THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Abstract

Plants produce a wide variety of natural products that can be exploited for medicinal purposes. Paclitaxel is a key anti-cancer drug originally isolated from the bark of *Taxus spp*. that is currently approved for use in the treatment of breast, lung and non-small cell cancers, AIDS-related Kaposi's sarcoma and coronary artery disease. Worldwide demand for paclitaxel is high and plant cell culture (PCC) is an attractive production route. Cultured cambial meristematic cells (CMCs) provide a good platform from which to increase drug production, as they possess superior growth properties on an industrial scale compared to typical dedifferentiated cell culture. Elicitors, such as methyl-jasmonate (MeJA), can up-regulate paclitaxel production in PCC, however the effect is only transient. Identification and characterisation of the key transcriptional regulators that control MeJA induced metabolic reprogramming can provide potential tools to manipulate *Taxus* CMC culture to produce more paclitaxel.

Roche454 sequencing was employed to establish the basic transcriptomic profile of *Taxus cuspidata* CMCs, which was then utilized as a reference to observe the transcriptional profile of CMCs at three time points after MeJA elicitation (0.5, 2 and 12 h). Analysis of the transcriptional regulatory network identified 19 transcription factors (TFs) that were significantly up-regulated at an early time point (0.5 h) after elicitation. These TFs came from five families – AP2, MYB, NAC, bHLH and WRKY – that are well known to regulate secondary plant metabolism. An *Arabidopsis thaliana* transient expression assay (TEA) was employed to investigate the regulatory activity of these 19TFs against 10 paclitaxel biosynthetic promoters.

The TEA screen identified 79 significant interactions with every promoter interacting with at least three TFs, which could activate or repress activity. A MYB TF was identified that could up-regulate eight out of the ten promoters tested, indicating it maybe a potential overall regulator of paclitaxel biosynthesis. *In vitro* electromobility shift assays established the possible binding site for this TF as an AC element, with the consensus sequence of A(A/C)C. Repressors of promoter activity were also identified, for example an AP2 TF which contains the well-established ERF associated amphiphilic repression (EAR) motif. The activity of the EAR domain was explored *in vivo* using a TEA assay and site directed mutagenesis mutants. Activity was lost when the mutation occurred within the domain suggesting the TF was working as an active repressor.

TFs can work individually or in combination to achieve metabolic reprogramming after MeJA elicitation. One of the best characterised examples of plant combinatorial control is between particular sub classes of MYB and bHLH TFs. However investigation into possible interactions between the *T. cuspidata* MYB and bHLH TFs *in vivo* using yeast two hybrid and TEAs found few combinations that led to a significant change in regulatory activity. The regulatory activity of WRKY TFs was shown to be post-translationally controlled when the TEAs were treated with MeJA, however the mechanism by which this occurs remains to be

elucidated.

The interactions identified between the 19 TFs and the paclitaxel biosynthetic promoters can be exploited in the future to produce superior *Taxus* CMC lines with increased paclitaxel yields.

Lay summary

Plants produce a variety of compounds that can be exploited for medicinal purposes. A classic example is Aspirin, found in willow trees, which has been used since Egyptians times as a painkiller. The anti-cancer drug paclitaxel is extracted from the bark of yew trees and can be used to treat a number of cancers including breast and lung cancer. Worldwide sales of the drug exceeded $1 billion in 2014 and its medicinal uses are continually expanding. Producing sufficient quantities of paclitaxel to keep up with this rising demand is problematic as the drug is found at very low amounts within the tree.

Plants can be grown in liquid cultures and this is seen as an attractive production route for paclitaxel. This thesis is based on cultures made from yew tree cambial meristematic cells. These cells have stem cell like properties, which mean they have higher growth rates and product yields compared to traditional cultures.

When plants are stressed, such as when they are being eaten by animals or insects, they produce hormones to signal the whole plant to respond to the threat. One of the defence responses is to produce compounds, like paclitaxel, which deter the animal from continuing to feed. A hormone called methyl-jasmonate can be used to increase yields of paclitaxel in plant cell culture, but the effect is only short-lived. The aim of this project is to understand how the plant regulates its response to methyl-jasmonate and to exploit this knowledge in efforts to increase the production of paclitaxel in plant cell culture.

Acknowledgements

I would like to thank the BBSRC and Unhwa for funding my PhD and the James Rennie Bequest for awarding me with a travel grant to attend and present my work at the 52nd Annual meeting of the Phytochemical Society of North America.

I would like to thank my supervisor Gary Loake and all the members of the Paclitaxel project past and present, in particular Marisol Ochoa Villarreal.

I am thankful to Unhwa for proving the plant material utilised in this thesis

I am grateful for all the help and advice from all the members of the Loake Lab past and present and the support of all the IMPS staff.

I would also like to thank my friends and family for their support and encouragement over the last four years; in particular my parents and great friends Pumi Perera, Rebecca Dewhirst and Georgie Ablett.

Declaration

I hereby declare that the work presented in this thesis is my own, completed as part of a larger group project with the contribution of other members clearly stated in the text and has not been submitted any other degree or professional qualification at this or any other university.

Susan Howat

Publications

S Howat, B Park, I.S Oh, Y-W Jin, E-K Lee and G.J Loake (**2014**) Paclitaxel: Biosynthesis, production and future prospects, *New biotechnology* 31 (3) 242-245

M Ochoa Villarreal, S Howat, M.O Jang, I.S Kim, Y.W Jin, E.K Lee and G.J Loake (**2015**) Cambial meristematic cells: a platform for the production of plant natural products *New biotechnology* doi:10.1016/j.nbt.2015.02.003

Table of Contents

Abbreviations

(cis-(+)-OPDA)	(9S,13S)-12-oxo-phytodienoic acid
10-DAB	10-deacetylbaccatin III
12,13(S)-EOT	12,13(S)-epoxy-octadecatrienoic acid
13(S)-HPOT	13(S)-hydroperoxy-octadecatrienoic acid
3AT	3-amino-1,2,4-triazole
ACC	Aminocyclopropane-1-carboxylic acid synthase
ACT	Aspartate kinase, chorismate mutase and TyrA
ACX	Acyl-CoA oxidase
ANR	Anthocyanidin reductase
AOC	Allene oxide cyclase
AOS	Allene oxide synthase
AP2	APETALA2
ARF	Auxin response actor
Aux/IAA	Auxin/Indole-3-Acetic Acid
BAPT	Baccatin III: 3-amino, 13-phenylpropanoyl transferase
bHLH	BASIC HELIX LOOP HELIX
BLAST	Basic local alignment search tool
BMS	Bristol-Myers Squibb
bp	Base pair
BS	Binding site
C1	COLORED1
CASTing	Cyclic amplification and selection of targets
CDP-ME	1,4-diphosphocytidyl-2*C*-methyl-D-erythritol
CHI	Chalcone isomerase
CHS	Chalcone synthase
CHX	Cycloheximide
CMC	Cambial meristematic stem cells
CMK	CDP-ME kinase
CMS	1,4-diphosphocytidyl-2*C*-methyl-D-erythritol(CDP-ME) synthase
CoA-Ligase	β -phenylalanoyl-CoA Ligase
COI1	Coronatine insensitive 1
cP450	cytochome cP450
CUC	CUP SHAPED COTYLEDON
DAD1	Defective in anther dehiscence 1
Dat	4-O-deacetylvindoline 4-O-acetyltransferase
DBAT	10-deacetylbaccatin III-10-*O*-acetyltransferase
DBBT	Taxane-2α-ol-*O*-benzoyl transferase
DBTNBT	N-benzoyltransferase
DDC	Dedifferentiated cell culture
DEGs	Differential expressed genes
Del	DELILA
DFR	Dihydroflavonol 4-reductase
DGL	DONGLE
DMAPP	Dimethylallyl pyrophosphate
DMSO	Dimethyl sulfoxide
DREB	Dehydration-responsive element-binding protein
DTT	Dithiothreitol
DXR	1-deoxy-D-xylulose 5-phosphate reductoisomerase
DXS	1-deoxy-D-xylulose 5-phosphate synthase
EAR	Ethylene responsive factor–associated amphiphilic repression
EBGs	Early biosynthetic genes
EGL3	ENHANCER of GLABRA 3
EIL1	EIN3-LIKE1
EIN3	ETHYLENE INSENSITIVE3
EMSA	Electromobility shift assay
ERF	Ethylene responsive factor
F3H	Flavanone 3-hydroxylase
F3'H	Flavanone 3'-hydroxylase
FDA	Federal Department of Agriculture
FPS	Farensyl diphoshate synthase
G10H	Geraniol 10-hydroxylase
GAI	GIBBERELLIC ACID INSENSITIVE
GAI	Gibberellic acid
GFP	Green fluorescent protein
GGPP	Geranyl geranyl diphoshate
GGPPS	Geranylgeranyl diphosphate synthase
GL1	GLABRA 1
GL3	GLABRA 3
GPS	Geranyl diphoshate synthase

GTP	Guanosine-5'-triphosphate
GUS	β-glucuronidase
h	Hours
HAD	Histone deacetylase
HDR	HMBPP reductase
HDS	4-Hydroxy-3-methylbut-2-enyl diphosphate (HMBPP) synthase
HMG-CoA	3-Hydroxy-3-methylglutaryl-CoA
IAA-aaH	IAA amino acid hydrolase
IAA-aaS	IAA-amido synthetase
IPP	Isopentenyl pyrophosphate
IPPI	Isopentenyl pyrophosphate isomerase
JA	Jasmonic acid
JA-Ile	Jasmonyl-L-isoleucine
JAM1	Jasmonic acid inducible MAPK kinase
JAR1	JA-amino acid synthetase
JAZ	JA ZIM domain
JERE	Jasmonate- and elicitor-responsive
JID	JAZ interaction domain
JMT	JA methyl transferase
KAT	L-3-ketoacyl-CoA thiolase
LBGs	Late biosynthetic genes
LDOX	leucoanthocyanidin dioxygenase
logFC	Log of fold change
LOX	Lipoxygenase
MAPK	Mitogen-activated protein kinase
MBW complex	R23-MYB-bHLH-WD-40 complex
MCS	2-*C*-Methyl-ᴅ-erythritol 2,4-cyclodiphosphate synthase
MeJA	Methyl jasmonate
MEP	Methylerythritol phosphate
MFP	Multifunctional protein
MT	Microtubule
MVA	Mevalonic pathway
MVA	Mevalonate
MVAP	Phosphomevalonate
MVAPP	Diphosphomevalonate
MYB	MYELOBLASTOSIS
NAC	NAM/ ATAF1,2/CUC2
NACRS	NAC recognition sequence
NARD	NAC repression domain
NARD	NAC repression domain
NC	Negative control
NCI	National Cancer Institute
NINJA	NOVEL INTERACTOR OF JAZ
NP	Natural Product
OPC-8	3-oxo-2-(2-pentenyl)-cyclopentane-1-octanoic acid
OPR3	OPDA reductase 3
ORCA	Octadecanoid-derivative Responsive Catharanthus AP2-domain
ORF	Open reading frame
PAM	Phenylalanine amminomutase
PAP1	PRODUCTION OF ANTHOCYANIN PIGMENT 1
PCC	Plant cell culture
PEG	Polyethylene glycol
PLACE	Plant Cis-acting Regulatory DNA Elements
PLD	Phospholipase D
PNP	Plant natural product
ppm	Part per million
PXY	Phloem intercalated with xylem
RAV	Related to ABI3/VP1
RGA	REPRESSOR OF GA
RGL1	RGA-LIKE1
RLU	Relative luminescence unit
Rluc	*Renilla* luciferase
RNAi	RNA interference
Ros1	ROSEA1
SA	Salicylic acid
SAM	*S*-adenosylmethionine-methyltransferase
SAR	Structure activity relationship
SCF^{COI1}	Skp–Cullin–F-box-type E3 ubiquitin ligase complex
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SIB	Sigma factor binding proteins
SNO	*S*-nitrosylation
Sp3	SPECIFIC PROTEIN 3
STR	Strictosidine synthase
T10βH	Taxane-10β-hydrolase
T13αH	Taxadiene-13α-hydrolase
T14βH	Taxane-14β-hydrolase
T1βH	Taxane-1β-hydrolase
T2'αH	Taxane 2'α-hydroxylase

T2αH	Taxane-2α-hydrolase
T5αH	Taxadiene-5α-hydrolase
T7βH	Taxane-7β-hydrolase
T9αH	Taxane-9α-hydroxylase
TASY	Taxadiene synthase
TCA	Tricarboxylic acid
TDAT	Taxadiene-5α-ol-*O*-acetyl transferase
TDR1	TRANSCRIPTIONAL REGULATOR OF DEFENCE RESPONSE
TE	Thioesterase
TEA	Transient expression assay
TF	Transcription factor
TIA	Terpenoid indole alkaloids
TIR1	TRANSPORT INHIBITOR RESPONSE PROTEIN 1
T-O-AT	Taxoid-*O*-acetyl transferase
TPL	TOPLESS
TPR	TPL-related proteins
TT12	TRANSPARENT TESTA 12
TT8	TRANSPARENT TESTA 8
USDA	US Department of Agriculture
WD-40	WD-repeat protein
WER	WEREWOLF
WHO	World Health Organisation
WOL	Wooden leg
Y2H	Yeast-two-hybrid

List of Figures

List of Tables

Chapter 1 – Introduction

1.1 Plant natural products

Plants can synthesise a plethora of small organic compounds, known as natural products (NPs). They are products of secondary metabolism and can be loosely defined as compounds that are not essential for growth and development, but play an important role in the plant's response to environmental stimuli (1). NPs accumulate at low levels within the plant, usually <1% dry weight, and are frequently produced in specialised cells in response to biotic and abiotic stresses. There are three main categories of plant natural products (PNPs): 1) Isoprenoids (also known as terpenoids) comprise the largest class of PNPs containing an estimated 40 000 compounds that have a wide variety of uses including food colourants, flavourants and pharmaceuticals. They are based on the 5 carbon building blocks isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) and have a diverse range of physiological and ecological roles within the plant (Figure 1a); 2) Alkaloids, the second largest group, are comprised of low molecular weight nitrogen containing compounds that are mostly derived from amino acids and can act as neurological agents (Figure 1b); and 3) Phenylpropanoids (also known as flavonoids) that are produced via the shikimate pathway, are ubiquitous across terrestrial plants and possess anti-oxidant activity (Figure 1c) (2).

