# Dissection of domain function for tandem AP2 plant transcription factors

Author: Harald Berg
Examiner: Leif Bülow
Faculty supervisor: Nélida Leiva Eriksson
SLU supervisors: Per Hofvander, Åsa Grimberg, Per Snell

Master of Science Thesis
Applied biochemistry
LTH faculty of chemistry

02-06-2018

# Abstract

Oils are some of the most energy dense compounds produced in nature and play a vital role in society as a source of food, energy and as raw material in many chemical processes. Plant produced oils (triacylglycerols) are a key asset in the transition away from fossil oils.
The progress made in recent years in molecular biology has greatly increased our understanding of plant oils biosynthesis. However, the regulatory mechanisms are not fully understood. Knowledge about these mechanisms could potentially lead to the development of new high yielding oil crops with tailored oil compositions.
In this report, three related transcription factors associated with oil biosynthesis and embryo development: WRINKLED1 (WRI1), WRINKLED4 (WRI4), and BABY BOOM (BBM), are studied for their effect on oil biosynthesis in relation to their structure. To this purpose, two hybrid transcription factors (HTFs) were created by swapping the DNA binding domain of WRI1 with those of BBM and WRI4.
The transcription factors were shown to have highly similar DNA binding domains, but only WRI1, WRI4 and WIR4 HTF induce oil production in leaf tissue.
This report identifies two potential regions within the DNA binding domain that are possible explanations for the difference seen in binding selectivity between the factors.

# Popular science abstract

Our dependence on fossil oil is one of the main problems facing humanity today, alternatives must be developed to stop the record levels of CO2 from rising any higher. But oil is central to so much in society; our transportation, our energy, our food and production of plastics and other materials are all dependant on it.

Plant oil is an environmentally friendly alternative to fossil oil and is therefore an important asset in the shift towards a more sustainable society. However, even with large scale cultivation of oil crops it would be hard to cover the global demand for oil through agricultural means alone. The use of arable land to grow oil crops for industrial purposes also rises an ethical dilemma; should the land be used to produce “green” oil or should it instead be used to feed earth's growing population. Consequently, there is a need for sustainably grown, high yielding oil crops that can efficiently use the land they are grown on. Furthermore, the progress seen in molecular biology has made it possible to engineer the composition of oil to better suit the needs of the industry. To this purpose, researchers have studied oil production in plants in order to figure out how to efficiently control it and possibly create new high-yielding oil crops with desirable lipid profiles.

Genetic modification is associated with the transfer of one or a few genes that bestow specific traits upon a target organism. For commercially grown crops that are genetically modified, the chosen genes often code for specific proteins that improve certain properties of the plants. Examples of such properties are protection against pests or pesticides or enzymes which alter specific characteristics of the plant. However, many traits seen in plants are the result of large numbers of genes being regulated simultaneously e.g. oil production. When these traits are the target for modification it is not possible to simply add a gene with the specific trait, instead the expression of a large panel of genes must be altered and this can potentially be done using DNA binding proteins called transcription factors (TFs).

In this report, the function of three similar TFs, WRINKLED1 (WRI1), WRINKLED4 (WRI4), and BABY BOOM (BBM), were studied for their effect on oil biosynthesis. Additionally, two hybrid transcription factors were created that are mixtures of WRI1 and the other two factors.
The results show that WRI1 and WRI4 could makes leaves produce oil but BBM could not, even though the DNA binding regions are very similar. Two potentials sequences within the genes of the TFs that could explain this difference were identified. Hopefully these regions will unlock ways to engineer new and better transcription factors, giving scientist more control over oil production in plants on a genetic level.

# Acknowledgements

When I first started my education in biotechnology at LTH, I had no idea what the subject entailed. But thanks to a few inspirational lectures by Leif Bülow, I quickly found my interest in genetics and molecular biology. In particular, I found genetic engineering of plants to be a fascinating subject, partly I think, because of its controversial nature. However, it was always just an interesting topic to talk and think about, never something I thought that I would do myself.

Therefore, it seems a bit surreal to have done this thesis. It might have been a lot of work from time to time, and when you're in the middle of it all, it becomes just regular lab work. But when I try to look at it from an outside perspective, to have created unique genes and see them in action in a living plant, I almost can't comprehend how cool that sounds to me. Therefore, I would like to make an honourable mention of those who made this possible.

To Nélida Leiva Eriksson for helping me along the way and to Leif Bülow for introducing me to this subject in the first place. I would also like to thank all the people at the plant breeding department of SLU who welcomed me and supported me in my work. Special thanks to Åsa Grimberg and Per Hofvander for making this project a reality and my sincerest gratitude to Per Snell, who has helped me each step of the way and taught me so much.

As this thesis marks the end of my five years at LTH, I would also like to thank all my friends from Karbokatt for sticking with me through these years and Mimmi Castmo for everything she has given me.

# Contents

# List of Abbreviations

| | |
|---|---|
| ACP | Acyl Carrier Protein |
| AP2 | Apetala2 |
| AtBBM | BABY BOOM from *A. thaliana* |
| AtWRI1 | WRINKLED1 from *A. thaliana* |
| AtWRI4 | WRINKLED4 from *A. thaliana* |
| BBM | BABY BOOM |
| BPM | BTB/POZMATH |
| CoA | Coenzyme-A |
| DAG | Diacylglycerol |
| DNA | Deoxyribonucleic Acid |
| ER | Endoplasmic Reticulum |
| FAS | Fatty Acid Synthase |
| GC | Gas Chromatography |
| GFP | Green Fluorescent Protein |
| HTF | Hybrid Transcription Factor |
| IDR | Intrinsically Disordered Region |
| MAG | Monoacylglycerol |
| PCR | Polymerase Chain Reaction |
| PLT | PLETHORA |
| RNA | Ribonucleic acid |
| TAG | Triacylglycerol |
| T-DNA | Transfer-DNA |
| TF | Transcription Factor |
| TLC | Thin Layer Chromatography |
| WRI1 | WRINKLED1 |
| WRI4 | WRINKLED4 |

# 1 Background

Oils are some of the most energy dense compounds produced in nature. This property makes them highly important for society, as a source of food, energy and as raw material in many chemical processes.

Plant oil is a de facto alternative to fossil oil and is therefore an important asset in the shift towards a more sustainable society. However, even with large scale cultivation of oil crops it would be hard to cover the global demand for oil through agricultural means alone. The use of arable land to grow oil crops for industrial purposes also rises an ethical dilemma; should the land be used to produce “green” oil or should it instead be used to feed earth's growing population. Consequently, there is a need for sustainably grown, high yielding oil crops that can efficiently use the land they are grown on. Furthermore, the progress seen in molecular biology has made it possible to engineer the composition of oil to better suit the needs of the industry. To this purpose, researchers have studied lipid metabolism in plants in order to figure out how to efficiently control oil production and possibly create new high-yielding oil crops with desirable lipid profiles.

WRINKLED1 (WRI1), a member of the Apetala2/ethylene responsive element-binding protein (AP2/EREBP) transcription factor (TF) family, has been shown to regulate expression of genes encoding glycolytic and fatty acid biosynthetic enzymes during seed development in *Arabidopsis thaliana* (L.). Mutants lacking the *wri1* gene has been shown to have a large deficiency in seed oil. (Cernac and Benning, 2004; Focks and Benning, 1998). However, overexpression of *WRI1* has been shown to have an opposite effect and increase the amount of oil in tissues of non-oil crops, even if the gene is from a distantly related plant. Studies have shown that WRI1 can induce oil accumulation in potato tubers (Hofvander et.al, 2016) and sugarcane (Zale et al., 2015). This TF shows promise in the search for new high-yielding oil crop, but the significance of its physical structure is still largely unknown.

To understand the structural significance of the domains that make up WRI1 this report compares the oil inducing effect of WRI1 and two closely related TFs, WRINKLED4 (WRI4) and BABYBOOM (BBM), via agroinfiltration of *Nicotiana benthamiana* Domin. Furthermore, Hybrid TFs (HTF) where created by swapping the DNA binding domain of WRI1 with the other factors, HTFs were injected along the native TFs to give insight into domain function. The effects on lipid accumulation was studied in relation to the primary and secondary structure of the proteins. Furthermore, the seeds of Arabidopsis knockouts lacking one of the three TFs was analysed for changes in lipid content.

## 1.1 Plant lipids

The word lipid is a loosely defined term used to describe a wide range of biomolecules that are soluble in nonpolar solvents. The Lipid Metabolites and Pathways Strategy (LIPID MAPS) consortium, outlined by Fahy et al. (2011), describes eight different categories of lipids: **Fatty acyls** are essentially hydrophobic carbon chains bound by an acyl functional group, these includes fatty acids, fatty alcohols and fatty esters to name a few. Fatty acyls are important building blocks in many biochemical processes not least in the formation of wax esters which protect plants from pathogens and other external stresses. Furthermore, fatty acyls are constituent in some of the other lipid categories. **Glycerolipids** consist of hydrophobic fragments, mainly acyl lipids, bound to a glycerol molecule. Mono- and diacylglycerols (MAG, DAG) can be further processed to produce constituents of the other lipid categories or be bound to galactose forming mono- and digalactosyldiacylglycerol, which are a major component in plastid membranes. Furthermore triacylglycerol (TAG) is an important carbon and energy storage compound. **Glycerophospholipids** are as the name suggests closely related to glycerolipids, however they differ in the presence of a phosphate group bound to the glycerol backbone. They are a major component in lipid membranes due to their amphiphilic nature. **Sphingolipids** contain a long chain nitrogenous base as backbone and serve as constituents in cell membranes. Additionally, they serve as signalling molecules in various regulatory processes (Bartke and Hannun, 2008). **Sterols** are important as they regulate the permeability and fluidity of cell membranes (Piironen et al., 2000). **Prenols** make up a minor part of plant lipid production but contain important members like terpenes and carotenoids. **Saccharolipids** and **Polyketides** are relatively unstudied in the field of plant lipid research as they are mostly associated with bacteria and fungi. The fatty acyls, glycerolipids, glycerophospholipids and sphingolipids are also known as the acyl lipids and they make up most of plant lipids biomass (Li-Beisson et al. 2013; Nakamura and Li-Beisson, 2016). TAG has by far the most economical impact as it is the main storage compound for almost all important oil crops (Ruuska et al., 2002). TAG metabolism will therefore be a main focus in the following text.

### 1.1.1 Acyl lipid metabolism in plants

Plant fatty acid synthesis occurs mainly in the plastid, but also in the mitochondria, and is dependent on pyruvate as a primary building block. Pyruvate is formed from carbohydrates via the glycolytic pathway that takes place in the plastid as well as the cytosol. However, there is evidence that plastidic glycolysis is not solely responsible providing building materials for fatty acid synthesis. Instead, there seems to exist a high degree of interplay between the two glycolytic pathways where glycolytic intermediaries are transported back and forth between the plastid and the cytosol. In particular, transfer of phosphoenolpyruvate and other 3-carbon molecules, from the cytosol to the plastid, has been pointed out as a key feature in oil accumulating tissues (Andriotis et al., 2009; Grimberg, et.al, 2015). A rough schematic of the biosynthetic pathway from carbohydrate to acyl lipid and finally TAG can be seen in figure 1.

In the plastid, pyruvate is converted into malonyl-CoA a building block for the acyl lipid metabolism, this is the first committed step in acyl lipid synthesis. The malonyl residue is transferred from coenzyme A to an acyl carrier protein (ACP) forming malonyl-ACP which subsequently enters the fatty acid synthesis process carried out by fatty acid synthase (FAS). FAS is a large protein complex consisting of multiple subunits which together catalyse a sequence of four reactions: condensation, reduction, dehydration, and reduction. The reactions results in the addition of a two-carbon fragment on to a growing acyl lipid backbone. After seven reaction cycles within the fatty acid synthase a 16:0-ACP molecule is formed. The 16:0-ACP molecule can then be elongated once more to form 18:0-ACP and further desaturated to form 18:1-ACP. The long chain acyl lipid/ACP moieties can then take part in the production of phospho-, sulfo-, and galactolipids as well as other compounds in the plastid or be hydrolysed and activated to CoA esters which can be transferred to the endoplasmic reticulum (ER) for further processing (Li-Beisson et al., 2013; Nakamura and Li-Beisson, 2016).

In the ER the acyl-CoA can partake in TAG biosynthesis which involves a series of acylations and subsequent phosphorylation of glycerol-3-phosphate forming a number of intermediaries before being converted to DAG and finally TAG. The phospholipid pathway shares several intermediaries with TAG assembly in the ER and is therefore a possible destination for the acyl-CoAs. Furthermore, single acyl-CoAs can be further elongated into very long fatty acids (mostly 24:0 longer) and/or enter the sphingolipid pathway or cuticle biosynthesis via the cutin polyester or the cuticular wax biosynthetic pathway in the ER. DAG intermediaries can also be transported back to the plastids for galacto and sulfolipid synthesis.

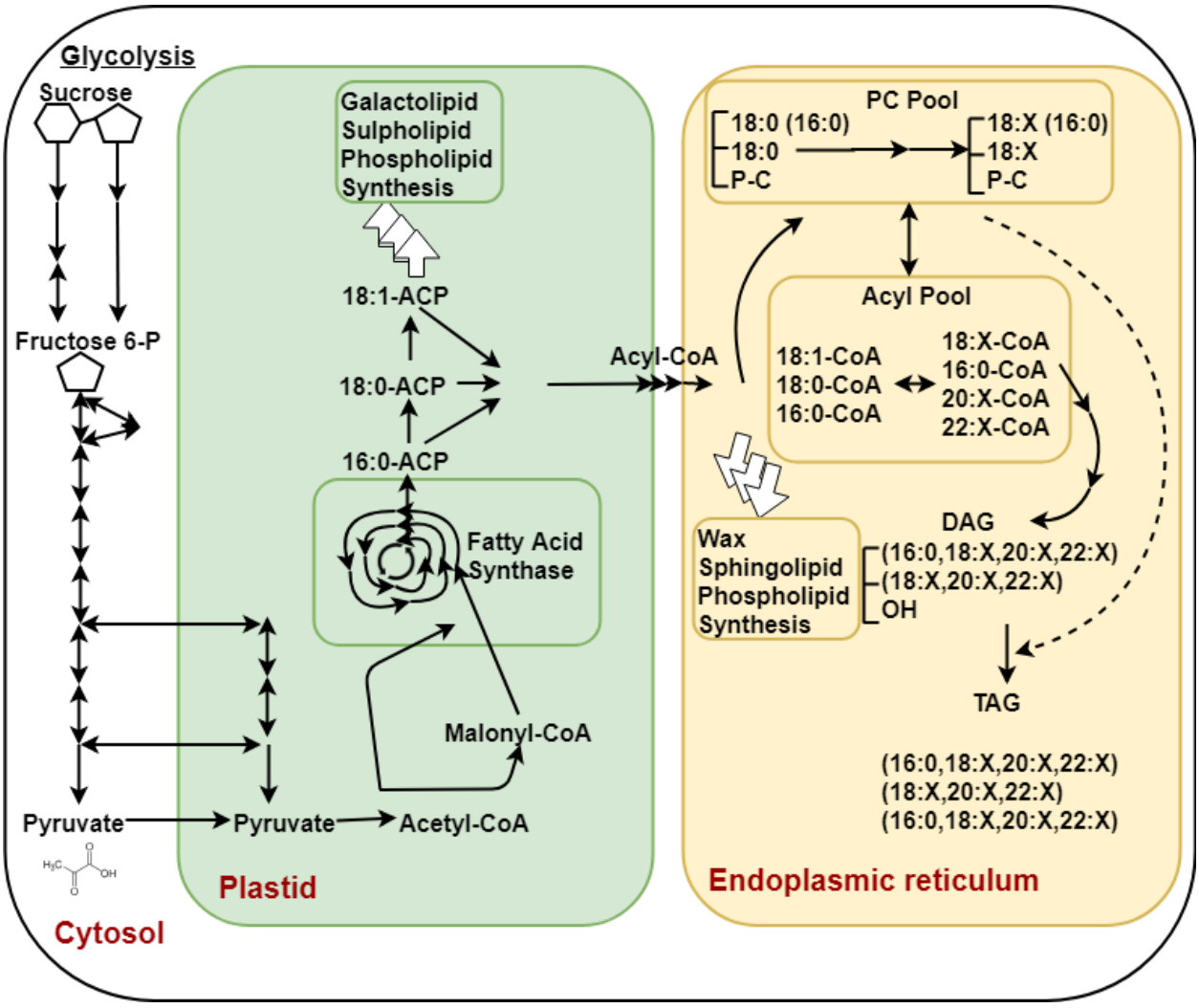


*Figure 1: Schematic view of triacylglycerol biosynthetic pathway. Glycolytic degradation of sucrose in the cytosol produces pyruvate and other 3-carbon intermediaries which are transported into the plastid. In the plastid Malonyl-CoA are fused into fatty acid and acyl carrier protein moieties (ACP) by fatty acid synthase. The 16:0-ACPs can be further elongated and desaturated before being transported into the endoplasmic reticulum (ER). Acyl-ACPs can be further elongated and unsaturated in the phosphatidylcholine (PC) pool or be added to a glycer-3-phosphate molecule forming triacylglycerol (TAG). Adapted from Li-Beisson et al. 2013 and Grimberg, et.al, 2015*

## 1.2 Transcription and transcription factors

Transcription is an important step in gene expression, and it is the process in which genetic information is transcribed from DNA to RNA. Gene expression signifies the flow of genetic information from DNA to a functional gene product, often in the form of a protein. The reaction itself is carried out by RNA-polymerase, which is a large protein made up of multiple subunits. Eukaryotes have multiple RNA-polymerases, of which RNA-polymerase II might be the most noteworthy as it is responsible for transcription of all protein coding genes as well as some small RNA molecules. There are three main strategies that regulate transcription within the DNA. The first involving the accessibility of the gene in question and is the result of the way DNA is packaged into a structure called chromatin by the use of packaging proteins called histones. The second involves the binding of activators or repressors to non-coding sequences within the genome called cis-acting regulatory elements. The third involves methylation of the DNA which interferes with the binding of proteins to the DNA sequence. However, the three mechanisms are not separate phenomenon, instead, they are all interconnected and work together to create a specific expression pattern, the common denominator for all mechanisms being the presence of small DNA binding proteins called transcription factors that participate in regulation of transcription (Berg et al., 2012; Cooper, 2000; Lee and Young, 2000).

