible to acquire sufficient quantities of 7 DAP embryos to achieve reliable dry weight and metabolite or phytohormone

measurements. The wild type and *val1* embryos are unaltered with respect to both timing and accumulation of all detectable metabolites (Figure A.7) with the exception of the raffinose family oligosaccharides (RFO) galactinol and raffinose (Figure 4A). Both compounds began accumulating at 13 DAP in both wild type and *val1* mutant embryos, but had accumulated to lower levels in *val1* than the wild type by 17 DAP. To investigate whether the reduction in RFO sugars observed in *val1* embryos is carried over into mature seeds, we analyzed the levels of raffinose and galactinol and found no differences between the wild type and *val1* mutant (n = 8, p-value = 0.259 (galactinol) and p-value = 0.741 (raffinose)).

Several isoforms of GALACTINOL SYNTHASE (GOLS) are found in Arabidopsis. GOLS1 and 2 are known to be involved in galactinol synthesis. *GOLS1* and *2* were expressed at lower levels in *val1* than in the wild type (Figure 4B) while *RAFFINOSE SYNTHASE 5* (*RS5*) transcript levels were not affected in the *val1* mutant (Figure A.4). To investigate whether the reduction in RFO sugars affects the ability of *val1* seeds to withstand desiccation, developing wild type and *val1* seeds (collected at 15, 17, 19, and 21 DAP) were subjected to two days of intense drying to mimic the natural desiccation process as described by (Ooms et al., 1993). No differences between the wild type and *val1* mutant embryos were observed, as all seeds, regardless of age and genotype, were capable of germination. Furthermore, we assayed the desiccation tolerance of after-ripened wild type and *val1* seeds exposed to 10 days of intense drying, and 100% of the seeds germinated.

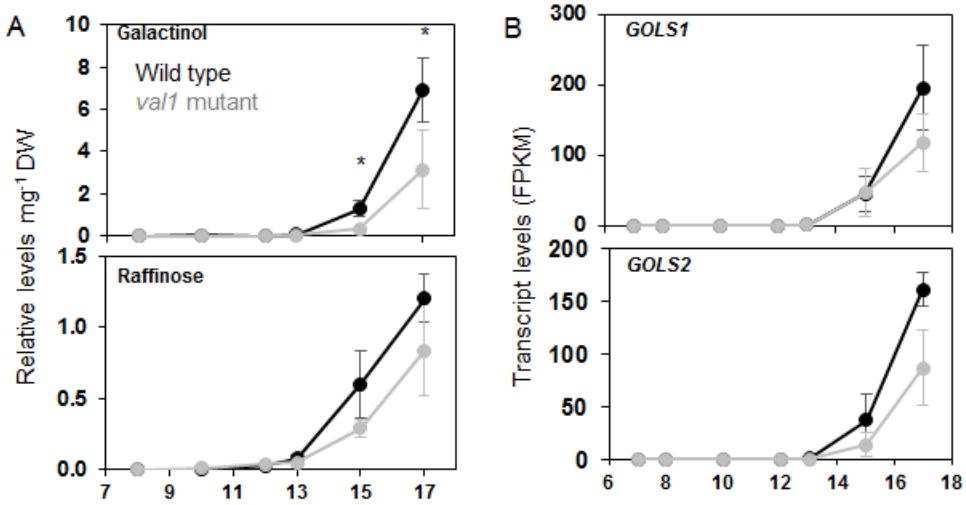

Figure 4. Relative levels of RFO sugars and transcripts encoding galactinol synthase isoforms. A) Levels of galactinol and raffinose are reduced in *val1* embryos during the transition to the desiccation phase. B) Levels of two galactinol synthase enzymes are reduced during this same transition. Asterisk indicates *p*-value < 0.05.

2.3.8 The *val1* mutant phenotypes are not caused by alterations in ABA or auxin levels

Seed development involves several phase transitions that are regulated by phytohormones, in particular the ratio of ABA and GA (Finkelstein et al., 2002; Nambara and Marion-Poll, 2005; Finkelstein, 2013). Based on the results of our transcriptomics analysis, *VAL1* transcript levels peak in the embryo during the transition from early to middle maturation, suggesting that VAL1 may be involved in helping to regulate this transition. This timing also represents a period during which maternal ABA levels increase dramatically to promote seed maturation and to inhibit precocious germination (Kanno et al., 2010). To investigate if differences in levels of ABA, JA or auxin could be contributing to the transcriptional changes and metabolic phenotypes observed in the *val1* mutant, the levels of ABA, auxin, OPDA, JA, and JA-Ile were analyzed in developing embryos. No differences in the levels of these phytohormones were identified between developing wild type and *val1* embryos (Figure A.8), suggesting that the phenotypes observed in *val1* are independent of these phytohormones. However, substantial upregulation in the expression of several GA catabolic and biosynthetic

genes was observed throughout the time course (Table 3). Potential changes in GA levels could alter the ABA/GA ratios and affect the seed maturation program.

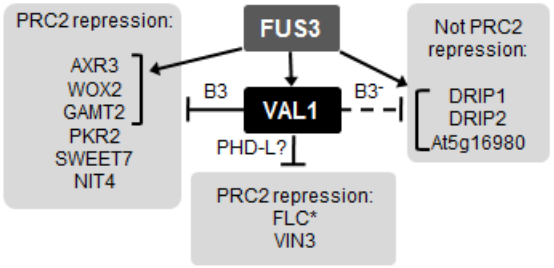

Figure 5. VAL1 represses genes in the FUS3 regulon through several potential mechanisms. The VAL1 B3 domain could target several genes for PRC2-mediated deposition of H3K27me3 marks, including the FUS3-regulated AXR3 and GAMT2. In addition, the CHD3 repressor PKR2 is repressed completely in wild type embryos and upregulated in the *val1* mutant. In addition, the putative sucrose transporter SWEET7 and the nitrilase NIT4 are repressed by PRC2 and have the RY/Sph motif in their promoters. Genes encoding the members of the FUS3 regulon FLC and VIN3 lack the RY/Sph motif, but are repressed through PRC2 activity. These two genes could be targeted by the VAL1 PHD-L domain. DRIP1 and 2, and At5g16980 are regulated by FUS3, but do not have the RY/Sph motif, nor they are repressed by PRC2. They could be indirectly repressed by VAL1 or targeted through a different, B3-independent ($B3^-$) mechanism. In the *val1* mutant, these genes are de-repressed to different degrees that are consistent with whether they are regulated by VAL1 alone or both VAL1 and FUS3. FUS3-activated genes are indicated by brackets and the asterisk for FLC.

2.4 Discussion

2.4.1 VAL1 can recognize target genes through two distinct mechanisms

The full-length VAL1 protein contains two domains that are known in other proteins to bind to chromatin or DNA (the PHD-L and B3 domains, respectively), although direct binding of the VAL1 domains have not been detected. Taken together, these two domains provide VAL1 with both epigenetic and direct transcriptional repressor capabilities. However, the question of which domain is involved in the repression of any given gene remains open. Perhaps both domains are required to repress the same genes, or each domain could repress an independent group of genes, with a possible overlap. Our results suggest that the VAL1 protein takes on two distinct forms, due to alternative splicing, that could repress genes by different mechanisms (Figure 5). The novel protein variant lacks the PHD-L domain, leaving a protein containing intact B3 and CW domains, and the EAR motif. This protein variant could only target genes through the B3 domain by binding directly to the RY/Sph motif. However, the VAL1 CW domain has been demonstrated to bind to H3K3me3 (Hoppmann et al., 2011) and this domain could provide chromatin binding capability to the novel splice variant or it could interact with other chromatin modifiers, as VAL2 does (Zhou et al., 2013). The full-length protein contains an intact PHD-L domain and has the capacity to target genes through binding to H3K4me3. In addition to seed development, VAL1 is expressed during seedling growth and vegetative tissues when the chromatin state is different than in the seed (van Zanten et al., 2011). The particular binding mechanism utilized by VAL1 may be changed depending on the chromatin state.

2.4.2 VAL1 is a global epigenetic and transcriptional regulator acting downstream of core LAFL transcriptional regulators in developing Arabidopsis embryos

In this study, we have shown that VAL1 has a novel, global role in regulating embryo development, as 5,320 transcripts were differentially expressed in *val1* embryos in at least one time point relative to the wild type. Although VAL1 has been shown to regulate the core LAFL TFs in seedlings in conjunction with VAL2 (Suzuki et al., 2007), expression of these genes was not affected in *val1* embryos, suggesting that VAL1 activity is not a consequence of direct repression of these TFs in developing seeds, in agreement with the conclusions of Jia et al. (2013). However, VAL1 does regulate other genes in the LAFL network. For example, transcripts of the FUS3 target *WOX2* are elevated in *val1* embryos. WOX2, in turn, activates the *CUC1* gene, which does not contain the methylation mark or the RY motif. As WOX2 is repressed through the action of PRC2-catalyzed H3K27me3 and it contains the RY-binding motif, it is a candidate for direct interaction with VAL1. VIN3, which is also repressed through the action of the PHD domain, and whose activity is dependent on histone acetylation, is another candidate for a VAL1 target, and may be activated in the *val1* mutant through increased histone acetylation, as evidenced by an increased expression of a histone acetyltransferase (At1g16705). In addition, although VIN3 represses *FLC* expression in response to cold by promoting PRC2-mediated repression, the elevated *VIN3* transcript levels in *val1* embryos do not appear to be a compensatory mechanism for the elevated *FLC* expression. Indeed, *FLC* transcript levels remained elevated for several days in *val1* embryos after *VIN3* transcript levels returned to wild type levels.

VAL1 is activated by FUS3, making it part of the LAFL network (Wang and Perry, 2013). VAL1 appears to repress FUS3 targets, as 701 transcripts that were upregulated in *val1*

mutant embryos also belonged to the FUS3 regulon. Overall, 38.3% of genes in the FUS3 regulon were upregulated in *val1* embryos at least one time point. In addition, several major targets of FUS3 and LEC2, including several TFs and phytohormone synthesis and response genes, were upregulated in *val1* embryos (Table 3). Several of these genes contained the RY/Sph motif and also carried the H3K27me3 mark, which could enable their repression.

TTG1 is another repressor of seed maturation gene expression that was recently reported to be repressed by FUS3 (Chen et al., 2015). TTG1 was found to repress the expression of the *ABI3* and *LEC2* genes, both encoding activators of *FUS3* expression. The model proposed by (Chen et al., 2015) suggests that *TTG1* transcription is repressed by FUS3 to maintain the expression of seed storage protein genes. When FUS3 activity is reduced, TTG1 expression will be elevated, and the genes of ABI3, LEC2, and several other seed-specific regulatory proteins will be indirectly repressed, thus forming a mechanism to repress the seed maturation program and maintain the appropriate level of seed storage protein synthesis. Our results suggest that VAL1 performs a somewhat similar role with respect to FUS3 targets, repressing a subset of them when they are no longer required.

TTG1 mutant seeds had elevated protein levels and the corresponding siliques showed increased expression of four of the five albumin genes (Chen et al., 2015). Similar to TTG1, *val1* dry seeds have elevated protein levels. However, there were no major changes in transcript and protein levels of any of the seed storage proteins in *val1* mutant embryos. As such, cell-specific involvement of VAL1 cannot be ruled out and the modest increase in protein levels that was observed in dry seeds could be attributed to the endosperm and seed coat, which contribute to over 30% of seed DW in Arabidopsis and are known to accumulate seed storage compounds (Li et al., 2006). Overall, involvement of VAL1 in suppressing the accumulation of seed storage

proteins is not entirely unexpected, as seed storage protein levels were found to be elevated in *val1*/*val2* mutant seedlings (Tsukagoshi et al., 2007) and several seed storage protein genes were upregulated in *val1* seedlings (Suzuki et al., 2007).

2.4.3 VAL1 is not essential for embryo development and metabolism and acquisition of seed desiccation tolerance

The physical appearance, growth, and metabolomes of developing wild type and *val1* mutant embryos were remarkably similar, which was unexpected considering the extensive global changes detected in the corresponding transcriptomes. However, moderate, but consistent decreases in the levels of the RFO sugars galactinol and raffinose and transcripts relevant to the synthesis of galactinol were observed in the *val1* mutant when compared to the wild type embryos. Galactinol is synthesized from UDP-α-D-galactose through the activity of three known GOLS (Taji et al., 2002; Nishizawa et al., 2008). Subsequently, RS5 converts galactinol to raffinose (Anderson and Kohorn, 2001). The synthesis of galactinol is the regulatory step in the synthesis of raffinose (Taji et al., 2002), and *GOLS1* and *GOLS2* transcript levels were reduced in *val1* embryos compared to the wild type, which could explain the observed reduction in raffinose levels without alterations in *RS5* transcript levels.

Furthermore, DRIP1 and DRIP2 are implicated as negative regulators of *GOLS1* and *2* expression (Qin et al., 2008). Both *DRIP1* and *2* expression levels are increased at all time points, which could result in the observed reductions in the levels of *GOLS1* and *2* transcripts. DRIP1 and 2 also both interact with VAL1 in seedlings (Yang et al., 2013), where they are required for ubiquitination of histone protein 2A (a repressive epigenetic mark) along with VAL2. Triple knockout mutants of DRIP1, 2 and the DRIP homolog BMI1c have the same

seedling phenotype as *val1/2* double mutants. The model proposed by Yang et al. (2013) is that VAL1/2 interact with DRIP1, 2 and BMI1c to catalyze ubiquitination of histone H2A, followed by PRC2-mediated histone trimethylation (H3K27me3). In *val1* embryos, the VAL1 protein is non-functional, while VAL2 still functions. The increased expression of *DRIP1* and *DRIP2* could be a mechanism to compensate for the absence of VAL1, with an additional, and, perhaps unrelated effect on galactinol, and subsequently raffinose levels. Consistent with this hypothesis, no reductions in desiccation tolerance was observed in developing and mature seeds, suggesting that the synthesis of other desiccation-related compounds is unaffected in *val1* embryos.

2.4.4 Robust responses among stress-related genes in the *val1* mutant connects VAL1 to various types of stresses

The diversity within the group of stress-related genes in the *val1* mutant suggests that a common mechanism underlies their increased expression. A comparable result was obtained by Sharma et al. (2013) in a study of effects of drought on *val1* seedlings, where upregulation of a number of stress-responsive genes was observed in the unstressed double mutant. The most obvious core mechanism involved in responses to all stresses, abiotic and biotic, is an increase in reactive oxygen species (ROS). Sharma et al. (2013) suggested that an increase in ROS could underlie the observed upregulation, this being a central initial event in stress signaling. However, the repression by VAL1 of stress-responsive genes (Figure 6) in the absence of stress may represent a hitherto unrecognized regulatory function, which is released when stress is applied. Upon the imposition of stress, the repression would be abolished, perhaps by the inactivation of the repressors, and stress signaling would be initiated. If this is the case, ROS levels may not be

altered in VAL1 mutants, since the signaling events involved would be downstream of ROS generation. Further experimentation is needed to test this hypothesis.

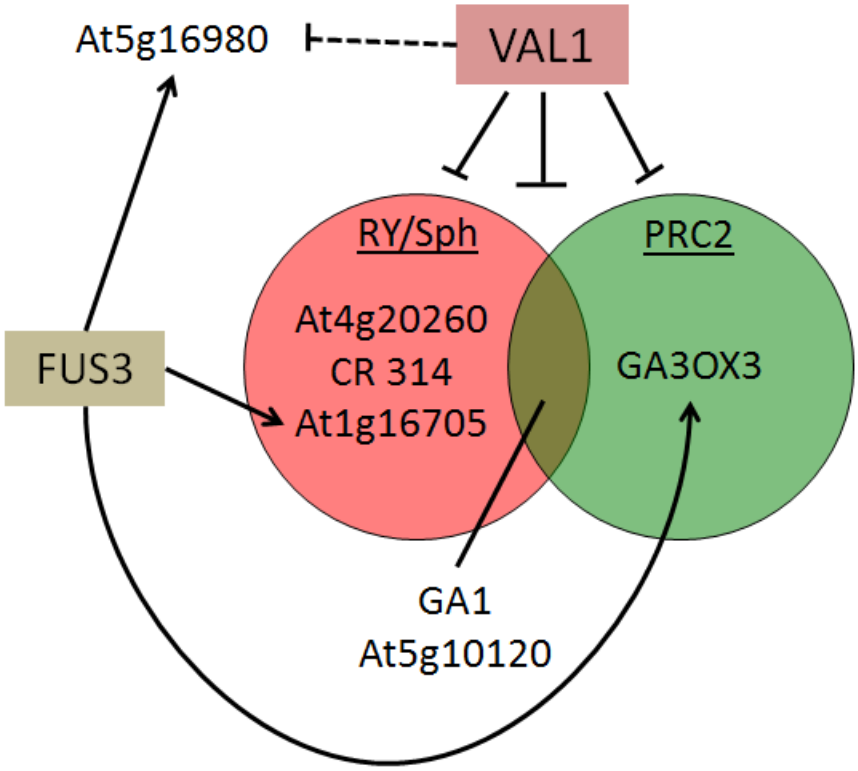

Figure 6. Selected hormone and stress-response genes repressed by VAL1 through several possible mechanisms. Three genes contain the RY/Sph motif in their promoters, while one carries the PRC2-mediated repression mark. Two genes, At4g02780 and At5g10120, have both the RY/motif and carry the PRC2 mark. One gene, At5g16980, contains neither mark, and may be indirectly regulated by VAL1. Genes activated by FUS3 are indicated with arrows. Solid lines indicate possible direct regulation; dashed line indicates possible indirect regulation. CR314 (At1g29395), GA1 (At4g02780), GA3OX3 (At4g21690).

2.5 Conclusions

2.5.1 A model for VAL1 function in developing embryos

VAL1 has an established role in gene repression through chromatin modification. Embryo development depends heavily on the activity of the core LAFL network TFs. VAL1 does not interfere with expression of the core LAFL network. However, considering that these four TFs activate genes required for different phases of seed development, some factor(s) is (are) required to repress LAFL target genes when they are no longer necessary. Such repression may require alteration to the chromatin state provided by covalent post-translational modifications as opposed to that provided for by active repression, such as a repressor binding to a promoter. Chromatin modification would be required to repress certain genes as the LAFL TF(s) responsible for activating these genes are still very much expressed and enabling active transcription. By repressing specific targets of LAFL, genes that are no longer required can be effectively repressed without interfering with LAFL expression. One major target of VAL1 seems to be genes belonging to the FUS3 regulon, though a subset of genes belonging to LEC1 and 2 regulons was also de-repressed in the absence of VAL1. As many of the LAFL targets are activated by binding to the RY motif, VAL1 may utilize its B3 domain to repress these targets through competition with activators of gene expression, as has been suggested by (Guerriero et al., 2009), and subsequently recruit chromatin modifiers. The VAL1 splice variant missing the PHD-L domain was expressed at higher levels than the full-length transcript, which suggests that this may be the predominant functional form of VAL1 in developing Arabidopsis embryos. The full-length protein containing the PHD-L domain may be required to repress genes that are not part of the LAFL network, or genes that do not contain an RY motif.

2.6 Materials and methods

2.6.1 Chemicals

All reagents and metabolite standards were of analytical purity and purchased from Sigma-Aldrich (St. Louis, MO) or Thermo Fisher Scientific (Waltham, MA). Solvents used for embryo work, extracting metabolites, and LC-MS/MS were MS grade, while solvents used for UPLC, GC, and other applications were LC or GC grade. Multiply labeled internal standards were obtained from Cambridge Isotope Laboratories, Inc. (Tewksbury, MA).