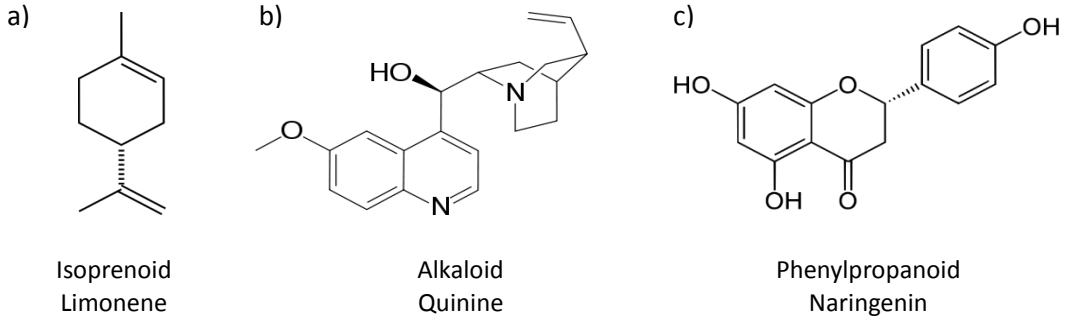
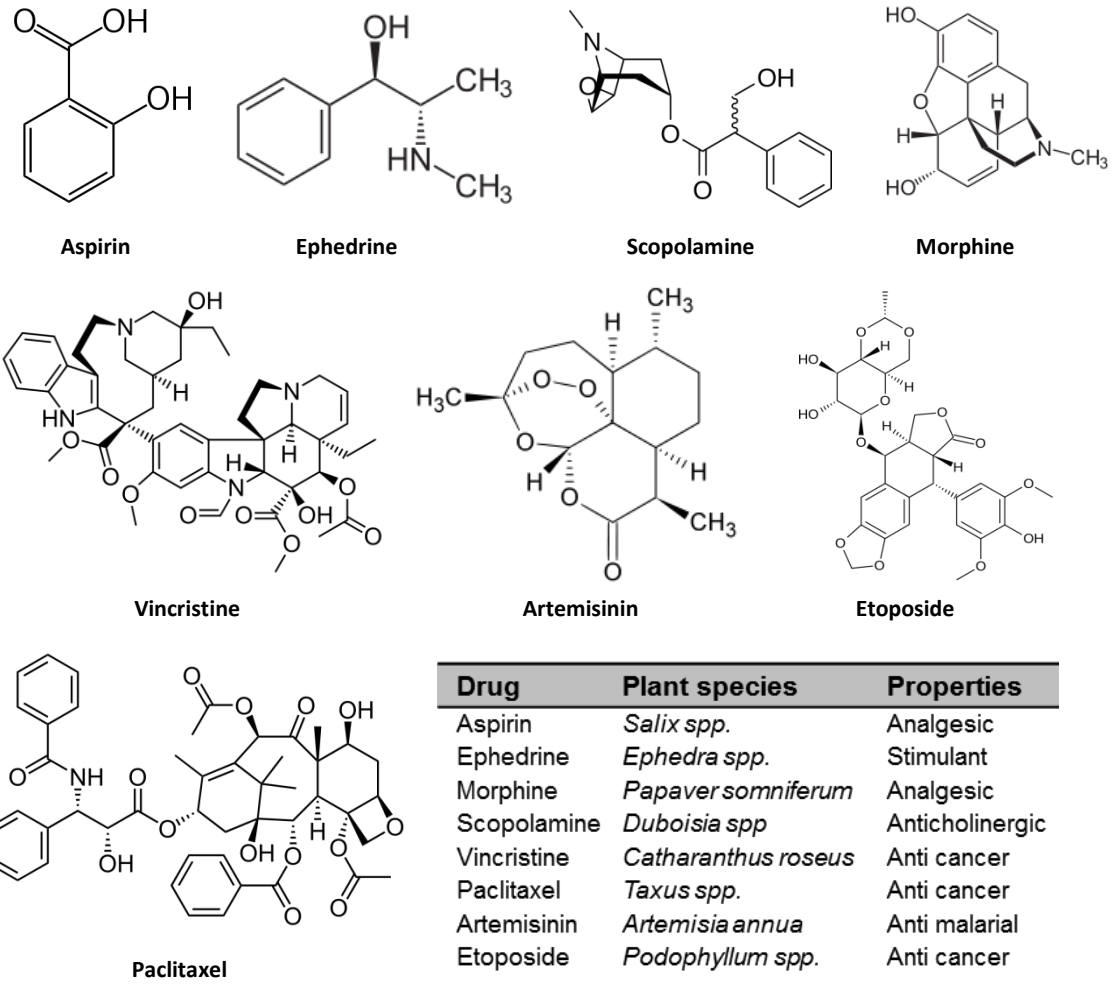


Figure 1 – Example structures of the three main classes of plant natural products, a) Isoprenoids, b) Alkaloids and c) Phenylpropanoids. a) Limonene is a cyclic isoprenoid comprised of two isopentyl pyrophosphate 5 carbon building blocks; b) Quinine is an alkaloid that contains two nitrogen atoms and is derived from tryptophan (3); and c) Naringenin is a phenylpropanoid (flavonoid) synthesised from the amino acids phenylalanine and tyrosine produced via the shikimate pathway.

1.1.1 The use of PNPs as pharmaceuticals

A major commercial use of PNPs is as pharmaceuticals. Approximately 25% of the drugs used worldwide are derived from plants and 10% of the drugs classified by the World Health Organisation (WHO) as "basic and essential" have a plant origin (4). The proportion of anti-bacterial and anti-cancer drugs that are natural products (NPs) or based on NPs is high at 75% and 65% respectively (5). PNPs have a wide range of chemical structures and pharmaceutical uses (Figure 2). Some valuable PNPs have simple structures, such as aspirin and ephedrine, allowing facile chemical synthesis, but many have complex structures containing multiple chiral

centres and heterocyclic rings that can make total chemical synthesis difficult and commercially unviable (Figure 2) (6). Semi-synthetic derivatives of NPs can also be employed as drugs, for example the anticancer drug Etoposide is a derivative of podophyllotoxin which in its original form was highly toxic (1).



Drug	Plant species	Properties
Aspirin	*Salix spp.*	Analgesic
Ephedrine	*Ephedra spp.*	Stimulant
Morphine	*Papaver somniferum*	Analgesic
Scopolamine	*Duboisia spp*	Anticholinergic
Vincristine	*Catharanthus roseus*	Anti cancer
Paclitaxel	*Taxus spp.*	Anti cancer
Artemisinin	*Artemisia annua*	Anti malarial
Etoposide	*Podophyllum spp.*	Anti cancer

Figure 2 – Examples of pharmaceutical plant natural products from a variety of species. The structure, plant species and major pharmacological property are stated for each drug.

The chemical diversity of plants makes them a rich source of pharmaceuticals, much greater than that achieved by humans using combinatorial chemistry. In previous decades a large amount of time and money was expended by pharmaceutical companies employing combinatorial chemistry to try and identify new lead compounds. However to date only one *de novo* drug has made it to market, Sorafenib (Bayer) an antitumour compound approved for the treatment for renal and liver carcinoma (5). Due to the low accumulation of PNPs and their high structural complexity ensuring sufficient supply of a compound can be problematic as is the case for the anti-cancer drug paclitaxel.

1.2 Paclitaxel

Paclitaxel is a key anti-cancer drug originally isolated from the bark of *Taxus spp.* with worldwide sales exceeding $1 billion in 2014. In this section the origin, mechanism of action,

structure-activity relationships, biosynthetic pathway and routes of paclitaxel production are described

1.2.1 History

The first sample of *Taxus brevifolia* was collected in 1962 by the US Department of Agriculture (USDA) as part of the plant screening program undertaken by the National Cancer Institute (NCI). Between 1963–64 extracts from the bark of *T. brevifolia* were observed to show significant signs of both cytotoxic and anti-leukemia activity in KB cytotoxicity assays (7). The active constituent, paclitaxel, was identified by Dr Monroe Wall in 1969 and in collaboration with Dr Mansukh Wani its structure was elucidated in 1971 (Figure 2) (8). Paclitaxel was not seen as a promising drug candidate until additional models were introduced by the NCI in 1975, when it showed strong activity against B16 melanoma. Studies by Dr Susan Horwitz and co-workers discovered that paclitaxel had at the time a unique mode of action promoting tubulin polymerization and stabilization of microtubules (see 1.2.2 Mode of action) (9, 10). Preclinical studies were initiated in 1980 but ran into major toxicity issues due to the compounds extremely low solubility, leading to higher dose requirements. After 12 years the drug was approved by the Federal Department of Agriculture (FDA) against ovarian cancer and over the years its uses have expanded to include breast, pancreatic and non-small-cell lung cancer, AIDS-related Kaposi's sarcoma and coronary artery disease (11–13). Prior to FDA approval the production of paclitaxel was passed to the company Bristol-Myers Squibb (BMS), which went on to trademark the compound as Taxol®. As can be seen in Table 1, the uses of paclitaxel are still expanding and recent studies have shown that the pathological changes induced by mutant-human-tau expression in human neurons can be rescued by 10 nM paclitaxel, suggesting that in the future paclitaxel could be used as a treatment for Alzheimer's disease (14).

Year	Event	Reference
1962	*Taxus brevifolia* bark sample collected by USDA & NCI	
1964	Cytotoxicity of extract identified	
1971	Paclitaxel isolated and structure elucidated	(8)
1975	Strong activity observed against the B16 melanoma model	
1979	Paclitaxel's unique mechanism of action identified	(9)
1983	Paclitaxel enters clinical trials against ovarian cancer	
1992	FDA approves paclitaxel for treatment of ovarian cancer	
1994	FDA approves paclitaxel for treatment metastatic breast cancer	
1997	FDA approves paclitaxel for treatment Kaposi's AIDs related sarcoma	
2011	FDA approves paclitaxel for use in coronary artery shunt	
2012	FDA approves Abraxane (paclitaxel protein bound) for treatment in advanced non-small cell lung cancer	
2013	FDA approves Abraxane for treatment in late stage Pancreatic cancer	

Table 1 – Key dates in the history of paclitaxel drug development and its expanding pharmacological uses with associated references. Abbreviations: Federal Department of Agriculture (FDA), US Department of Agriculture (USDA), National Cancer Institute (NCI).

Originally identified in *T. brevifolia* the production of paclitaxel has been identified in other

Taxus spp. including but not limited to *Taxus baccata*, *Taxus cuspidata*, *Taxus chinensis*, *Taxus canadensis*, *Taxus yunnanensis* and *Taxus x media* (a cross of *T. baccata* and *T. cuspidata*) (15). The first reported case of a paclitaxel producing fungus, *Taxomyces andreanae,* was reported in 1993 which was isolated from *T. brevifolia* (16). Subsequent studies have identified more paclitaxel producing endophytes, however their yields are generally low in the range of 24–70 ng/L. Production of taxanes, including paclitaxel, has also been described in hazel (*Corylus avellana*) (17) suggesting that the compounds are not unique to *Taxus spp*; however *Taxus spp.* are the current major source.

