The downstream targets of MYB-type transcription factors involved in suberin biosynthesis

by

Daniel Klein

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

Carleton University
Ottawa, Ontario

ABSTRACT

Suberin is a lipid- and phenolic-based heteropolymer that is deposited in the cell walls of certain plant tissues to act as a physical barrier against unregulated gas, water, and solute movement. Suberin is deposited, for example, in root endodermis and periderm. It is also deposited in stem periderms (including tree bark), seed coats, and at wound sites. Suberin deposition in root endodermis of the model plant *Arabidopsis thaliana* is controlled by MYB-type transcription factors AtMYB53, AtMYB92 and AtMYB93 under non-stress conditions.. The transcriptional targets of these MYBs have not been previously reported. These gene targets may include suberin biosynthetic enzymes, transporters, and co-regulators not previously characterized. In this thesis, the transcriptomes of two Arabidopsis triple knockout *myb53 myb92 myb93* (TKO) mutants were compared to 14-day old wild-type seedlings using RNA-seq analyses. Additionally, the transcriptome of a steroid-inducible *AtMYB53* overexpression line was compared to an uninduced control in 10-day old seedlings. From these analyses, a set of new candidate suberin-associated genes was generated, including GDSL-type esterases / acyltransferases, an HXXXD-type (BAHD) acyl-transferase, additional MYB-type transcription factors, a WRKY transcription factor, ABCG-type transporters, and CASP-like assembly proteins. Sequential C-terminal deletions of each of these MYBs revealed important regions for their activity and may facilitate future experiments aimed at elucidating direct gene targets. Identification of potential gene targets of AtMYB53, AtMYB92, and AtMYB93 gene targets presents new opportunities to add to our understanding of suberin biosynthesis and its regulated deposition.

ACKNOWLEDGEMENTS

I would like to first thank my supervisor Dr. Owen Rowland for his continuous guidance and support throughout this project. I would also like to thank Sollapura Vishwanath from Agriculture and Agri-Food Canada for his help with setting up the RNA-seq analysis work-flow and for his continuous support in this part of the project. I'd like to thank Caroline Vergette (StemCore Laboratories, Ottawa Hospital Research Institute) for answering all my questions about sequencing and for the services she provided in a professional and timely manner. I'd like to thank Hefeng (Jasmine) Hu and Rebecca Kalinger, who were sources of continuous mutual support. Also, I'd like to express my gratitude to Jhadeswar Murmu, who initially mentored me and shared his endless expertise and technical knowledge with me. Lastly, I would like to thank the rest of the Rowland and Hepworth labs, family, and friends for their support and words of encouragement.

LIST OF TABLES

LIST OF FIGURES

LIST OF SUPPLEMENTARY MATERIALS

CHAPTER 1 - INTRODUCTION

1.1: Suberin is a protective hydrophobic barrier deposited at plant border tissues

The interactions between a plant and its environment are highly complex. Plants are constantly challenged by various biotic (e.g. insects and microbes) and abiotic (e.g. salt, drought, and temperature) environmental stressors (Franke and Schreiber, 2007). As one way to cope with these factors, cell wall-associated lipid-based polymers (i.e. cutin and suberin with their associated waxes) are deposited by the plant at environmental interfaces (Schreiber, 2010). A major function of these chemically similar polymers is to serve as hydrophobic barriers to prevent uncontrolled water loss. Cutin is deposited on the aerial surfaces of plants, for example as part of the stem and leaf cuticles (Schreiber, 2010). Meanwhile, suberin is deposited in various external and internal tissues, especially roots (Vishwanath *et al.*, 2015). Besides modulating water transport, both cutin and suberin have different tissue-specific functions and distinct chemical features. In terms of chemistry, cutin usually lacks the aromatic compounds and very-long-chain acyl chains (≥C20) typically present in suberin (Schreiber, 2010; Kosma *et al.*, 2014). Cutin and suberin composition also varies between plants and even between different organs of the same plant (e.g. root suberin versus seed coat suberin).

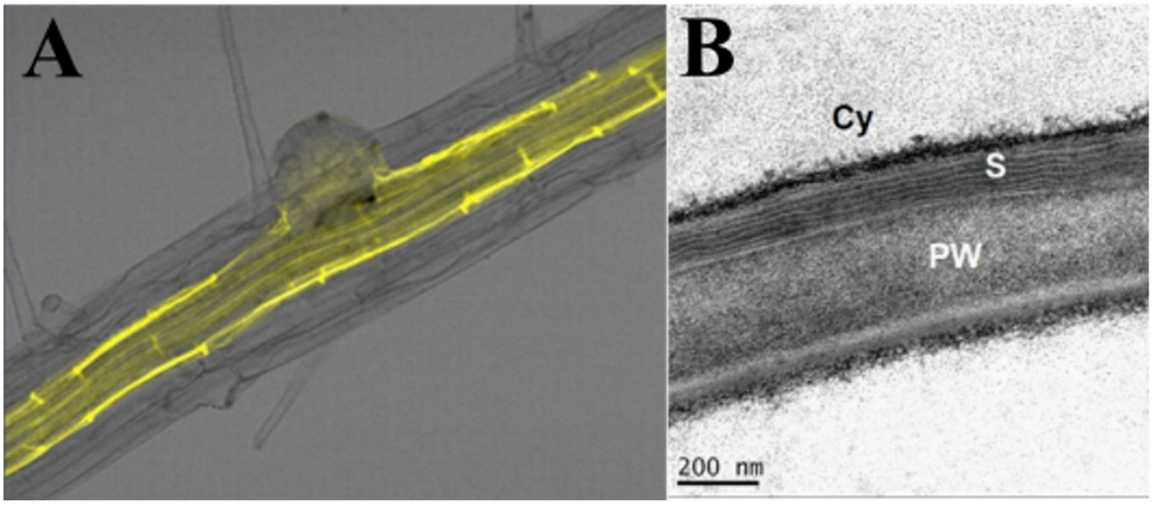


Figure 1: Patterns of suberin deposition in plant roots at the tissue and cellular levels. A) Fluoryl yellow staining of a 6-day old *Arabidopsis thaliana* root indicating the deposition of suberin in root endodermal cell walls. The image was obtained using laser scanning confocal microscopy. **B)** Transmission electron microscope (TEM) image showing the arrangement of suberin lamellae (S) on the inner face of the plant cell wall (PW) (Abbr. Cytoplasm (Cy)). Images courtesy of Nayana de Silva, Rowland Lab, unpublished data.

Suberin is a hydrophobic heteropolymer comprised of fatty acid derivatives, glycerol, and ferulate (Beisson *et al*., 2007). It is deposited on the inner face of the cell wall of specific cell types, including: 1) root endodermis (Figure 1A) and hypodermis, 2) the periderms of shoots, roots and tubers, 3) seed coat, 4) bundle sheath cells, 5) tree bark, and 6) wound sites (Schreiber, 2010). The primary function of suberin in these tissues is to act as a hydrophobic barrier to control the movement of water, solutes, and gases (Franke and Schreiber, 2007). While suberin may appear as a single layer, it has a multi-layered arrangement of alternating dark and light bands when imaged by transmission electron microscopy (TEM) (Figure 1B) (Schreiber, 2010). These structures are referred to as suberin lamellae.

1.2: Role of suberin in water transport

There are three different pathways for radial water transport within plant tissue: the apoplastic pathway, the symplastic pathway, and the transcellular pathway (Barberon and Geldner, 2014). Each of these pathways describes a specific mode by which water is transported at the cellular level across a tissue. The apoplastic pathway involves the transport of water within the apoplast, the collective term for the cell walls and the intercellular spaces. The symplastic pathway describes the cell-to-cell transport of water through inter-connected plasmodesmata. Finally, the transcellular pathway describes water transport across cell plasma membranes via influx and efflux transporters.

In plant roots, cellular barriers, for example, suberin, can contribute to the regulation of non-specific water transport. In roots, the water and solutes that are absorbed at root tips are transported radially into the stele (Nawrath *et al*., 2013). As they are conducted longitudinally to the xylem, the hydrophobic nature of suberin deposited in

the cell walls of the surrounding endodermal cells serves as a bi-directional transcellular barrier to the transport of water and solutes between the endodermal cells and the apoplastic space (Franke *et al.*, 2005; Barberon, 2017).

Suberin also functions as a protective barrier in both aerial and root tissues. The accumulation of suberin in the secondary tissues of mature stem periderms protects the plant from desiccation (Vishwanath *et al*., 2015). As plant roots mature, the deposition of suberin on the outer surface of the root protects the plant from up-taking toxic solutes from the soil, as well as from soil-borne pathogens (Nawrath *et al*., 2013). Additionally, the wound-induced deposition of suberin in aerial tissues or tubers protects against opportunistic pathogens (Ranathunge *et al*., 2011).

1.3: Coordinated developmental processes produce suberized cells

In both the root and shoot tissues, suberization of cells is regulated spatially and temporally. For example, in endodermal root tissues, there are multiple cell wall modification events that lead to the formation of suberized cells (Doblas *et al*., 2017). Endodermal cells are first derived from periclinal divisions of the endodermal-cortical initial cells near the quiescent center of root meristematic tissue (Dolan *et al*., 1993). As these cells continue to mature, Casparian bands, defined as localized rings of lignin or a lignin-like polymer, are deposited to block the apoplastic junctions between adjacent endodermal cells (Naseer *et al*., 2012). The Casparian bands are deposited to limit the uncontrolled transport of water and solutes via the apoplastic pathway, known as Stage I endodermal tissue development (Doblas *et al*., 2017). In Stage II, the endodermal cells continue to differentiate and features the deposition of suberin, typically forming lamellae. Suberin covers the whole cell until the tissue is suberized to form the

transcellular barrier with the functions described above. In some plants, the development of the endodermis progresses further into Stage III, resulting in the formation of a lignified or suberized tertiary cell wall inside the suberin lamellae (Peterson and Enstone, 1996).

The progression of endodermal cell differentiation is spatially separated along the length of the root. The younger cells near the root tip are mostly newly differentiated endodermal cells without cell wall modifications (i.e. no Casparian bands) or in Stage I, while the older tissues distal to the root tip will be mostly entering Stage II (Doblas *et al*., 2017). However, such development is not always uniform within tissues of the same age. In Stage I tissues, areas where the Casparian strip is disrupted (Doblas *et al*., 2017), become sites for lateral root formation that originate from the pericycle – the next layer of cells inside the endodermis (Gibbs *et al*., 2014). In Stage II tissues undergoing suberization, some cells remain un-suberized long after the other cells have completed stage II. With only Casparian bands present as cell wall modifications, they impart some permeability of the mature endodermis to water and solutes. These cells are called endodermal passage cells (Peterson and Enstone, 1996), and while they do not appear to follow any pattern along the longitudinal axis of the root, they may exist as clusters of un-suberized cells (Andersen *et al*., 2018). However, with time, as the tissue enters Stage III differentiation, even these endodermal passage cells become suberized (Peterson and Enstone, 1996). Recent research suggests that phytohormones auxin and cytokinin signaling participates in suppressing the formation of such passage cells (Andersen *et al*., 2018).

1.4: Structural models of suberin

One model for the overall structure and organization of suberin has two distinct domains (Figure 3) (Bernards, 2002). The first domain, sometimes called polyaliphatic suberin, is comprised of various long-chain (C16 and C18) and very-long-chain (≥C20) fatty acid derivatives that are linked with each other and glycerol. Examples of the types of compounds found in suberin include unsubstituted fatty acids, α,ω – dicarboxylic acids, ω – hydroxy fatty acids, primary fatty alcohols, and α,ω – diols (Vishwanath *et al*., 2015) (Figure 2). These fatty acid derivatives may be esterified to glycerol (as monoacylglyceryl esters) and ferulate (e.g. as ω – hydroxy alkyl ferulates) (Figure 3). The polyphenolic domain of suberin is formed by various *n*-alkylated aromatic compounds that are shuttled from the phenylpropanoid pathway, including ferulate, coumarate and caffeate (Vishwanath *et al*., 2015). This domain is similar to lignin but contains more of the acidic forms of the phenolics than typical lignin, which is mainly comprised of the alcohol forms of the phenolics (monolignols).

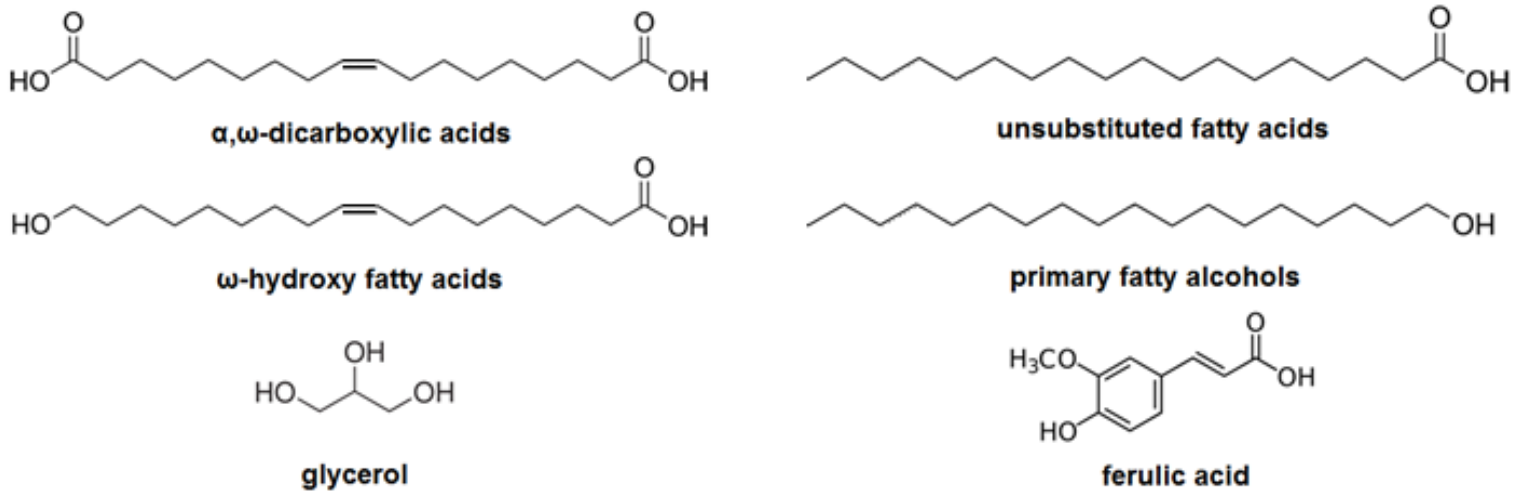


Figure 2: Chemical structures of common C18 aliphatic suberin monomers generated in the endoplasmic reticulum using plastid-borne fatty acids. Aliphatic suberin is enriched in C18 mono-unsaturated (Δ9) α,ω-dicarboxylic and ω-hydroxy fatty acids, as shown. These compounds can be generalized for the C16 – C26 monomers that are observed in suberin. Primary fatty alcohols, glycerol and ferulate are the three other important components found in aliphatic suberin.

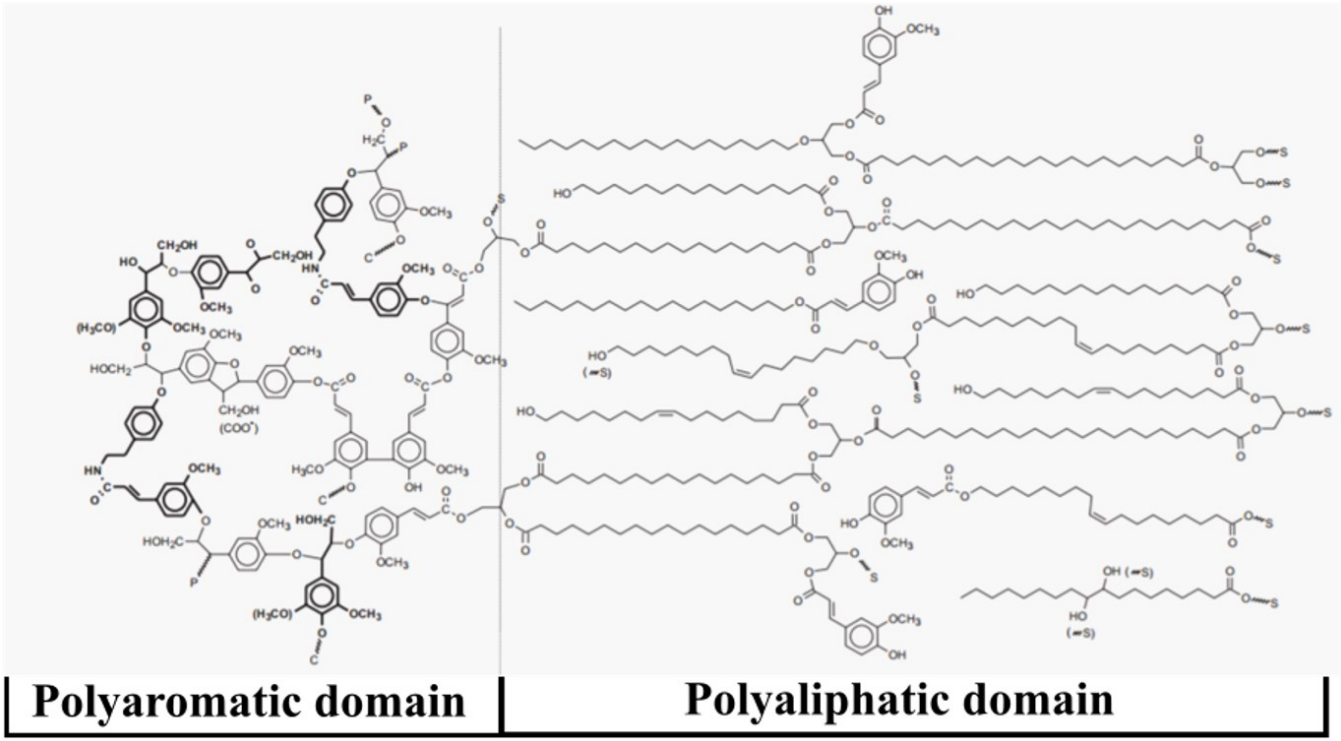


Figure 3: **Proposed model for the structure of suberin in cell walls (adapted from Bernards, 2002).** The polyphenolic domain is enriched in *n*-alkylated aromatic compounds, while the aliphatic domain is made up of mostly long-chain and very-long-chain fatty acid derivatives linked with glycerol and ferulate.

In another model for the structure of the suberin polymer, suberin is defined as the polyaliphatic domain described by the Bernards model (Figure 3 and 4) and classifies the polyaromatic domain separately as suberin-associated lignin (Graça, 2015). This lignin was found to be mostly guiacyl (G-type) in the tree bark of *Quercus suber*, *Q. cerris*, and *Betula pendula*; while in the potato periderm of *Solanum tuberosum*, the suberin-associated lignin contained both S- (syringyl) and G-types.

According to Graça *et al.* (2015), the polyaromatic domain as described in Bernards (2002) should be classified as lignin due to the distinct spatial, chemical, and structural characteristics of the aromatic components as compared to the polyaliphatic components, with ferulic acid derivatives forming the interface between these two distinct polymers. Furthermore, restricting the definition of suberin to the polyaliphatic domain is convenient for analytical methods, since it is difficult to co-extract the polyaromatic and polyaliphatic components due to their distinct chemical differences.

These objections to the Bernard suberin model proposed by Graça *et al.* are also supported in a review of apoplastic barriers by Geldner (2013), where it is argued that the polyaromatic domain proposed by Bernard's model is lignin-like and is therefore its own substance. Therefore, by restricting the definition of suberin to include only the polyaliphatic components is more specific and leads to less confusion and ambiguity in the literature. However, the polyaliphatic and polyaromatic components are likely to be functionally connected.

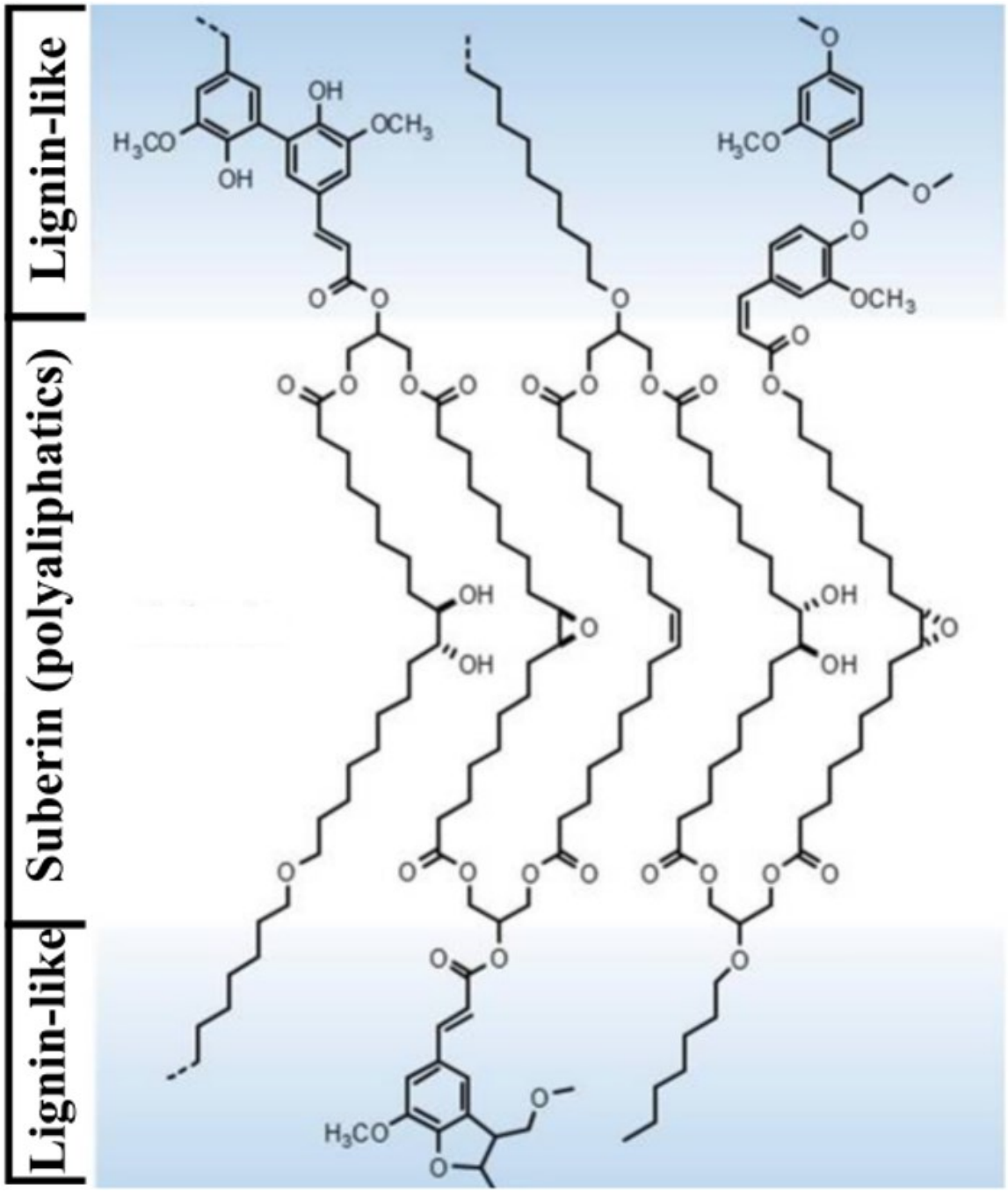


Figure 4: Proposed model for the structure of suberin in cell walls (adapted from Graça, 2015). The polyphenolic domain is enriched in *n*-alkylated aromatic compounds and esterified with ferulate. The suberin polymer is often found to be closely associated and covalently linked to lignin.

1.5: Suberin biosynthesis involves several distinct pathways

The biosynthesis of suberin begins in the plastids of suberin-depositing cells, where 16:0, 18:0, and 18:1 fatty acid precursors are synthesized *de novo* and exported to the endoplasmic reticulum (ER) where they are converted to Coenzyme A activated fatty acyl thioesters. (Vishwanath *et al.*, 2015). At the ER they have three possible fates for suberin monomer biosynthesis: (1) elongation by the fatty acid elongase (FAE) complex, (2) enter an oxidative pathway, or (3) enter a reductive pathway. The acyl lipids can also enter the oxidative or reductive pathways after fatty acid elongation.

In elongation by the FAE complex (Figure 5), the activated C16 and C18 fatty acyl-CoA thioesters are elongated into ≥C20 substrates in a 4-step enzymatic conversion (Lee *et al.*, 2009; Joubès *et al*., 2008). The fatty acyl-CoA thioesters are elongated by two carbon units per reaction cycle using malonyl-CoA, which is synthesized by acyl-CoA carboxylase (ACC) (Samuels *et al*., 2008). The initial acyl-CoA thioester is condensed with malonyl-CoA to produce a 3-ketoacyl-CoA intermediate by a β-ketoacyl-CoA synthase (KCS) (Millar and Kunst, 1997). This intermediate is then reduced and dehydrated by 3-ketoacyl-CoA reductase (KCR) and 3-hydroxacyl-CoA dehydratase (HCD), respectively (Joubès *et al*., 2008). The fourth and final step of elongation is another reduction reaction by *trans*-2,3-enoyl-CoA reductase (ECR). This cycle can be repeated, elongating the acyl chain by two carbons each time (e.g. 18:0 -> 20:0 -> 22:0 -> 24:0) .

From the 21 members of the β-ketoacyl-CoA synthases (*KCS*) gene family in Arabidopsis (Joubès *et al*., 2008), *KCS2* and *KCS20* are functionally redundant genes that participate in the elongation of the fatty acyl-CoA thioesters in suberin biosynthesis

(Franke *et al.*, 2009; Lee *et al.*, 2009). KCS2 demonstrates specificity for the elongation of 18:1-CoA, 18:2-CoA, and 20:0-CoA substrates (Lee *et al.*, 2009; Paul *et al.*, 2006), while KCS20 demonstrates substrate specificity for 16:0-CoA, 16:1-CoA, 18:0-CoA, and 20:0-CoA chain lengths (Lee *et al.*, 2009; Paul *et al.*, 2006). In addition, KCS9 was also reported to be specifically involved in the elongation of C22 fatty acyl-CoA thioesters into their C24 counterparts (Kim *et al*., 2013).

In the oxidative pathway (Figure 5), some of the fatty acyl-CoA pool is directed to the formation of ω – hydroxy fatty acids, mediated by the cytochrome P450 enzymes *CYP86A1* (specific for C16 and C18 acyl chains) (Höfer *et al.*, 2008) and *CYP86B1* (specific for >C18 elongated chains) from *Arabidopsis* (Compagnon *et al*., 2009; Molina *et al.*, 2009). To produce α,ω – dicarboxylic fatty acids, some ω – hydroxy fatty acids are thought to be further oxidized by the same CYP86A1 or CYP86B1 enzymes, although this has not been proven (Li-Beisson *et al*., 2013). To form monoacylglyceryl esters, either ω – hydroxy fatty acyl chains or α,ω – dicarboxylic acyl chains are esterified to the *sn-2* position of glycerol-3-phosphate (G3P) via the catalytic action of glycerol 3-phosphate acyltransferases (GPAT5 and GPAT7 in the case of suberin in Arabidopsis) (Beisson et al., 2007; Yang *et al.*, 2012).

In the reductive pathway (Figure 5), some of the fatty acyl-CoA pool is used by the cell to produce 18:0, 20:0, and 22:0 primary fatty alcohols via catalysis by FAR5, FAR4, and FAR1, respectively (Domergue *et al.*, 2010; Vishwanath *et al.*, 2013). It is possible that a fatty acid reductase (FAR) enzyme catalyzes the reaction converting some ω – hydroxy fatty acyls into α,ω – diols, but this has not been proven (Vishwanath *et al.*, 2015).

In the production of suberin-associated waxes, a large proportion (80%) of the primary fatty alcohols produced by the FAR enzymes are combined with ferulates, coumarates, and caffeates to form *n*-alkylated aromatic compounds that are deposited in the cell wall, while the rest of the primary fatty alcohols are polymerized with other suberin aliphatics (Vishwanath *et al.*, 2013; Delude *et al.*, 2016). The production of n- or ω-OH alkylated ferulates is mediated by ALIPHATIC SUBERIN FERULOYL TRANSFERASE (ASFT*)* (Figure 5), which is classed as a BAHD-type O-acyltransferase (Molina *et al*., 2009). *ASFT* knock-out mutants exhibit a 90% reduction in ferulate monomers measured in root suberin extracts. Similarly, the production of n-alkyl caffeates is mediated by *FATTY ALCOHOL:CAFFEOYL-CoA CAFFEOYL TRANSFERASE* (*FACT*), another member of the BAHD-type O-acyltransferases (Kosma *et al.*, 2012). Knock-out mutants for *FACT* were reported to have a nearly complete lack of alkyl caffeate esters in root wax extracts.

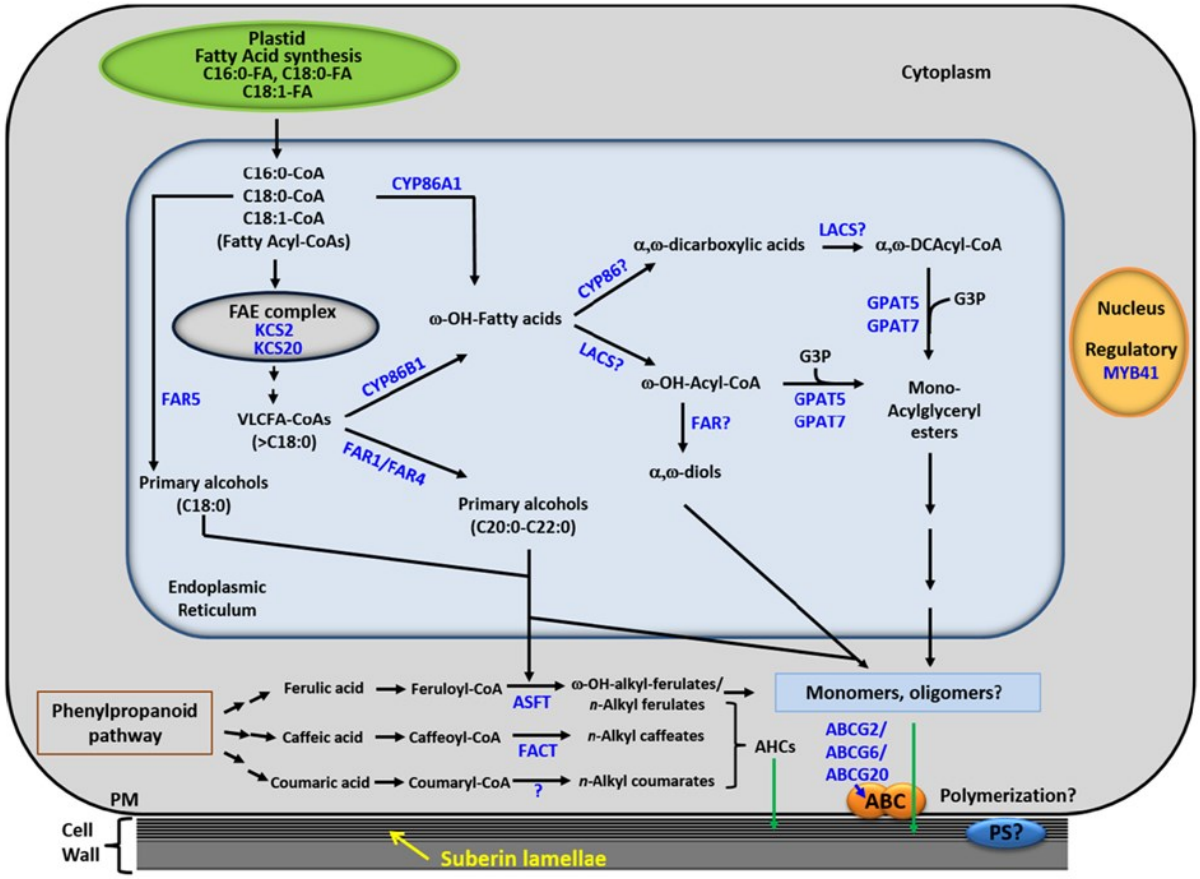


Figure 5: Model of suberin biosynthesis in plant cells from plastid fatty acid synthesis to suberin polymerization of suberin monomers. Suberin aliphatic compounds are produced in the endoplasmic reticulum, while suberin phenolic compounds are produced in the cytoplasm. The monomers are polymerized, including with glycerol, and deposited together to form suberin lamellae within the cell wall. Abbreviations: fatty acid elongation complex (FAE complex), fatty acid reductase (FAR), β-ketoacyl-CoA synthases (KCS), cytochrome P450 (CYP), long-chain acyl-CoA synthetase (LACS), glycerol 3-phosphate acyltransferases (GPAT), Aliphatic suberin feruloyl transferase (ASFT), Fatty alcohol: caffeoyl-CoA transferase (FACT), ATP-binding cassette transporters (ABC), polyester synthase (PS). (Vishwanath *et al*., 2015)

Collectively, the end products of the suberin monomer biosynthetic pathways are exported from the endoplasmic reticulum to the plasma membrane (Vishwanath *et al.*, 2015) (Figure 5). This is possibly mediated by secretory vesicles via the Golgi apparatus. In cutin literature, there is evidence supporting this secretory mechanism of cutin biosynthesis (McFarlane *et al*., 2014), however the same has not been tested yet for suberin biosynthesis. Concerning the mechanism by which these products are exported to the cell wall, initial studies show that several members of the G-subfamily of ATP-binding cassette transporters (ABCG2, ABCG6, and ABCG20 in Arabidopsis) are probably involved (Yadav *et al*., 2014), although the details for this transport have not yet been revealed. Furthermore, the enzymes that are involved in polymerizing these monomers to form suberin have not yet been identified (Vishwanath *et al.*, 2015), and this is a major gap in our knowledge in suberin biosynthesis.

1.6: Select R2R3-type MYB transcription factors regulate suberin biosynthesis

While there has been much progress in the past 10 years in identifying the genes encoding the key suberin biosynthetic enzymes, not much is known about how their expression is coordinately regulated. In plants, there exists a large family of MYELOBLASTOSIS (MYB)-type transcription factors that regulate a wide array of plant development and metabolic processes (Dubos *et al.,* 2010). The members of this protein family are named after their observed homology to the avian viral oncogene dubbed *v*-MYB, which originated from the ubiquitous vertebrate gene *c*-MYB (Ambawat, 2013). There are 196 predicted MYB-type transcription factors in Arabidopsis that are involved in both primary and secondary metabolism (Stracke *et al*., 2001). For example, these transcription factors regulate genes involved in anthocyanin

biosynthesis, secondary growth, cell fate and identity, development, and stress responses (Dubos *et al.*, 2010; Kosma *et al.*, 2014).

The DNA-binding "MYB domains" of these transcription factors have conserved basic helix-loop-helix (bHLH) motifs with regularly spaced tryptophan residues, forming a cluster in three-dimensional space, characteristic to this type of transcription factor (Dubos *et al.*, 2010; Stracke *et al.*, 2001). In plants, the MYB protein family is organized into subgroups based on the number and types of the three different MYB domain repeats, identified as conserved domains R1, R2, and R3 from *c*-MYB. The R2R3 subfamily is the largest of the four classes of MYB transcription factors in plants (126 members in Arabidopsis), featuring single repeats of the R2 and R3 MYB domains forming the N-terminal DNA-binding domain. (Stracke *et al*., 2001). Based on sequence similarity of the remaining C-terminal portions of the proteins, the R2R3 subfamily is further categorized into 22 unique sub-families (Figure 7).

Recently, the R2R3-type transcription factor *AtMYB41* (Figure 5; Figure 6) was identified by the Rowland lab and collaborators to be associated with suberin production (Kosma *et al.*, 2014). When *AtMYB41* was overexpressed in either *A. thaliana* or *N. benthamiana* lamellar structures were found, similar to those found in suberin, on the inner face of the primary cell walls of leaf epidermal cells (Kosma *et al.*, 2014). With further experimentation, *AtMYB41* was found to regulate suberin biosynthesis in response to signaling by abscisic acid (ABA), a major phytohormone that coordinates plant responses to salt and drought stresses (Kosma *et al.*, 2014). Using transgenic *A. thaliana* lines expressing the GUS reporter gene under the control of the *AtMYB41* promoter, *AtMYB41* expression was only observed in 5 day-old roots that were treated with 10 μM

ABA or 200 mM NaCl for 24 hours in the endodermis and surrounding cortical tissues. No such expression was observed on untreated roots of the same age. Upregulation of *AtMYB41* by ABA in plant roots led to the activation of many suberin biosynthetic genes, *CYP86A1*, *CYP86B1*, *GPAT5*, *FAR1, FAR5, FAR5*, and *ASFT*, presumably leading to increased accumulation of suberin in those tissues (Kosma *et al.*, 2014; Barberon *et al.*, 2016).

A pair of Arabidopsis MYB transcription factors were found to be linked to suberin biosynthesis in seed coats, AtMYB9 and AtMYB107 (Lashbrooke *et al.*, 2016) (Figure 6). Reductions in suberin were observed in *myb9* and *myb107* mutants, accompanied by increased seed permeability in both mutants, and abnormal morphology of the seed coat layers in *myb107* (Gou *et al.*, 2017; Lashbrooke *et al.*, 2016). Notable suberin-related genes that are downregulated in *myb107* include *GPAT5* and *ASFT*, while *CYP86A1* was found to be downregulated in *myb9* (Lashbrooke *et al.*, 2016).

Three other Arabidopsis R2R3-type MYB transcription factors have been shown to regulate suberin biosynthesis in roots during normal development (unstressed conditions), AtMYB53, AtMYB92, and AtMYB93 (Rowland Lab, unpublished) (Figures 6-8). These three transcription factors are closely related to one another and are grouped together under the S24 subclade of R2R3 *MYB* proteins (Dubos *et al*., 2010) (Figure 6). Early work in characterizing *AtMYB93* indicated that this transcription factor also inhibits lateral root formation (Gibbs *et al*., 2014). It is co-expressed with other important suberin-related genes in root endodermal cells, such as *FAR4*, *CYP86A1*, *CYP86B1*, and *GPAT5* (Nemanja Mladevic, Rowland Lab, unpublished). The relationship between *AtMYB93* and suberin biosynthesis was confirmed when mutant lines for this gene

showed a reduction in total aliphatic suberin, while suberin deposition was induced in a transient over-expression assay (Fakhria M. Razeq and Jhadeswar Murmu, Rowland lab, unpublished data).

When both *AtMYB92* and *AtMYB93* genes are deleted, there is a 50% reduction in total aliphatic suberin in 4-week old roots relative to wild-type seedlings (Jhadeswar Murmu, Rowland lab, unpublished). This would then suggest that *AtMYB53*, which is closely related to both genes and expressed in root endodermis, is also involved in regulating suberin deposition in a partially redundant manner. Indeed, as part of my Undergraduate Honour's Thesis at Carleton University, using transgenic promoter:GUS lines, I was able to demonstrate that all three genes are co-expressed together in root endodermal cells (Figure 7), where suberin is known to be deposited. Furthermore, I also showed that two independent alleles of *myb53* knockout mutants have a reduction of suberin in 2-week old roots compared to wild-type (Figure 8).