Transcription Factors (TFs) are proteins that bind to cis-acting regulatory elements and alter the rate of transcription. A common feature for all TFs is that they possess a DNA binding domain. TFs that share a certain type of DNA binding domain are considered members of the same TF family. Furthermore, there is one group called the general TFs which are essential constituents of the preinitiation complex and thereby the expression of all genes. For protein coding genes, the general TFs include RNA polymerase II as well as several other factors that bind to the basal promoter near the transcription start site. The promoter region, illustrated in figure 2, is located upstream of the coding sequence of a gene and contains some key elements that influence transcription. The remaining TFs, also known as specific TFs, are highly diverse and can bind to cis-regulatory elements anywhere from within the gene itself to several thousand base pairs in either direction of the transcription start site. They regulate the transcription rate by interactions with the preinitiation complex via coactivators, such as the mediator complex, or corepressor as well as by affecting the accessibility of the gene in question. Accessibility of a gene can be altered by interactions with chromatin through histone modifying enzymes or chromatin remodelling factors (Gonzalez, 2016; Cooper, 2000; Lee and Young, 2000).
By associating cis-regulatory elements with large panels of genes eukaryotes can efficiently control transcriptional transitions in response to stress or to mediate cell specification and development. This ability makes them invaluable and is a reason for the large variation and abundance observed in eukaryotic genomes where they make up 3–10% of all genes (Levine and Tjian, 2003; Harbison et al., 2004; Reece-Hoyes et al., 2005).

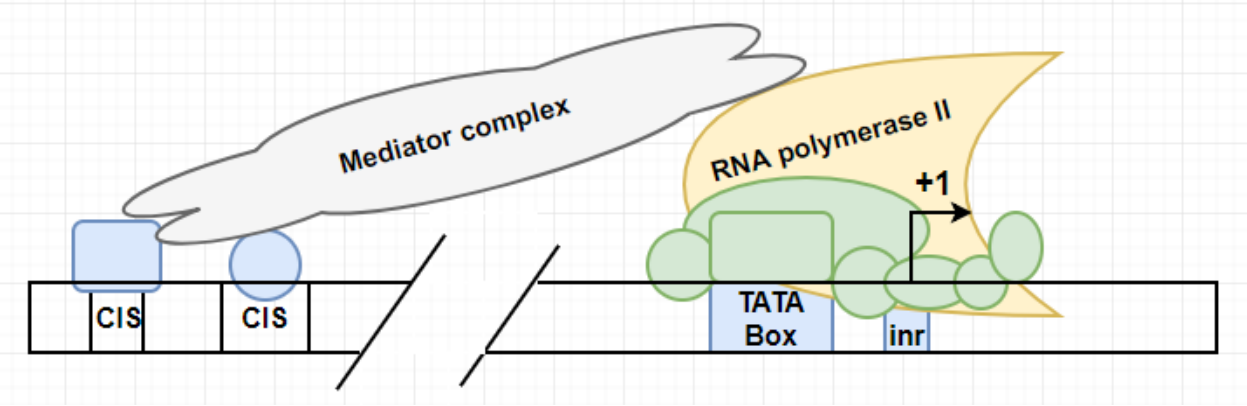


*Figure 2: A rough schematic of the promotor region and interactions with the preinitiation complex. Green figures signify the general transcription factors that are essential constituents of the preinitiation complex, blue figures are specific transcription factors that bind to upstream cis-regulatory elements and interact with the preinitiation complex through the large multicomponent mediator complex. As adapted from Berg et al., 2012*

### 1.2.1 The Apetala2 family of transcription factors

AP2/ERF is a plant specific TF family containing AP2/ERF DNA binding domains, the family is encoded in 145 loci in the *A. thaliana* genome and is associated with various developmental processes in plants (Riechmann and Meyerowitz, 1998; Jong, 2016). 14 of these belong to a subfamily referred to as the AP2 sub family which is distinguished by the presence of two AP2/ERF domains (Jong, 2016).

The AP2 domain can be divided into two blocks, the first consists of around 20 amino acids and a highly conserved YRG motif while the second consists of around 40 amino acids and a RAYD motif. The DNA binding ability of the AP2 domain is thought to be linked to a 20 amino acid amphipathic alpha-helix present within the RAYD block. Additionally, the AP2 subfamily contains a highly conserved 25-26 amino acid linker region between the two AP2 domains (Okamuro et al., 1997).

- WRI1

WRINKLED1 (WRI1), a member of the AP2 TF subfamily has been shown to induce expression of genes encoding for glycolytic and fatty acid biosynthetic enzymes during seed development in *A. thaliana*. (Focks and Benning, 1998; Cernacand and Benning, 2004) Loss of function mutants of this gene in *A. thaliana* have a large deficiency in seed oil accumulation giving the seeds a wrinkled phenotype. WRI1 derived from *A. thaliana* (AtWRI1) and other homologues have been thoroughly studied for their oil inducing effect as potential targets in the development of new high yielding oil crops (Hofvander et.al, 2016; Vanhercke et al., 2013; Grimberg, et.al, 2015). The TAG accumulating effect of WRI1 is mainly attributed to upregulation of glycolytic enzymes in the cytosol, pyruvate transport from the cytosol to the plastids and upregulation of enzymes connected to fatty acid synthesis in the plastid (Baud et al., 2007; Grimberg, et al, 2015). WRI1 function as a key regulator of fatty acid synthesis in oil seeds has been shown to be

a downstream target of master regulators involved in seed and embryo development, such as the LAFL (LEC1, ABI3, FUS3 and LEC2) network (Baud et al., 2007).

Several studies have been performed that try to uncover the structural role of the different domains of AtWRI1. A study by Ma et al. (2015) examined the role of intrinsically disordered regions (IDRs) on the stability and transactivational ability of AtWRI1. The TF was described as a structured double AP2 domain encompassed by one IDR on the N-terminal side (IDR1) and two IDRs on the C-terminal side (IDR2 and IDR3), a schematic of the primary structure of WRI1 can be seen in figure 3 below. The study showed that removal of part of IDR3 (amino acids 398-430) increased the stability of AtWRI1. The results were ascribed to the removal of a PEST motif. A PEST motif is a sequence rich in the amino acids proline (P), glutamic acid (E), serine (S), and threonine (T), which has previously been shown to play an important role in BTB/POZMATH (BPM) mediated protein degradation of a related TF (Gregorio et al., 2014) and WRI1 (Chen et al., 2013). Furthermore, Ma et al. showed that the transactivational ability of AtWRI1 was dependant the region between residues 307-397 and pointed to the high occurrence of acidic residues in the region as an indicator of a transactivating domain (Ma et al., 2015; Mitchell and Tjian, 1989) Residues 78-92 of AtWRI1 have been identified as a potential location for interaction with BPM and might therefore influence its stability (Chen et al., 2013). Furthermore, this region has been shown to interact with a 14-3-3 protein which belongs to a group of proteins that are known regulators of various biological processes, including effects on TF stability in mammalian cells and nucleocytoplasmic transport of plant TFs. Transient coexpression of a 14-3-3 proteins and AtWRI1 in *N. benthamiana* resulted in higher levels of TAG accumulation than expression of AtWRI1 alone and it is therefore theorized that the 14-3-3 proteins have a stabilizing effect on AtWRI1 expression (Ma et al., 2016).

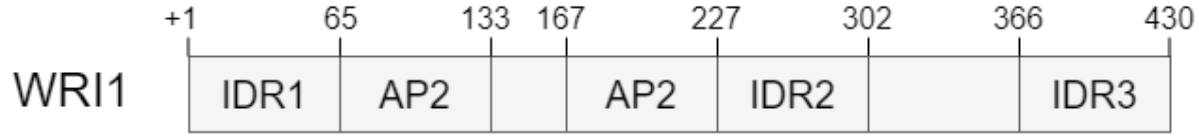


*Figure 3: The location of different domains of WRI1 within the amino acid sequence. Three intrinsically disordered regions are present, numbered from the N-terminal (+1). A tandem AP2 domain is located in between IDR1 and IDR2. Adapted from Ma et al., 2015*

- WRI4

WRINKLED4 (WRI4) is a member of the AP2 TF subfamily and a relative to WRI1. WRI4 seem to have the same ability for activation of the glycolytic and lipid biosynthetic pathway as WRI1 and has been associated with wax biosynthesis in stem epidermis (Park et al., 2016). However, while high activity of WRI1 is seen in seed development, WRI4 levels remain low, implying a large difference in gene regulation if not function of the two factors (Park et al., 2016). This is further evident when looking at mutants with non-functional versions of *wri1* that are rescued with WRI4 constructs behind a promoter sequence associated with seed development. The rescued seeds showed a reversion from the wrinkled phenotype providing

additional evidence that the WRI4 and WRI1 differ mainly in the regulation of their expression (To et al., 2012).

- BBM

BABYBOOM (BBM), also known as PLT4, is often grouped with the AINTEGUMENTA-LIKE clave of TFs and a member of the AP2 TF subfamily. The group contains the PLETHORA (PLT) TFs that are associated with embryonic development and maintenance of meristem tissue. The meristem is a region of undifferentiated plant cells that provide new cells for tissue and organ development thereby allowing for growth. (Aida et al., 2004; Galinha et al., 2007; Santuari et al., 2016). Overexpression of BBM causes somatic embryogenesis, which is the development of embryos in plant tissue that is not normally involved in the development of embryos e.g. leaves (Horstman et al., 2017). The transcriptional transitions in tissues exposed to BBM is thought to arise from interactions with the LAFL network of master regulators (Horstman et al., 2017; Boutilier, 2002). The LAFL network has been shown to be crucial regulators of embryo development, storage compound accumulation and seed maturation. Linking BBM to WRI1 as a potential upstream regulator for LAFL which in turn regulates WRI1(Fatihi et al., 2016). Limited research has been done on the significance of the structure of BBM and why it affects an entirely different panel of genes when its DNA binding domain is highly similar to that of WRI1.

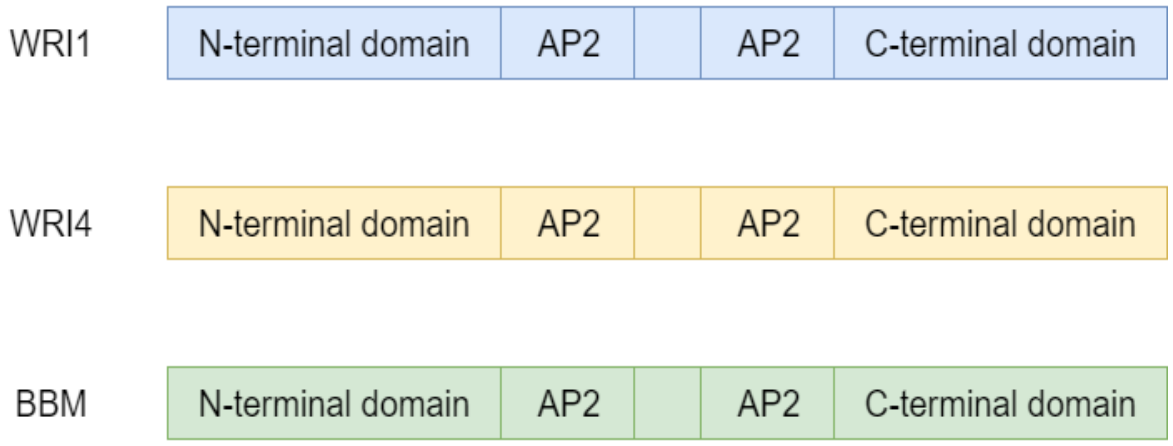


*Figure 4: The basic domains of WRI1, WRI4 and BBM. Not evident from this figure is that the N- and C- terminal varies greatly in length between the three factors but the AP2 region is in the middle and show a high degree of conservation between the three factors.*

## 1.4 Gateway cloning

The Gateway cloning system is a method for transferal of DNA-fragments into plasmids. While traditional methods often utilize restriction enzymes and ligases to introduce new genetic material in to vectors, Gateway cloning utilize two enzyme mixtures called BP and LR clonase. The clonases are derived from the lambda bacteriophage which use them to introduce its DNA in to the *Escherichia coli* chromosome. The target gene and its intended position in the target vector is encompassed by *att*-sites, that are recognized by one of the clonases. When active, the clonase will cut out the gene at a specific position in its *att*-sites and incorporate the gene between two positions in the vectors corresponding *att*-sites. The incorporated gene is now encompassed by a combination of the previous *att*-sites that together form a third kind of *att*-site, that is recognized by the second clonase, the process is illustrated in figure 5. The gene can therefore be transferred in a similar way by the second clonase and the resulting att-sites from the second reactions recreate the original sites that were encompassing the gene making it easy to change vectors multiple times.

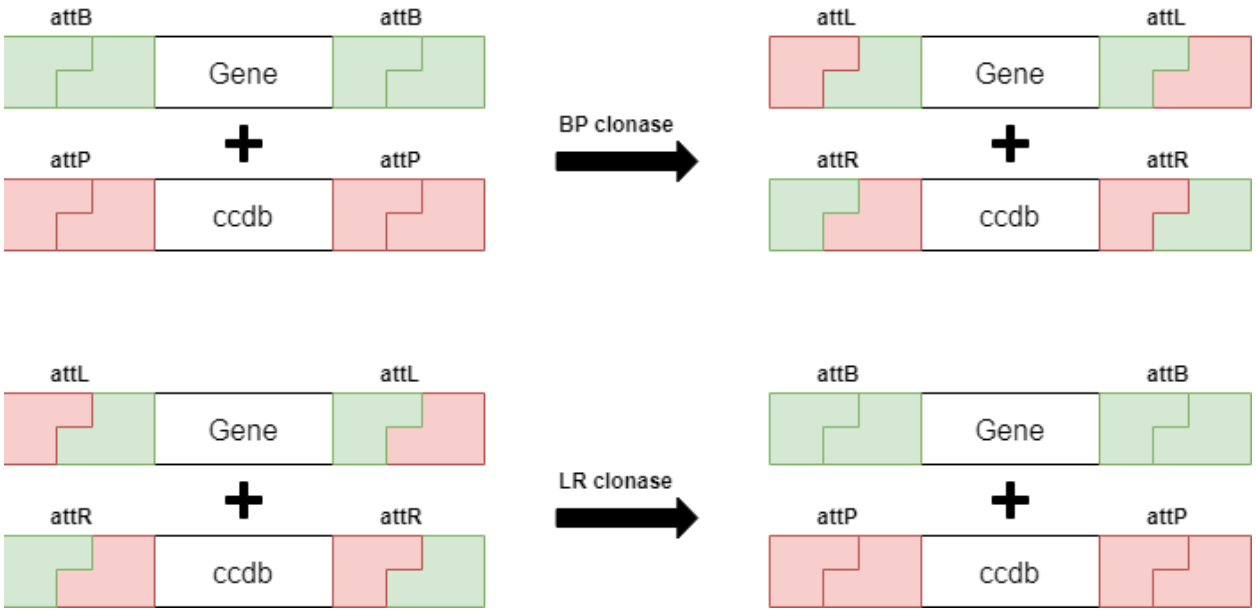


*Figure 5:An illustration on the two main steps of the gateway system. First step is the insertion of a gene in to a donor vector, originally containing the ccdb gene which is lethal to the bacteria that is transformed with the products of the BP reaction. This results in that only bacteria with vectors containing a successfully transferred gene will proliferate and amplify the correct construct. The plasmids can then be transferred to a destination vector via LR clonase which essentially does the reverse reaction of BP clonase.*

Some advantages with this technology is that the orientation of the gene is controlled and later screening for correctly transformed bacteria is simplified by the use of ccdB as a negative selection marker in the destination vector. It also limits the complications associated with restriction enzymes and results in a more efficient cloning procedure than traditional techniques. (User guide to: Gateway® Technology A universal technology to clone DNA sequences for functional analysis and expression in multiple systems, 2018)

## 1.5 Transient gene expression in *Nicotiana benthamiana*

*A. tumefaciens* has been well studied in plant science ever since its discovery as a causative agent for the crown gall disease (Smith and Townsend, 1907). The disease is characterized by tumorous growths on the stem of dicotyledonous plants and has a negative impact on many agriculturally important crops (Hoekema et al., 1983; Sheng and Citovsky, 1996). The mechanism of action is explained in a review by Zupan et al, *A. tumefaciens* infects its host with transfer DNA (T-DNA), which includes genes coding for growth promoting plant hormones, leading to tumour development. The transferred genes also code for enzymes that induce production of amino acid-like compounds called opines, that can subsequently be used as a carbon and nitrogen source for the bacteria. The T-DNA as well as the virulence genes needed for transfection has been located in the tumour inducing plasmid (Ti-plasmid) present in pathogenic strains of the bacteria (Sheng and Citovsky, 1996). Two direct repeats have been found to border the T-DNA region of the Ti plasmid, by replacing the native genes between these repeats, the virulence system can be used insert any replacing genes in to the nucleus of the host (Zupan et al., 2000).