1.2.2 Mode of action

In 1979 paclitaxel was shown to have a unique mechanism of action. Previously known anticancer drugs affecting the mitotic spindle (e.g. vinblastine and vincristine) acted by rapidly depolymerizing microtubules preventing spindle formation during mitosis. Paclitaxel has an opposite mode of action stabilizing tubulin polymers which prevents correct spindle formation leading to inhibition of cell replication at the G_2/M phase (18, 19).

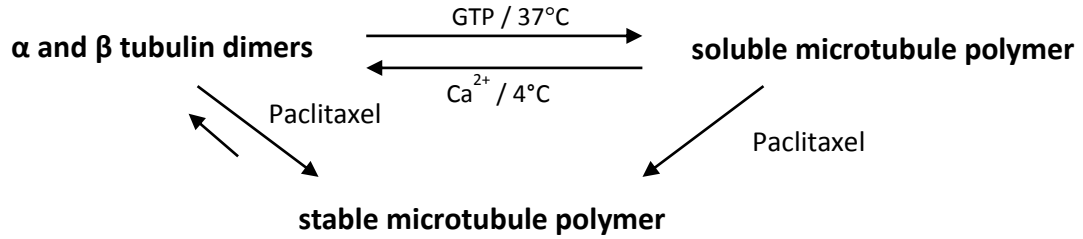


Figure 3 – A schematic showing the mode of action of paclitaxel. Normally there is an equilibrium between tubulin dimers and microtubule (MT) polymers. Higher temperatures or the addition of Guanosine-5'-triphosphate (GTP) favours MT polymer formation, while the depolymerisation of MTs can be accomplished by the addition of calcium or cold temperature. The addition of paclitaxel alters the equilibrium in favour of MT formation and stabilises the MTs once formed making them resistant to depolymerisation by calcium or cold temperature. Schematic modified from Horwitz *et al*. 1992 (20).

Microtubules (MTs) are an essential component of the mitotic bundle and are required for cell shape and cell motility. *In vivo* MTs are in a dynamic equilibrium between soluble tubulin dimers and MT polymers. Paclitaxel alters the equilibrium in favour of MTs by lowering the critical concentration of tubulin required for polymer formation and stabilizing formed MTs (Figure 3). Paclitaxel binds specifically and reversibly to MTs, preferentially through the β tubulin subunit, with a stoichiometry approaching 1 (tubulin dimer to paclitaxel) (20). The enhanced rate and extent of MT formation is independent of exogenous GTP application and is resistant to depolymerisation by cold treatment or calcium chloride (10).

1.2.3 Structure–activity relationships

Paclitaxel is a complex tetracyclic terpenoid with a large number of functional groups that are important for activity. Numerous research groups have synthesised a large number of synthetic analogues in order to elucidate its structure–activity relationships (SARs). These studies are summarized in Figure 4 and show that the C-13 side chain, C-2 benzoate and oxetane ring are

essential for cytotoxic activity (21–23).

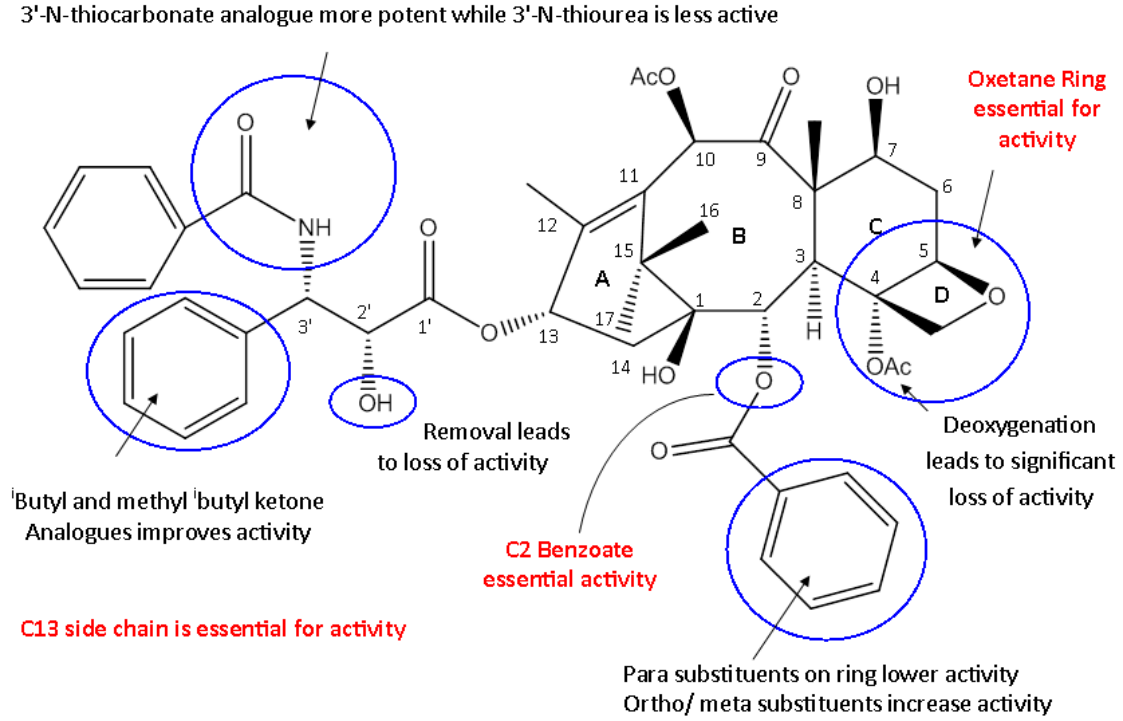


Figure 4 – Summary of the structure–activity relationships of paclitaxel. The majority of the cytotoxic activity comes from the southern hemisphere of the molecule involving C1–C5 and C13. Carbon positions and the four rings are indicated. Highlighted in red are parts of the molecule shown to be essential for activity: the C13 side chain, C2 benzoate group and the oxetane ring. Highlighted in the blue rings are specific changes in substituents that alter the cytotoxic activity.

1.2.3.1 C-13 side chain

Analogues with altered side chains are synthesised either by the addition of modified side chains to baccatin III or by altering existing side chains, however the latter is harder to accomplish as there are many functional groups on the molecule that require selective protection (23). Changing the 3'-phenyl group to an 3'-isobutenyl or 3'-isobutyl leads to an increase in activity, while the introduction of an epoxide produces an enantiospecific response. The 3'-(*R*)-epoxide is extremely potent while the 3'-(*S*)-epoxide is 10–30 times less active (21), proving that paclitaxel binding *in vivo* is very specific. The amide bond and the 2'-hydroxyl group of the side chain are also essential for activity (23).

1.2.3.2 C-2 benzoate

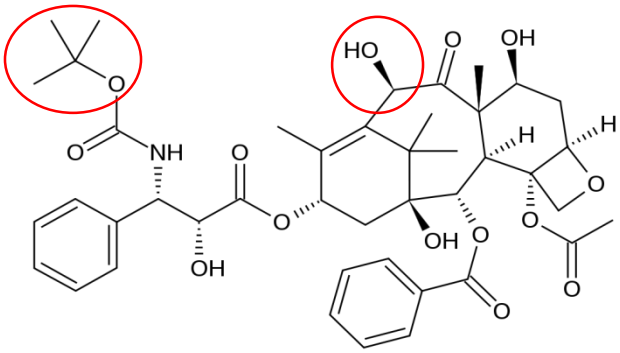
The Kingston group produced a variety of substituted benzoate groups, which showed correlation between the position of groups on the benzene ring and cytotoxity (22). *Meta*-substituted compounds were more active than paclitaxel, whereas *para* and *ortho* substituents led to a decrease in activity. Introduction of a second group on the ring led to a reduction in activity (21).

1.2.3.3 Rings

Paclitaxel contains four rings that are important for activity. Opening of the oxetane ring leads to loss of activity, however, the physiological explanation for this is unclear. Sulphur analogues have at least a 10-fold reduction in activity compared to paclitaxel, indicating the ability of the oxygen to act as a hydrogen acceptor is essential for cytotoxicity. The introduction of a sulphur atom also alters the sterics of the molecule, which may lead to steric congestion affecting the compounds ability to bind to tubulin (23). Analogues of the other rings also lead to a reduction or loss in activity. The analogues nor-seotaxel and nor-secotaxotere have a bridged oxygen in the **B** ring and an opened **A** ring that leads to a 20-40 fold reduction in activity (24), contraction of the **C** ring also reduces the cytotoxicity of the compound and the introduction of a bridged system at C4–C6 leads to complete loss of activity (21, 25).

1.2.3.4 Other important functional groups

Various other functional groups are important for toxicity, including a number of hydroxyl groups in the molecule. Deoxygenation of the C2 and C4 positions leads to a significant loss in activity, whereas the C7 hydroxyl group is not required for function (23). The trend in the SARs of paclitaxel is that changes to the southern hemisphere of the molecule (involving C1–C5 and C13) have a greater impact on the compound's activity compared to modification made in the northern hemisphere (C6–C12) (22). This knowledge allowed the development of the paclitaxel analogue Docetaxel (Taxotere®) (Structure 1), which is twice as effective against certain cancers compared to paclitaxel and has a higher solubility.



Structure 1 – The structure of the Docetaxel (Taxotere®) an analogue of paclitaxel with the regions of molecule that are altered highlighted in red. The 3'-phenyl group has changed to 3'-isobutyl and the C10 acetoxy group has changed to a hydroxyl group.

1.2.4 Paclitaxel biosynthetic pathway

This section is adapted from a review on Paclitaxel biosynthesis and production by Howat *et al.* (26). Paclitaxel has a highly complex tetracyclic structure ($C_{47}H_{51}NO_{14}$) containing three ring systems and numerous functional groups including hydroxyl, benzoyl and acetyl groups, an oxetane ring and a side chain with additional hydroxyl and benzoyl functionality, attached at C-13 (Figure 2) (11). The biosynthetic pathway of paclitaxel has not been fully elucidated with only 15 of the 19 postulated enzymes characterised and the order of some transformations remains unclear.

A proposed route for paclitaxel biosynthetic synthesis and the enzymes involved are summarised in Figure 6 and Table 2 .

As a terpenoid paclitaxel is formed from the C5 precursors IPP and DMAPP that undergo three rounds of condensation to produce geranyl geranyl diphoshate (GGPP). Plants are unique in producing IPP and DMAPP *via* the cytosolic mevalonic pathway (MVA) and plastidial methylerythritol phosphate (MEP) pathway (Figure 15) (27), however the origin of the precursors involved in paclitaxel biosynthesis remains unclear (28). Studies have shown involvement of both routes and research in *T. baccata* observed that paclitaxel production can be reduced by inhibitors of both the cytosolic and plastid pathways (29–32).

Enzyme	Abbreviation	GenBank Ac No.	Ref
Geranylgeranyl diphosphate synthase	GGPPS	AF081514	(33)
Taxadiene synthase	TASY	AY364469	(34)
Taxadiene-5α-hydrolase	T5αH	AY289209	(35)
Taxadiene-5α-ol-*O*-acetyl transferase	TDAT	AF190130	(36)
Taxadiene-13α-hydrolase	T13αH	AY056019	(37)
Taxane-10β-hydrolase	T10βH	AF318211	(38)
Taxane-14β-hydrolase	T14βH	AY188177	(39)
Taxane-2α-hydrolase	T2αH	AY518383	(40)
Taxane-7β-hydrolase	T7βH	AY307951	(40)
Taxane-1β-hydrolase	T1βH	-	-
C4β,C20-epoxidase	-		
Oxonmutase	-		
Taxane-9α-hydroxylase	T9αH		
Taxane-2α-ol-*O*-benzoyl transferase	DBBT	AF297618	(41)
10-deacetylbaccatin III-10-*O*-acetyltransferase	DBAT	AF193765	(42)
Phenylalanine amminomutase	PAM	AY582743	(43)
β-phenylalanine-CoA Ligase	CoA-Ligase	KM593667	(44)
Baccatin III: 3-amino, 13-phenylpropanoyl transferase	BAPT	AY082804	(45)
Taxane 2'α-hydroxylase	T2'αH		
N-benzoyltransferase	DBTNBT	AF466397	(46)

Table 2 – Summary of postulated enzymes involved in paclitaxel biosynthesis, abbreviations, GenBank Accession Number (GenBank Ac No.) and the first published report. Table adapted from Onrubia *et al.* 2013. (47).