2.6.2 Plant growth, silique and seed harvesting, and embryo dissections

Seeds were cold stratified (4 ^{o}C) for 3 – 5 days in 0.1% agarose and sown onto moist Farfard superfine germinating mix (Sun Gro Horticulture, Vancouver, British Columbia, Canada) in 2-inch pots. Plants were grown in a controlled environment at ~115 μE, 20/18 ^{o}C day/night, 65% humidity, and a 16-hour photoperiod. Watering was done with deionized water every 3 – 4 days until they were 3 – 4 weeks old. After that they received Miracle Gro All-Purpose Plant Food (1 tsp per 2 gallons) with every watering. For storage compound and metabolite analyses in dry seeds, a newly opened flower on the main stem was marked (0 DAP) and after 21 days, the siliques were dry. The four siliques above and below this marked silique were collected to retrieve sufficient amount of material for dry seed analyses.

All three time course experiments (transcriptomics, metabolomics, and phytohormone level analysis) were set up in the same manner. To ensure similar nutrient allocation and light exposure to the developing seeds, the time course was set up in stages: stage 1 (7 and 8 DAP), stage 2 (10, 12, and 13 DAP), and stage 3 (15 and 17 DAP), such that embryos belonging to the same stage were harvested on the same day and individual stages within days apart. Flowers

were tagged, starting with the latest time point within each stage. Each set of stages represented a biological replicate. Seeds and embryos from the tagged siliques were dissected at low temperatures to slow down any metabolic processes. This was achieved by performing these dissections on wet glass-fiber filters placed on Petri dishes completely filled with frozen water and ensuring that temperatures were maintained between -5 and 5 ^{o}C by frequent measurements with an infrared thermometer. We found that these temperatures are the lowest that can be used without freezing the seeds, which would cause cell damage that would be problematic for the subsequent embryo washing steps.

For different applications, slightly different approaches in embryo harvesting were taken. For example, transcriptomics did not require DW determination and embryos could be harvested a little bit faster than for metabolomics. Depending on the developmental stage of the embryos and application, embryo dissection from a single silique (40 – 60 seeds) took between 10 and 25 minutes. We acknowledge that these are long times for metabolomics purposes, as metabolites turn over within seconds to minutes (Fiehn, 2002; Collakova et al., 2008; Ma et al., 2014). However, care was taken to perform dissections at low temperatures and wild type and *val1* mutant embryos were dissected in parallel by two persons. In addition, the metabolomics experiment was performed three times and all conclusions are based on comparing wild type and *val1* mutant embryos. Therefore, it is reasonable to assume that any changes to the metabolomes due to sample preparations would be to the same degree in both wild type and *val1* mutant embryos. It is noteworthy that, collectively, over 35,000 embryos were collected for the entire study. The following text provides descriptions of two different approaches undertaken to dissect sufficient amounts of Arabidopsis embryos for transcriptomics, and metabolomics and phytohormone analyses.

For transcriptomics, embryos were rinsed off from the glass-fiber filter into ice-cold MS-grade water, briefly centrifuged at 4 ^{o}C prior to an additional wash to remove any remaining endosperm, snap-frozen in liquid nitrogen, and stored at -80 ^{o}C until processed. For metabolomics and hormone analyses, the procedure was similar, except that dissected embryos were transferred with a metal probe into ice-cold MS-grade water to prevent transfer of glass fibers into the sample, which would interfere with the subsequent DW determination on freeze-dried embryos. The lyophilization was performed on a FreeZone 2.5 freeze drier (Labconco, Kansas City, MO) for two days.

2.6.3 Analyses of seed storage compounds in dry seeds

For dry seed studies, 1.000 mg (± 5%) of dry seed was measured on an XP-26 analytical microbalance (Mettler Toledo, Columbus, OH) and lipids and proteins were extracted and analyzed by GC-FID and a fluorescent hydrophobic protein assay as described in Collakova et al. (2013). Briefly, lipid-derived fatty acids were analyzed as fatty acid methyl esters using heptadecanoic acid as an internal standard on an Agilent 7890A series GC-FID (Agilent Technologies, Santa Clara, CA) equipped with a 30-m DB-23 column (0.25 mm x 0.25 µm, Agilent Technologies). Proteins were solubilized according to Hou et al. (2005) and a plate-reader-based Marker Gene Hydrophobic Protein Analysis Kit (Marker Gene Technologies, Inc., Eugene, OR) using bovine serum albumin as a standard was used according to the manufacturer's recommendations. The fluorescent compound 6-(*p*-toluidino)-2-naphthalenesulfonic acid used in this kit binds only to hydrophobic regions of proteins and is not influenced by high concentrations of strong detergents needed to solubilize highly hydrophobic

seed storage proteins. Protein composition was determined on 15% gels by using SDS-PAGE according to Hou et al. (2005).

2.6.4 Metabonomics of Arabidopsis embryo development

Metabonomics (a time-course metabolomics), including data processing and analysis, was performed on Arabidopsis embryos and dry seeds as described for developing soybean embryos in Collakova et al. (2013) with minor modifications. Briefly, due to their small size, it was not possible to obtain a milligram of dry material for analysis of all stages of embryo development and depending on the stage, between 0.3 and 1.4 mg of DW was collected for each sample. Biphasic extractions were used to separate polar and non-polar metabolites and insoluble proteins. Internal standards included (i) heptadecanoic acid for free and lipid-derived fatty acids, (ii) [U-$^{13}C_6$]-glucose for sugars, sugar alcohols, sugar acids, and sugar-phosphates, (iii) [2,2,4,4-D_4]-citrate for carboxylic and organic acids, and (iv) norvaline for amino acids and organic amines. Organic phase was analyzed for fatty acids by GC-FID as described for dry seeds. A portion of aqueous phase (5%) was used for amino acid and organic amine analysis using Waters AccQ-Tag^{TM} Ultra Kit and an H-class Acquity UPLC-FLD equipped with a 10-cm Waters AccQ-Tag^{TM} Ultra C18 (1.7 µm x 2.1 mm) column (Waters Corporation, Milford, MA) using Waters 10.2-min method for analyzing free amino acids as described in Collakova et al. (2013). The remaining aqueous phase was used to analyze other polar metabolites, which were separated as trimethyl silyl derivatives on an Agilent 7890A series GC equipped with a DB-5MS-DG column (30 m length × 0.25 mm × 0.25 µm with a 10-m pre-column, Agilent Technologies) and analyzed on an Agilent 5975C series single quadrupole MS (Collakova et al., 2013). The remaining insoluble interphase contained hydrophobic proteins and cell-wall material and was

used to analyze relative seed storage protein levels. We tested that seed storage proteins partitioned in the interphase and not in the aqueous phase (Figure A.9).

2.6.5 Phytohormone analysis using LC-MS/MS

Absolute levels of ABA, IAA, OPDA, JA, and JA-Ile were quantified in developing embryos (three biological replicates, seven time points as for the metabonomics experiment) by UPLC coupled to electrospray ionization tandem (ESI) tandem MS. We were unable to analyze GA levels due to their low levels in developing embryos and rather large amounts of material (e.g., 200 mg) needed for their analyses (Varbanova et al., 2007) that could not be obtained. Phytohormone extractions and quantification were based on previously established methods (Muller and Munne-Bosch, 2011; Koo et al., 2014) with minor modifications. Briefly, phytohormones were extracted from 100 - 300 lyophilized embryos in a 300-µl volume with a mixture of methanol:isopropanol:glacial acetic acid at the 40/59/1 (v/v/v) ratios at 4 °C for 30 minutes. Multiply labeled authentic compounds were used for each phytohormone as internal standards to obtain absolute levels. Five µL of extract were injected on a UPLC BEH C18 column (1.7 µm, 2.1 x 50 mm; Waters) maintained at 40 °C and attached to an H-class Acquity UPLC system (Waters, Milford, MA). A 3-min gradient program was applied using methanol and 0.15% aqueous formic acid as mobile phases with a 0.4 ml min^{-1} flow rate. Multiple reaction monitoring (MRM) was employed to detect characteristic precursor to product ion transitions for phytohormones and their internal standards: ABA (m/z 263 > 153), d_6-ABA (269 > 159), OPDA (291 > 165), d_5-OPDA (296 > 170), JA (209 > 59), dihydro-JA (211 > 59), JA-Ile (322 > 130), and $^{13}C_6$-JA-Ile (328 > 136) using Waters Xevo TQ-S MS operated at ESI negative ion mode. A second channel was set to monitor IAA (m/z 176 > 130) and d_5-IAA (181 > 135) simultaneously

in a positive ion mode. Data were acquired under the control of MassLynx 4.1 software and chromatographic peaks were integrated using TargetLynx Application Manager (Waters). Absolute quantification of analytes was based on standard curves comparing analyte responses to the corresponding authentic internal standards.

2.6.6 RNAseq analysis

RNA was extracted from frozen embryos according to Onate-Sanchez and Vicente-Carbajosa (2008) and purified by using the RNeasy kit (Qiagen, Limburg, Netherlands). RNA quality and integrity analysis was performed by the Virginia Bioinformatics Institute (Virginia Tech, Blacksburg, VA). Reverse transcription, library preparation, and paired-end RNAseq generating 75-bp reads (HiSeq 2500 Ultra-High-Throughput sequencer, Illumina, San Diego, CA), including removing low quality reads and adapter sequence trimming, was performed by Beckman Coulter Genomics (Danvers, MA) using established protocols. Any additional data processing was done in house. Each library was sequenced four times and samples were randomly multiplexed among eight lanes. Therefore, the abundances of a transcript in these four different sequencing runs are expected to be very close, unless there is a problem with a lane. If abundance of a transcript was significantly different in a particular sequencing run (out of four runs) from the average abundance of all the runs, it was captured and removed using t-test (*p*-value < 0.05). The abundance of a transcript in a sample was calculated using the average abundance of the runs, which passed the t-test. This approach enabled retaining high quality data without sacrificing the number of biological replicates.

All subsequent steps, starting with mapping reads to the TAIR10 version of the Arabidopsis genome using Tophat2, were performed using the Tuxedo Suite (Trapnell et al.,

2009; Trapnell et al., 2010; Roberts et al., 2011) as described previously (Aghamirzaie et al., 2013), except that StringTie (Pertea et al., 2015) instead of Cufflinks was used for transcript assembly. On average, 95% of the reads mapped to the genome and 1.5% of the reads were multi-mapped. The only defective sample was the 17 DAP replicate 3 for *val1*, in which only 5.9% reads mapped to the genome, and it was removed from further analyses. Assembled transcripts from all samples were merged using Cuffmerge to generate a reference transcriptome GTF file (transcriptome.gtf) using Arabidopsis reference GTF as a guide, which facilitates identification of different types of transcripts using Cuffcompare class codes. StringTie-B, together with transcriptome.gtf, was used to prepare read coverage tables from aligned reads.

Limma was chosen for performing differential expression analysis (Ritchie et al., 2015). Limma accepts raw counts as input, but StringTie only provides raw counts for exons and introns of a transcript. Therefore, an in-house script was written to calculate raw count abundance of each transcript. These raw counts were normalized using the Voom package (Law et al., 2014), as required for Limma prior to performing differential gene expression analysis. A moderated t-statistic was used to assess the differential transcript expression between the wild type and *val1* mutant embryos at each time point. Empirical Bayes method was used to shrink the probe-wise sample variances towards a common value (Smyth, 2004). F-statistic was used to test if a transcript is differentially expressed at any time point. The *p*-values were adjusted for multiple testing using the Benjamini and Hochberg method to control the false discovery rate (Benjamini and Hochberg, 1995). Transcripts with adjusted *p*-value < 0.05 were declared to be differentially expressed. Raw count abundance values were normalized to the library size and transcript sizes to obtain Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values. Database analyses and overlap determinations (e.g., how many transcripts were included in the

FUS3 regulon) were performed using a combination of Excel (Microsoft) and in-house scripting. GO enrichment analysis was performed at http://bar.utoronto.ca/ntools/cgi-bin/ntools_classification_superviewer.cgi.

2.6.7 Semi-quantitative real-time reverse transcription PCR (qPCR)

The RNAseq data was validated using TaqMan qPCR. Briefly, RNA was converted to cDNA using the TaqMan Reverse Transcription kit (Life Technologies, Carlsbad, CA) following the manufacturer's instructions. TaqMan reactions were performed with commercially available primers and probe kits acquired from Life Technologies. *VAL1* full-length and novel splice variant were validated with custom made primers and probes for each time point in both wild type and *val1* mutant embryos. Other selected transcripts were validated for selected time points. The expression of many commonly used reference genes (e.g., tubulin, elongation factor alpha, etc.), change dramatically in developing embryos (Figure A.10), including those that are seed specific (Dekkers et al., 2012). To identify suitable reference transcripts, the average expression of each time point was taken and the fold change between one time point and the next was compared across the time course. Positive and negative cutoffs of 1.15 and 0.85, respectively, were used to identify 16 transcripts that did not change. The At2g3053 transcript had the most stable expression across the time course and was selected for transcript expression normalization purposes. Normalization of each sample was performed with the $2^{-\Delta CT}$ method (Schmittgen and Livak, 2008).

References

Aghamirzaie, D., Nabiyouni, M., Fang, Y., Klumas, C., Heath, L.S., Grene, R., and Collakova, E. (2013). Changes in RNA splicing in developing soybean (*Glycine max*) embryos. Biology 2, 1311-1337.

Aichinger, E., Villar, C.B., Farrona, S., Reyes, J.C., Hennig, L., and Kohler, C. (2009). CHD3 proteins and polycomb group proteins antagonistically determine cell identity in Arabidopsis. PLoS genetics 5, e1000605.

Anderson, C.M., and Kohorn, B.D. (2001). Inactivation of Arabidopsis SIP1 leads to reduced levels of sugars and drought tolerance. Journal of Plant Physiology 158, 1215–1219.

Babiychuk, E., Kushnir, S., Belles-Boix, E., Van Montagu, M., and Inze, D. (1995). Arabidopsis thaliana NADPH oxidoreductase homologs confer tolerance of yeasts toward the thiol-oxidizing drug diamide. The Journal of biological chemistry 270, 26224-26231.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Nobel, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic acids research 37, 202-208.

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

Baud, S., Dubreucq, B., Miquel, M., Rochat, C., and Lepiniec, L. (2008). Storage reserve accumulation in Arabidopsis: metabolic and developmental control of seed filling. The Arabidopsis book / American Society of Plant Biologists 6, e0113.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57, 289-300.

Bond, D.M., Dennis, E.S., Pogson, B.J., and Finnegan, E.J. (2009). Histone acetylation, VERNALIZATION INSENSITIVE 3, FLOWERING LOCUS C, and the vernalization response. Molecular plant 2, 724-737.

Braybrook, S.A., Stone, S.L., Park, S., Bui, A.Q., Le, B.H., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (2006). Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. Proceedings of the National Academy of Sciences of the United States of America 103, 3468-3473.

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.H., Sung, Z.R., and Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. Development 131, 5263-5276.

Chen, M., Zhang, B., Li, C., Kulaveerasingam, H., Chew, F.T., and Yu, H. (2015). TRANSPARENT TESTA GLABRA 1 regulates the accumulation of seed storage reserves in Arabidopsis. Plant Physiol.

Collakova, E., Goyer, A., Naponelli, V., Krassovskaya, I., Gregory, J.F., 3rd, Hanson, A.D., and Shachar-Hill, Y. (2008). Arabidopsis 10-formyl tetrahydrofolate deformylases are essential for photorespiration. Plant Cell **20,** 1818-1832.

Collakova, E., Aghamirzaie, D., Fang, Y., Klumas, C., Tabataba, F., Kakumanu, A., Myers, E., Heath, L.S., and Grene, R. (2013). Metabolic and transcriptional reprogramming in developing soybean (*Glycine max*) embryos. Metabolites 3, 347-372.

Dekkers, B.J., Willems, L., Bassel, G.W., van Bolderen-Veldkamp, R.P., Ligterink, W., Hilhorst, H.W., and Bentsink, L. (2012). Identification of reference genes for RT-qPCR expression analysis in Arabidopsis and tomato seeds. Plant & cell physiology 53, 28-37.

Fiehn, O. (2002). Metabolomics--the link between genotypes and phenotypes. Plant Mol Biol **48,** 155-171.

Finkelstein, R. (2013). Abscisic Acid synthesis and response. The Arabidopsis book / American Society of Plant Biologists 11, e0166.

Finkelstein, R.R., Gampala, S.S., and Rock, C.D. (2002). Abscisic acid signaling in seeds and seedlings. Plant Cell 14 Suppl, S15-45.

Guerriero, G., Martin, N., Golovko, A., Sundstrom, J.F., Rask, L., and Ezcurra, I. (2009). The RY/Sph element mediates transcriptional repression of maturation genes from late maturation to early seedling growth. The New phytologist 184, 552-565.

Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L., and Luo, J. (2005). DATF: a database of Arabidopsis transcription factors. Bioinformatics 21, 2568-2569.

Higashi, Y., Hirai, M.Y., Fujiwara, T., Naito, S., Noji, M., and Saito, K. (2006). Proteomic and transcriptomic analysis of Arabidopsis seeds: molecular evidence for successive processing of seed proteins and its implication in the stress response to sulfur nutrition. The Plant journal : for cell and molecular biology 48, 557-571.

Hoppmann, V., Thorstensen, T., Kristiansen, P.E., Veiseth, S.V., Rahman, M.A., Finne, K., Aalen, R.B., and Aasland, R. (2011). The CW domain, a new histone recognition module in chromatin proteins. EMBO J 30, 1939-1952.

Hou, A., Liu, K., Catawatcharakul, N., Tang, X., Nguyen, V., Keller, W.A., Tsang, E.W., and Cui, Y. (2005). Two naturally occurring deletion mutants of 12S seed storage proteins in Arabidopsis thaliana. Planta 222, 512-520.

Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., Konagaya, A., and Shinozaki, K. (2005). RARTF: database and tools for complete sets of Arabidopsis transcription factors. DNA research : an international journal for rapid publication of reports on genes and genomes 12, 247-256.

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

Jia, H., Suzuki, M., and McCarty, D.R. (2014). Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. WIREs Dev Biol 3, 135-145.

Junker, A., Monke, G., Rutten, T., Keilwagen, J., Seifert, M., Thi, T.M., Renou, J.P., Balzergue, S., Viehover, P., Hahnel, U., Ludwig-Muller, J., Altschmied, L., Conrad, U., Weisshaar, B., and Baumlein, H. (2012). Elongation-related functions of LEAFY COTYLEDON1 during the development of Arabidopsis thaliana. The Plant journal : for cell and molecular biology 71, 427-442.

Kagale, S., and Rozwadowski, K. (2011). EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression. Epigenetics : official journal of the DNA Methylation Society 6, 141-146.

Kanno, Y., Jikumaru, Y., Hanada, A., Nambara, E., Abrams, S.R., Kamiya, Y., and Seo, M. (2010). Comprehensive hormone profiling in developing Arabidopsis seeds: examination of the site of ABA biosynthesis, ABA transport and hormone interactions. Plant & cell physiology 51, 1988-2001.

Koo, A.J., Thireault, C., Zemelis, S., Poudel, A.N., Zhang, T., Kitaoka, N., Brandizzi, F., Matsuura, H., and Howe, G.A. (2014). Endoplasmic reticulum-associated inactivation of the hormone jasmonoyl-L-isoleucine by multiple members of the cytochrome P450 94 family in Arabidopsis. The Journal of biological chemistry 289, 29728-29738.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome biology 15, R29.