A method for transient expression in plants using infiltration of *Agrobacterium tumefaciens* was developed by Kapila et al., (1997). Transient gene expression, unlike stable gene expression, is the process in which an organism is transfected with a gene that will not be passed on to subsequent generations. The gene can be incorporated into the genome of the infected cells or be transiently present in the nucleus where it exhibits higher levels of transcription than a stable transformation would. Unintegrated DNA will be gradually degraded, and expression levels will therefore decrease with time, but the method has other advantages. Some of these advantages are: results can be measured shortly after transfection, expression is not affected by where the gene is inserted in the genome, and much of the laborious process of regeneration from a transformed cell into a transgenic plant is bypassed by transformation of already mature plants.

The method was further developed by Wood et al. (2009), who used it to assemble multicomponent metabolic pathways in *N. benthamiana*. The successful result would have been laborious to achieve through stable transformation. The method has since been widely used in plant science. One example of its use is in the study of transcriptional transitions and metabolic changes of WRI1 induction (Grimberg et al., 2015).
Transfection of *N. benthamiana* is achieved by injection of an infiltration mixture into the intercellular space of a developing leaf where there is still a high level of protein synthesis. The infiltration mixture contains *A. tumefaciens* transformants with modified Ti-plasmids containing the gene construct or constructs of interest as well as transformants containing a virulence gene (*p19*) that suppresses RNA silencing. Identification of successfully transfected regions can be facilitated by co-infiltration with green fluorescent protein (GFP) (Wood et al., 2009).

# 2 Material and Methods

## 2.1 Bioinformatic analysis

A comparative analysis of the primary structures of AtBBM, AtWRI4 and AtWRI1 was performed using the sequence alignments tool available in CLC main workbench 8.0 software (Qiagen, Hilden), with the default settings.
The results were compared with the sequences of WRI1 homologues as well as PLT factors from *A. thaliana*.

Secondary structures of the TFs, in particular the AP2 domains, were compared using the secondary structure prediction tool in CLC main workbench 8.0 software (Qiagen, Hilden).
Intrinsically disordered regions (IDRs) were predicted using the online tool PONDR (Molecular Kinetics, Inc., Indianapolis, USA) and the VL3-BA prediction algorithm developed by Obradovic et al., (2003).

Sequences used in primer design and bioinformatic analysis were obtained from full length coding sequences provided by The Arabidopsis Information Resource (TAIR) for *AtWRI1* (AT3G54320.1), *AtWRI4* (AT1G79700.1), *AtBBM* (AT5G17430.1) as where the full-length coding sequences for *A. thaliana PLT*-1,2,3,5 and 7 (Arabidopsis.org, 2018). The sequence for homologous *WRI1* sequences from potato embryo (*StWRI1em*), tobacco (*NbWRI1*), poplar stem (*PtWRI1ca*), oat endosperm (*AsWRI1es*), and nutsedge tuber parenchyma (*CeWRI1tp*) were obtained from the cDNA sequences published by Grimberg et al., (2015).

All sequences are shown in appendix 1

## 2.2 Plant material

All plants were grown in climate-controlled growth chambers (Biotronen, SLU, Alnarp, Sweden) with a light intensity of 250 µmol/$m^2$s. The plants grew under 70% relative humidity and a 16h/ 8h light/dark period at 25°C/23°C respectively.

Three batches of around 15 *N. benthamiana* plants each were grown for agroinfiltration purposes.

*A. thaliana wri1* (CS69538 from Per Snell), *wri4* (SALK_048920 from Joseph Ecker), *bbm* (SALK_097021 Joseph Ecker) knockouts were grown for seed TAG analysis to be compared with wild type *A. thaliana* col-0.

*A. thaliana wri1* knockout seeds were surface sterilized by soaking once with 70% ethanol for 2 min, and once with 50% bleach and a small amount of Triton X-100 for 5 min. The seeds were washed three times with sterile MQ-$H_2O$ before being plated on petri dishes containing solid half strength Murashige and Skoog with sucrose medium.

The Murashige and Skoog with sucrose medium was prepared by adding 0.55 g MS-salts, 0.1 g Mes, 2.5 g sucrose to 250 ml water and adjusting pH to 5,8 before the final addition of 3,75 g phyto agar. After sterilization in autoclave, 25-30 ml of medium was poured in to empty petri dishes under sterile conditions.

## 2.3 Gene constructs

### 2.3.1 Acquisition of genetic material

*Escherichia coli* stocks with pENTR/TOPO vectors containing the coding sequences for WRI4 (stock# TOPO-U10-G01) and BBM (stock# TOPO-U15-C07) were ordered (Arabidopsis Biological Resource Center, The Ohio State University, Columbus OH, USA). The stocks were transferred to agar plates containing LB medium with kanamycin (50 µg/ml) and incubated overnight at 37°C. Single colonies were isolated for amplification in 5 ml LB medium containing kanamycin and incubated at 37°C with 225 RPM agitation overnight. Plasmids were recovered, se section 2.3.3 - *Recovery and Verification*, and used as a template in subsequent PCR reactions. 1 ml of each culture was not used for recovery but instead transferred to a new eppendorf tube. After addition of 1 ml 50% glycerol both cultures were immediately snap frozen in liquid nitrogen and put into a storage freezer at -80°C.

Purified plasmids containing *AtWRI1*, TAIR accession number: AT3G54320.3, were obtained from existing stock to be used in synthesis of hybrid constructs.

Primers were designed and synthetically constructed by an external producer (Integrated DNA Technologies, Inc. Skokie, USA)

### 2.3.2 Creation of gene constructs

*WRI4* and *BBM* genes were amplified by PCR using loose end primers containing the *att*B sites used in the Gateway system. Pre-existing stocks of *A. tumefaciens* with pXZP393 plasmids containing *AtWRI1* were readily available, therefore only *AtBBM* and *AtWRI4* were continued throughout the gateway process.

Two hybrid gene constructs were synthesised using fusion PCR, the constructs contained the double AP2 region from *AtBBM* and *AtWRI4* respectively, encompassed by the N- and C-terminal domain of AtWRI1. A simplified illustration of the hybrid gene constructs is shown in figure 6 below.

HTFs were created by separate amplification of the N-terminal, AP2 and C-terminal domains using loose end primers which overlapped on to the next gene segment. The primer length was designed so that primer pairs had matching melting temperatures. The separate domains were later fused by PCR of the three overlapping domains and the two end primers in the same reaction mixture. The process is explained further in figure 6 below where the red arrows indicate the primers used for amplification of the individual gene segments, while the black arrows indicate the primers used for both native and hybrid constructs as well as in the final fusion PCR.

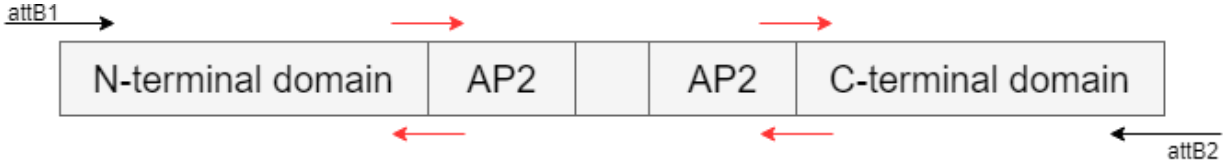


*Figure 6: An illustration of the construction of hybrid transcription factor (HTF) constructs. The N- and C-terminal Domain of AtWRI1 were amplified using lose end primers that overlapped with the sequences of the AP2 domains of AtWRI4 and AtBBM. Similarly, the AP2 domains of AtWRI4 and AtBBM were amplified with open end primers that overlapped with the sequence of AtWRI1. N- and C-terminal domains were combined with their matching AP2 domain, one with AtWRI4 AP2 domain and one with AtBBM AP2 domain, and a PCR reaction was performed together with the attB1 and attB2 primers to fuse the three domains. Red arrows signify primers used in the amplification of each domain. Black arrows signify primers used in both the amplification of the N- and C-terminal domains as well as in the final PCR to fuse the HTF construct.*

A total of 10 PCR reactions were performed, two for the full length *AtWRI4* and *AtBBM* genes one for each AP2 domain and two for the N and C domains of *AtWRI1* and two final PCRs to fuse the HTF genes. All PCR reaction mixtures were produced according to the Phusion High-Fidelity DNA Polymerase instruction sheet (Thermo Scientific, Waltham, USA) for 20 µl mixtures with 4 ng template per reaction mixture. The only exception being the final fusion reaction of the HTFs were volumes corresponding to 100ng of each gene segment were added. The length of the coding sequence of each hybrid gene construct as well as for *AtBBM*, *AtWRI4*, and *AtWRI1* are shown in table 1.

*Table 1: Lengths of the coding sequences for AtBBM, AtWRI4, AtWRI1 and hybrid transcription factors BBM HTF and WRI4 HTF*

| Gene | *AtBBM* | *AtWRI4* | *BBM HTF* | *WRI4 HTF* | *AtWRI1* |
|---|---|---|---|---|---|
| Length bp | 1755 | 942 | 1287 | 1287 | 1293 |

A combined 3-step and 2-step PCR reaction protocol was used in the first amplification of *AtWRI4* and *AtBBM*

The reaction for amplification of each domain was run on a more time efficient two step protocol involving 35 cycles of; 10s denaturation(98°C) and 60 s combined annealing and elongation (72°C). The fusion reaction of the HTFs was run on a combined 3-step and 2-step protocol. The protocol first 3-step part of the protocol involved 14 cycles of; 10s denaturation (98°C), 15 s

annealing (59°C for *BBM* and 62°C for *WRI4*) and 60 s elongation (72°C). The second part involved 17 cycles of 10 s denaturation (98°C) and 60 s annealing and elongation (72°C). All reactions were finalized with a 5 min final extension (72°C) and stored in the PCR machine (4°C) until further processing. The sequences for all primers used are presented in table 2.

*Table 2: The name, sequence and annealing temperature of all primers used to construct the gateway compatible gene constructs. Annealing temperature was calculated using an online tool (Thermofisher.com, 2018)*

| Primer name | Nucleotide sequence | Annealing Temp °C |
|---|---|---|
| attB1-AtWRI1_f | 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAAGAAGCGCTTAACCACTTCC-3' | 66.2 |
| attB2-AtWRI1_r | 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTCAGACCAAATAGTTACAAGAAACCGA-3' | 65.3 |
| attB1-AtBBM_f | 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAACTCGATGAATAACTGGTT-3' | 59.4 |
| attB2-AtBBM_r | 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTAAGTGTCGTTCCAAACTG-3' | 56.4 |
| attB1-AtWRI4_f | 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCAAAAGTCTCTGGGA-3' | 61.2 |
| attB2-AtWRI4_r | 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTTAAGGTCCATAATCAAACTCG-3' | 58.6 |
| WRI1(N):BBM(AP2)_f | 5'-CTACCCGACGCAGCTCTATC:TACCGCGGTGTTACAAGG-3' | WRI1 62.7 : 61.1 BBM |
| WRI1(N):BBM(AP2)_r | 5'-TGCCTTGTAACACCGCGGTA:GATAGAGCTGCGTCGGGTAG-3' | WRI1 67.3 : 62.7 BBM |
| BBM(AP2):WRI1(C)_f | 5'-ACATGAACAGATACAATGTT:CGGTTAAAGAAGAAAGGTGTTTTCC-3' | WRI1 64.6 : 50.6 BBM |
| BBM(AP2):WRI1(C)_r | 5'-AACACCTTTCTTCTTTAACCG:AACATTGTATCTGTTCATGTCGAAG-3' | WRI1 57.9 : 61.9 BBM |
| WRI1(N):WRI4(AP2)_f | 5'-CTACCCGACGCAGCTCTATC:TACAGAGGC | WRI1 62.7 : 58.9 WRI4 |

| | | |
|---|---|---|
| | GTCACAAGG-3’ | |
| WRI1(N):WRI4(AP2)_r | 5’-TGCCTTGTGACGCCTCTGTAGATAG:AGCTGCGTCGGGTAGAAG-3’ | WRI1 68.1 : 62.1 WRI4 |
| WRI4(AP2):WRI1(C)_f | 5’-ACGTCAGCCGTTATCTAAAC:CGGTTAAAGAAGAAAGGTGTTTTCC-3’ | WRI1 64.6 : 57.4 WRI4 |
| WRI4(AP2):WRI1(C)_r | 5’-ACACCTTTCTTCTTTAACCG:GTTTAGATAACGGCTGACGTCG-3’ | WRI1 56.5 : 63.5 WRI4 |

### 2.3.3 Recovery and Verification

All plasmids were recovered using the GeneJET Plasmid Miniprep kit (Thermo Scientific, Waltham, USA). Following the kit provided protocol, 4 ml of cell culture was harvested by centrifugation at 6800 g for 2 min. The supernatant was decanted, and the remaining pellet was resuspended in 250 µl resuspension buffer supplied by the kit. 250 µl lysis solution was added to the cell suspension and after careful mixing 350 µl of neutralization solution was added and the tube was immediately inverted 4 times to mix the solutions thoroughly. A 5 min centrifugation at 12000 g followed to pellet all cell debris and genomic DNA. The supernatant was transferred to a GeneJET spin column and subsequently centrifuged to bind plasmid DNA to the column while removing the remaining solution. Two washing steps with wash solution followed and the flow through was discarded after each centrifugation at 12000 g. In the final step the column was transferred to a new Eppendorf tube and 50 µl elution buffer was added to the column. After a final centrifugation at 12000 g, a plasmid solution was obtained and put into storage at -20°C.

All PCR products were purified using the Thermo Scientific GeneJET PCR Purification Kit (Thermo Scientific, Waltham, USA) and following the supplied protocol. First 20 µl kit provided Bunding Buffer was added to the PCR reaction volume of 20 µl, the mixture turns yellow if pH is suitable for DNA binding otherwise the pH must be adjusted. The mixture was transferred to a GeneJET purification column and centrifuged at 12000 g for 60 s and the flow through was discarded. 700 µl of Wash Buffer was added to the column and the centrifugation was repeated. A centrifugation of the empty column followed to remove any residual Wash Buffer and the column was transferred to a new eppendorf tube. 50 µl of Elution Buffer was added to the column and the centrifugation step was repeated yielding a 50 µl PCR product suspension which was stored in -20°C.

The products were separated on a 1% agarose gel at 80-100 V containing GelRed (Biotium, Fremont, USA) and compared to GeneRuler 1kb DNA ladder (Thermo Scientific, Waltham, USA) to confirm that each gene fragment had the expected size.

## 2.4 Gene cloning and bacterial procedures

### 2.4.1 Construction of transformation vector

The transformation vector was created using Gateway recombination cloning technology. The gene constructs were inserted into the pDONR221 by BP gateway reaction. The reaction involved the addition of 2 µl BP-clonase II enzyme mix (Invitrogen, Carlsbad, USA) to 150 ng attB flanked PCR product, 150 ng donor vector and TE-buffer to a volume of 8 µl. The final 10 µl mixture was incubated for 18 hours (25°C) before being terminated by the addition of 1 µl Proteinase K (Invitrogen, Carlsbad, USA) and subsequent incubation for 10 min (37°C).
The BP product was transformed into chemically competent *E.coli* TOP10 cells by heat shock transformation method. 1 µl BP product was added to 25 µl bacterial suspension, the mixture was incubated on ice for 30 min before 30s heat shock (42°C) and immediately back on ice for 2 min. 250 µl SOC medium was added to the transformed cells and the suspension was incubated for one hour (37°C) before being plated on LB plates containing kanamycin (50 µg/ml). Single colonies were isolated for further amplification in liquid LB-medium containing kanamycin (50 µg/ml). The plasmids were harvested using GeneJET Plasmid Miniprep kit (Thermo scientific). Samples of the purified plasmids were controlled by restriction enzyme digestion and gel electrophoresis: EcoRV + XhoI for *AtBBM* and EcoRV for *AtWRI4*. The HTFs contained only one identical restriction site, EcoRV, for the available nucleases therefore only a linearization and no fragmentation could be expected. Candidate plasmids were sent for external sanger sequencing by Eurofins genomics (Eurofins Scientific, Brussels) using the m13 universal (-21) primer (TGT AAA ACG ACG GCC AGT).