The first committing step of the pathway is the cyclization of GGPP to the tricyclic intermediate taxadiene, catalysed by taxadiene synthethase (TASY) in a "one step" cyclisation, forming three carbon–carbon bonds, three stereogenic centres and leading to the loss of a hydrogen atom (48). Taxadiene subsequently undergoes numerous regio- and stereospecific oxygenations and acylations mediated by a variety of cytochrome P450 (cP450) oxygenases and acyltransferases. Taxadiene is first hydroxylated at the C5-position that leads to the migration of the C4-C5 double bond to C4-C20, by cP450 taxadiene-5α-hydroxylase (T5αH) to produce taxadien-5α-ol.

Taxadien-5α-ol can then undergo two possible transformations to either produce taxadien-5α-yl-acetate, by taxadiene-5α-ol-*O*-acetyl transferase (TDAT), or taxadien-5α-13α-diol, by cP450 taxadiene-13α-hydroxylase (T13αH). The taxadien-5α-yl-acetate intermediate is then hydroxylated at C10 by the cP450-dependent monooxygenase taxoid-10β-hydroxylase (T10βH) (Figure 6). The

intermediate then undergoes four further hydroxylations at the C1, C2, C4 and C7 positions, an oxidation at C9 and an epoxidation between C4 and C5. The enzymes for these steps have not been fully characterised but the hydroxylations are mediated by cP450s. The enzyme involved in the production of the essential oxetane ring remains unknown, however a number of hypothetical mechanisms have been proposed (22). One of the preferred mechanisms involves the epoxidation of the C4-C20 double bond of the 5α-acetate intermediate, followed by a rearrangement reaction where the α-acetoxy group migrates from C5 to C4, accompanied by the oxirane-to-oxetane ring expansion (Figure 5) (49).

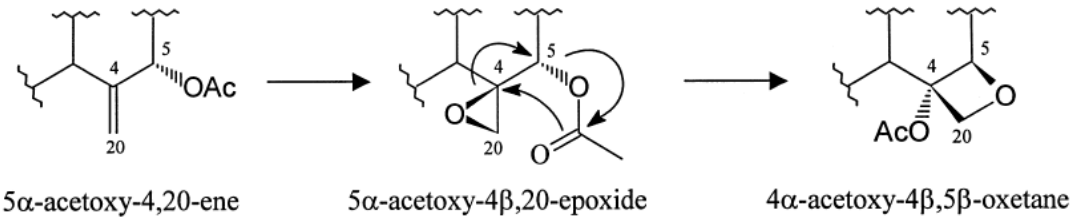


Figure 5 – A postulated mechanism for the formation of the oxetane ring. Image adapted from Walker *et al.* 2001) (50).

Four ester functionalities are also attached; a benzonate at C2, an acetate at C4 and C10 and an *N*-benzoyl-3-phenylisoserinoyl at C13. The order in which these transformations occur is still unknown and elucidating the sequence is more problematic than for the oxygenations as acylations lead to a large number of side products, many of which are known to have multiple acylation sites (51).

These transformations produce the hypothetical precursor 2-debenzoyltaxane for the enzyme taxane 2α-*O*-benzoyltransferse (DBBT) to convert into the intermediate 10-deacetylbaccatin III (10-DAB), which is subsequently acetylated at the C10 position by 10-deacetylbaccatin III-10-*O*-acetyltransferse (DBAT). The C13 side chain is formed from α-phenylalanine which is converted to β-phenylalanine by phenylalanine aminomutase (PAM) and subsequently esterified by a β-phenylalanineCoA ligase (CoA-Ligase). This is then attached to the taxane core by baccatin III-13-*O*-phenylpropanoyl transferase (BAPT) to produce 3'-*N*-debenzoyl-2'-deoxytaxol. This intermediate is hydroxylated at the C2 positon by an unknown cP450 and finally converted into paclitaxel by 3'-*N*-debenzoyl-2'-deoxytaxol-*N*-benzoyl transferase (DBTNBT) (Figure 6).

Taxus spp. are capable of producing a variety of different taxanes (52). The production of paclitaxel is likely to occur *via* a branched rather than linear pathway suggested in Figure 1, generating a wide variety of different taxanes. The identification of enzymes that are not involved in transformations to produce paclitaxel e.g. taxoid 14β-hydroxylase and the observation of multiple enzymes with the same substrate (i.e. T13αH and TDAT) supports this hypothesis (51). Work is currently on going to identify the missing enzymes in the pathway. As recently as April 2015 the Palazon Barandela group identified and characterised CoA-Ligase and suggested a number of candidates for the enzymes C1 and C9 hydroxylase, C4-C20 epoxidase, C9 oxidase and the oxomutase (44); however these enzymes still need to be isolated and characterised.

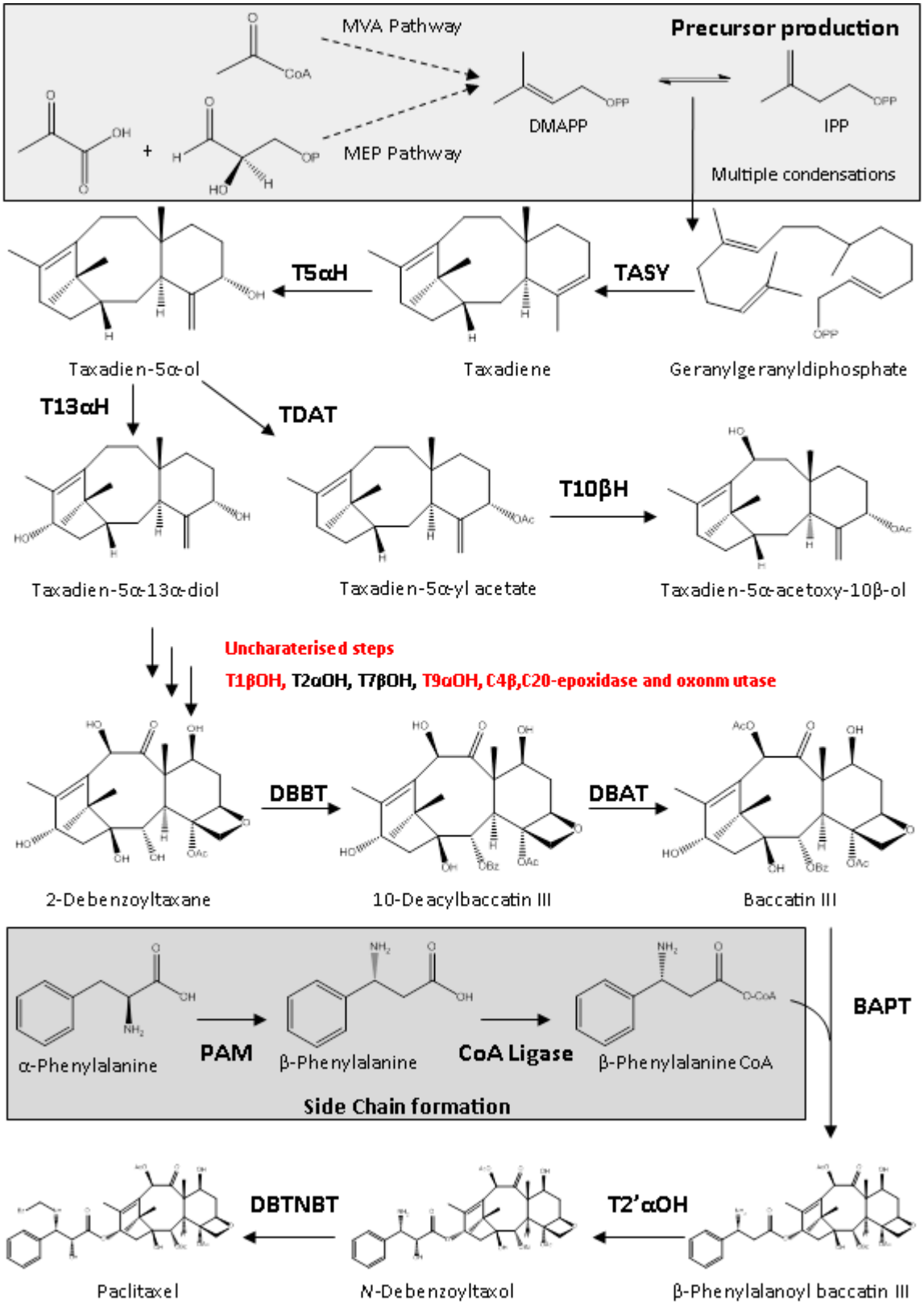


Figure 6 – A postulated paclitaxel biosynthetic pathway. Enzymes abbreviations are stated in Table 2. Terpenoid C5 precursors isoprenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are produced *via* two possible routes, the cytosolic mevalonic pathway (MVA) and plastidial methylerythritol phosphate (MEP) pathway which undergo multiple rounds of condensation to produce the intermediate geranylgeranyl diphosphate (GGPP) see Figure 15 and Figure 16 for further details. GGPP undergoes 19 transformations to produce the final product, paclitaxel. The enzymes involved in the middle of the pathway are yet to be characterised and are highlighted in red, with hypothesized enzymes stated. Image adapted from Howat *et al.* 2014 (26).

1.2.5 Production of Paclitaxel

The demand for paclitaxel is increasing as its therapeutic uses expand, however the drug exists at very low concentrations within the plant. One course of paclitaxel requires 2.5–3 g, which involves harvesting bark from eight mature (60 year old) yew trees (11). The process is destructive and production *via* natural harvest is also compounded by the slow growth of *Taxus spp.* and their low distribution in old-growth forests (49). Extraction of paclitaxel also requires complex purification techniques which are difficult and expensive. In 1992 the US Pacific Yew Act was signed into law by President George Bush Sr to ensure careful management during the collection of yew tree bark for medical purposes to ensure the long term survival of the species (53). Over 30 million yew trees were grown commercially in nursery plantations by Weyerhaeuser Company in collaboration with BMS in an attempt to limit the environmental impact of paclitaxel production (54). With the demand for paclitaxel increasing, production *via* natural harvest was not sustainable and other sources were explored (Table 3).

Year	Event	Reference
1988	First semi-synthetic route published	(55)
1990	Phyton Biotechnology Inc. founded	
1991	NCI enters into CRADA with Bristol-Myers Squibb (BMS)	
1991	First patent for PCC production of paclitaxel is issued, using fungal extracts as an elicitor	(56)
1992	Pacific Yew Act is signed into law by George Bush Sr	(53)
1994	FDA approves semi-synthetic route to paclitaxel	
1994	Total chemical syntheses published independently by two different groups	(57–61)
1995	Phyton Biotech licenses its PCC process to BMS for paclitaxel production	
1995	A semi-synthetic route using intermediates from isolation from *Taxus baccata* needles published	(62)
1995	Samyang Genex begins work on PCC process to produce paclitaxel, under the name Genexol ®	
1997	Paclitaxel produced *via* PCC is approved in South Korea	
2004	FDA approves PCC route to paclitaxel	
2007	Phyton Biotech patents elicitation strategies to increase titers to 900 mg/L	(63)
2010	*Taxus cuspidata* cell culture using cambial meristematic cells published	(64)

Table 3 – Key dates in the history of paclitaxel production with associated references. Paclitaxel can be produced by natural harvest, total and semi synthesis and plant cell culture. Overtime paclitaxel is produced commercially first by natural harvest, then semi-synthesis and subsequently plant cell culture. PCC, plant cell culture, NCI, National cancer institute, FDA, Federal Department of Agriculture, BMS, Bristol-Myers Squibb, CRADA, Cooperative research and development agreement. Table based on Leone *et al* 2013. (19)

1.2.5.1 Total chemical synthesis

The complex tricyclic core, oxetane ring and multiple chiral centres of paclitaxel make total chemical synthesis very difficult. Two total chemical syntheses were first published by two different groups in 1994, Holton *et al.* and Nicolaou *et al.* (57–59). The Nicolaou group proposed a convergent route by synthesizing two fragments containing the **A** and **C** rings and coupled them together using a Shapiro reaction and McMurry coupling to create the tricyclic skeleton. While many steps in the schema have high yields, the McMurry like cyclisation has yields of only 23%, which dramatically reduces the efficiency of the overall reaction scheme (59). Holton and co-workers meanwhile employed a linear scheme with overall yields of 4–5% (57, 58).

Both syntheses and subsequent schema are amazing achievements and in 2010 the Nobel Prize for chemistry was awarded for research using palladium catalysts that were used in the chemical synthesis of paclitaxel. However none of these syntheses are commercially viable because they involve complex multi-step pathways, some steps of which produce low yields and toxic by-products (26, 65).