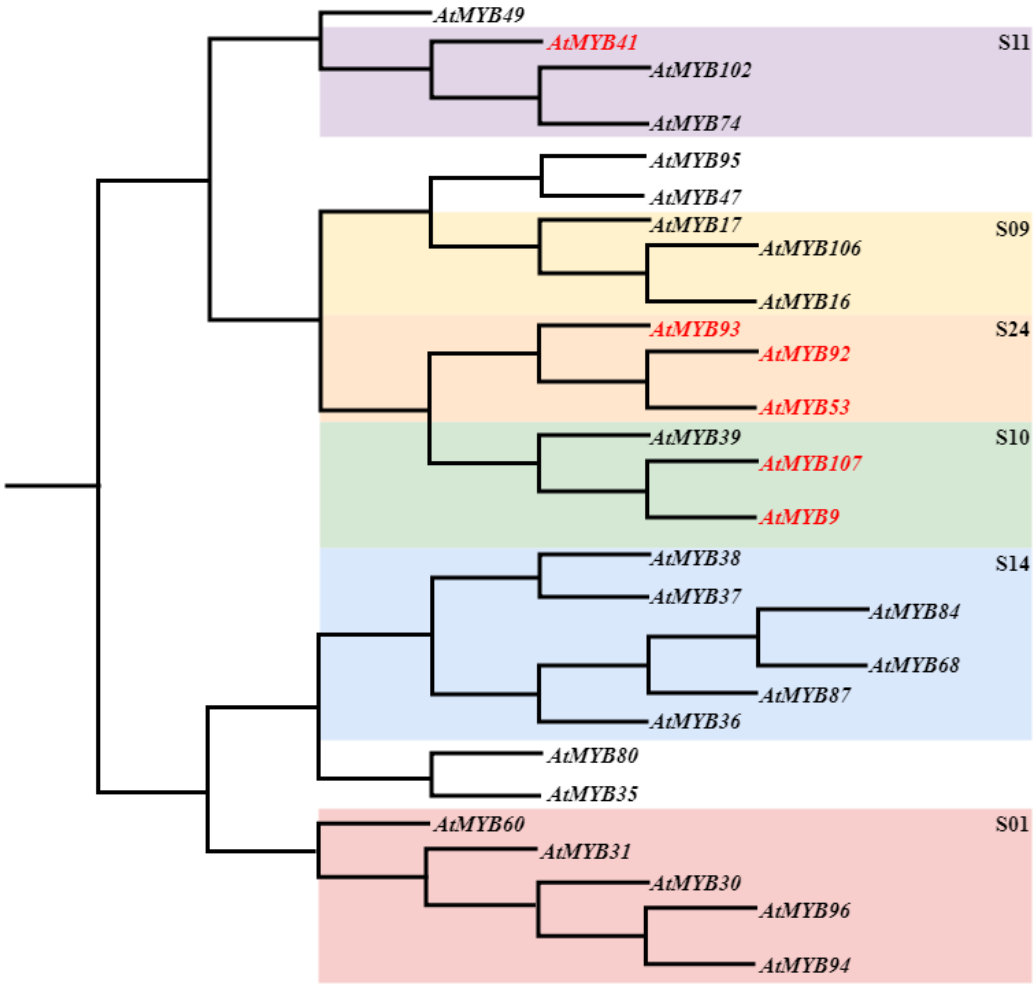


Figure 6: Portion of a neighbour-joining tree for all 126 members of the R2R3 family of MYB transcription factors, indicating their respective groupings into one of the 22 subfamilies. Shown are subfamilies 11, 9, 24, 10, 14, and 1. The tree highlights members that have been documented in the literature to be regulating suberin biosynthesis in red, illustrating their close relationships within clades and closely related clades. Adapted from (Stracke *et al.*, 2001)

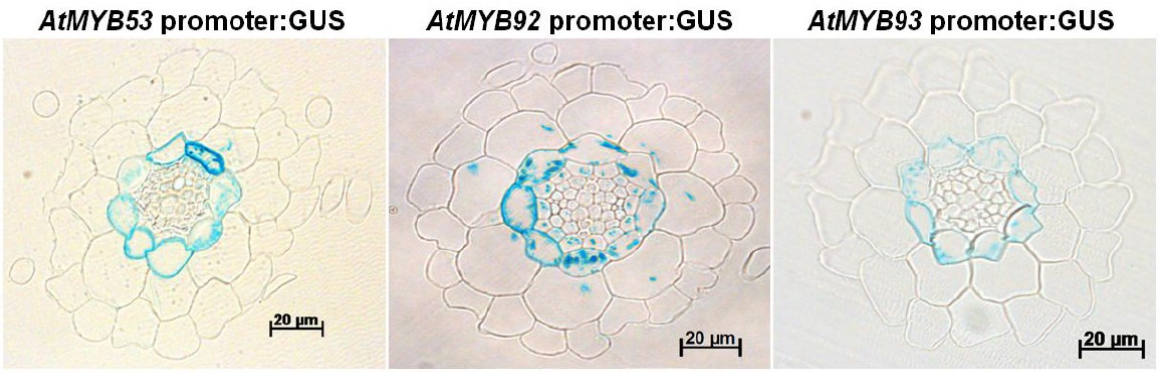


Figure 7: . Cross-sections (3 µm thick) of 6-day old roots of *Arabidopsis thaliana* indicating the expression patterns of promoter *AtMYB53:GUS, AtMYB92:GUS,* and *AtMYB93:GUS* reporter genes. Sections were embedded in LR White Medium Grade resin, mounted in 20% (v/v) glycerol, using the promoter:GUS reporter gene assay. *AtMYB53, AtMYB92,* and *AtMYB93* are all co-expressed in the root endodermal cells. *AtMYB53* and *AtMYB93* are both endodermal-specific, while *AtMYB92* is also expressed in the pericycle and possibly cortical cell layer. Image acquired using AxioVision 4.8 software (bright-field light microscopy) (Daniel Klein, undergraduate research thesis, and Jhadeswar Murmu, Rowland lab, unpublished).

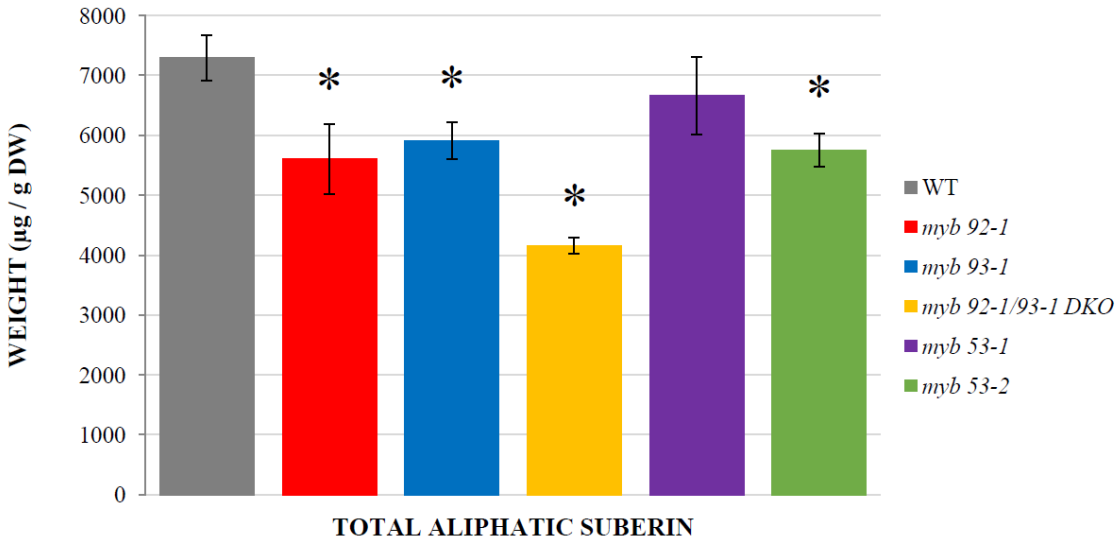


Figure 8: The amount of total aliphatic suberin per gram of dry weight for Arabidopsis *myb53-2*, *myb92-1*, and *myb93-1*, and *myb92-1 myb93-1* knockout lines as detected by GC-FID. The sample size was n=4 for all lines, with the exception of *myb53-2* (n=2). Asterisk (*) indicates significant differences with respect to wild type at the $\alpha = 0.05$ confidence interval (one-tailed Student t-test, Data Analysis Package, Excel 2007) (Daniel Klein, undergraduate research thesis, Rowland lab, unpublished data).

1.7: Project rationale and objectives

There are several unanswered key questions about suberin biosynthesis and its regulation that may be addressed by studying these MYB-type transcription factors. First, the molecular mechanisms by which the suberin biosynthetic genes are regulated are unclear (Vishwanath *et al*., 2015). We hypothesize that *AtMYB53*, *AtMYB92*, and *AtMYB93* transcription factors directly control the transcriptional regulation of the suberin biosynthetic genes by binding their promoters (Figure 9). Alternatively, there could be a set of intermediate regulatory proteins targeted by these MYB's, which in turn directly control the expression of the suberin biosynthetic genes. Both cases could explain the observed link between *AtMYB53*, *AtMYB92*, and *AtMYB93* and the suberin biosynthetic genes and the observed mutant suberin phenotypes. As *AtMYB53*, *AtMYB92*, and *AtMYB93* are involved in suberin deposition during normal root development (Rowland Lab, unpublished data), there are still questions about what role they play in maturing roots to coordinate the suberization of: the endodermal tissues, between Stages I, II, and III of endodermal differentiation, and exodermal tissues, where applicable.

Another area of study in suberin research is to identify the other remaining genes/proteins involved in suberin monomer synthesis, transport, and polymerization (Figure 9). Specifically, while a number of key enzymes have been characterized to be involved in suberin biosynthesis, as described above, there are some intermediate catalytic steps that have no assigned enzyme. For example, the involvement of LACS-type enzymes is speculated to process oxidized suberin monomers before esterification with glycerol based on analogous findings in cutin research (Schnurr *et al*., 2004), however, no such LACS enzymes have been identified for suberin biosynthesis.

Intracellular trafficking of suberin monomers from the ER to the plasma membrane is uncharacterized to date. The ABCG2, ABCG6, and ABCG20 transporters have been proposed to mediate transport across the plasma membrane (Yadav *et al*., 2014). However, there are likely more transporters related to suberin deposition that have yet to be discovered, such as members of the lipid transfer protein (LTP) family located in the apoplasm (Vishwanath *et al*., 2015; Salminen *et al*., 2016). There are 79 LTP members in Arabidopsis and homologs are found in all land plants. (Salminen *et* al., 2016). A suberin polyester synthase is proposed to mediate the polymerization of the suberin monomers at the cell wall, but the identity of the synthase and mechanism by which polymerization occurs is unknown. However, through the analogous catalytic steps identified in cutin research (Yeats *et* al., 2012; Kurdyukov *et al*., 2006; Nawrath, 2002), the suberin polymer assembly is proposed to involve catalytic enzymes belonging to GDSL-motif lipase/acyltransferase, α/β lipase/hydrolase and peroxidase families, but no such enzymes have been identified to be involved in suberin biosynthesis to date. With so many unidentified suberin-related genes, exploiting the properties of mutant lines for *AtMYB53*, *AtMYB92*, and *AtMYB93* may provide the necessary tools for gene discovery to supplement our current knowledge of suberin biosynthesis and for understanding the mechanistic roles that these *MYB* transcription factors play in suberin biosynthesis (Figure 10).

To address the above research questions, a combination of mutant knockout lines and over-expression lines have been developed by a previous Rowland lab graduate student, Hefeng Hu. Knowing that AtMYB53, AtMYB92, and AtMYB93 are all regulators of suberin biosynthesis in normal root development, Hu generated two

independent T-DNA alleles of triple knockout (TKO) lines of *myb53*, *myb92*, and *myb93*, where it was reported that each TKO line (*myb53-1 myb92-1 myb93-1* and *myb53-2 myb92-1 myb93-1*) demonstrated a reduction of suberin by about 72% (Hu, 2018).

Hu also developed a steroid-inducible over-expression (OE) TRANSPLANTA line of *MYB53* (Coego *et al*., 2014; Hu, 2018). This homozygous transgenic line contains a constitutively expressed chimeric VP16-type transactivator fused to the bacterial repressor LexA DNA-binding domain and human estrogen receptor (hER) regulatory domain (Zuo *et al*., 2000). Upon hER binding to an applied exogenous dose of β-estradiol, the transactivator complex binds in the nucleus to an eight-repeat LexA operator upstream from the *MYB53* coding region to drive the chemical-inducible overexpression of MYB53 transcripts. Characterization of this *AtMYB53*-OE line (Hu, 2018) showed that accumulation of large amounts of root suberin occurs 4 days after β-estradiol treatment compared to negative controls (Figure 10). To answer the above research questions, the objective of this thesis was to make use of these different suberin triple *myb53 myb92 myb93* mutant and the *AtMYB53* overexpression lines by using RNA-seq analysis on the Illumina platform. First, transcriptomic profiles obtained from the *myb53 myb92 myb93* triple knock-out lines were compared to wild-type (WT) plants to generate a list of candidate genes contributing to the observed mutant suberin phenotype. Second, using the β-estradiol inducible *AtMYB53* TRANSPLANTA over-expression line, RNA-seq analysis was performed to generate a list of differentially expressed genes at two time-points to identify the early and late responsive genes of *AtMYB53*. By combining the results of these two experiments, new hypotheses about the overall regulation and production of suberin were generated from these datasets to inspire new

research initiatives in this field.

The choice of using *AtMYB53* in an over-expression line over *AtMYB92* or *AtMYB93*, or a combination of all three, is the result of several factors. First, the over-expression line for *AtMYB53* had already been characterized (Hu, 2018), while the other two were not. Second, the *AtMYB92* over-expression line was available, but could not be successfully validated (Hefeng Hu and Owen Rowland, personal communication). Third, including all three overexpression lines would not have been feasible for the scope and timeline of this project. However, the choice of using only *AtMYB53* for over-expression is justifiable since the *myb53 myb92 myb93* triple knock-out dataset was used to complement the over-expression dataset in the RNA-seq experiment.

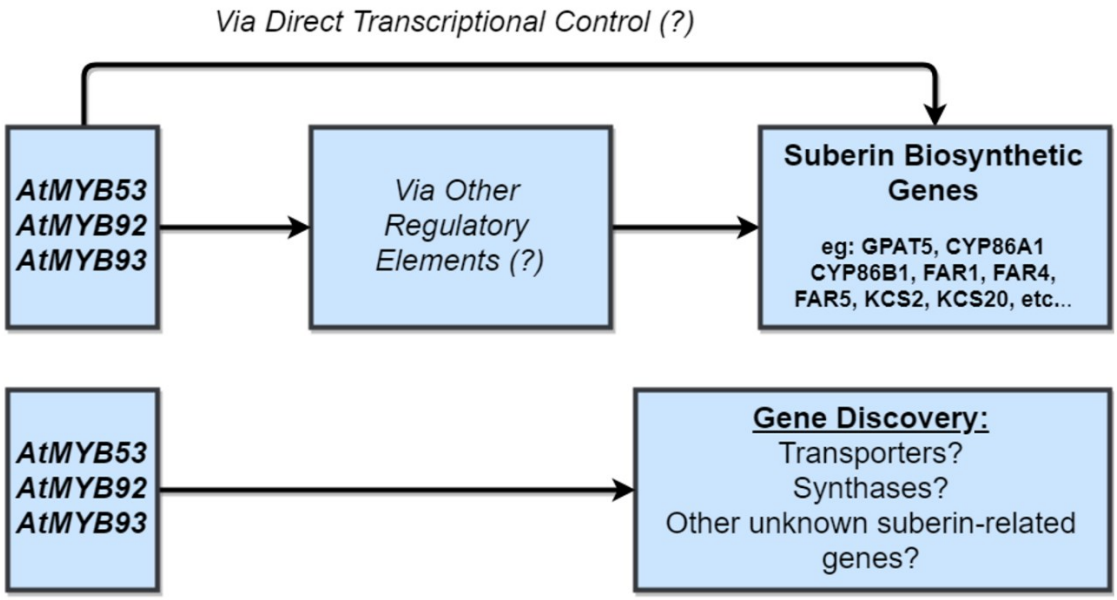


Figure 9: Summary of the thesis rationale and objectives. The transcription factors *AtMYB53*, *AtMYB92*, and *AtMYB93* are all known to play a significant role in the suberization of endodermal tissues during normal root development (Rowland lab, unpublished data), but the question remains whether the suberin biosynthesis genes are their direct transcriptional targets, or if they regulate other intermediate factors. Furthermore, significant gaps remain in our knowledge of suberin production, intracellular trafficking, secretion and assembly.

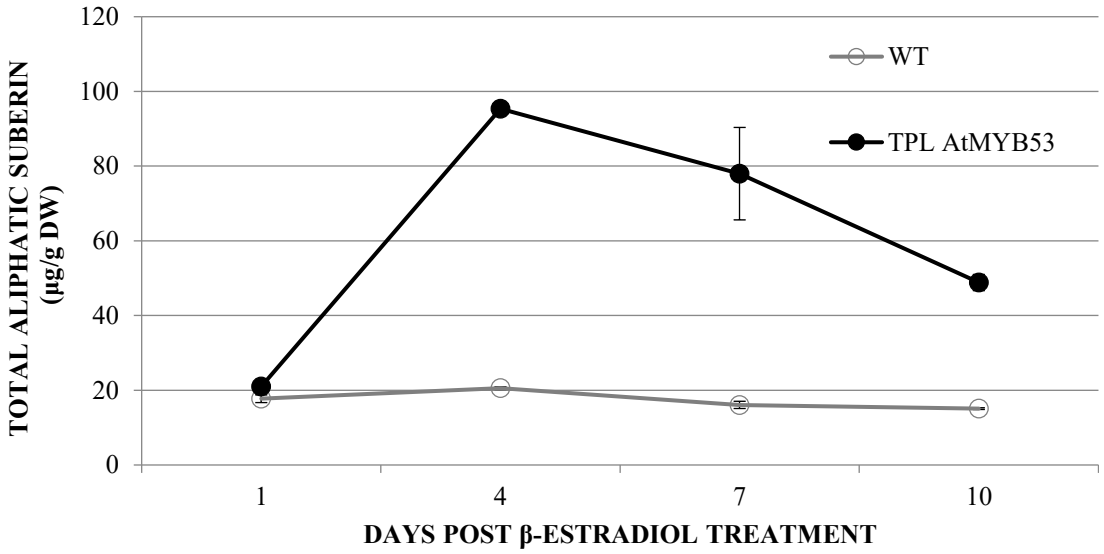


Figure 10: A time-course experiment using the *AtMYB53* TRANSPLANTA over-expression line (black) with β-estradiol compared to WT seedlings treated the same way (grey). The amount of total aliphatic suberin per gram of dry weight is reported, as quantified by GC-FID, showing maximum suberin induction in roots 4 days post-treatment. The sample size was n=4 for each treatment. Data provided courtesy of Hefeng (Jasmine) Hu and Dr. Owen Rowland (Hu, 2018).

.

CHAPTER 2 – MATERIALS AND METHODS

2.1: Seed sterilization procedure

Seeds of *Nicotiana benthamiana* and *Arabidopsis thaliana* were surface sterilized with 70% (v/v) ethanol for 2 minutes, followed by incubation in a bleach solution (2.14% sodium hypochlorite, 0.1% (v/v) Tween 20) for 5 minutes. Following sterilization, the seeds were rinsed 4-5 times with distilled sterile water until all traces of the bleach solution were eliminated. The seeds were then suspended in a 0.1% (w/v) agar solution.

2.2: Growth conditions of *Nicotiana benthamiana* plants for transient over-expression assay

Wild-type *N. benthamiana* seeds were sterilized as described above. The suspended seeds were sown directly onto autoclaved soil (Pro-mix BX-General Purpose Growing Medium) and allowed to germinate over 7 days under a plastic cover in an environmental growth chamber (Econair, SP-56) set at 22°C and with a light intensity of 150-170 $\mu mol \cdot m^2 \cdot s^{-1}$ The soil was supplemented with fertilizer solution (0.2% (w/v) 20-20-20 macronutrient mix [Plant-Prod], 0.0125% (w/v) chelated micronutrient mix (Plant-Prod). The plants were then allowed to grow uncovered for an additional 7 weeks under the same growth conditions and fertilized twice a week with the same nutrient mix.

2.2: Growth of T-DNA knockout (KO) mutants for RNA extraction and lipid polyester analysis

Seeds of *A. thaliana* wild-type (Col-0) and triple *myb53 myb92 myb93* mutant lines (*myb53-1 myb92-1 myb93-1* (TKO1) and *myb53-2 myb92-1 myb93-1* (TKO2)) were previously obtained and genotyped by Hefeng Hu (Hu, 2018). The seeds were sterilized as described above, stratified in the dark at 4°C for 3 days, and suspended in a 0.1%

(w/v) agar solution. The seeds were then sown onto AT media (0.6% (w/v) agar, 2 mM $Ca(NO_3)_2$, 2 mM $MgSO_4$, 5 mM KNO_3, 2.5 mM KH_2PO_4, 1 mM micronutrients, and 50 μM Fe(EDTA)) in petri dishes (10 mm x 150 mm) at a density of 40 seeds per plate and then placed in an environmental growth chamber (Econair, SP-56) set at 22°C and a light intensity of 150-170 $\mu mol \cdot m^2 \cdot s^{-1}$ for 14 days. Roots pooled from a single plate were harvested as one bio-replicate, discarding all leaf tissue, flash frozen in liquid nitrogen, and stored at -80°C until RNA extraction (described below). For analysis of lipid polyester content, roots harvested from 6 plates were pooled and transferred to $CHCl_3$-rinsed glass GC extraction tubes as one bio-replicate, and then proceeding directly to the delipidation and depolymerization protocol described below.

2.3: Growth of *A. thaliana MYB53* steroid-inducible over-expression lines for RNA extraction

Seeds of wild-type (Col-0) and *AtMYB53* steroid-inducible overexpression line (TRANSPLANTA-559) were sterilized as described above, stratified in the dark at 4°C for 3 days, and suspended in a 0.1% (w/v) agar solution. The seeds were then sown onto AT media in 100 mm x 15 mm square petri plates on strips of sterilized 20 μm filter paper at a density of 20 seeds per strip/plate. The plates were then placed in vertical orientation in an environmental growth chamber (Econair, SP-56) set at 22°C and a light intensity of 150-170 $\mu mol \cdot m^2 \cdot s^{-1}$ for 10 days. The seedlings were then divided into two groups and exposed to two treatments. The first treatment involved the transfer of one group of seedlings to AT media supplemented with 20 μM β-estradiol (E–8875, Sigma-Aldrich) solubilized in DMSO (4100-1-05, Caledon Laboratory Chemicals) at a final concentration of 0.2% (v/v). The second treatment, serving as the control group, was transferred to AT plates with only DMSO at a concentration of 0.2% (v/v). Over two

separate time course experiments, roots were harvested at 2 h, 6 h, 15 h, and 24 h post-treatment (first time-course) or at 6 h, 8 h, 10 h, 12 h, and 15 h post-treatment (second time-course). Roots were pooled from two plates (40 seedlings) as a single bio-replicate, flash frozen in liquid nitrogen, and stored at -80°C until RNA extraction.

2.4: Extraction and purification of total RNA from *A. thaliana* roots

Frozen roots were transferred to a sterilized and liquid-nitrogen cooled mortar, and then thoroughly ground into a fine powder using a pre-cooled pestle in liquid nitrogen. The powder was transferred to a sterile and nuclease-free microcentrifuge tube, and total RNA was prepared using the NORGEN Biotek Plant/Fungi Total RNA Extraction Kit (Cat# 25800) according to the manufacturer's protocol. The total RNA was further purified using the Invitrogen DNA-free DNA Removal Kit (Cat# AM1906). The crude and further purified total RNA extracts were analyzed using RNA quality metrics measured on a ThermoFisher Nanodrop 1000, and via agarose gel electrophoresis on a 1.5% (w/v) agarose gel with ethidium bromide staining. The samples were loaded into the gel with 1X Orange G DNA loading dye (50% (v/v) glycerol, 5 mM Tris-HCl, 0.5 mM EDTA), 0.5 mg/mL Orange G dye). The purified total RNA extracts were diluted to 150 ng/µL in nuclease-free water, aliquoted, and stored at -80°C until further use.

2.5: cDNA synthesis and quantitative PCR (qPCR)

Three hundred ng of purified total RNA was as template for cDNA synthesis in 20 µL reactions using the BIORAD iScript™ cDNA Synthesis Kit (Cat# 1708890), which uses a blend of oligo dT and random hexamer primers, according to the manufacturer's protocol. Negative controls for each sample were prepared using a Master Mix without the reverse transcriptase enzyme (-RT control), and a no template reaction

was prepared using nuclease-free water as a negative control (NTC). The reactions were carried out in a BIORAD T100™ Thermal Cycler using the following cycling program: priming for 5 minutes at 25°C, cDNA synthesis for 20 minutes at 46°C, and RT inactivation for 1 minute at 95°C. The resulting cDNA was aliquoted and stored at -20°C until further use.

The qPCR mastermixes (final reaction: 1X SsoAdvanced™ Universal SYBR® Green Supermix, 400 nM forward primer, 400 nM reverse primer) and cDNA samples (including (-RT) and (NTC) negative controls) were combined in 10 µL reactions (6 µL of mastermix, 4 µL of cDNA) in sealed plates and spun-down at 3000 RPM for 1 minute in a centrifuge (Eppendorf Centrifuge 5804 R). The qPCR reactions were carried out with an Applied Biosystems StepOne Plus qPCR thermocycler using the following cycling program over 40 cycles: preheating for 3 minutes at 95°C, denaturation for 30 seconds at 95°C, annealing and extension for 30 seconds at 59°C. A melt curve was generated by recording the fluorescence over a 55 – 95°C temperature range with 0.5°C increments (5 seconds/step). The genes assayed were *GAPDH* (At1g16300) (reference gene), *PP2A* (At1g69960) (reference gene), *GPAT5* (At3g11430), *FAR1* (At5g22500), *FAR4* (At3g44540), *FAR5* (At3g44550), *CYP86A1* (At5g58860), *CYP86B1* (At5g23190), *FACT* (At5g63560), and *ASFT* (At5g41040). The primer sequences used for each tested gene are listed in Appendix A.

2.6: Next Generation Sequencing using the NextSeq500 Illumina platform

The transcriptomes from *myb53-1 myb92 myb93* TKO line*, myb53-2 myb92 myb93* TKO line*,* wild-type (WT), MYB53-OE (6 h post-treatment with β-estradiol), MYB53-OE (15 h post-treatment with β-estradiol), MYB53-OE (6 h post-treatment with

DMSO), and MYB53-OE (15 h post-treatment with DMSO) were sequenced. Aliquots of 100 ng of purified RNA were processed by StemCore Laboratories (OHRI) for RNA-seq analysis. mRNA libraries were prepared using the Illumina TrueSeq RNA Library Prep Kit v2 following the manufacturer's protocol. The integrity of the RNA and the resulting mRNA library was determined using the ThermoFisher Qubit system and the Advanced Analytical Technologies Fragment Analyzer and visualized using the PROSize 3.0 software package (3.0.1.5). The mRNA libraries were sequenced using a High Output Flow Cell on the Illumina NextSeq500 system generating ~50 million 75 base-pair reads per sample and using PhiX as an internal control.

2.7: RNA-Seq bioinformatics

The resulting Illumina NextSeq500 75 bp reads were trimmed for low quality scores using Trimmomatic (v0.36) (Bolger *et al*., 2014) with default settings. Quality Assurance metrics were generated for each sample using FastQC, including basic statistics, per base sequence quality, per sequence quality scores, per base sequence content, per base GC content, per base N content, sequence length distribution, sequence duplication levels, over-represented sequences, and k-mer content.

The reference transcriptome was downloaded from the TAIR site, sourced from the TAIR 10 cDNA transcript release. An indexed transcriptome, which is a required input for quantification built by Kallisto (v0.43.1) (Bray *et al*., 2016), was generated by using the "index" command as instructed in the Kallisto documentation.

Quantification of the transcripts was achieved using the bootstrapping method of the Kallisto software package. Running the single-end settings of the Kallisto software

requires manual determination of the average transcript length, which was curated from the Qubit results of the mRNA libraries that were prepared by StemCore Laboratories.

The pseudo-alignment was performed using Kallisto software for 41,671 responsive genes, with 47,760,969 k-mers (of length 31) and 115,798 equivalence classes. The pseudo-alignment approach used by Kallisto addresses the common RNA-seq issue of dealing with ambiguous reads (Bray *et al*., 2016). To assign a particular read to a transcript, put simply, the possible k-mers of that read are compared to the possible k-mers generated by a transcript, called "k-compatibility classes", therefore skipping the internal alignment of a read to a transcript.

The variance of read counts was estimated using 100 bootstrapping cycles by sampling equivalence classes, with replacement. The transcript abundances were then determined using the expectation-maximization algorithm by the Kallisto software. The Kallisto transcript counts were compared between the treatment and control sample groups using the DESeq2 (v1.20.0) statistical software package in R (TKO versus WT and β-estradiol treatment versus DMSO negative control). The results were then filtered based on the following parameters: Base Mean > 5, Fold-Change (FC) > 2, False Discovery Rate (FDR) p-value < 0.10.

2.8: Gene Ontology (GO) term enrichment analysis

With the complete set of differentially expressed genes, lists corresponding to genes present in the *myb53 myb92 myb93* knock-out dataset and the inducible *AtMYB53*-OE dataset (at 6 h, 6 and 15 h, and 15 h), were each filtered into lists of genes that are up- and down-regulated for each treatment. Those lists were then uploaded to Princeton University's Generic GO Term Finder web tool (https://go.princeton.edu/cgi-

bin/GOTermFinder), with default statistics settings, and with custom gene annotation file from TAIR10 (2018-07-03). The results were filtered to include only GO Terms that were found to be enriched with an FDR p-value cut-off of 0.10.

2.9: Construction of plasmids to express full-length and truncated MYB proteins

Bacteria harboring plasmids with the full-length protein-coding sequence of *AtMYB53, AtMYB92,* or *AtMYB93* were grown overnight at 37°C on LB Media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride; Cat# LBL407.1; Bioshop) with 15 g/L agar (Cat# AGR001.500; Bioshop). The plasmids were purified using the Gene-Aid High-Speed Plasmid Mini Kit (Cat# PD100) following the manufacturer's instructions. The plasmids were used as templates for PCR with primers (Appendix A) that generated DNA sequences encoding the following C-terminal truncated proteins: AtMYB53 (1 – 183), AtMYB53 (1 – 276), AtMYB53 (1 – 289), AtMYB53 (1 – 301), AtMYB53 (1 – 310), AtMYB92 (1 – 184), AtMYB92 (1 – 304), AtMYB92 (1 – 317), AtMYB92 (1 – 328), AtMYB92 (1 – 334), AtMYB93 (1 – 172), AtMYB93 (1 – 335), AtMYB93 (1 – 348), AtMYB93 (1 – 359), and AtMYB93 (1 – 365). The plasmids were also used as templates for overlap PCR with primer sets (Appendix A) that generated DNA sequences encoding the following C-terminal tagged (3 x HA) proteins: AtMYB53 (1 – 276), AtMYB53 (1 – 310), AtMYB92 (1 – 304), AtMYB92 (1 – 334), AtMYB93 (1 – 335), and AtMYB93 (1 – 365). The amplicons were generated using the BIORAD iProof™ High-Fidelity PCR Kit (Cat# 1725330) according to the manufacturer's protocol. The

PCR products were gel purified from a 0.8% (w/v) agarose gel using the QIAGEN Gel Extraction Kit (Cat # 28704) according to the manufacturer's protocol. The purified PCR products were analyzed for quantity and purity using a ThermoFisher Nanodrop 1000. The purified PCR products were inserted into the pENTR-D-TOPO entry vector using the Invitrogen pENTR Directional TOPO Cloning Kit according to the manufacturer's instructions, but in a 2.5 µL reaction volume. The reactions were transformed into DH5α Stellar (Cat # 636763; Takara Bio USA, Inc.) chemical competent *E. coli* cells by heat shock at 42°C for 30 seconds. Five hundred µL of SOC liquid media was then added to each transformation reaction and incubated at 37°C for 1 hour at 200 RPM (Innova 40 Incubator Shaker Series; New Brunswick Scientific) before plating on LB solid agar with selection antibiotics (50 µg/mL kanamycin (Cat# KAN201.5; Bioshop)). The plated bacteria were grown overnight (~16 h) at 37°C.

To identify transformants with plasmids containing the desired inserts, single colonies were suspended in 20 µL of sterile milliQ water for colony PCR screening. PCR reactions were prepared with 2 µL of the bacterial suspension in 20 µL reaction volumes. The reactions were performed using a BIORAD T100 Thermocycler with the following cycling conditions: preheat 3 minutes at 95°C, denature for 30 seconds at 95°C, primer annealing at 60°C for 15 seconds, extension for 1 minute at 72°C, 35 total cycles of denaturation/annealing/extension, and a final extension of 10 minutes at 72°C. The amplicons were run on a 0.8% (w/v) agarose gel to identify transformants with plasmids containing the expected insert size.

Overnight cultures of positive transformants identified by colony PCR were prepared by inoculating 5 mL of LB liquid media with selection antibiotic (10 g/L

tryptone, 5 g/L yeast extract, 10 g/L sodium chloride, 50 µg/mL kanamycin) overnight (~16 h) at 37°C and at 200 RPM (Innova 40 Incubator Shaker Series; New Brunswick Scientific). The plasmids of these bacteria were purified using the Gene-Aid High-Speed Plasmid Mini Kit (Cat# PD300) following the manufacturer's protocol. Plasmids were further screened to contain the desired inserts using diagnostic restriction digests. Fifty µL reactions were set up using 500 – 1000 ng of plasmid DNA and incubated for 3 hours at 37°C before loading the digested DNA on a 0.8% (w/v) agarose gel as described above. Finally, plasmids were sent for Sanger DNA sequencing at Eurofins Genomics (Toronto, Canada) to confirm the constructs were correct and that the inserts contained no unexpected mutations, for example from the PCR amplification.

Plasmid sequences (pENTR-D-TOPO entry vector containing the target construct sequence) that were verified by diagnostic restriction digests and Sanger DNA sequencing were then used in a Gateway reaction to insert the target sequences into the pK7WG2D expression vector (Karimi *et al*., 2002). The Invitrogen Gateway LR Clonase II Enzyme Mix kit (Cat # 11791020) was used according to the manufacturer's protocol but with reduced volumes. The LR Clonase reaction contained 10 ng of the pENTR-D-TOPO entry vector and 140 ng of the pK7WG2D expression vector in a 3 µL reaction volume. The reactions were transformed into DH5α Stellar chemically competent *E. coli* cells by heat shock at 42°C for 30 seconds. Five hundred µL of SOC liquid media was added to each suspension, and then incubated at 37°C for 1 hour at 200 RPM (Innova 40 Incubator Shaker Series; New Brunswick Scientific) before plating on LB solid agar with selection antibiotics (80 µg/mL spectomycin; Cat# SPE201.5; Bioshop). The plated

bacteria were incubated overnight (~16 h) before performing colony PCR, plasmid purification, diagnostic restriction digests, and Sanger sequencing, as described above.

2.10: Agrobacterium-mediated transient over-expression of MYBs in leaves of *Nicotiana benthamiana*

The pK7WG2D expression plasmids containing MYB coding sequences were transformed into GV3101 (C58 – rif R– Ti pMP90– pTiC58DT-DNA – gentR – Nopaline) *Agrobacterium tumefaciens* cells by electroporation. The cells were suspended in fresh liquid LB media without antibiotics, then incubated at 30°C for 1 hour at 220 RPM (E24 Incubator Shaker Series; New Brunswick Scientific) before plating on LB solid media with selection antibiotics (100 μg/mL rifampicin (Cat# RIF222.5; Bioshop), 40 μg/mL gentamycin (Cat # GTA202.5; Bioshop), 80 μg/mL spectomycin). The plated bacteria were incubated at 30°C for 48 hours before performing colony PCR as described above.

Overnight *A. tumefaciens* cultures of successful transformants were prepared using 5 mL of liquid LB media with the selection antibiotics specified above and incubated at 30°C for 16 hours at 220 RPM (E24 Incubator Shaker Series; New Brunswick Scientific). Overnight cultures containing the pBIN19-p19 and pBIN19 empty vector were also prepared in LB media containing 100 μg/mL rifampicin, 40 μg/mL gentamycin, and 5 μg/mL tetracycline, and incubated under the same conditions. The cultures were centrifuged at 2000 RPM (Eppendorf Centrifuge 5804 R) for 5 minutes and decanted. The resulting bacterial pellets were rinsed 3 times with an Infiltration Solution (50 mM MES buffer (pH = 5.6), 2 mM $Na_3PO_4 \cdot 12\ H_2O$, 5 mg/mL glucose, and 0.1 mM acetylsyringone (dissolved in DMSO)), centrifuging at 2000 RPM (Eppendorf Centrifuge

5804 R) for 5 minutes between each rinse. The rinsed bacterial cells were resuspended in the Infiltration Solution to an OD_{600nm} of 0.3. Each bacterial suspension was then combined with the pBIN19-p19 bacterial suspension, and the Infiltration Solution, in a ratio of (1:1:1). Using a needle-less syringe, the undersides of the leaves of 8-week old *N. benthamiana* plants were infiltrated with each bacterial mixture and the plants returned to growth chambers for 6 days before harvesting the infiltrated tissue.

2.11: Delipidation of tissues to obtain lipid polyester free of soluble lipids

Two frozen 2 cm diameter leaf discs from the transient over-expression assay, or the pooled root material from the knockout analysis, were macerated with a razor blade and then transferred to $CHCl_3$-rinsed and pre-weighed ~8 mL glass extraction tubes. For each sample, four such bio-replicates were prepared. The tubes containing the plant tissue were filled with heated (80°C) isopropanol (ACS grade; Cat# 76406-540; Anachemia [VWR]) and vortexed, and incubated at 80°C for 15 minutes. After allowing the samples to cool slightly, half of the volume of isopropanol was discarded and the plant tissue was ground into a fine powder using a hand-held homogenizer. The samples were topped off again with isopropanol and left to incubate on a tilting platform at room temperature for 12 hours. After centrifugation at 2000 RPM at room temperature for 10 minutes in an Eppendorf Centrifuge 5804 R, the spent organic liquid was decanted and replaced with fresh isopropanol. After vortexing the samples to resuspend the plant material, the samples were left to incubate again on a tilting platform at room temperature for 12 hours. Following the same decanting procedure, the samples were incubated in chloroform:methanol solutions of varying ratios (2:1, 1:1, 1:2, 0:1, in that order) for 24 hours each, changing the solution once in that timeframe. HPLC-grade methanol (Cat#

BDH20864.400; VWR BDH Chemicals) and chloroform (Cat# 234911-540; Anachemia [VWR]) were used in the delipidation process. After the delipidation process was completed, the methanol contained in the tubes was decanted. The delipidated leaf tissue was allowed to dry for 24 hours in a fume hood, and then for 3 days in a sealed desiccator filled with calcium sulfate (Drierite™). After the plant material was thoroughly dried out, the tubes were weighed to obtain the dry weight (DW) of the plant material.

2.12: Base-catalyzed depolymerization and acetylation

For quantification of the suberin monomers, the samples were spiked with 10 µg each of 17:0 methyl ester (Cat # H4515-1G; Sigma Life Science) and ω-pentadecalactone (OPL) (Cat # 76530; Fluka) as internal standards, dissolved in toluene. The samples were then suspended in a 1.5 mL aliquot of a Depolymerization Solution (60% (v/v) methanol, 25% (v/v) sodium methoxide (Cat# 156256; Sigma-Alderich, 15% (v/v) methyl acetate (Cat# 296996; Sigma-Alderich). The samples were then incubated for 2 hours at 60°C, vortexing every 30 minutes, and then acidified by the addition of one mL of ACS grade glacial acetic acid (Cat# 00598-463; Anachemia [VWR]). The fatty acid methyl esters were extracted by adding 2.5 mL of ACS grade dichloromethane (Cat# 32667-360; Anachemia [VWR]) and vortexing thoroughly. A series of liquid-liquid extractions were performed using 0.5 M NaCl (Cat# SOD002.1; Bioshop) (3 times), centrifuging the samples between extractions for 5 minutes at 2000 RPM at room temperature to separate the aqueous and organic layers. The lower organic phase was transferred to a clean tube after the first extraction, and at subsequent extractions, the upper aqueous layer was removed. To ensure complete dehydration of the samples, a generous amount of anhydrous sodium sulphate (Cat# 85330-380; Anachemia [VWR]) was added to the

samples to absorb any residual aqueous solution and incubated overnight at room temperature. Following centrifugation for 5 minutes at 2000 RPM at room temperature to compact the salt pellet, the organic phase was transferred to a clean tube. The samples were dried under nitrogen gas with gentle heating. Acetylated derivatives of the extracted FAME's were prepared by adding 200 µL of a 1:1 mixture of pyridine:acetic anhydride (Cat# 270970; Sigma-Alderich and Cat# 351001-100; Fisherbrand, respectively) to each sample and incubated at 60°C for one hour, vortexing thoroughly every 15 minutes. The samples were dried once more under nitrogen gas. After resuspending in 100 µL of HPLC grade n-hexane (Cat# BDH24575.400; VWR BDH Analytical), the samples were analyzed by gas chromatography (GC).

2.13: Gas chromatography and quantification of aliphatic suberin monomers

A Varian 450 gas chromatograph equipped with a flame ionization detector (FID) was used to analyze aliphatic suberin following depolymerization (released suberin monomers). For each sample, one µL was injected in splitless mode at an injector temperature of 300°C. Compounds were separated using an Agilent HP5-MS capillary column (dimensions: 30 m length; 0.25 mm diameter; 0.25 µm film thickness). The initial oven temperature was 140°C for 3 minutes, then ramped up to 310°C at a rate of $10°C \cdot min^{-1}$, and held for 10 minutes. The eluted compounds were detected by the FID set at 350°C.

Representative samples for infiltrated leaf tissue were also analyzed by GC-MS using the same GC parameters at the Carleton Mass Spectrometry Facility to confirm identifications (peak patterns corresponded between GC-MS and GC-FID analyses). Quantification was done relative to the internal standards using the GC-FID data. Peak

areas were converted to micrograms per gram of plant dry weight. Statistical analysis across samples were carried out using the two-tailed multiple t-test, with Holm-Sidak method for multiple comparison corrections, using GraphPad Prism 8.1.0 Software.