The gene constructs were transferred from the pDONR221 plasmid into *A. tumefaciens* compatible binary vector pXZP393 and positioned behind a constitutive promoter (CaMV-35S), using the gateway LR-reaction. The LR-reaction followed a similar protocol as the BP-reaction explained above, except that LR-clonase II enzyme mix (Invitrogen, Carlsbad, USA) was used. The LR-reaction product was transformed into *E. coli* following the previously mentioned procedure and transformed cells were grown on selective media containing spectinomycin (50 µg/ml). The amplified pXZP393 plasmids were harvested and confirmed through external sanger sequencing by Eurofins genomics (Eurofins Scientific, Brussels) using 35Sf (AATCCCACTATCCTTCGCAAG) as primer.

*A. tumefaciens* strain GV3101 mP90 were transformed with respective construct by electroporation. 50-500 ng plasmid was added to 50 µl competent cell suspension in an electroporation cuvette. Electroporation was performed at 2.5kV, 200Ω and 25 µFD. 250 µl LB medium was added to the cuvette and the content was transferred to an Eppendorf tube and incubated for 60 min at 25°C. After incubation the cells were plated on to agar plates containing LB medium rifampicin, spectinomycin and gentamicin (50 µg/ml each) and incubated at 25°C for two days before colonies could be identified and selected.

## 2.5 Agrobacterium infiltration of *Nicotiana benthamiana*

### 2.5.1 Preparation of infiltration cultures

Colonies of *A. tumefaciens* ,GV3101mP90, containing pXZP393 plasmids with all relevant constructs (*P19*, *GFP*, *BBM*, *WRI4*, *BBM HTF*, *WRI4 HTF* and *WRI1*) were inoculated separately into 5 ml of LB medium containing rifampicin, gentamicin and spectinomycin (50 µg/ml each) to serve as starter cultures. The starter cultures were incubated with open lids at 28°C and agitated at 225 RPM overnight. Starter cultures were scaled up the following afternoon by inoculating 20 ml LB medium containing rifampicin, gentamicin and spectinomycin (50 µg/ml each) with 20 µl of each starter culture in 50 ml falcon tubes. Two cultures containing 35S::GFP construct and three containing 35S::p19 were made since they were needed for all infiltrations. The lids were loosely attached with a single layer of parafilm to allow air exchange with the cultures, cultures were then incubated at 28°C and agitated at 225 RPM overnight.

### 2.5.2 Infiltration

On the day of the infiltration, in order to minimize the number of dead cells, $OD_{600}$ measurements were taken to confirm that the cultures were still in exponential growth phase. Samples of each culture were taken and diluted 10 times, an undiluted $OD_{600}$ of around 1 abs indicated that the cultures were still in exponential phase, this was based on previous work where the OD had been measured over time to establish a cell density versus time diagram from which the exponential phase could be confirmed. 20 µl acetosyringone (100mM) was added to each culture to a total concentration 0.1 mM, which were subsequently incubated for an additional two hours to induce all virulence genes.

The cells were isolated by centrifugation at 3000 g for 5 min in room temperature. The supernatant, LB medium and antibiotics, was removed and the pellets were carefully resuspended in 10 µl infiltration buffer containing 5 mM $MgCl_2$, 5 mM MES and 0.1 mM acetosyringone. The optical density (abs) was measured and later used for the infiltration mixtures. Acetosyringone needs to be added right before as it is quickly degraded.

Infiltration mixtures were prepared by mixing infiltration buffer with a volume of each culture, containing a pXZP393 plasmids with a TF-construct insert, and cultures containing pXZP393 plasmids with p*19* and *GFP* inserts to a final volume of 10 ml (final concentrations corresponding to 0.2 abs TF, 0.2 abs p19 and 0.05 abs GFP). A control mixture was created with final concentrations corresponding to 0.2 abs p19 and 0.25 abs GFP. The control was used to compare basal levels of stress induced by *A. tumefaciens* and expression levels with respect to other leaves.

Four week old *N. benthamiana* plants which were right before flowering stage were thoroughly watered in the morning to facilitate infiltration. One or two developing leaves with a bulging surface and a size of approximately 6x6 cm were chosen for infiltration. The leaves were

infiltrated using a 1 ml syringe. Triplicates were performed for each infiltration mixture, approximately 3 ml mixture were injected through the abaxial side of each leaf. The infiltrated plants were watered and returned into the growth chamber under the same growth conditions as mentioned in section 2.2.

### 2.5.3 Sampling

Approximately 96 hours after infiltration the plants were removed from the growth chamber for sampling. Two leaves at a time were brought to a dark room where the infiltrated zone was marked under UV light which made infiltrated areas visible due to the expression of GFP in successfully infiltrated areas. The leaves where subsequently pulled from the stem and the marked region was cut out. All thick veins were removed as they would represent a big part of the dry weight while they might not have been fully infiltrated. Approximately 2-3 $cm^2$ was taken for RNA extraction and remaining leaf matter was saved for lipid analysis. Two samples of leaves infiltrated with a control mixture were accidentally pooled during sampling. Thus, there is no technical triplicate for control samples during the lipid extraction and analysis. All leaf samples were snap frozen in liquid nitrogen immediately after being cut out. Samples for lipid extraction were freeze dried two weeks after sampling.

## 2.6 Lipid extraction and analysis

The lipid extraction protocol is based on a method described by Bligh and Dyer (1959). Around 100 mg of freeze dried leaf matter, 5 mg for the seed analysis, were weighed and subsequently homogenized in a mixture of 3.75 ml methanol:$CHCl_3$ (2:1) and 1 ml acetic acid (0.15 M), in a glass homogenizer. After addition of 1.25 ml pure $CHCl_3$, and 1.25 ml MQ-$H_2O$, the mixture was transferred in to a glass tube, vortexed and centrifuged at 3000 RPM for 2 min. The bottom phase containing $CHCl_3$ and the nonpolar fraction was recovered using a Pasteur pipette and stored overnight at -20°C.

Volumes corresponding to 10 mg dried leaf matter (1 mg for seed analysis) were loaded on to a thin layer chromatography (TLC) silica plate (Merck, Darmstadt) along with a reference mixture of 18:1 mono-, di- and triacylglycerols as well as 18:1 free fatty acids and 18:1 fatty acid methyl esters. The samples were subsequently separated using TLC with a heptane: diethyl ether: acetic acid (70:30:1 ratio) mobile phase.

After 60 min incubation the plate was removed from the TLC vessel and sprayed with a primulin and acetone mixture. The primulin made the lipid fractions visible under UV light (White et al., 1998) and the different fractions were divided into three sections; triacylglycerols, polar lipids and rest, based on the reference mixture. Each section of the TLC silica plate was scraped off after softening the silica with water. Only the TAG fraction was isolated from the lipid analysis of A. thaliana knockout seeds.

All samples were dried on a heated sand bed under a protective N2 atmosphere to minimize oxidation of the lipids. The dry silica fractions were then incubated at 90°C for 45 min with 2%

H2SO4 in methanol (2 ml) to obtain methylated free fatty acids. 200 µl 17:0 internal standard (0,84 nmol/ml), 1 ml heptane and 2 ml MQ-H2O were added to the reaction mixture, the heptane phase was isolated using a Pasteur pipette. 200 ul of the nonpolar phase was then transferred to gas chromatography (GC) vials for analysis in a GC-17A model (Shimadzu Corp. Kyoto) with a CP-Wax 58 FFAP CB column (50 m, 0.32 mm inner diameter, 0.20 µm film)(Agilent, Santa Klara, USA). The parameters for the GC analysis are located in appendix 2.

Fatty acid methyl esters were identified by retention times of fatty acid methyl standards and quantified using internal standard (heptadecanoic acid methyl ester, 17:0). Total fatty acid content was derived from the sum of all major peaks, the threshold for peak inclusion was determined by its area (150 for TAG, 500 for Polar lipids, 100 for Rest in leaves and 200 for TAG in seeds; measured in area units)

## 2.7 Statistical analysis of GC data

A one-way analysis of variance was performed on the total fatty acid content of each fraction and Tukey's pairwise comparison was used to compare the means of the different infiltrations using the Minitab (Minitab Inc., State college, USA) built in function. Statistical significance was determined at confidence level $\alpha$=0,05. The data was tested for normality by determining its linear appearance on a normal probability plot.

# 3 Results

## 3.1 Bioinformatic analysis

The intrinsically disordered regions for native TFs were compared using the Pondr VL3-BA predictor and are shown in the order AtBBM, AtWRI4, AtWRI1 from top to bottom in figure 7. AtWRI1 and AtWRI4 show very similar predictions except for IDR3 on the C-terminal of AtWRI1 which is absent from AtWRI4. The prediction for AtBBM is less clear although the highly flexible linker between the two AP2 regions show some similarity.

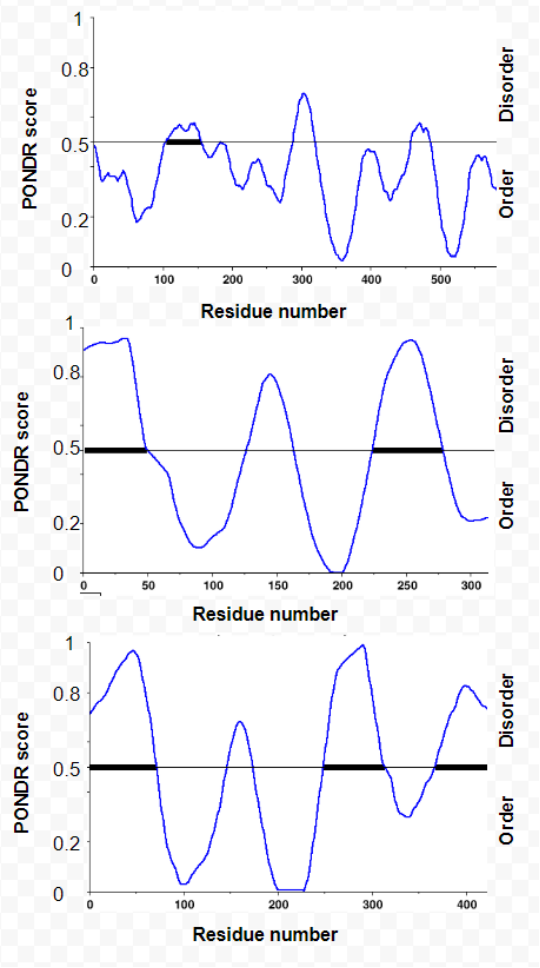


*Figure 7: Intrinsically disordered region (IDR) prediction for AtBBM (top), AtWRI4(middle) and AtWRI1(bottom). Values higher on the y axis signify higher probability of an IDR. predicted using the online tool PONDR (Molecular Kinetics, Inc., Indianapolis, USA) and the VL3-BA prediction algorithm developed by (Obradovic et al., 2003).*

An amino acid sequence alignment between the TFs, sequences are available in appendix 3, reveals a roughly 200 amino acid long N-terminal domain on AtBBM that show little resemblance to the corresponding regions on AtWRI1 and AtWRI4. No remarkable patterns can be seen except perhaps a somewhat elevated amount of asparagine and threonine in some areas. Contrarily, the N terminal of AtWRI1 show a high frequency of serine and threonine residues, adding up to 28 of the 65 residues, most of which are located in the first half of the domain. The frequency of positively charged residues on AtWRI1 N-terminal domain is also noteworthy, making up 13 of the 65 residues. The N-terminal of AtWRI4 exhibit half as many serine and threonine residues as AtWRI1 while containing slightly more charged residues, it is also 13 residues shorter.

The C terminal of AtBBM again show no immediate patterns except perhaps residues AtBBM 536-546 that contain un almost unbroken series of 8 Glutamates. A dense region of 12 glutamate residues is located at position 279-299 of AtWRI1, a homologous region can be found on AtWRI4 although it contains less glutamate and more proline and serine. The C-terminal domain of AtWRI1, 397-431, exhibit a high frequency of proline, serine and threonine.

The AP2 regions, shown in figure 8 below, are far more conserved between the different TFs than the N and C terminal domains. The first YRG motif and the following 19 residues are almost identical. However, residues 20-29, counted from the YRG motif, show some deviation.

Notably, the AtBBM sequence contains a cysteine residue followed by two positively charged residues where AtWRI1 and AtWRI4 contain an identical sequence of uncharged residues serine, tryptophan and asparagine. Furthermore, the cysteine residue on BBM is the only cysteine residue on all three AP2 domains. The RAYD motif seen at positions 47-70 is notably different for atWRI1 which starts with a histidine and threonine instead of arginine and alanine. Surrounding the RAYD motif is a rather conserved region except perhaps for two positively charged residues at position 41 and 44 of BBM. A series of 5 threonines stands out on position 59-63 of AtBBM were only one is present in both AtWRI1 and AtWRI4. 66-83 contains a less conserved region with a high density of polar residues. 84-marks the start of a highly conserved region past the YRG motif at position 103 and onwards to position 160 with no significantly differences between the three factors. The end of the selected double AP2 domain.

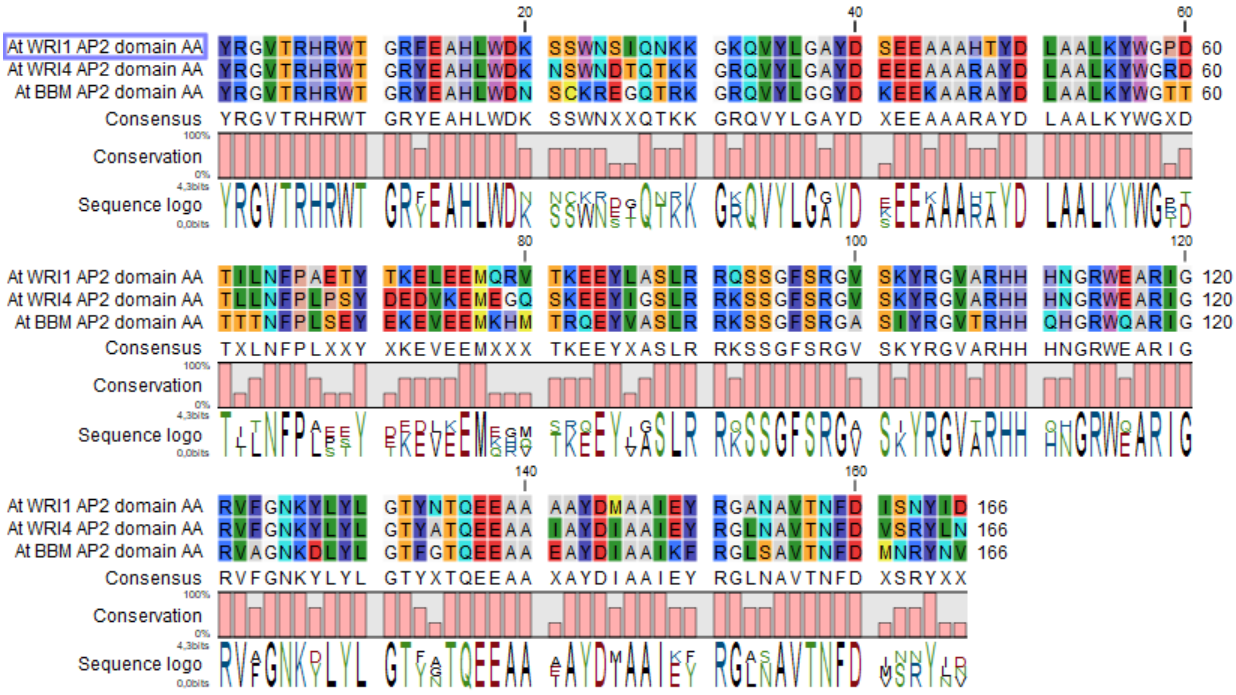


*Figure 8: Primary sequence alignment of the AP2 domains of AtWRI1(top), AtWRI4 (middle) and AtBBM (bottom).*

Primary structure comparison of the AP2 domains to WRI1 homologues as seen in figure 9 further underlines the presence of the two positively charged amino acids at position 23-24, as counted from the YRG motif, on AtBBM. Furthermore, it can be noted that the cysteine residue seen in AtBBM is shared by two of the homologs. Additionally, the WRI1 homologues show high consensus for positions 25-26 consisting of a serine and a short hydrophobic residue while AtBBM and AtWRI4 exhibit a negatively charged residue on position 25. As previously pointed out the threonine sequence at positions 59-63 is unique to AtBBM. While all WRI1 homologues and WRI4 show highly conserved regions surrounding the second YTG motif, BBM show a few singular exceptions to this pattern.