1.2.5.2 Semi-synthesis

Taxus spp. produce a variety of taxanes, with paclitaxel being only a minor component. Compounds such as baccatin III and 10-DAB (Figure 6) can act as a building block in paclitaxel production, because they contain the tetracyclic core. Complex mixtures of taxanes can be converted (with yields of nearly 100%) into 10-DAB by the action of hydrolysing enzymes C-13 taxolase and C-10 deacetylase – isolated from *Nocardioides albus* and *Nocardioides luetus* respectively. 10-DAB can also be isolated from the needles of *T. baccata* in a non-destructive process (54), which only requires acetylation at the C-10 position and the addition of the C-13 side chain to produce paclitaxel.

The first reported semi-synthesis of paclitaxel was published by the Potier group, however it had low yields and was commercially unfeasible (55). Subsequent viable syntheses were developed by Holton and others employing a variety of side chains and reaction conditions (66). β-lactams can act as excellent precursors for side chains, which are *N*-acylated and subsequently reacted with baccatin III as shown in Figure 7. In 1990 BMS bought the licensing agreement for the semi-synthetic production of paclitaxel from Florida State University, which in 1997–1998 produced a revenue of $45 million (66, 67).

The BMS semisynthetic process is complex however, requiring 11 chemical transformations and seven isolations that has several disadvantages including: (1) production is dependent on epigenetic and environmental factors, (2) *Taxus* is a slow growing species, (3) intermediates require expensive purification and (4) the process requires 13 solvents and organic reagents that are hazardous and environmentally unfriendly (26).

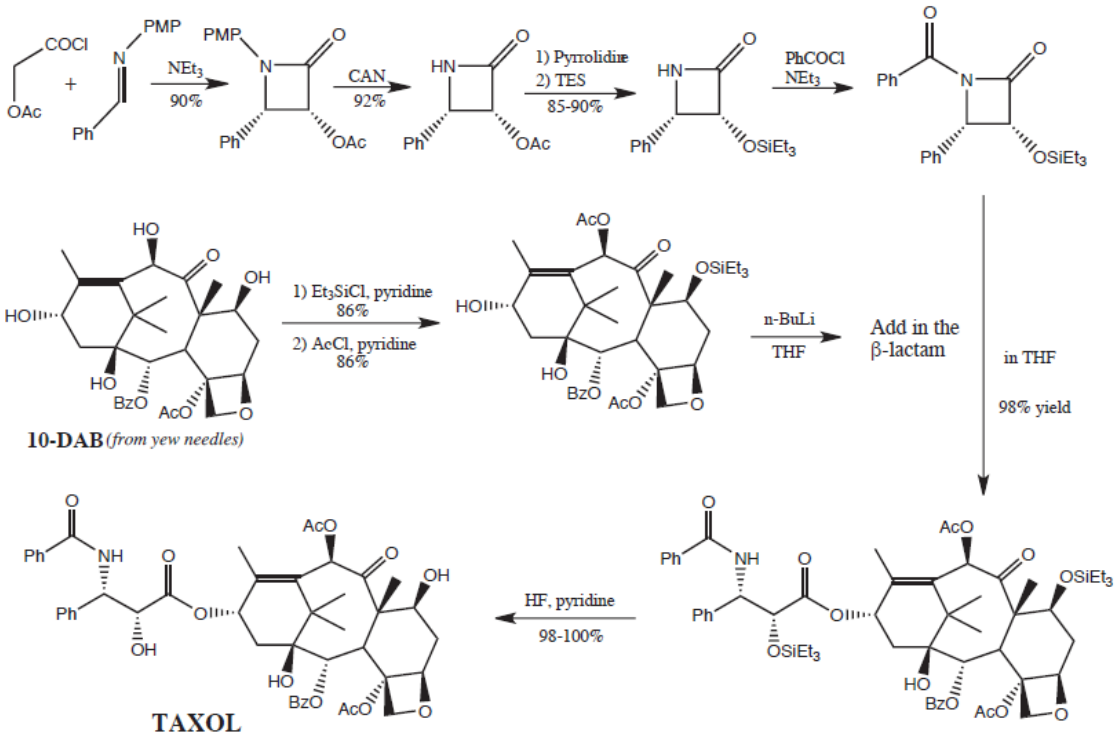


Figure 7 – An example of a semi-synthetic schema for the production of paclitaxel using 10-DAB (10-deacetylbaccatin III) extracted from *Taxus baccata* needles and using a β-lactam to produce the side chain.

1.2.5.3 Synthetic biology

This section is adapted from the review by Howat *et al.* (26). The production of PNP can be achieved in heterologous expression systems with the aid of synthetic biology. The artemisinin precursor artemisinic acid was successfully produced in *Saccharomyces cerevisiae* through the heterologous expression of three *Artemisia annua* enzymes, amorpha-4,11-diene synthase, a cP450 monooxygenase and its respective reductase, producing yields of 2.5 g/L (68). Other examples include: expressing isoprene synthase from poplar along with the MVA pathway from *S. cerevisiae* in *Escherichia coli*, which led to isoprene yields of 532 mg/L (69); and the expression of flavonoid biosynthesis genes –chalcone synthase, tyrosine ammonia-lyase, 4-coumaroyl CoA ligase and chalcone isomerase – in *E. coli* that led to the production of 84 mg/L naringenin, when the enzymes involved in the fatty acid biosynthesis were also suppressed (70, 71).

Attempts have been made to express some steps of the paclitaxel biosynthetic pathway in heterologous expression systems, including *E. coli*, *S. cerevisiae* and various plants species. The overexpression of IPP isomerase, GGPP synthase and TASY in *E. coli* led to taxadiene production levels of 1.3 mg/L in cell culture (72). However further development of the pathway is limited due to the difficulties of expressing cP450s in microbial systems, which include loss of functionality due to incorrect folding, translation and insertion into the cell membrane. Further problems with cofactor availability together with the absence of the specific cP450 reductases required to enzymatically recycle each cP450 represent significant obstacles (73). Five sequential paclitaxel biosynthetic genes has been expressed in *S. cerevisiae* with the intention of producing taxadien-5α-acetoxy-10β-ol, however only high (mg/L) levels of taxadiene were observed (74). The pathway was

restricted at the first cP450 hydroxylation, which may require the co-expression with its cognate cP450 reductase for efficient function (75).

A number of plant hosts have also been exploited to produce paclitaxel intermediates. GGPP synthase and TASY were overexpressed in *Arabidopsis thaliana* leading to the accumulation 600 ng/gram of dry weight of taxadiene but the introduction of these enzymes led to growth retardation (76). Tomato plants overexpressing TASY and lacking the ability to utilize GGPP for carotenoid synthesis produced 160 mg of highly pure taxadiene from 1 kg of freeze dried fruit (77), while overexpression of TASY in *Physcomitrella patens* produced 0.05% fresh weight of taxadiene with no growth inhibition (78). Collectively, these attempts to synthesise paclitaxel in heterologous hosts highlights the complexity of PNP biosynthetic pathways, which are frequently not fully elucidated, and the difficulty of expressing plant enzymes, especially cP450s, in microbial systems.

1.2.5.4 Plant cell culture

Plant cell culture (PCC) is seen as a viable alternative to the previously mentioned production routes. Advantages of PCC include: (1) production is independent of geographical and seasonal variations and can be more easily controlled compared to whole plant systems; (2) the system produces a continuous and uniform supply; (3) it is a renewable and environmentally friendly resource, (4) metabolic engineering can be employed to produce novel compounds and genetically modified cultures are isolated from the environment eliminating the risk of transgene migration (12, 79). Numerous PNPs are produced commercially using PCC including arbtin in *Catharanthus roseus*, berberine in *Coptis japonica*, scopolamine *Duboisia spp.*, ginseng in *Panag ginseng* and shikonin in *Lithospermum erythrorhizon* (80). Currently two companies – Phyton Biotech (USA) and Samyang Genex (South Korea) – produce paclitaxel on an industrial scale using PCC (26).

A plant suspension cell culture is initiated from an explant isolated from plant material, such as needles or an embryo. The explant is grown on solid medium and under the correct conditions it grows into a callus which is a mass of dedifferentiated cells. The callus is then transferred to liquid medium where it eventually results in a suspension culture (80). The first example of *Taxus* cell culture was published in 1991 using *T. brevifolia* and could produce 1–3 mg/L of paclitaxel (56). Since then a number of research groups have extensively investigated optimization of this process using a variety of different *Taxus spp.* and altering a number of biological and operational conditions summarised in Table 4 (15, 79).

Adsorbants and two-phase culture, for example, can be employed to improve paclitaxel production levels by shifting the intracellular/extracellular equilibrium, thereby reducing internal feedback inhibition and product degradation (11, 81). Adsorbants remove paclitaxel from the medium and are subsequently extracted using organic solvents, such as 10% dibutylphthalate, in two-phase culture. Adsorbents, such as Amberlite XAD, can increase paclitaxel production in *T. cuspidata* by 40–70% (82).

Strategy
Optimization of culture conditions - including media composition and two stage culture
Immobilization
In situ product removal by two-phase culture/adsorbants
Addition of Precursors
Selection of high yielding cell lines
Elicitation

Table 4 – A table summarising the strategies researched in attempts to increase paclitaxel production in plant cell culture. These include operational conditions such as optimizing the culture media and removing produced paclitaxel to prevent negative feedback inhibition and possible toxic side effects; and biological conditions for example treating the culture with elicitors such as methyl jasmonate.

1.2.5.4.1 Elicitation

Elicitation is a process that enhances the biosynthesis of secondary metabolites by the addition of low amounts of a compound that triggers extensive reprogramming in a cell; initiating or improving the expression of PNP related genes (83–85). This strategy has been employed to great effect using a variety of abiotic (e.g. vanadyl sulphate (86), lanthanum (87)) and biotic elicitors (e.g. fungal extracts (56, 88), salicylic acid (89), ethylene (90)) individually and in combination (91) to increase paclitaxel production levels in PCC; however using methyl jasmonate (MeJA) is seen as one of the most effective elicitors (47). Yukimune and coworkers were the first to employ MeJA as an elicitor in *T. media* and *T. baccata* cultures, increasing paclitaxel production from 28–110 mg/l and from 0.4–8 mg/l respectively (92). More recently Bentebibel *et al.* used MeJA to achieve paclitaxel production yields of 2.71mg/L/day in a stirred bioreactor with *T. baccata* cell culture (93). In recent years a new elicitor coronatine, a molecular mimic of isoleucine-jasmonic acid (a conjugated form of jasmonic acid), has been shown to act as more powerful elicitor than MeJA at increasing total taxane production in *T. media* cultures but MeJA is still the major elicitor employed to increase paclitaxel production in PCC (94).

1.2.5.4.2 Cambial Meristematic Cell suspension culture

There are a number of issues associated with traditional dedifferentiated cell (DDC) culture including slow growth rates, aggregation and low variable yields (26, 95). When PCC is initiated an explant proliferates into a callus of dedifferentiated cells, forming a heterologous mixture of cells in liquid culture. Evidence suggests that this process is not a simple reversal of cell programming to a stem cell like state and affects cells at a genetic level, leading to variable growth rates and inconsistent yields (96).

Variation is also introduced by aggregation, which creates distinct subpopulations within the culture producing microenvironments with different nutrient and gas compositions that leads to differences in cell morphology and metabolism (12). Aggregation occurs within the culture due to the natural method of plant cell division. Daughter cells remain connected to their parent cell through the cell wall and this can lead to large aggregates when cultures are grown on an industrial scale (80). The growth rate of PCC is also lower compared to mammalian or microbial cultures

and due to their large size are more susceptible to shear stresses especially in large bioreactors. Some of the problems associated with traditional DDC culture have been circumvented by the initiation of PCC from undifferentiated cambial meristematic stem cells (CMCs) completed by Lee *et al.* (64). CMCs are obtained from the cambium tissue and are multipotent plant cells that give rise to the vascular xylem and phloem tissues (Figure 8a) (95). Their stem-cell-like properties were authenticated through their ability to differentiate into tracheary elements, high sensitivity to genetic damage by γ-irradiation and high expression of procambium marker genes WOODEN LEG (WOL) and PHLOEM INTERCALATED WITH XYLEM (PXY).

CMCs produce a homogenous culture with significantly faster growth rates, smaller aggregate size and lower sensitivity to shear stress in stirred tank and air lift bioreactors (Figure 8c-d). CMCs can produce higher levels of paclitaxel compared to DDCs with or without MeJA elicitation. In a 3 litre air-lift bioreactor CMC produced 98 mg/kg fresh cell weight compared to only 11 mg/kg and 13 mg/kg for DDC derived from needles and embryos respectively. CMCs also secrete a higher percentage of paclitaxel into the cell medium reducing the need for downstream processing and reducing the possible toxic effects, such as impaired cell growth and division (Figure 8e) (64, 97).