CHAPTER 3 - RESULTS

In this study, two previously described independent triple knockout lines (TKO-1 and TKO-2) of the *MYB53*, *MYB92* and *MYB93* genes (Hu, 2018) were used in a genome-wide expression profiling (RNA-seq) experiment to identify differentially expressed genes in comparison to wild-type. The two triple mutant lines, *myb53-1 myb92-1 myb93-1* (TKO-1) and *myb53-2 myb92-1 myb93-1* (TKO-2), differ in the T-DNA insertion positions of *MYB53* (Figure 11). We also made use of a previously described β-estradiol-inducible overexpression line for *AtMYB53* (Hu, 2018), and performed RNA-seq experiments to identify the differentially expressed genes in comparison to control. The combined transcriptome datasets of the triple *MYB* knockout lines and steroid-inducible *MYB53* overexpression line were deemed necessary for two reasons. First, the data-mining possibilities would be increased by having complementary loss-of-function and overexpression experiments to uncover novel suberin-associated genes, and second, be able to filter out the most likely gene targets of these transcription factors by first focusing on genes that were differentially regulated in both experiments.

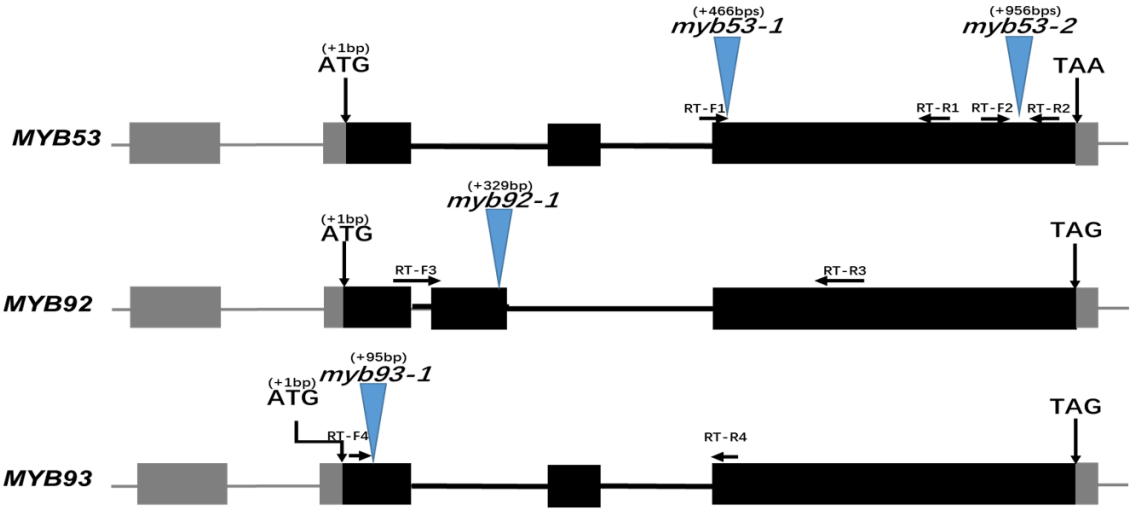


Figure 11: Summary of the locations of the T-DNA insertions in the alleles of *MYB53*, *MYB92*, and *MYB93* gene regions that make up the triple knock-out lines. Each insertion is labelled with the position relative to the start codon (in base-pairs), and the primer annealing sites used to validate the successful knock-out for each gene. The primer pairs RT-F1/RT-R1 and RT-F2/RT-R2 were used to verify the T-DNA insertion for the *myb53-1* and *myb53-2* alleles, respectively. The primer pairs RT-F3/RT-R3 and RT-F4/RT-R4 were used to verify the T-DNA insertion in the *myb92-1* and *myb93-1* alleles. Each schematic includes the introns (represented by solid black lines), exons (represented by black boxes), and 5' and 3' UTR indicated in gray.

Prior to the construction of mRNA libraries for Illumina NextSeq RNA sequencing, we biochemically verified the phenotypes of the starting plant materials because the suberin phenotypes in either the loss-of-function or overexpression systems are not readily visible. The suberin content in the *AtMYB53/92/93* triple knock-out mutants were chemically determined using base-catalyzed depolymerization of the suberin polymer and analysis of constituents by gas chromatography. Additionally, the expression levels of select suberin biosynthetic genes were analyzed by quantitative PCR (qPCR) using total RNA extracts from both the triple *AtMYB53/92/93* loss-of-function lines and the *AtMYB53* overexpression line after steroid induction.

Previous research showed that the reference genes *GAPDH* and *PP2A* are suitable to use as housekeeping genes for qPCR assays related to both the triple knock-out mutant lines and the MYB53-OE lines (Hu, 2018). Additionally, previous testing demonstrated that the effect of the DMSO control treatment in the over-expression line experiments has no effect in the qPCR assay compared to no treatment at all (Hefeng Hu, unpublished data). Therefore, for this reason, the DMSO control treatment was considered here as an appropriate negative control in these qPCR experiments, and for the RNA-seq experiments described further below.

3.1: Verification of the *myb53 myb92 myb93* triple knockout lines by chemical and gene expression analyses

The root suberin content of two-week old seedlings of wild-type (WT) and the triple knockout mutants, TKO-1 and TKO-2, were compared (Figure 12). The TKO-1 and TKO-2 lines had 73.8% and 76.6% reductions, respectively, in total suberin content compared to WT (1196 ± 59 μg suberin/g root DW and 1068 ± 70 μg suberin/g root DW

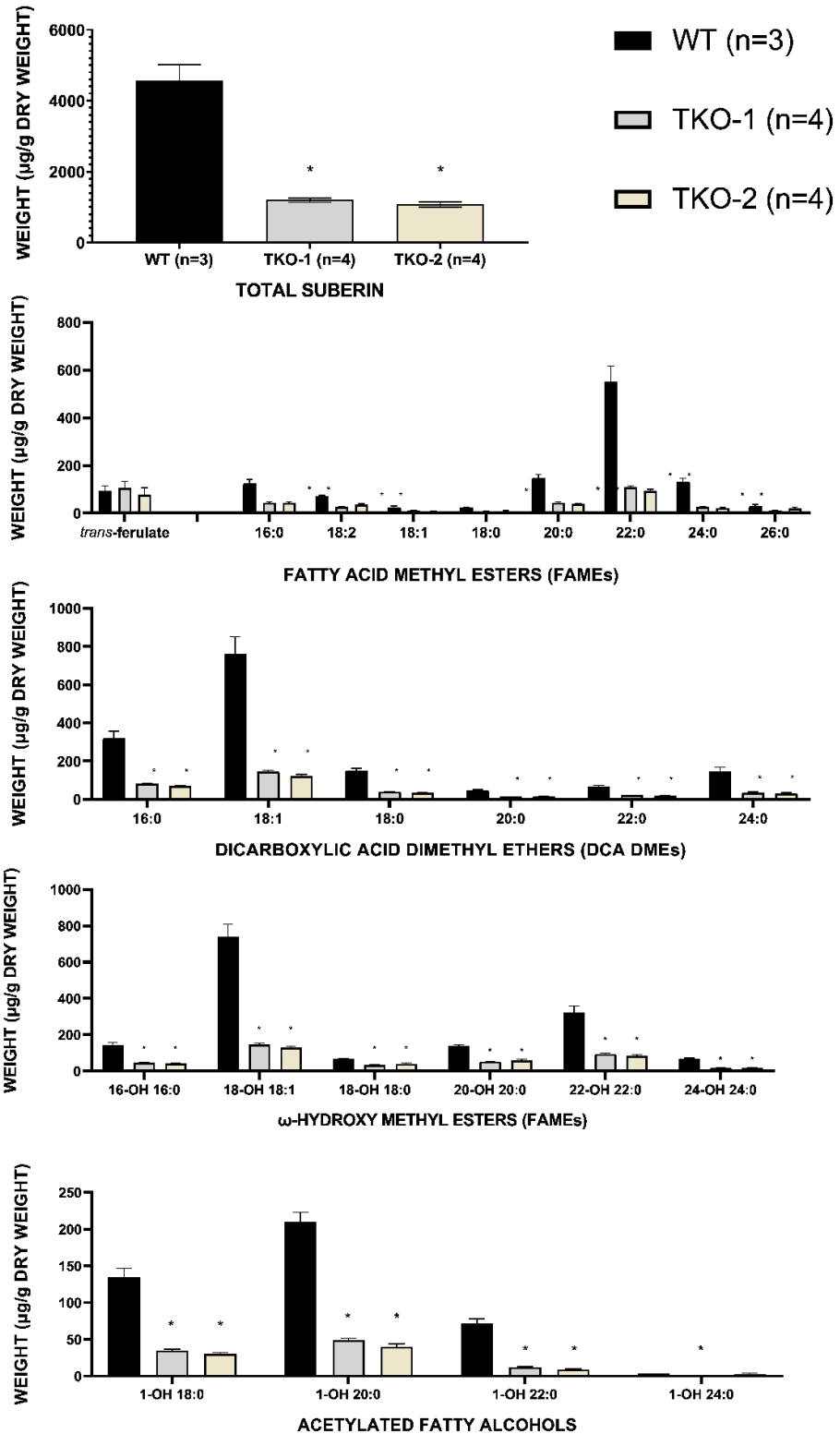
for TKO-1 and TKO-2, respectively, compared to 4567 ± 446 µg suberin/g root DW for wild-type). There was no significant difference in the total root suberin content between the two triple knockout mutant lines.

With regards to suberin monomer composition, each chemical class of suberin compound had similar proportional reductions in the triple knockout lines compared to the wild-type control (Figure 12). All chemical classes of suberin monomers were reduced in the triple knockout lines except for ferulate. This included unsubstituted fatty acids (FAs), dicarboxylic fatty acids (DCAs), ω-hydroxy FAs, and primary fatty alcohols. In particular, significant reductions in the TKO-1 and TKO-2 mutant lines were observed in the most abundant root suberin monomers, such as 22:0 FA (80.3% and 83.3% reduced, respectively), 18:1 DCA (81.1% and 84.2% reduced, respectively), and 18-OH 18:1 FA (80.2% and 82.6% reduced, respectively). However, *trans*-ferulate, the most abundant aromatic component of suberin, was found to not be significantly different between the two triple *MYB* mutant lines and wild-type (105 ± 28 and 77 ± 29 µg suberin/g root DW for TKO-1 and TKO-2, respectively, versus 93 ± 22 µg suberin/g root DW for WT).

The transcript levels of select known suberin biosynthetic genes (Vishwanath *et al.,* 2015) were then analyzed in two-week old seedlings of the triple *MYB* mutant lines in comparison to wild-type using qPCR (Figure 13). Specifically, the transcript levels of three of the major suberin biosynthetic enzymes were analyzed: *FAR1*, *GPAT5*, and *CYP86A1*. These genes were chosen as representative genes involved in either the reductive or oxidative parts of the suberin biosynthetic pathways. Both TKO-1 and TKO-2 alleles of the triple MYB mutant showed strong down-regulation of gene expression for

all three of the suberin biosynthetic genes tested (Figure 13). The *FAR1* gene had the largest decrease in relative gene expression in the triple MYB knockout, resulting in percent decreases of 79 ± 9 % and 69 ± 6 % in TKO-1 and TKO-2, respectively, when compared to the wild-type (WT). However, while *GPAT5* dropped significantly in relative gene expression in the TKO-1 by about 4.8-fold (79 ± 13 %) fold change compared to WT), the differences observed in the second allele (TKO-2) were found to be non-significant in a student's two-tailed t-test. In both TKO-1 and TKO-2, down-regulation of the *CYP86A1* gene was observed when compared to the expression of the same gene in WT (84 ± 15 % and 72 ± 18 % fold changes in TKO-1 and TKO-2, respectively, compared to WT).

Figure 12 (next page): The amount of total aliphatic suberin and suberin component break-down per gram of dry weight for both alleles (TKO-1 and TKO-2) of the Arabidopsis *myb53 myb92 myb93* triple knockout lines compared to wild-type (WT) as detected by GC-FID. The sample size was n = 4 for TKO-1 and TKO-2 mutant lines, and n = 3 for the WT line. (*) signifies statistical significance between the mutant line compared to the WT at the $\alpha = 0.05$ level (student's t test; GraphPad Prism 8.1.0).


WEIGHT (µg/g DRY WEIGHT)
6000
4000
2000
0
WT (n=3)
TKO-1 (n=4)
TKO-2 (n=4)
TOTAL SUBERIN
WT (n=3)
TKO-1 (n=4)
TKO-2 (n=4)
800
600
400
200
0
trans-ferulate
16:0
18:2
18:1
18:0
20:0
22:0
24:0
26:0
FATTY ACID METHYL ESTERS (FAMEs)
1000
800
600
400
200
0
16:0
18:1
18:0
20:0
22:0
24:0
DICARBOXYLIC ACID DIMETHYL ETHERS (DCA DMEs)
16-OH 16:0
18-OH 18:1
18-OH 18:0
20-OH 20:0
22-OH 22:0
24-OH 24:0
ω-HYDROXY METHYL ESTERS (FAMEs)
250
200
150
100
50
0
1-OH 18:0
1-OH 20:0
1-OH 22:0
1-OH 24:0
ACETYLATED FATTY ALCOHOLS

(Figure 12)

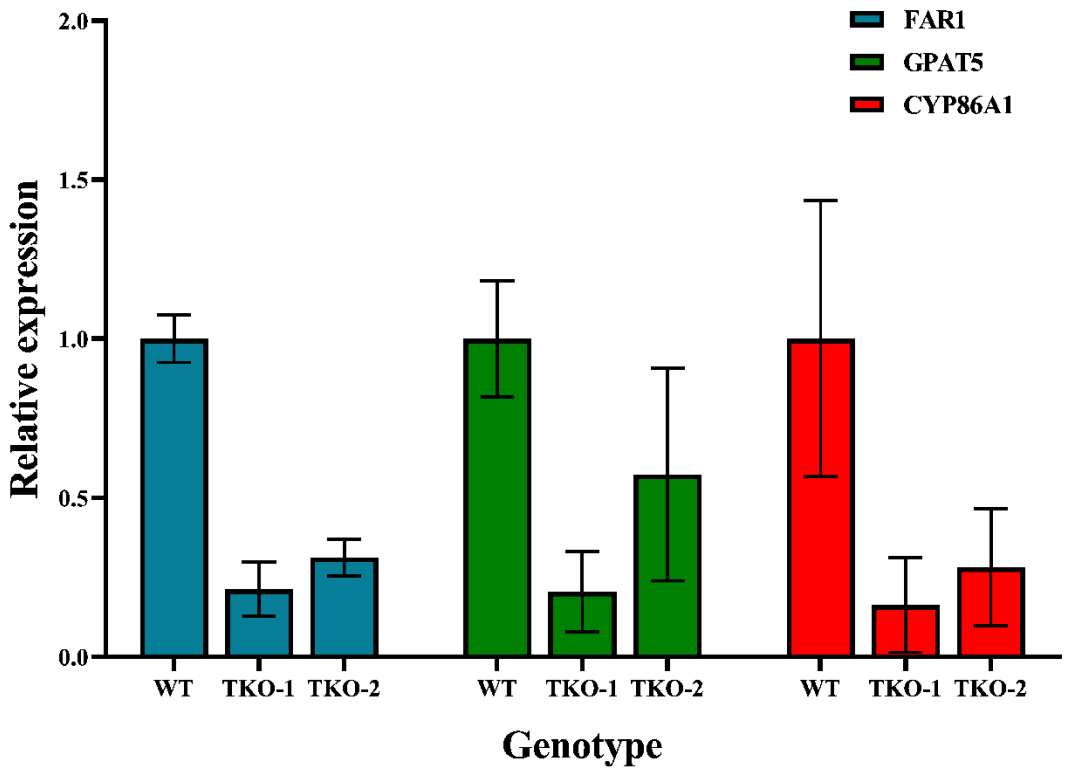


Figure 13: Relative expression of select suberin biosynthetic genes for the Arabidopsis *myb53 myb92 myb93* triple mutant lines (TKO-1 and TKO-2) compared to WT using two-week old seedlings. Transcript abundance was quantified by qPCR and normalized using the *GAPDH* reference gene. The differences relative to the WT group were statistically significant at the α=0.05 level using the student's t-test. Error bars represent the SD across 3 independent bio-replicates (each with three technical replicates).

3.2: Validation of the starting plant material for the TRANSPLANTA *MYB53* over-expression line by gene expression analysis

While previous research had analyzed the accumulation of suberin in response to β-estradiol-induced *AtMYB53* over-expression (Hu, 2018) (Figure 10), the transcript profiles of suberin-associated genes in this system were not analyzed. The previous study had shown that suberin does not accumulate at 24 hours following steroid induction, therefore collecting suberin chemical data on extracts harvested before this time would be uninformative. However, we wanted to examine the transcriptomes at very early time-points, prior to full assembly of the suberin polymer. The roots of 10-day old *Arabidopsis thaliana* seedlings were exposed to β-estradiol or a DMSO mock control for a period between 2 hours and 24 hours to observe transcriptional changes (Figures 14 and 15). RNA obtained from these treated roots were assayed using three major suberin biosynthetic genes, *GPAT5*, *CYP86A1*, and *FAR5.*

For all three genes, there were no significant changes in transcript levels in roots that were exposed to β-estradiol versus the DMSO mock treatment after 2 hours of steroid induction (Figure 14). At 6 hours post-treatment, *GPAT5* transcript levels were induced 2.99 ± 0.74 fold in the β-estradiol treatment compared to the DMSO treatment, while the transcript levels remained stable between the treatments for *CYP86A1* and *FAR5*. At 15 hours post-steroid treatment, major increases in transcript levels were observed for all three genes, resulting in up-regulation of *GPAT5*, *CYP86A1*, and *FAR5* by fold-changes of 4.74 ± 2.24, 5.08 ± 0.42, and 10.3 ± 2.8, respectively. At 24 hours post-treatment, the level of induction decreased slightly for *GPAT5* and *FAR5* compared to 15 hours post-treatment, while the induction level of *CYP86A1* was the same at 15 and

24 hours post-treatment.

A second time-course experiment with 10-day old seedlings was performed, this time for a period ranging from 6 to 15 hours following overexpression of *MYB53* to narrow the window of gene expression following induction of *MYB53* (Figure 15). Additional suberin-associated genes were assayed by qPCR: *GPAT5*, *CYP86A1*, *CYP86B1*, *FAR4,* and *ASFT*. Between 6 and 15 hours, the magnitude of gene expression of *GPAT5* increased slightly, resulting in fold changes increasing from 2.64 ± 0.56 at 6 hours to 3.52 ± 1.39 at 10 hours, and then peaked thereafter at 12 hours, with a fold change of 8.07 ± 2.07. Finally, at 15 hours post-treatment, the induced gene expression tapered off to an induction fold change of 6.8 ± 2.6. Similarly, the expression of *ASFT*, *FAR4,* and *CYP86B1* peaked at 12 hours, with fold-changes of 12.1 ± 3.8, 4.20 ± 1.74, and 10.9 ± 3.9 respectively. The expression levels of these genes began decreasing again at the 15 hour mark.

In both *AtMYB53* overexpression time-course experiments (Figures 14 and 15), 6 hours marked the time-point with the lowest induced expression above the DMSO control levels. This was true for *GPAT5*, *CYP86A1*, *CYP86B1*, *FAR4,* and *ASFT*, while in the case of *FAR5*, there was no significant difference compared with the DMSO control treatment at 6 hours. With the majority of transcriptional changes beginning to occur, the 6 hour time-point was chosen for capturing very early events following MYB53 overexpression, including genes that may be induced prior to the suberin biosynthetic genes (e.g. possibly intermediary regulatory factors). At 15 hours, the expression of all suberin-associated biosynthetic genes tested were still highly induced, but just beginning to decrease. Therefore, this time-point was chosen to broadly identify genes induced by

MYB53 even if they are indirectly regulated. Thus, RNA from the 6 and 15 hour time-points were chosen for transcriptomics analysis by Illumina NextSeq RNA-sequencing.

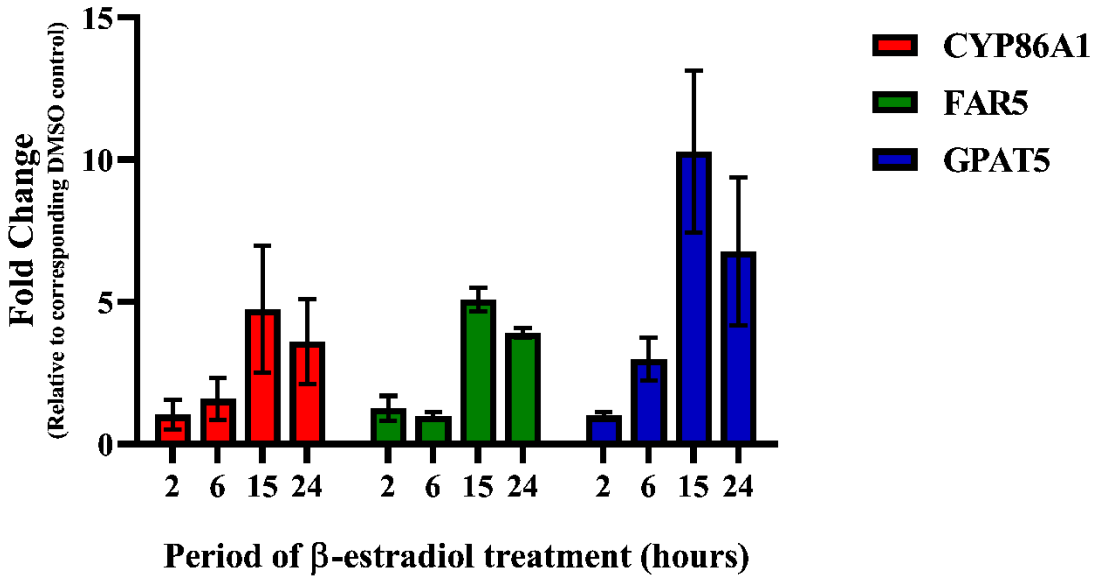


Figure 14: Relative expression of select suberin biosynthetic genes in the steroid-inducible *AtMYB53* overexpression line, measured at 2, 6, 15, and 24 hours post treatment with 20 µM β-estradiol. Transcript abundance was quantified by qPCR and normalized using the *GAPDH* reference gene, and fold changes calculated relative to the corresponding relative gene expression of the corresponding DMSO control. Error bars represent the SD across 3 independent bio-replicates (each with three technical replicates).

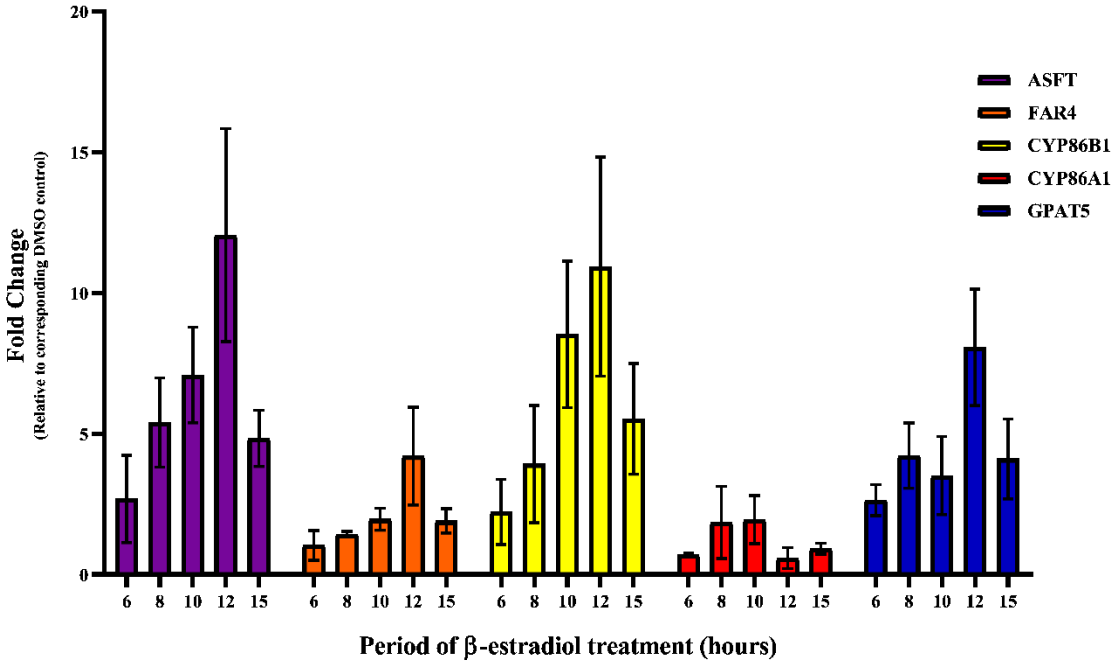


Figure 15: Relative expression of select suberin biosynthetic genes in the steroid-inducible *AtMYB53* overexpression line, measured at 6, 8, 10, 12, and 15 hours post treatment with 20 µM β-estradiol. Transcript abundance was quantified by qPCR and normalized using the *GAPDH* reference gene, and fold changes calculated relative to the corresponding relative gene expression of the corresponding DMSO control. The differences relative to the control group were statistically significant at the α=0.05 level using the student's t-test. Error bars represent the SD across 3 independent bio-replicates (each with three technical replicates).

3.3: Processing of the raw-reads generated by Illumina NextSeq 500 sequencing

The RNA-sequencing reads were trimmed using Trimmomatic v0.36 software using the default settings to remove the Illumina sequencing adaptors and to exclude low-quality reads before proceeding to the next step of the RNA-seq pipeline (Table 1). This operation involved removing the leading and trailing bases that fell below a quality score of 3, scanning the remaining part of the read with a 4-base long scanning window, and removing the read when the overall quality score was below 15. Finally, reads that were less than 36 bases long were removed.

Prior to trimming, more than 75 million reads each were sequenced for the samples belonging to the experimental group (WT, TKO-1, and TKO-2), while the over-expression experimental group (DMSO and β-estradiol treatments at 6H and 15H each) generated more than 44 million reads each (Table 1). The trimming removed between 1.48% to 2.99% of reads in the triple *myb* knockout experimental group, while 1.02% to 1.67% of reads were removed in the over-expression experimental group. In the end, between 76.7 and 96.5 million high quality reads remained in the triple knock-out experiment samples, while in the over-expression experiment samples 44.0 to 66.0 million high quality reads remained.

Table 1. Summary of the number of single-end reads generated by next-generation sequencing of the mRNA libraries prepared from total RNA extracts, using the Illumina NextSeq500 platform. The raw reads provided were trimmed from low-quality data using Trimmomatic (v0.36) software using the default settings, and the resulting number of high-quality reads used for downstream transcriptome alignment and quantification are reported below.

Sample genotype/treatment	**Bio-replicate**	**Number of Raw Reads ($\times 10^6$)**	**Number of un-trimmed reads ($\times 10^6$)**	**Number of trimmed reads ($\times 10^6$)**	**Percentage of reads removed**
Triple knock-out (TKO) samples:					
Wild-type (WT)	A	86.2	83.9	2.3	2.62%
	B	87.2	84.6	2.6	2.99%
TKO-1	A	77.9	76.7	1.2	1.48%
	B	98.7	96.5	2.2	2.20%
TKO-2	A	94.3	92.7	1.6	1.71%
	B	95.0	93.4	1.6	1.65%
MYB53 over-expression (OE) samples:					
DMSO Control (6H)	A	44.6	44.0	0.7	1.48%
	B	48.6	48.0	0.6	1.25%
DMSO Control (15H)	A	56.1	55.6	0.6	1.02%
	B	51.3	50.7	0.6	1.09%
β-estradiol treatment (6H)	A	57.1	56.1	1.0	1.67%
	B	57.2	56.6	0.7	1.15%
β-estradiol treatment (15H)	A	66.7	66.0	0.7	1.10%
	B	44.6	44.1	0.5	1.09%

3.4: Alignment and processing of the sequencing reads

Using Kallisto v0.43.1 software, the coverage of the pseudo-alignment of the reads to the indexed reference transcriptome (TAIR10) was consistently between 81.0% to 85.0% for the samples in both TKO and overexpression samples (Table 2), except for bio-replicate A from the 6 hour β-estradiol treatment sample, which only had coverage of about 70.8%.

Count matrices that are required inputs for the DESeq2 software package were constructed by identifying each sample as a member of either a control or a treatment group, accompanied by their corresponding read counts for each transcript in the indexed reference transcriptome. These count matrices were populated using the count data from the Kallisto quantification algorithm and analyzed using DESeq2. The results were then filtered to obtain statistically significant, differentially expressed responsive genes as described below. A base mean cutoff of 5 reads averaged over the treatment and control groups (TKO mutant line versus WT line, and β-estradiol treatment versus DMSO mock treatment, respectively), and fold changes between the experimental groups that are at least 2-fold were used as filters. These were selected to increase the probability that the observed differences could be considered relevant at the biological level. Furthermore, the corrected p-value statistic, denoted as the false-discovery rate (FDR) adjusted p-value, was capped at 0.10 to allow flexibility for gene discovery.

According to the above described parameters, 124 genes were identified to be DE genes in the loss-of-function mutant lines (Table 3). Among these, 58 and 68 genes were up-regulated and downregulated, respectively, in the mutant line compared to the control. In the *AtMYB53* over-expression line experiment, a much larger number of DE genes

were identified. A total of 527 early-responsive genes were differentially expressed when the β-estradiol exposure period lasted for 6 hours; from these, 212 genes were found to be up-regulated, and the remaining 315 genes were found to be down-regulated (Table 3). A total of 1099 late-responsive DE genes were identified at 15 hours of exposure to the hormone; and among these, 647 genes were found to be up-regulated, and the remaining 452 genes were found to be down-regulated. Between both the early- and late-responsive genes, 174 genes were found to be shared, leaving 353 genes differentially expressed exclusively at 6 hours of β-estradiol exposure, while 925 genes were exclusive to the 15 hour treatment. From the 353 genes expressed at 6 hours exclusively, 139 and 214 genes were found to be up- and down-regulated, respectively, while from the 925 genes expressed at 15 hours exclusively, 567 and 358 genes were found to be up- and down-regulated, respectively. From the 174 DE genes shared between the two time-points, 66 and 87 genes were found to be up- and down-regulated, respectively. There were 42 genes that exhibit differential expression patterns at 6 hours and 15 hours post-treatment: 14 genes were down-regulated in the 6 hour treatment and up-regulated in the 15 hour treatment, while 7 genes were up-regulated in the 6 hour treatment and down-regulated in the 15 hour treatment. Finally, a comparison between the 124 DE genes in the triple *myb* knock-out experiment and the 1452 DE genes in the over-expression experiment (both time-points), there were 43 genes that were differentially expressed in both experiments (Figure 17).

Table 2. Results of the pseudo-alignment between the high-quality single-end reads of the indicated sample data and the indexed reference *Arabidopsis thaliana* transcriptome obtained from the TAIR10 genome release.

Sample genotype/treatment	Bio-replicate	Average read length (bp)	Number of reads processed ($\times 10^6$)	Number of reads pseudo-aligned ($\times 10^6$)	Percentage of reads aligned
Triple knock-out (TKO) samples:					
Wild-Type (WT)	A	331 ± 66	83.9	70.1	83.5%
	B	314 ± 57	84.6	71.8	84.8%
TKO-1	A	347 ± 66	76.7	64.1	83.5%
	B	327 ± 66	96.5	80.4	83.3%
TKO-2	A	311 ± 46	92.7	78.4	84.6%
	B	332 ± 62	93.4	78.0	83.5%
MYB53 over-expression (OE) samples:					
DMSO Control (6H)	A	348 ± 76	44.0	36.0	81.8%
	B	338 ± 76	48.0	40.2	83.7%
DMSO Control (15H)	A	339 ± 76	55.6	46.2	83.1%
	B	334 ± 71	50.7	42.5	83.8%
β-estradiol treatment (6H)	A	330 ± 72	56.1	39.7	70.8%
	B	330 ± 70	56.6	48.0	84.9%
β-estradiol treatment (15H)	A	348 ± 83	66.0	56.0	84.9%
	B	345 ± 80	44.1	37.0	83.9%

Table 3: Summary of the reported number of differentially expressed (DE) genes for the Arabidopsis *myb53 myb92 myb93* triple-knockout lines and the *AtMYB53* over-expression line. DE gene counts were determined by the statistical analysis of the transcript abundances derived from the Kallisto software (v0.43.1), by using the DESeq2 software package (v1.20.0). Responsive genes are determined to be statistically significant and classed as DE genes where the following parameters are applied: base mean > 5, FC > 2, FDR p-value < 0.10.

Sample genotype/treatment	Total number of DE Genes	Number of up-regulated DE genes	Number of down-regulated DE genes
Triple knock-out (TKO) analysis	124	58	66
MYB53 over-expression (OE) analysis - 6 H	527	212	315
MYB53 over-expression (OE) analysis - 15 H	1099	647	452
Overlapping DE genes in the knock-out and the entire over-expression dataset	43	--	--

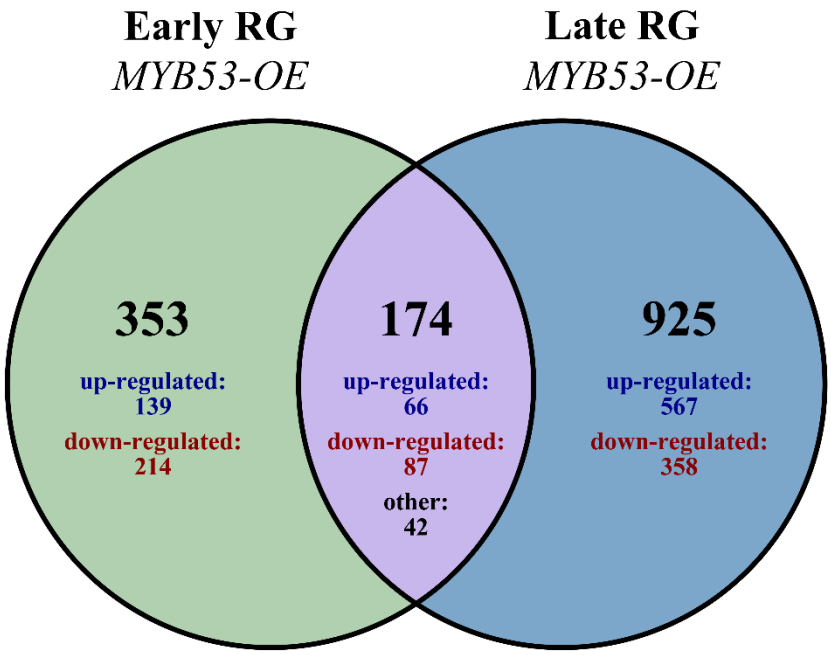


Figure 16: Venn diagram showing the total number (1452 genes) of responsive genes (RG) in the 6 hour (blue) and 15 hour (green) timepoints in the MYB53 over-expression experiment (TRANSPLANTA). Analysis was performed using the DESeq2 statistical software (v1.20.0). Responsive genes that are identified as DE genes strictly at 6 hours are classed as early gene targets, while those that are identified strictly at 15 hours are classed as later responsive genes. The "other" category in the intersecting section refers to genes with opposite gene expression trends between the two time periods (ex: up-regulation at 6 hours and down-regulation at 15 hours). The purple overlapping region identifies the number of common DE genes between the two time-points (base mean > 5, FC > 2, FDR p-value < 0.10).

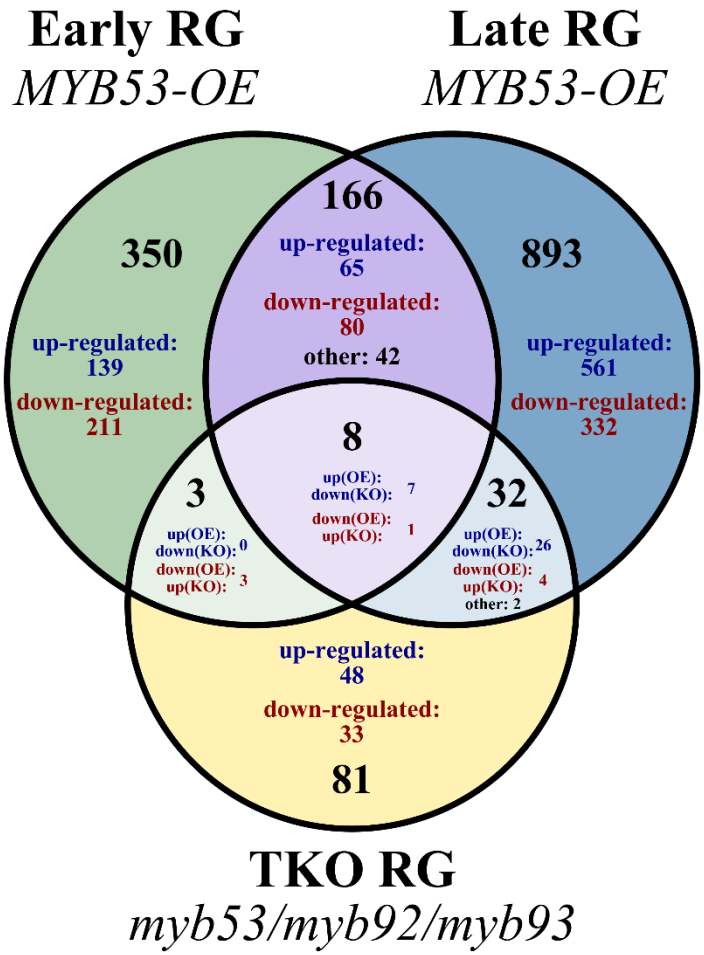


Figure 17: Venn diagram showing the full dataset (1533 genes) of responsive genes (RG) in the 6 hour (blue) and 15 hour (green) timepoints in the MYB53 over-expression experiment (TRANSPLANTA), and the DE genes in the triple knockout mutant (yellow). Analysis was performed using the DESeq2 statistical software (v1.20.0). Responsive genes that are identified as DE genes strictly at 6 hours are classed as early responsive genes , while those that are identified strictly at 15 hours are classed as later responsive genes. The “other” category in the intersecting section refers to genes with opposite gene expression trends between the two time periods (ex: up-regulation at 6 hours and down-regulation at 15 hours). The purple overlapping region identifies the number of common DE genes between the two time-points (base mean > 5, FC > 2, FDR p-value < 0.10).

3.5: Summary of differentially expressed genes that overlap in the KO and OE datasets

Within this group of 43 overlapping differentially expressed responsive genes (Table 4; Figure 18), only 3 genes were found to be differentially expressed exclusively at 6 hours of exposure to β-estradiol. RBOHB, an NADPH-oxidase (At1g09090), was up-regulated in the knock-out line (2.14 ± 0.15 FC for both TKO-1 and TKO-2, respectively) and down-regulated in the over-expression line (0.47 ± 0.06 $\log_2$ FC at 6 h) (Table 4). Two other genes, a leucine-rich family protein (At1g51890) and a Tesmin/TSO1-like CXC-domain protein (At2g20110) had the opposite trend. They were down-regulated in the triple knockout lines, and up-regulated in the over-expression line at 6 hours post β-estradiol treatment.

There were eight genes that were overlapping in the triple knock-out lines and over-expression line at both 6 and 15 hours post-induction (Table 4; Figure 18). These genes encode predicted enzymes and transporters that could be involved in suberin biosynthesis based on knowledge of cutin biosynthesis. Two GDSL-motif esterases/acyltransferases/ lipases were differentially expressed. One of them, GLIP2 (At1g53940), was up-regulated in the triple *myb* knock-out line and down-regulated in the over-expression line at both time-points (2.64 ± 0.37, 2.83 ± 0.39, 0.38 ± 0.13, and 0.33 ± 0.07 FC for TKO-1, TKO-2, MYB53-OE (6 h), MYB53-OE (15 h), respectively), while the other predicted GDSL lipase (At3g50400) had the opposite trend (0.29 ± 0.06, 0.20 ± 0.04, 2.64 ± 0.37, and 9.19 ±1.27 FC for TKO-1, TKO-2, MYB53-OE (6 h), MYB53-OE (15 h), respectively). A predicted α/β-hydrolase (At4g36610) was also down-regulated in the knock-out line and up-regulated at both time-points in the over-

expression line. Finally, two predicted ABC transporters exhibited this same behaviour, ABCG6 (At5g13580) and ABCG23 (At5g19410).