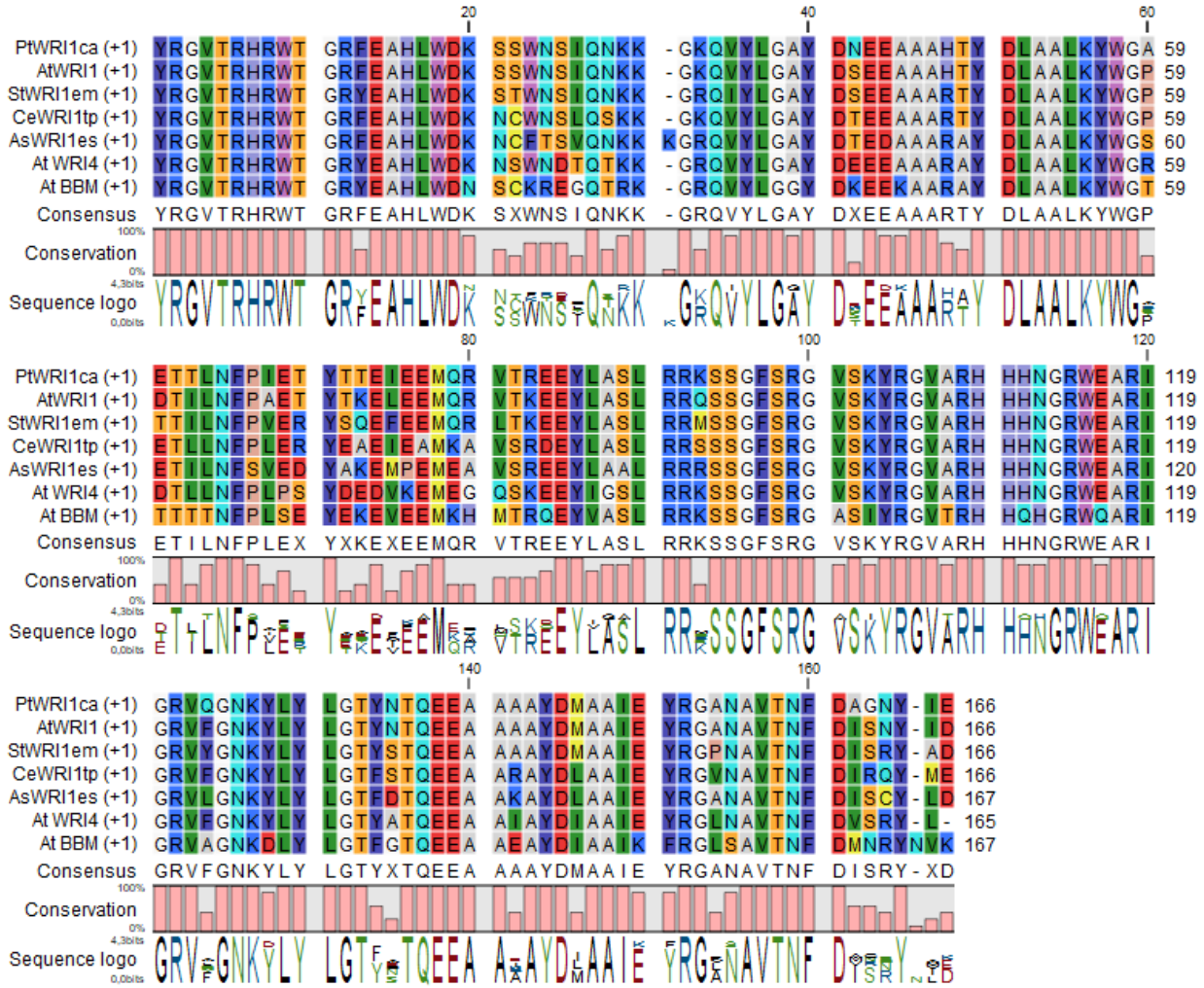


*Figure 9: Primary sequence alignment of the AP2 domains of AtWRI1 (row 2), AtWRI4 (row 6) and AtBBM(row 7) as well as the WRI1 homologues(rows 1,3,4,5)*

Comparison with the PLT factors AP2 domains, as seen in figure 10 below, show high conservation of the region between the two groups. The primary structure alignment further emphasizes the previous observation that the two positively charged residues at position 23-24,

as counted from the YRG motif. The surrounding region seem to be highly conserved within both groups and point of divergence between PLT and WRINKLED factors. AtWRI4 is a bit of an exception as it closely resembles the WRI1 factors at positions 20-24 while position 25-28 diverges from both the other groups. Furthermore, the 5 threonine sequence seems to be conserved within the PLT group although some substitutions with serine, proline and alanine present.

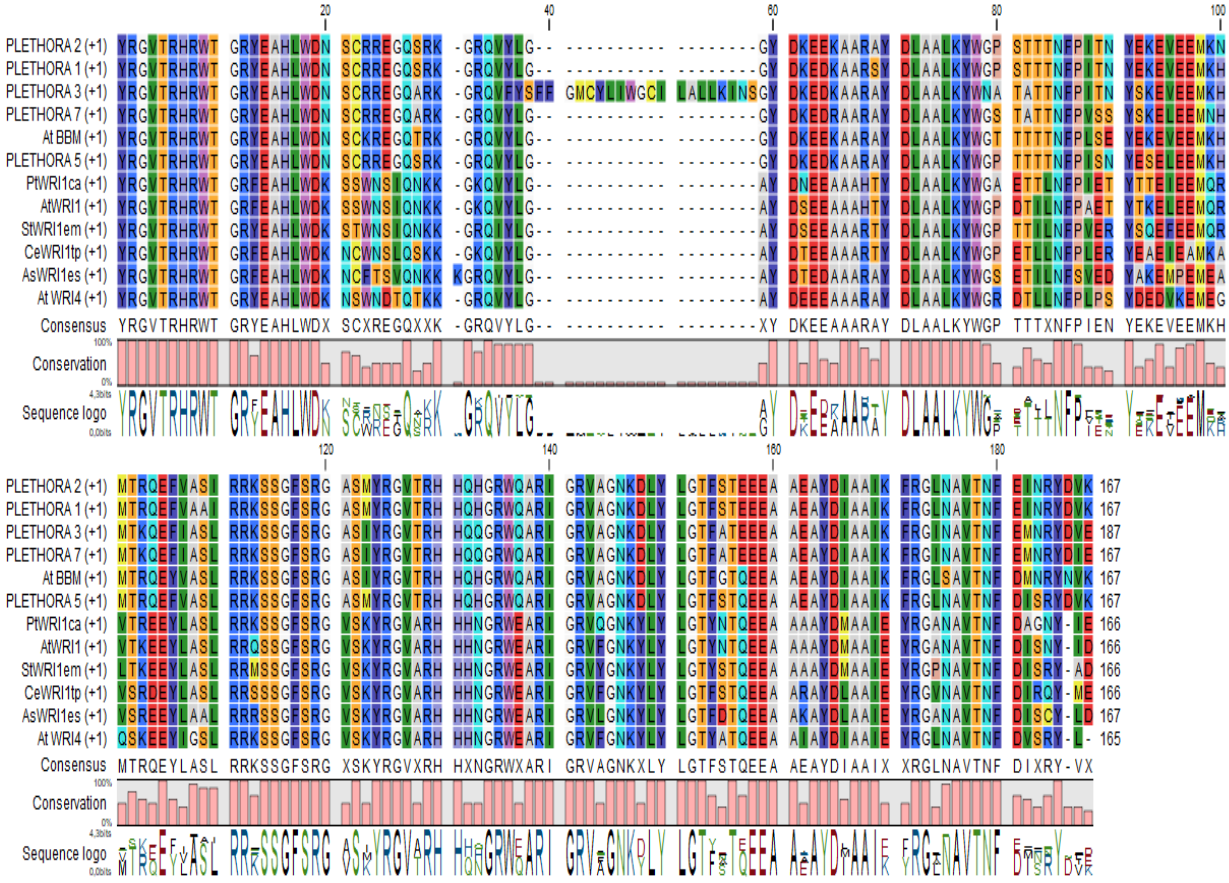


*Figure 10: Primary sequence alignment of the AP2 domains of AtWRI1 (row 8), AtWRI4 (row 12) and AtBBM(row 5) as well as the WRI1 homologues(rows 7,9,10,11) and PLT homologues(rows 1,2,3,4,6)*

The secondary structure predictions, figure 11, reveal a number of predicted beta strands on the N terminal domain of AtBBM that are not present on AtWRI1, AtWRI4. There are no immediate differences within the AP2 domains, the main figure being alpha helices that are predicted on the first and second RAYD element for all three TFs.

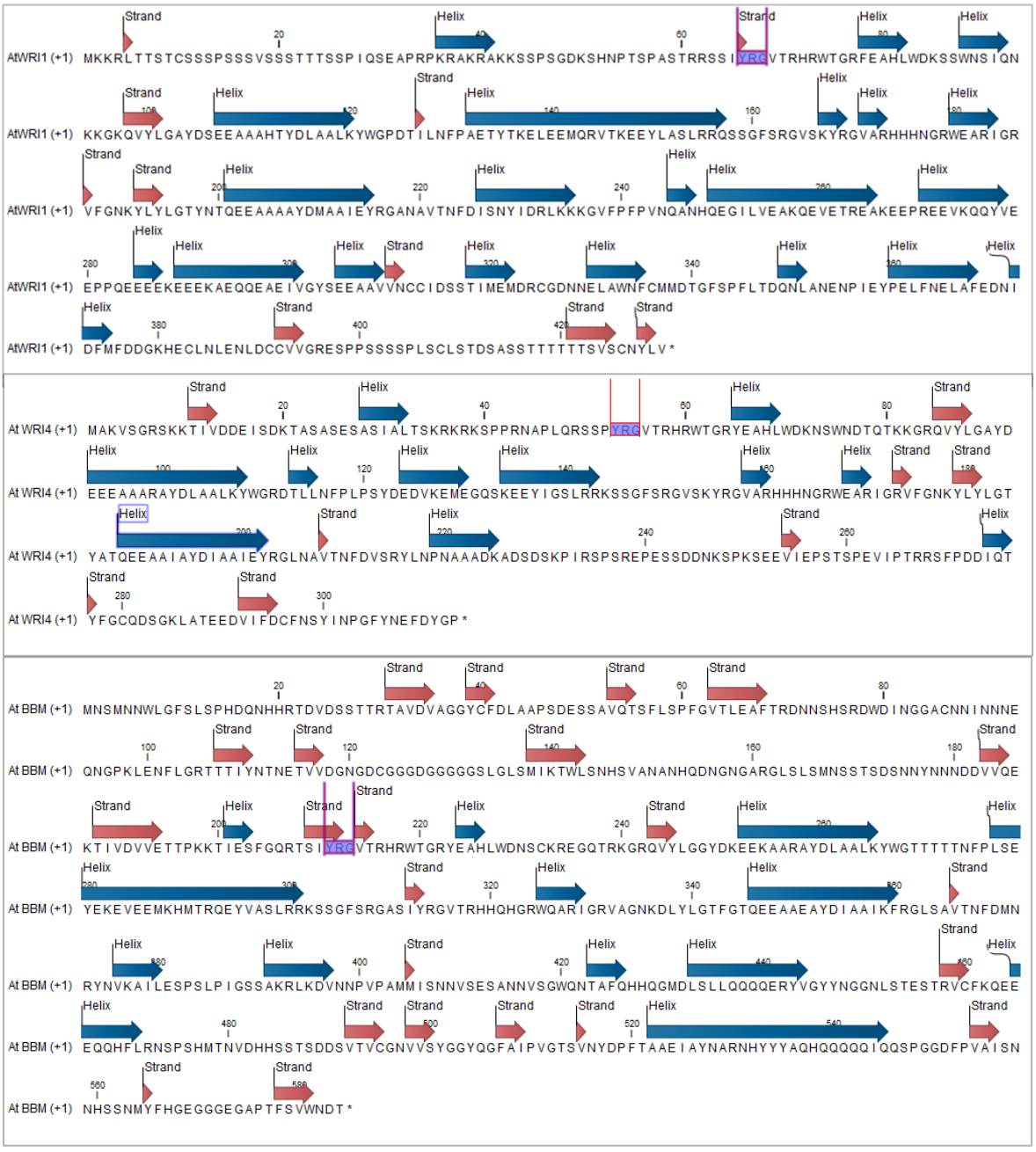


*Figure 11: Secondary structure predictions of WRI1(top), WRI4(middle) and BBM (bottom). The start of the first AP2 domain, the YRG motif, is highlighted in purple. Blue arrows indicate predicted α-helices while red arrows indicate predicted β-strands.*

## 3.2 Molecular cloning

Hybrid transcription factors were manufactured by multiple PCRs, figure 13 illustrate the domains of the HTFs.

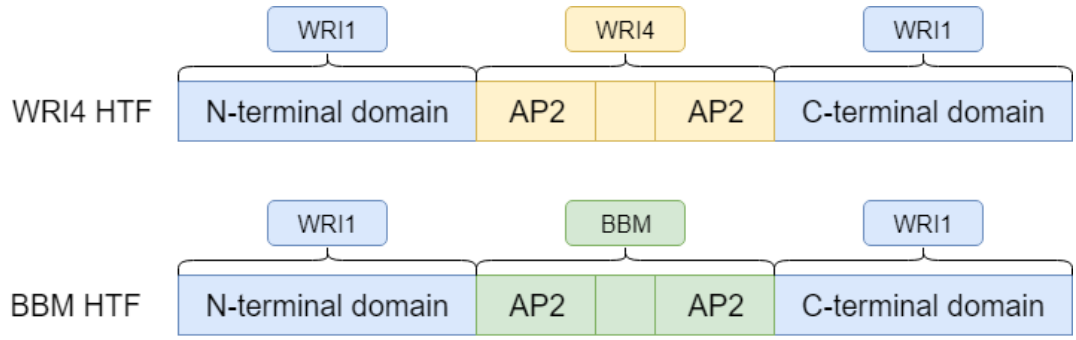


*Figure 13: Schematic figure of the HTFs.*

The fusion of PCR products was performed, figure 12 show the unpurified HTF products of the fusion PCR *BBM HTF* on the left and *WRI4 HTF* on the right. Both seem to have a similar size less than 1500 bp when compared to ladder. The expected size for the fusion products being 1287 bp.

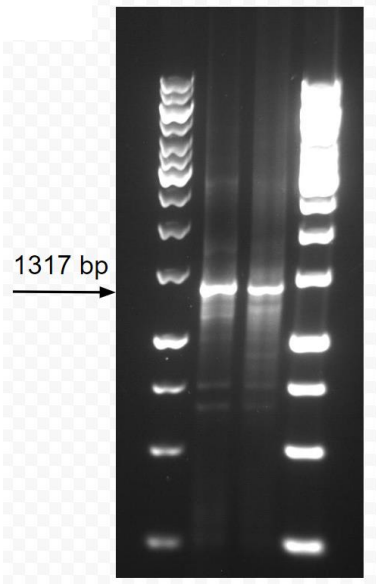


*Figure 12: Agarose (1%) gel electrophoresis of HTF products from fusion PCR. Ladder (well 1 and 4) BBM HTF (well 2) and WRI4 HTF in (well 3). The length of the HTF bands are indicated by an arrow on the left*

Purified plasmids from separate cultures of *E.coli* transformed with BP product, pDONR221 with *BBM* (wells 2-9) and *WRI4* (wells 10-17) inserts, are shown in an undigested/digested pattern in figure 13 below. Sample 1 and 2 of digested plasmid with *BBM* insert (well 3 and 5) show single bands of shorter length than other digested products, while sample 3 and 4 with *BBM* insert (well 5 and 9) exhibit a similar large size fragment as the samples containing plasmids with *WRI4* inserts. Additionally, sample 3 and 4 has a smaller fragment close to the 1kb band in ladder while the smaller fragments for all plasmids with *WRI4* inserts are somewhere in the range of the 250 bp band on ladder. PCR product of HTF constructs in well 18 and 19 exhibit a strong band at around 1500 bp as well as some evidence of impurities in the samples.
Plasmids from wells 6-7 and 10-11 were chosen for further processing and sequence analysis as they exhibited the expected fragment size (731 fragment, total 4299 for BBM and 242 fragment, total 3486 for WRI4).

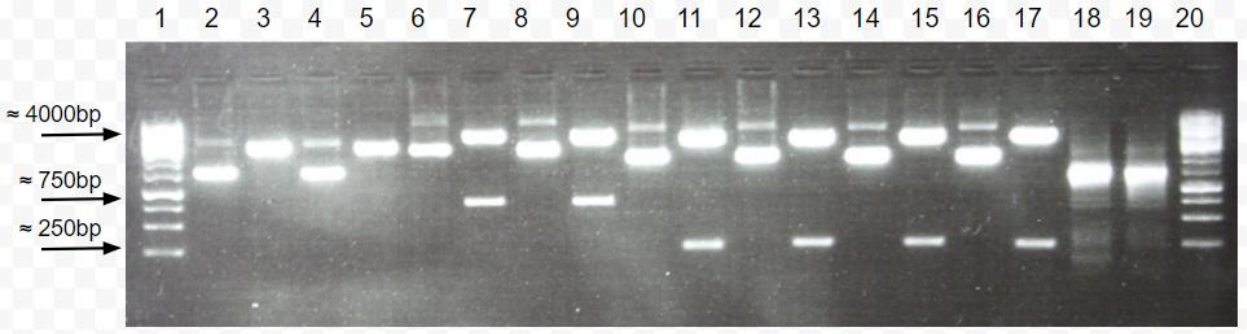


*Figure 14: Agarose (1%) gel electrophoresis of pDONR221 containing BBM and WRI4 inserts in an undigested/digested pattern. Undigested/digested (EcoRV and XhoI) plasmid containing BBM (2-9), undigested/digested (EcoRV) plasmid containing WRI4 (10-17) ladder (1 and 20) BBM HTF (18) WRI4 HTF (19). Sizes of fragments from digested plasmids are indicated with arrows on the right-hand side.*

The sequencing results of the of the chosen plasmids show evidence that the genes were present (appendix 4). However, the analysis of plasmid containing the *BBM* gene showed poor quality, 215 bp clipped length compared to 956 bp for *WRI4*. Since the gene was present the decision was made to go forward with the LR reaction and get confirmation in the sequencing of the LR product.