Le, B.H., Cheng, C., Bui, A.Q., Wagmaister, J.A., Henry, K.F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S., Drews, G.N., Fischer, R.L., Okamuro, J.K., Harada, J.J., and Goldberg, R.B. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proceedings of the National Academy of Sciences of the United States of America 107, 8063-8070.

Lee, J., Yun, J.Y., Zhao, W., Shen, W.H., and Amasino, R.M. (2015). A methyltransferase required for proper timing of the vernalization response in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 112, 2269-2274.

Li, Y., Beisson, F., Pollard, M., and Ohlrogge, J. (2006). Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. Phytochemistry 67, 904-915.

Lie, C., Kelsom, C., and Wu, X. (2012). WOX2 and STIMPY-LIKE/WOX8 promote cotyledon boundary formation in Arabidopsis. The Plant journal : for cell and molecular biology 72, 674-682.

Ma, F., Jazmin, L.J., Young, J.D., and Allen, D.K. (2014). Isotopically nonstationary 13C flux analysis of changes in Arabidopsis thaliana leaf metabolism due to high light acclimation. Proceedings of the National Academy of Sciences of the United States of America **111,** 16967-16972.

Meinke, D.W. (1995). Molecular Genetics of Plant Embryogenesis. Annual Review of Plant Physiology and Plant Molecular Biology 46, 369-394.

Mendes, A., Kelly, A.A., van Erp, H., Shaw, E., Powers, S.J., Kurup, S., and Eastmond, P.J. (2013). bZIP67 regulates the omega-3 fatty acid content of Arabidopsis seed oil by activating fatty acid desaturase3. Plant Cell 25, 3104-3116.

Monke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H.P., Baumlein, H., and Conrad, U. (2004). Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. Planta 219, 158-166.

Monke, G., Seifert, M., Keilwagen, J., Mohr, M., Grosse, I., Hahnel, U., Junker, A., Weisshaar, B., Conrad, U., Baumlein, H., and Altschmied, L. (2012). Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. Nucleic acids research 40, 8240-8254.

Mouriz, A., Lopez-Gonzalez, L., Jarillo, J.A., and Pineiro, M. (2015). PHDs govern plant development. Plant signaling & behavior, 10, e993253.

Muller, M., and Munne-Bosch, S. (2011). Rapid and sensitive hormonal profiling of complex plant samples by liquid chromatography coupled to electrospray ionization tandem mass spectrometry. Plant methods 7, 37.

Nambara, E., and Marion-Poll, A. (2005). Abscisic acid biosynthesis and catabolism. Annual review of plant biology 56, 165-185.

Nishizawa, A., Yabuta, Y., and Shigeoka, S. (2008). Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. Plant Physiol 147, 1251-1263.

Onate-Sanchez, L., and Vicente-Carbajosa, J. (2008). DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. BMC Res Notes 1, 93.

Ooms, J., Leon-Kloosterziel, K.M., Bartels, D., Koornneef, M., and Karssen, C.M. (1993). Acquisition of Desiccation Tolerance and Longevity in Seeds of Arabidopsis thaliana (A Comparative Study Using Abscisic Acid-Insensitive abi3 Mutants). Plant Physiology 102, 1185-1191.

Penfield, S., Pinfield-Wells, H.M., and Graham, I.A. (2006). Storage reserve mobilisation and seedling establishment in Arabidopsis. The Arabidopsis book / American Society of Plant Biologists 4, e0100.

Perez-Rodriguez, P., Riano-Pachon, D.M., Correa, L.G., Rensing, S.A., Kersten, B., and Mueller-Roeber, B. (2010). PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic acids research 38, D822-827.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature biotechnology 33, 290-295.

Qin, F., Sakuma, Y., Tran, L.S., Maruyama, K., Kidokoro, S., Fujita, Y., Fujita, M., Umezawa, T., Sawano, Y., Miyazono, K., Tanokura, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2008). Arabidopsis DREB2A-interacting proteins function as RING E3 ligases and negatively regulate plant drought stress-responsive gene expression. Plant Cell 20, 1693-1707.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 43, e47.

Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics **27,** 2325-2329.

Ruuska, S.A., Girke, T., Benning, C., and Ohlrogge, J.B. (2002). Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell 14, 1191-1206.

Sanchez, R., and Zhou, M.M. (2011). The PHD finger: a versatile epigenome reader. Trends in biochemical sciences 36, 364-372.

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

Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T., and Goodrich, J. (2006). Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. EMBO J 25, 4638-4649.

Shaikhali, J., Heiber, I., Seidel, T., Stroher, E., Hiltscher, H., Birkmann, S., Dietz, K.J., and Baier, M. (2008). The redox-sensitive transcription factor Rap2.4a controls nuclear expression of 2-Cys peroxiredoxin A and other chloroplast antioxidant enzymes. BMC plant biology 8, 48.

Sharma, N., Bender, Y., Boyle, K., and Fobert, P.R. (2013). High-level expression of sugar inducible gene2 (HSI2) is a negative regulator of drought stress tolerance in Arabidopsis. BMC plant biology 13, 170.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative CT method. Nature Protocols 3, 1101-1108.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology 3, Article3.

Suzuki, M., Wang, H.H., and McCarty, D.R. (2007). Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. Plant Physiol 143, 902-911.

Swaminathan, K., Peterson, K., and Jack, T. (2008). The plant B3 superfamily. Trends in plant science 13, 647-655.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P., and Huala, E. (2008). The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res 36, D1009-1014.

Taji, T., Ohsumi, C., Iuchi, S., Seki, M., Kasuga, M., Kobayashi, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2002). Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. The Plant journal : for cell and molecular biology 29, 417-426.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25,** 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology **28,** 511-515.

Tsukagoshi, H., Morikami, A., and 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, 2543-2547.

Tsukagoshi, H., Saijo, T., Shibata, D., Morikami, A., and 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 Physiol 138, 675-685.

van Zanten, M., Koini, M.A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., Koornneef, M., Fransz, P., and Soppe, W.J. (2011). Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. Proceedings of the National Academy of Sciences of the United States of America 108, 20219-20224.

Varbanova, M., Yamaguchi, S., Yang, Y., McKelvey, K., Hanada, A., Borochov, R., Yu, F., Jikumaru, Y., Ross, J., Cortes, D., Ma, C.J., Noel, J.P., Mander, L., Shulaev, V., Kamiya, Y., Rodermel, S., Weiss, D., and Pichersky, E. (2007). Methylation of gibberellins by Arabidopsis GAMT1 and GAMT2. Plant Cell 19, 32-45.

Veerappan, V., Chen, N., Reichert, A.I., and Allen, R.D. (2014). HSI2/VAL1 PHD-like domain promotes H3K27 trimethylation to repress the expression of seed maturation genes and complex transgenes in Arabidopsis seedlings. BMC plant biology 14, 293.

Veerappan, V., Wang, J., Kang, M., Lee, J., Tang, Y., Jha, A.K., Shi, H., Palanivelu, R., and 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, 1-17.

Wang, F., and Perry, S.E. (2013). Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. Plant Physiol 161, 1251-1264.

Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M. (2013). VAL- and AtBMI1-mediated H2Aub initiate the switch from embryonic to postgerminative growth in Arabidopsis. Current biology : CB 23, 1324-1329.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS biology 5, e129.

Zhou, Y., Tan, B., Luo, M., Li, Y., Liu, C., Chen, C., Yu, C.W., Yang, S., Dong, S., Ruan, J., Yuan, L., Zhang, Z., Zhao, L., Li, C., Chen, H., Cui, Y., Wu, K., and Huang, S. (2013). HISTONE DEACETYLASE19 interacts with HSL1 and participates in the repression of seed maturation genes in Arabidopsis seedlings. The Plant cell 25, 134-148.

Appendix A

Supplemental figures

A.1 Supplemental figures

The following figures and diagrams are included as supplemental figures of the manuscript in Chapter 2.

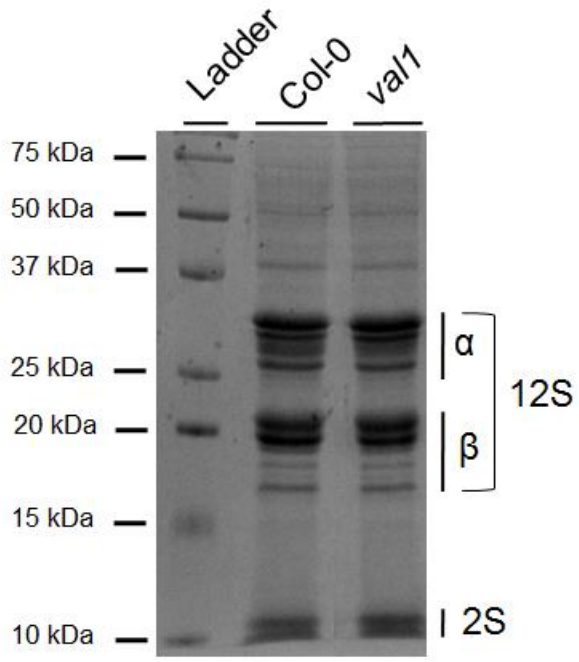

Figure A.1. SDS-PAGE analysis of seed storage protein composition in dry wild type and *val1* seeds. 2S albumins and subunits of 12S globulins are indicated.

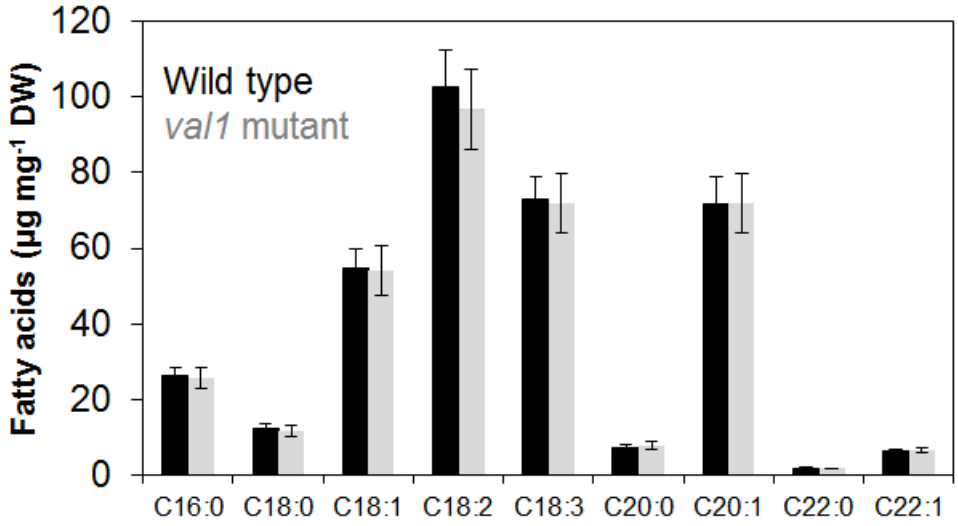

Figure A.2. Levels and composition of fatty acid methyl esters from wild type and *val1* dry seeds. Total levels and composition of fatty acids (C16 to C22) are unchanged between the wild type and *val1*. Total fatty acid levels are 356.35 ± 31.59 and 348.19 ± 37.55 mg mg^{-1} DW in wild type and *val1* embryos, respectively. Analysis was performed on 1.000 mg (± 5%) of dry seeds harvested from individual plants (n = 26 for wild type, n = 19 for *val1*).

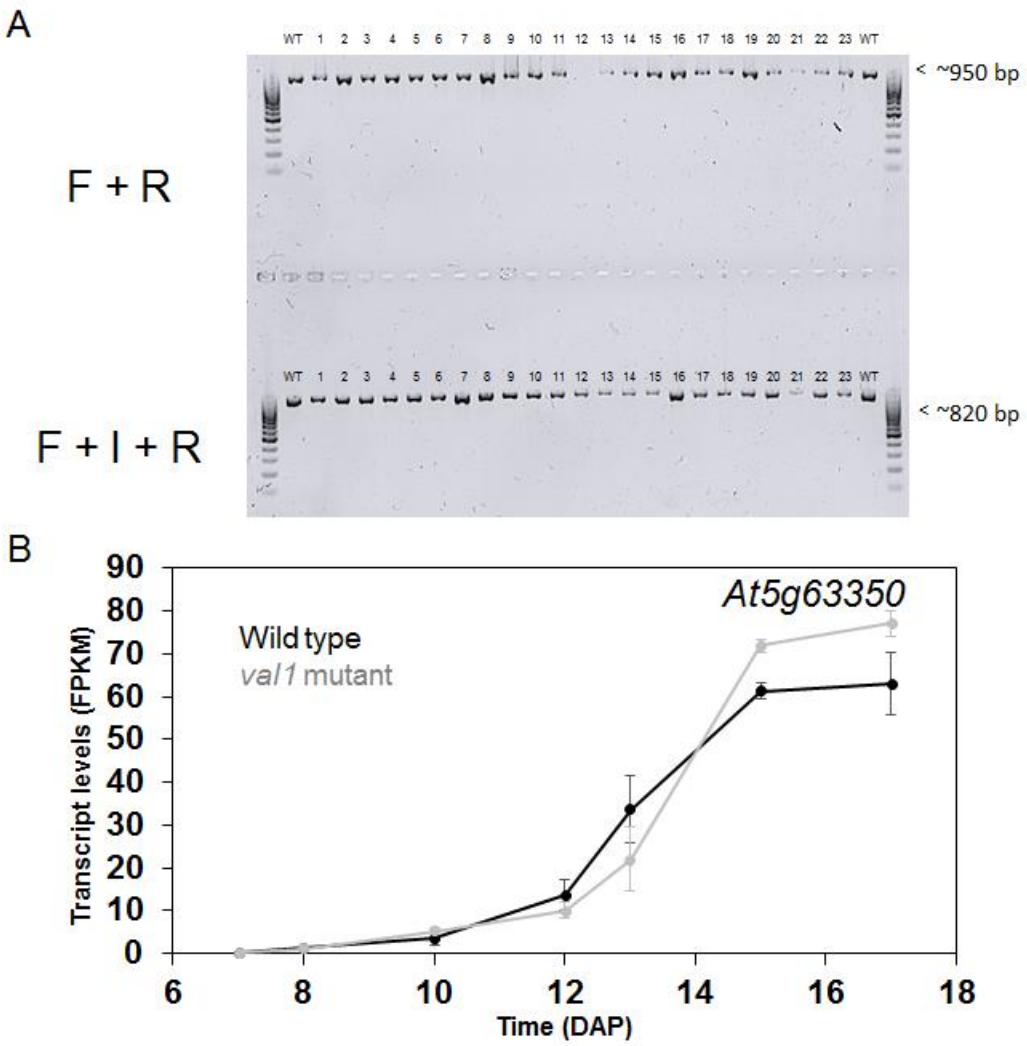

Figure A.3. Validation that SALK_088606C does not contain an additional T-DNA insertion in another locus. A) PCR amplification of the region (At5g63350 gene) potentially containing the insertion. F+R indicates PCR performed with primers annealing to regions flanking the insertion site. Lack of amplification would suggest that the insertion is present. F+I+R indicates that PCR was performed with the two flanking primers and a T-DNA-specific primer (I). In case the plants were homozygous for the T-DNA insertion, a band of the indicated size would appear, while a heterozygote would produce two bands. Presence of one band (~950 bp) that is the same size as that produced with the wild type FR primers indicates that all plants tested lack this type of insertion, which would produce a band of ~820 bp (expected location indicated). WT indicates wild type DNA sample, 1 – 23 are individual *val1* plants. Note that the F+R PCR reaction for sample 12 failed, but the F+I+R reaction produced one ~950 bp band. B) Expression of At5g63350 in *val1* is unaltered in its pattern compared to the wild type.

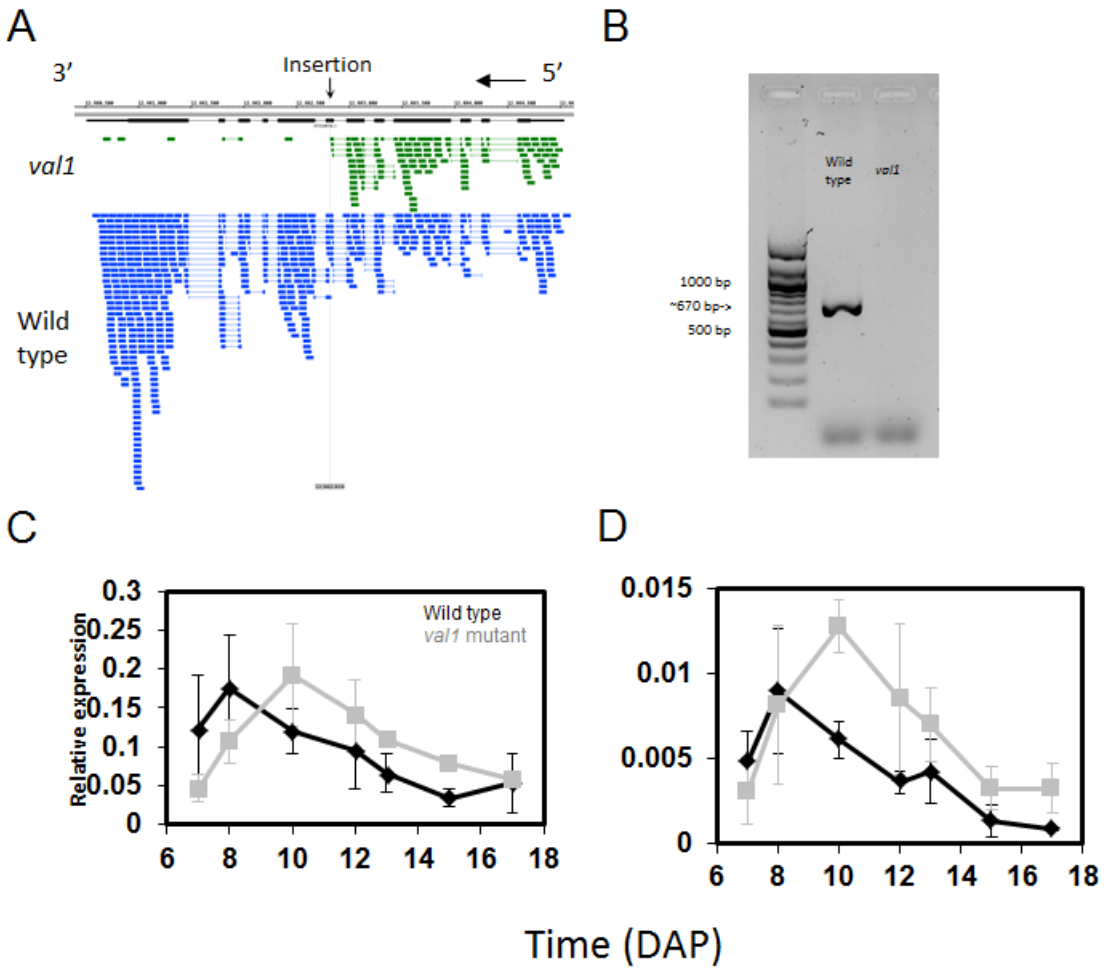

Figure A.4. Characterization of the VAL1 mutation. A. Mapping of reads to wild type (blue) and val1 (green) from a representative time point. Reads map to the region upstream of the insertion (indicated with the dashed line). Very few read map downstream of the insertion in *val1*. B. qPCR confirmation of full-length VAL1 expression in wild type and *val1*. Region targeted is the 5' end, upstream of the insertion. Wild type levels of expression in val1 confirms read-mapping and intact 5' end of the transcript is made in *val1* mutants. C. RT-PCR on the region containing the insertion suggests the insertion is retained in the VAL1 transcript in *val1*, as no band is detected. D. qPCR confirmation of the existence of the truncated splice variant. The pattern of expression closely follows that of the full-length splice variant in both wild type and *val1*.