The remaining 32 genes of this overlapping set of 43 genes were previously classed as late-responsive genes (Table 4; Figure 18). Within the group of genes that were up-regulated in the triple *myb* knockout analysis and down-regulated in the over-expression line, there is a peroxidase superfamily protein (At4g11290) that appears in this category (3.48 ± 0.48, 4 ± 0.55, and 0.5 ± 0.07 FC for TKO-1, TKO-2, and MYB53-OE (15 h), respectively). There is also a C_2H_2 and C_2HC zinc finger superfamily protein (At5g43540), with fold changes 2.00 ± 0.42, 2.64 ± 0.55, and 0.27 ± 0.11 FC for TKO-1, TKO-2, and MYB53-OE (15 h), respectively. In the case where down-regulation of genes in the triple knockout analysis coincides with an up-regulation of the over-expression line at 15 hours post-induction, there were many different classes of proteins that were found to be differentially expressed (Table 4; Figure 18). Among those, there are two known suberin biosynthetic genes, *FAR4* (At3g44540) and *CYP86A1* (At5g58860), and one known cutin biosynthetic gene, *LACS2* (At1g49430) that appear in this grouping. Moreover, two CASP-like proteins, *CASPL1B2* (At4g20390) (0.25 ± 0.07, 0.18 ± 0.05, and 3.03 ± 0.42 FC for TKO-1, TKO-2, and MYB53-OE (15h), respectively) and *CASPL1D2* (At3g06390) (0.41 ± 0.08, 0.23 ± 0.06, and 3.48 ± 0.48 FC for TKO-1, TKO-2, and MYB53-OE (15 h), respectively) also follow the same pattern. Another GDSL-motif esterase/acyltransferase/ lipase (At2g23540) is differentially expressed in the same manner (0.47 ± 0.1, 0.33 ± 0.07, and 3.25 ± 0.23 FC for TKO-1, TKO-2, and MYB53-OE (15 h), respectively). Other catalytic enzymes in this grouping include an O-acyltransferase (At5g12420), an SGNH hydrolase-type esterase (At5g37690), and an

HXXXD-type acyl-transferase (At1g78990). Three transcription factors, *MYB52* (At1g17950), *MYB54* (At1g73410), and *WRKY56* (At1g64000) are also down-regulated in the knockout line and up-regulated in the over-expression line.

Table 4. Summary of the fold-changes and statistics of the 43 responsive genes that were found to be overlapping between the triple-knockout analysis and the over-expression analysis (6 or 15 hours). The locus IDs were matched to the corresponding gene IDs and gene descriptions using the TAIR10 genome release. Records that appear as grey text represent data that was found to be statistically non-significant according to the following parameters: base mean > 5, FC > 2, FDR p-value < 0.10.

	LOCUS ID	GENE ID	BASE MEAN				LOG BASE 2 FOLD CHANGE (±SE) [Mutant line compared to control]				FDR p-VALUE				TAIR10 Gene Description
			KO Analysis		TPL-OE line		KO Analysis		TPL-OE line		KO Analysis		TPL-OE line		
			TKO-1	*TKO-2*	*6 H*	*15 H*	*TKO-1*	*TKO-2*	*6 H*	*15 H*	*TKO-1*	*TKO-2*	*6 H*	*15 H*	
Early-Responsive Genes - Differential Regulation at 6 H TPL-OE	***Up-regulated in KO analysis; down-regulated in TPL-OE analysis:***														
	AT1G09090.2	RBOHB	3701	3812	776	722	1.1 ± 0.1	1.1 ± 0.1	-1.1 ± 0.2	-0.8 ± 0.2	2.65E-10	5.40E-10	2.10E-05	3.65E-03	NADPH-oxidase
	AT1G51890.2	--	139	224	337	402	23.2 ± 4.8	11.2 ± 1.5	-1.4 ± 0.5	-0.6 ± 0.4	3.25E-04	5.40E-10	7.56E-02	5.51E-01	Leucine-rich repeat protein kinase family protein
	AT2G20110.2	--	32	28	10	14	8.4 ± 2.2	8.2 ± 2.3	-6.9 ± 2.2	1.4 ± 2	1.37E-02	2.99E-02	4.31E-02	1.00E+00	Tesmin/TSO1-like CXC domain-containing protein
	Down-regulated in KO analysis; up-regulated in TPL-OE analysis:														
	----------		----------				----------				----------				----------
Gene targets with both Early- and Late-Responsive Genes	***Up-regulated in KO analysis; down-regulated in TPL-OE analysis:***														
	AT1G53940.1	GLIP2	3233	3502	206	297	1.4 ± 0.2	1.5 ± 0.2	-1.4 ± 0.5	-1.6 ± 0.3	1.01E-09	4.87E-09	5.59E-02	1.81E-04	GDSL-motif esterase/acyltransferase/lipase
	Down-regulated in KO analysis; up-regulated in TPL-OE analysis:														
	AT3G50400.1	--	976	941	1923	4777	-1.8 ± 0.3	-2.3 ± 0.3	1.4 ± 0.2	3.2 ± 0.2	7.16E-07	2.60E-10	5.48E-09	1.03E-62	GDSL-motif esterase/acyltransferase/lipase
	AT4G36610.1	--	2739	2777	3820	8575	-1 ± 0.1	-1.1 ± 0.2	1.5 ± 0.2	3 ± 0.2	1.46E-09	4.91E-08	3.19E-13	7.79E-56	α/β-hydrolase protein
	AT1G73290.1	SCPL5	205	215	350	640	-1.8 ± 0.4	-1.7 ± 0.4	1.4 ± 0.2	3.4 ± 0.4	2.18E-04	2.83E-03	4.38E-07	1.25E-16	Serine carboxypeptidase-like 5
	AT2G41170.1	--	987	1007	1807	4950	-1.1 ± 0.3	-1.2 ± 0.3	1.9 ± 0.2	3.1 ± 0.2	5.54E-02	5.73E-04	3.51E-14	1.04E-50	F-box family protein
	AT2G43390.1	--	205	227	365	1170	-1.8 ± 0.5	-1.6 ± 0.5	1.4 ± 0.3	3.7 ± 0.3	9.85E-03	3.68E-02	8.88E-04	1.74E-31	Hypothetical protein
	AT5G13580.1	ABCG6	2981	2669	6684	18175	-1.1 ± 0.3	-1.8 ± 0.3	1.1 ± 0.2	3.1 ± 0.2	4.60E-02	3.17E-07	6.39E-07	2.09E-78	ATP-BINDING CASSETTE G6 half transporter protein
	AT5G19410.1	ABCG23	650	617	2069	6340	-1.4 ± 0.3	-1.8 ± 0.4	1 ± 0.2	3.2 ± 0.1	1.73E-04	1.62E-03	8.78E-04	5.59E-99	ABC-2 type transporter family protein

Table 4: (cont'd)

	LOCUS ID	GENE ID	BASE MEAN				LOG BASE 2 FOLD CHANGE (±SE) [Mutant line compared to control]				FDR p-VALUE				TAIR10 Gene Description
			KO Analysis		TPL-OE line		KO Analysis		TPL-OE line		KO Analysis		TPL-OE line		
			TKO-1	*TKO-2*	*6 H*	*15 H*	*TKO-1*	*TKO-2*	*6 H*	*15 H*	*TKO-1*	*TKO-2*	*6 H*	*15 H*	
Late-Responsive Genes - Differential Regulation at 15 H OE	***Up-regulated in KO analysis; down-regulated in TPL-OE analysis:***														
	AT4G25800.2	--	40	89	154	41	6.3 ± 1.9	7.4 ± 1.8	-0.5 ± 0.7	-4.8 ± 0.9	7.94E-02	3.45E-03	1.00E+00	8.41E-06	Calmodulin-binding protein
	AT4G11290.1	--	12770	15274	1888	2169	1.8 ± 0.2	2 ± 0.2	-0.7 ± 0.2	-1 ± 0.2	1.75E-21	2.09E-13	3.07E-03	4.83E-05	Peroxidase superfamily protein
	AT5G52710.1	--	498	547	27	61	1.6 ± 0.3	1.7 ± 0.3	-1.5 ± 0.8	2.4 ± 0.6	3.43E-05	7.32E-08	4.14E-01	7.41E-04	Copper transport protein family
	AT1G27730.1	STZ	3744	3684	1012	783	1.2 ± 0.2	1.1 ± 0.2	0 ± 0.3	2.1 ± 0.4	1.57E-05	2.82E-05	1.00E+00	1.59E-06	Cys2/His2-type zinc-finger related protein
	AT2G19970.1	--	2093	2543	439	488	1.2 ± 0.4	1.5 ± 0.4	-0.4 ± 0.3	-1.7 ± 0.3	9.89E-02	6.00E-02	7.90E-01	2.73E-07	CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1 protein) superfamily protein
	AT5G43540.1	--	455	578	58	59	1 ± 0.3	1.4 ± 0.3	-1.2 ± 0.6	-1.9 ± 0.6	2.54E-02	1.37E-04	3.58E-01	3.86E-02	C2H2 and C2HC zinc fingers superfamily protein
	Down-regulated in KO analysis; up-regulated in TPL-OE analysis:														
	AT3G44540.1	FAR4	2576	2384	2613	4889	-1 ± 0.2	-1.6 ± 0.4	0.3 ± 0.1	1.2 ± 0.1	2.44E-06	1.81E-03	1.73E-01	1.06E-18	Fatty Acid Reductase 4
	AT2G48140.1	EDA4	3903	3612	4976	11726	-1 ± 0.2	-1.6 ± 0.3	0.8 ± 0.1	2.6 ± 0.2	5.31E-07	1.22E-05	8.66E-07	1.59E-42	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
	AT4G29800.2	PLP8	903	907	1152	2026	-1 ± 0.2	-1.2 ± 0.2	0.7 ± 0.2	1.9 ± 0.2	4.51E-05	8.97E-06	9.03E-03	5.11E-21	PATATIN-like protein 8
	AT2G23540.1	--	4632	4296	9162	20039	-1.1 ± 0.3	-1.6 ± 0.3	0 ± 0.1	1.7 ± 0.1	9.89E-03	6.93E-07	1.00E+00	1.78E-39	GDSL-motif esterase/acyltransferase/lipase
	AT5G12420.1	--	5070	4878	6270	12483	-1.1 ± 0.2	-1.4 ± 0.2	0.7 ± 0.1	2.2 ± 0.2	2.30E-03	1.21E-07	1.65E-04	3.07E-41	O-acyltransferase
	AT2G40370.1	LAC5	904	867	2210	6275	-1.1 ± 0.3	-1.5 ± 0.4	0.8 ± 0.2	2.6 ± 0.2	2.99E-03	5.12E-02	7.08E-04	4.41E-59	Member of laccase family of genes
	AT1G49430.1	LACS2	2270	2126	3702	6839	-1.1 ± 0.3	-1.6 ± 0.3	0 ± 0.1	1.5 ± 0.1	1.58E-02	8.90E-07	1.00E+00	1.91E-35	Long-chain acyl-CoA synthetase 2
	AT2G22510.1	--	3465	3291	6896	22549	-1.3 ± 0.3	-1.9 ± 0.4	0.3 ± 0.1	2.8 ± 0.1	2.08E-03	3.02E-04	1.17E-01	1.11E-72	Hydroxyproline-rich glycoprotein family protein
	AT2G01580.1	--	237	223	277	491	-1.3 ± 0.4	-1.9 ± 0.4	0.3 ± 0.3	1.2 ± 0.3	3.13E-02	9.05E-04	9.78E-01	4.88E-04	Transmembrane protein
	AT3G06390.1	CASPL1D2	1714	1585	2639	5385	-1.3 ± 0.3	-2.1 ± 0.4	0.1 ± 0.1	1.8 ± 0.2	8.06E-04	3.73E-05	1.00E+00	3.01E-18	CASP-LIKE PROTEIN 1D2 uncharacterized protein family
	AT4G38080.1	--	8110	7744	13206	36992	-1.3 ± 0.2	-1.9 ± 0.4	0.2 ± 0.1	2.2 ± 0.1	8.45E-07	1.08E-04	8.69E-01	2.57E-50	Hydroxyproline-rich glycoprotein family protein
	AT1G73410.1	MYB54	352	342	338	1133	-1.4 ± 0.3	-1.7 ± 0.3	0.2 ± 0.2	1.9 ± 0.2	2.68E-03	6.47E-05	9.57E-01	1.30E-15	Myb domain protein 54
	AT5G58860.1	CYP86A1	4342	4142	9297	17669	-1.4 ± 0.3	-1.9 ± 0.3	-0.3 ± 0.1	1 ± 0.1	1.09E-03	8.67E-08	2.77E-01	1.20E-23	Member of the CYP86A subfamily of cytochrome p450 genes
	AT1G29000.1	--	147	130	218	531	-1.4 ± 0.4	-2.6 ± 0.5	-0.2 ± 0.3	1.1 ± 0.2	7.24E-02	7.49E-05	9.84E-01	1.69E-06	Heavy metal transport/detoxification superfamily protein
	AT1G17950.1	MYB52	316	319	375	1384	-1.4 ± 0.4	-1.5 ± 0.4	0.8 ± 0.3	2.5 ± 0.2	8.52E-02	6.58E-03	2.73E-02	1.50E-26	Myb domain protein 52
	AT1G55990.1	--	943	909	1529	3007	-1.4 ± 0.2	-2.1 ± 0.3	0.8 ± 0.2	3.4 ± 0.2	1.59E-06	1.40E-06	4.36E-03	3.17E-69	Glycine-rich protein
	AT5G09520.1	PELPK2	1106	1114	2120	2602	-1.4 ± 0.3	-1.7 ± 0.3	0.8 ± 0.2	2.4 ± 0.2	3.68E-04	1.00E-05	1.32E-03	2.72E-26	Hydroxyproline-rich glycoprotein family protein
	AT5G20860.1	--	208	204	559	1182	-1.6 ± 0.4	-1.9 ± 0.5	0.9 ± 0.2	3.7 ± 0.2	4.60E-03	6.71E-03	7.14E-05	7.72E-69	Plant invertase/pectin methylesterase inhibitor superfamily
	AT3G07970.1	QRT2	291	290	610	1215	-1.7 ± 0.3	-1.9 ± 0.5	0.2 ± 0.2	2.4 ± 0.2	1.39E-04	1.57E-02	8.93E-01	1.28E-26	QUARTET2 pollen related protein
	AT4G33610.1	--	3637	3540	5930	12934	-1.7 ± 0.2	-2.2 ± 0.3	0.8 ± 0.2	3.9 ± 0.2	1.02E-17	5.40E-10	1.52E-04	6.36E-119	Glycine-rich protein
	AT5G07130.1	LAC13	1216	1170	2763	6151	-1.7 ± 0.2	-2.2 ± 0.3	1.0 ± 0.2	3.2 ± 0.2	5.22E-19	9.81E-09	1.56E-05	1.08E-73	Laccase 13
	AT4G20390.1	CASPL1B2	2324	2273	3517	9915	-2 ± 0.4	-2.5 ± 0.4	0.0 ± 0.1	1.6 ± 0.2	2.50E-05	6.94E-09	1.00E+00	3.56E-17	CASP-LIKE PROTEIN 1B2 Uncharacterized protein family
	AT1G64000.1	WRKY56	242	229	476	976	-2.1 ± 0.5	-3 ± 0.6	0.5 ± 0.2	1.7 ± 0.2	9.79E-03	2.33E-04	1.46E-01	6.17E-11	Member of WRKY Transcription Factor; Group II-c
	AT2G19200.1	--	311	309	632	1769	-2.1 ± 0.3	-2.4 ± 0.4	0.7 ± 0.2	2.5 ± 0.2	5.02E-07	5.54E-08	1.42E-02	3.62E-28	Pseudogene of hypothetical protein
	AT5G37690.1	--	2670	2700	6478	17505	-2.8 ± 0.2	-3 ± 0.2	-0.2 ± 0.1	1.4 ± 0.1	1.32E-38	9.86E-41	5.11E-01	3.28E-31	SGNH hydrolase-type esterase
	AT1G78990.1	--	672	667	2064	5120	-3.8 ± 0.3	-5.3 ± 0.4	-0.3 ± 0.2	1.8 ± 0.2	3.19E-33	3.99E-29	8.27E-01	8.38E-30	HXXXD-type acyl-transferase

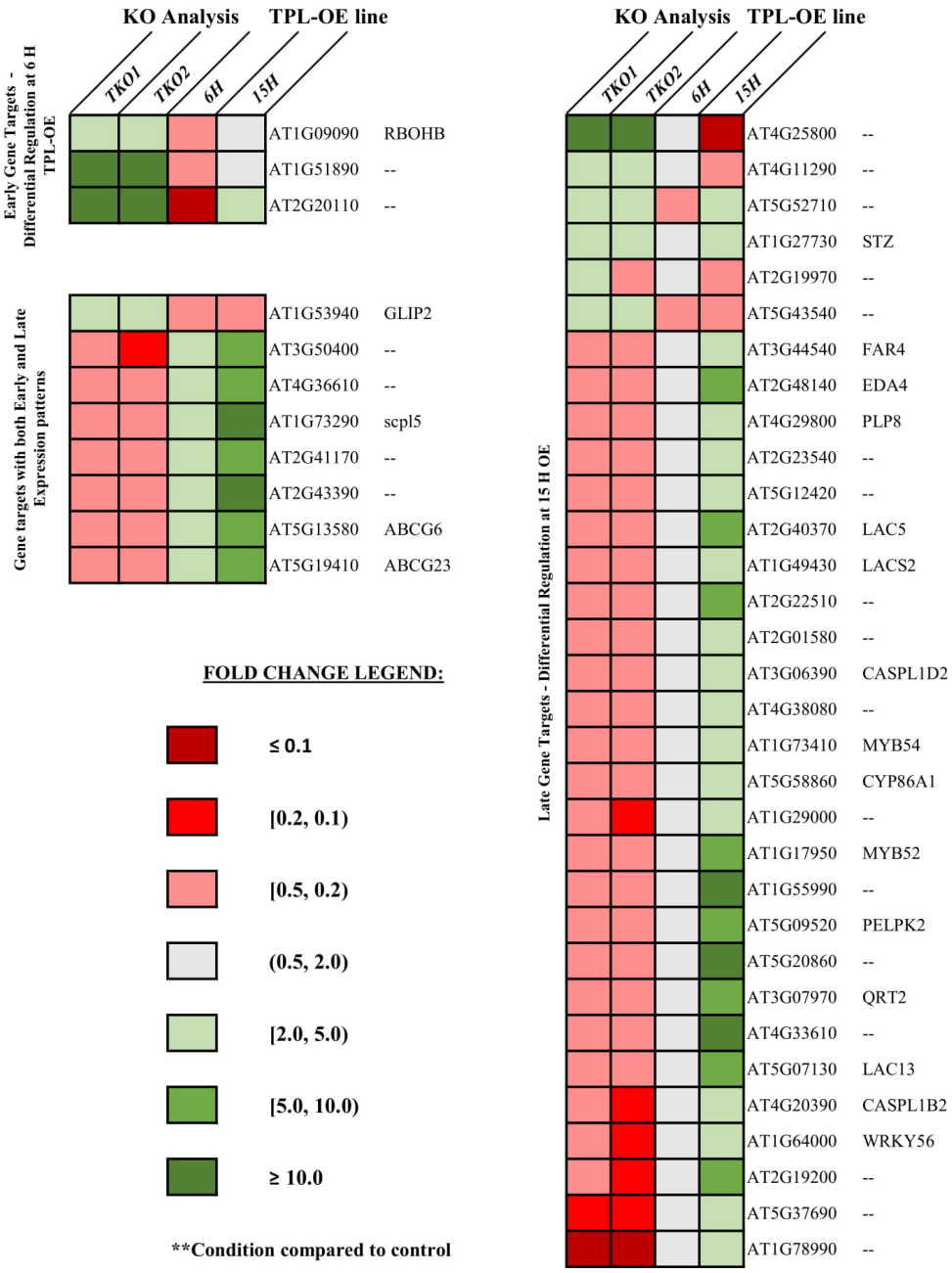


Figure 18: Summary of the fold changes measured from RNA-seq data generated using the NestSeq500 Illumina platform, resulting in a set of 43 overlapping DE genes in the knock-out and over-expression mutant lines. RNA was extracted from 14-day old roots (TKO) and 10-day old roots (MYB53-OE) after 6 or 15 hours of *MYB53* overexpression. Responsive genes are grouped relative to temporal class as assigned by the over-expression data: base mean > 5, FC > 2, FDR p-value < 0.10.

3.6: Using the GO-Term Enrichment Analysis method to analyze entire dataset of DE genes

While the overlapping list of 43 genes between the knockout and over-expression datasets gives great confidence about the biological relevance of those observed differences in expression, it only represents a small subset of the total 1533 genes that were initially filtered from the sequencing results. In order to fully take advantage of this large dataset, gene ontologies (GO) were used to annotate the filtered dataset for GO Enrichment Analysis across all three categories: Biological Processes (BP) (Table S2), Molecular Function (MF) (Table S3) and Cellular Components (CC) (Table S4). Additional details on the statistics and genes assigned to each annotation can be found in the indicated supplementary tables.

3.6.1: GO-Term Enrichment analysis of Biological Processes

Within the *myb53 myb92 myb93* triple knockout dataset relating to Biological Processes (Table S2), the enrichment analysis resulted in many groupings of up-regulated genes relating to plant responses to stimuli, including “Response to stimulus” [GO:0050896] (31 genes), “Response to stress” [GO:0006950] (23 genes), “Response to wounding” [GO:0009611] (7 genes), and “Response to ethylene” [GO:0009723] (7 genes). In contrast, the down-regulated genes in the triple knockout dataset were mostly related to cell wall biogenesis; for example, “Regulation of secondary cell wall biogenesis” [GO:2000652] (4 genes) and “Regulation of cell wall organization or biogenesis” [GO:1903338] (4 genes). Also, the triple knockout GO Enrichment revealed that many terms linked metabolic processes were down-regulated, such as

"Phenylpropanoid metabolic process" [GO:0009698] (6 genes), "Secondary metabolic process" [GO:0019748] (8 genes), "Suberin biosynthetic process" [GO:0010345] (3 genes), and "Lipid metabolic process" [GO:0006629] (11 genes).

In the over-expression line, each time-point has distinct GO terms that were found to be over-represented in the enrichment analysis for Biological Processes. In the up-regulated early-responsive genes (Table S2), the term represented with the highest statistical significance was "Ornithine biosynthetic process" [GO:0006592] (2 genes), while in the up-regulated late-responsive genes (Figure 19, panel D), there are many more, and with more variety, with GO terms such as "Fatty acid metabolic process" [GO:0006631] (28 genes), "Carboxylic acid metabolic process" [GO:0019752] (52 genes), "Suberin biosynthetic process" [GO:0010345] (7 genes), "Wax biosynthetic process" [GO:0010025] (9 genes), and "Lipid metabolic process" [GO:0006629] (44 genes). Also in the late-responsive genes, there are genes associated with plant responses that are over-represented, most notably "Response to osmotic stress" [GO:0006970] (29 genes). In the dataset overlapping the up-regulated early and late-responsive genes (Figure 19, panel C), the GO terms "Suberin biosynthetic process" [GO:0010345] (3 genes) was over-represented.

In the down-regulated genes in the over-expression line, the categories represented are again distinct for each time-point examined. In the early-responsive genes (Figure 19, panel B), there are a wide variety of small molecule pathways that are over-represented, such as "Glycosyl compound biosynthetic process" [GO:1901659] (9 genes) and "S-glycoside biosynthetic process" [GO:0016144] (7 genes). In the down-regulated late-responsive genes (Table S2), GO terms related to catabolic processes appear as over-

represented; for example, “Hydrogen peroxide catabolic process” [GO:0042744] (13 genes) and “Cofactor catabolic process” [GO:0051187] (13 genes). Finally, the lists containing down-regulated overlapping responsive genes from the early- and late-responsive genes (Table S2) is dominated by terms relating specifically to plant root development, like “Root development” [GO:0048364] (12 genes) and “Root morphogenesis” [GO:0414015] (8 genes).

3.6.2: GO-Term Enrichment analysis of Molecular Functions

In the Molecular Function (MF) GO-term enrichment analysis for the *myb53 myb92 myb93* triple knockout dataset (Table S3), the most prominent finding is that there are many GO terms relating to ion- and metal- -binding activities that are up-regulated in the knockout line, more specifically, “Manganese ion binding” [GO:0030145] (4 genes), and “Calcium ion binding” [GO:0005509] (5 genes).

The over-expression line for *AtMYB53* had only a few Molecular Function GO terms corresponding to up-regulated genes when analyzed exclusively at 6 hours post-treatment (Table S3), however the overlapping dataset between 6 and 15 hours post-treatment and the dataset exclusive to 15 hours did have numerous enriched GO terms ((Table S3). The term “Hydrolase activity” [GO:0016787] (20 genes) and “Transition metal ion binding” [GO:0046914] (11 genes) were represented in the former category, while the vast majority of significantly enriched terms found in the latter category relate directly to lipid metabolism. For example, “Long-chain fatty acid-CoA ligase activity” [GO:0004467] (4 genes) and “Fatty acid elongase activity” [GO:0009922] (4 genes) both appear in the 15 hour post-treatment dataset (Table S3). It should be noted however that the remaining GO-terms represented are redundant in terms of the lists of genes that are

annotated to those terms. There are also a large number of genes annotated to the "Catalytic activity" [GO:0003824] term (162 genes).

For the Molecular Function GO terms for the over-expression line, there was significant enrichment in down-regulated genes observed in both 6 hour and 15 hour post-treatment categories. At 6 hours, "Catalytic activity" [GO:0003824] was found to be over-represented (96 genes) (Table S3). Overlapping the 6 hour and 15 hour datasets, "Structural constituent of cell wall" [GO:0005199] (5 genes) was the only over-represented group of down-regulated genes (Table S3). At 15 hours post-treatment, the majority of the over-represented classes include terms related to binding, such as "Ion binding" [GO:0043167] (87 genes,), "Small molecule binding" [GO:0036094] (55 genes). Also, the term "Ion transmembrane transporter activity" [GO:0015075] (25 genes) was found to be enriched in this dataset as well.

3.6.3: GO-Term Enrichment analysis of Cellular Components

In the Cellular Component (CC) GO-term enrichment analysis for the *myb53 myb92 myb93* triple knockout line dataset and over-expression line dataset revealed that, in general, terms relating to the cell periphery and associated structures were significantly over-represented and up-regulated in the knockout line, while they are down-regulated in the over-expression line (Figure 21). For example, the term "Cell wall" [GO:0005618] was found to follow this trend in the knock-out line (10 genes annotated), the overlapping 6 hour and 15 hour over-expression line sets (12 genes), and in the 15 hour over-expression set (24 genes) (Table S4). The same enrichment trend is observed for the term "Extracellular region" [GO:0005576], with 19, 28, and 76 genes that are differentially regulated in the triple knockout, 6 hour and 15 hour overlapping, and 15 hour post-

treatment datasets. Furthermore, terms related directly to the plasma membrane appear to be affected in this way as well; “Plasma membrane” [GO:0005886] (6H; 49 genes), “Membrane” [GO:0016020] (6H; 83 genes), and “Integral component of membrane” [GO:0016021] (15H; 75 genes) all are enriched GO-terms in the downregulated dataset.

3.7: Justification for generating protein truncations of MYB proteins

The experimental design of the RNA-seq experiments described above took advantage of triple *myb53 myb92 myb93* mutant lines and a steroid-inducible AtMYB53 overexpression line to generate a large database of transcriptomic data. With this dataset, gene discovery for suberin-related genes and functions were possible using both an extremely stringent filtering process resulting in the discovery of 43 putative suberin-related responsive genes, and also a thorough analysis of all of the available data to identify other interesting gene candidates associated with suberin deposition. However, the drawback of this RNA-seq experimentation is that it cannot confirm, nor deny, the identity of the direct gene targets for *MYB53*, *MYB92* or *MYB93*. A natural next step is to develop epitope-tagged *MYB* protein constructs destined for chromatin immunoprecipitation experiments (ChIP) to identify the direct gene targets for each respective *MYB* transcription factor of interest.

In our repeated efforts to identify gene targets of the MYB transcription factors by chromatin immunoprecipitation (ChIP) failed due to difficulties with epitope antibody methods. First, epitope tags at the N-terminus of AtMYB53, AtMYB92 or AtMYB93eliminated the activity of each protein.(Jhadeswar Murmu, personal communication). Second, when the epitope tags were placed at the C-terminus of AtMYB53, AtMYB92 or AtMYB93, the protein fusions appeared to have full function, but the protein was no longer able to be recognized by the corresponding antibodies. It is possible that proteolysis of the epitope tag from the C-terminus occurred (Jhadeswar Murmu and Owen Rowland, personal communication). I therefore explored whether C-

terminal truncated versions of AtMYB53, AtMYB92, and AtMYB93 were active and whether these truncated versions could retain an epitope tag.

3.8: Development and testing "Series1" and "Series 2" truncations

The development of such epitope tagged constructs begins with the validation of a number of truncated MYB constructs with C-terminal Human influenza hemagglutinin (HA) tags and testing the extent of their activity *in vivo*. A multiple sequence alignment (Figure 22) of the protein sequences for *AtMYB53*, *AtMYB92*, and *AtMYB93* was used to determine conserved amino acid sequences in all three proteins. In addition to the conserved R2 and R3 repeat domains at the N-termini required for DNA binding and likely some protein-protein interactions, conserved regions in the C-terminus may represent recognition sequences for post-translational protein processing or other functions. Truncated versions of each MYB were tested for activity using the *Nicotiana benthamiana* leaf transient assay as measured by ectopic suberin production (Table 5; Figures 22 and 23). Simultaneously, these experiments enable characterization of the importance of the amino acid sequences at the C-terminus of each protein in activating downstream gene expression.

Table 5: List of the truncations cloned into pENTR-D-TOPO and pK7WG2D for the *Agrobacterium*-mediated transient overexpression assay in *N. benthamiana* leaves based on the amino acid sequence alignment for *AtMYB53*, *AtMYB92*, and *AtMYB93*.

Transcription Factor	Full Length Protein	**"Series 1"** After SUB – II Motif	**"Series 2"** Before SUB – III Motif	**"Series 3"** Within SUB – III Motif	**"Series 4"** After SUB – III Motif
AtMYB53	(1-310)	1-183	1-276	1-289	1-301
AtMYB92	(1-334)	1-184	1-304	1-317	1-328
AtMYB93	(1-365)	1-172	1-335	1-348	1-359

The “Series1”truncations were chosen such that the protein retained both R2 and R3 domains that are required for DNA binding to target sequences, and the first two conserved SUB-type motifs (SUB – I and SUB – II) that are proximal to the DNA binding domain (Figures 22 and 23) (Lashbrooke *et al.*, 2016). In the case of *AtMYB93*, the truncation site was chosen 5 amino acids before the end of the SUB – II motif since an alignment between this sequence and that of its apple homolog (*MdMYB93*) (Legay *et al.*, 2016) shows that their sequences diverge at this residue. The “Series 2” truncations were chosen to exclude the SUB – III motif and its N-terminal polyserine repeat. Polyserine motifs have disordered structure in polypeptides and often act as linking segments between protein domains. The disordered region may also be a proteolytic cleavage site. The “Series 3” truncations falls before a conserved WPD motif found in all three sequences within the SUB – III domain, and the “Series 4”truncation terminates the protein after the SUB – III motif, the last known region conserved across these suberin-associated *MYB*-type transcription factors.

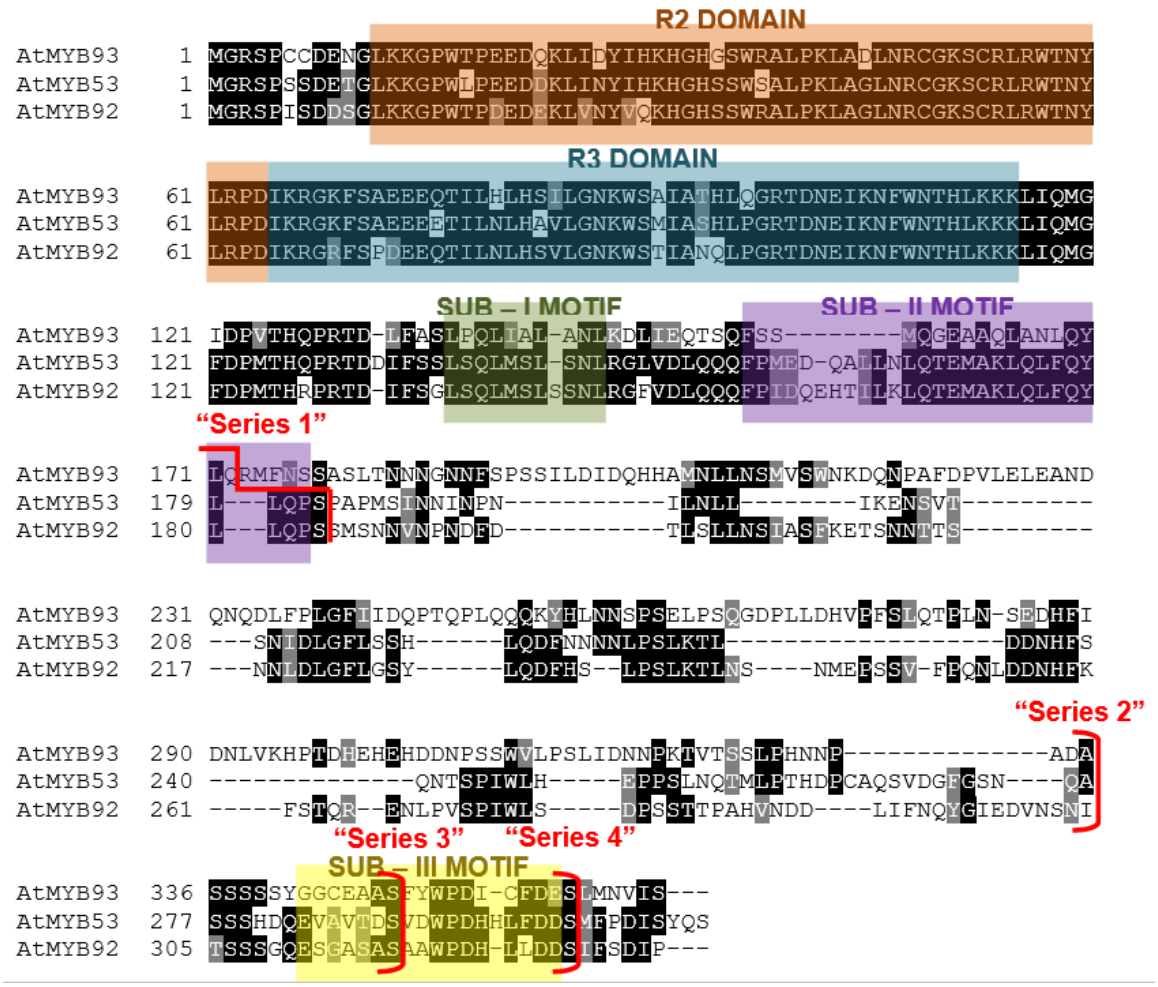


Figure 19: Multiple Sequence Alignment of MYB53, MYB92, and MYB93, indicating the positions of the generated truncated proteins in relation to the R2 and R3 domains, and the conserved SUB-type motifs. The alignment was constructed from protein sequences from TAIR and processed using Clustal Omega and BOXSHADE. Conserved amino acid sequences are indicated by black boxes, while grey boxes refer to residues with conserved chemistry. Conserved SUB domains were annotated using (Lashbrooke *et al.*, 2016) as a reference.

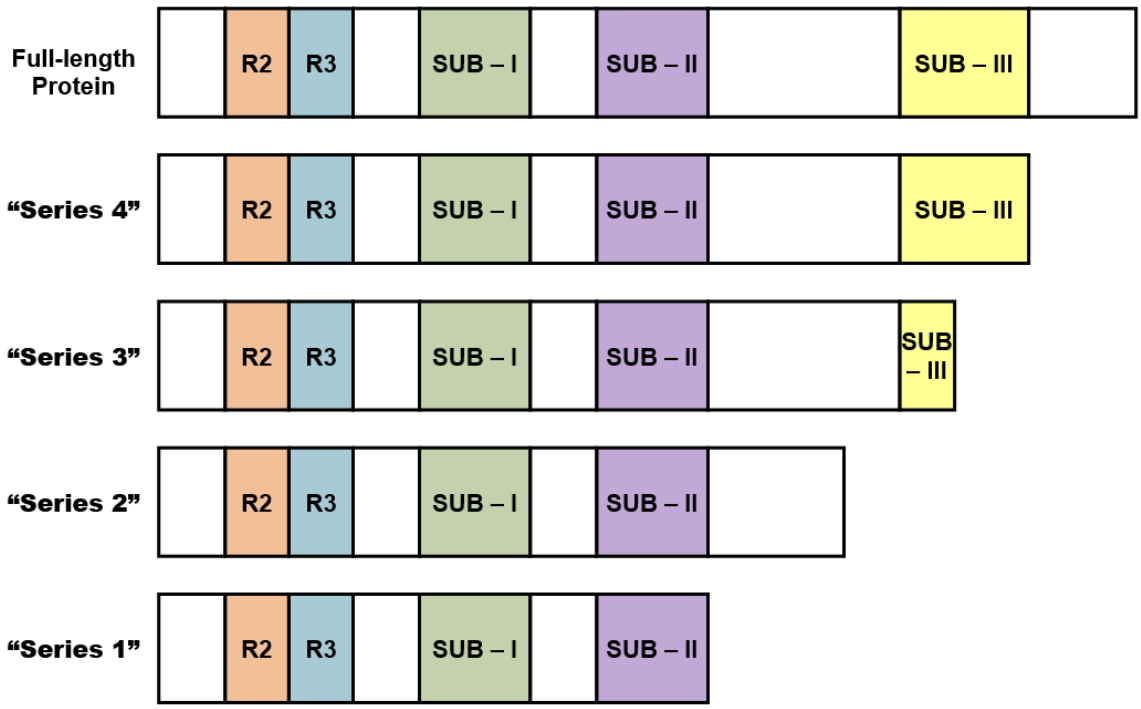


Figure 20: Schematic representation of the MYB truncations that were tested for suberin induction using the transient over-expression assay in *N. benthamiana*. Full-length: *AtMYB53* (1-310), *AtMYB92* (1-334), *AtMYB93* (1-365). "**Series 4"**: *AtMYB53* (1-301), *AtMYB92* (1-328), *AtMYB93* (1-359). **"Series 3"**: *AtMYB53* (1-289), *AtMYB92* (1-317), *AtMYB93* (1-348). **"Series 2":** *AtMYB53* (1-276), *AtMYB92* (1-304), *AtMYB93* (1-335). **"Series 1"**: *AtMYB53* (1-183), *AtMYB92* (1-184), *AtMYB93* (1-172).

To determine if these truncated proteins were functional *in vivo*, the constructs were transiently expressed under the control of a constitutive 35S promoter, in 8-week old *N. benthamiana* leaves (Figures 24 to 35). The leaf areas were excised 6 days after infiltration and chemically examined for ectopic suberin production. The negative empty vector controls pBIN19-p19 and pBIN19 had a low abundance of aliphatic suberin monomers, in accordance with visual inspection of the infiltration sites (no chlorosis). The pBIN19-p19 and pBIN19-empty vector controls were found to have 1465 ± 491 and 1675 ± 187 μg/g dry weight (DW) of total lipid polyesters (including both cutin and suberin), respectively. Among the full-length constructs, AtMYB53 was found to have the most induction of aliphatic suberin (78118 ± 5649 μg lipid polyester/g DW) compared to *AtMYB92* (44664 ± 3455 μg lipid polyester /g DW) and *AtMYB93* (51176 ± 1582 μg lipid polyester /g DW). Compared to their full-length counterparts, the "Series 2" truncations, which lacks SUB-III and part of region in between SUB-II and SUB-III, had suberin production that was reduced by 45%, 31%, and 44%, respectively, for AtMYB53 (42822 ± 3455 μg/g DW), AtMYB92 (30649 ± 3679 μg/g DW), and AtMYB93 (28404 ± 1811 μg/g DW). Each "Series 1" truncation, which lacks everything after SUB-II, was observed to have negligible suberin induction compared to their full-length counterparts (Figure 24).

The general trends described above are mirrored when the major aliphatic suberin-specific monomers were analyzed individually (Figures 25 - 27). The top five most abundant monomers were found to be *trans*-ferulate, 22:0 fatty acid, 18:1 – α,ω-dicarboxylic fatty acid, 18:1 ω – hydroxy fatty acid, and 22:0 ω – hydroxy fatty acid. For each of these major monomers, the "Series 2" truncations had moderate reductions in

abundance compared to their full length counterparts, while the amounts were greatly reduced with each of the "Series 1" truncations.