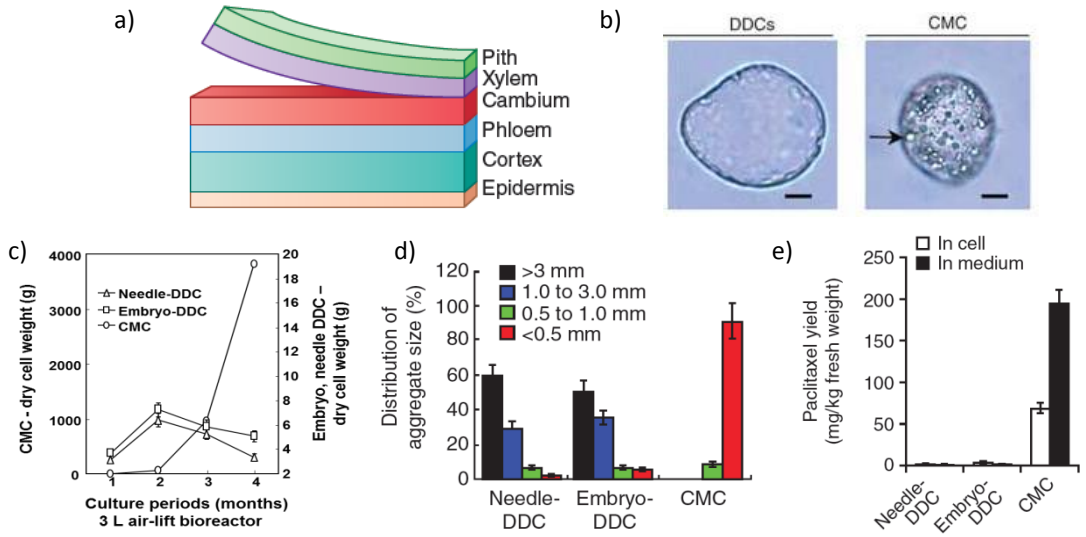


Figure 8 – The advantages of cambial meristematic cell (CMC) culture compared to traditional dedifferentiated cell (DDC) cultures initiated from needles or embryos. a) A schematic showing the cross section of a twig illustrating the location of the cambial layer (image adapted from Roberts and Kolewe 2010 (95)). b) Micrograph of a DDC and a CMC, CMCs are smaller than DDCs and possess numerous, small vacuole-like structures indicated by a black arrow. Scale bars, 20 µm. c) Growth of CMC and DDC suspension cultures in a 3 litre air-lift bioreactor as dry cell weight. CMCs grow to significantly higher weights compared to DDC culture. d) Bar graph showing the difference in cell aggregation in DDC initiated from needles or embryos and CMC suspension cultures. CMC have significantly smaller aggregate sizes compared to DDCs. e) The intracellular and extracellular paclitaxel yield of DDCs and CMCs in a 3 litre air-lift bioreactor. The majority of the paclitaxel produced by CMCs is secreted into the medium. Error bars represent 95% confidence limits. Image adapted from results published in Lee *et al.* 2010 (64).

CMCs have been isolated from a number of different species and are currently employed industrially by Unwha to produce ginseng from *Panax ginseng* and raw materials for nutrition and cosmetic products from *T. cuspidata*, *Ginkgo biloba* and *Solanum lycopersicum* (12). Collectively, the advantages of CMCs make them an attractive platform from which to further increase paclitaxel

production levels.

1.3 Jasmonate signalling

Jasmonates are oxylipin phytohormones, consisting of jasmonic acid (JA) and its derivatives, which are involved in regulating a variety of plant stress and developmental responses (98, 99). Originally designated as secondary metabolites identified in the floral scent of *Jasminum spp*. it later became clear that these compounds act as elicitors and are ubiquitous across the plant kingdom (100). Jasmonates have been found to play important roles in a range of developmental processes including root growth, senescence, seed germination, flower development and trichrome formation; and act as signals for biotic (herbivore/pathogen attack) and abiotic (osmotic/wounding) stresses (101, 102). JA can also act as a systemic signal immunizing the plant against subsequent attack (99), and triggers extensive transcriptional reprogramming leading to the concerted activation of a plethora of secondary metabolites from all three PNP classes (i.e. terpenoids, alkaloids and phenylpropanoids) (85, 100).

1.3.1 Jasmonic acid biosynthesis

The production of JA by 7 different plant species was first published in 1984 by Vick and Zimmermann and subsequent research has elucidated the enzymes involved in the transformations (Figure 9) (101, 103). JA originates from α-linolenic acid (18:3) released from galactolipids in the chloroplast membrane by the action of a phospholipase 1 however the specific enzyme has yet to be determined with the reported involvement of Defective in anther dehiscence 1 (DAD1) and Dongle (DGL) (99). The released α-linolenic acid is then oxygenated at the C-13 position by a 13-lipoxygenase (LOX) to produce 13(*S*)-hydroperoxy-octadecatrienoic acid (13(*S*)-HPOT), which is the substrate for the first committing step in the JA biosynthetic pathway allene oxide synthase (AOS) to create 12,13(*S*)-epoxy-octadecatrienoic acid (12,13(*S*)-EOT) (98). The cyclisation of this allylic epoxide by allene oxide cyclase (AOC) produces (9*S*,13*S*)-12-oxo-phytodienoic acid (cis-(+)-OPDA) and is the final step to occur in the plastid, with subsequent transformations located in the peroxisome (101). OPDA is reduced by OPDA reductase 3 (OPR3) yielding 3-oxo-2-(2-pentenyl)-cyclopentane-1-octanoic acid (OPC-8), which is then activated by the addition of a CoA moiety to undergo three rounds of β-oxidation performed by acyl-CoA oxidase (ACX), multifunctional protein (MFP), and L-3-ketoacyl-CoA thiolase (KAT). These transformations shorten the chain and produce jasmonoyl-CoA which is cleaved by an as yet unknown thioesterase (TE) to produce (+)-7-iso-JA which epimerises to the more stable form of (-)-JA (98). JA can be subsequently metabolised into a number of different conjugates including the signalling molecule MeJA and the most biologically active amino acid conjugate jasmonoyl-L-iso-leucine (JA-Ile), by the enzymes JA methyl transferase (JMT) and JA-amino acid synthetase (JAR1) respectively (99).

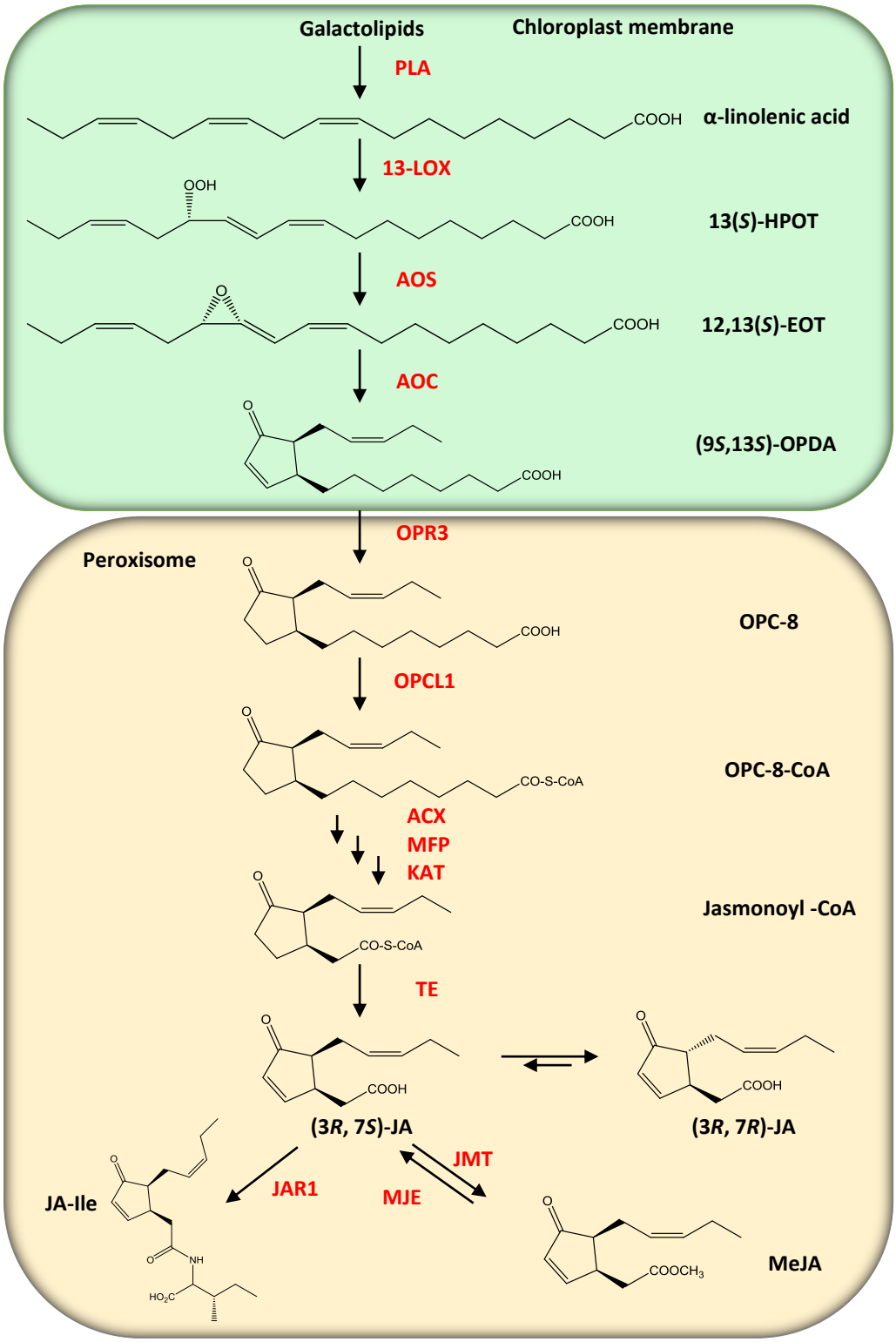


Figure 9 – The jasmonic acid biosynthetic pathway. α-Linolenic acid is cleaved from galactolipids in the chloroplast membrane and is then transformed by the action of three enzymes 13-lipoxygenase (13-LOX), allene oxide synthase (AOS) and allene oxide cyclase (AOC) to produce 12-oxo-phytodienoic acid (OPDA). OPDA is transferred to the peroxisome and undergoes six transformations to produce jasmonic acid ((3*R*, 7*S*)-JA). This can epimerises to the more stable form ((3*R*, 7*R*)-JA) or be metabolised to the conjugate forms methyl jasmonate (MeJA) and jasmonoyl-L-iso-leucine (JA-Ile). Enzyme are shown in red and abbreviations are phospholipases (PLA), OPDA reductase 3 (OPR3), OPC-8:CoA ligase 1 (OPCL1), acyl-CoA oxidase (ACX), multifunctional protein (MFP), and L-3-ketoacyl-CoA thiolase (KAT), thioesterase (TE), JA methyl transferase (JMT) and JA-amino acid synthetase (JAR1), methyl-jasmonate esterase (MJE). Substrates abbreviations: 13(*S*)-hydroperoxy-octadecatrienoic acid (13(*S*)-HPOT), 12,13(*S*)-epoxy-octadecatrienoic acid (12,13(*S*)-EOT), (9*S*,13*S*)-12-oxo-phytodienoic acid ((9*S*,13*S*)-OPDA), 3-oxo-2-(2-pentenyl)-cyclopentane-1-octanoic

acid (OPC-8).

1.3.2 JA perception and signalling

A central regulator of plant hormone signalling is the ubiquitin-protease system and JA signalling is no exception (99). A core signalling module is thought to be conserved across the plant kingdom and its components have been identified in numerous medicinal plants including *C. roseus*, *Nicotiana tabacum* and *A. annua* (100). The F box protein CORONATINE INSENSITIVE1 (COI1) is an essential part of the Skp–Cullin–F-box-type E3 ubiquitin ligase complex (SCF^{COI1}) whose function is to provide substrate specificity, targeting proteins for degradation by the 26S proteasome. One of the targets of COI1 are JA ZIM domain (JAZ) repressor proteins which interact with JA responsive TFs, such as MYC2, inhibiting their function (Figure 10) (104).