Purified pDONR221 containing HTF constructs from separate cultures of *E.coli* transformed with BP product are shown in an undigested/digested pattern in figure 15 below, *BBM HTF* inserts (wells 2-9) and *WRI4 HTF* inserts (wells 10-17). Digested sample 3 and 4 of *BBM HTF* inserts (well 7 and 9) show a single band of similar size as digested sample 2, 3 and 4 of the *WRI4 HTF* insert (well 13, 15 and 17). Sample 1 and 2 of *BBM HTF* as well as sample 1 of *WRI4 HTF* show somewhat smaller fragment size. All digested products show bands higher up on the gel than the large band on their respective undigested partner although a faint band can be seen for all undigested products of equal length to their digested counterpart, this is probably due to supercoiling which makes the unnicked circular plasmids move faster through the gel. The expected total length for both constructs is 3861 bp.

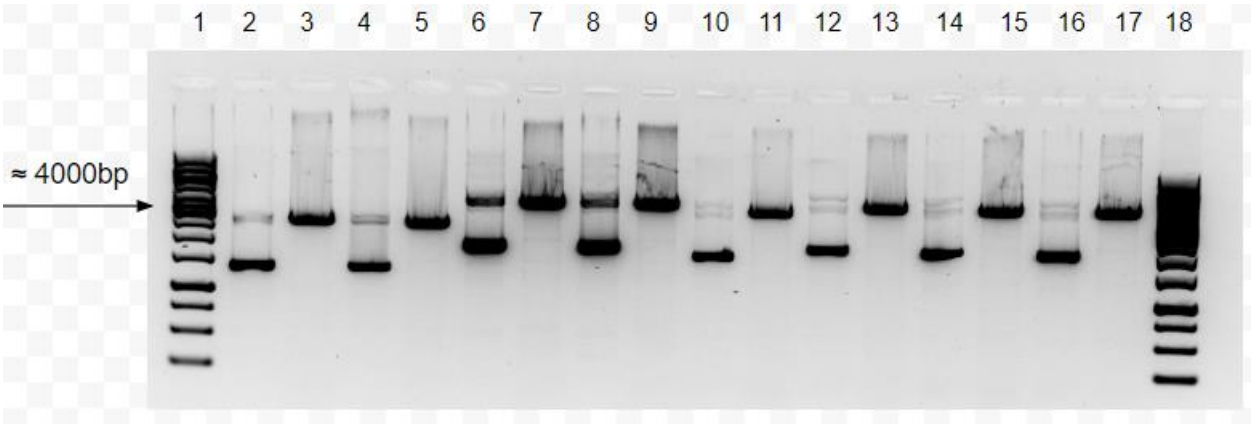


*Figure 15: Agarose (1%) gel electrophoresis of pDONR221 containing BBM HTF and WRI4 HTF inserts in an undigested/digested pattern. Undigested/digested (EcoRV) plasmid containing BBM HTF(2-9), undigested/digested (EcoRV) plasmid containing WRI4 HTF(10-17) ladder (1 and 18) Size of digested plasmid fragment is indicated with an arrow to the right-hand side.*

Sequencing of pXZP393 plasmids containing inserts of *BBM*, *WRI4*, *BBM HTF* and *WRI4 HTF* all showed successful insertion of the correct gene when aligning the results with gene sequences. The quality of the sequencing results were good with a clipped length of 912-1006 bp. The sequencing results are presented in appendix 4.

## 3.3 Lipid extraction

### 3.3.1 Leaf

The order on the TLC plates from the lipid separation seen in figure 16 are as follows: Control (1x), BBM (2x), WRI4 (3x), BBM HTF (4x), WRI4 HTF (5x), WRI1 (6x). The reference mixture can be seen to the right side of plate two and three but was left out of plate one. A small area approximately 4 mm to each side of the loading line showed limited movement in the nonpolar mobile phase and was therefore identified as the polar lipid fraction. The dark areas on the bottom were of green pigment while the dark areas on the top were yellow pigment.

The TAG stain For WRI1 is very distinct for all three samples (61, 62, 63) compared to the other infiltrations. The TAG fraction for WRI4 HTF (5) show a similar distinctness on all plates and that the same is true for WRI4 (3) on the first and second TLC plate. The separations for control (1), BBM (2) and BBM HTF (4) have weak TAG stains for all plates.

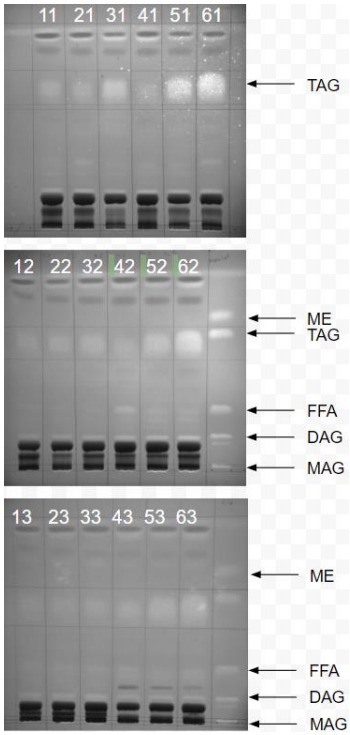


*Figure 16: TLC sepparation of lipids, from leaf infiltrations, using a heptane:diethyl ether:acetate (70:30:1) mobile phase.Samples in the order Control (11,12,13), AtBBM (21,22,23), AtWRI4(31,32,33), BBM HTF(41,42,43), WRI4 HTF (51,52,53) and WRI1(61,62,63). The different fractions are indicated on the right-hand side*

### 3.3.2 Seed

The separation of lipids from knockout seeds can be seen in figure 17, samples 1-3 designates *bbm* mutant, 4-6 *wri4* mutant, 7-9 *wri1* mutant and 10-12 is control. Thin layer chromatography of seed extracts shows much more distinct TAG stains than those from leaves except for wri1 extractions which have weaker stains.

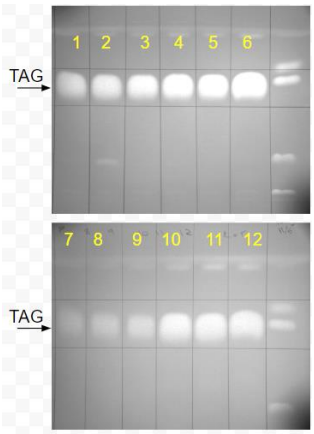


*Figure 17: TLC sepparation of lipids, from mutant seeds, using a heptane:diethyl ether:acetate (70:30:1) mobile phase. Samples in the order: bbm(1,2,3), wri4(4,5,6), wri1(7,8,9) and control (10,11,12)*

## 3.4 Lipid analysis

### 3.4.1 Lipid analysis of leaf matter

Leaves infiltrated with BBM and BBM HTF show no increased TAG accumulation while WRI1 show a significant increase (44,1 nmol /mg), the same is true for WRI4 HTF (25.1 nmol /mg). No significant differences can be found between WRI4 (12.5 nmol/mg), BBM HTF (4.7nmol/mg), BBM (5.7 nmol/mg) and Control (5.0 nmol/mg) at significance level $\alpha = 0.05$. However, the results hint at an increase in TAG production in the WRI4 infiltrated leaves. The total TAG amounts from GC analysis seem to fit with the observations made on the TLC separation results. A comparison of the TAG levels for each construct can be seen in figure 18. A, B and C denotes groups of significant differences. Raw data from the lipid analyses is available in appendix 5 and 6.

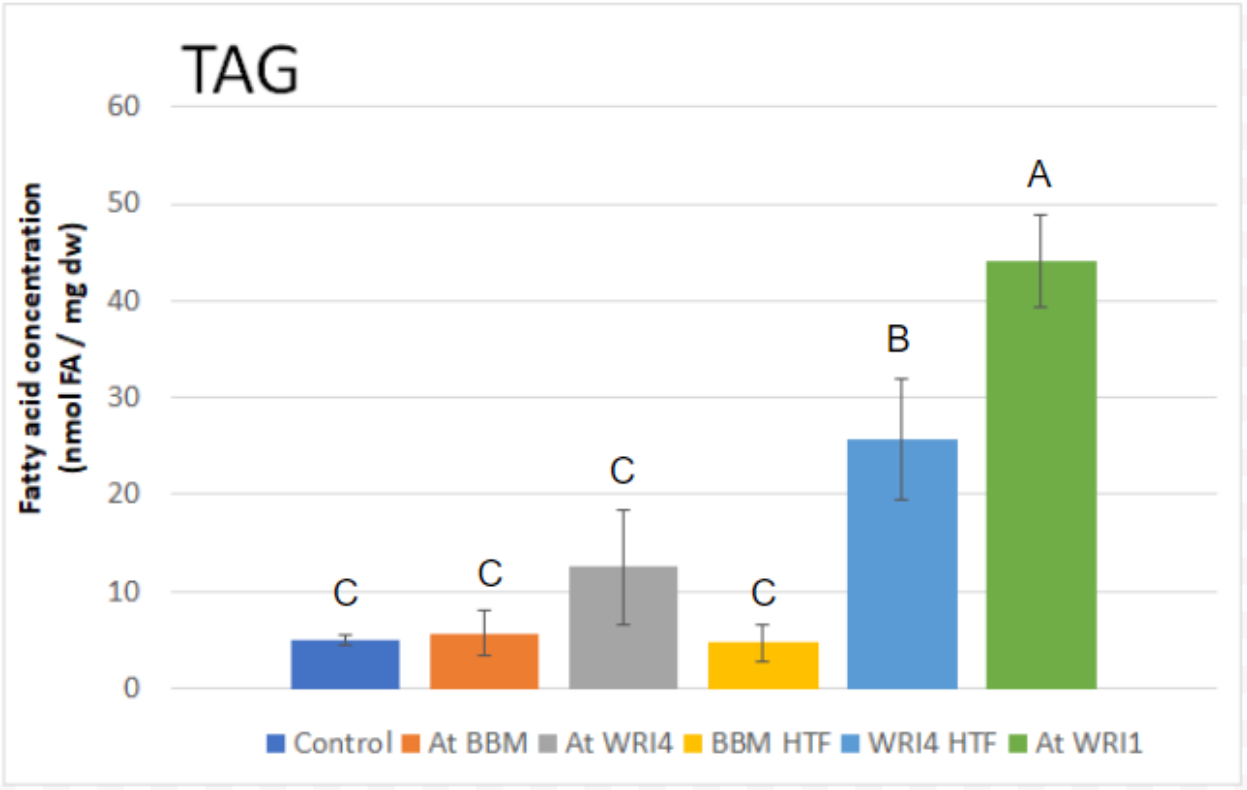


*Figure 18: Total triacylglycerol (TAG) content of N. benthamiana leaves infiltrated with genes coding for (control) AtBBM, AtWRI4, BBM HTF, WRI4 HTF and WRI1. Results are shown as mean of three biological replicates ± standard deviation. The letters A, B and C designate groups with significantly different means as calculated using a one way anova and Tukey's test at significance level α=0,05*

The change in fatty acid profile between the samples are shown in figure 19. Perhaps the most notable change is the transition from 16:0 to 18:3 and 24:0 that can be seen for the BBM HTF infiltrations. Contrarily, the relative amount of 24:0 is decreased while 18:1 is increased for WRI1, WR4-HTF and to some extent WRI4. WRI1 infiltrations also exhibit a decrease in the

relative amount of 18:3 compared to the other constructs. X1-X3 denotes a number of unidentified recurring peaks with retention times between 16:0 and 17:0 for X1 and X2 while X3 is positioned between 17:0 and 18:0. Z is a summation of unidentified peaks of significant amplitude.

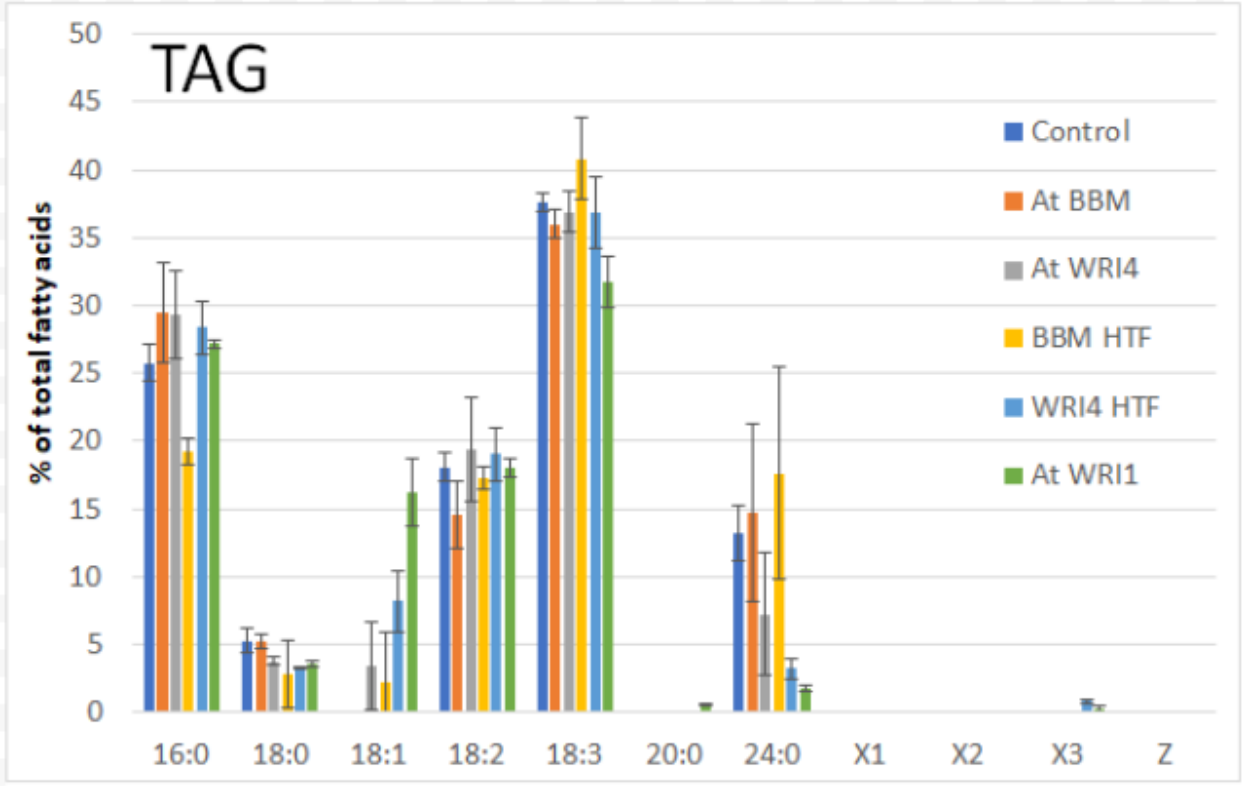


*Figure 19: Lipid profile. Triacylglycerol lipid profile of N. benthamiana leaves infiltrated with genes coding for (control) AtBBM, AtWRI4, BBM HTF, WRI4 HTF and WRI1. Results are shown as mean of three biological replicates ± standard deviation.*

Total levels of Polar Lipids and Rest fractions were relatively equal between samples (figure 20). The deviation seen for total polar lipid content of *WRI4* infiltrations is thought to have resulted from an error in a lipid extraction step where most of one sample was lost.

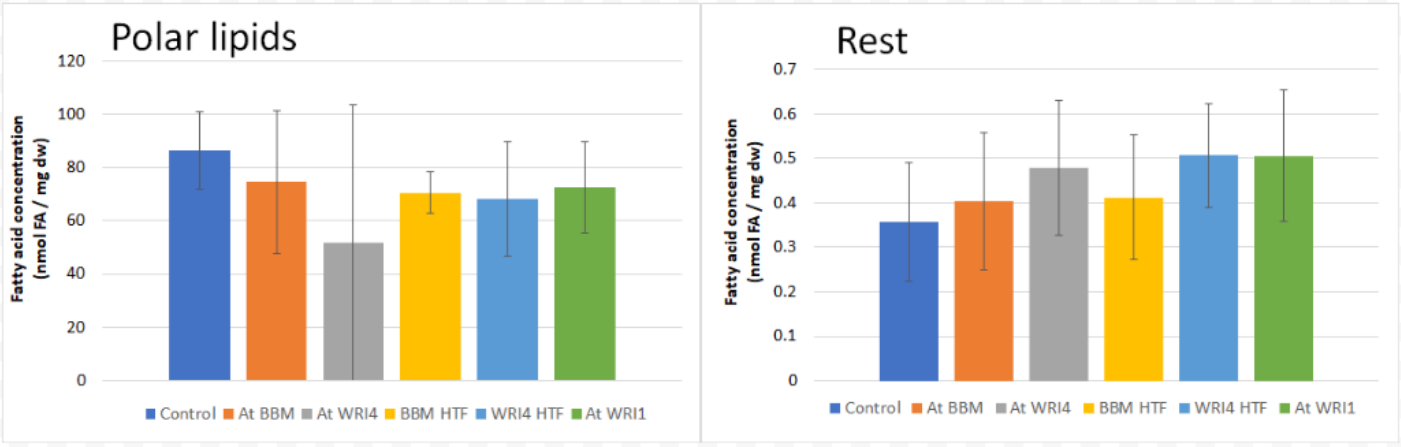


*Figure 20: Total polar lipid (left)and rest (right) content of N. benthamiana leaves infiltrated with genes coding for (control) AtBBM, AtWRI4, BBM HTF, WRI4 HTF and WRI1. Results are shown as mean of three biological replicates ± standard deviation. No significantly different means were found when calculated using a one-way anova and Tukey's test at significance level α=0,05*

There were no dramatic differences in fatty acid profile for the polar lipid fraction between samples, shown in figure 21 (top). Most notably there seem to be a general decrease in 18:3, especially for the *WRI1* infiltrated samples which also show an indication of higher levels of 18:1. In contrast, large differences can be seen in the fatty acid profiles of the Rest fraction shown in figure 21 (bottom). The variation is also seen between the samples with the same construct which is evident from the larger error bars. However, an increase in 18:1 can be seen yet again in the WRI1 infiltrations. Contrary to the TAG results, the lipid profile for *BBM HTF* infiltrations shows lower levels of 24:0 compared to the other samples and a possible increase in 18:0 and 18:3 for *BBM HTF* infiltrations can be interpreted from the Rest fraction results.