3.9: Testing of "Series 3" and "Series 4" truncations

Next, two more truncation sets were prepared to specifically evaluate the role of the SUB III domain for transcription factor activity using the same transient over-expression assay (Table 5; Figures 22 and 23). Following the relative success in generating a physiologically active MYB truncation ("Series 2"), the possibility of generating a protein truncation with nearly identical activity as the full-length protein led to the development of the "Series 3" and "Series 4" truncation sets for AtMYB53, AtMYB92, and AtMYB93. The latter, "Series 4", was chosen to determine if the three SUB – type domains are required for full protein function, while the "Series 3" was designed to truncate the protein in between (Δ11000) and (Δ11110).

The results "Series 4" truncations were mixed (Figure 28, top panel). In the case of AtMYB53 (1-301), the effect of truncating the last 9 amino acids had no significant effect on the total amount of suberin deposited (205632 ± 13758 μg/g DW versus 170681 ± 35592 μg/g DW, for AtMYB53 (1-310, full-length) and AtMYB53 (1-301), respectively). However, in the case of AtMYB92 and AtMYB93, their respective "Series 4" truncations were each found to have higher activity than the full-length protein (141000 ± 22186 μg/g DW and 277875 ± 66589 μg/g DW for AtMYB92 (1-334, full-length) and AtMYB92 (1-328), respectively, and 112077 ± 23078 μg/g DW and 199999 ± 33847 μg/g DW, for AtMYB93 (1-365, full length) and AtMYB93 (1-359), respectively).

The next set of truncations, "Series 3", also demonstrated high levels of activity

when transiently expressed in *N. benthamiana* (Figure 28, top panel). The corresponding truncation in this series for AtMYB53 showed about 50% less suberin induction than full-length (205632 ± 13758 μg/g DW versus 103643 ± 18520 μg/g DW, for AtMYB53 (1-310 – full-length) and AtMYB53 (1-289), respectively). In the case of AtMYB92 and AtMYB93, their respective "Series 3" truncations were found to have no significant difference in activity compared to the full-length protein; (141000 ± 22186 μg/g DW, 155645 ± 31061 μg/g DW, for AtMYB92 (1-334 – full length) and AtMYB92 (1-317), respectively); (112077 ± 23078 μg/g DW, 120384 ± 19559 μg/g DW, for AtMYB93 (1-365 – full-length) and AtMYB93 (1-348), respectively).

Given how many suberin monomers exist and the large dataset presented, the observed differences between all of the truncations may be difficult for the reader to visualize from the graphs alone (Figures 29, 31, 33). For this reason, a visual aid (Figures 30, 32, 34, and 35) has been prepared to summarize the statistics shown. The differences between the abundance of each monomer have been plotted relative to the log of the adjusted p-value to highlight which suberin monomers are significantly different for each truncation. The plot area is divided into a blue area ($p < 0.05$) and a red area ($p > 0.05$). By observing which points appear in the red area, it is much easier to recognize which of the corresponding suberin monomers were not found to be statistically different from the corresponding control.

3.10: Comparison of HA-tagged version of full-length MYB proteins with HA-tagged truncations from "Series 2"

The addition of a C-terminal HA epitope tag to the full-length protein or the truncated proteins from "Series 2" (lacking SUB–III), did not significantly affect the

protein compared to untagged versions (Figure 28, bottom panel). The total amount of suberin did not change between any of the pairs of HA-tagged and un-tagged proteins; (205632 ± 13758 µg/g DW versus 171093 ± 20879 µg/g DW un-tagged and tagged AtMYB53 (1-310), respectively), (124793 ± 11958 µg/g DW versus 162379 ± 41083 µg/g DW un-tagged and tagged AtMYB53 (1-276), respectively), (141000 ± 22186 µg/g DW versus 196711 ± 33185 µg/g DW un-tagged and tagged AtMYB92 (1-334), respectively), (254643 ± 44369 µg/g DW versus 166077 ± 23961 µg/g DW un-tagged and tagged AtMYB92 (1-304), respectively), (112077 ± 23078 µg/g DW versus 151496 ± 30302 µg/g DW un-tagged and tagged AtMYB93 (1-365), respectively) (110328 ± 15513 µg/g DW versus 141072 ± 15513 µg/g DW un-tagged and tagged AtMYB93 (1-335), respectively).

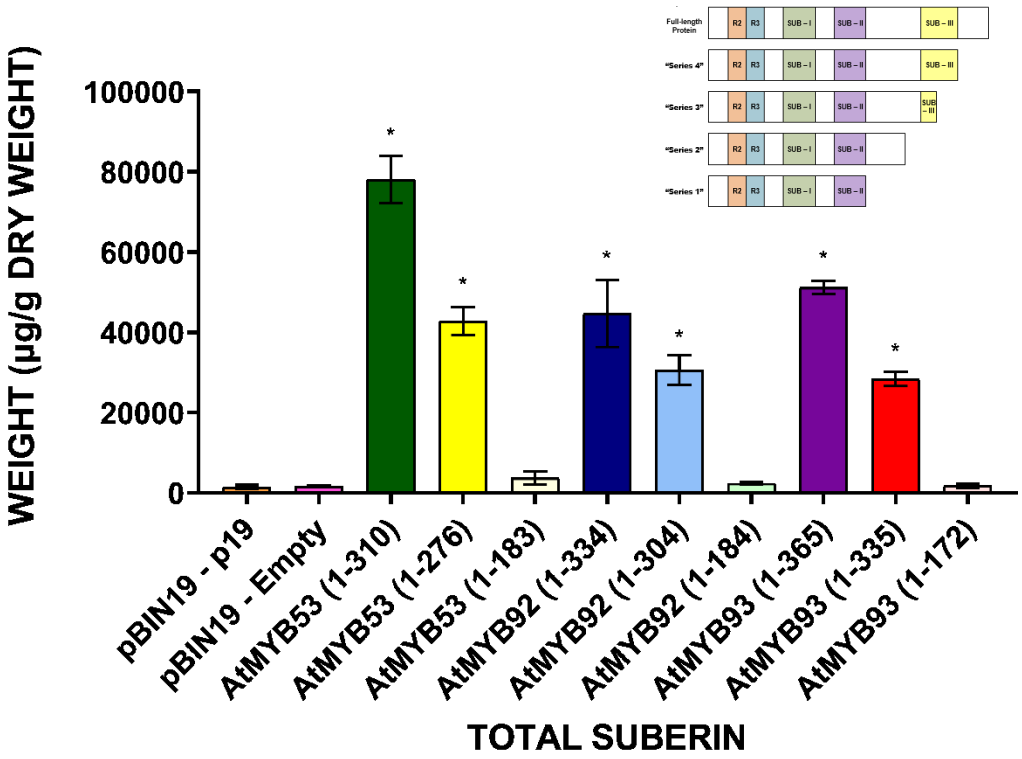
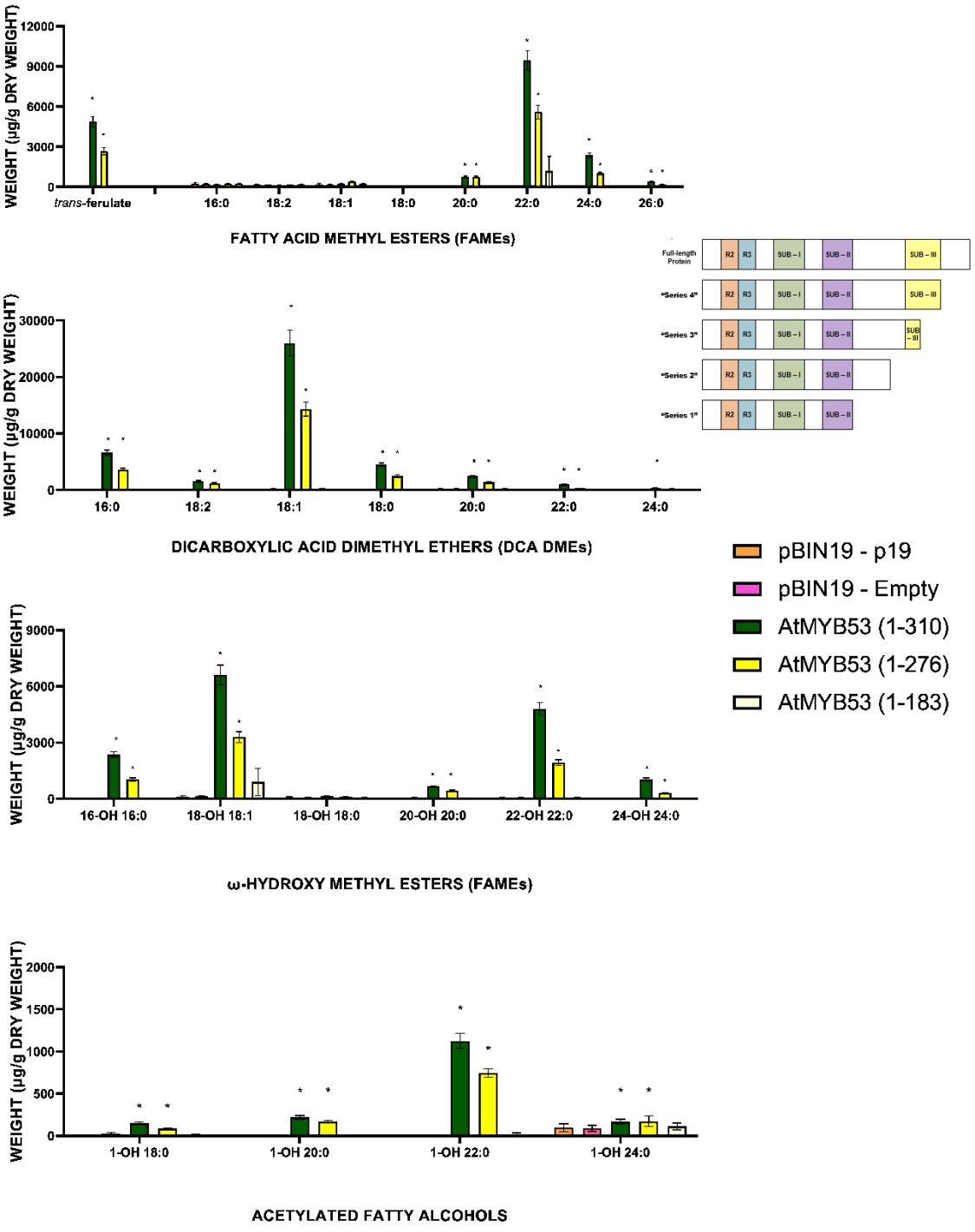


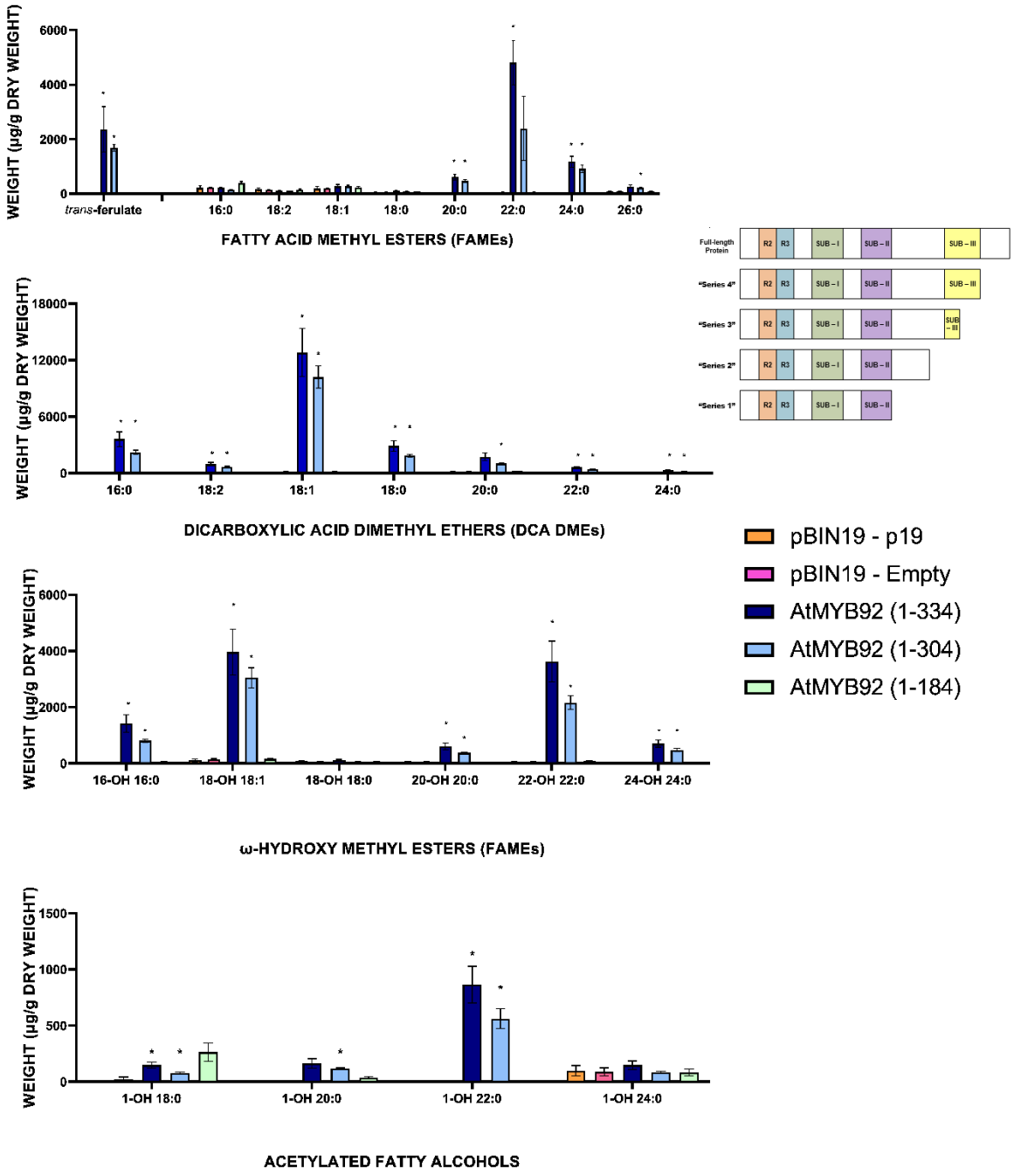
Figure 21: The total amount of aliphatic suberin per gram of dry weight extracted from AtMYB53, AtMYB92, and AtMYB93 truncations (full-length, "Series 1", and "Series 2"). Six-day old infiltration spots were collected from 8-week old transiently transformed *N. benthamiana* leaves with truncated MYB proteins. These constructs were cloned into the pK7WG2D expression vector and transformed into electrocompetent *Agrobacterium tumefaciens* for the Agroinfiltration assay. Alongside the experimental conditions, pBIN-p19 and pBIN-empty vector negative controls were also infiltrated into *N. benthamiana* leaves of the same age. The sample size, n=4, was used for each construct, except for AtMYB92 (1-184), AtMYB92 (1-304), AtMYB93 (1-365), and AtMYB93 (1-335) (n=3). (*) signifies statistical significance between the mutant line compared to the pBIN19-p19 at the $\alpha = 0.05$ level (student's t test; GraphPad Prism 8.1.0).

Figure 22 (next page): The amount of each aliphatic suberin monomer extracted per gram of the dry weight from truncated AtMYB53 proteins (full-length, "Series 1", and "Series 2"). Six-day old infiltration spots were collected from 8-week old transiently transformed *N. benthamiana* leaves with truncated AtMYB53 proteins **(**AtMYB53 (1-310), AtMYB53 (1-276), AtMYB53 (1-183)). These constructs were cloned into the pK7WG2D expression vector and transformed into electrocompetent *Agrobacterium tumefaciens* for the Agroinfiltration assay. Alongside the experimental conditions, pBIN-p19 and pBIN-empty vector negative controls were also infiltrated into *N. benthamiana* leaves of the same age. The sample size, n=4, was used for each construct. (*) signifies statistical significance between the mutant line compared to the pBIN19-p19 at the $\alpha = 0.05$ level (student's t test; GraphPad Prism 8.1.0).


WEIGHT (µg/g DRY WEIGHT)
12000
9000
6000
3000
0
trans-ferulate
16:0
18:2
18:1
18:0
20:0
22:0
24:0
26:0
FATTY ACID METHYL ESTERS (FAMEs)
Full-length Protein
"Series 4"
"Series 3"
"Series 2"
"Series 1"
R2
R3
SUB – I
SUB – II
SUB – III
WEIGHT (µg/g DRY WEIGHT)
30000
20000
10000
0
16:0
18:2
18:1
18:0
20:0
22:0
24:0
DICARBOXYLIC ACID DIMETHYL ETHERS (DCA DMEs)
pBIN19 - p19
pBIN19 - Empty
AtMYB53 (1-310)
AtMYB53 (1-276)
AtMYB53 (1-183)
WEIGHT (µg/g DRY WEIGHT)
9000
6000
3000
0
16-OH 16:0
18-OH 18:1
18-OH 18:0
20-OH 20:0
22-OH 22:0
24-OH 24:0
ω-HYDROXY METHYL ESTERS (FAMEs)
WEIGHT (µg/g DRY WEIGHT)
2000
1500
1000
500
0
1-OH 18:0
1-OH 20:0
1-OH 22:0
1-OH 24:0
ACETYLATED FATTY ALCOHOLS

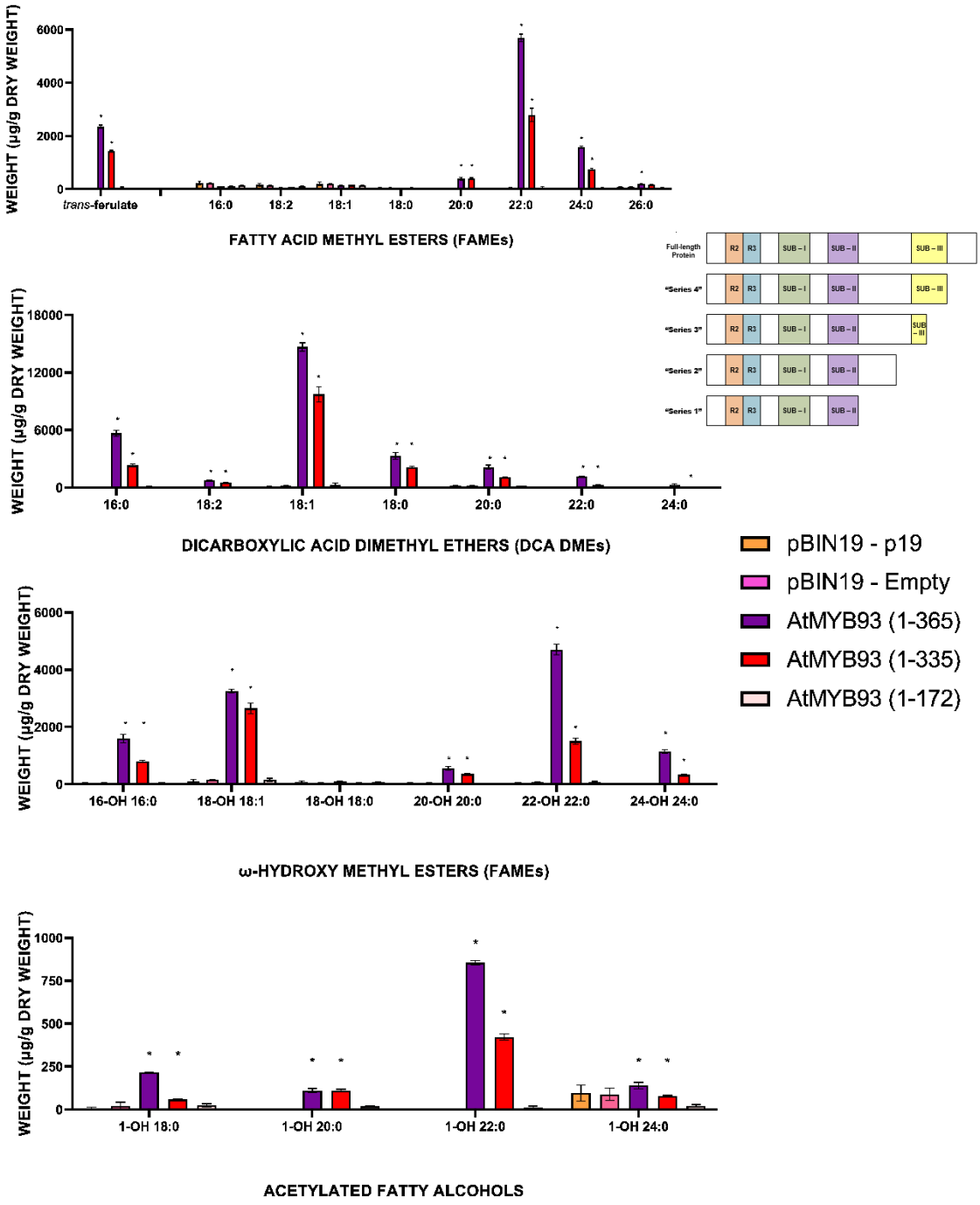
(Figure 22)

Figure 23 (next page): The amount of each aliphatic suberin monomer extracted per gram of the dry weight from truncated AtMYB92 proteins (full-length, "Series 1", and "Series 2"). Six-day old infiltration spots were collected from 8-week old transiently transformed *N. benthamiana* leaves with truncated AtMYB92 proteins (AtMYB92 (1-334), AtMYB92 (1-304), AtMYB92 (1-184)). These constructs were cloned into the pK7WG2D expression vector and transformed into electrocompetent *Agrobacterium tumefaciens* for the Agroinfiltration assay. Alongside the experimental conditions, pBIN-p19 and pBIN-empty vector negative controls were also infiltrated into *N. benthamiana* leaves of the same age. The sample size, n=4, was used for each construct, except for AtMYB92 (1-184), and AtMYB92 (1-304) (n=3). (*) signifies statistical significance between the mutant line compared to the WT at the $\alpha = 0.05$ level (student's t test; GraphPad Prism 8.1.0).


WEIGHT (µg/g DRY WEIGHT)
6000
4000
2000
0
trans-ferulate
16:0
18:2
18:1
18:0
20:0
22:0
24:0
26:0
FATTY ACID METHYL ESTERS (FAMEs)
Full-length Protein
"Series 4"
"Series 3"
"Series 2"
"Series 1"
R2
R3
SUB – I
SUB – II
SUB – III
WEIGHT (µg/g DRY WEIGHT)
18000
12000
6000
0
16:0
18:2
18:1
18:0
20:0
22:0
24:0
DICARBOXYLIC ACID DIMETHYL ETHERS (DCA DMEs)
pBIN19 - p19
pBIN19 - Empty
AtMYB92 (1-334)
AtMYB92 (1-304)
AtMYB92 (1-184)
WEIGHT (µg/g DRY WEIGHT)
6000
4000
2000
0
16-OH 16:0
18-OH 18:1
18-OH 18:0
20-OH 20:0
22-OH 22:0
24-OH 24:0
ω-HYDROXY METHYL ESTERS (FAMEs)
WEIGHT (µg/g DRY WEIGHT)
1500
1000
500
0
1-OH 18:0
1-OH 20:0
1-OH 22:0
1-OH 24:0
ACETYLATED FATTY ALCOHOLS

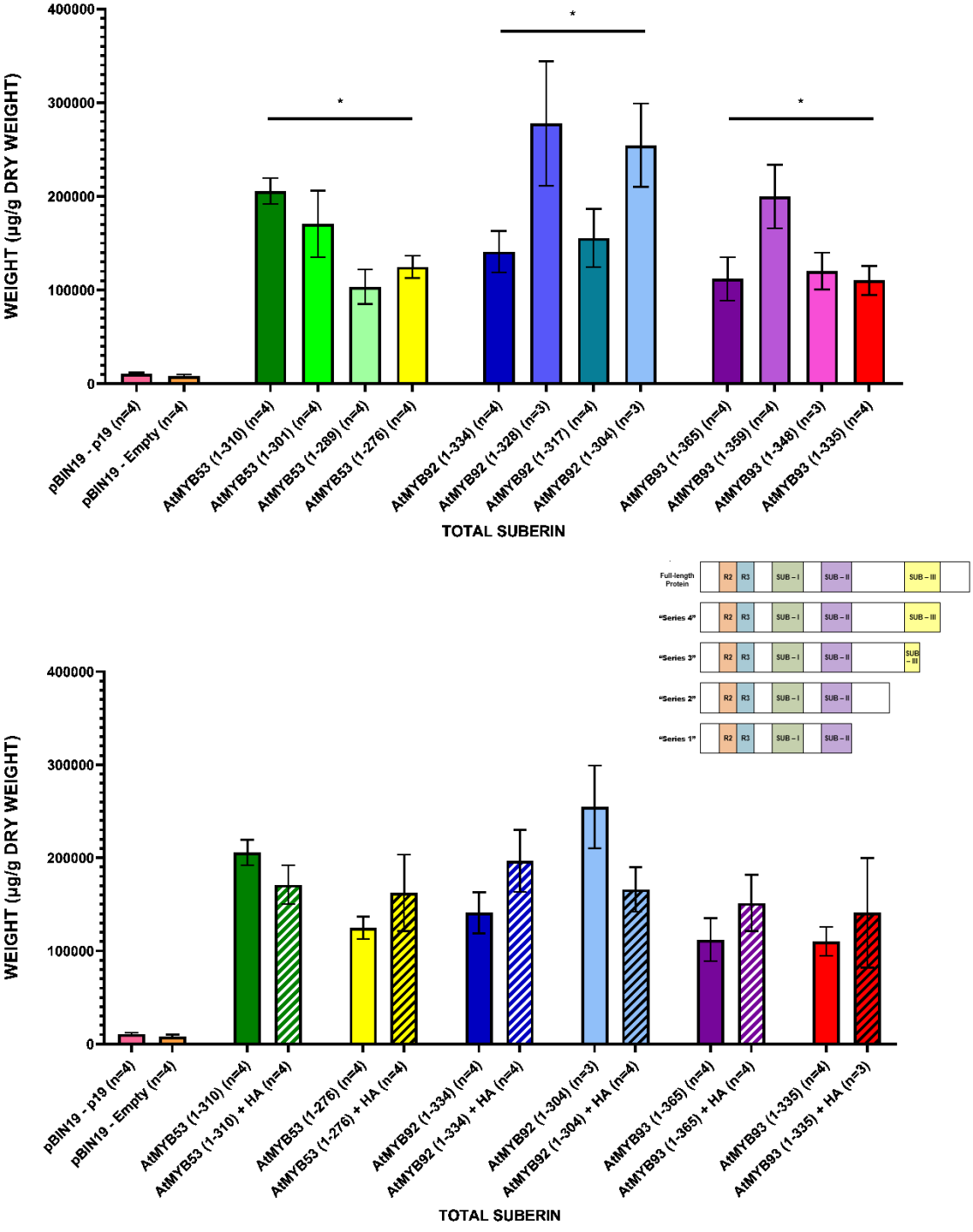
(Figure 23)

Figure 24 (next page): The amount of each aliphatic suberin monomer extracted per gram of the dry weight from truncated AtMYB93 proteins (full-length, "Series 1", and "Series 2"). Six-day old infiltration spots were collected from 8-week old transiently transformed *N. benthamiana* leaves with truncated AtMYB93 proteins (AtMYB93 (1-365), AtMYB93 (1-335), AtMYB93 (1-176)). These constructs were cloned into the pK7WG2D expression vector and transformed into electrocompetent *Agrobacterium tumefaciens* for the Agroinfiltration assay. Alongside the experimental conditions, pBIN-p19 and pBIN-empty vector negative controls were also infiltrated into *N. benthamiana* leaves of the same age. The sample size, n=4, was used for each construct, except for AtMYB93 (1-365), and AtMYB93 (1-335) (n=3). (*) signifies statistical significance between the mutant line compared to the pBIN19-p19 at the $\alpha = 0.05$ level (student's t test; GraphPad Prism 8.1.0).


WEIGHT (µg/g DRY WEIGHT)
6000
4000
2000
0
trans-ferulate
16:0
18:2
18:1
18:0
20:0
22:0
24:0
26:0
FATTY ACID METHYL ESTERS (FAMEs)
Full-length Protein
R2
R3
SUB – I
SUB – II
SUB – III
"Series 4"
"Series 3"
"Series 2"
"Series 1"
WEIGHT (µg/g DRY WEIGHT)
18000
12000
6000
0
16:0
18:2
18:1
18:0
20:0
22:0
24:0
DICARBOXYLIC ACID DIMETHYL ETHERS (DCA DMEs)
pBIN19 - p19
pBIN19 - Empty
AtMYB93 (1-365)
AtMYB93 (1-335)
AtMYB93 (1-172)
WEIGHT (µg/g DRY WEIGHT)
6000
4000
2000
0
16-OH 16:0
18-OH 18:1
18-OH 18:0
20-OH 20:0
22-OH 22:0
24-OH 24:0
ω-HYDROXY METHYL ESTERS (FAMEs)
WEIGHT (µg/g DRY WEIGHT)
1000
750
500
250
0
1-OH 18:0
1-OH 20:0
1-OH 22:0
1-OH 24:0
ACETYLATED FATTY ALCOHOLS

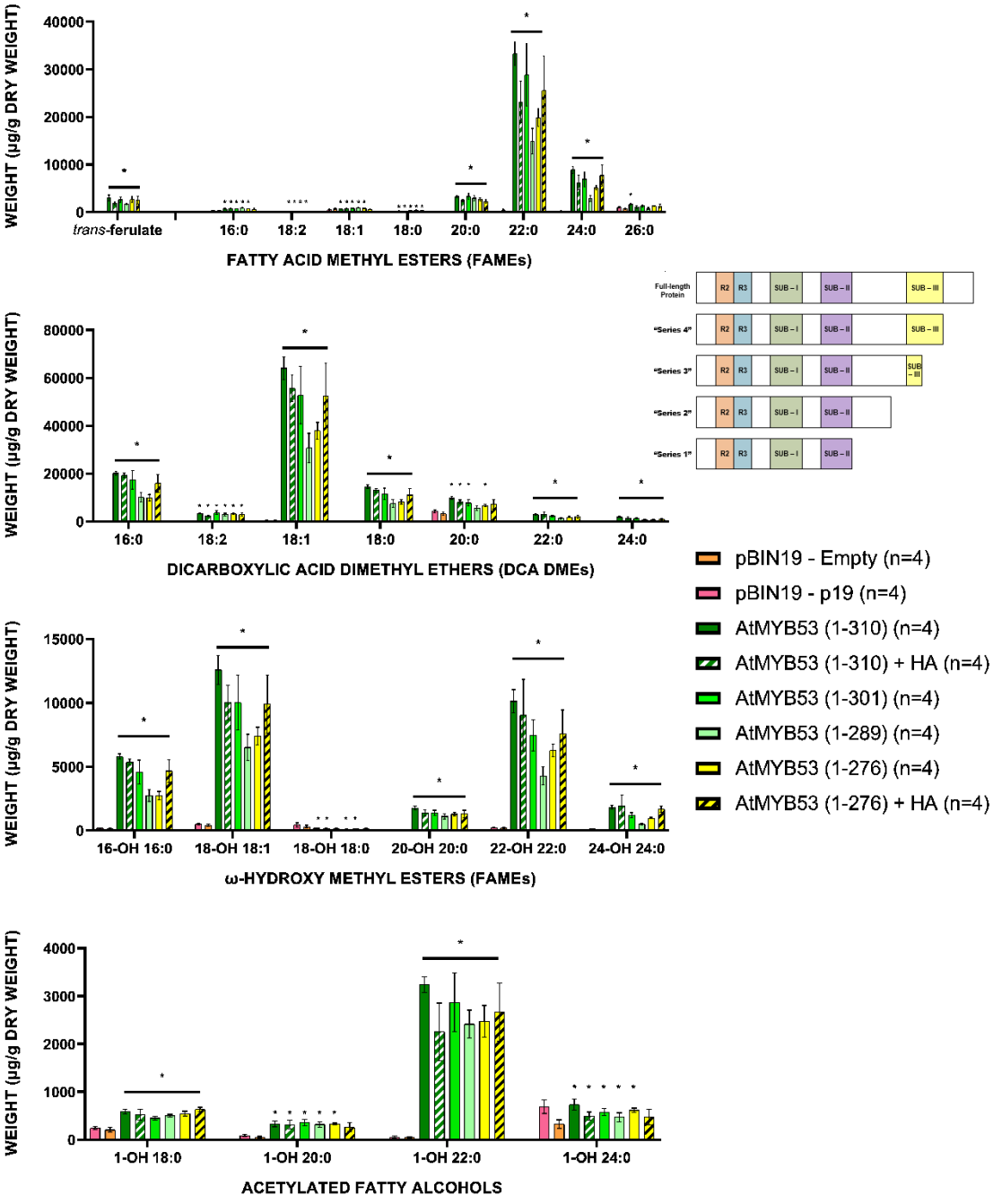
(Figure 24)

Figure 25 (next page): The total amount of aliphatic suberin per gram of dry weight extracted from 6-day old infiltration spots from 8-week old transiently transformed *N. benthamiana* leaves with MYB protein truncation constructs (full length, "Series 2", "Series 3", and "Series 4"). The corresponding truncated protein constructs for the transcription factors AtMYB53, AtMYB92, and AtMYB93 were cloned and transformed into electrocompetent *A. tumefaciens* for the infiltration assay, and total suberin was measured by GC-FID. Alongside the experimental conditions, pBIN19-p19 (pink) and pBIN19-empty vector (orange) negative controls were also infiltrated into leaves of *N. benthamiana* plants of the same age. **Top panel:** Total aliphatic suberin for each un-tagged protein presented in order of sequential deletions, from full length, "Series 2", "Series 3", and "Series 4" truncations. **Bottom panel:** Total aliphatic suberin measured for constructs lacking or containing the HA tag on the C-terminus of each full-length and "Series 2" truncation. The sample size was n=4 for each construct, except for "AtMYB92 (1-328)", "AtMYB92 (1-304)", "AtMYB93 (1-348)", and "AtMYB93 (1-335) + HA" (n=3 for all). (*) signifies statistical significance between the mutant line compared to the pBIN19-p19 at the $\alpha = 0.05$ level (student's t test; GraphPad Prism 8.1.0).


WEIGHT (µg/g DRY WEIGHT)
400000
300000
200000
100000
0
*
*
*
pBIN19 - p19 (n=4)
pBIN19 - Empty (n=4)
AtMYB53 (1-310) (n=4)
AtMYB53 (1-301) (n=4)
AtMYB53 (1-289) (n=4)
AtMYB53 (1-276) (n=4)
AtMYB92 (1-334) (n=4)
AtMYB92 (1-328) (n=3)
AtMYB92 (1-317) (n=4)
AtMYB92 (1-304) (n=3)
AtMYB93 (1-365) (n=4)
AtMYB93 (1-359) (n=4)
AtMYB93 (1-348) (n=3)
AtMYB93 (1-335) (n=4)
TOTAL SUBERIN
Full-length Protein
"Series 4"
"Series 3"
"Series 2"
"Series 1"
R2
R3
SUB – I
SUB – II
SUB – III
WEIGHT (µg/g DRY WEIGHT)
400000
300000
200000
100000
0
pBIN19 - p19 (n=4)
pBIN19 - Empty (n=4)
AtMYB53 (1-310) (n=4)
AtMYB53 (1-310) + HA (n=4)
AtMYB53 (1-276) (n=4)
AtMYB53 (1-276) + HA (n=4)
AtMYB92 (1-334) (n=4)
AtMYB92 (1-334) + HA (n=4)
AtMYB92 (1-304) (n=3)
AtMYB92 (1-304) + HA (n=4)
AtMYB93 (1-365) (n=4)
AtMYB93 (1-365) + HA (n=4)
AtMYB93 (1-335) (n=4)
AtMYB93 (1-335) + HA (n=3)
TOTAL SUBERIN

(Figure 25)

Figure 26 (next page): The breakdown of individual suberin monomer content per gram of the dry weight extracted from AtMYB53 protein truncation constructs (full length, "Series 2", "Series 3", and "Series 4"). Six-day old infiltration spots were collected from 8-week old transiently transformed *N. benthamiana* leaves with truncated MYB proteins. The corresponding constructs for the transcription factors *AtMYB53* were cloned and transformed into electrocompetent *A. tumefaciens* for the infiltration assay, and total suberin was measured by GC-FID. Alongside the experimental conditions, pBIN19-p19 (pink) and pBIN19-empty vector (orange) negative controls were also infiltrated into leaves of *N. benthamiana* leaves of the same age. The sample size was n=4 for each construct. (*) signifies statistical significance between the mutant line compared to the pBIN19-p19 at the $\alpha = 0.05$ level (student's t test; GraphPad Prism 8.1.0).