JAZ proteins contain a highly conserved C-terminal jas motif that mediates its interaction with a variety of transcription factors (TFs) in *A. thaliana* including the bHLH TFs MYC2, MYC3, MYC4, GLABRA 3 (GL3), ENHANCER OF GLABRA 3 (EGL3) and TRANSPARENT TESTA 8 (TT8); the R2R3 MYB TFs PRODUCTION OF ANTHOCYANIN PIGMENT 1 (PAP1), GLABRA 1 (GL1), MYB 21 and MYB 24; and TFs from other hormone signalling pathways ETHYLENE INSENSITIVE3 (EIN3), EIN3-LIKE1 (EIL1), GIBBERELLIC ACID INSENSITIVE (GAI), REPRESSOR OF GA (RGA), and RGA-LIKE1 (RGL1) (99, 105). JAZ proteins also contain a ZIM domain with a highly conserved TIFY motif which mediates JAZ protein dimerization and interaction with the adapter protein NOVEL INTERACTOR OF JAZ (NINJA); the N-terminal domain of which contains an ethylene responsive factor–associated amphiphilic repression (EAR) motif by which it can interact with the co-repressors TOPLESS (TPL) and TPL-related proteins (TPR) (105, 106).

In a resting state the JAZ protein complex interacts with TFs such as MYC2 through its JAZ interaction domain (JID). MYC2 resides on the promoters of JA-responsive genes and JAZ proteins suppress their ability to activate downstream processes due to chromatin modifications performed by histone deacetylases (HDAs) (99). During a stress response JA-Ile is rapidly produced by JAR1 which binds with COI1 activating it to interact with JAZ proteins in a complex that requires inositol pentakisphosphate (104). JAZ proteins are subsequently ubiquitinated and degraded by the 26S proteasome, releasing MYC2 to activate its downstream targets (Figure 10) (107, 108).

The JA signalling pathway has similarities with the auxin pathway as they both employ a SCF-ubiquitin ligase complex to degrade repressors and COI1 has homology with the auxin receptor TRANSPORT INHIBITOR RESPONSE PROTEIN 1 (TIR1), which also targets AUXIN/INDOLE-3-ACETIC ACID (AUX/IAA) repressors that bind to transcriptional activators Auxin Response Factors (ARFs) (98, 109). Crosstalk can occur between JA and other hormone signalling pathways, for example TPL acts as a corepressor in the auxin and ABA pathways interacting with IAA12 and ABA insensitive 5 binding proteins (AFP) proteins (106, 110). JAZ proteins can interact with the gibberellic acid (GA) DELLA repressors GAI, RGA and RGL1 and the positive regulators of ethylene response EIN3 and EIL1 (105, 111).

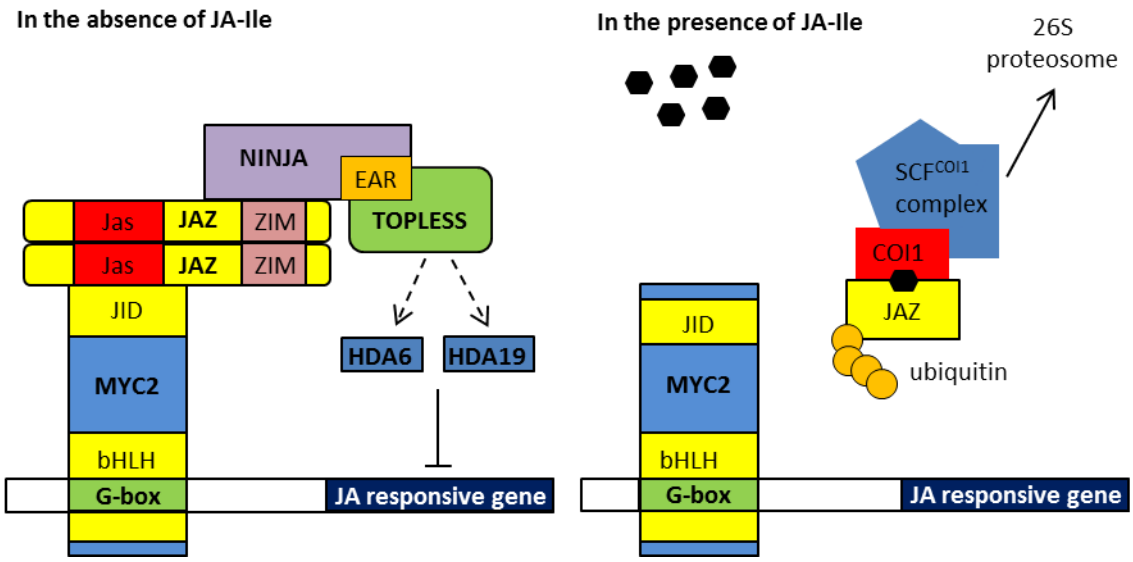


Figure 10 – Regulation of jasmonic acid (JA) responsive gene expression. In the absence of the biologically active amino acid conjugate form of JA – jasmonoyl-L-iso-leucine (JA-Ile) – MYC2, which is bound via a G-box to the promoter of a JA-responsive gene, does not activate transcription due to binding through a JAZ-interacting domain (JID) domain to the Jas domain of the JA ZIM domain (JAZ) repressor proteins. In concert with co-repressors NOVEL INTERACTOR OF JAZ (NINJA) and TOPLESS (TPL) JAZ proteins repress transcription via histone deacetylases HDA6 and HDA19. The ZIM domain of JAZ interacts with NINJA, while an EAR domain in NINJA facilitates its interaction with TPL. Upon stimulation by high levels of JA-Ile (black hexagons) JAZ proteins are recruited by CORONATINE INSENSITIVE 1 (COI1), part of the Skp–Cullin–F-box-type E3 ubiquitin ligase complex (SCF^{COI1}), which ubiquitinates JAZ targeting it for degradation by the 26S proteasome. MYC2 is therefore free to activate transcription of early JA-responsive genes.

1.3.3 JA inducible gene expression

Even though a central core JA signalling module has been hypothesized across plant species there is a significant degree of species specificity in the metabolic pathways that are altered after JA elicitation (100). The extensive reprogramming of gene expression after JA elicitation and subsequent increase in secondary metabolite production has been well documented in numerous species, including cell suspension cultures of *N. tabacum* (112), *C. roseus* (113) and *A. thaliana* (114), where it led to the increased production of phenylpropanoid conjugates & nicotine alkaloids, terpenoid indole alkaloids (TIAs) and monolignols respectively. The genes involved in each of these distinct pathways were tightly co-regulated and the induction in their gene expression occurred quickly, within 1–4 hours (h) (85). Transcriptional cascades control genetic reprogramming, for example in *A. thaliana* cell suspension culture after MeJA elicitation the number of genes with altered expression increased between 0.5 and 6 h from 75 to 495 (114). The expression of metabolic pathways is also temporally regulated as observed in *N. tabacum* where transcripts involved in nicotine biosynthetic genes were up-regulated at 1–2 h after MeJA elicitation, while genes involved in phenylpropanoid metabolism were most highly expressed between 2–6 h (112).

1.4 Transcription factors

Transcription factors (TFs) play a vital role in ensuring tight coordination of gene expression and a number of families have been implicated in the MeJA mediated regulation of secondary

metabolism including the AP2 (APETALA2), MYB (MYELOBLASTOSIS), NAC (NAM/ATAF1,2/CUC2), bHLH (BASIC HELIX-LOOP-HELIX) and WRKY families.

1.4.1 AP2 Transcription Factor Family

The AP2/ERF (APETALA2/Ethylene responsive factor) superfamily includes four major subfamilies ERF, DREB (dehydration-responsive element-binding protein), AP2 and RAV (Related to ABI3/VP1) which all contain at least one conserved 60 amino acid DNA binding domain (115). The first reports of ERF TFs implicated them in ethylene signalling, hence the acronym, however this not a universal feature of all proteins in this class. The ERF and DREB sub-families contain only one AP2 binding domain, while the RAV sub-family has an additional B3 DNA binding domain and the AP2 subfamily can contain 1 or 2 AP2 domains (116). The AP2 domain has a 3D structure similar to zinc fingers, containing three antiparallel β-sheets and an approximately parallel α-helix. However, in contrast to zinc fingers the β-sheet is important in DNA binding containing arginine and tryptophan residues that interact with the major groove of the DNA (117, 118). Most plant TFs activate target genes *via* acidic regions located C and N terminally and recently a strong activation motif EDLL has been identified in AtERF98/TDR1 (TRANSCRIPTIONAL REGULATOR OF DEFENCE RESPONSE 1) (119). AP2 repression can occur *via* two repression domains, the EAR motif ($^{L}/_{F}$DLN$^{L}/_{F}$XP) (120) and the B3 repression domain (RLFGV), and can involve the co-repressors TPL and TPR (116). AP2 TFs have been reported to recognise a number of different cis elements summarised in Table 5, but most include a GC rich region such as a GCC box.

1.4.2 MYB Transcription Factor Family

MYB TFs are named after the first published gene containing the MYB domain in the avian myeloblastosis virus and have subsequently been identified in all eukaryotic organisms (121, 122). Compared to animals the MYB superfamily has dramatically expanded in plants with 10% of all TFs identified in the *A. thaliana* genome members of this family that are able to regulate a range of metabolic, cellular and developmental processes (123). The first identified plant TF was a MYB gene, COLORED1 (C1), in *Zea mays* which controls anthocyanin pigmentation of seed kernals (124). MYB proteins are characterised by a highly conserved DNA binding domain that consists of 1–4 imperfect amino acid repeats (125). Each repeat is approximately 52 amino acids in length and encodes three α-helixes, the second and third of which form a helix-turn-helix structure with three spaced tryptophan residues that form a hydrophobic core that intercalates with the major groove of the DNA (126, 127).

Family	Binding motif		Species	TF	Ref.
AP2	GCC- Box	TAAGAGCCGCC	*Nicotiana tabacum*	NtERF1-4	(128)
		AGCCGCC	*Arabidopsis thaliana*	AtERF1-5	(129)
		GCCGCC	*Solanum lycopersicum*	Pti4/5/6	(130)
		Two GCC boxes required GCCGCC and GCAGCCGCT	*Arabidopsis thaliana*	ORA59	(131)
	DRE element	A/GCCGAC	*Arabidopsis thaliana*	DREB1A DREB2A	(132)
	JERE motif	CTCTTAGACCGCCTTCTT	*Catharanthus roseus*	ORCA3	(133)
	RAV bipartite motif	CAACA-(N)n –CACCTG n is predominantly 5	*Arabidopsis thaliana*	RAV1	(134)
MYB	CNGTT(A/G) consensus	CNGTT(A/G)G	*Arabidopsis thaliana*	*At*MYB66 (WER)	(135)
	ACC(A/T)A(A/C) consensus	Recognises three forms ACCTACC ACCAACC ACCTAAC	*Pinus taeda*	PtMYB1/4	(136, 137)
		ACC(A/T)ACC(A/C/T)	*Zea Mays*	P	(138)
	TTAGGG motif	$(TTTAGGG)_2$	*Arabidopsis thaliana*	AtTBP1	(139)
	AAAATATCT motif	AAAATATCT	*Arabidopsis thaliana*	LHY	(140)
	GATA motif	GGATA	*Solanum tuberosum*	MYBSt1	(141)
NAC	CGT[G/A] motif	TTNCGTA	*Arabidopsis thaliana*	ANAC092	(142)
		TTGCGTGT	*Arabidopsis thaliana*	ANAC019	(142)
	CACG	ANNNNNTCNNNNNNNACA		ANAC019	(143)
		CGCATGT	*Arabidopsis thaliana*	ANAC055 ANAC072	
bHLH	E-box - CANNTG	CACGTG – G box	*Zea Mays*	R	(144)
		CACGTTG G-box like	*Nicotiana tabacum*	NtMYC2a NtMYC2b	(145)
WRKY	W- box TTGAC(C/T)	TTGACC	*Arabidopsis thaliana*	WRKY26	(146)
		CAAAAATTTGACTATACTTGACTATTAGTG	*Catharanthus roseus*	CrWRKY1	(147)

Table 5 – A table summarising the known binding sites of the AP2 (APETALA2), MYB (MYELOBLASTOSIS), NAC (NAM/ ATAF1,2/CUC2), bHLH (BASIC HELIX LOOP HELIX) and WRKY families. The table shows for each TF family the name of the motif or the consensus sequence with reported examples, stating the TF, species and specific nucleotide binding sequence with associated references. AP2 and MYB TFs have multiple known binding sites, whereas NAC bHLH and WRKY TFs recognise one general motif.

MYB TFs are sub-divided into four classes (R2R3, 1R, 3R (R1+R2+R3) and 4R ($R1R2_2R1/2$)) based on the number of adjacent repeats they contain (1-4), named R1/R2/R3 after the repeats in prototypic MYB protein c-MYB (125). All four classes have been identified in plants, however, the largest sub-family is R2R3-MYB. R2R3-MYB TFs have a modular structure comprising an N-terminal DNA binding domain and a variable C-terminal region that can activate or repress gene expression and can contain conserved serine and threonine residues that can be post-translationally modified (122).