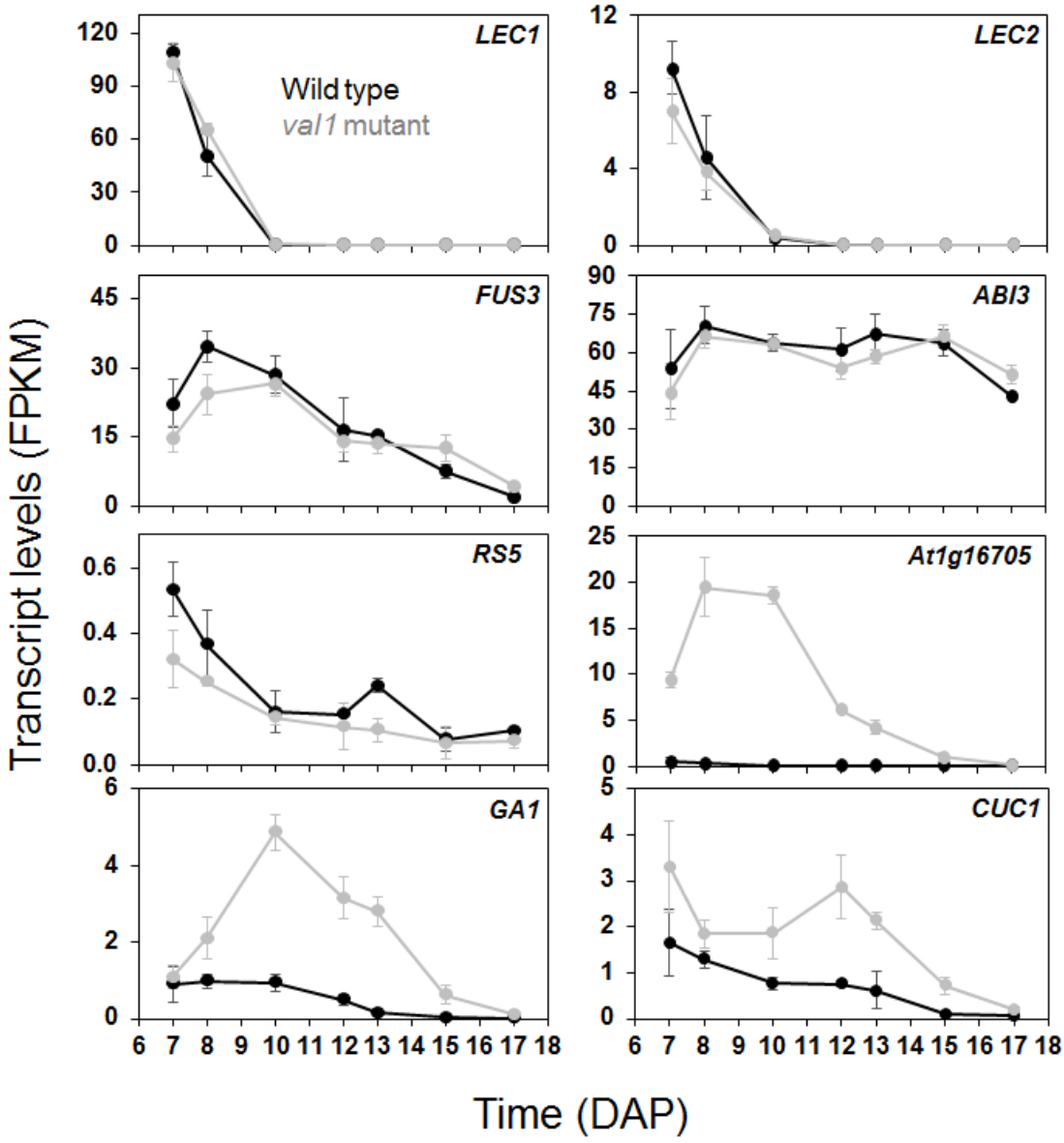

Figure A.5. Expression of core LAFL genes and selected transcripts in wild type and *val1* embryos. In cases where multiple splice variants were detected (e.g., *FUS3*), expression of only canonical, full length splice variants are shown.

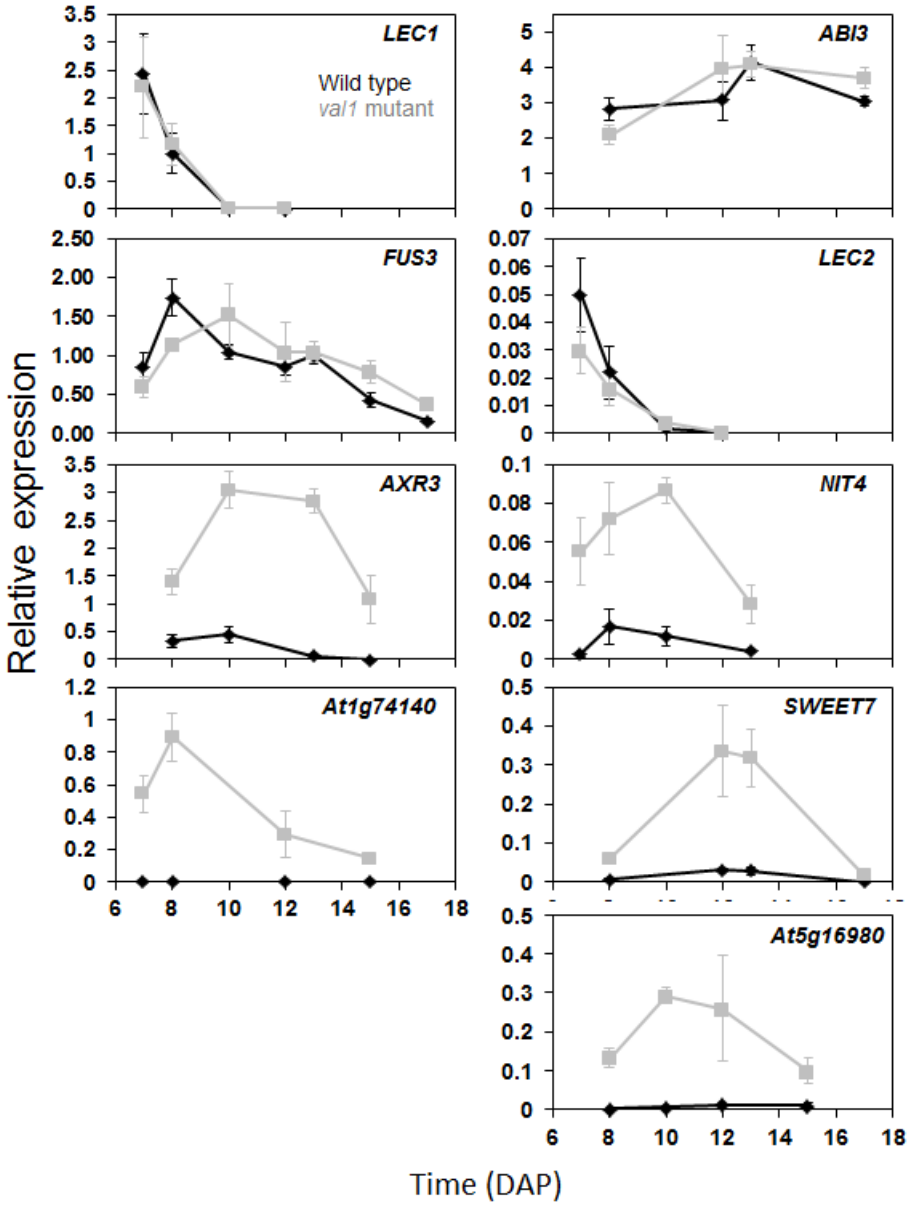

Figure A.6. Validation of RNAseq data through qPCR of selected genes.

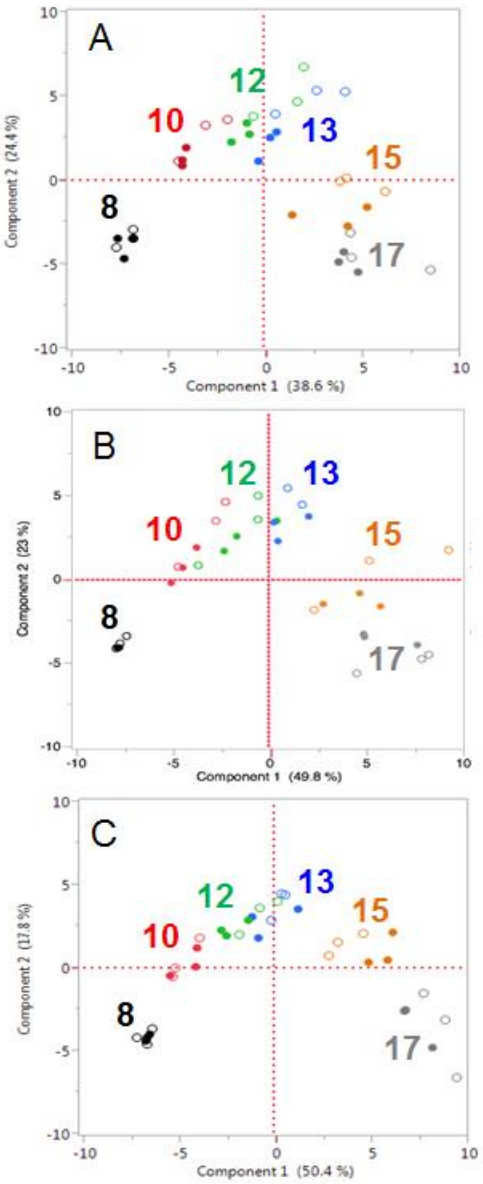

Figure A.7. Principal component analysis (PCA) on the correlations of metabolite, fatty acid, and hydrophobic protein levels in developing wild type and *val1* embryos. Metabolites from embryos (8 – 17 DAP) harvested from 3 experiments (A - C; 3 biological replicates each) were analyzed and the resulting differences among samples visualized by PCA score plots. Closed circles represent the wild type, while open circles represent *val1* mutant. The numbers represent days after pollination.

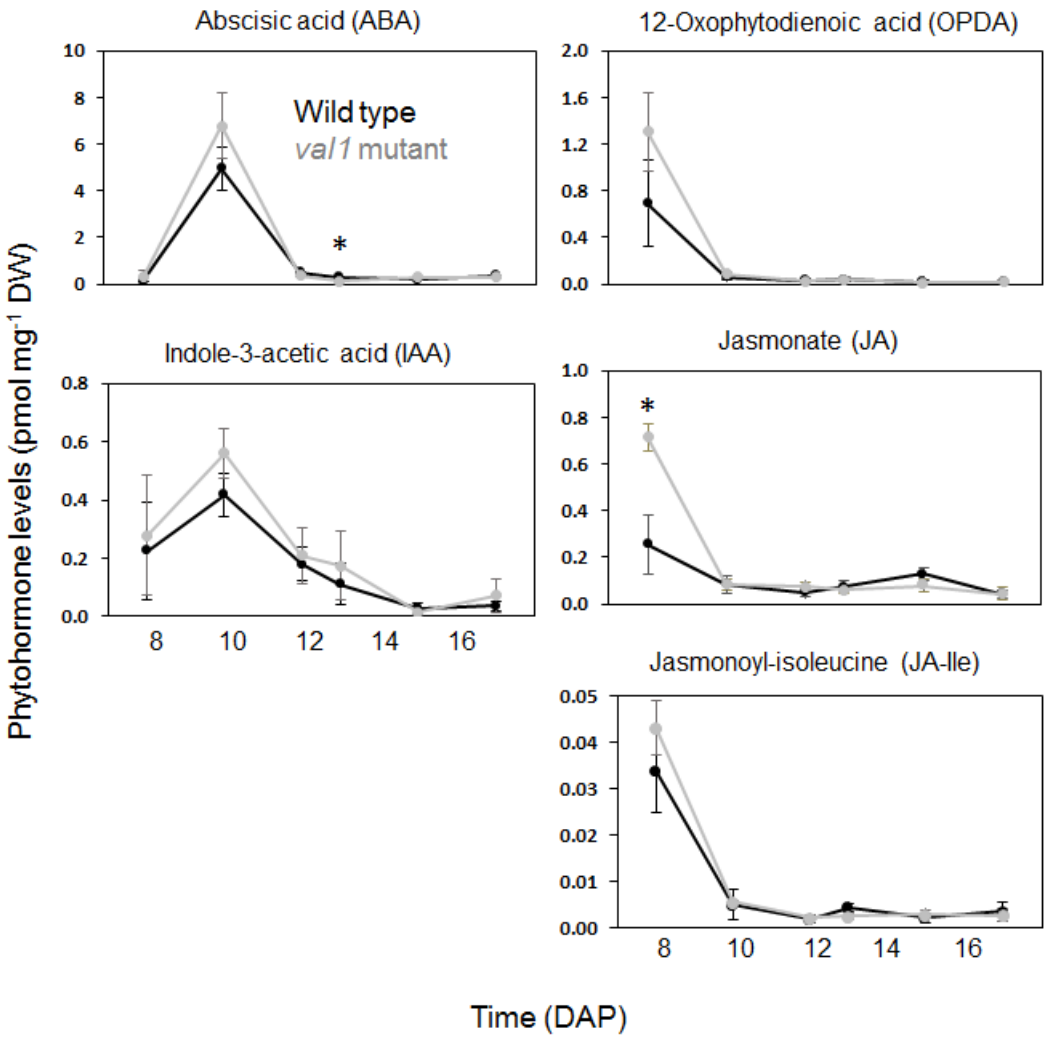

Figure A.8. Levels of phytohormones in developing wild type and *val1* mutant embryos.

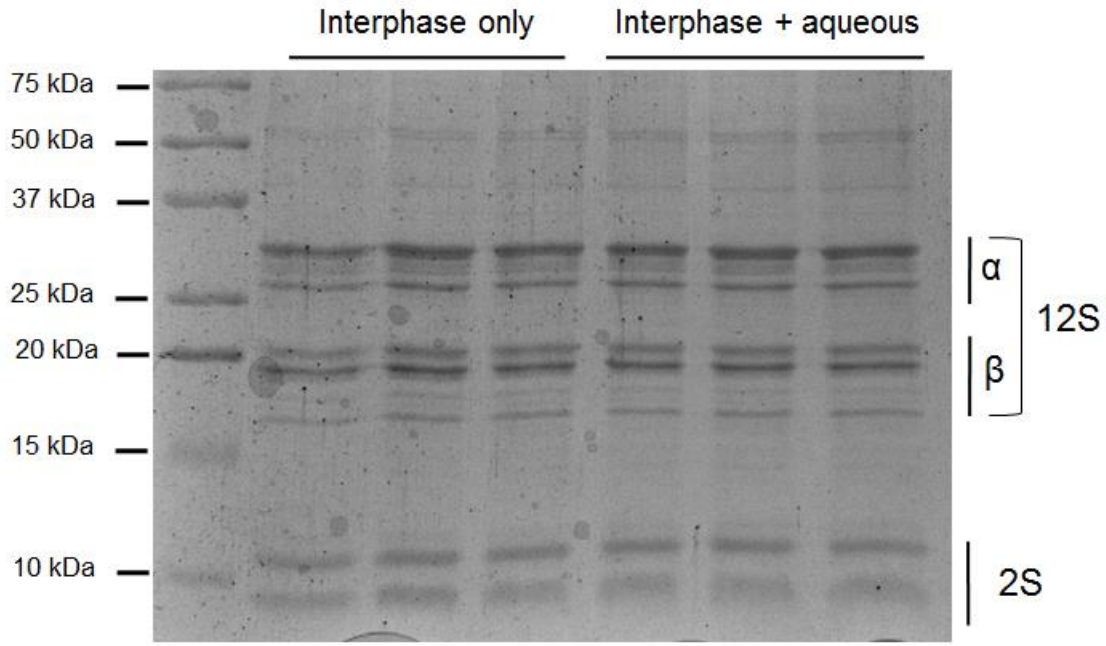

Figure A.9. SDS PAGE analysis of SSP location in the biphasic extraction.

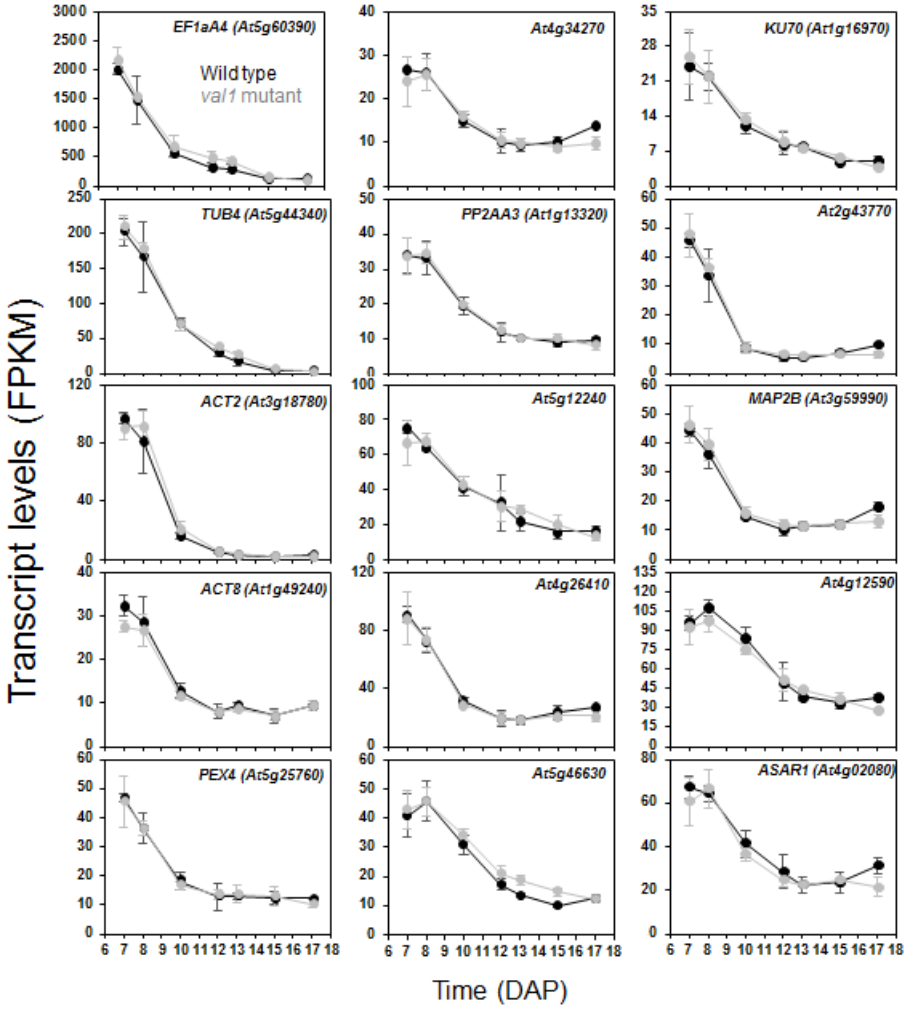

Figure A.10. Changes in expression of commonly used seed house-keeping genes. Commonly used housekeeping genes, including those specific to seeds, change their expression dramatically during embryo development. In addition, many of these genes have more than one splice variant (SV): At5g60390 (4 SVs), At5g44340 (1), At3g18780 (2), At1g49240 (1), At5g25760 (2), At4g34270 (1), At1g13320 (4), At5g12240 (4), At4g26410 (1), At5g46630 (2), At1g16970 (3), At2g43770 (1), At3g59990 (10), At4g12590 (1), At4g02080 (1).