WEIGHT (µg/g DRY WEIGHT)
40000
30000
20000
10000
0
trans-ferulate
16:0
18:2
18:1
18:0
20:0
22:0
24:0
26:0
FATTY ACID METHYL ESTERS (FAMEs)
80000
60000
40000
20000
DICARBOXYLIC ACID DIMETHYL ETHERS (DCA DMEs)
15000
10000
5000
16-OH 16:0
18-OH 18:1
18-OH 18:0
20-OH 20:0
22-OH 22:0
24-OH 24:0
ω-HYDROXY METHYL ESTERS (FAMEs)
4000
3000
2000
1000
1-OH 18:0
1-OH 20:0
1-OH 22:0
1-OH 24:0
ACETYLATED FATTY ALCOHOLS
Full-length Protein
"Series 4"
"Series 3"
"Series 2"
"Series 1"
R2
R3
SUB – I
SUB – II
SUB – III
pBIN19 - Empty (n=4)
pBIN19 - p19 (n=4)
AtMYB53 (1-310) (n=4)
AtMYB53 (1-310) + HA (n=4)
AtMYB53 (1-301) (n=4)
AtMYB53 (1-289) (n=4)
AtMYB53 (1-276) (n=4)
AtMYB53 (1-276) + HA (n=4)

(Figure 26)

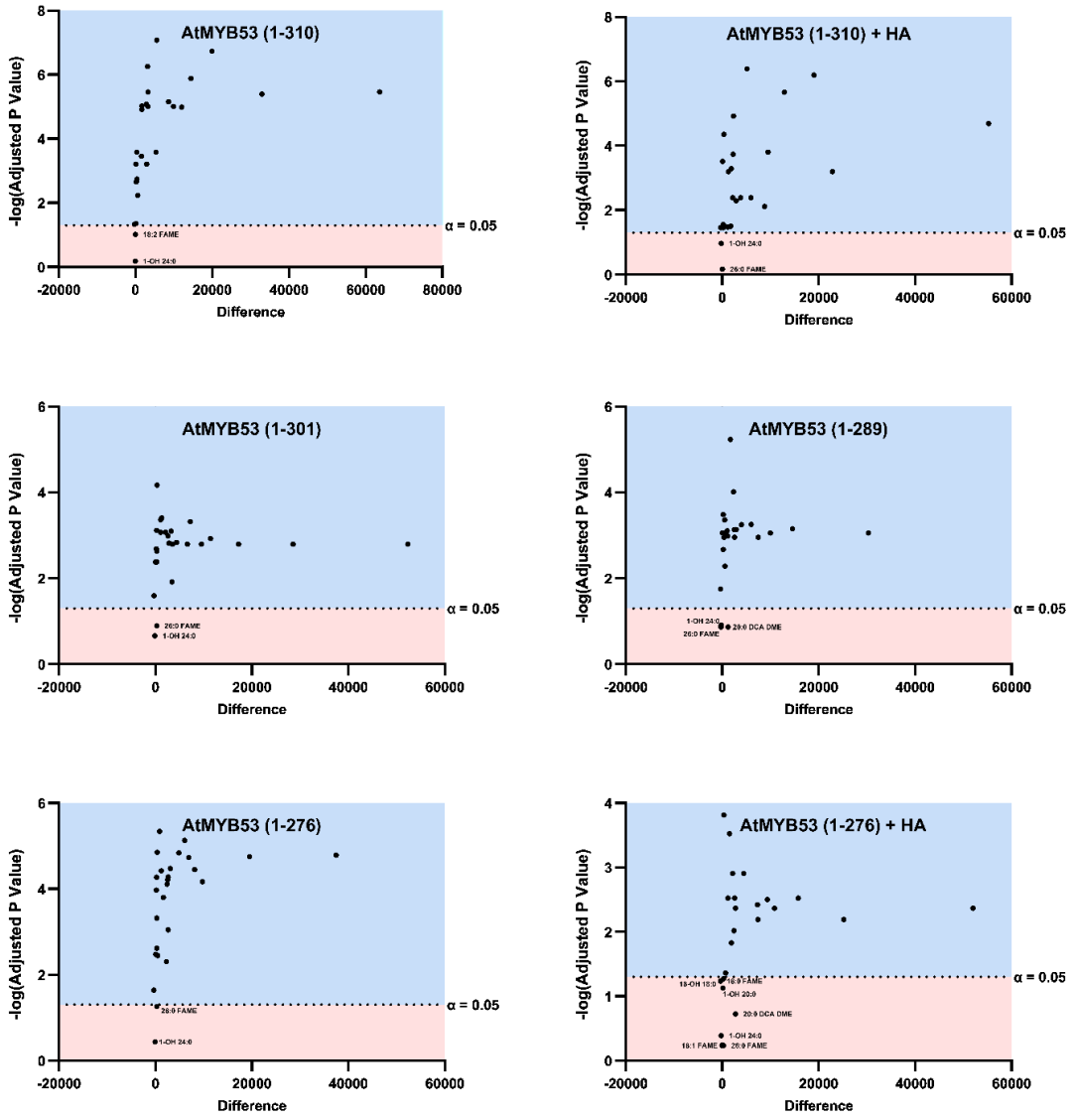
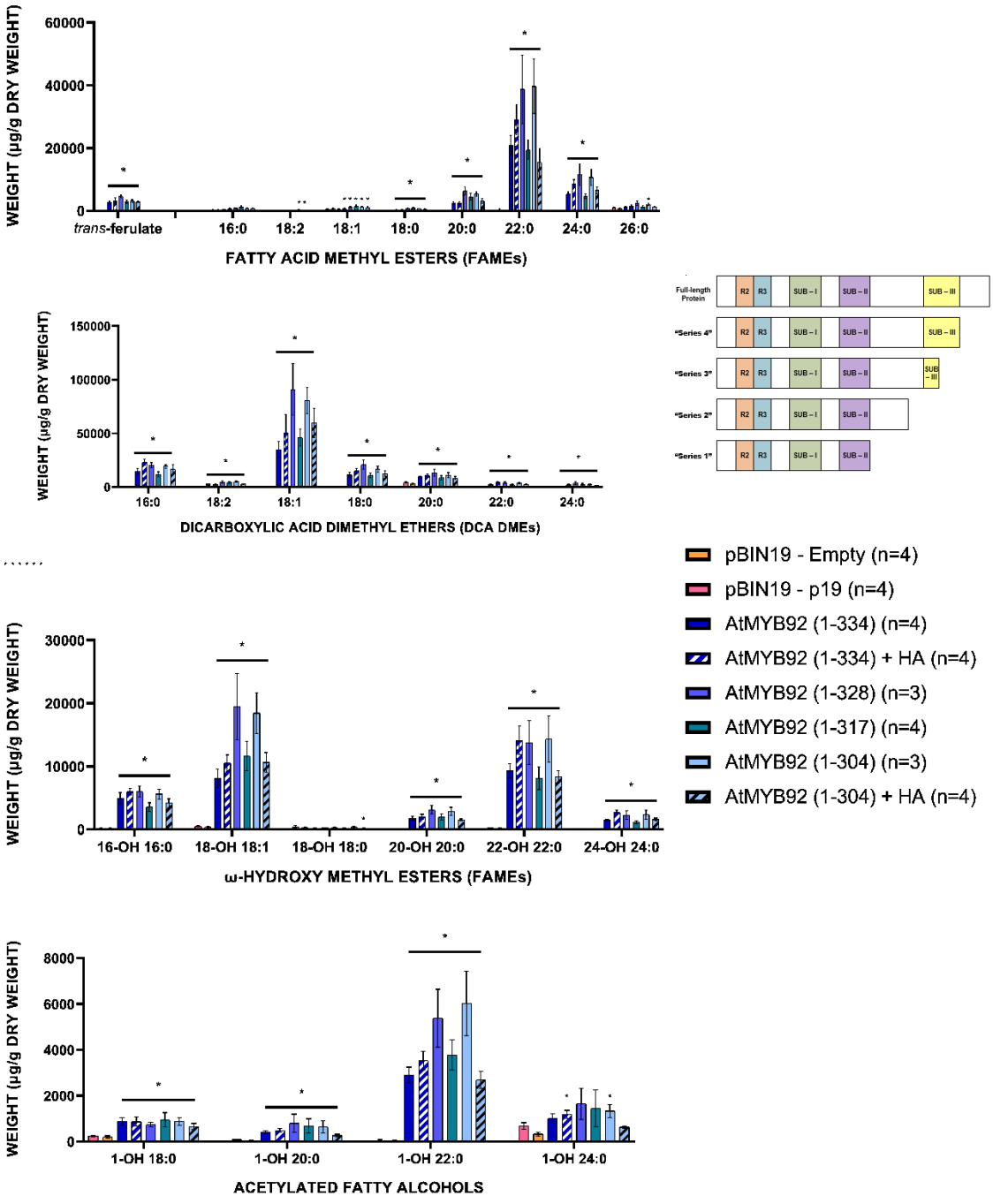


Figure 27: Distributions of adjusted p-values comparing the relative abundance of each suberin monomer type measured by GC-FID for each truncation, relative to pBIN19-p19 negative control. The HA-tagged versions of the constructs are also represented, where applicable. The p-values were obtained from t-tests using GraphPad Prism 8.1.0 software with the statistical significance determined using the Holm-Sidak method, with cut-off value $\alpha = 0.05$. The -log transformed p-values were plotted relative to the observed suberin monomer differences between the treatments. Statistically significant results are plotted in the blue area while results that are not statistically significant are plotted in the red area.

Figure 28 (next page): The breakdown of individual suberin monomer content per gram of the dry weight extracted from AtMYB92 protein truncation constructs (full length, "Series 2", "Series 3", and "Series 4"). Six-day old infiltration spots were collected from 8-week old transiently transformed *N. benthamiana* leaves with truncated MYB proteins. The corresponding constructs for the transcription factors AtMYB92 were cloned and transformed into electrocompetent *A. tumefaciens* for the infiltration assay, and total suberin was measured by GC-FID. Alongside the experimental conditions, pBIN19-p19 (pink) and pBIN19-empty vector (orange) negative controls were also infiltrated into leaves of *N. benthamiana* leaves of the same age. The sample size was n=4 for each construct, except for "AtMYB92 (1-328)", and "AtMYB92 (1-304) (n=3 for both). (*) signifies statistical significance between the mutant line compared to the pBIN19-p19 at the $\alpha = 0.05$ level (student's t test; GraphPad Prism 8.1.0).


WEIGHT (µg/g DRY WEIGHT)
60000
40000
20000
0
trans-ferulate
16:0
18:2
18:1
18:0
20:0
22:0
24:0
26:0
FATTY ACID METHYL ESTERS (FAMEs)
150000
100000
50000
0
DICARBOXYLIC ACID DIMETHYL ETHERS (DCA DMEs)
Full-length Protein
"Series 4"
"Series 3"
"Series 2"
"Series 1"
R2
R3
SUB – I
SUB – II
SUB – III
pBIN19 - Empty (n=4)
pBIN19 - p19 (n=4)
AtMYB92 (1-334) (n=4)
AtMYB92 (1-334) + HA (n=4)
AtMYB92 (1-328) (n=3)
AtMYB92 (1-317) (n=4)
AtMYB92 (1-304) (n=3)
AtMYB92 (1-304) + HA (n=4)
30000
20000
10000
0
16-OH 16:0
18-OH 18:1
18-OH 18:0
20-OH 20:0
22-OH 22:0
24-OH 24:0
ω-HYDROXY METHYL ESTERS (FAMEs)
8000
6000
4000
2000
0
1-OH 18:0
1-OH 20:0
1-OH 22:0
1-OH 24:0
ACETYLATED FATTY ALCOHOLS

(Figure 28)

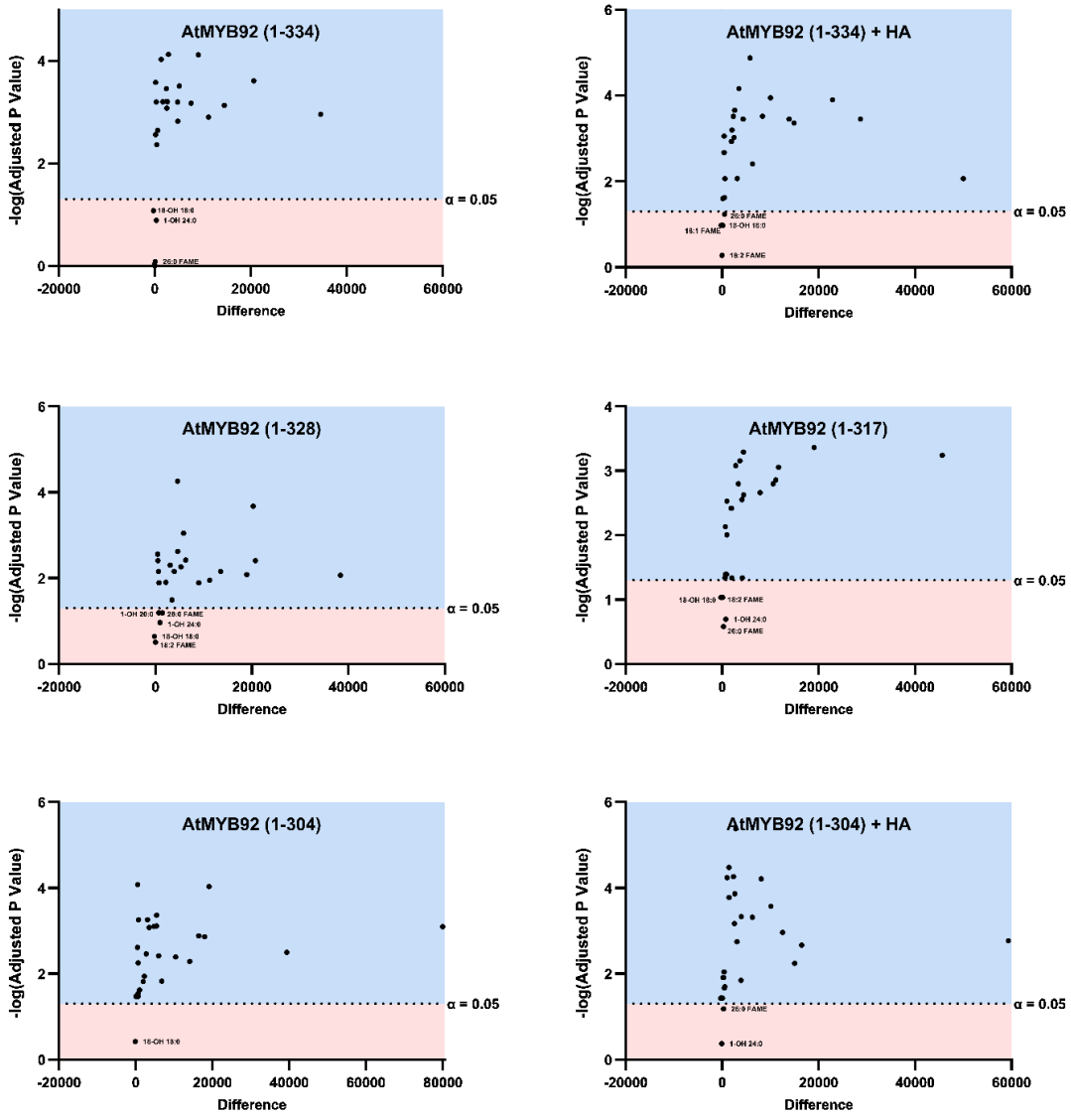
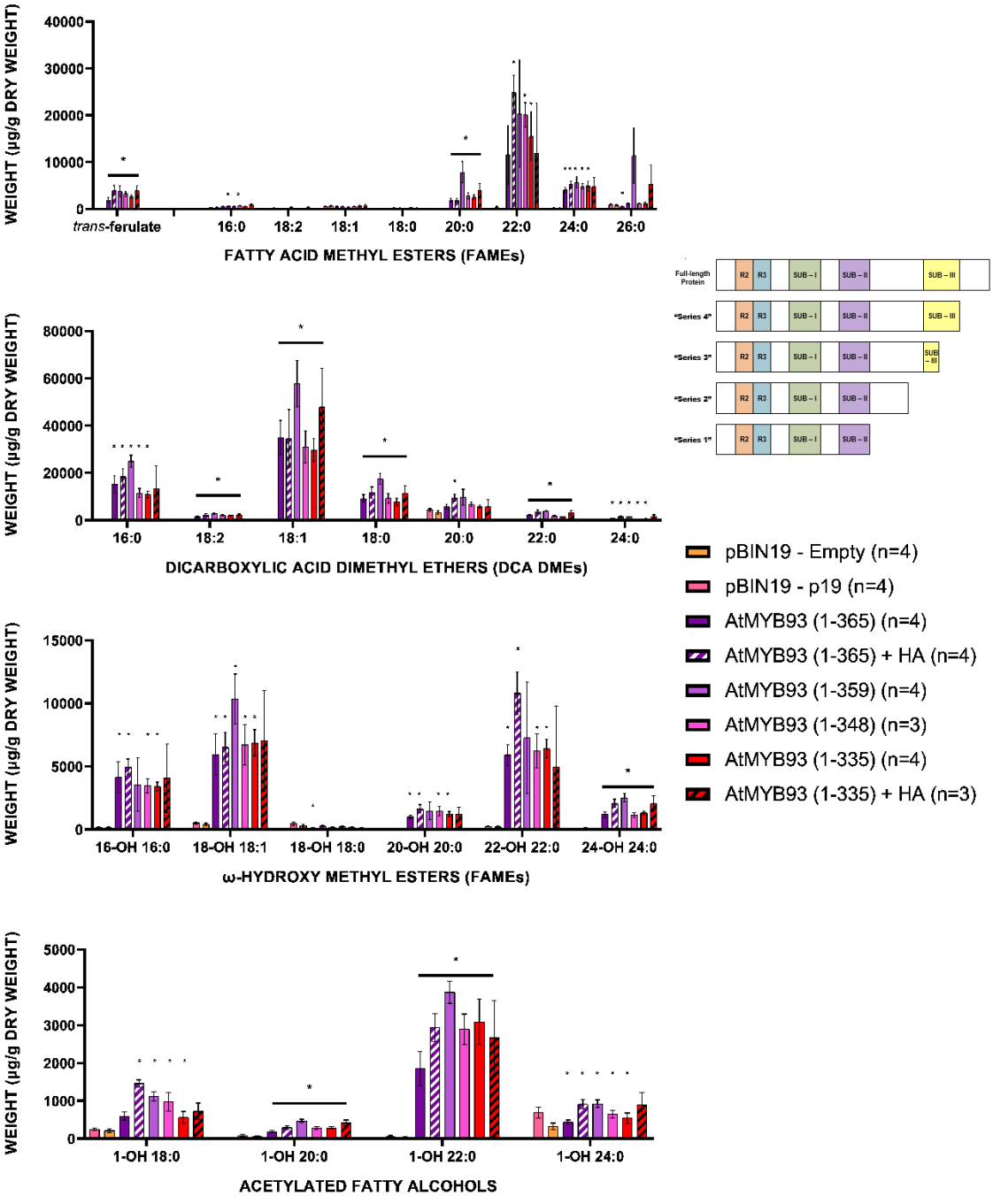


Figure 29: Distributions of adjusted p-values comparing the relative abundance of each suberin monomer type measured by GC-FID for each truncation, relative to pBIN19-p19 negative control. The HA-tagged versions of the constructs are also represented, where applicable. The p-values were obtained from t-tests using GraphPad Prism 8.1.0 software with the statistical significance determined using the Holm-Sidak method, with cut-off value $\alpha = 0.05$. The -log transformed p-values were plotted relative to the observed suberin monomer differences between the treatments. Statistically significant results are plotted in the blue area while results that are not statistically significant are plotted in the red area.

Figure 30 (next page): The breakdown of individual suberin monomer content per gram of the dry weight extracted from AtMYB93 protein truncation constructs (full length, "Series 2", "Series 3", and "Series 4"). Six-day old infiltration spots were collected from 8-week old transiently transformed *N. benthamiana* leaves with truncated MYB proteins. The corresponding constructs for the transcription factors *AtMYB93* were cloned and transformed into electrocompetent *A. tumefaciens* for the infiltration assay, and total suberin was measured by GC-FID. Alongside the experimental conditions, pBIN19-p19 (pink) and pBIN19-empty vector (orange) negative controls were also infiltrated into leaves of *N. benthamiana* leaves of the same age. The sample size was n=4 for each construct, except for "AtMYB93 (1-348)", and "AtMYB93 (1-335) + HA" (n=3 for both). (*) signifies statistical significance between the mutant line compared to the pBIN19-p19 at the $\alpha = 0.05$ level (student's t test; GraphPad Prism 8.1.0).


WEIGHT (µg/g DRY WEIGHT)
40000
30000
20000
10000
0
trans-ferulate
16:0
18:2
18:1
18:0
20:0
22:0
24:0
26:0
FATTY ACID METHYL ESTERS (FAMEs)
Full-length Protein
"Series 4"
"Series 3"
"Series 2"
"Series 1"
R2
R3
SUB – I
SUB – II
SUB – III
WEIGHT (µg/g DRY WEIGHT)
80000
60000
40000
20000
0
16:0
18:2
18:1
18:0
20:0
22:0
24:0
DICARBOXYLIC ACID DIMETHYL ETHERS (DCA DMEs)
pBIN19 - Empty (n=4)
pBIN19 - p19 (n=4)
AtMYB93 (1-365) (n=4)
AtMYB93 (1-365) + HA (n=4)
AtMYB93 (1-359) (n=4)
AtMYB93 (1-348) (n=3)
AtMYB93 (1-335) (n=4)
AtMYB93 (1-335) + HA (n=3)
WEIGHT (µg/g DRY WEIGHT)
15000
10000
5000
0
16-OH 16:0
18-OH 18:1
18-OH 18:0
20-OH 20:0
22-OH 22:0
24-OH 24:0
ω-HYDROXY METHYL ESTERS (FAMEs)
WEIGHT (µg/g DRY WEIGHT)
5000
4000
3000
2000
1000
0
1-OH 18:0
1-OH 20:0
1-OH 22:0
1-OH 24:0
ACETYLATED FATTY ALCOHOLS

(Figure 30)

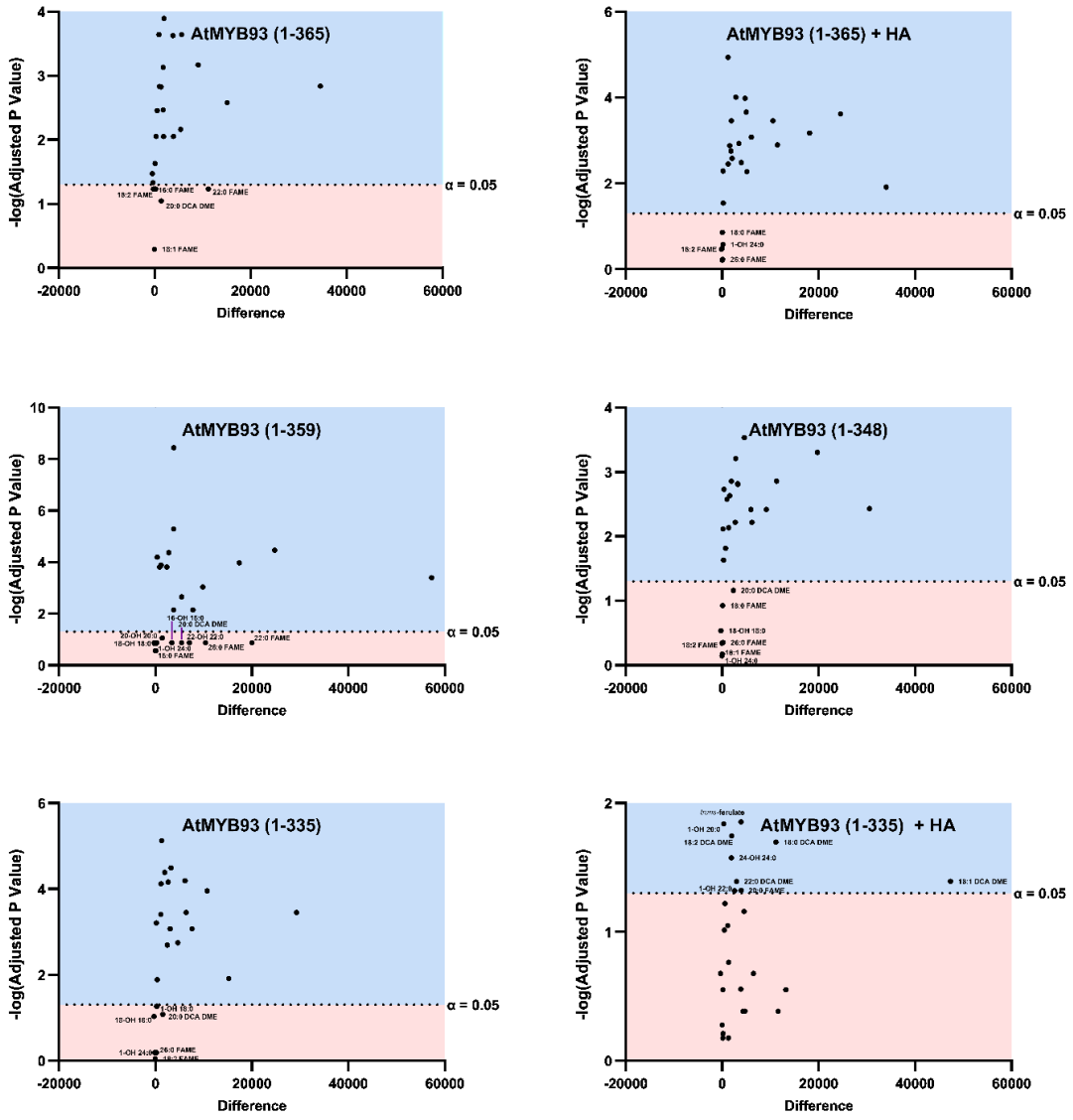


Figure 31: Distributions of adjusted p-values comparing the relative abundance of each suberin monomer type measured by GC-FID for each AtMYB93 truncation, relative to pBIN19-p19 negative control. The HA-tagged versions of the constructs are also represented, where applicable. The p-values were obtained from t-tests using GraphPad Prism 8.1.0 software with the statistical significance determined using the Holm-Sidak method, with cut-off value $\alpha = 0.05$. The -log transformed p-values were plotted relative to the observed suberin monomer differences between the treatments. Statistically significant results are plotted in the blue area while results that are not statistically significant are plotted in the red area.

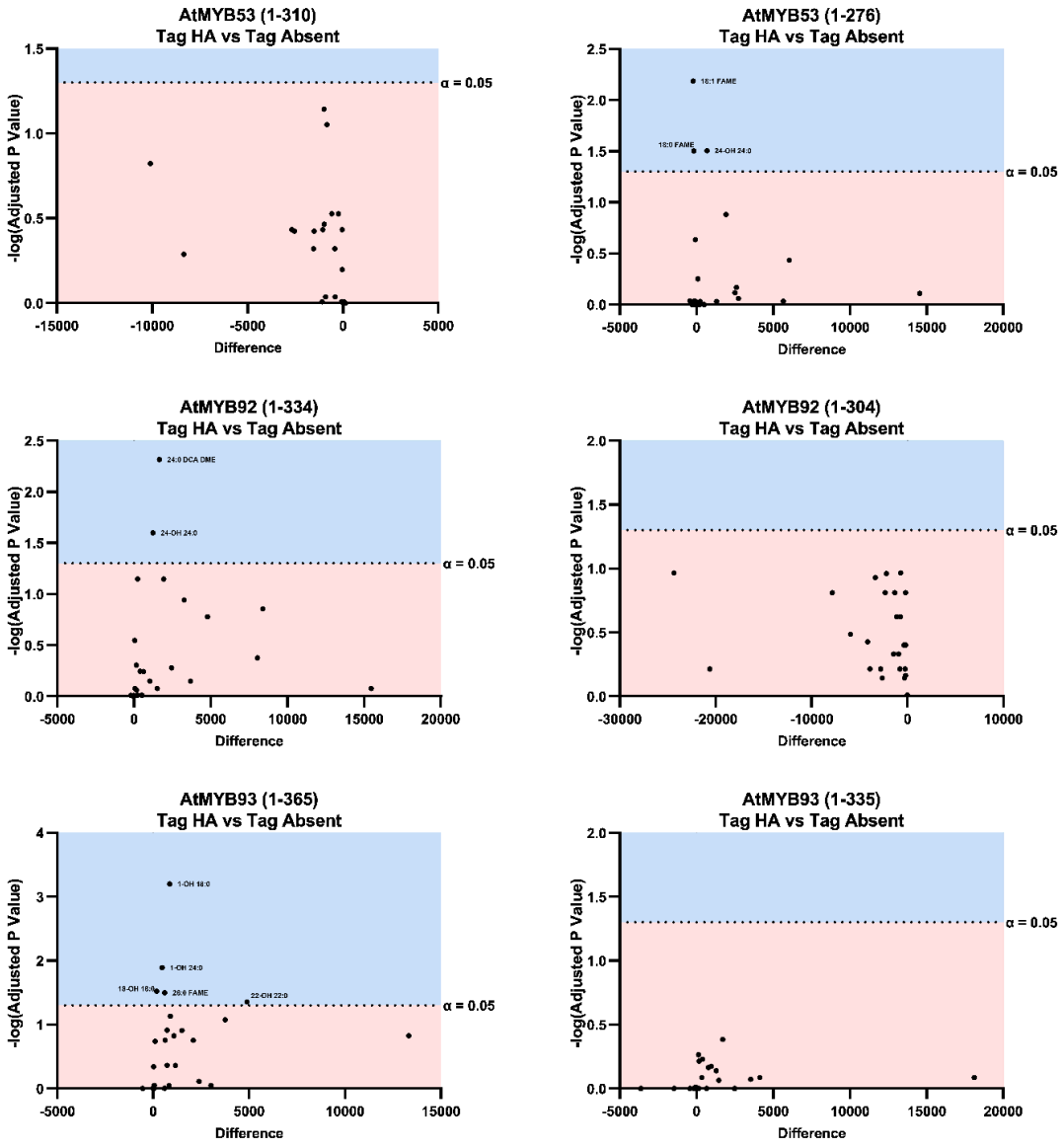


Figure 32: Distributions of adjusted p-values comparing the relative abundance of each suberin monomer type measured by GC-FID from the transient over-expression of HA-tagged versus un-tagged versions of full length and "Series 2" MYB truncations in *N. benthamiana* leaves. The p-values were obtained from t-tests using GraphPad Prism 8.1.0 software with the statistical significance determined using the Holm-Sidak method, with cut-off value $\alpha = 0.05$. The -log transformed p-values were plotted relative to the observed suberin monomer differences between the treatments. Statistically significant results are plotted in the blue area while results that are not statistically significant are plotted in the red area.

CHAPTER 4 – DISCUSSION

Suberin is a complex and protective heteropolymer that is deposited on the inner face of the primary cell walls of select plant tissues (Schreiber, 2010). A precise understanding of the regulatory mechanisms governing suberin production is important to understand its biosynthesis and functions. However, our current knowledge of the transcriptional control of suberin biosynthetic genes is limited, other than knowing that R2R3 MYB-type transcription factors are involved. Also, many key players in the biosynthetic and transport pathways remain to be discovered and characterizing master regulators of suberin deposition may reveal the missing genes. Here, an in-depth transcriptomic analysis of the suberin regulators *MYB53, MYB92,* and *MYB93* from *Arabidopsis thaliana* was conducted using RNA-seq on mutant triple knockout (TKO) lines and a steroid-inducible over-expression (OE) line for *AtMYB53*.

4.1: Root extracts shows mutant suberin phenotype in *myb53 myb92 myb93* triple knockout lines

In agreement with previous research using the *myb53 myb92 myb93* triple knock-out lines, there were large reductions in total lipid polyesters in the mutant lines compared to wild-type (WT). Reductions by 73.8% and 76.6% were observed in the two independent TKO mutant alleles, which is comparable to the average 72% overall reduction in suberin that was reported by Hu (2018). Despite knocking-out all three putative master regulators of suberin in normal root development, there was still some suberin remaining. It is worth noting, however, that a decrease of this magnitude of suberin in *Arabidopsis thaliana* roots is unprecedented. Nonetheless, there must be

additional transcription factors, partially redundant with these three MYB factors, which activate downstream suberin biosynthetic genes to account for the remaining amount of root suberin. It could be the case that these other transcription factors are activated by the initial osmotic stress caused by the absence of suberin within the root to compensate. A complete loss of suberin biosynthesis in roots may also have lethal consequences due to major defects in nutrient acquisition and water relations.

In the previous line of research presented by Hu (2018), it was found that suberin monomers of every chemical class were decreased in both TKO lines. This included 20:0 to 24:0 fatty acids, 16:0 to 22:0 dicarboxylic fatty acids, 16:0 to 24:0 ω-hydroxy fatty acids, 18:0 to 22:0 primary fatty alcohols, and ferulate (reported as *cis*- and *trans*-ferulate isomers combined). The current work also found similar across-the-board decreases, with the exception of *trans*-ferulate. Compared to the 37 – 51% decrease in total ferulate previously reported, I observed no significant decrease in the *trans*-ferulate configuration. While this could be because the *cis*-ferulate configuration was not detected here, it is commonly known that the *trans*-ferulate configuration is vastly more abundant in suberin than its *cis*- ferulate counterpart. Furthermore, the enzyme ASFT (as also known as RWP1 or At5g41010), which catalyzes the acyl transfer of feruloyl-CoA for form alkyl ferulates, was found in the RNA-seq analysis to not be differentially regulated in either TKO allele. Previous studies concluded that the *asft* knock-out line in *Arabidopsis thaliana* results in the absence of ferulate in aliphatic suberin (Molina *et al*., 2009). Therefore, it is unlikely that the production of alkyl ferulates via ASFT is under the sole transcriptional control of *AtMYB53, AtMYB92,* and *AtMYB93*. On the other hand, there was a significant up-regulation of two splice variants of the *ASFT* gene (At5g41010.1 and

At5g41010.2) at both 6 and 15 hours post-treatment using the *AtMYB53*-OE steroid-inducible line. This indicates that the transcriptional regulation of the *ASFT* gene could in part be provided by *AtMYB53* (also possibly *AtMYB92* and *AtMYB93*) in addition to other transcription factors. The existence of other such transcription factors may explain the unaffected levels of ferulate in the *myb53 myb92 myb93* triple knock-out lines.

4.2: RNA-seq analysis identifies potential suberin polyester synthase

One of the most relevant findings from the overlapping set of differentially expressed genes from the TKO and *AtMYB53*-OE experiments is that there are a multitude of new candidate genes that could be involved in suberin biosynthesis.

Three GDSL-motif esterases/acyltransferases/lipases, At1g53940 (*GLIP2*), At2g23540, and At3g50400, and an SGNH hydrolase-type esterase (At5g37690) were identified as being differentially expressed in both experiments. From previous research, these four genes are designated *AtGELP20*, *AtGELP51*, *AtGELP73*, and *AtGELP96*, respectively (Lai *et al.*, 2017). The GDSL-motif enzymes are described primarily as lipolytic enzymes, catalyzing the hydrolysis and/or synthesis of lipid-containing esters. In *Arabidopsis thaliana*, there are 105 documented proteins, organized into 12 subclades within 4 larger clades that have the conserved GDSL catalytic (Lai *et al.*, 2017). This potentially represents a wide diversity of chemical specificity within this family. The SGNH hydrolase-type esterases are members of the larger GDSL-motif esterase/acyltransferase/lipase enzyme family, since they contain four conserved SGNH-motifs in addition to the GDSL-containing catalytic site (Akoh *et al.*, 2004). The discovery of these enzyme classes within the RNA-seq dataset is particularly striking because it has been reported that at least one member of the GDSL-motif enzyme family,

CUTIN SYNTHASE 1 (*CUS1*), is involved with the polymerization of monomers to form polyesters in cutin development in the fruits of tomato (*Solanum lycopersicum*) (Bakan and Marion, 2017). The overall function and biosynthesis of cutin is analogous to that of suberin, making members of the GDSL-motif esterase/acyltransferase/lipase family attractive candidates as suberin synthase enzymes.

Three of the GDSL-motif esterases/acyltransferases/lipases represented above, *AtGELP51* (At2g23540), *AtGELP73* (At3g50400), *AtGELP96* (At5g37690) are all members of the same IIc subclade (Lai *et al.*, 2017). Interestingly, *AtGELP51* and *AtGELP73* are very closely related, appearing together within the phylogenetic tree. The closest other member, *AtGELP98* (At5g41890), does not appear to be differentially regulated in the triple knock-out mutant experiment nor in the *AtMYB53* over-expression experiment. From gene expression analyses across organs, *AtGELP51* and *AtGELP73* are both expressed in seedlings and roots (Lai *et al.*, 2017). *AtGELP51* is also expressed in siliques and seeds, which are examples of tissues where suberin is also deposited. A recent RNA-seq study using siliques of mixed developmental stages examining the role of *AtMYB107* in the production of seed coat suberin corroborates this finding (Gou *et al.*, 2017). Our transcriptomics dataset showed a significant 2.24-fold down-regulation of *AtGELP51* in siliques of the *myb107* knock-out line compared to wild-type, while *AtGELP73* did not have any significant differential expression. The other GDSL-motif enzyme identified here, *AtGELP96*, is not closely related to either *AtGELP51* or *AtGELP73* (Lai *et al.*, 2017). However, *AtGELP38* (At1g74460) is very closely related to *AtGELP96*, and appears in the *AtMYB53* over-expression dataset, reporting a 4.66 ± 1.11 fold change in the mutant line at 15 hours post-treatment compared to the DMSO control.

Both *AtGELP38* and *AtGELP96* are expressed in seedlings, roots, stems, siliques, and seeds, while *AtGELP38* is additionally expressed in rosette leaves (Lai *et al.*, 2017). In agreement with the RNA-seq data presented in this thesis, *AtGELP96*, but not *AtGELP38*, was found to be differentially expressed in the *myb107* knock-out line affecting seed coat suberin (3.14 to 3.19 fold change) (Gou *et al.*, 2017). Therefore, the GDSL-motif esterases/acyltransferases/lipases AtGELP51, AtGELP73, and AtGELP96 are all candidate enzymes worth investigating as putative suberin polyester synthases, while investigation into the closely related members *AtGELP38* and *AtGELP98* would be worthwhile to establish whether or not they are involved in suberin-related processes. In the future, T-DNA knock-out mutants for these genes could be screened by measuring the root suberin extracted from 2-week old roots as described here to determine if there is an abnormal suberin phenotype associated with these mutants.

From the set of GDSL-motif esterases/acyltransferases/lipases, *GLIP2* (*AtGELP20*; At1g53940) exhibited a unique gene expression profile. It was the only member from this enzyme family to be up-regulated in the triple knock-out analysis and down-regulated in the *AtMYB53* over-expression analysis. While the other candidates were all in the same phylogenic clade, *GLIP2* is by itself in clade IIIa (Lai *et al.*, 2017). In previous research, it was found that expression of *GLIP2* is induced by ethylene signaling, as well as by the salicylic acid and jasmonic acid signaling pathways, to suppress auxin responses and as a defense response to pathogenic bacteria (Lee *et al.*, 2009). Knock-out mutants of *GLIP2* are also cited to have an increase in the number of lateral roots formed. Since *AtMYB93* is a negative regulator of lateral root formation (Gibbs *et al*., 2014) and is inactive in the triple knock-out mutant line, then the induction

of *GLIP2* expression may be the plant attenuating the number of lateral roots formed in response. The production of so many extra lateral roots in this mutant line may be too metabolically expensive for the plant to maintain, while stressed by other internal factors, especially in a growing media that is already well saturated with water to meet the plant's physiological needs. In the case of *AtMYB53* over-expression resulting in a decrease in *GLIP2* expression, this could be the result of overlapping functions with *AtMYB53* and *AtMYB93*, since they are closely related, phylogenetically (Stracke *et al.*, 2001). Following this logic, it seems unlikely that GLIP2 could represent a suberin polyester synthase enzyme since its expression pattern does not coincide with the development of suberin-rich roots.

4.3: RNA-seq analysis identifies potential suberin biosynthetic enzymes

An HXXXD-type (BAHD) acyl-transferase (At1g78990) was up-regulated at 15 hours post-treatment in the *AtMYB53* OE line and down-regulated in the TKO mutant lines. This family of enzyme uses acyl-CoA substrates to produce secondary metabolic products with ester or amide linkages (for O-acyltransferases and N-acyltransferases, respectively) (Bontpart *et al*., 2015). The particular enzyme described here is a member of the sub-clade IIIb of BAHD acyl-transferases (Tuominen *et al.*, 2011). Unfortunately, there are no published data describing any role for this enzyme. While BAHD acyltransferases are known to catalyze the formation of a wide variety of secondary metabolites, such as terpenoids and alkaloids (Tuominen *et al.*, 2011), the association of this enzyme with suberin development hints at a possible role in synthesizing phenolic compounds for the suberin-associated lignin-like polymer. Since ASFT (Molina *et al*., 2009) and FACT (Kosma *et al*., 2012) are also BAHD acyltransferases themselves, this

new candidate gene may have a similar activity to contribute to the synthesis of suberin-associated waxes.

The identification of the cutin biosynthetic gene *LONG-CHAIN ACYL-CoA SYNTHETASE 2* (*LACS2*; At1g49430) as a potential transcriptional target of *AtMYB53, AtMYB92,* and *AtMYB93* also suggests that there is significant overlap between the suberin and cutin biosynthetic pathways. *LACS* enzymes have been well documented to catalyze the conversion of free fatty acids into fatty acyl-CoAs for further chemical conversions, eventually leading to the incorporation of acyl chains into cutin or its associated cuticular waxes (Schnurr *et al.*, 2004; Li-Beisson *et al.*, 2013). Preliminary results indicate an overlapping role for LACS2 in suberin and cutin biosynthesis, although such data has not yet been published (Isabel Molina, personal communication). The inclusion of *LACS2* in the list of differentially expressed genes in the knock-out and over-expression experiments performed here provides additional supporting evidence to indicate its involvement in both cutin and suberin biosynthesis.

4.4: Oxidative proteins may have a key role in suberin biosynthesis.

Two redox reaction proteins, NADPH oxidase (*RBOHB*; At1g09090) and a peroxidase protein (At4g11290), were a surprise pair of suberin-related candidate genes that were differentially expressed in our RNA-seq dataset. Lignin biosynthesis is heavily dependent on the production of reactive oxygen species (ROS) for lignification of monolignols to occur, and NADPH oxidases and peroxidases are known to work in tandem in this process (Lee *et al.*, 2013). It has been reported that a protein with similar activity, RBOHF, is one of the NADPH oxidases that is responsible for the production of ROS species in Casparian strip assembly (Lee *et al.*, 2013).

Both *RBOHB* and peroxidase (At1g09090) were up-regulated in the *myb53 myb92 myb93* knock-out analysis and down-regulated in the *AtMYB53*-OE line (*RBOHB* expression was repressed at 6H, while the peroxidase was repressed at 15H post-treatment). One possible explanation for this induction of oxidative genes in the knock-out mutant plants is to compensate and replace the missing aliphatic barriers in the root with more lignin or lignin-like compounds, such as suberin-associated polyphenolics. The reverse logic for the down-regulation of such genes in the over-expression line applies; the sheer abundance of aliphatic barriers accumulating in the root endodermis may not benefit from added polyphenolic compounds. The expression of these genes may suggest that the *AtMYB53*, *AtMYB92*, and *AtMYB93* proteins act as regulators for endodermal development, controlling the transition from Stage I to Stage II root differentiation, where the knock-out of MYB functions slows the transition between the two stages.

4.5: CASP-like proteins are novel key players identified by RNA-Seq analysis

The uncharacterized protein family members CASPARIAN STRIP MEMBRANE DOMAIN- (CASP) LIKE PROTEINS 1B2 and 1D2 (At4g20390 and At3g06390, respectively) form an interesting set of candidate genes for suberin deposition in roots. CASP proteins are described as protein scaffolding proteins, coordinating the positions of important catalytic machinery, including NADPH oxidases and peroxidases, in close proximity to each other to enable the polymerization of lignin (Tobimatsu and Schuetz, 2019). Both *CASP1B2* and *CASP1D2* were found to be down-regulated in the *myb53 myb92 myb93* triple knock-out line and up-regulated in the *MYB53*-OE line, indicating that they may serve an analogous role in suberin biosynthesis to arrange the important catalytic protein players at the cell wall for suberin polymerization. From the same

protein family, *CASP1D1* and *CASP1B1* were found to be up-regulated in response to *MdMYB93* over-expression in apple leaves (Legay *et al*., 2016), and *myb9* and *myb107* mutants were shown to exhibit down-regulation of *CASP1B2* in mature seeds (Lashbrooke *et al.*, 2016). Characterizing T-DNA knock-out lines for these genes may prove fruitful in understanding the process of cell wall-localized suberin polymerization.