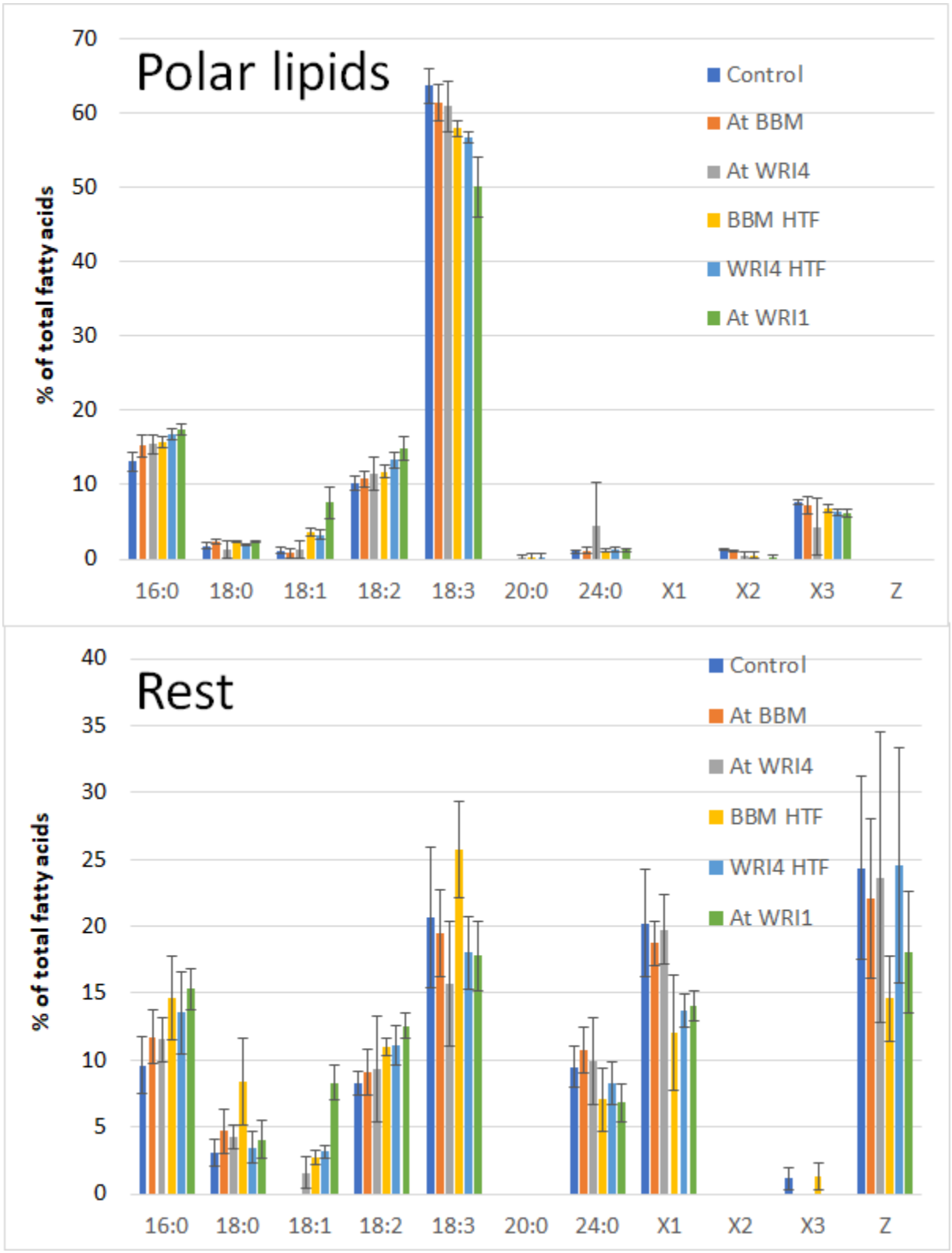


*Figure 21: Lipid profile. Polar lipid (top) and Rest (bottom) fraction lipid profile of N. benthamiana leaves infiltrated with genes coding for (control) AtBBM, AtWRI4, BBM HTF, WRI4 HTF and WRI1. Results are shown as mean of three biological replicates ± standard deviance.*

### 3.4.2 Lipid analysis of mutant Arabidopsis seeds

Seeds of *wri1* and *bbm* show a significant decrease, 80% reduction for *wri1* and 30% for *bbm*, in TAG accumulation when compared to *A. thaliana* col-0 control. The total seed TAG content is shown in figure 22.

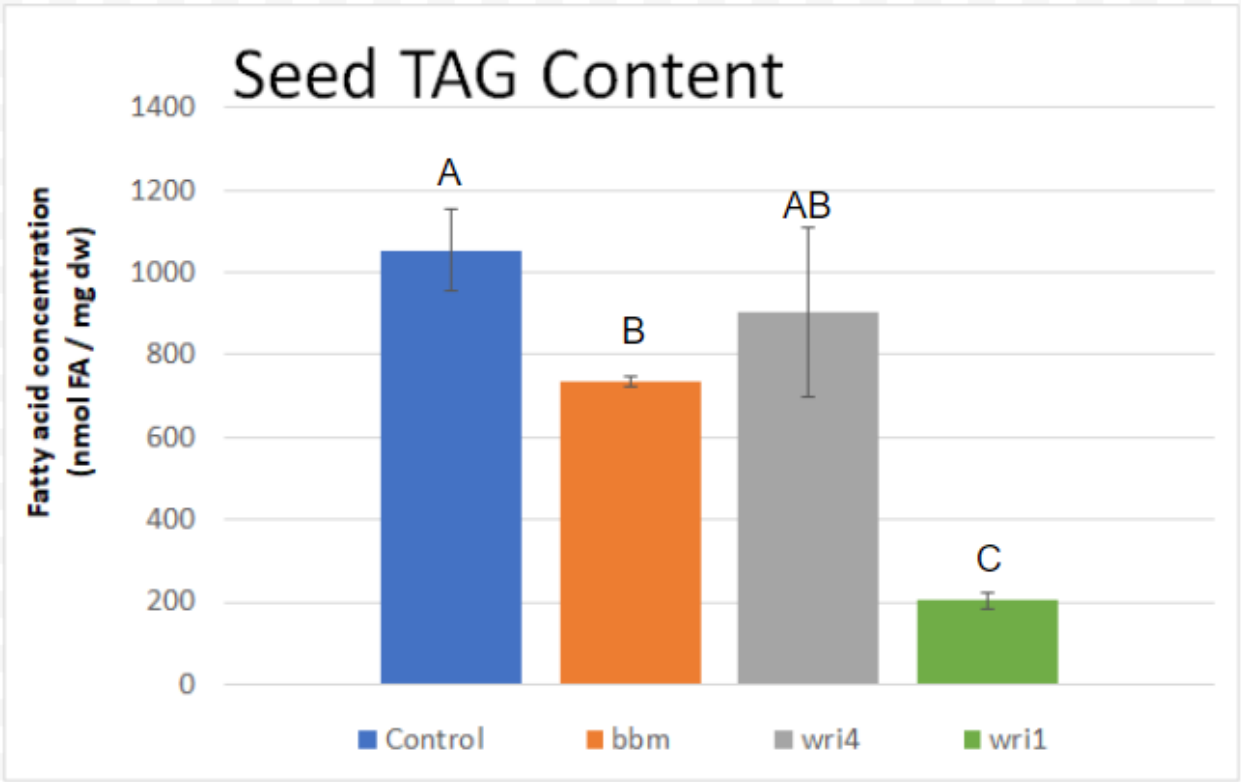


*Figure 22: Total triacylglycerol (TAG) content of seeds from (A thaliana col-0 as control) A thaliana mutants with non-functional genes for bbm, wri4 and wri1. Results are shown as mean of three biological replicates ± standard deviation. The letters A, B and C designate groups with significantly different means as calculated using a one way anova and Tukey's test at significance level α=0,05*

The lipid profile of the TAG fraction for knockout seeds is shown in figure 23. The relative amounts of 18:1 and 20:1 are dramatically reduced for *wri1* while 18:3, 20:2, 22:1 and 24:0 are all increased. The increase in relative amount of 22:1 is notably 4 times higher for *wri1* than for the other genotypes. Over all lipid species, *bbm* and *wri4* show a remarkable similarity in their lipid profiles and differ only slightly in 18:1 and 18:2 when compared to control. Interestingly 20:1, 20:2, and 22:1 are all present in seeds of all genotypes while they were absent in the analysis of all lipid fractions of the transiently expressed TFs in *N. benthamiana* leaves.

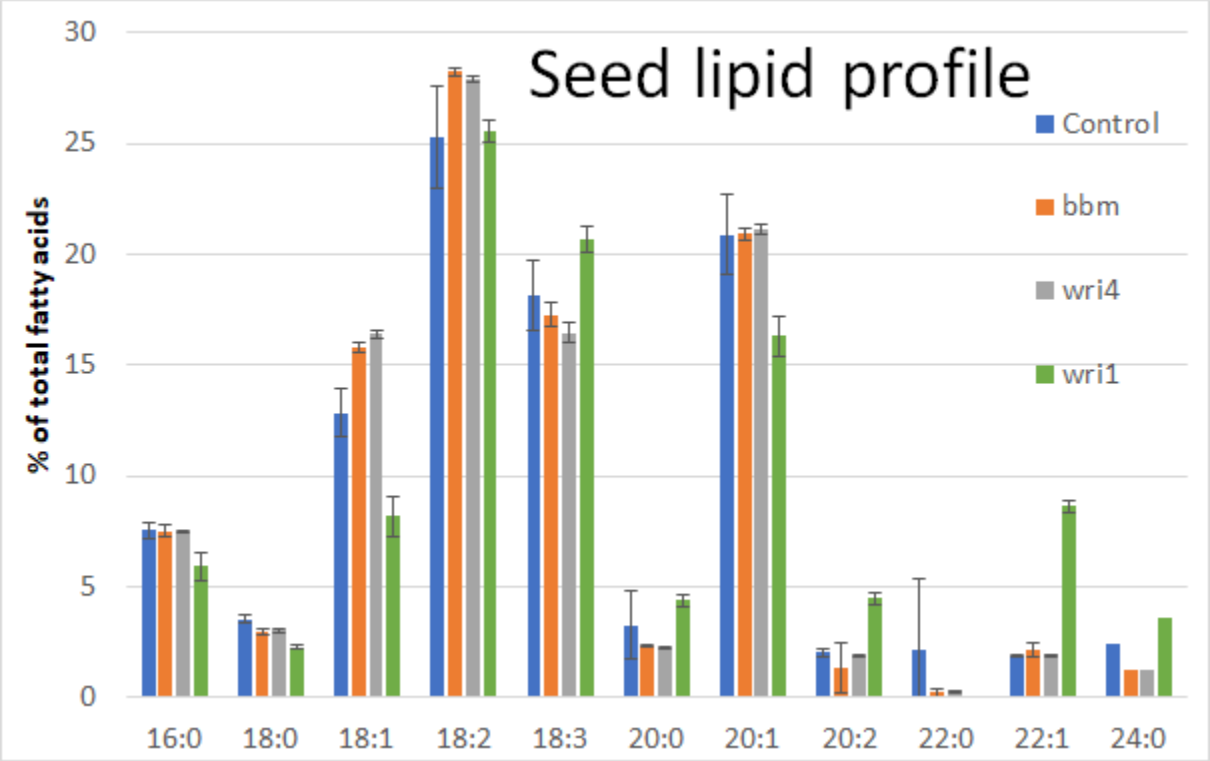


*Figure 23:Lipid profile. Triacylglycerol (TAG) lipid profile of seeds from (A thaliana col-0 as control) A thaliana mutants with non-functional genes for bbm, wri4 and wri1. Results are shown as mean of three biological replicates ± standard deviation.*

# 4 Discussion

The purpose of this work is to examine the relationship, in terms of effects on oil accumulation, between three related transcription factors of the AP2 subfamily, WRI1, WRI4 and BBM. To this purpose, the effects of AtWRI1, AtWRI4 and AtBBM on acyl lipid accumulation by transient expression in *N. benthamiana* leaves has been documented. As well as the effects on TAG accumulation in mutant *A. thaliana* seeds with a non-functional version of each gene. Additionally, this work strives to identify possible structural difference that explain how the factors can target different gene panels when the common DNA binding domain show a high degree of similarity. For this reason, hybrid transcription factors have been produced and the lipid analysis of the two, TFs BBM HTF and WRI4 HTF, provides insight into the function the AP2 domain.

A preliminary bioinformatic comparison of the three transcription factors reveals a close similarity between WRI1 and WRI4 while BBM differs substantially in its N- and C-terminal domains but not in the AP2 domains. Perhaps indicating a more distant evolutionary relationship of BBM in relation to the other factors.

## 4.1 Lipid content from infiltrated leaves and mutant seeds

The high TAG content of leaves infiltrated with AtWRI1 confirms its oil accumulation inducing effect as has been demonstrated in previous studies (Cernacand and Benning, 2004; Grimberg et al., 2015; Hofvander et.al, 2016). The fatty acid profile in leaves transfected with *AtWRI1* show a similar increase in the relative level of 18:1 and decrease in 18:3 which had been noted before (Grimberg, et al. 2015; Vanhercke et al., 2013). This shift away from highly desaturated fatty acids could be an effect of an increased flow of fatty acid in to the ER resulting in less interactions between each fatty acid and the desaturation machinery before being converted to TAG. A similar shift in fatty acid profiles has been observed both in *Avena sativa L.* where strains producing more oil tended to contain higher relative levels of 18:1 (Holland, Frey and Hammond, 2001) and in leaf lipid profiles during day/night cycles of high/low fatty acid synthesis (Browse, Roughan and Slack, 1981). However, the lipid profile differed from what Grimberg, et.al (2015) reported as there was only a small reduction in 18:0 in the TAG fraction. Seeds from *wri1* mutant show a similar lack in TAG accumulation, 80% reduction, as was seen in the original study that first described WRI1 (Focks and Benning, 1998). Furthermore, the lipid profile of the *wri1* mutant seeds shows a tendency to favour highly unsaturated and elongated fatty acids. This is perhaps a reverse effect of the previously mentioned correlation between high levels of 18:1 in tissues exhibiting high levels of fatty acid synthesis. Therefore, low production of new fatty acids could result in more interactions between modifying enzymes and the fatty acids that are produced.

Lipid content data of AtWRI4 infiltrations indicates a similar but weaker activation of acyl lipid biosynthesis as AtWRI1 which is in line with previous observations (To et al., 2012). The

finding is not statistically significant ($\alpha = 0.05$); however, this could be a result of an outlier as two of the three samples saw a threefold increase in total lipid content when compared to control. The previously mentioned samples also exhibit a similar increase in 18:1, further indicating an AtWRI1-like effect. There are no indications of changes in fatty acid profile towards longer chain acyl lipids which could be expected if WRI4 alone promoted wax synthesis as previously described (Park, Go and Suh, 2016). This could be the result of unsuitable lipid extraction and gas chromatography methods for wax identification or possibly indicate that AtWRI4 needs to be expressed in combination with other factors to activate the wax biosynthetic pathway. Wri4 knockouts show no significant decrease in TAG content which is in line with previous results (To et al., 2012). Although, there are somewhat higher levels of 18:1 and 18:2 in seed TAG profile when compared to WRI1 and control. Since no measurement were done on the Rest or Polar Lipid fraction in the seed acyl lipid analysis, no further conclusions can be made regarding the wax synthesis.

AtBBM show no effect on total TAG levels when transiently expressed in leaves and its acyl lipid profile closely resembles the control for all three fractions. This is expected as BBM seem to target a different set of genes than those involved in acyl lipid biosynthesis. Interestingly, there is a significant decrease in seed TAG accumulation for *bbm* mutants. The decrease in TAG accumulation could potentially be a result of BBMs interactions with the LAFL network and its role in embryo development (Boutilier, 2002; Horstman et al., 2017). Lec2, a member of the LAFL network, has in turn been identified as an upstream regulator for WRI1. Therefore, the reduction of TAG content might be an indirect effect caused by the lack of BBMs activity on upstream regulators of WRI1. However, similar to wri4 the lipid profile shows an increase in 18:1 and 18:2 compared to WRI1 indicating a high rate of fatty acid synthesis. This could be an indication that the reduction in total fatty content is a result of an underdeveloped embryo rather than being connected to WRI1. Furthermore, if BBM had the capacity to induce oil production through a secondary effect on WRI1 then there would be some indication of this in the leaf lipid analysis.