Appendix B

Detailed experimental procedures

B.1 Experimental Procedures

In this section, procedures used in this study are described in enough detail such that all experiments presented can be repeated using these protocols without the necessity of consulting other publications.

B.1.1 Genotyping

DNA extraction

DNA was extracted from 2-3 week old leaf tissue in one of two ways. The first method utilized leaf tissue pressed onto Whatman FTA cards (GE Healthcare, UK) and was used as a fast extraction method. In brief, a leaf was placed onto an FTA card and pressed hard to smear tissue into the card. After drying, a small circle (~2 mm) was cut out, washed twice with 1% Triton solution and once with water. The second method involved extraction of DNA using ethanol precipitation. A small section of leaf was cut and placed in a 1.5 mL-tube containing three 3-mm glass beads. Extraction buffer (200 µL TE buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0)) was added and the leaf tissue was homogenized for 10 minutes on a paintshaker, followed by a quick centrifugation. Next, 100 µL DNA grinding buffer (200 mM Tris-HCl (pH 8.0), 250 mM NaCl, 25 mM EDTA, 0.5% SDS) and 100 µL 3 M potassium acetate (pH 5.2) was added, followed by vortexing. The mixture was centrifuged for 5 minutes at maximum speed and 200 µL supernatant was transferred into a new tube containing 200 µL isopropanol. DNA was precipitated at -20 °C for 30 minutes, and centrifuged at maximum speed for 30 minutes at 4 °C. The DNA pellet was washed with 80% ethanol and centrifuged at maximum speed for 5 minutes. Ethanol was removed and the DNA pellet was dried in a vacuum

concentrator (10 minutes at 45 °C). The pellet was resuspended in 20 µL TE buffer and used for PCR or stored at -20 °C.

PCR and DNA gel analysis

The Salk insertion lines used in this study (listed in Appendix D along with all primers used) must be genotyped to ensure each mutant is homozygous for the insertion. This is easily performed through a PCR strategy involving forward (F) and reverse (R) primers unique to the region flanking the insertion and one insertion specific (I) primer. Each PCR reaction contained 10 µl of GoTaq Green PCR master mix (Promega, Madison, WI), 1 mM of each primer (F, R, and I when applicable), either the small disc from the FTA card or 1 µL of DNA template, and was brought up to 20 µL with water. When 96-well plate format was used, 1 drop of mineral oil was added to the top of the reaction to prevent evaporation. The PCR program started with 5 minutes denaturing at 95 °C, followed by 35 cycles consisting of 30 seconds at 95 °C, 40 seconds at 58 °C, and 70 seconds at 72 °C. Another 5 minutes at 72 °C ensured final extension. The primers were designed to have the annealing temperature close to 58 °C to use the same PCR program, facilitating high-throughput screening. Amplicons were analyzed by electrophoresis on a 1% agarose gel, stained with 0.5 µL (for 30 mL gels, undiluted) or 1.5 µL (for 100 mL gels, undiluted) of Gel Red fluorescent dye (Biotium Inc, Hayward, CA), and visualized on a Bio-Rad ChemiDoc XRS+ imager with Image Lab software (Bio-Rad).

B.1.2 General plant growing conditions

The following is the general strategy developed for growing Arabidopsis plants for the purposes of acquiring high quality seeds. Seeds were stored at room temperature in boxes and to

facilitate germination, they were tapped into 1.5 mL-tubes and cold stratified (4 ^{o}C) for 3 – 5 days in 0.1% agarose. Seeds were sown onto moist Farfard superfine germinating mix (Sun Gro Horticulture, MA). To mix soil, approximately 2 scoops of soil with 1 – 1.5 L of water was used to make one 32 pot flat. Seeds (5 – 6) were sown into the middle of each pot. Flats were placed in a growth chamber set to 20 ^{o}C day/18 ^{o}C night, 65% humidity and 16-hour photoperiod. A clear cover was kept over the flat for one week and removed. The first application of water contained insecticide (Bayer 3-in-1 Insect, Disease and Mite Control, approximately 5 mL per 2 gallons) and was typically added about one week after placing the flat in the growth chamber. The plants required watering at different intervals: between the time they were sown to approximately 3 – 4 weeks, the plants received deionized water every 3 – 4 days. After the plants were 3 – 4 weeks old, they received Miracle Gro All-Purpose Plant Food (approximately 1 tsp per 2 gallons) every 2 – 3 days until they were approximately 5 – 6 weeks old, when they received the Miracle Gro mix every 1 – 2 days. Plants received this mixture until they senesced. Plants grown for storage compound screening were grown in duplicate, with the same arrangement of mutants and with a wild type plant in each flat as a control. For seed storage compound confirmations, plants were all grown in the same flat and randomized within it. For flux profiling and metabolic flux analysis, plants were grown in the same flat without randomizing. For time course studies, separate flats were used.

B.1.3 Analyses of SSCs in dry seeds and embryos

Harvesting dry seeds for SSC analyses

Seeds used for all analyses of SSCs were collected from the main stem of senescing plants. The first 1 – 2 siliques were not collected as they often grow abnormally and likely have

alterations in composition. For all analyses (screening and other experiments), a newly opened flower was marked at 0 days after pollination (DAP) and after 21 days, the siliques were dry. The four siliques above and up to four siliques below this marked silique were collected in a 1.5 mL-tube for a total of nine siliques. Seeds were liberated by application of pressure with a metal probe and silique detritus was removed. A small hole was punctured in the tube lid to prevent moisture build up and subsequent fungus growth.

Biphasic extraction of oils and proteins for biomass quantification

The following protocol was used for quantification of oils and proteins from both dry seeds and cultured embryos. For dry seed studies, approximately 1.000 mg (± 5%) of dry seed was measured on an analytical microbalance (XP-26, Mettler Toledo) and added to a 1.5-mL tube. For cultured embryos, 0.5 mL-tubes were used and no target mass (e.g. whatever mass was collected was used). To the tube, 100 µL of heptane containing 2 – 5 µg heptadecanoic acid (C17:0) internal standard was added along with two 3-mm borosilicate glass beads (four 2-mm borosilicate glass beads for 0.5 mL-tubes). Seeds/embryos were homogenized on a paintshaker for 10 minutes. To bring ground material down from the top of the tube, the tubes were centrifuged for 10 seconds. To separate polar metabolites and oils, 100 µL of water (99 µL water and 1 µL norvaline (3.35 µM) for embryos cultured for biomass) and 100 µL of diethyl ether were added, respectively. Tubes were vortexed for 15 seconds and centrifuged for 5 minutes at 16,000 x g, producing two phases. The top organic phase contained fatty acids and the bottom aqueous phase contained water-soluble proteins. The storage proteins partitioned into the insoluble interphase along with cell walls and other polysaccharides. The organic phase was transferred to a 2 mL-glass reaction vial (black capped) and the walls of the 1.5 mL-tube were

washed with 200 – 400 µL diethyl ether. The washing was added to the 2 mL-glass vial and the remainder of the diethyl ether in the 1.5 mL-tube evaporated, leaving behind the aqueous phase and interphase.

Generation and analysis of fatty acid methyl esters (FAME)

The organic phase was dried under a stream of N_2 gas at 55 ^{o}C for 15 – 20 minutes, and 500 µL of 1 N methanolic-HCl was added. The vials were re-capped and vortexed for 15 seconds. The derivatization reaction was carried out for 2 hours at 75 ^{o}C, followed by drying under N_2 as before. The FAMEs were dissolved in 30 – 50 µL heptane and the FAME solution was separated and analyzed on an Agilent 7890A series gas chromatographer coupled with flame ionization detector (GC-FID, Agilent Technologies, Santa Clara, CA) equipped with a 30-m DB-23 column (0.25 mm x 0.25 µm, Agilent Technologies). By comparing the area of the abundance peak to that of the added C17:0 internal standard, the amount of fatty acid present in the seed sample was determined. Integration of the peak area was performed with MSD ChemStation software (Agilent).

Quantification and analysis of SSPs from dry seeds and embryos

The following protocol was used to estimate the protein content of dry seeds and embryos harvested for metabolomics. For dry seeds, the interphase and aqueous phase were used. For time course embryos, only the interphase was used as the aqueous phase was used for metabolomics. Depending on the experiment, the phase(s) was dried on a Savant SPD131DDA speedvac concentrator (Thermo Scientific) at 45 ^{o}C for approximately 45 minutes, followed by addition of 100 µL of protein extraction buffer (100 mM Tris-HCl (pH 8.0), 0.5% SDS, 2% 2-

mercaptoethanol) and vortexing for 15 seconds. The samples were boiled for 5 minutes, vortexed again for 15 seconds and boiled for another 5 minutes, followed by centrifugation for 20 minutes (the supernatant contained solubilized proteins). Protein content was estimated using a "homemade" version of the MarkerGene Hydrophobic Protein Analysis Kit (Marker Gene Technologies, Inc., Eugene, OR). A 20-µL aliquot of the supernatant was added to 80 µL water in a white bottom 96-well plate, producing a 100 µL total volume. 100 µL of 1 mM 6-p-toluidino-2-naphthalenesulfonic acid was added to each well and the plate was covered and incubated for 1.5 – 2 hours at room temperature. Fluorescence was measured using a plate reader (Synergy 4, Biotek), with excitation set to 460 nm, and emission set to 365 nm. A standard curve using 0, 2, 5, 10, 20, and 40 µg of bovine serum albumin (BSA) was constructed for comparison. Storage protein composition was qualitatively determined using SDS PAGE. A small aliquot (12 µL) of the solubilized protein was mixed with 4 µL SDS PAGE loading buffer. Following a quick centrifugation (10 seconds at 16,000 x g), the entire mixture was loaded onto a gel (15% resolving, 6% stacking gels) and stained with Coomassie brilliant blue.

Hydrolysis of total proteins from cultured embryos and analysis of labeling in amino acids through generation of *tert*-butyl dimethyl-silyl derivatives

Proteins present in the aqueous and insoluble interphase were hydrolyzed in one of two ways. For flux profiling, 100 µL of concentrated hydrochloric acid (12 – 13 M) was added to each tube and the tubes were incubated in an oven set to 105 – 110 ^{o}C for 16 hours. For MFA, tubes were placed in a custom-made hydrolysis chamber and 4 mL concentrated hydrochloric acid was added directly to the chamber. The chamber was sealed, the air was removed with a vacuum pump and the chamber was incubated in an oven set to 115 ^{o}C for 16 hours. After

cooling, the acid needed to be neutralized. For labeling studies, samples were first dried under nitrogen gas and then 50 µL of 1 M ammonium hydroxide was added, followed by vortexing. Samples were then dried under nitrogen for derivatization as described below. Embryos cultured for biomass were dried under nitrogen gas and neutralized with 200 µL of borate buffer (Waters) as ammonium hydroxide interferes with the subsequent derivatization, and the samples were vortexed for 10 seconds. The samples were ready for derivatization or storage.

Quantification of amino acids in cultured embryos and embryos collected for metabolomics through ultra-performance liquid chromatography (UPLC)

For both embryos cultured for biomass and embryos harvested for metabolomics, the 5 µL of the aqueous sample volume was added to 15 µL of borate buffer, to which 5 µL AccQ-Tag derivatization reagent (Waters) was added, followed immediately by vortexing (if in a glass vial) or pipette mixing (if in a 96-well plate). The samples were separated and detected on a Waters Acquity UPLC equipped with a Waters AccQ-Tag Ultra C18 1.7 µm (2.1 x 100 mm) column coupled to an fluorescence detector (Waters). Amino acid quantity was determined by comparison to a standard curve of amino acids.

B.1.4 Protocols specific to flux profiling and metabolic flux analysis experiments

Embryo culturing conditions

Embryos for culturing need to be approximately 7 DAP, when they appear as “walking sticks”. Siliques of this approximate age were cut and surface sterilized with bleach for 5 minutes and washed several times to remove the bleach. Seeds were removed on a sterile plate with fine forceps and a metal probe (both sterilized in 70% ethanol for 5 minutes). Embryos were

dissected on a sterile 35-mm glass fiber Whatman filter pad (Grade 934-AH, GE Healthcare) with fine forceps and a sterile needle (0.3-mm). For flux profiling studies, 10 – 20 embryos were used per replicate. For MFA uptake studies (both biomass quantification and CO_2 release) 30 – 35 embryos per replicate were used to accurately measure dry weight. For MFA labeling studies, 20 embryos were cultured per replicate. The embryos from each replicate were placed into one well of a six well tissue culture plate (Costar 3516 with 9.5 cm^2 well size, Corning Life Sciences) for all studies except CO_2 release experiments, which used 14 mL-glass vials sealed with rubber septa and crimped shut. Each well (or glass vial) contained two 30-mm Whatman filter pads (Grade 934-AH, GE Healthcare) and 1 mL of culture media (24 mM KCl, 15 mM $MgSO_4$, 1.5 mM K_2HPO_4/KH_2PO_4, 30 mM $CaCl_2$, 40 µM Na_2EDTA, 40 µM $FeSO_4$, 2 µM H_3BO_3, 199 µM $MnSO_4$-H_2O, 73 µM $ZnSO_4$-H_2O, 10 µM KI, 2 µM Na_2MoO_4-$2H_2O$, 200 nM $CuSO_4$-$5H_2O$, 200 nM $CoCl_2$-$6H_2O$, 6 mM glucose, 3 mM sucrose, 5 mM alanine, 5 mM glutamine, 9.8 mM KNO_3, 5.2 mM NH_4NO_3, 2 µM abscisic acid, and Gamborg's vitamins (5.5µM inositol, 5 mg/L nicotinate, 500 µg/L pyridoxine HCl, 500 µg/L thiamine HCl, 500 µg/L folate, 50 µg/L biotin)). Depending on the labeling strategy, one of the carbon sources was replaced with labeled carbon: uniformly labeled [U-$^{13}C_6$] glucose, [1,2-$^{13}C_2$]-glucose, [$^{13}C_3$]-alanine, or [$^{13}C_5$]-glutamine. For nitrogen labeling experiments, [alpha-^{15}N]-glutamine, [amide-^{15}N]-glutamine and [^{15}N]-alanine replaced their non-labeled counterparts. The culture plates have two air spaces between the wells and 1 mL sterile water was added to each to maintain high internal humidity. Once this setup was complete, the plate was sealed with surgical tape and placed in a growth chamber (set to 15 – 18 µE of green filtered continuous light, 20 ^oC, 60% humidity) for seven days. Following culturing, embryos were collected in 0.5-mL tubes with 400 µL of high-performance liquid

chromatography (UPLC)-grade water and rinsed two times. Tubes were centrifuged briefly, snap-frozen in liquid nitrogen and stored at -80 ^{o}C until processed.

Nutrient uptake measurements

Nutrient uptake was calculated by comparing the amount of glucose, sucrose, glutamine and alanine remaining in the culture wells to the initial mixture. After 1 week in the growth chamber, the culture plate was unwrapped and 500 µL of a mixture of ribitol/norvaline (5 mM each) was added to the culture media. The entire content of the well was collected in a 50 mL-tube. The filter pads supporting the embryos were placed in a vacuum filter apparatus (Autofil) attached to the same 50 mL tube. The filter apparatus was attached to the vacuum line and the liquid on the filter pads was pulled across the filter membrane. The wells of the culture plate were rinsed three times with 5 mL UPLC-grade water and the rinse was carefully applied to the filter pads. The filter pads were then rinsed three times with 5 mL UPLC-grade water directly. The filter pads were removed and the filter membrane was rinsed once with 5 mL UPLC-grade water. 1 mL of the initial media was collected in a new tube, and 500 µL of the ribitol/norvaline mixture was added. The 50 mL-tubes containing the media (both the initial and the used media) were capped, frozen in liquid nitrogen and lyophilized. The embryos were collected, rinsed and frozen in liquid nitrogen until processed.

After lyophilization, the dry powder was resuspended in 1 mL UPLC-grade water. For quantification of amino acids, 10 µL was added to 990 µL of UPLC-grade water, and 5 µL was used in a 25 µL AccQ Tag reaction (5 µL sample, 15 µL borate buffer, 5 µL AccQ Tag reagent). Separation and analysis was performed on a Waters UPLC equipped with a Waters AccQ-Tag Ultra C18 1.7 µm (2.1 x 100 mm) column coupled to an Acquity fluorescence (FLR) detector

(Waters). Quantification was performed by comparison of the area of each amino acid to the added norvaline internal standard. For sugar analysis, 15 µL of the initial resuspended solution was added to 15 µL of acetonitrile. Separation and analysis was performed on a Waters UPLC equipped with an Acquity UPLC BEH Amide 1.7 µm (2.1 x 100-mm) column attached to an Acquity evaporative light scattering (ELSD) detector (Waters). Quantification was performed by comparison of the area of each sugar to the ribitol internal standard.

CO_2 release and O_2 usage measurements

The levels of CO_2 and O_2 were measured by GCMS. Developing embryos contain CO_2 bound within them, which was released by the addition of 150 µL of a 1 M sulfuric acid solution through the rubber septum with a syringe. After 15 – 20 minutes, the embryos turned yellow, indicating the acid had penetrated the tissues. Levels of CO_2 and O_2 were quantified by injecting 100 µL of the air from the vial onto the GCMS equipped with a DB5-MS column (Agilent). Embryo gas release and usage was quantified by comparison to a blank using naturally occurring argon gas as an internal standard.

Analysis of labeling in fatty acids through generation of butyl amides

Frozen cultured embryos were thawed and four 2-mm borosilicate glass beads and 100 µL of heptane were added to each sample tube. Samples were homogenized by placing the tubes on a paintshaker for 5 – 10 minutes, followed by a quick centrifugation and addition of 100 µL each water and diethyl ether. The tubes were vortexed for 10 seconds and centrifuged for 5 minutes at the maximum speed, resulting in aqueous and hydrophobic phases. The upper hydrophobic phase was removed and transferred to a 2 mL-glass vial and dried under nitrogen

gas, followed by derivitization with 200 µL of *n*-butylamine, which generates fatty acid butylamides (FABA). Vials were placed in an oven overnight at 75 ^{o}C, followed by drying under nitrogen. FABA were resuspended in 50 µL of heptane. Derivatized fatty acids were analyzed on an Agilent 7890A gas chromatography mass spectrometer (GC-MS, Agilent Technologies) equipped with a DB-5MS-DG column (30 m length × 0.25 mm × 0.25 µm with a 10-m pre-column, Agilent Technologies) and analyzed on an Agilent 5975C series single quadrupole mass spectrometer (MS). Analysis of the mass spectra of the major fragments yields information regarding the positional labeling. For example, the major fragments for fatty acids have a mass-to-charge ratio (m/z) of 115 and 128, which correspond to unlabeled two carbon and three carbon fragments, respectively. A fragment with m/z of 116 means that one of the two carbons is labeled (present as ^{13}C), and both are labeled in m/z of 117. Analysis of labeling was determined by calculating the percentage of the labeled fragments compared to the unlabeled fragments out of the total fragment amount. For flux profiling and MFA, the samples were derivatized with 30 µL of N-methyl-N-(*tert*-butyldimethylsilyl)-trifluoroacetamide) and 1% (v/v) *tert*-butyldimethylchlorosilane diluted 1:1 with pyridine. The derivatization reaction was carried out at 50 ^{o}C for 30 minutes. Amino acid labeling was analyzed on the GC-MS as performed for FABA.