4.6: *ABCG23* is a novel transporter associated with suberin biosynthesis

Previous studies have shown that an *abcg2 abcg6 abcg20* triple knock-out mutant has severely altered suberin lamellae structure and altered suberin composition (Yadav *et al.*, 2014). This mutant line exhibited a reduction of ferulate, 22:0 primary fatty alcohol, and 22:0 fatty acid. This is accompanied by an increase in dicarboxylic acids ranging from 18:0 to 22:0, indicating that the *ABCG2*, *ABCG6*, and *ABCG20* set is not an exhaustive list of transporters involved with suberin monomer export to the cell wall. The same study reports that two more transporters, ABCG1 and ABCG16, are required for correct pollen wall formation, but are not reported to have an effect on the assembly of suberin in the root endodermis. Up until now, the question remains whether there are additional transporters involved in endodermal root suberin formation.

While there are already three ABC-type transporters associated with suberin biosynthesis, the ABC-2 type transporter family protein *ABCG23* (At5g19410) is a new candidate transporter for suberin monomers across the plasma membrane. This transporter was found to be down-regulated in the *myb53 myb92 myb93* triple knock-out line, and up-regulated at both time-points in the *AtMYB53* over-expression line, giving a solid foundation for proposing the involvement of ABCG23 in suberin monomer transport across the plasma membrane. At this time, no other function has been proposed

for this transporter. Since it is a completely novel transporter, a quadruple *abcg2 abcg6 abcg20 abcg23* knock-out mutant could easily be developed from the existing triple knock-out mutant and tested using the suberin extraction assay to confirm its putative role in suberin transport. Assuming this mutant line would be viable, it would also be possible to attempt to observe the effect that this mutant has on the deposition of the patterned suberin lamellae on endodermal cells by transmission electron microscopy (TEM), and compare them to see if there are additional effects versus the triple mutant.

There are other ABC-type transporters that were identified in the RNA-seq dataset with roles related to suberin-associated transport. Firstly, the transporter *ABCG40* appears as an up-regulated responsive gene at 6 hour post-treatment in the *AtMYB53* over-expression analysis. This transporter is reported to be responsible for cell-to-cell transport of abscisic acid, along with *ABCG16* and others. *ABCG16*, which was previously reported to be involved in pollen wall formation (Yadav *et al.*, 2014), was also up-regulated in the over-expression line for *AtMYB53* at both 6 and 15 hours post-treatment. It remains unclear whether the functions of either of these transporters is related explicitly to suberin polymer assembly, as it is unknown if the specificity of the *ABCG* transporters could accept such diverse substrates (Borghi *et al.*, 2015). Lastly, *ABCG10* was also found to be up-regulated in the over-expression line at both 6 and 15 hours post-treatment. All that has been reported about this transporter is that it is down-regulated in a *myb11 myb12 myb111* triple knock-out mutant, where it's proposed to facilitate the transport of flavonols across the tonoplast membrane (Stracke *et al*., 2007), although this hypothesis was never tested.

4.7: *AtMYB53, AtMYB92,* and *AtMYB93* regulate other transcription factors

From these analyses, it would appear that MYB-type transcription factors may regulate the transcription of other MYB-type transcription factors. *MYB52* and *MYB54* (At1g17950 and At1g73410, respectively) may be two such examples of down-stream transcriptional targets of *AtMYB53, AtMYB92,* and *AtMYB93.* However, it is not clear from the RNA-seq analysis if the observed correlation is direct, since there could be other signaling events triggering the changes in *AtMYB52* and *AtMYB54* expression. Both genes were found to be down-regulated in the triple knock-out lines and up-regulated in the *AtMYB53* over-expression experiment at 15 hours post-treatment. It was previously reported that *MYB52* and *MYB54* are transcriptionally activated at least in part by *MYB46* and *MYB103*, which coincides with the up-regulation of genes related to secondary cell wall biogenesis, including the formation of cellulose, xylan, and lignin (Zhong *et al.*, 2012). It was reported that activation of *MYB46* was sufficient to induce gene expression of these secondary cell wall biosynthetic genes, but that *MYB52* and *MYB54*, among other transcription factors, may also participate in regulating the downstream genes. Clarifying the roles of *MYB52* and *MYB54* in the context of *AtMYB53*, *AtMYB92*, and *AtMYB93* regulation in root endodermal cells may shed light on how the transition between Casparian strip deposition (Stage I), to ectopic suberin deposition (Stage II), and to lignified or suberized tertiary cell wall development (Stage III), is coordinated by the plant (Doblas *et al*., 2017; Peterson and Enstone, 1996).

Another MYB-responsive gene identified was the Group II-c *WRKY56* (Eulgem *et al.*, 2000) transcription factor (At1g64000), which was down-regulated in the triple knock-out lines, but up-regulated in the *AtMYB53* over-expression experiment at 15 hours post-treatment. Exclusive to the plant kingdom, WRKY transcription factors form a 72

member family in *Arabidopsis thaliana* are involved, for example, in regulating hormonal responses to manage plant responses to abiotic and biotic stresses, and to coordinate plant development (Eulgem *et al.*, 2000). The WRKY transcription factors are so-named as they all contain a conserved WRKYGQK N-terminal amino acid motif and a C-terminal zinc finger-like motif that confers high affinity binding to a consensus DNA sequence. So far, no reports about the possible function of *WRKY56* has been reported.

4.8: The depletion of suberin in roots simulates a stress-related response

It is notable that the triple knock-out lines for *myb53 myb92 myb93* appear healthy under non-stress conditions despite a ~70% reduction in suberin. It would have been expected that the plants would be under great osmotic stress, which would manifest physically. While they appear to thrive in their carefully controlled growth chambers, the Gene Ontology studies presented in this work give some insight to the internal state of the plant.

Under the category of "Biological Processes" (Table S2), GO enrichment analysis of the triple knock-out line may reveal clues about the physiological state of the plant in the absence of *AtMYB53*, *AtMYB92*, and *AtMYB93.* When the list of down-regulated responsive genes for the triple-knock-out dataset was analyzed by GO Enrichment, the result was that the main gene ontology represented were suberin-related terms. Also, when the list of up-regulated responsive genes in the triple knock-out dataset, the result was that there were many GO terms related to stress-responsive genes. Therefore, a correlation was observed between enhanced stress responses with diminished suberin function. Specifically, the terms "Regulation of cell wall organization or biogenesis", "Phenylpropanoid metabolic process" [GO:0009698], "Suberin biosynthetic process"

[GO:0010345], and “Lipid metabolic process” [GO:0006629] are all obviously connected to suberin biosynthesis, and these are correlated with the gene ontologies “Response to stimulus” [GO:0050896], “Response to stress” [GO:0006950], “Response to wounding” [GO:0009611], and “Response to ethylene” [GO:0009723]. One interpretation of these results is that despite the triple knock-out mutant showing no physical symptoms of stress, there are stress-responsive pathways that are activated in this mutant line even under adequate growing conditions.

Among the many genes represented in this grouping of stress-responsive genes, there are a few genes with interesting properties within the context of plant survival. One of these genes, *AtPRx72* (At5g66390), encodes a peroxidase that has been previously shown to participate in lignification and secondary cell wall thickening (Herrero *et al.,* 2013). A number of genes related to cold and freezing tolerance, such as *ERF105* were also represented, however *ERF105* expression is shown to increase the cold-tolerance of plants (Bolt *et al.*, 2017). It could be speculated that there is a fine-tuning of expression of cold-responsive genes to help plants cope in an osmotically challenging environment. There is documentation that there may exist cross-talk between the cold-response and drought-response pathways. In one such case, *DREB1A*/*CBF3* is reported to affect downstream gene expression in response to both drought and freezing stress (Sakuma *et al.*, 2006). While *DREB1A*/*CBF3* was not identified in the dataset, *DREB1B*/*CBF1* was found to be up-regulated in the triple knock-out line (At4g25490), which has previously been linked to improve drought tolerance and carbohydrate metabolism when ectopically expressed in the Chinese plant *Salvia miltiorrhiza* (Wei *et al.*, 2016). Another ethylene-related gene, *ERF11*, is known to be a negative regulator of ethylene synthesis. The

aforementioned *GLIP2* and *RBOHB* genes are also annotated to be related to stress responses. Therefore, the lack of *AtMYB53, AtMYB92,* and *AtMYB93* gene activities results in the induction of stress-response genes; despite the lack of visual cues about the health of the triple knock-out mutants with 70% reduction in suberin, such that the plants are actively experiencing stress despite growing in a well-maintained and well-supplemented environment.

4.9: The profiles of expressed genes as a function of time in response to MYB53 induction is dynamic

One of the main goals of this research was to determine the possible mechanism by which the activation of the suberin biosynthetic genes take place in response to overexpression of *AtMYB53*. The Enrichment Analysis of the steroid-inducible *AtMYB53* over-expression line shows that the profiles of genes that are transcriptionally regulated are distinct between the initial activation of the gene and the long-term gene expression changes that it induces. While this analysis only provides a limited snapshot of the events that unfold post-expression of *AtMYB53*, the differences observed are informative enough to mention here.

Firstly, in terms of Cellular Component gene ontologies (Table S5), there is a consistent down-regulation of gene products that are annotated to the periphery of the cell (plasma membrane and cell wall). At 6 and 15 hours post-treatment together, and at 15 hours post-treatment, the gene ontology is enriched in down-regulated genes annotated to the cell wall. Meanwhile, at 6 hours only, the ‘cell wall’ GO term was absent from the enrichment analysis, and instead replaced by the GO term related to the ‘plasma membrane’. It is unexpected to observe such down-regulation of genes in the *AtMYB53*

over-expression line, but this widespread occurrence may represent shifting of general cell wall resources from other metabolic pathways into suberin biosynthesis. Alternatively, this boosted cell activity at the periphery of the triple *myb53 myb92 myb93* knock-out lines may represent the plant compensating for the lack of a cell wall-associated barrier, while the active suppression of activity in the same sub-cellular location in the *AtMYB53*-OE line indicates that the cell is attempting to attenuate the production of excessive suberin in a relatively low osmotic stress environment.

Next, the Molecular Function gene ontology (Table S4) shows that the overall cell activities at 6 hours post-treatment for the *AtMYB53* over-expression line is much different from the longer-term effects at 15 hours. In the short-term, there is a striking down-regulation in general catalytic activity, while in the long term, a number of up-regulated lipid-related functions are represented in the Enrichment Analysis. Among these, there are terms related to fatty acid elongase, lipase, and acyltransferase activities. Furthermore, at 15 hours post-treatment, catalytic activity appears as both enriched in the analysis as a grouping of down-regulated and up-regulated genes, which emphasizes the massive re-direction of resources by the cell, coordinated by *AtMYB53*, towards suberin biosynthesis and other related functions. Within the dataset of shared responsive genes between the 6 and 15 hour treatments, the Enrichment Analysis reveals that the transcriptional activation of hydrolase enzymes may be central to the role that *AtMYB53* plays in suberin biosynthesis, and therefore some of its members may be good candidates as direct transcriptional targets by *AtMYB53.*

Finally, the Biological Process gene ontology (Table S2) shows that the massive accumulation of suberin in the *MYB53*-OE line results in a multitude of metabolic

pathways that are altered. The evidence presented in this analysis shows that the increased amount suberin- and lipid-related genes expressed late after *MYB53* induction was coupled to a significant decrease in small molecule synthesis and plant tissue development and differentiation, highlighting the high energetic cost of suberin production. The decrease in metabolic activity may also be a consequence of the reduced need to boost root development through primary metabolism while water and solute transport may be more efficient in highly suberized root tissues.

This work highlights that there are more factors involved from a physiological point of view relative to suberin production. These analyses suggest that imbalanced accumulation of suberin simulates a stress response in the plant. Therefore, future experimentation should focus on characterizing these mutant lines using physiological methods, such as measuring the transpiration rates, measuring the root length and density of lateral roots, and a battery of stress exposure experiments (drought tolerance, salt tolerance, heavy metal tolerance) to evaluate the tolerance of these suberin mutant lines compared to wild-type plants.

4:10: Domains required for functional activity of MYB proteins

The transcriptomics (RNA-seq) experiments identified genes regulated by *AtMYB53*, *AtMYB92*, and *AtMYB93*. However, it is currently unknown which genes are direct targets of these transcription factors. A future ChIP experiment would help identify which genes are the direct gene targets of these transcription factors. Here, I examined five types of truncations of these MYBs to identify the minimal protein sequence required for activity and that could be tagged for future ChIP experiments.

Originally, while only the “Series 1” and “Series 2” truncations (Figure 24) were considered for initial tests, it was apparent that the “Series 1” truncations for each transcription factor are not ideal since they do not appear to have all the regions necessary for full activity. Since the truncations contained the DNA-binding domain formed by the R2/R3 domains are present, the remaining part of the proteins must contain domains that are critical for transcription factor activity. In contrast, the “Series 2” truncations are only 30%-45% less active relative to their full-length counterparts. These larger proteins include conserved protein motifs near their C-termini, but they do not have the third conserved SUB – type motif found across several suberin gene-inducing MYB transcription factors. The “Series 3” and “Series 4” constructs were developed to further examining the role of the SUB III motif in protein function.

Of the five types of truncations, the “Series 3” truncated constructs are the most attractive candidates to be tested in the future. This construct boasted high activity when tested *in vivo* in the *N. benthamiana* transient over-expression assay when the epitope tags were absent. Surprisingly, the longest AtMYB92 and AtMYB93 truncations available, “Series 4”, generated much more suberin than even the full-length proteins, just by truncating the last 6 amino acids. With high activity, epitope-tagged versions of these constructs may be more active for a ChIP experiment, although this high activity may result in non-specific activity. The truncation that most closely imitates full-length (wild-type) protein should be used to minimize data artifacts. The closest to wild-type activity were the “Series 3” set of truncations, which overall matched the activities of the full-length proteins. The “Series 3” truncation constructs have the last half of the SUB III domain missing, yet there is no obvious difference between the level of activity between

these constructs and the full-length proteins, at least in the transient overexpression experiment using *N. benthamiana* leaves. There may not be a major structural or functional purpose to the SUB III conserved domain, although the protein-protein interactions at the C-terminus are yet to be characterized. Moving forward, further validation involving the construction of an HA epitope-tagged version of the "Series 3" truncation will be required, along with Western Blot analysis of the chosen epitope-tagged constructs before a ChIP-Seq experiment is carried out.

Between the first and second iteration of the transient over-expression experiments, the behaviour of the "Series 2" truncations altered slightly. At first, the truncated proteins exhibited moderate suppression of transcription factor activity, based on the chemical data, compared to their full-length counterparts, by anywhere between 30 – 40 %. However, in the second experiment, the "Series 2" truncations were found to be significantly more transcriptionally active than the full-length protein in the case of *AtMYB92*, while the truncated version of *AtMYB93* was just as active as its native form. Only the "Series 2" truncation of *AtMYB53* was found to replicate its previous result. The ability of these variations of MYB transcription factors to be active may rely heavily on the internal state of the plants themselves at the time of infiltration and sampling.

CHAPTER 5 – CONCLUSION

Suberin is a glycerol-based heteropolymer that is deposited in the inner face of the plant cell wall to serve as a physical barrier to unregulated gas, water, and solute movement (Franke and Schreiber, 2007; Schreiber, 2010). Here, the R2/R3-type MYB transcription factors AtMYB53, AtMYB92, and AtMYB93 that are known to be involved in suberin biosynthesis were analyzed using transcriptomics, with the objective of determining the identity of unknown suberin polyester synthases, biosynthetic enzymes, transporters, and co-regulators.

To meet these objectives, a combination of previously characterized *myb53 myb92 myb93* triple knock-out and *AtMYB53* over-expression lines in *Arabidopsis thaliana* were analyzed. The GDSL esterases / acyltransferases / lipases *AtGELP51*, *AtGELP73*, and *AtGELP96* were revealed as potential suberin synthases, mirroring advances in the cutin literature (Bakan and Marion, 2017). The HXXXD-type (BAHD) acyl-transferase At1g78990 was one of several catalytic acyltransferase enzymes identified; while their function is unknown, other enzymes with similar activity, such as FACT and ASFT (Molina *et al*., 2009; Kosma *et al*., 2012), are known to participate in the formation of suberin-associated waxes. The CASP-like proteins *CASP1B2* and *CASP1D2* proteins may serve an analogous role in suberin formation as the CASP scaffolding proteins in Casparian strip assembly (Tobimatsu and Schuetz, 2019). *ABCG23* is added to the list of potential suberin monomer membrane transporters. Finally, the identification of transcription factor *WRKY56* expands our current scope of regulatory players in suberin biosynthesis.

To determine the direct gene targets of each MYB-type transcription factor, work is underway to develop epitope-tagged versions of these proteins that could be used in a ChIP-seq experiment. Here, it is shown that the inclusion of the HA tag at the C terminus of each protein does not significantly affect their activities *in vivo*. Current results suggest that cleaving the protein in the middle of the SUB – III domain (corresponding to the "Series 3" set of truncations) was the most suitable truncation for this purpose, since the un-altered full-length predicted proteins cannot be stably expressed with the tag. Further validation is required to test the expression of such fusion proteins using Western blot analysis.

These initial steps taken to understand the mechanism behind the regulation of suberin deposition by these MYB-type transcription factors are important. Plants must tolerate and adapt to drought conditions, salinity, heavy metal contamination, and pathogenic interactions (Franke and Schreiber, 2007). Knowledge of suberin regulation may be employed when developing stress resistant crops in the future.

REFERENCES

Alves, M. S., S. P. Dadalto, A. B. Gonçalves, G. B. de Souza, V. A. Barros, and L. G. Fietto. 2014. Transcription factor functional protein-protein interactions in plant defense responses. *Proteomes*. **2**: 85 – 106.

Akoh, C. C., G. C. Lee, Y. C. Liaw, T. H. Huang, and J. F. Shaw. 2004. GDSL family of serine esterases/lipases. *Progress in Lipid Research*. **43**: 534 – 552.

Ambawat, S., P. Sharma, N. R. Yadav, and R. C. Yadav. 2013. MYB transcription factor genes as regulators for plant responses: an overview. *Physiology and Molecular Biology of Plants*. **19**: 307 – 321.

Bakan, B., and D. Marion. 2017. Assembly of the cutin polyester: from cells to extracellular cell walls. *Plants*. **6**: 57.

Barberon, M., and N. Geldner. 2014. Radial transport of nutrients: the plant root as a polarized epithelium. *Plant Physiology*. **166**: 528 – 537.

Barberon M., J.E. Vermeer, D. De Bellis, P. Wang, S. Naseer, T.G Andersen, B.M. Humbel, C. Nawrath, J. Takano, D.E. Salt, and N. Geldner. 2016. Adaptation of root function by nutrient-induced plasticity of endodermal differentiation. *Cell*. **164**: 447 – 459.

Barberon, M. 2017. The endodermis as a checkpoint for nutrients. *New Phytologist*. **213**: 1604 – 1610.

Baxter I, P.S. Hosmani, A. Rus, B. Lahner, J.O Borevitz, B. Muthukamar, M.V. Mickelbart, L. Schreiber, R. Franke, and D.E. Salt. 2009. Root suberin forms an extracellular barrier that affects water relations and mineral nutrition in *Arabidopsis thaliana*. *PLoS Genetics*. **5**: e1000492.

Beisson F, Y. Li-Beisson, G. Bonaventure, M. Pollard, and J.B. Ohlrogge. 2007. The acyltransferase *GPAT5* is required for the synthesis of suberin in the seed coat and root of *Arabidopsis*. *Plant Cell*. **19**: 351 – 368.

Bernards, M.A. 2002. Demystifying suberin. *Canadian Journal of Botany*. **80**: 227-240.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*. **30**: 2114 – 2120.

Bolt, S., E. Zuther, S. Zintl, D. K. Hincha, and T. Schmulling. 2017. ERF105 is a transcription factor gene of *Arabidopsis thaliana* required for freezing tolerance and cold acclimation. *Plant, Cell and Environment*. **40**: 108 - 120.

Bontpart, T., V. Cheynier, A. Ageorges, and N. Terrier. 2015. *BAHD* or *SCPL* acyltransferase? What a dilemma for acylation in the world of plant phenolic compounds. *New Phytologist*. **208**: 695 – 707.

Borghi, L., J. Kang, D. Ko, Y. Lee, and E. Martinoia. 2015. The role of *ABCG*-type ABC transporters in phytohormone transport. *Biochemical Society Transactions*. **43**: 924 – 930.

Bray, N. L., H. Pimentel, P. Melsted, and L. Pachter. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*. **34**: 525 – 527.

Coego, A., E. Brizuela, P. Castillejo, S. Ruíz, C.Koncz, J.C. Pozo, M. Piñeiro, J.A. Jarillo, J. Paz-Ares, J. León. 2014. The TRANSPLANTA collection of *Arabidopsis* lines: a resource for functional analysis of transcription factors based on their conditional over-expression. *The Plant Journal*. **77**: 944 – 953

Compagnon V., P. Diehl, I. Benveniste, D. Meyer, H. Schaller, L. Schreiber, R. Franke, and F. Pinot. 2009. *CYP86B1* is required for very long chain ω-hydroxy acid and α,ω-dicarboxylic acid synthesis in root and seed suberin polyester. *Plant Physiology.***150**: 1831 – 1843.

D'Auria, J. 2006. Acyltransferases in plants: A good time to be *BAHD*. *Current Opinion in Plant Biology*. **9**: 331 – 340.

Delude C., L. Fouillen, P. Bhar, M.J. Cardinal, S. Pascal, P. Santos, D.K. Kosma, J. Joubès, O. Rowland, and F. Domergue. 2016. Primary fatty alcohols are major components of suberized root tissues of *Arabidopsis* in the form of alkyl hydroxycinnamates. *Plant Physiology*. **171**: 1934 – 1950.

Doblas, V.G., N. Geldner, and M. Barberon. 2017. The endodermis, a tightly controlled barrier for nutrients. *Current Opinion in Plant Biology*. **39**: 136 – 143.

Domergue F, S.J. Vishwanath, J. Joube`s, J. Ono, J. Lee, M. Bourdon, R. Alhattab, C. Lowe, S. Pascal, R. Lessire, O. Rowland. 2010. Three *Arabidopsis* fatty acyl-CoA reductases, *FAR1*, *FAR4*, and *FAR5*, generate primary fatty alcohols associated with suberin deposition. *Plant Physiology*. **153**: 1539 – 1554.

Dubos C., R. Stracke, E. Grotewold, B. Weisshaar, C. Martin, and L. Lepiniec. 2010. *MYB* transcription factors in *Arabidopsis thaliana*. *Trends in Plant Science*. **15**: 573 – 581.

Eulgem, T., P. J. Rushton, S. Robatzek, and I. E. Somssich. 2000. The *WRKY* superfamily of plant transcription factors. *Trends in Plant Science*. **5**: 199 – 206.

Franke, R., I. Briesen, T. Wojciechowski, A. Faust, A. Yephremov, C. Nawrath, and L. Schreiber. 2005. Apoplastic polyesters in *Arabidopsis* surface tissues - A typical suberin and a particular cutin. *Phytochemistry*. **66**: 2643 – 2658.

Franke, R., and L. Schreiber. 2007. Suberin — a biopolyester forming apoplastic plant interfaces. *Current Opinion in Plant Biology*. **10**: 252 – 259.

Franke, R., R. Höfer, I. Brieson, M. Emsermann, N. Efremova, A., Yephremov, and L. Schreiber. 2009. The DAISY gene from *Arabidopsis* encodes a fatty acid elongase condensing enzyme involved in the biosynthesis of aliphatic suberin in roots and the chalaza-micropyle region of seeds. *The Plant Journal*. **57**: 80 – 95.

Garcia, O., P. Bouige, C. Forestier, and E. Dassa. 2004. Inventory and comparative analysis of rice and Arabidopsis ATP-binding cassette (ABC) systems. *Journal of Molecular Biology*. **343**: 249 – 265.

Geldner, N. 2013. The endodermis. *Annual Review of Plant Biology*. **64**: 531 – 558.

Gibbs D.J., U. Vo , S.A. Harding, J. Fannon, L.A. Moody, E. Yamada, K. Swarup, C. Nibau, G.W. Bassel, A. Choudhary, J. Lavenus, S.J. Bradshaw, D.J. Stekel, M.J. Bennet, and J.C. Coates. 2014. *AtMYB93* is a novel negative regulator of lateral root development in *Arabidopsis*. *New Phytologist*. **203**: 1194 – 1207.

Gou, M., G. Hou, H. Yang, X. Zhang, Y. Cai, G. Kai, and C.J. Liu. 2017. The *MYB107* transcription factor positively regulates suberin biosynthesis. *Plant Physiology* **173**: 1045 – 1058.

Graça, J. 2015. Suberin: the biopolyester at the frontier of plants. *Frontiers in chemistry*. **3**: 62

Herrero, J., F. Fernández-Pérez, T. Yebra, E. Novo-Uzal, F. Pomar, A. Pedreño, J. Cuello, A. Guéra, A. Esteban-Carrasco, and J. M. Zapata. 2013. Bioinformatic and functional characterization of the basic peroxidase 72 from *Arabidopsis thaliana* involved in lignin biosynthesis. *Planta*. **237**: 1599 – 1612.

Höfer R, I. Briesen, M. Beck, F. Pinot, L. Schreiber, and R. Franke. 2008. The *Arabidopsis* cytochrome P450 *CYP86A1* encodes a fatty acid omega-hydroxylase involved in suberin monomer biosynthesis. *Journal of Experimental B*otany. **59**: 2347 – 2360.

Hu, Hefeng. 2018. The role of transcription factor MYB53 from *Arabidopsis thaliana* in the regulated production of suberin. M. Sc. Thesis. Carleton University. 1 – 120.

Joubès, J., S. Raffaele, B. Bourdenx, C. Garcia, J. Laroche-Traineau, P. Moreau, F. Domergue, and R. Lessire. 2008. The VLCFA elongase gene family in *Arabidopsis thaliana*: phylogenetic analysis, 3D modelling and expression profiling. *Plant Molecular Biology*. **67**: 547 – 566.

Karimi, M., D. Inzé, and A. Depicker. 2002. GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science*. **7**: 193 – 195.

Kim, J., J. H. Jung, S. B. Lee, Y. S. Go, H. J. Kim, R. Cahoon, J. E. Markham, E. B. Cahoon, M. C. Suh. 2013. *Arabidopsis* 3-ketoacyl-coenzyme A synthase 9 is involved in the synthesis of tetracosanoic acids as precursors of cuticular waxes, suberins, sphingolipids, and phospholipids. *Plant Physiology*. **162**: 567 – 580.

Kosma, D. K., I. Molina, J.B. Ohlrogge, and M. Pollard. 2012. Identification of an *Arabidopsis* fatty alcohol:caffeoyl-Coenzyme A acyltransferase required for the synthesis of alkyl hydroxycinnamates in root waxes. *Plant Physiology.* **160**: 237 – 248.

Kosma, D. K., J. Murmu, F.M. Razeq, P. Santos, R. Bourgault, I. Molina, and O. Rowland. 2014. *AtMYB41* activates ectopic suberin synthesis and assembly in multiple plant species and cell types. *The Plant Journal*. **80**: 216 – 229.

Kurdyukov, S., A. Faust, C. Nawrath, S. Bar, D. Voisin, N. Efremova, R. Franke, L. Schreiber, H. Saedler, J.P. Metraux, and A. Yephremov. 2006. The epidermis-specific extracellular *BODYGUARD* controls cuticle development and morphogenesis in *Arabidopsis*. *Plant Cell*. **18**: 321 – 339.

Lai, C. P., L. M. Huang, L. F. O. Chen, M. T. Chan, and J. F. Shaw. 2017. Genome-wide analysis of *GDSL*-type esterases/lipases in Arabidopsis. *Plant Molecular Biology* **95**: 181 – 197.

Lashbrooke, J., H.Cohen, D.L. Samocha, O. Tzfadia, I. Panizel, V. Zeisler, H. Massalha, A. Stern, L. Trainotti, L. Schreiber, F. Costa, and A. Aharoni. 2016. *MYB107* and *MYB9* homologs regulate suberin deposition in angiosperms. *Plant Cell*. **28**: 2097 – 2116.

Lee S. B., S.J. Jung, Y.S. Go, H.U. Kim, J.K. Kim, H.J. Cho, O.K. Park, and M.C. Suh. 2009. Two *Arabidopsis* 3-ketoacyl CoA synthase genes, *KCS20* and *KCS2/DAISY*, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. *The Plant Journal*. **60**: 462 – 475.

Lee, D.S., B.K. Kim, S.J. Kwon, H. C. Jin, and O.K. Park. 2009. Arabidopsis GDSL lipase 2 plays a role in pathogen defense via negative regulation of auxin signaling. *Biochemical and Biophysical Research Communications*. **379**: 1038 – 1042.

Lee, Y., M.C. Rubio, J. Alassimone, and N. Geldner. 2013. A mechanism for localized lignin deposition in the endodermis. *Cell*. **153**: 402 – 412.

Legay, S., G. Guerriero1, C. André, C. Guignard, E. Cocco, S. Charton, M. Boutry, and O. Rowland, and J.F. Hausman. 2016. *MdMyb93* is a regulator of suberin deposition in russeted apple fruit skins. *New Phytologist*. **212**: 977 – 991.

Li-Beisson, Y., B. Shorrosh, F. Beisson, M. X. Andersson, V. Arondel, P. D. Bates, S. Baud, D. Bird, A. Debono, T. P. Durrett, R. B. Franke, I. A. Graham, K. Katayama, A. A Kelly, T. Larson, J. E. Markham, M. Miquel, I. Molina, I. Nishida, O. Rowland, L. Samuels, K. M. Schmid, H. Wada, R. Welti, C. Xu, R. Zallot, and J. Ohlrogge. 2013. Acyl-lipid metabolism. *The Arabidopsis Book*. **11**: e0161.

Millar, A. A., and L. Kunst. 1997. Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. *The Plant Journal.* **12**: 121 – 131.

Molina, I., Y. Li-Beisson, F. Beisson, J.B. Ohlrogge, and M. Pollard. 2009. Identification of an Arabidopsis feruloyl-coenzyme A transferase required for suberin synthesis. *Plant Physiology*. **151**: 1317 – 1328.

McFarlane H.E., Y. Watanabe, W. Yang, Y. Huang, J. Ohlrogge, and A.L. Samuels. 2014. Golgi- and trans-Golgi network-mediated vesicle trafficking is required for wax secretion from epidermal cells. *Plant Physiology*. **164**: 1250 – 1260.

Naseer, S., Y. Lee, C. Lapierre, R. Franke, C. Nawrath, and N. Geldner. 2012. Casparian strip diffusion barrier in *Arabidopsis* is made of a lignin polymer without suberin. *Proceedings of the National Academy of Sciences USA*. **109**: 10101 – 10106.

Nawrath, C. 2002. The biopolymers cutin and suberin. *The Arabidopsis Book*. **1**: e0021.

Nawrath, C., L. Schreiber, R.B. Franke, N. Geldner, J.J. Reina-Pinto, and L. Kunst. 2013. Apoplastic diffusion barriers in *Arabidopsis*. *The Arabidopsis Book* **11**: e0167.

Paul, S., K. Gable, F. Beaudoin, E. Cahoon, J. Jaworski, J.A. Napier, and T.M. Dunn. 2006. Members of the *Arabidopsis FAE1*-like 3-ketoacyl-CoA synthase gene family substitute for the Elop proteins of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*. **281**: 9018 – 9029.

Peterson, S.A., and D.E. Enstone. 1996. Functions of passage cells in the endodermis and exodermis of roots. *Physiologia Plantarum*. **97**: 592 – 598.

Ranathunge K., L. Schreiber, and R. Franke. 2011. Suberin research in the genomics era—New interest for an old polymer. *Plant Science*. **180**: 399 – 413.

Sakuma, Y., K. Maruyama, Y. Osakabe, F. Qin, M. Seki, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2006. Functional analysis of an *Arabidopsis* transcription factor, *DREB2A*, involved in drought-responsive gene expression. *The Plant Cell*. **18**: 1292–1309.

Salminen, T. A., K. Blomqvist, and J. Edqvist. 2016. Lipid transfer proteins: classification, nomenclature, structure, and function. *Planta*, **244**: 971 – 997.

Samuels, L., L. Kunst, and R. Jetter. 2008. Sealing plant surfaces: cuticular wax formation by epidermal cells. *Annual Reviews in Plant Biology*. **59**: 683 – 707

Schreiber, L. 2010. Transport barriers made of cutin, suberin and associated waxes. *Trends in Plant Science*. **15**: 546 – 553.

Schnurr, J., J. Shockey, and J. Browse. 2004. The acyl-CoA synthetase encoded by LACS2 is essential for normal cuticle development in *Arabidopsis*. *Plant Cell*. **16**: 629 – 642.

Stracke, R., M. Werber, and B. Weisshaar. 2001 The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*. **4**: 447 – 456.

Stracke, R., H. Ishihara, G. Huep, A. Barsch, F. Mehrtens, K. Niehaus, and B. Weisshaar. 2007. Differential regulation of closely related R2R3-*MYB* transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *The Plant Journal* **50**: 660 – 677.

Tobimatsu, Y., and M. Schuetz. 2019. Lignin polymerization: how do plants manage the chemistry so well? *Current Opinion in Biotechnology*. **56**: 75 – 81.

Tuominen, L. K., V. E. Johnson, and C. J. Tsai. 2011. Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among Populus paralogues. *BMC Genomics*. **12**: 236.

Verrier, P. J., D. Bird, B. Burla, E. Dassa, C. Forestier, M. Geisler, M. Klein, Ü. Kolukisaoglu, Y. Lee, E. Martinoia, A. Murphy, P. A. Rea, L. Samuels, B. Schulz, E. P. Spalding, K. Yazaki, and F. L. Theodoulou. 2008. Plant ABC proteins – a unified nomenclature and updated inventory. *Trends in Plant Science*. **13**: 151 – 159.

Vishwanath, S.J., D.K. Kosma, I.P. Pulsifer, S. Scandola, S. Pascal, J. Joubes, F. Dittrich-Domergue, R. Lessire, O. Rowland, and F. Domergue. 2013. Suberin-associated fatty alcohols in *Arabidopsis thaliana*: distributions in roots and contributions to seed coat barrier properties. *Plant Physiology.* **163**: 1118 – 1132.

Vishwanath, S. J., C. Delude, F. Domergue, and O. Rowland. 2015. Suberin: biosynthesis, regulation, and polymer assembly of a protective extracellular barrier. *Plant Cell Reports*. **34**: 573 – 586.

Wang, P., Z. Wang, Y. Dou, X. Zhang, M. Wang, X. Tian. 2013. Genome-wide identification and analysis of membrane-bound O-acyltransferase (*MBOAT*) gene family in plants. *Planta*. **5**: 907 – 922.

Wei, Tao, K. Deng, Y. Gao, Y. Liu, M. Yang, L. Zhang, X. Zheng, C. Wang, W. Song, C. Chen, and Y. Zhang. Arabidopsis *DREB1B* in transgenic *Salvia miltiorrhiza* increased tolerance to drought stress without stunting growth. *Plant Physiology*. **104**: 17 – 28.

Yadav V., I. Molina, K. Ranathunge, I.Q. Castillo, S.J. Rothstein, and J.W. Reed. 2014. *ABCG* transporters are required for suberin and pollen wall extracellular barriers in *Arabidopsis*. *The Plant Cell*. **26**: 3569 – 3588.

Yang W, J.P. Simpson, Y. Li-Beisson, F. Beisson, M. Pollard, J.B. Ohlrogge. 2012. A land-plant-specific glycerol-3-phosphate acyltransferase family in *Arabidopsis*: substrate specificity, sn-2 preference, and evolution. *Plant Physiology.***160**: 638 – 652.

Yeats, T.H., B.B. Laetitia, H.M.F. Viart, T. Isaacson, Y. He, L. Zhao, A.J. Matas, G.J. Buda, S.D. Domozych, M.H. Clausen, and J.K.C. Rose. 2012. The identification of cutin synthase: formation of the plant polyester cutin. *Nature Chemical Biology.* **8**: 609 – 611.

Zhong, R., C. Lee, J. Zhou, R. L. McCarthy, and Z. H. Ye. 2008. A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *The Plant Cell*. **20**: 2763 – 2782.

Zhong, R., and Z. H. Ye. 2012. *MYB46* and *MYB83* bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes. *Plant and Cell Physiology*. **53**: 368 – 380.

Zuo, J., Q.W. Niu, and N.H. Chua. 2000. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *The Plant Journal*. **24**: 265 – 273.

APPENDIX A – PRIMERS USED IN THIS STUDY

Table 6: Summary of qPCR primers used for RNA extract validation

Primer Name	Primer Sequence (5' to 3')
FAR1 Forward	ACAGCTCATTCGGGAGACAC
FAR1 Reverse	GAGCCGTGAAATCGTGAAGT
FAR4 Forward	AGTCCTTGATCTTATACCTGTGG
FAR4 Reverse	GCTTCCCTGCGTGTATTGC
FAR5 Forward	AACTCATGGGGTCAAGTCGC
FAR5 Reverse	CTTCTTAAGCACGTGTGTGACG
GPAT5 Forward	TGAGGGAACCACTTGTCGTG
GPAT5 Reverse	ATCGCAACCGGAACAATCCT
ASFT Forward	AACTCATGGGGTCAAGTCGC
ASFT Reverse	CTTTGGAGGGTTTCGAGCATTGAG
CYP86B1 Forward	ATCCAGGATGTCTCGGTCCA
CYP86B1 Reverse	TGACGAATCTCACAACCGCA
CYP86A1 Forward	CGCTGCGTTTATACCCTTCTGTGC
CYP86A2 Forward	CTTGGCACGAAAGTCCCGTC
GAPDH Forward	TTGGTGACAACAGGTCAAGCA
GAPDH Reverse	AAACTTGTCGCTCAATGCAATC

Table 7: Summary of PCR primers for generating MYB protein truncations

Primer Name	Primer Sequence (5' to 3')
MYB53 Forward	CACCATGGGAAGATCTCCTAGCTCA
MYB53 (1-301) Rev	TTAGGAATCATCAAATAGATGATGATCAGG
MYB53 (1-289) Rev	TTAAGAGTCTGTGACTGCTACTACTTCTTGATC
MYB53 (1-276) Rev	TTAGGCTTGGTTGCTGCCGAAA
MYB53 (1-183) Rev	TTAAGAAGGTTGAAGAAGGTATTGG
MYB92 Forward	CACCATGGGAAGATCTCCTATCTCT
MYB92 (1-328) Rev	CTAAGAATCATCCAATAGATGATCAGGCC
MYB92 (1-317) Rev	CTATGAAGCTGAAGCTCCTGATTCTTG
MYB92 (1-304) Rev	CTAGATATTGCTATTGACATCTTCAA
MYB92 (1-184) Rev	CTAAGACGGTTGAAGAAGGTATTG
MYB93 Forward	CACCATGGGGAGGTCGCCTTGTT
MYB93 (1-359) Rev	CTAGCTTTCGTCGAAGCAAATGTCAGGC
MYB93 (1-348) Rev	CTACGACGCAGCTTCACAACCTCCATAG
MYB93 (1-335) Rev	CTATGCATCCGCCGGATTGTTGT
MYB93 (1-176) Rev	CTATTGGAGATATTGTAGATTGGCTAGT

Table 8: Summary of PCR primers for generating HA epitope-tagged MYB protein truncations

Primer Name	Primer Sequence (5' to 3')
HA Reverse	TCAGCCGGCAGCGTAATCTGGAAC
MYB53 FL HA For	CCATGTTTCCAGACATTTCTTATCAATCTATT CTTTACCCATACGATGTTCCTG
MYB53 FL HA Rev	CAGGAACATCGTATGGGTAAAAGATAGATTG ATAAGAAATGTCTGGAAACATGG
MYB53 L HA For	GATGGTTTCGGCAGCAACCAAGCCATCTTTTA CCCATACGATGTTCCTG
MYB53 L HA Rev	CAGGAACATCGTATGGGTAAAAGATGGCTTGG TTGCCGAAACCATC
MYB92 FL HA For	GATGATTCTATATTTTCCGACATTCCTGGGATG ATCTTTTACCCATACGATG
MYB92 FL HA Rev	CATCGTATGGGTAAAAGATCATCCCAGGAATG TCGGAAAATATAGAATCATC
MYB92 L HA For	TACGGTATTGAAGATGTCAATAGCAATATCATC TTTCCATACGATGTTCCTG
MYB92 L HA Rev	CAGGAACATCGTATGGGTAAAAGATGATATTG CTATTGACATCTTCAATACCGTA
MYB93 FL HA For	GACGAAAGCCTCATGAACGTTATATCTGGGAT GATCTTTTACCCATACGATG
MYB93 FL HA Rev	CATCGTATGGGTAAAAGATCATCCCAGATATA ACGTTCATGAGGCTTTCGTC
MYB93 L HA For	CACACAACAATCCGGCGGATGCAATCTTTTAC CCATACGATGTTCCTG
MYB93 L HA Rev	CAGGAACATCGTATGGGTAAAAGATTGCATCC GCCGGATTGTTGTGTG

APPENDIX B – TRIPLE MYB KNOCK-OUT SUBERIN ANALYSIS

Table 9 (page 138): The amount suberin broken-down into components per gram of dry weight for both alleles (TKO-1 and TKO-2) of the Arabidopsis *myb53 myb92 myb93* triple knockout lines compared to wild-type (WT) as detected by GC-FID (±SE). The sample size was n = 4 for the knock-out mutant lines, and n = 3 for the WT line.