The result for BBM HTF proves that the oil accumulating ability of AtWRI1, and AtBBMs lack thereof, is directly connected to their respective double AP2 domain. This is in part surprising, because of the highly conserved nature of the region but not wholly unexpected. The native factors target completely different sets of genes, and since the double AP2 domain is thought to be the DNA binding domain of both proteins there is expected to be some key element that differ between the two. WRI4 HTF result show greater oil accumulating effect than native AtWRI4 but less than WRI1. This provides evidence that the C- and N-terminals of AtWRI1 have greater oil inducing ability than those native to AtWRI4. Furthermore, the reduced activity of WRI4 HTF compared to native AtWRI1 means that the AP2 domain of AtWRI1 has a greater ability to activate lipid biosynthesis than WRI4 does.

## 4.2 Regions in AP2 domains that could explain differences between TFs

A difference in binding selectivity is a possible explanation for why the two groups have such different transactivation patterns. With this in mind, the primary structure comparison between AtBBM, AtWRI4, AtWRI1 and homologues of WRI1 and BBM reveals at least two regions of the AP2 domain that could explain the differences seen between the WRI and PLT proteins.

The first one, located in a region 20-28 residues from the YRG motif, contains three charged residues in the PLT proteins, the WRI1 homologues have no polar residues present while WRI4 contains only one acidic residue. If this region is important to the DNA binding selectivity of the domain, the charged residues on the PLT group would probably interfere with the site that the corresponding WRI1 region interacts with. This interference is likely considering the large hydrophobic tryptophan that is present on most of the WRI1 homologues. This would also provide an explanation as to why WRI4 less strongly induce oil accumulation as the aspartic acid residue might interfere somewhat with the region. The reverse situation is also possible, where the tryptophan and surrounding region interferes with binding of the PLT group. However, this would not explain why WRI4 has a weaker activity on oil accumulation, although that could be a result of a different interaction. Residues 12-21 counted from the YRG motif has previously been identified as the main binding site for a 14-3-3 protein and its coexpression with WRI1 resulted in higher levels of lipid accumulation in seeds than WRI1 expression alone (Ma et al., 2016). Furthermore, the previous study identified the same region as a potential binding site for BPM proteins that has previously been shown to be a negative regulator of WRI1 by means of proteolytic degradation (Chen et al., 2013). This led to the hypothesis that 14-3-3 shields WRI1 from degradation by interfering with the binding of BPM proteins (Ma et al., 2016). Disruption of 14-3-3s binding or a more favourable interaction with BPM could explain the difference seen between the activities of WRI1 and WRI4. However, this would not explain the difference in binding selectivity seen between the WRI1 and PLT clave of TFs. This means that there is another part of the AP2 domain that explains the difference in binding selectivity seen between the groups, or that there are additional interactions between the region that explains the phenomenon.

The second region that is markedly different between BBM and WRI1, WRI4 as well as between the PLT and WRI claves is the region 59-63 residues from the YRG motif. This region is dense in threonine but also contain some proline and serine for the PLT factors while WRI4 and WRI1 has two hydrophobic residues as well as an aspartic acid in the same region. This could indicate a PEST region for the PLT factors and act as a signal for proteolytic degradation. However, this would not explain the sequence specificity discrepancy seen between the two groups. Furthermore, this region exists outside the 18 residue region surrounding the RAYD element that has previously been indicated as highly important to the function of the AP2 domain (Okamuro et al., 1997). The reason for the increase in activity seen in WRI4 HTF as compared to native WRI4 could be due to the high incidence of negatively charged residues on the C-terminal which have previously been pointed out as important for its transactivational capacity (Ma et al., 2015; Cooper, 2000).

## 4.3 Experimental procedures

The domain swapping of the hybrid constructs seems to have been successful based on the length of the constructs seen in the gel electrophoresis and later in the sequencing results of the pXZP393 plasmids. Furthermore, the sequencing results demonstrate successful integration of the *A. thaliana* native TFs as well as the hybrid constructs. The successful integration into expression clones and subsequent infiltration of *N. benthamiana* is supported by the effects seen on lipid levels in the lipid analysis.

## 4.4 Limitations

Because of the small sample sizes, it is hard to be certain about trends seen in the data. Larger sample sizes would make small patterns more noticeable, especially in the lipid profile measurements. Plants grown in the biotron were not controlled for positional effects that could influence the results. The bioinformatical analysis could be performed much more thoroughly and scientifically by comparing with other DNA binding motifs. More knowledge about the 3d structure of the tandem AP2 domain would also facilitate the bioinformatical analysis.

The start of the AP2 domain was set to the start of the YRG motif and equal length WRI4 and BBM were used for the AP2 domain of the hybrid constructs; since these boundaries are arbitrarily set based on homology between WRI1 and WRI4, it is possible that another result would be achieved with different boundaries.

There is no proof that any of the constructs were expressed in leaf tissue. A transcriptome analysis would provide important information about the gene regulatory effects of the constructs.

# 5 Conclusion

The differences seen between AtWRI1 and AtBBM are a direct result of differences within the AP2 domain. Two regions that might be responsible for these differences have been highlighted in this report. Both regions show motifs that could be involved in proteolytic degradation and one is reported to interact with a 14-3-3 protein. The difference in binding selectivity indicates additional interactions on these regions than degradation alone.

AtBBM does not induce lipid production when transiently expressed in leaf tissue, however, a lack of AtBBM leads to a lower level of total lipids in seeds. The results of this study hints that this could be the result of an underdeveloped embryo and not from any secondary effect on WRI1 expression.

The differences seen in the activity of AtWRI1 and AtWRI4 is a result of interactions both within the AP2 domain and from other parts of the proteins.

## 5.1 Future research

This report has not actively contributed to the development of high yielding oil crops or given solutions into engineering fatty acid composition. However, it has provided some insight into the domain function of a key regulator of fatty acid metabolism in plants and will hopefully inspire more studies in this area. It might therefore contribute to achieving the previously mentioned technologies or perhaps lead to other discoveries further down the road.

A more complete comparison of the transcription factors can be made once the samples taken for RNA sequencing are analysed. In particular it would be interesting to see what the effect of the HTFs would be.

To further understand the DNA binding domain of the AP2 subfamily and to be able to design transcription factors with novel function, it would be interesting to see how amino acid substitutions of the regions highlighted in this report would affect their function.

# References

Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., Nussaume, L., Noh, Y.,Amasino, R. and Scheres, B. (2004). The PLETHORA Genes Mediate Patterning of theArabidopsis Root Stem Cell Niche. *Cell*, 119(1), pp.109-120

Andriotis, V., Kruger, N., Pike, M. and Smith, A. (2009). Plastidial glycolysis in developing Arabidopsis embryos. *New Phytologist*, 185(3), pp.649-662.

Arabidopsis.org. (2018). TAIR - Home Page. Available at: https://www.arabidopsis.org [Accessed 16 May 2018].

Bartke, N. and Hannun, Y. (2008). Bioactive sphingolipids: metabolism and function. *Journal of Lipid Research*, 50(Supplement), pp.S91-S96.

Baud, S., Mendoza, M., To, A., Harscoët, E., Lepiniec, L. and Dubreucq, B. (2007). WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. *The Plant Journal*, 50(5), pp.825-838.

Berg, J., Gatto, G., Stryer, L. and Tymoczko, J. (2012). Biochemistry. 7th ed. New York: W.H. Freeman and Co., pp.883-987.

Bligh, E. and Dyer, W. (1959). A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION. *Canadian Journal of Biochemistry and Physiology*, 37(1), pp.911-917.

Boutilier, K. (2002). Ectopic Expression of BABY BOOM Triggers a Conversion from Vegetative to Embryonic Growth. *THE PLANT CELL ONLINE*, 14(8), pp.1737-1749.

Browse, J., Roughan, P. and Slack, C. (1981). Light control of fatty acid synthesis and diurnal fluctuations of fatty acid composition in leaves. *Biochemical Journal*, 196(1), pp.347-354.

Cernac, A. and Benning, C. (2004). WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *The Plant Journal*, 40(4), pp.575-585.

Chen, L., Lee, J., Weber, H., Tohge, T., Witt, S., Roje, S., Fernie, A. and Hellmann, H. (2013). Arabidopsis BPM Proteins Function as Substrate Adaptors to a CULLIN3-Based E3 Ligase to Affect Fatty Acid Metabolism in Plants. *The Plant Cell*, 25(6), pp.2253-2264.

Cooper GM. (2000). The Cell: A Molecular Approach. 2nd ed. [ebook] Sunderland (MA): Sinauer Associates, Regulation of Transcription in Eukaryotes. Available at: https://www.ncbi.nlm.nih.gov/books/NBK9904/#_ncbi_dlg_citbx_NBK9904 [Accessed 20 May 2018].

Cramer, P., Armache, K., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G., Dengl, S., Geiger, S., Jasiak, A., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C., Lehmann, E., Leike, K., Sydow, J. and Vannini, A. (2008). Structure of Eukaryotic RNA Polymerases. *Annual Review of Biophysics*, 37(1), pp.337-352.

Fahy, E., Cotter, D., Sud, M. and Subramaniam, S. (2011). Lipid classification, structures and tools. Biochimica et Biophysica Acta (BBA) - *Molecular and Cell Biology of Lipids*, 1811(11), pp.637-647.

Fatihi, A., Boulard, C., Bouyer, D., Baud, S., Dubreucq, B. and Lepiniec, L. (2016). Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds. *Plant Science*, 250, pp.198-204.

Focks, N. and Benning, C. (1998). wrinkled1: A Novel, Low-Seed-Oil Mutant of Arabidopsis with a Deficiency in the Seed-Specific Regulation of Carbohydrate Metabolism. *Plant Physiology*, 118(1), pp.91-101.

Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R. and Scheres, B. (2007). PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. *Nature*, 449(7165), pp.1053-1057.

Gonzalez, D. (2016). Chapter 1 - Introduction to Transcription Factor Structure and Function. In: D. Gonzalez, ed., Plant Transcription Factors. Boston: Academic Press, pp.3-11. ISBN 9780128008546

Gregorio, J., Hernández-Bernal, A., Cordoba, E. and León, P. (2014). Characterization of Evolutionarily Conserved Motifs Involved in Activity and Regulation of the ABA-INSENSITIVE (ABI) 4 Transcription Factor. *Molecular Plant*, 7(2), pp.422-436.

Grimberg, Å., Carlsson, A., Marttila, S., Bhalerao, R. and Hofvander, P. (2015). Transcriptional transitions in Nicotiana benthamiana leaves upon induction of oil synthesis by WRINKLED1 homologs from diverse species and tissues. *BMC Plant Biology*, 15(1).

Hoekema, A., Hirsch, P., Hooykaas, P. and Schilperoort, R. (1983). A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid. *Nature*, 303(5913), pp.179-180.

Holland, J., Frey, K. and Hammond, E. (2001). Correlated responses of fatty acid composition, grain quality, and agronomic traits to nine cycles of recurrent selection for increased oil content in oat. Euphytica, 122(1), pp.69-79.

Horstman, A., Bemer, M. and Boutilier, K. (2017). A transcriptional view on somatic embryogenesis. *Regeneration*, 4(4), pp.201-216.

Horstman, A., Li, M., Heidmann, I., Weemen, M., Chen, B., Muiño, J., Angenent, G. and Boutilier, K. (2017). The BABY BOOM transcription factor activates the LEC1-ABI3-FUS3-LEC2 network to induce somatic embryogenesis. *Plant Physiology*, 175 (2) pp.848-857

Jong, C. (2016). Chapter 3 - General Aspects of Plant Transcription Factor Families. In: D. Gonzalez, ed., Plant Transcription Factors. Academic Press, pp. 35-56, ISBN 9780128008546

Kapila, J., De Rycke, R., Van Montagu, M. and Angenon, G. (1997). An Agrobacterium-mediated transient gene expression system for intact leaves. *Plant Science*, 122(1), pp.101-108.

Lee, T. and Young, R. (2000). Transcription of Eukaryotic Protein-Coding Genes. *Annual Review of Genetics*, 34(1), pp.77-137.

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

Ma, W., Kong, Q., Grix, M., Mantyla, J., Yang, Y., Benning, C. and Ohlrogge, J. (2015). Deletion of a C–terminal intrinsically disordered region of WRINKLED1 affects its stability and enhances oil accumulation in Arabidopsis. *The Plant Journal*, 83(5), pp.864-874.

Ma, W., Kong, Q., Mantyla, J., Yang, Y., Ohlrogge, J. and Benning, C. (2016). 14-3-3 protein mediates plant seed oil biosynthesis through interaction with AtWRI1. *The Plant Journal*, 88(2), pp.228-235.

Mitchell, P. and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, 245(4916), pp.371-378.

Nakamura, Y. and Li-Beisson, Y. (2016). Lipids in Plant and Algae Development. London: Springer International Publishing.

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C. and Dunker, A. (2003). Predicting intrinsic disorder from amino acid sequence. *Proteins: Structure, Function, and Genetics*, 53(S6), pp.566-572.

Okamuro, J., Caster, B., Villarroel, R., Van Montagu, M. and Jofuku, K. (1997). The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. *Proceedings of the National Academy of Sciences*, 94(13), pp.7076-7081.

Park, C., Go, Y. and Suh, M. (2016). Cuticular wax biosynthesis is positively regulated by WRINKLED4, an AP2/ERF-type transcription factor, in Arabidopsis stems. *The Plant Journal*, 88(2), pp.257-270.

Piironen, V., Lindsay, D., Miettinen, T., Toivo, J. and Lampi, A. (2000). Plant sterols: biosynthesis, biological function and their importance to human nutrition. *Journal of the Science of Food and Agriculture*, 80(7), pp.939-966.

Riechmann, J. and Meyerowitz, E. (1998). The AP2/EREBP Family of Plant Transcription Factors. biological chemistry, 379(1), pp.633 – 646

Ruuska, S., Girke, T., Benning, C. and Ohlroggea, J. (2002). Contrapuntal Networks of Gene Expression during Arabidopsis Seed Filling. *THE PLANT CELL ONLINE*, 14(6), pp.1191-1206.

Santuari, L., Sanchez-Perez, G., Luijten, M., Rutjens, B., Terpstra, I., Berke, L., Gorte, M., Prasad, K., Bao, D., Timmermans-Hereijgers, J., Maeo, K., Nakamura, K., Shimotohno, A., Pencik, A., Novak, O., Ljung, K., van Heesch, S., de Bruijn, E., Cuppen, E., Willemsen, V., Mähönen, A., Lukowitz, W., Snel, B., de Ridder, D., Scheres, B. and Heidstra, R. (2016). The PLETHORA Gene Regulatory Network Guides Growth and Cell Differentiation in Arabidopsis Roots. *The Plant Cell*, 28(12), pp.2937-2951.

Sheng, J. and Citovsky, V. (1996). Agrobacterium-Plant Cell DNA Transport: Have Virulence Proteins, Will Travel. *THE PLANT CELL ONLINE*, 8(10), pp.1699-1710.

Shigyo, M. and Ito, M. (2004). Analysis of gymnosperm two-AP2-domain-containing genes. *Development Genes and Evolution*, 214(3), pp.105-114.

Smith, E. and Townsend, C. (1907). A PLANT-TUMOR OF BACTERIAL ORIGIN. *Science*, 25(643), pp.671-673.

Thermofisher.com. (2018). Tm Calculator | Thermo Fisher Scientific. Available at: https://www.thermofisher.com/se/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html#instruction [Accessed 23 May 2018].

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

User guide to: Gateway® Technology A universal technology to clone DNA sequences for functional analysis and expression in multiple systems. (2018). 1st ed. [ebook] Invitrogen. Available at: https://tools.thermofisher.com/content/sfs/manuals/gatewayman.pdf [Accessed 19 Apr. 2018].

Vanhercke, T., El Tahchy, A., Shrestha, P., Zhou, X., Singh, S. and Petrie, J. (2013). Synergistic effect of WRI1 and DGAT1 coexpression on triacylglycerol biosynthesis in plants. FEBS Letters, 587(4), pp.364-369.

White, T., Bursten, S., Federighi, D., Lewis, R. and Nudelman, E. (1998). High-Resolution Separation and Quantification of Neutral Lipid and Phospholipid Species in Mammalian Cells and Sera by Multi-One-Dimensional Thin-Layer Chromatography. *Analytical Biochemistry*, 258(1), pp.109-117.

Zale, J., Jung, J., Kim, J., Pathak, B., Karan, R., Liu, H., Chen, X., Wu, H., Candreva, J., Zhai, Z., Shanklin, J. and Altpeter, F. (2015). Metabolic engineering of sugarcane to accumulate energy-dense triacylglycerols in vegetative biomass. *Plant Biotechnology Journal*, 14(2), pp.661-669.

Zupan, J., Muth, T., Draper, O. and Zambryski, P. (2000). The transfer of DNA from Agrobacterium tumefaciens into plants: a feast of fundamental insights. *The Plant Journal*, 23(1), pp.11-28.

# Appendix