B.1.5 Experiments specific to time course studies

Time course: set up and embryo harvesting

All three time course experiments were set up the same way. To ensure similar nutrient allocation and light exposure to the developing seeds, the time course was set up in stages: stage 1 (7 and 8 DAP), stage 2 (10, 12, 13 DAP) and stage 3 (15 and 17 DAP). Flowers were tagged

starting with the oldest time point first (e.g., on a stage 2 plant, 13 DAP was marked first, followed by 12 DAP, etc.). After the requisite time passed, tagged siliques were cut from the plant individually and transported on ice to a dissecting microscope. The seeds were removed and embryos were dissected with forceps and a fine needle on filter pads on large Petri dishes (145 x 20-mm) filled with frozen water. Dissections were performed between -5 and 0 ^{o}C and temperature was frequently measured with an infrared thermometer (Ryobi Tools). For transcriptomics, the seed coats were removed from the filter pad and the embryos were rinsed from the filter pad into 50 mL-tubes containing ice-cold water. The tubes were then centrifuged in a cold Allegra X-15R tabletop centrifuge (Beckman Coulter) and the majority of the water was drawn off. The embryos were poured into a 1.5 mL-tube and washed with 1 mL ice-cold water. The embryos were centrifuged for 3 minutes in a cold (Thermo Scientific). The water was removed and the embryos were snap-frozen in liquid nitrogen and stored at -80 ^{o}C until processed. For metabolomics and phytohormone analyses, embryos dissected on filter pads were transferred with a metal probe into a 1.5 mL-tube containing ice cold LCMS grade water. The tubes were centrifuged for 2 minutes in a cold Sorvall Legend Micro 21R centrifuge and rinsed with ice-cold LCMS grade water. Embryos were snap frozen in liquid nitrogen, lyophilized overnight and stored in a desiccator until processed.

RNA extraction and quantification

RNA was extracted from frozen embryos in the following way. Two 3-mm borosilicate glass beads were added to each tube and the frozen embryos were homogenized by placing the tubes in a cold box maintained at 0 ^{o}C onto a paintshaker. After a quick centrifugation, 550 µL of RNA extraction buffer (0.4 M LiCl, 0.2 M Tris (pH 8.0), 25 mM EDTA, 1% SDS) and 550 µL

of chloroform were added, followed by vortexing for 10 seconds and centrifugation for 3 minutes at maximum speed in a cold centrifuge. The supernatant was transferred to a 2.0 mL-tube and 500 µL saturated phenol was added and pipette-mixed, followed by 200 µL of chloroform. The tubes were inverted 10 times and placed on ice for 15 minutes. The tubes were centrifuged for 3 minutes at maximum speed in a cold centrifuge and the supernatant was transferred to a new 2.0-µL tube. The transferred volume would typically be around 450 µL, and 1/3 volume of 8M LiCl was added and pipette mixed. The samples were incubated for at least 1 hour at -20 ^{o}C to precipitate RNA and centrifuged for 30 minutes at maximum speed in a cold centrifuge. The supernatant was discarded and the pellet was resuspended in 500 µL DEPC-treated water. 7 µL of 3 M potassium acetate and 250 µL of 100% ethanol were added and pipette mixed, followed by centrifugation for 10 minutes at maximum speed in a cold centrifuge. The supernatant was transferred to a new 2.0 mL-tube and 43 µL 3 M potassium acetate and 750 µL 100% ethanol were added, followed by pipette mixing. The samples were incubated for 1 hour at -20 ^{o}C and centrifuged for 20 minutes at maximum speed in a cold centrifuge. The resulting pellet is difficult to see and the majority of the supernatant was removed, leaving behind ~50 µL. This was used in an on-column clean up and DNase digestion using the RNeasy kit and following the manufacturer's instructions. The resulting RNA was eluted in 32 µL of water and quantified using a Nanodrop. RNA integrity analysis was performed by the Virginia Bioinformatics Institute (Blacksburg, Virginia).

Biphasic extraction and metabolite analysis for time course embryos

Metabolomics on harvested embryos was performed with an alternative biphasic extraction method. To each sample, 99 µL of chloroform and 1 µL of 2 – 5 µg/µL C17:0

internal standard was added along with four 2-mm borosilicate glass beads. The embryos were ground on a paintshaker for 5 – 10 minutes followed by a quick centrifugation. To each sample, 98 µL LCMS-grade water was added, along with 1 µL of 1 mM norvaline and 1 µL of a 5 mM citrate and glucose mix (both uniformly labeled). Samples were vortexed for 10 seconds and centrifuged for 5 minutes at maximum speed, producing two layers. The bottom layer was hydrophobic and contained the fatty acids; the top layer was aqueous and contained polar metabolites, soluble proteins, etc. The interphase contained cell walls and storage proteins. The bottom layer was carefully transferred (to avoid removing the interphase) to a 2 mL-glass vial and the fatty acids were analyzed as described above. The interphase was dried in a speedvac concentrator and the proteins were analyzed as described above. The aqueous layer contained free amino acids, and 5 µL was used for quantification of free amino acids as described above. The remainder of the aqueous layer was transferred to a 250 µL-glass GCMS vial, dried in a speedvac concentrator set to 45 ^{o}C for 45 minutes, followed by further drying under a nitrogen stream. The samples need to be very dry, as the derivatization reagents are sensitive to moisture. To each sample, 15 µL methoxyamine-HCl (MOX) was added, vortexed for 40 seconds, and incubated for 2 hours at 50 ^{o}C. 15 µL of N-methyl-N-trimethylsilyl-trifluoroacetamide and 1% trimethylchorosilane was added to each sample, followed by vortexing and incubation for 30 minutes at 50 ^{o}C. Derivatized metabolites (primarily sugars and organic acids) were analyzed on an Agilent 7890A GC-MS (Agilent Technologies) equipped with a DB-5MS-DG column (30 m length × 0.25 mm × 0.25 µm with a 10-m pre-column, Agilent Technologies) and analyzed on an Agilent 5975C series single quadrupole mass spectrometer (MS). Relative levels of metabolites were determined by comparison to the levels of citrate and glucose internal standards.

Appendix C

Results from the high-throughput screen

Table C.1 Results of the high-throughput screen, expression of genes in embryos and observations made during the course of the investigation.

The initial stage of this work involved a phenotypic screen of dry seeds and cultured embryos. The results of that screen, the expression of the TFs in the embryo, and any observations made are listed in the following table for future reference. Briefly, Tier 1 refers between 5 – 10% difference in the indicated storage compound, Tier 2 is between 10 – 15%, and Tier 3 is greater than 15%. WT indicates there was no difference between the mutant and the wild type control.

SALK Line	AGI	Gene Name	Oil Tier	Protein Tier	Flux	Expression in embryos	Notes
SALK_000196C	At4g00730	ANTHOCYANINLESS 2	1	WT	WT	Yes/no difference	
SALK_000247	At2g24500	REI1-LIKE 2	1	WT	WT	Yes/no difference	
SALK_000287C	At1g28470	NAC DOMAIN PROTEIN 10	WT	WT	WT	No	
SALK_001014C	At1g43700	SUE 3	2	WT	Possible	Yes/no difference	
SALK_001597C	At1g34180	NAC DOMAIN PROTEIN 16	2	WT	Yes	No	
SALK_001658	At4g24150	GROWTH-REGULATING FACTOR 8	3	WT	WT	Yes/ difference	
SALK_002516C	At1g50420	SCARECROW-LIKE 3	n/a	n/a	n/a	Yes/no difference	
SALK_003155	At3g26744	INDUCER OF CBF EXPRESSION 1	n/a	n/a	n/a	No	
SALK_003943C	At2g47260	WRKY DNA-BINDING PROTEIN 23	n/a	n/a	n/a	Yes/ difference	
SALK_004037C	At5g57180	CHLOROPLAST IMPORT APPARATUS 2	3	1	WT	Yes/ no difference	
SALK_005260	At5g35550	TRANSPARENT TESTA 2	3	1	WT	No	
SALK_006229C	At2g03710	SEPELLATA 4	1	WT	WT	No	
SALK_006473	At5g57520	ZINC FINGER PROTEIN 2	WT	WT	WT	No	
SALK_006751	At5g07680	NAC DOMAIN PROTEIN 80	WT	2	WT	No	
SALK_008040C	At1g05380	Unknown	3	n/a	WT	Yes/ no difference	
SALK_008559C	At3g54320	WRINKLED 1	3	3	WT	Yes/ no difference	
SALK_008606C	At5g67300	MYB DOMAIN PROTEIN 44	3	1	WT	Yes/ no difference	
SALK_010653C	At4g36900	RELATED TO AP2 10	3	3	WT	Yes/ no difference	
SALK_011023C	At4g34610	BEL1-LIKE HOMEODOMAIN 6	WT	1	WT	Yes/ no difference	
SALK_011042C	At1g30810	JUMONJI 18	WT	WT	WT	Yes/ no difference	
SALK_011069C	At3g15500	NAC DOMAIN PROTEIN 3	1	1	WT	No	
SALK_011502C	At5g18270	NAC DOMAIN PROTEIN 87	2	WT	WT	No	
SALK_011957C	At1g19180	JAZ1	3	3	Possible	Yes/ no difference	
SALK_012997C	At1g62300	WRKY6	3	1	WT	Yes/ difference	
SALK_013163C	At2g36270	ABA INSENSITIVE 5	1	WT	WT	Yes/ no difference	
SALK_000196C	At4g00730	ANTHOCYANINLESS 2	1	WT	WT	Yes/ no difference	

SALK Line	AGI	Gene Name	Oil Tier	Protein Tier	Flux	Expression in embryos	Notes
SALK_014315	At3g10590	Unknown	WT	1	WT	No	
SALK_014999	At4g18960	AGAMOUS	n/a	n/a	n/a	No	Sterile/no seeds produced
SALK_015495C	At1g32770	NAC DOMAIN PROTEIN 12	WT	1	WT	No	
SALK_016142C	At3g26790	FUSCA3	n/a	n/a	n/a	Yes/ no difference	
SALK_016183C	At1g76590	Unknown	3	n/a	WT	Yes/ no difference	
SALK_016333C	At3g09600	REVEILLE 8	n/a	n/a	n/a	Yes/ no difference	
SALK_016554	At1g19000	Unknown	WT	1	WT	Yes/ no difference	
SALK_017963C	At4g24660	ZINC FINGER HOMEODOMAIN 2	WT	2	Possible	Yes/ no difference	
SALK_018618C	At2g35550	BASIC PENTACYSTEINE 7	1	1	WT	Yes/ no difference	
SALK_020767C	At1g78080	WOUND INDUCED 1	3	n/a	WT	Yes/ difference	
SALK_020840C	At5g46880	Unknown	3	WT	WT	Yes/ no difference	
SALK_021965C	At2g41070	bZIP12	3	3	WT	Yes/ no difference	
SALK_022174C	At1g34190	NAC DOMAIN PROTEIN 17	WT	WT	WT	Yes/ no difference	
SALK_022235C	At5g43270	SQUAMOSA PROTEIN-LIKE 2	3	3	WT	Yes/ no difference	
SALK_022648	At3g23030	HOMEOBOX-7	n/a	n/a	n/a	Yes/ difference	
SALK_022661C	At2g21240	BASIC PENTACYSTEINE 4	WT	n/a	WT	Yes/ no difference	
SALK_023411C	At3g24650	ABA INSENSITIVE 3	3	3	n/a	Yes/ no difference	
SALK_024494C	At1g22590	Unknown	2	WT	WT	Yes/ no difference	
SALK_024760C	At5g01310	AGAMOUS-LIKE 87	WT	WT	WT	Yes/ no difference	
SALK_024800C	At2g40140	SALT-INDUCIBLE ZINC FINGER 2	WT	WT	WT	Yes/ no difference	
SALK_025198C	At3g56400	WRKY70	1	1	WT	No	
SALK_025279C	At1g74950	JASMONATE-ZIM-DOMAIN PROTEIN 2	WT	1	WT	Yes/ no difference	
SALK_025754C	At5g17300	REVEILLE 1	2	n/a	WT	Yes/ no difference	
SALK_025839C	At3g59580	Unknown	n/a	n/a	n/a	Yes/ difference	
SALK_028204	At5g65210	NIN-LIKE PROTEIN 9	n/a	n/a	n/a	Yes/ slight difference	
SALK_030459C	At3g16770	ETHYLENE RESPONSE FACTOR 72	3	n/a	WT	Yes/ no difference	

SALK Line	Gene	Gene Name	Oil Tier	Protein Tier	Flux	Expression in embryo	Notes
SALK_030799C	At3g53370	Unknown	1	WT	WT	Yes/ difference	
SALK_030876C	At1g43640	Unknown	1	WT	WT	Yes/ difference	
SALK_030989C	At5g12840	NUCLEAR FACTOR Y SUBUNIT A1	1	n/a	WT	Yes/ no difference	
SALK_031064C	At2g43500	NIN-LIKE PROTEIN 8	1	WT	WT	Yes/ no difference	
SALK_032192	At3g54220	SCARECROW	n/a	n/a	n/a	Yes/ no difference	
SALK_033462C	At2g32370	HOMEODOMAIN GLABROUS 3	3	n/a	WT	No	
SALK_033647	At5g60910	AGAMOUS-LIKE 8/FRUITFUL	n/a	3	WT	No	Produces shriveled siliques
SALK_038543C	At4g29080	INDOLE-3-ACETIC ACID INDUCIBLE 27	3	2	WT	Yes/ no difference	
SALK_039825C	At1g79840	GLABRA 2	3	n/a	WT	Yes/ no difference	
SALK_040513	At2g33860	AUXIN RESPONSE FACTOR 3	3	n/a	WT	Yes/ no difference	
SALK_040812C	At5g04410	NAC DOMAIN PROTEIN 2	WT	WT	WT	Yes/ no difference	
SALK_040835C	At2g34600	JASMONATE-ZIM-DOMAIN PROTEIN 7	n/a	n/a	n/a	Yes/ difference	
SALK_042113C	At2g27050	ETHYLENE-INSENSITIVE3-LIKE 1	3	3	WT	Yes/ difference	
SALK_042196C	At3g16857	RESPONSE REGULATOR 1	WT	3	WT	Yes/ no difference	
SALK_042715	At2g42400	ZINC FINGER PROTEIN 2	1	WT	WT	Yes/ no difference	
SALK_042853C	At1g17520	Unknown	WT	2	WT	Yes/ no difference	
SALK_043433	At3g58680	MULTIPROTEIN BRIDGING FACTOR 1B	2	1	WT	Yes/ difference	
SALK_044860	At1g06180	MYB DOMAIN PROTEIN 13	n/a	n/a	n/a	Yes/ no difference	
SALK_045370C	At5g67480	Unknown	2	2	WT	Yes/ no difference	
SALK_045791C	At1g55970	HISTONE ACETYLTRANSFERASE 6	3	WT	WT	Yes/ no difference	
SALK_046891C	At5g64530	NAC DOMAIN CONTAINING PROTEIN 104	WT	n/a	WT	Yes/ no difference	
SALK_047120C	At4g01250	WRKY22	n/a	n/a	n/a	Yes/ no difference	
SALK_047306C	At1g53910	RELATED TO AP2 12	WT	1	WT	Yes / difference	
SALK_047772C	At2g42830	AGAMOUS-LIKE 5	2	WT	WT	No	
SALK_049142C	At1g20693	HIGH MOBILITY GROUP B2	1	2	WT	Yes / difference	
SALK_050488C	At5g59340	WOX2	1	WT	WT	Yes / difference	

SALK Line	Gene	Gene Name	Oil Tier	Protein Tier	Flux	Expression in embryo	Notes
SALK_051138C	At3g06380	TUBBY-LIKE PROTEIN 9	1	2	WT	Yes/ no difference	
SALK_051647C	At2g38090	Unknown	n/a	n/a	n/a	No	
SALK_051936C	At5g16680	Unknown	1	n/a	WT	Yes/ no difference	
SALK_052158C	At4g17880	MYC4	n/a	n/a	WT	Yes/ difference	
SALK_053198C	At5g24470	PSEUDO-RESPONSE REGULATOR 5	1	n/a	Possible	Yes/ no difference	
SALK_054092C	At1g27730	SALT TOLERANCE ZINC FINGER	3	WT	Possible	Yes/ no difference	
SALK_054113	At2g42150	Unknown	1	n/a	WT	Yes/ no difference	
SALK_055032C	At5g20900	JASMONATE-ZIM-DOMAIN PROTEIN 12	WT	WT	WT	Yes/ no difference	
SALK_055989C	At1g77950	AGAMOUS-LIKE 67	1	1	WT	Yes/ difference	
SALK_056605	At1g15050	INDOLE-3-ACETIC ACID INDUCIBLE 34	n/a	n/a	n/a	Yes/ no difference	
SALK_058642	At2g47070	SQUAMOSA -LIKE 1	2	n/a	WT	Yes/ no difference	
SALK_059705C	At2g03340	WRKY3	n/a	n/a	n/a	Yes/ no difference	
SALK_059823C	At5g43170	ZINC-FINGER PROTEIN 3	1	WT	WT	Yes/ no difference	
SALK_061079C	At3g56850	ABRE 3	WT	WT	WT	Yes/ no difference	
SALK_062413C	At5g17800	MYB DOMAIN PROTEIN 56	1	WT	WT	Yes/ no difference	
SALK_063280C	At3g57800	Unknown	n/a	n/a	n/a	Yes/ no difference	
SALK_063576C	At4g27410	RESPONSIVE TO DESICCATION 26	WT	WT	WT	Yes/ no difference	
SALK_063778C	At2g46530	AUXIN RESPONSE FACTOR 11	3	n/a	WT	Yes/ difference	
SALK_063950C	At5g54070	HEAT SHOCK TF A9	1	2	WT	Yes/ difference	
SALK_065223	At1g24260	SEPALLATA3	n/a	n/a	n/a	No	
SALK_065384C	At3g62100	INDOLE-3-ACETIC ACID INDUCIBLE 30	2	n/a	WT	Yes/ difference	
SALK_065473	At5g06100	MYB DOMAIN PROTEIN 33	WT	WT	WT	Yes/ no difference	
SALK_065697C	At1g04250	INDOLE-3-ACETIC ACID INDUCIBLE 17	WT	WT	WT	Yes/ difference	
SALK_067627	At1g01720	NAC DOMAIN PROTEIN 2	n/a	n/a	n/a	Yes/ no difference	
SALK_067825C	At3g17860	JASMONATE-INSENSITIVE 3	3	1	Possible	Yes/ difference	
SALK_069339C	At2g22840	GROWTH-REGULATING FACTOR 1	1	WT	WT	Yes/ no difference	