Table 10 (page 139): Statistical analysis of the amount of total aliphatic suberin and suberin component break-down per gram of dry weight for triple knockout lines compared to wild-type (WT). The sample size was n = 4 (df=6) for mutant lines, and n = 3 (df=5) for the WT line. Multiple t-testing correction were applied with the Holm-Sidak method. Red entries signify non-statistical significance between the mutant line compared to the WT at the $\alpha = 0.05$ level (student's t test; GraphPad Prism 8.1.0).

Table 9:

ALIPHATIC SUBERIN MONOMER		AVERAGED WEIGHT OF ALIPHATIC SUBERIN COMPONENT (µg / g DRY WEIGHT)		
		WT (n=3)	*TKO-1 (n=4)*	*TKO-2 (n=4)*
HCAs	***trans*- ferulate**	93 ± 22	105 ± 28	77 ± 29
FAMEs	**16:0**	124 ± 17	44 ± 3	43 ± 4
	18:2	70 ± 4	26 ± 2	36 ± 4
	18:1	23 ± 6	12 ± 1	7 ± 1
	18:0	22 ± 3	6 ± 1	8 ± 3
	20:0	146 ± 16	44 ± 3	38 ± 2
	22:0	551 ± 66	108 ± 6	92 ± 8
	24:0	131 ± 15	26 ± 1	21 ± 2
	26:0	28 ± 9	10 ± 1	19 ± 7
DCA DMEs	**16:0**	319 ± 37	81 ± 4	69 ± 5
	18:2	0 ± 0	0 ± 0	0 ± 0
	18:1	762 ± 88	144 ± 7	121 ± 9
	18:0	149 ± 14	39 ± 2	33 ± 3
	20:0	45 ± 6	15 ± 0	13 ± 2
	22:0	65 ± 7	20 ± 1	17 ± 2
	24:0	146 ± 22	34 ± 4	30 ± 4
Hydroxy FAMEs	**16-OH 16:0**	142 ± 15	46 ± 2	40 ± 3
	18-OH 18:1	737 ± 72	146 ± 8	128 ± 9
	18-OH 18:0	67 ± 3	31 ± 2	39 ± 5
	20-OH 20:0	137 ± 9	51 ± 2	58 ± 7
	22-OH 22:0	323 ± 36	93 ± 6	81 ± 8
	24-OH 24:0	65 ± 7	18 ± 1	17 ± 2
Fatty Alcohols	**1-OH 18:0**	135 ± 12	35 ± 2	30 ± 2
	1-OH 20:0	210 ± 13	49 ± 3	40 ± 4
	1-OH 22:0	72 ± 6	12 ± 1	9 ± 1
	1-OH 24:0	4 ± 0	1 ± 0	3 ± 1
TOTAL		4567 ± 446	1196 ± 59	1068 ± 70

Table 10:

ALIPHATIC SUBERIN MONOMER		Adjusted p-Values	
		TKO-1 (n=4)	*TKO-2 (n=4)*
HCAs	***trans*-ferulate**	0.764	0.821
FAMEs	**16:0**	0.013	0.020
	18:2	0.003	0.018
	18:1	0.182	0.112
	18:0	0.013	0.112
	20:0	0.007	0.009
	22:0	0.007	0.008
	24:0	0.006	0.007
	26:0	0.182	0.821
DCA DMEs	**16:0**	0.007	0.009
	18:2	NA	NA
	18:1	0.006	0.007
	18:0	0.005	0.005
	20:0	0.013	0.018
	22:0	0.007	0.009
	24:0	0.013	0.017
Hydroxy FAMEs	**16-OH 16:0**	0.007	0.009
	18-OH 18:1	0.004	0.004
	18-OH 18:0	0.003	0.043
	20-OH 20:0	0.003	0.010
	22-OH 22:0	0.007	0.009
	24-OH 24:0	0.007	0.009
Fatty Alcohols	**1-OH 18:0**	0.004	0.004
	1-OH 20:0	0.001	0.001
	1-OH 22:0	0.002	0.002
	1-OH 24:0	NA	0.821
TOTAL		0.005	0.005

APPENDIX C – "Series 1" and "Series 2" MYB TRUNCATION SUBERIN ANALYSIS

Table 11: The amount suberin broken-down into components per gram of dry weight for MYB truncation lines compared to pBIN19-p19 negative control as detected by GC-FID (±SE). The sample size was n = 4 for all lines, except for *AtMYB92* (1-304), *AtMYB92* (1-184), *AtMYB93* (1-365), and *AtMYB93* (1-335) (n = 3).

ALIPHATIC SUBERIN MONOMER		AVERAGED WEIGHT OF ALIPHATIC SUBERIN COMPONENT (µg / g DRY WEIGHT)																							
		pBIN19 - p19 (n=4)			*pBIN19 - Empty (n=4)*			*AtMYB53 (1-310) (n=4)*			*AtMYB53 (1-276) (n=4)*			*AtMYB53 (1-183) (n=4)*			*AtMYB92 (1-334) (n=4)*			*AtMYB92 (1-304) (n=3)*			*AtMYB92 (1-184) (n=3)*			*AtMYB93 (1-365) (n=3)*			*AtMYB93 (1-335) (n=3)*			*AtMYB93 (1-172) (n=4)*		
HCAs	*trans*- ferulate	0	±	0	0	±	0	4864	±	382	2651	±	260	0	±	0	2363	±	833	1676	±	126	0	±	0	2351	±	57	1423	±	40	58	±	25
FAMEs	16:0	229	±	78	223	±	18	175	±	20	237	±	19	219	±	30	217	±	30	152	±	5	396	±	55	103	±	2	112	±	3	136	±	9
	18:2	164	±	59	149	±	15	132	±	23	156	±	12	184	±	29	120	±	15	106	±	3	145	±	25	60	±	7	72	±	5	115	±	7
	18:1	196	±	70	200	±	15	239	±	33	380	±	30	211	±	34	273	±	65	273	±	41	223	±	41	139	±	13	151	±	5	146	±	11
	18:0	58	±	20	57	±	9	83	±	9	102	±	8	50	±	8	114	±	21	91	±	12	72	±	7	44	±	3	56	±	2	33	±	3
	20:0	0	±	0	6	±	6	757	±	56	731	±	82	14	±	5	611	±	105	468	±	56	38	±	6	390	±	53	404	±	23	18	±	7
	22:0	6	±	4	37	±	37	9439	±	709	5586	±	495	1164	±	1103	4809	±	812	2392	±	1182	64	±	17	5689	±	144	2796	±	246	60	±	45
	24:0	21	±	11	42	±	6	2351	±	183	1001	±	86	27	±	11	1170	±	201	910	±	144	41	±	5	1569	±	47	739	±	43	40	±	11
	26:0	71	±	22	83	±	12	409	±	35	185	±	17	61	±	14	266	±	51	219	±	21	90	±	8	207	±	13	169	±	8	60	±	4
DCA DMEs	16:0	8	±	7	30	±	30	6575	±	478	3605	±	266	50	±	11	3622	±	781	2228	±	225	72	±	13	5668	±	321	2321	±	174	74	±	37
	18:2	0	±	0	9	±	9	1527	±	151	1202	±	87	14	±	8	946	±	185	650	±	73	0	±	0	749	±	16	524	±	18	12	±	12
	18:1	82	±	24	156	±	75	25984	±	2271	14318	±	1233	159	±	50	12808	±	2554	10197	±	1174	164	±	41	14670	±	441	9733	±	795	294	±	189
	18:0	12	±	12	17	±	17	4493	±	285	2453	±	252	20	±	12	2908	±	558	1864	±	147	28	±	16	3287	±	358	2106	±	133	64	±	29
	20:0	166	±	57	201	±	26	2434	±	155	1376	±	160	232	±	63	1751	±	387	1031	±	81	210	±	25	2122	±	231	1066	±	52	157	±	10
	22:0	47	±	17	12	±	12	984	±	104	295	±	41	31	±	11	617	±	68	410	±	47	43	±	10	1151	±	45	276	±	19	18	±	3
	24:0	0	±	0	4	±	4	414	±	40	263	±	71	16	±	9	305	±	37	205	±	28	0	±	0	292	±	103	111	±	6	6	±	6
Hydroxy FAMEs	16-OH 16:0	30	±	16	43	±	6	2360	±	148	1032	±	90	34	±	13	1426	±	309	805	±	62	49	±	10	1599	±	149	794	±	42	44	±	13
	18-OH 18:1	113	±	51	153	±	17	6615	±	531	3299	±	293	898	±	719	3964	±	815	3050	±	358	156	±	29	3259	±	62	2653	±	186	148	±	55
	18-OH 18:0	82	±	28	30	±	30	132	±	19	107	±	10	60	±	10	118	±	31	69	±	11	67	±	4	113	±	9	45	±	6	62	±	8
	20-OH 20:0	29	±	27	49	±	20	661	±	34	426	±	46	37	±	19	606	±	114	374	±	29	14	±	14	554	±	66	354	±	14	26	±	6
	22-OH 22:0	47	±	15	65	±	16	4803	±	345	1934	±	157	52	±	14	3624	±	731	2159	±	249	81	±	18	4704	±	187	1501	±	99	73	±	27
	24-OH 24:0	0	±	0	0	±	0	1030	±	82	315	±	23	0	±	0	700	±	125	477	±	59	0	±	0	1136	±	71	329	±	13	10	±	6
Fatty Alcohols	1-OH 18:0	7	±	6	21	±	21	147	±	18	84	±	8	11	±	7	148	±	26	77	±	10	264	±	81	216	±	3	59	±	3	24	±	10
	1-OH 20:0	0	±	0	0	±	0	219	±	24	168	±	14	6	±	6	163	±	42	117	±	6	37	±	9	110	±	13	110	±	9	19	±	5
	1-OH 22:0	0	±	0	0	±	0	1124	±	90	745	±	48	17	±	17	865	±	163	562	±	89	7	±	7	855	±	14	422	±	18	12	±	8
	1-OH 24:0	97	±	47	88	±	36	166	±	29	173	±	62	110	±	40	148	±	37	85	±	9	82	±	31	140	±	19	79	±	4	22	±	7
TOTAL		1465	±	491	1674	±	187	78118	±	5949	42822	±	3455	3677	±	1630	44664	±	8310	30649	±	3679	2341	±	305	51176	±	1582	28404	±	1811	1728	±	500

Table 12: Statistical analysis of the amount of total aliphatic suberin and suberin component break-down per gram of dry weight for MYB truncation lines compared to pBIN19-p19 negative control as detected by GC-FID. The sample size was n = 4 (df=6) for all lines, except for *AtMYB92* (1-304), *AtMYB92* (1-184), *AtMYB93* (1-365), and *AtMYB93* (1-335) (n = 3; df=5). Multiple t-testing correction were applied with the Holm-Sidak method. Red entries signify non-statistical significance between the mutant line compared to the WT at the α = 0.05 level (student's t test; df=5 GraphPad Prism 8.1.0).

ALIPHATIC SUBERIN MONOMER		T-TEST: TWO-SAMPLE ASSUMING UNEQUAL VARIANCES - TWO-TAILED PROBABILITY								
		AtMYB53 (1-310) (n=4)	*AtMYB53 (1-276) (n=4)*	*AtMYB53 (1-183) (n=4)*	*AtMYB92 (1-334) (n=4)*	*AtMYB92 (1-304) (n=3)*	*AtMYB92 (1-184) (n=3)*	*AtMYB93 (1-365) (n=3)*	*AtMYB93 (1-335) (n=3)*	*AtMYB93 (1-172) (n=4)*
HCAs	*trans*- ferulate	2.60E-04	7.77E-04	NA	1.90E-01	4.66E-04	NA	2.00E-06	4.00E-06	4.00E-06
FAMEs	16:0	8.94E-01	9.90E-01	1.00E+00	9.35E-01	9.40E-01	9.55E-01	7.33E-01	8.15E-01	8.15E-01
	18:2	8.94E-01	9.90E-01	1.00E+00	9.35E-01	9.40E-01	9.99E-01	7.33E-01	8.15E-01	8.15E-01
	18:1	8.94E-01	2.75E-01	1.00E+00	9.35E-01	9.40E-01	9.99E-01	8.73E-01	9.41E-01	9.41E-01
	18:0	7.75E-01	3.66E-01	1.00E+00	4.75E-01	8.30E-01	9.99E-01	8.73E-01	9.42E-01	9.42E-01
	20:0	2.19E-04	1.44E-03	5.32E-01	2.68E-02	3.26E-03	1.46E-02	3.15E-03	1.00E-04	1.00E-04
	22:0	2.23E-04	5.45E-04	9.99E-01	2.57E-02	3.55E-01	2.11E-01	2.00E-06	4.70E-04	4.70E-04
	24:0	2.60E-04	5.45E-04	1.00E+00	2.84E-02	7.36E-03	9.74E-01	5.00E-06	1.58E-04	1.58E-04
	26:0	1.44E-03	4.97E-02	1.00E+00	9.69E-02	4.14E-02	9.99E-01	3.77E-02	9.92E-02	9.92E-02
DCA DMEs	16:0	2.19E-04	2.44E-04	3.67E-01	4.22E-02	1.78E-03	1.08E-01	7.10E-05	2.90E-04	2.90E-04
	18:2	5.98E-04	2.33E-04	9.55E-01	3.80E-02	2.78E-03	NA	<0.000001	1.00E-05	1.00E-05
	18:1	3.54E-04	5.08E-04	9.94E-01	3.92E-02	3.12E-03	9.09E-01	4.00E-06	3.80E-04	3.80E-04
	18:0	1.14E-04	9.78E-04	1.00E+00	3.80E-02	6.04E-04	9.99E-01	1.45E-03	1.58E-04	1.58E-04
	20:0	2.19E-04	3.84E-03	1.00E+00	6.51E-02	4.13E-03	9.99E-01	2.34E-03	1.08E-03	1.08E-03
	22:0	1.01E-03	1.25E-02	1.00E+00	4.82E-03	5.20E-03	9.99E-01	2.70E-05	2.63E-03	2.63E-03
	24:0	5.71E-04	6.82E-02	9.55E-01	4.64E-03	4.51E-03	NA	1.29E-01	8.10E-05	8.10E-05
Hydroxy FAMEs	16-OH 16:0	1.14E-04	5.82E-04	1.00E+00	4.37E-02	7.92E-04	9.98E-01	8.32E-04	1.51E-04	1.51E-04
	18-OH 18:1	2.60E-04	6.25E-04	9.99E-01	4.17E-02	3.62E-03	9.99E-01	4.00E-06	3.33E-04	3.33E-04
	18-OH 18:0	7.18E-01	8.39E-01	1.00E+00	9.35E-01	9.40E-01	9.99E-01	8.73E-01	8.15E-01	8.15E-01
	20-OH 20:0	1.65E-04	3.33E-03	1.00E+00	3.92E-02	4.55E-03	9.99E-01	3.88E-03	2.13E-03	2.13E-03
	22-OH 22:0	2.19E-04	4.67E-04	1.00E+00	3.92E-02	3.25E-03	9.74E-01	1.60E-05	2.21E-04	2.21E-04
	24-OH 24:0	2.60E-04	2.35E-04	NA	2.99E-02	3.62E-03	NA	1.08E-04	1.90E-05	1.90E-05
Fatty Alcohols	1-OH 18:0	2.22E-03	3.01E-03	1.00E+00	3.75E-02	1.25E-02	2.20E-01	2.00E-05	7.78E-03	7.78E-03
	1-OH 20:0	9.73E-04	4.67E-04	9.99E-01	7.11E-02	7.30E-05	9.29E-02	1.94E-03	3.80E-04	3.80E-04
	1-OH 22:0	2.60E-04	1.22E-04	9.99E-01	3.75E-02	7.09E-03	9.87E-01	<0.000001	2.60E-05	2.60E-05
	1-OH 24:0	7.75E-01	8.39E-01	1.00E+00	9.35E-01	9.40E-01	9.99E-01	8.73E-01	9.42E-01	9.42E-01
TOTAL		2.60E-04	4.67E-04	9.96E-01	3.80E-02	3.83E-03	9.74E-01	8.00E-06	2.46E-04	2.46E-04

APPENDIX D – "Series 2" – HA, "Series 3", and "Series 4" MYB TRUNCATION SUBERIN ANALYSIS

Table 13 (next page): The amount suberin broken-down into components per gram of dry weight for MYB truncation lines tagged with HA compared to pBIN19-p19 negative control as detected by GC-FID. The sample size was n = 4 for all lines, except for *AtMYB92* (1-328), *AtMYB92* (1-304), *AtMYB93* (1-348), and *AtMYB93* (1-335) + HA (n = 3).

Table 13:

ALIPHATIC SUBERIN MONOMER		AVERAGED WEIGHT OF ALIPHATIC SUBERIN COMPONENT (µg / g DRY WEIGHT)									
		pBIN19 - p19 (n=4)	*pBIN19 - Empty (n=4)*	*AtMYB53 (1-310) (n=4)*	*AtMYB53 (1-310) + HA (n=4)*	*AtMYB53 (1-301) (n=4)*	*AtMYB53 (1-289) (n=4)*	*AtMYB53 (1-276) (n=4)*	*AtMYB53 (1-276) + HA (n=4)*	*AtMYB92 (1-334) (n=4)*	*AtMYB92 (1-334) + HA (n=4)*
HCAs	***trans*- ferulate**	12 ± 7	29 ± 7	2965 ± 610	1904 ± 338	2605 ± 516	1732 ± 133	2672 ± 598	2471 ± 791	2639 ± 486	3155 ± 1045
FAMEs	**16:0**	327 ± 27	335 ± 72	717 ± 98	725 ± 36	659 ± 28	875 ± 86	723 ± 24	626 ± 142	509 ± 6	753 ± 102
	18:2	141 ± 10	177 ± 47	173 ± 24	187 ± 20	180 ± 7	198 ± 6	193 ± 12	117 ± 41	137 ± 20	136 ± 11
	18:1	513 ± 74	697 ± 167	685 ± 60	722 ± 82	817 ± 64	917 ± 52	789 ± 17	571 ± 54	514 ± 75	677 ± 99
	18:0	117 ± 12	100 ± 26	283 ± 32	246 ± 17	340 ± 39	442 ± 64	331 ± 29	158 ± 54	296 ± 49	272 ± 71
	20:0	34 ± 19	51 ± 11	3264 ± 296	2426 ± 213	3287 ± 603	2953 ± 543	2693 ± 324	2255 ± 435	2510 ± 447	2488 ± 511
	22:0	321 ± 254	53 ± 15	33267 ± 2494	23174 ± 4323	28875 ± 6534	14905 ± 2657	19860 ± 1876	25523 ± 7305	20960 ± 3070	29015 ± 4783
	24:0	239 ± 35	202 ± 48	8890 ± 728	6207 ± 1612	6882 ± 1533	2854 ± 608	5093 ± 446	7694 ± 2179	5360 ± 807	8637 ± 1323
	26:0	1000 ± 175	729 ± 158	1666 ± 149	1079 ± 316	1307 ± 210	760 ± 174	1291 ± 100	1281 ± 433	1119 ± 224	1524 ± 256
DCA DMEs	**16:0**	160 ± 24	101 ± 22	20144 ± 873	19225 ± 1027	17434 ± 3896	10161 ± 2013	9914 ± 1289	15965 ± 3710	14678 ± 2892	23078 ± 3048
	18:2	128 ± 26	83 ± 20	3406 ± 239	2411 ± 334	3667 ± 808	2909 ± 517	3199 ± 336	2897 ± 719	2652 ± 521	2444 ± 363
	18:1	420 ± 138	442 ± 107	64035 ± 4645	55696 ± 5445	52761 ± 12007	30770 ± 6124	37901 ± 3551	52452 ± 13651	34991 ± 7624	50456 ± 17297
	18:0	93 ± 17	73 ± 22	14540 ± 889	13001 ± 860	11556 ± 2368	7582 ± 1679	8231 ± 910	10969 ± 2887	11304 ± 2571	14995 ± 2612
	20:0	4328 ± 630	3256 ± 712	9703 ± 655	8187 ± 816	7773 ± 1293	5570 ± 938	6627 ± 498	7123 ± 1948	9148 ± 978	10662 ± 1670
	22:0	117 ± 9	51 ± 19	2952 ± 248	3027 ± 846	2226 ± 402	1295 ± 256	1779 ± 265	2038 ± 684	2543 ± 394	4478 ± 716
	24:0	229 ± 57	107 ± 19	1805 ± 283	1387 ± 547	1300 ± 165	815 ± 190	712 ± 141	953 ± 322	653 ± 121	2316 ± 391
Hydroxy FAMEs	**16-OH 16:0**	186 ± 34	154 ± 35	5788 ± 211	5363 ± 255	4566 ± 948	2744 ± 483	2732 ± 324	4662 ± 876	4948 ± 895	5980 ± 515
	18-OH 18:1	497 ± 79	399 ± 86	12581 ± 1142	10040 ± 1350	10034 ± 2160	6498 ± 1034	7405 ± 676	9917 ± 2268	8055 ± 1460	10510 ± 1297
	18-OH 18:0	443 ± 158	300 ± 120	179 ± 13	139 ± 23	135 ± 26	92 ± 14	128 ± 26	120 ± 41	184 ± 18	229 ± 21
	20-OH 20:0	68 ± 11	43 ± 11	1764 ± 167	1384 ± 246	1381 ± 209	1099 ± 200	1290 ± 139	1293 ± 287	1772 ± 312	1982 ± 419
	22-OH 22:0	232 ± 36	193 ± 43	10127 ± 907	9034 ± 2806	7454 ± 1207	4270 ± 704	6289 ± 491	7597 ± 1852	9333 ± 1096	14127 ± 2280
	24-OH 24:0	121 ± 18	44 ± 18	1815 ± 152	1940 ± 824	1193 ± 202	502 ± 84	976 ± 61	1664 ± 234	1481 ± 170	2728 ± 384
Fatty Alcohols	**1-OH 18:0**	247 ± 29	206 ± 51	585 ± 49	524 ± 119	446 ± 43	502 ± 27	545 ± 51	632 ± 43	882 ± 166	866 ± 207
	1-OH 20:0	85 ± 26	52 ± 19	330 ± 63	317 ± 94	355 ± 69	318 ± 59	332 ± 16	262 ± 94	410 ± 55	486 ± 77
	1-OH 22:0	53 ± 26	42 ± 11	3239 ± 167	2257 ± 600	2870 ± 620	2414 ± 295	2474 ± 330	2663 ± 606	2906 ± 339	3525 ± 410
	1-OH 24:0	688 ± 142	325 ± 89	731 ± 116	492 ± 84	579 ± 69	467 ± 97	615 ± 45	477 ± 164	1016 ± 193	1192 ± 172
TOTAL		10798 ± 1495	8245 ± 1763	205632 ± 13758	171093 ± 20879	170681 ± 35592	103643 ± 18520	124793 ± 11958	162379 ± 41083	141000 ± 22186	196711 ± 33185

Table 13 (cont'd):

ALIPHATIC SUBERIN MONOMER		AVERAGED WEIGHT OF ALIPHATIC SUBERIN COMPONENT (µg / g DRY WEIGHT)									
		AtMYB92 (1-328) (n=3)	*AtMYB92 (1-317) (n=4)*	*AtMYB92 (1-304) (n=3)*	*AtMYB92 (1-304) + HA (n=4)*	*AtMYB93 (1-365) (n=4)*	*AtMYB93 (1-365) + HA (n=4)*	*AtMYB93 (1-359) (n=4)*	*AtMYB93 (1-348) (n=3)*	*AtMYB93 (1-335) (n=4)*	*AtMYB93 (1-335) + HA (n=3)*
HCAs	*trans*- ferulate	4594 ± 388	2822 ± 521	3116 ± 429	2820 ± 206	1880 ± 584	3972 ± 1043	3807 ± 1133	3233 ± 558	2666 ± 321	3946 ± 1039
FAMEs	16:0	811 ± 88	1231 ± 421	799 ± 91	794 ± 208	555 ± 116	577 ± 108	459 ± 157	689 ± 58	557 ± 108	936 ± 243
	18:2	148 ± 4	274 ± 88	209 ± 35	215 ± 38	101 ± 20	127 ± 10	369 ± 149	126 ± 16	139 ± 33	315 ± 126
	18:1	1238 ± 225	1499 ± 322	1251 ± 49	1059 ± 221	471 ± 95	534 ± 26	286 ± 167	565 ± 86	598 ± 101	649 ± 326
	18:0	775 ± 167	818 ± 218	632 ± 45	433 ± 112	199 ± 41	185 ± 44	100 ± 22	243 ± 77	188 ± 96	158 ± 51
	20:0	6324 ± 1315	4482 ± 1050	5447 ± 706	3116 ± 718	1853 ± 333	1879 ± 423	7817 ± 2306	2778 ± 702	2512 ± 599	3970 ± 1462
	22:0	38698 ± 10956	19416 ± 3068	39702 ± 8657	15359 ± 4462	11553 ± 6200	24876 ± 3669	20347 ± 11494	20067 ± 2607	15546 ± 5215	11924 ± 10611
	24:0	11488 ± 3473	4677 ± 740	10666 ± 2539	6523 ± 1109	4085 ± 569	5242 ± 731	5682 ± 1286	4832 ± 542	4865 ± 1077	4799 ± 1927
	26:0	2440 ± 764	1284 ± 425	2042 ± 393	1299 ± 129	516 ± 119	1125 ± 183	11365 ± 5961	1195 ± 78	1190 ± 327	5322 ± 4096
DCA DMEs	16:0	20441 ± 2263	11886 ± 2218	19292 ± 1812	16654 ± 4035	15292 ± 3604	18312 ± 3332	24892 ± 2623	11485 ± 1890	10908 ± 1432	13401 ± 9630
	18:2	4754 ± 863	4236 ± 1011	4887 ± 728	2699 ± 486	1484 ± 288	2210 ± 519	2918 ± 308	2078 ± 327	2047 ± 208	2093 ± 556
	18:1	90911 ± 24073	46088 ± 7823	80409 ± 12409	59796 ± 13542	34993 ± 7197	34441 ± 12394	57700 ± 9785	30958 ± 6730	29685 ± 4846	47791 ± 16570
	18:0	20806 ± 4438	10683 ± 2269	16540 ± 2885	12642 ± 2613	9191 ± 1630	11584 ± 2428	17491 ± 2291	9281 ± 2068	7726 ± 1529	11259 ± 3275
	20:0	13346 ± 2980	8548 ± 1984	11082 ± 2280	8282 ± 1390	5709 ± 903	9465 ± 1430	9758 ± 3309	6673 ± 930	5844 ± 584	5652 ± 3047
	22:0	4023 ± 985	2062 ± 520	3597 ± 551	2474 ± 270	2079 ± 256	3589 ± 715	3908 ± 85	1713 ± 312	1409 ± 104	3129 ± 1055
	24:0	3670 ± 1519	2257 ± 1026	2209 ± 731	1292 ± 110	765 ± 125	1494 ± 339	1385 ± 148	498 ± 76	613 ± 123	1574 ± 755
Hydroxy FAMEs	16-OH 16:0	6015 ± 880	3527 ± 712	5600 ± 815	4171 ± 691	4140 ± 1236	4965 ± 611	3566 ± 2112	3461 ± 574	3403 ± 333	4064 ± 2727
	18-OH 18:1	19453 ± 5274	11628 ± 2295	18443 ± 3222	10607 ± 1579	5954 ± 1602	6545 ± 1161	10355 ± 1964	6701 ± 1602	6875 ± 1043	7026 ± 3988
	18-OH 18:0	255 ± 77	185 ± 75	344 ± 86	163 ± 17	108 ± 31	295 ± 58	177 ± 21	212 ± 67	168 ± 50	117 ± 47
	20-OH 20:0	3096 ± 694	1944 ± 507	2847 ± 631	1524 ± 152	996 ± 135	1627 ± 336	1477 ± 715	1444 ± 380	1215 ± 195	1231 ± 531
	22-OH 22:0	13770 ± 3467	8121 ± 1815	14295 ± 3678	8352 ± 963	5900 ± 823	10793 ± 1699	7261 ± 4424	6228 ± 1360	6402 ± 731	4952 ± 4810
	24-OH 24:0	2265 ± 695	1108 ± 248	2347 ± 726	1568 ± 206	1179 ± 222	2068 ± 315	2504 ± 342	1140 ± 206	1292 ± 144	2066 ± 612
Fatty Alcohols	1-OH 18:0	736 ± 98	949 ± 318	885 ± 170	657 ± 132	598 ± 107	1455 ± 101	1115 ± 120	971 ± 243	563 ± 154	721 ± 222
	1-OH 20:0	796 ± 395	688 ± 310	637 ± 261	263 ± 59	185 ± 30	290 ± 55	471 ± 38	273 ± 43	286 ± 27	413 ± 82
	1-OH 22:0	5381 ± 1260	3778 ± 665	6032 ± 1410	2692 ± 361	1850 ± 453	2938 ± 362	3867 ± 291	2891 ± 397	3087 ± 603	2661 ± 982
	1-OH 24:0	1641 ± 679	1455 ± 797	1332 ± 294	624 ± 44	441 ± 56	909 ± 127	922 ± 102	650 ± 105	543 ± 136	903 ± 316
TOTAL		277875 ± 66589	155645 ± 31061	254643 ± 44369	166077 ± 23961	112077 ± 23078	151496 ± 30302	199999 ± 33847	120384 ± 19559	110328 ± 15513	141072 ± 59007

Table 14: Statistical analysis of the amount of total aliphatic suberin and suberin component break-down per gram of dry weight for MYB truncation lines tagged with HA compared to pBIN19-p19 negative control as detected by GC-FID. The sample size was n = 4 (df=6) for all lines, except for *AtMYB92* (1-328), *AtMYB92* (1-304), *AtMYB93* (1-348), and *AtMYB93* (1-335) + HA (n = 3, df=5). Multiple t-testing correction were applied with the Holm-Sidak method. Red entries signify non-statistical significance between the mutant line compared to the WT at the α = 0.05 level (student's t test; df=5 GraphPad Prism 8.1.0).

ALIPHATIC SUBERIN MONOMERS		Adjusted p-Values									
		AtMYB53 (1-310) (n=4)	*AtMYB53 (1-310) + HA (n=4)*	*AtMYB53 (1-301) (n=4)*	*AtMYB53 (1-289) (n=4)*	*AtMYB53 (1-276) (n=4)*	*AtMYB53 (1-276) + HA (n=4)*	*AtMYB92 (1-334) (n=4)*	*AtMYB92 (1-334) + HA (n=4)*	*AtMYB92 (1-328) (n=3)*	*AtMYB92 (1-317) (n=4)*
HCAs	*trans*- ferulate	6.15E-04	5.16E-04	1.01E-03	6.00E-06	8.99E-04	9.55E-03	6.31E-04	8.53E-03	5.60E-05	8.26E-04
FAMEs	16:0	1.79E-03	4.40E-05	6.70E-05	4.33E-04	1.40E-05	5.36E-02	2.61E-04	2.12E-03	2.76E-03	4.06E-02
	18:2	9.56E-02	3.32E-02	4.14E-03	8.77E-04	3.33E-03	5.82E-01	9.29E-01	5.26E-01	3.12E-01	9.22E-02
	18:1	4.41E-02	3.51E-02	4.14E-03	1.10E-03	2.41E-03	5.82E-01	9.85E-01	1.05E-01	1.28E-02	9.87E-03
	18:0	6.15E-04	3.02E-04	7.65E-04	8.77E-04	1.06E-04	5.67E-01	2.75E-03	2.51E-02	6.97E-03	7.38E-03
	20:0	1.00E-05	1.20E-05	7.89E-04	7.29E-04	5.30E-05	1.23E-03	6.15E-04	9.51E-04	3.76E-03	2.36E-03
	22:0	4.00E-06	6.38E-04	1.60E-03	6.97E-04	1.80E-05	6.41E-03	2.46E-04	3.53E-04	8.55E-03	4.35E-04
	24:0	7.00E-06	4.10E-03	1.60E-03	1.10E-03	1.40E-05	6.41E-03	3.10E-04	3.03E-04	1.11E-02	5.12E-04
	26:0	5.77E-03	6.77E-01	1.27E-01	1.36E-01	5.48E-02	5.82E-01	8.19E-01	5.82E-02	6.43E-02	2.63E-01
DCA DMEs	16:0	<0.000001	<0.000001	1.60E-03	8.77E-04	6.80E-05	2.98E-03	7.36E-04	1.25E-04	2.09E-04	8.84E-04
	18:2	3.00E-06	1.84E-04	1.60E-03	7.29E-04	3.30E-05	4.27E-03	8.38E-04	3.03E-04	2.39E-03	2.80E-03
	18:1	3.00E-06	2.00E-05	1.60E-03	8.77E-04	1.60E-05	4.27E-03	1.11E-03	8.53E-03	7.27E-03	5.72E-04
	18:0	1.00E-06	2.00E-06	1.18E-03	1.10E-03	3.50E-05	4.27E-03	1.26E-03	4.35E-04	3.93E-03	1.59E-03
	20:0	2.59E-04	4.10E-03	1.21E-02	1.36E-01	4.92E-03	1.88E-01	1.51E-03	3.92E-03	1.28E-02	4.59E-02
	22:0	8.00E-06	5.11E-03	8.43E-04	1.02E-03	1.58E-04	1.49E-02	3.50E-04	3.53E-04	6.97E-03	3.83E-03
	24:0	3.50E-04	3.32E-02	4.28E-04	5.22E-03	3.56E-03	4.35E-02	4.32E-03	6.35E-04	3.20E-02	4.59E-02
Hydroxy FAMEs	16-OH 16:0	<0.000001	<0.000001	1.46E-03	7.29E-04	6.10E-05	1.23E-03	6.34E-04	1.30E-05	8.93E-04	1.59E-03
	18-OH 18:1	1.00E-05	1.58E-04	1.60E-03	5.51E-04	1.90E-05	3.14E-03	6.70E-04	1.13E-04	8.27E-03	1.38E-03
	18-OH 18:0	4.66E-02	3.51E-02	2.53E-02	1.77E-02	2.28E-02	5.82E-02	8.36E-02	1.05E-01	2.27E-01	9.22E-02
	20-OH 20:0	1.20E-05	6.33E-04	3.92E-04	7.85E-04	3.80E-05	2.98E-03	6.31E-04	1.16E-03	4.99E-03	3.83E-03
	22-OH 22:0	1.00E-05	7.60E-03	4.76E-04	5.57E-04	7.00E-06	3.78E-03	7.60E-05	3.53E-04	6.97E-03	2.18E-03
	24-OH 24:0	9.00E-06	3.16E-02	8.43E-04	1.10E-03	5.00E-06	2.98E-04	9.40E-05	2.19E-04	1.25E-02	2.97E-03
Fatty Alcohols	1-OH 18:0	2.59E-04	3.16E-02	2.05E-03	3.24E-04	4.76E-04	1.53E-04	2.26E-03	8.53E-03	3.93E-03	4.05E-02
	1-OH 20:0	2.20E-03	2.79E-02	2.32E-03	2.13E-03	5.30E-05	7.43E-02	6.34E-04	8.74E-04	6.43E-02	4.59E-02
	1-OH 22:0	<0.000001	4.10E-03	1.50E-03	9.70E-05	7.70E-05	2.98E-03	7.40E-05	6.90E-05	5.48E-03	6.97E-04
	1-OH 24:0	6.56E-01	1.07E-01	2.17E-01	1.22E-01	3.65E-01	4.09E-01	1.29E-01	2.39E-02	1.08E-01	2.02E-01
TOTAL		2.00E-06	7.80E-05	7.68E-04	5.81E-04	2.40E-05	9.56E-04	3.04E-04	3.64E-04	9.56E-04	7.68E-04

Table 14 (cont'd)

ALIPHATIC SUBERIN MONOMERS		Adjusted p-Values							
		AtMYB92 (1-304) (n=3)	*AtMYB92 (1-304) + HA (n=4)*	*AtMYB93 (1-365) (n=4)*	*AtMYB93 (1-365) + HA (n=4)*	*AtMYB93 (1-359) (n=4)*	*AtMYB93 (1-348) (n=3)*	*AtMYB93 (1-335) (n=4)*	*AtMYB93 (1-335) + HA (n=3)*
HCAs	*trans*- ferulate	5.53E-04	4.00E-06	8.90E-03	3.25E-03	7.19E-03	1.52E-03	6.90E-05	1.40E-02
FAMEs	16:0	2.45E-03	2.14E-02	5.85E-02	2.86E-02	2.75E-01	1.86E-03	5.39E-02	6.02E-02
	18:2	3.37E-02	3.68E-02	5.85E-02	3.29E-01	1.36E-01	4.56E-01	9.11E-01	2.81E-01
	18:1	5.55E-04	2.01E-02	5.12E-01	6.12E-01	1.39E-01	6.73E-01	6.53E-01	6.67E-01
	18:0	8.30E-05	1.22E-02	5.85E-02	1.39E-01	2.75E-01	1.19E-01	6.53E-01	5.27E-01
	20:0	4.35E-04	1.79E-03	7.40E-04	1.77E-03	7.19E-03	6.07E-03	2.03E-03	4.76E-02
	22:0	3.19E-03	5.75E-03	5.85E-02	2.40E-04	1.35E-01	4.96E-04	1.22E-02	4.12E-01
	24:0	4.07E-03	4.82E-04	2.36E-04	2.19E-04	2.23E-03	2.91E-04	1.78E-03	6.94E-02
	26:0	2.38E-02	6.54E-02	3.36E-02	5.92E-01	1.35E-01	4.45E-01	6.53E-01	4.12E-01
DCA DMEs	16:0	9.30E-05	2.16E-03	2.64E-03	6.73E-04	3.40E-05	1.39E-03	1.10E-04	2.81E-01
	18:2	7.93E-04	6.77E-04	1.50E-03	2.62E-03	4.30E-05	1.39E-03	4.10E-05	1.79E-02
	18:1	8.03E-04	1.71E-03	1.46E-03	1.22E-02	3.98E-04	3.69E-03	3.53E-04	4.04E-02
	18:0	1.31E-03	1.09E-03	6.78E-04	1.27E-03	1.08E-04	3.82E-03	8.42E-04	2.01E-02
	20:0	1.48E-02	1.43E-02	8.99E-02	5.33E-03	1.36E-01	6.89E-02	8.35E-02	6.67E-01
	22:0	8.41E-04	5.50E-05	1.27E-04	1.16E-03	<0.000001	2.34E-03	7.00E-06	4.04E-02
	24:0	1.52E-02	5.80E-05	3.50E-03	3.54E-03	1.31E-04	2.34E-02	1.29E-02	1.72E-01
Hydroxy FAMEs	16-OH 16:0	7.71E-04	4.63E-04	8.90E-03	1.05E-04	1.36E-01	1.55E-03	3.20E-05	2.79E-01
	18-OH 18:1	1.38E-03	2.67E-04	6.89E-03	8.35E-04	9.10E-04	6.07E-03	3.51E-04	2.10E-01
	18-OH 18:0	3.78E-01	3.69E-02	4.65E-02	3.40E-01	1.35E-01	2.92E-01	9.24E-02	2.10E-01
	20-OH 20:0	3.45E-03	3.30E-05	2.29E-04	1.33E-03	8.77E-02	7.34E-03	3.91E-04	8.92E-02
	22-OH 22:0	5.17E-03	6.10E-05	2.29E-04	3.48E-04	1.36E-01	3.82E-03	6.40E-05	4.12E-01
	24-OH 24:0	1.14E-02	1.66E-04	1.47E-03	3.48E-04	1.54E-04	2.66E-03	7.60E-05	2.65E-02
Fatty Alcohols	1-OH 18:0	5.61E-03	9.06E-03	8.90E-03	1.20E-05	1.53E-04	1.53E-02	5.39E-02	9.70E-02
	1-OH 20:0	2.92E-02	1.22E-02	2.34E-02	5.19E-03	6.30E-05	7.65E-03	6.21E-04	1.45E-02
	1-OH 22:0	3.82E-03	1.37E-04	3.44E-03	9.80E-05	5.00E-06	6.21E-04	8.42E-04	4.79E-02
	1-OH 24:0	3.37E-02	4.22E-01	5.85E-02	2.64E-01	1.39E-01	7.15E-01	6.53E-01	6.14E-01
TOTAL		7.68E-04	1.97E-04	7.68E-04	7.68E-04	3.64E-04	7.68E-04	1.98E-04	6.01E-03