SALK Line	Gene	Gene Name	Oil Tier	Protein Tier	Flux	Expression in embryo	Notes
SALK_069978C	At5g52660	Unknown	3	3	WT	Yes/ no difference	
SALK_070506C	At5g60450	AUXIN RESPONSE FACTOR 4	WT	2	WT	Yes/ no difference	
SALK_070715C	At5g13330	RELATED TO AP2 6L	2	WT	WT	Yes/ no difference	
SALK_072590C	At5g10140	AGAMOUS-LIKE 25	n/a	n/a	n/a	Yes/ difference	
SALK_074601C	At5g46690	BETA HLH PROTEIN 71	1	1	n/a	Yes/ no difference	
SALK_074642C	At4g28190	ULTRA PETALA 1	WT	WT	WT	Yes/ no difference	
SALK_074896C	At5g37260	REVEILLE 2	3	1	WT	Yes/ difference	
SALK_080380C	At1g79000	HISTONE ACETYLTRANSFERASE 1	2	WT	WT	Yes/ no difference	
SALK_080724C	At2g24260	LJRHL1-LIKE 1	WT	1	WT	Yes/ difference	
SALK_081983C	At3g14180	ASIL2	3	WT	WT	Yes/ no difference	
SALK_082546C	At5g17490	RGA-LIKE PROTEIN 3	WT	WT	WT	Yes/ no difference	
SALK_082840C	At2g46270	G-BOX BINDING FACTOR 3	3	3	WT	Yes/ difference	
SALK_083813C	At3g24500	MULTIPROTEIN BRIDGING FACTOR 1C	3	n/a	WT	Yes/ difference	
SALK_085482C	At3g28920	HOMEOBOX PROTEIN 34	3	n/a	WT	Yes/ difference	
SALK_085497C	At3g44460	bZIP67	WT	WT	Yes	Yes/ no difference	
SALK_086597C	At3g13960	GROWTH-REGULATING FACTOR 5	n/a	n/a	n/a	Yes/ no difference	
SALK_086776C	At4g00180	YABBY3	2	WT	WT	Yes/ no difference	
SALK_087012C	At3g59060	PIL6	3	2	Possible	No	
SALK_087702C	At1g77450	NAC DOMAIN PROTEIN 32	n/a	n/a	n/a	Yes/ no difference	
SALK_088606C	At2g30470	VP1/ABI3-LIKE 1	WT	2	Yes	Yes/ difference	
SALK_089129C	At1g73730	ETHYLENE-INSENSITIVE3-LIKE 3	n/a	n/a	n/a	Yes/ no difference	
SALK_089483C	At5g44800	PICKLE RELATED 1	n/a	n/a	n/a	Yes/ no difference	
SALK_089809C	At3g23050	INDOLE-3-ACETIC ACID INDUCIBLE 7	1	2	WT	Yes/ no difference	
SALK_090154C	At5g39610	NAC DOMAIN CONTAINING PROTEIN 92	WT	WT	WT	No	
SALK_090239C	At3g62090	PIL 2	2	2	WT	Yes/ difference	
SALK_090261C	At5g60100	PSEUDO-RESPONSE REGULATOR 3	1	n/a	WT	Yes/ difference	

SALK Line	Gene	Gene Name	Oil Tier	Protein Tier	Flux	Expression in embryo	Notes
SALK_090665C	At2g18090	Unknown	3	2	n/a	Yes/ no difference	
SALK_090859C	At3g58630	Unknown	2	WT	Possible	Yes/ no difference	
SALK_091352	At4g30935	WRKY32	3	2	WT	Yes/ difference	
SALK_091887C	At5g47370	HAT2	3	WT	WT	Yes/ no difference	
SALK_091933C	At4g32280	INDOLE-3-ACETIC ACID INDUCIBLE 29	n/a	n/a	n/a	No	
SALK_092176C	At2g40770	Unknown	n/a	n/a	n/a	Yes/ no difference	
SALK_092889C	At1g46768	RELATED TO AP2 1	3	WT	WT	Yes/ difference	
SALK_092897C	At5g42780	HOMEOBOX PROTEIN 27	WT	1	WT	Yes/ no difference	
SALK_094943	At3g02990	HEAT SHOCK TF A1E	2	1	WT	Yes/ no difference	
SALK_095699C	At1g21970	LEC1	2	n/a	Yes	Yes/ no difference	
SALK_095791C	At2g42430	ASYMMETRIC LEAVES2-LIKE 18	n/a	n/a	n/a	Yes/ no difference	
SALK_096150C	At1g32870	NAC DOMAIN PROTEIN 13	WT	1	WT	Yes/ difference	
SALK_096310C	At1g52890	NAC DOMAIN PROTEIN 19	n/a	n/a	n/a	Yes/ difference	
SALK_099250C	At1g28420	HOMEOBOX-1	WT	2	Possible	Yes/ no difference	
SALK_099786	At5g22290	NAC DOMAIN PROTEIN 89	WT	1	WT	Yes/ no difference	
SALK_101961C	At5g07500	PEI1	n/a	n/a	WT	Yes/ no difference	
SALK_102071C	At1g21450	SCARECROW-LIKE 1	1	WT	Possible	Yes/ no difference	
SALK_102383C	At2g38880	NUCLEAR FACTOR Y SUBUNIT B1	WT	1	WT	Yes/ no difference	
SALK_102687C	At3g11020	DREB2B	2	2	WT	Yes/ no difference	
SALK_103138C	At5g22220	E2F TRANSCRIPTION FACTOR 1	2	1	WT	Yes/ no difference	
SALK_103619C	At2g18280	TUBBY LIKE PROTEIN 2	WT	n/a	WT	Yes/ difference	
SALK_103716C	At2g24430	NAC DOMAIN PROTEIN 39	1	n/a	WT	No	
SALK_103775	At5g67110	ALCATRAZ	n/a	n/a	n/a	Yes/ no difference	
SALK_103823C	At3g49530	NAC DOMAIN PROTEIN 62	1	1	n/a	Yes/ no difference	
SALK_104032	At5g46910	Unknown	n/a	n/a	n/a	Yes/ no difference	
SALK_108221C	At5g13790	AGAMOUS-LIKE 15	WT	2	WT	Yes/ no difference	

SALK Line	Gene	Gene Name	Oil Tier	Protein Tier	Flux Phenotype	Expression in embryo	Notes
SALK_109425C	At4g04890	PROTODERMAL FACTOR 2	WT	WT	WT	Yes/ no difference	
SALK_111422C	At1g56170	NUCLEAR FACTOR Y, SUBUNIT C2 (NF-YC2)	3	3	WT	Yes/ no difference	
SALK_111812C	At3g27785	MYB DOMAIN PROTEIN 118	2	1	WT	No	
SALK_118611C	At3g11440	MYB DOMAIN PROTEIN 65	n/a	n/a	n/a	Yes/ no difference	
SALK_118938C	At1g17950	MYB DOMAIN PROTEIN 52	2	WT	WT	No	
SALK_123839C	At1g27650	ATU2AF35A	2	2	WT	Yes/ no difference	
SALK_126332C	At1g33420	Unknown	n/a	n/a	n/a	Yes/ difference	
SALK_130355C	At5g06960	OCS-ELEMENT BINDING FACTOR 5	2	n/a	n/a	Yes/ no difference	
SALK_130430C	At4g32570	TIFY DOMAIN PROTEIN 8	n/a	n/a	n/a	Yes/ difference	
SALK_130848C	At3g58710	WRKY69	2	3	WT	Yes/ no difference	
SALK_131796C	At3g54340	APETALA 3	2	2	WT	No	
SALK_132120C	At4g29230	NAC DOMAIN PROTEIN 75	2	WT	WT	Yes/ no difference	
SALK_132562C	At3g19580	ZINC-FINGER PROTEIN 2	WT	2	WT	Yes/ no difference	
SALK_132571	At3g04060	NAC DOMAIN PROTEIN 46	n/a	n/a	n/a	No	
SALK_132592	At3g10040	HYPOXIA RESPONSE ATTENUATOR1	3	WT	WT	Yes/ no difference	
SALK_133100C	At2g17560	CHROMATIN ASSEMBLY FACTOR D4	3	1	WT	Yes/ no difference	
SALK_133857C	At5g65410	HOMEOBOX PROTEIN 25	1	1	Yes	Yes/ no difference	
SALK_135824C	At3g54990	SCHLAFMUTZE	n/a	n/a	n/a	Yes/ difference	
SALK_137131C	At3g15510	NAC DOMAIN PROTEIN 56	3	WT	WT	Yes/ no difference	
SALK_141414C	At5g50915	Unknown	2	1	WT	No	
SALK_141618C	At1g08320	Unknown	WT	1	WT	No	
SALK_142727C	At1g77800	bZIP21	n/a	n/a	n/a	Yes/ no difference	
SALK_143721C	At2g25900	Unknown	WT	WT	Possible	Yes/ no difference	
SALK_148646C	At1g08810	MYB DOMAIN PROTEIN 60	3	WT	n/a	No	
SALK_149402C	At2g23760	SAWTOOTH 2	3	WT	WT	Yes/ no difference	
SALK_149765C	At5g16770	MYB DOMAIN PROTEIN 9	3	2	WT	No	

SALK Line	Gene	Gene Name	Oil Tier	Protein Tier	Flux Phenotype	Expression in embryo	Notes
SALK_149951	At2g24790	CONSTANS-LIKE 3	n/a	n/a	WT	Yes/ no difference	
SALK_150356C	At5g24590	NAC DOMAIN PROTEIN 91	WT	1	WT	Yes/ no difference	
SALK_150387	At3g07670	Unknown	3	1	Possible	Yes/ no difference	

Appendix D

Primers used in this study and genotyping results

Table D.1 Results of the genotyping project and the primers used.

Prior to the screen, SALK T-DNA insertion mutants were genotyped to ensure homozygosity. The results from that project are listed.

SALK Line	AGI	Homozygote Acquired (Y/N)	Forward Primer	Reverse Primer	Band Size (F+R)	Band Size (FIR)
SALK_000196C	At4g00730	Y	TTCTCGATCCATCTTTTCGTG	TAAACGAGCGTTTTCGATACG	1073	465-765
SALK_000247	At2g24500	Y	AAGCTCGAACAAAGGAACATC	ATCCAAATCCTCATCCGAATC	1173	591-891
SALK_000287C	At1g28470	Y	TGGTTATCGCGATTTCATTTC	CTCGAGGTTAAAGTTACGCCC	1166	563-863
SALK_001014C	At1g43700	Y	GACGCTAACAACACAAAAGCC	TAGCGAATAGACAATCTGCGG	1092	511-811
SALK_001597C	At1g34180	Y	CTGATGAGAACTGGCTCCTTG	TCTCAATGAAATCCCAGATGC	1065	492-792
SALK_001658	At4g24150	Y	TTGGCAATACTTGGACCTGTC	GTCCTCGTTCAAGAAAGCATG	1004	440-740
SALK_002516C	At1g50420	N	TTCCTCTGTTCTTTAACCCCC	AGCGCAGTTCTTTCTCATGAG	998	473-773
SALK_003155	At3g26744	N	ATTCTTTGCTCTGCCTCTTCC	TTTGTAGGGCCTTTGTTGTTG	1231	562-862
SALK_003943C	At2g47260	N	CGGTGGGTTTATCAACAAATG	ATTCTTGATCACGATTCACGG	1043	460-760
SALK_004037C	At5g57180	Y	CTTCACCATTCGCTCACTCTC	TTTTCGATTTCACCACCGTAG	1099	456-756
SALK_005260	At5g35550	Y	GCGGGTCAGAATCTAGTTTCC	AAGCAGATGGTCGTTGATAGC	1162	501-801
SALK_006229C	At2g03710	Y	TTGATTTGGATCCATTGTTGC	GGAGGAGAGTTTTCAGTTGGG	1076	519-819
SALK_006473	At5g57520	Y	GCACGTGTGGACCTCTCTTAG	CTTTATCTCTTCTCCGGCGTC	1224	576-876
SALK_006751	At5g07680	Y	TGTTGTTGTTGCAGCTTTAGC	CCCCTTTTAAAATGAAGTTTGC	1205	579-879
SALK_008040C	At1g05380	Y	GCCATCTTCACATGATGATCC	TTTGGATGTCTTGCATTTTCC	1067	540-840
SALK_008559C	At3g54320	Y	AAAAACAGGCATCACCACAAC	TTTGTTCGGAGGACAAAGATG	1146	521-821
SALK_008606C	At5g67300	Y	TTGTCAATTTGTCATGCACTG	AATCGTCGCCCATTTATTACC	1059	446-746
SALK_010653C	At4g36900	Y	GCTACTGGAATCAAAGCAACG	TGGAAAATGTCATAAAGTGCC	1242	561-861
SALK_011023C	At4g34610	Y	GTGCTCTTTGCTATCATGCAG	GATCATGCTAGCAAGACAAACG	1081	511-811
SALK_011042C	At1g30810	Y	TTTGAAACATTGTGCTTTTTGG	TTTGTTTTGGCCCAATTAATG	1141	506-806
SALK_011069C	At3g15500	Y	TTATGAATCGCTGATTCGTTG	AGTTATCTCGACGGAGGGAAG	1111	496-796
SALK_011502C	At5g18270	Y	GAAAAGAGGAGATTTCTTAGGGC	TTGGATGAAACCTGAAACCAG	960	437-737
SALK_011957C	At1g19180	Y	AGGTAAATGCGGAGAGAGAGG	AGGCACCGCTAATAGCTTAGC	1078	490-790

SALK Line	AGI	Homozygote Acquired (Y/N)	Forward Primer	Reverse Primer	Band Size (F+R)	Band Size (FIR)
SALK_012997C	At1g62300	Y	GAACGTATTAGCCAATCACGC	TGTGGACGTGTCATAATTTGG	1202	536-836
SALK_013163C	At2g36270	Y	CAATGGAAGTTCGGAATCATG	CACTCGTTTTCTTCTTAAAGCG	1203	551-851
SALK_014315	At3g10590	Y	TTCTGAGTCAAGTTGTGCGTG	TATGTCAAAAATACTCCGGCG	1015	453-753
SALK_014999	At4g18960	N	GTATTTGGAAATTGCGTGTCG	TTTACGAGGGTTTTCGATCTTC	1211	575-875
SALK_015495C	At1g32770	Y	GATTCTGGAAAGCTACCGGAC	GTTCAATTGCCAAGCTACGAG	1047	441-741
SALK_016142C	At3g26790	-	GCTTTCTTGGGTCAATTTTCTG	TTCTTGGTCGCAAAAATGTTC	1140	492-792
SALK_016183C	At1g76590	Y	ATCATGAGCATTTCAAATGCC	GAATGAAGCGTGATCCAAGTC	1025	457-757
SALK_016333C	At3g09600	N	TGAAACGTCCTTTGATGGAAC	GCCATGTGGACTTTTCTTTTTC	1105	521-821
SALK_016554	At1g19000	Y	AATCACATTTCATCTTCATGCG	TTCTGCAACCCAATCAAGAAC	1072	476-776
SALK_017963C	At4g24660	Y	TACAATCGTCCATGGCTTCTC	TTGTTGCGAGTTTATGCCTTC	1082	547-847
SALK_018618C	At2g35550	Y	AGCAGGAGGTTCCTGAGAAAG	CCAAAGTCCAAAGACTTTCCC	1031	504-804
SALK_020767C	At1g78080	Y	GCAGGTGAGGTTAGTTCATCG	TTGTACGCCGCCTTATCATAG	1148	530-830
SALK_020840C	At5g46880	Y	ACCATTATGACCCTCCCTCAC	TAAGGACACCGTGAGGATCAC	1066	520-820
SALK_021965C	At2g41070	Y	TCCACTTAGGATCTGGTGGTG	ATACTGCATGCAAATTCCCTG	1047	509-809
SALK_022174C	At1g34190	Y	TGTTACGTAGATGGCGGATTC	TTCTCAGTTGAAGAAGCTCCG	1248	605-905
SALK_022235C	At5g43270	Y	CTTTAAACCGAGAACCGGATC	TTTCTTCTTCCTCTGGGATCC	1154	608-908
SALK_022648	At3g23030	N	AACCCATTTTGAAGGGAAATG	TTTGGAGAAAAGGGGAGAGAG	1115	558-858
SALK_022661C	At2g21240	Y	CAGTTCGTTTAGTCGTCTCGG	ATTGGGACAGTCAAGATGTCG	1038	471-771
SALK_023411C	At3g24650	Y	AATGGAAAGCAAAGGCTACAAG	TCCATTTCAAAACCCTAACCC	1034	516-816
SALK_024494C	At1g22590	Y	AGAGAAGAGGATTTTGGGTGG	TTACACACACACACCTCGTATG	1114	432-732
SALK_024760C	At5g01310	Y	GAAGACATGGATTGGGACAAG	CCCAAATTTACATTAAAACGCC	1179	605-905
SALK_024800C	At2g40140	Y	CAATTTGCGTTGCTCTTTCTC	TCTGTTCTGCCATAAACCACC	1191	598-898
SALK_025198C	At3g56400	-	TGATCTTCGGAATCCATGAAG	CAAACCACACCAAGAGGAAAG	1119	487-787

SALK Line	AGI	Homozygote Acquired (Y/N)	Forward Primer	Reverse Primer	Band Size (F+R)	Band Size (FIR)
SALK_025279C	At1g74950	Y	AGCCTGGTCTGATCTACTCCAC	TCTACGGTGGTCGAGTTATGG	1280	590-890
SALK_025754C	At5g17300	Y	AAGTGGAGATGAATCTCATGCTC	CAAAGACCGCAGTTCAGATTC	1162	605-905
SALK_025839C	At3g59580	Y	ACATCCAAGGATCAGCAATTG	TATGTTTTTCAGGCCGTGAAC	1085	481-781
SALK_028204	At5g65210	-	AAACCTGGATTCATGGTTTCC	GTGTCCCCTCTGGTTTCTTTC	1091	435-735
SALK_030459C	At3g16770	Y	ATTCTGGTTCAACGAGCACTG	GTACATACCTTGGTTGGTGGG	1015	439-739
SALK_030799C	At3g53370	Y	TGTACCGGTTGTGCCTAAGTC	CGATTAGGAACACAAGAAGCG	1104	569-869
SALK_030876C	At1g43640	Y	CCCACTTTCAACAAGTATGGC	TTGTTTTGCTACCCACTCAAAG	1191	504-804
SALK_030989C	At5g12840	Y	AAGACGATCTGGCCAAAACTC	GGCTATGAAGCAATGCAGAAG	1161	550-850
SALK_031064C	At2g43500	Y	TTGACCCCACTTCTGAAACAG	GCGGTTATTTCATTCTCCCAC	1084	512-812
SALK_032192	At3g54220	N	AACGGAAGACGTCAAACACAC	GGAGGAGAAGGATTGTTGGAG	1158	489-789
SALK_033462C	At2g32370	Y	CTGTGCAAAAACGTTGTGTTG	CTGCAAAAATCCAAATCTTGG	957	445-745
SALK_033647	At5g60910	Y	AATTGTCCTTCTTGCTGACCC	CGATCGAGAAGTTGAGTTTGG	1086	441-741
SALK_038543C	At4g29080	Y	GTGGATCCAGGAGAGAAGACC	CACATGATTAAATGGTTCACCG	1054	522-822
SALK_039825C	At1g79840	Y	ACCACCGATCAGATCAGACAC	GGAGTTTTCGAGGTGGAGATC	1019	486-786
SALK_040513	At2g33860	Y	AACGGTAGTCCAAATCTTCGC	TCTGCTTTGAGAGCAGAGGAG	1153	499-799
SALK_040812C	At5g04410	Y	GGTCAGCGTCTGATGAAGAAG	AAGAAGCCCTTGACCAGAAAG	1215	529-829
SALK_040835C	At2g34600	N	CAAACAAACTTCACATCAATCCC	GTTGGGATCGAAGATAGAGGG	1062	464-764
SALK_042113C	At2g27050	Y	GCATGAGAATGATGTTCAATGG	TGGTTTACCAAATGAGCAAGG	1277	598-898
SALK_042196C	At3g16857	Y	GATCAAACCCATTCAATGTCG	GAGATGGCATTGTCTCTGCTC	1009	438-738
SALK_042715	At2g42400	Y	AAGAGAGAAGCAAAACTGGGC	TGGAGGAGAAGCTAAGGGAAC	1243	562-862
SALK_042853C	At1g17520	Y	ACCCAAAGCGAGAATCATACC	TTTGGTGTCGAAGTAAATGCC	1259	606-906
SALK_043433	At3g58680	Y	ATTTTTGTTTCTCCCCCACAG	TGTTGATGGGTTATTGGGATG	1178	509-809
SALK_044860	At1g06180	N	GGAAAGTTATTGAAGTTCATGCG	CAGCTTGTCATTTTCCTACGG	1203	606-906

SALK Line	AGI	Homozygote Acquired (Y/N)	Forward Primer	Reverse Primer	Band Size (F+R)	Band Size (FIR)
SALK_045370C	At5g67480	Y	GGGATGGCTTTTCTTCATAGC	TGCTGTGAAAATTGTTGTCTGAG	1011	494-794
SALK_045791C	At1g55970	Y	TCTTCATGCGTCCCATAGTTC	CATGTGTGACTGATCAGGTGG	1257	573-873
SALK_046891C	At5g64530	Y	GGTGTATCATGAGATCGCACC	TTTGCGTGGAACTATCCAATC	1245	588-888
SALK_047120C	At4g01250	N	TACTGCTGACGGATTATTCCG	CCTTTACCAAAAATGTAACGCAG	1125	545-845
SALK_047306C	At1g53910	Y	TTTGGAGTCGAAGTGAAAACG	CGAAACGACAAAAGCTTTGTC	1069	558-858
SALK_047772C	At2g42830	Y	AAGACCCTAATGTGTACTTCCCC	GAAAATGAATCAGGGACAAAGC	1041	438-738
SALK_049142C	At1g20693	Y	TGAAAACGAAGAAAGCACTGG	AAAGCAAGAGGATCCATGGAC	1191	570-870
SALK_050488C	At5g59340	Y	TGTTGACGACGTTTTCTTTCC	AGAATGTCTATGTCGGCCTCC	1108	478-778
SALK_051138C	At3g06380	Y	TTATGGCCGGTCAAAAGTATC	AGCTTTTGTTGAGCTGAGCAC	1052	537-837
SALK_051647C	At2g38090	N	TATGGCGCACAGGAAAATTAG	CCAAACATATTTTCTCTGCGC	1048	492-792
SALK_051936C	At5g16680	Y	CGAGGTGAAGCTAAAAGGAGG	CGCTAACGAGAATTCATCTGC	1126	471-771
SALK_052158C	At4g17880	Y	TTTATGCGATTCCGAAATGTC	TGAAAGCTTCCATTACCGATG	1106	515-815
SALK_053198C	At5g24470	Y	CGGAGGAAAAAGAGTACGAGG	CCAGCTTTTACACTGGAGCTG	1246	563-863
SALK_054092C	At1g27730	Y	TATTTTGTAAGGCGGCATCAG	AAGTCAAACCGAGGCTTCTTC	1092	493-793
SALK_054113	At2g42150	Y	TTTACTTGACGTGGTTCCCAG	ATCGTACTGTTCAACTTCGCG	1123	539-839
SALK_055032C	At5g20900	Y	AGTTATGGCACACTCCCATTG	AGCATCAGTCCTGTCTCATCG	1025	454-754
SALK_055989C	At1g77950	Y	TACAGTCCAGAAGCCACCAAC	AAGCCTCTAGCTAGTCAAACAGG	1030	455-755
SALK_056605	At1g15050	N	TTTCTGTATTTGGAGGGTGTTG	TTACCACTCATCTGGCCAAAG	1132	608-908
SALK_058642	At2g47070	Y	CTGATGAACAACATGTGACCG	TCTGGATCTCTTAGCCTGCTG	1067	553-853
SALK_059705C	At2g03340	N	TTCTTTGTTCAAACCCTCGTG	ATGGTATTTTTGGGATTTGGG	1148	510-810
SALK_059823C	At5g43170	Y	TCATGATTCTTGGTTGCAATG	AACAAACGTGAGATTTCACGG	1073	540-840
SALK_061079C	At3g56850	Y	TCTTCTTTGATGGGTGGTTTG	CAAGCTTGGCTATTGCAGAAC	1173	603-903
SALK_062413C	At5g17800	Y	TCCCTTAATCCCTAAACCCAG	CCTGATGCAAGGGTACTATCG	1117	458-758

SALK Line	AGI	Homozygote Acquired (Y/N)	Forward Primer	Reverse Primer	Band Size (F+R)	Band Size (FIR)
SALK_063280C	At3g57800	N	CGATACTACCCCTTCCTCACC	CTGAGACCGATTCATCTCAGC	1065	545-845
SALK_063576C	At4g27410	Y	AGTGATCGAGTGCTTCAGGAC	ACTCGTGCATAATCCAGTTGG	1222	593-893
SALK_063778C	At2g46530	Y	TAACGGTCCACATTCGTTCTC	TGGGTCAAGACTTGTTGGTTC	1165	504-804
SALK_063950C	At5g54070	Y	AATATTCACAAGCATGGCTGC	CTGCACAGTGACCATCTGATG	1166	498-798
SALK_065223	At1g24260	N	ACCAGTTTACAAACCCGAACC	GGTGGAAGTGGAGTGTCAAAG	1048	525-825
SALK_065384C	At3g62100	Y	CGGAACAATTGTAATATCTCCG	AGGGAGAAGCTCATCGTCTTC	1270	579-879
SALK_065473	At5g06100	Y	CTGACAGATTCTGAACGGCTC	TCCACAAACGATTTCCAAAAC	1033	512-812
SALK_065697C	At1g04250	Y	CGATTTTCCTCAAGTACGGTG	TTTCCTTCACTTGTGCTTTCG	1132	609-909
SALK_067627	At1g01720	N	ACCGATGAAGAGCTTGTCATG	TGAACAGAAATCTAATGGCCG	1124	538-838
SALK_067825C	At3g17860	Y	TCTTCCTCCTGGTGCATAATG	AATGTTGGAAAAGAGGGGATG	1222	575-875
SALK_069339C	At2g22840	Y	TTGGTTTTGTTGTGGTGTCAG	CATCATTGAATCATCAAGAAAAGG	1107	592-892
SALK_069978C	At5g52660	Y	CACCAACGACTTCTTCTACGG	AGAGCTCAACGAGCTCACTTG	1296	606-906
SALK_070506C	At5g60450	Y	GTGCGTCCTCTCATCTTTTTG	CGCGACAAATCAATAGCTCTC	1196	577-877
SALK_070715C	At5g13330	Y	AACCACCGATCAGCATACAAC	TGCATGTGCAATTCTACATGC	1143	541-841
SALK_072590C	At5g10140	N	GATCGGATTATACGAATTGCG	TACCGAACCACATACTTTGCC	1148	484-784
SALK_074601C	At5g46690	Y	GTGGCTTCTCTACAATCGCTG	ATCCAAGATCGAGAGCCTAGC	1062	480-780
SALK_074642C	At4g28190	Y	TTTGACAATGGAACCTTTTCG	TCTTCTCTTCTCCCGAAAAGC	1050	485-785
SALK_074896C	At5g37260	Y	CAGATGCCTTTTACCTCAAGG	ACGAACACTTCACCACAAAGG	1123	556-856
SALK_080380C	At1g79000	Y	AGGGAACATGTCATCCATGAG	AGGGAGAATCTGAGAACCTGC	1090	578-878
SALK_080724C	At2g24260	Y	TCAACGTCTTGCACGATTTATC	GATTCGTGCGATCCATAATTC	1105	498-798
SALK_081983C	At3g14180	Y	CACCTCTTTCCAGTGCTTCTG	ATTGTGCTAAACGCATTGGAC	1201	532-832
SALK_082546C	At5g17490	Y	GTCGGTCCAAATTCATTCATG	GTCCCTACCTGAAATTCGCTC	1203	598-898
SALK_082840C	At2g46270	Y	AAATGGGAATCCTGAAAATGG	GAGCATTTCAGTTTGTCCTGC	1007	461-761

SALK Line	AGI	Homozygote Acquired (Y/N)	Forward Primer	Reverse Primer	Band Size (F+R)	Band Size (FIR)
SALK_083813C	At3g24500	Y	ATCCAATGATAATAAGGCGGC	TAAAACCATTGAGCCAAATCG	1224	585-885
SALK_085482C	At3g28920	Y	CTTCTTCACGACGAAGACGAC	GCAACATTCATCAACCAGAGC	1176	550-850
SALK_085497C	At3g44460	Y	ACGATGTAATTTCAGCATCGG	CTCGGTTTTGGGAGAATCTTC	1202	524-824
SALK_086597C	At3g13960	N	GATGAAGTGACAAAGCGGAAG	CGAAACTCAGGAGCCCTAATC	1060	532-832
SALK_086776C	At4g00180	Y	CCCTAGATCGTAGATCCTAAGGC	CATCTACAAGAGATGCCTCGG	1101	498-798
SALK_087012C	At3g59060	Y	CGATTTGTTACCCATGGTTTG	CCTTGCTCGATTTTTGTTACG	1228	557-857
SALK_087702C	At1g77450	N	ATTACAGTCTGCCACGTGTCC	ACCGTTGCTACACGTGAATTC	1064	514-814
SALK_088606C	At2g30470	Y	CCTCTTTCCAAAGACAGGAGG	GTCGCATTGGTCGTCTAGTTC	1072	560-860
SALK_089129C	At1g73730	Y	GGACAATAAAACTCATTCATGAGC	GGCTCATTCTCCATCCTATCC	1229	569-869
SALK_089483C	At5g44800	N	AAAGAGCACTACACAGCGAGG	CAAGACTGATAAGAAACGCGC	1089	565-865
SALK_089809C	At3g23050	Y	CGCAGAACATAAATATTGTGAGG	TATGATGATGTGCCAGCTACG	1261	576-876
SALK_090154C	At5g39610	Y	CGATCTTAGGGTTACGTTGGG	ATCTTCCCCAAACAGCTAAGG	1109	487-787
SALK_090239C	At3g62090	Y	TTCGTTTTTGGAAACGACATC	ACTCTGCGTTGAGACAAAACC	1022	491-791
SALK_090261C	At5g60100	Y	GGAGTCGGAGATGATTTCTCC	TGTCTTTGAAATTTATGCCGG	1143	489-789
SALK_090665C	At2g18090	Y	TGATTGTCTGGGAAAGACGTC	ATGATGTCCAAACTCTCGTGC	1183	564-864
SALK_090859C	At3g58630	Y	AGTTCGCGTTACCCATTTAGC	CCTCTTGCCAGTGTTTCTGAC	985	440-740
SALK_091352	At4g30935'	Y	CCTTGTACCCTATCTCCTCCG	GGTGAAAGGAAATCCTCATCC	1119	558-858
SALK_091887C	At5g47370	Y	TTTCCATCAACGGTTAACTGC	CGGAAAAGCAATTGAAGTCAG	1258	567-867
SALK_091933C	At4g32280	Y	CAAGGAGTTAGTGAGGGACTCG	GCGTGTTGGTTAAGATACCATC	1019	489-789
SALK_092176C	At2g40770	N	CTGCGAGATTTGTTAAGACGG	CGCATTCAACTCAAAGCTTTC	1094	454-754
SALK_092889C	At1g46768	Y	ATGTTGAATCGCGTTTTTGAC	GTATCCCTCTGTACGGCTTCC	1092	481-781
SALK_092897C	At5g42780	Y	TTGCGAACAAAAACAAGATGG	GCGTTACTGAGACGGTTCTTG	1054	458-758
SALK_094943	At3g02990	Y	CGAACTTTGCAGGCTTATTTG	TCCATTATCTGTCTCGCATCC	1250	607-907

SALK Line	AGI	Homozygote Acquired (Y/N)	Forward Primer	Reverse Primer	Band Size (F+R)	Band Size (FIR)
SALK_095699C	At1g21970	Y	TTTGCATGCAACAAATCACTC	GCTGGTCCAAGTTCAATTCTG	1246	586-886
SALK_095791C	At2g42430	N	TTTCTTCCTTTTTGCTTTGCC	CAATGGCCAGTGACTTAAAGC	1149	523-823
SALK_096150C	At1g32870	Y	AACCCACAGTTTTTGTAGGCC	TTTGAAATGAAGGGTACCAGG	1036	519-819
SALK_096310C	At1g52890	N	GGGTATCCAAGAAACTGACCC	AACCCTTTTGAGCTAACTCGC	1095	550-850
SALK_099250C	At1g28420	Y	CTTCGCAGTGTTCTAACATGC	ATTCCTGCTGGAGAAGGACTC	1133	512-812
SALK_099786	At5g22290	Y	TGGTACAAGTCAGAAGGCACC	CAGCACAACAAAGCAAAATTG	1210	554-854
SALK_101961C	At5g07500	Y	TTACACGTGTGCCTTCTCATG	AAATCCAAATCCAAATCGGAC	1278	607-907
SALK_102071C	At1g21450	Y	GGCTTTTGGGAGATTACCAAG	CTATCCGTTCTTCTCCTTCGC	1191	588-888
SALK_102383C	At2g38880	Y	AGGCACTTTGAATCCACATTG	TGAAACATGATCAAGATACGGC	1059	537-837
SALK_102687C	At3g11020	Y	TGGTGAGCAAGTAAAACACCC	TTTACGTGGTTTTGGTCTTGG	1079	505-805
SALK_103138C	At5g22220	Y	TGCGAACTCTGTTATGCAATG	TTCATCAGCCTGAGGAATGTC	1149	585-885
SALK_103619C	At2g18280	Y	GGATAGGTCTGGTTTTCTGGG	ATGATTGCCTTTGCATGACTC	1091	580-880
SALK_103716C	At2g24430	Y	AATATCCTGTATTCGTCGCCC	ATGAACGGAAGTCATGCATTC	1025	500-800
SALK_103775	At5g67110	N	TACTTGGTGGTCCAAAAATGC	ACGGCTAAAGTCTGCATCATG	1118	545-845
SALK_103823C	At3g49530	Y	TCCCATTTGCAAATATCGATC	AGTATTCAAATTGGGGATGGC	1028	507-807
SALK_104032	At5g46910	N	ACACCAATCAATGGCTTCAAC	CGTCTCTGTTCTTTTGGTTCG	1231	556-856
SALK_108221C	At5g13790	Y	CCGGTGTTACAAAGCTGCTAC	CCGTATCTGGAAAGTGTTTGC	1019	503-803
SALK_109425C	At4g04890	Y	CTCGTTCGCTTGATCTTGAAG	AGATCACCAGGAATGTTGCTG	1103	438-738
SALK_111422C	At1g56170	Y	CATGACCTGTTTGGGATCATC	CCTCTTCTGACTGCTCCATTG	1110	439-739
SALK_111812C	At3g27785	Y	TGTTGCTACCAAGTGAAAGGG	ACCCAACGTTTCACATGTCTC	1156	530-830
SALK_118611C	At3g11440	N	CACTAAGTTGCGAGGGACTTG	AGATGGGAAATAAATGGGCAC	1145	461-761
SALK_118938C	At1g17950	Y	TGGTTTGATTGGATCTTTTGG	AACCCTGAGAGGCAGAGTTTC	1081	515-815
SALK_123839C	At1g27650	Y	TGGTGACAGAGGTAATCCAGG	CGAAGCAGTGCTCCAAAATAG	1053	506-806

SALK Line	AGI	Homozygote Acquired (Y/N)	Forward Primer	Reverse Primer	Band Size (F+R)	Band Size (FIR)
SALK_126332C	At1g33420	Y	CCGATTTCGGAAAACTAAAGC	ACGACACAGCAGTGTTCACAG	1172	493-793
SALK_130355C	At5g06960	Y	CGATTACTTCTCACGGCTACG	TTGAAATGAGAACCAACGGAC	1284	610-910
SALK_130430C	At4g32570	N	ATCATGGTGAATCAAAGCTCG	AAATTCGTCATGAGTGGTTGG	1135	573-873
SALK_130848C	At3g58710	Y	TTCAAGAATCGGATGACGAAG	CTTCTCTCCACCTGCTTCCTC	1069	529-829
SALK_131796C	At3g54340	Y	ATAGACAATGATGGCACCAGC	GAGTGTTTGGACGAGCTTGAC	1101	430-730
SALK_132120C	At4g29230	Y	ACCTGGTAATTTTCATTGGGG	CGAGCTTGAACGCTATAGGTG	1094	548-848
SALK_132562C	At3g19580	Y	TTGTTGCTAACAAGCATGTGC	GCCAGAATCAAAGAACCTTCC	1012	479-779
SALK_132571	At3g04060	N	TTCCAATAACCAGACACGGTC	ATGAAGGTCATGTGGCTATCG	1167	513-813
SALK_132592	At3g10040	N	ATTGACAGCGTGAGAGAATCC	TTGATCATGGCGGTTTTCTAC	1163	530-830
SALK_133100C	At2g17560	Y	TCAGTCTTTCTGCTCTCAGCC	ATGCGTCAATTGGTGAAGAAC	1093	465-765
SALK_133857C	At5g65410	Y	CCTAGTGTGTTTTAAGATGAGCCC	AAACTCACCACAACCATCGAC	983	439-739
SALK_135824C	At3g54990	N	TGAAGGTTAGGCTTTGAATGG	TGGAAAGCAAGTTTATTTGGG	1204	605-905
SALK_137131C	At3g15510	Y	GGCACTGCGTCGTTATATAGG	AGACTCCACCATTGATGCAAC	957	437-737
SALK_141414C	At5g50915	Y	GCCTTCCCTGTTACCTATTCG	TCTAACATAAATTACCCGCCG	1087	435-735
SALK_141618C	At1g08320	Y	TTTTGACCTCTCGAATCGTTG	CCAATTGGGAAAGCTTTATCC	1028	430-730
SALK_142727C	At1g77800	N	GGAAGGATCCGAAAGAAGATG	GTCTGACCCAGAGGAGATTCC	1042	472-772
SALK_143721C	At2g25900	Y	TCGGATGTTACGACTTTGACTG	TCCAATCATGAGATCGACCTC	1147	464-764
SALK_148646C	At1g08810	Y	CGTTACGTTTGGCTTCATCTC	CTACCTATGCTTCAAGCACCG	1033	468-768
SALK_149402C	At2g23760	Y	TTTTGGTTTTGCAGAATTTGC	TGTAGATCCCTCCAATGGATG	1129	473-773
SALK_149765C	At5g16770	Y	TGGAATCGCAGTTTTTCTAGC	CGTGCATAAATATTTGCATGTG	1083	451-751
SALK_149951	At2g24790	Y	ACTTTGCAAGGCACATACCAC	CAGAATAGAGTTGCTGCCGTC	1056	466-766
SALK_150356C	At5g24590	Y	ATCTTCCTTTTCGCTCGCTAC	GAAGCCTGGGAATTGCTAAAC	1061	469-769
SALK_150387	At3g07670	Y	AACCGAACCAAACCGTAAATC	ACTGCATCAATGATTTCCAGC	1060	462-762