MYB TFs have an inherent flexibility in the DNA binding sites that they can recognise (as can be seen in Table 5). Many R2R3-MYBs recognise AC rich elements residing approximately 500 bp upstream of the transcription initiation site and some MYB TFs are capable of recognising multiple binding sites for example, the *Petunia hyrida* MYB.Ph3 can bind to two sites MBSI A(A/G/T)(A/G/T)C(C/G)GTTA and MBSII AGTTAGTTA (123, 148).

1.4.3 NAC Transcription Factor Family

NAC TFs are plant specific and were initially implicated in plant development, however their function has expanded to include roles in senescence and stress responses, including drought and high salinity, and responding to the hormones ABA and JA (149). The NAC DNA binding domain is based on a consensus from NAM, ATAF1/2 and CUC2 and is located at the N-terminus (150). The structure of the NAC domain was determined *via* X-ray crystallography of ANAC019 and is different from the normal helix-turn-helix configuration of other TFs. The novel fold consists of a very twisted antiparallel β-sheet packed against two α-helixes (151). NAC proteins have the ability to form hetero and homo dimers through the NAC domain (152). ANAC019 preferentially forms dimers in solution and these dimers were observed when the protein was crystalized (150, 151). EMSA studies have shown that truncating the NAC domain prevents dimer formation and subsequent DNA binding (142).

NAC proteins can function as activators or repressors of gene expression. The *A. thaliana* TFs ATAF1/2 can activate the CaMV 35S promoter in yeast cells (152) and ANAC019, ANAC055 and ANAC072 are able to transactivate the *erd1* promoter (143); while a 35 amino acid active NAC repression (NARD) domain has been identified in the *Glycine max* TF GmNAC20 (153). NACs have two known core motifs CGT[A/G] and CACG but the surrounding nucleotides can differ (Table 5). For example, cyclic amplification and selection of targets (CASTing) and electromobility shift assay (EMSA) studies of ANAC091 and ANAC092 identified their consensus binding motifs as [TA][TG]N**CGT[GA]** and T[TAG][GA]**CGT[GA]**[GCA][TAG] respectively (142), while *in vivo* experiments with ANAC019, ANAC055 and ANAC072 determined their NAC recognition sequence (NACRS) as $TCN_7ACACGCATGT$ (143).

1.4.4 bHLH Transcription Factor Family

bHLH TFs act as transcriptional regulators in all eukaryotes and have a diverse range of functions including: phytochrome signalling, root development, anthocyanin and flavonoid biosynthesis, hormone signalling and stress responses (154). The highly conserved bHLH domain located at the N-terminus and is comprised of around 60 amino acids which form two regions. First is a region 13–17 amino acids in length that is mostly basic that binds to the E-box DNA motif CANNTG and can contain the highly conserved HER motif (His5-Glu9-Arg13) (155). This is followed by a helix-loop-helix region which is comprised of mostly hydrophobic residues that form two amphipathic α-helixes separated by a loop of variable length. A highly conserved leucine residue within these helixes is required for bHLH proteins to form homo and hetero dimers (122, 156). Sequence conservation outside of the bHLH domain is low but several conserved motifs have been identified such as the leucine zipper domain and the ACT (aspartate kinase, chorismate mutase and TyrA) fold. This region can also bind to other proteins such as the JID domain in MYC2 that binds to JAZ proteins (99). bHLH proteins can also work in combination with MYB TFs and WD-40 proteins to regulate gene expression see section 4.2 for further discussion.

bHLH proteins have a well conserved E-box DNA binding motif of CANNTG with much less variation in DNA sequence compared to the other TFs previously mentioned. A version of this motif, CACGTG, known as a G-box is the binding site of the *Zea mays* R protein (144) however, the MYC2 TFs of *N. tabacum* do not bind to the canonical CANNTG binding site but to a G-box like site (Table 5) (145).

1.4.5 WRKY Transcription Factor Family

Although only first discovered in the mid-1990s WRKY TFs have been implicated in a number of abiotic and biotic stresses, developmental responses, regulation of biosynthetic pathways and in hormone signalling (157). WRKY proteins belong to the WRKY-GCM1 superfamily of zinc finger TFs and contain a WRKY DNA binding domain approximately 60 amino acids that nearly always contain the invariant sequence WRKYGQK and a $CX_{4-5}CX_{22-23}HXH$ zinger finger motif (158). An NMR and crystal structure of the WRKY domain in AtWRKY5 and AtWRKY1 respectively have shown it consists of a 4–5 stranded β-sheet in which is located a novel zinc binding pocket formed by the conserved cysteine and histidine residues in the zinger finger motif. The DNA binding region (WRKYGQK) is located N-terminally on a β strand that protrudes out of the protein due to a kink produced by a glycine residue, allowing it to interact with the major groove of the DNA (159, 160).

WRKY TFs bind to the well conserved W box ($TTGAC^{T}/_{C}$) (158), however a detailed study by Ciolkowski *et al.* of five *A. thaliana* WRKY proteins showed that the adjacent bases and the surrounding sequence are important in determining DNA binding affinity. For example if the base adjacent to the 5' end of the W-box is a G residue only AtWRKY6 and AtWRKY11 bound whereas if the residue was a T, C or A AtWRKY26/38/43 bound (146). A large number of W-boxes have been identified in stress inducible promoters therefore the ability of WRKY TFs to discriminate between them provides protein specific regulatory activity. WRKY proteins capable of activating and repressing transcription and the regulatory can be promoter dependent, such as AtWRKY53 (158).

1.4.6 JA induced TFs involved in secondary metabolism

JA elicitation triggers extensive genetic reprogramming and TFs play a major role in regulating these changes in metabolism. A large number of JA inducible TFs work in concert to control secondary metabolism (see Table 6) and have been extensively reviewed by Geyter *et al.* and Afrin *et al.* (100, 161).

Species	Metabolite	Transcription Factor	Reference
Arabidopsis thaliana	Indole glucosinolates, flavonoids & anthocyanins	MYC2	(162)
	Anthocyanins	PAP1	(163)
		GL3	
		EGL3	
		TT8	
	Aliphatic glucosinolates	MYB29	(164)
	Camalexin	WRKY33	(165)
		ANAC042	(166)
Nicotiana spp.	Nicotine	NtMYC2	(145)
		NtMYBJS1	(167)
		NbbHLH1-2	(168)
		NtERF221/ORC1	(169, 170)
		ERF189	(169)
	Phenylpropanoids	NtMYB8	(171)
	Volatile terpenes	WRKY3	(172)
		WRKY6	
Catharanthus roseus	Terpenoid indole alkaloids	CrMYC1	(173)
		CrMYC2	(174)
		ORCA2	(175)
		ORCA3	(176)
		CrZCT1-3	(177)
		CrWRKY1	(147)
Gossypium arboretum	Gossypol	GaWRKY1	(178)
Artemisia annua	Artemisinin	AaERF1-2	(179)
		AaWRKY1	(180)
Pinus Taeda	Flavonoids & isoprenoids	MYB14	(181, 182)
Taxus spp.	Paclitaxel	TcWRKY1	(183)
		TcJAMYC1	(184)
		TcJAMYC2	
		TcJAMYC4	

Table 6 – A table summarising the jasmonic acid (JA) inducible transcription factors (TF) involved in regulating secondary metabolite production from the families AP2, MYB, NAC, bHLH and WRKY. The examples of TFs are broken down based on species, with the secondary metabolite pathway regulated and the associated reference. In *Arabidopsis thaliana* and *Nicotiana spp.* numerous natural product pathways are affected by JA elicitation and these are controlled by different sets of transcription factors. Table adapted from Geyter *et al.* (100) and Afrin *et al.* (161).

There are numerous examples of JA inducible TFs that regulate secondary metabolism in the AP2, MYB, bHLH and WRKY families but only one from the NAC family. Research into NAC TFs is currently limited and other NAC TFs are JA inducible, such as ANAC019, ANAC055 and ANAC072, but as yet they have not been directly shown to act as regulators of secondary metabolism (149). Important AP2 regulators include the ORCA (Octadecanoid-derivative Responsive Catharanthus AP2-domain) proteins from *C. roseus* which regulate the TIA pathway. ORCA3 can regulate five TIA biosynthetic genes but not activate the entire pathway as the expression of geraniol 10-hydroxylase (G10H) and acetyl-CoA: 4-O-deacetylvindoline 4-O-acetyltransferase (Dat) were not induced in the ORCA3-overexpression lines (176). WRKY proteins have been implicated in the regulation of a number of secondary metabolite pathways for example; AaWRKY1 can activate the expression of the artemisinin gene amopha-4,11-diene synthase (180), CrWRKY1 overexpressed in hairy root cells can increase the expression of the TIA biosynthesis gene tryptophan decarboxylase by 7–9 fold (147) and TcWRKY1 has been shown to activate DBAT expression (183).

MYB TFs can work individually or in combination with bHLH TFs to control transcript levels (see section 4.2). Examples of MYB TFs controlling secondary metabolism indivdually include MYBJS1 in *N. tabacum*, which when overexpressed induces the entire phenylpropanoid biosynthetic pathway (167) and *Pt*MYB14 that has been suggested to act as a putative regulator of flavonoid and isoprenoid biosynthesis in pine (182). The best known bHLH regulator is MYC2 that acts as a positive regulator of secondary metabolism in numerous species including: *At*MYC2 that positive regulates flavonoid biosynthesis (162); the *N. benthamiana* homologs *Nt*bHLH1 and *Nt*bHLH1 which activate nicotine biosynthesis (168); and the *C. roseus* homologs *Cr*MYC1 and *Cr*MYC2 which regulate TIA biosynthesis (173, 174).

1.4.7 Regulation of TF activity

The activity of TFs can be regulated in a number of ways including combinatorial control and post-translational modification.

1.4.7.1 Combinatorial regulation

TFs do not work in isolation but within networks to produce a co-ordinated stress response. MYC2 is known to activate expression of AP2 TFs ORCA2 and ORCA3 in *C. roseus* (174) and ORC1 in *N. tabacum* (185), thereby indirectly regulating alkaloid production in both species. The interaction of JAZ proteins with numerous TF including MYC2 is a good example of combinatorial control (see section 1.3.2). bHLH and MYB TFs are well known to cooperate to activate flavonoid biosynthesis in various species and this is discussed further in section 4.2 – Combinatorial action of MYB and bHLH.

1.4.7.2 Post-translational modification

The activity of TFs can be controlled at the post-translational level by modifications including but not limited to phosphorylation, acetylation, nitrosylation, disulphide bond formation and ubiquitination (186). These modifications can change the structural conformation of the protein, altering its activity or affecting its stability. Phosphorylation is a well characterised mechanism for regulating TF activity, for example phosphorylation of a serine in the C-terminal region of PtMYB4 by mitogen-activated protein kinase (MAPK) 6 positively regulates its activity (125). In *A. thaliana* 15 AP2 TFs have been shown to act as substrates for MAPKs (116) and the binding ability of the tomato AP2 TF Pti4 increases after phosphorylation (187). The ability of ORC1 and MYC-like bHLH1 to activate alkaloid biosynthesis in *N. tabacum* is increased by the JA inducible MAPK kinase (JAM1) (185). WRKY proteins are also known to be controlled by phosphorylation with MPK3 and MPK6 in *A. thaliana* regulating the expression and activity of WRKY33 (188).

TFs can also be controlled by redox regulation. In maize the P1 MYB TF has two cysteine residues (Cys-49 and Cys53) which under oxidising conditions produce an intramolecular disulphide bind, preventing DNA binding. If Cys49 is mutated to a serine or and alanine the TF is

able to bind DNA irrespective of the redox environment (189). Different post-translational controls can have opposing effects on the TF regulatory activity, for example the activity of SPECIFICITY PROTEIN 3 (Sp3) in *A. thaliana* is up-regulated by acetylation but down-regulated by sumoylation (190).

1.5 Bioengineering to improve secondary metabolite production

As discussed in section 1.2.5.3 Synthetic biology, the production of secondary metabolites can be accomplished by expressing the appropriate biosynthetic genes in a heterologous organism, such as *S. cerevisiae* and *E. coli*. Metabolic engineering can also be used to overexpress or silence enzymes in the host plant to increase NP production (191). The overexpression of the *P. hybrida* chalcone isomerase (CHI) gene in tomato led to a 78-fold increase in flavonoid levels in the peel and a 21-fold increase in the paste manufactured from the fruit (192). RNA interference (RNAi) technology has been used to regulate flower colour in numerous plant species by altering flavonoid accumulation (193). For example silencing CHI in *N. tabacum* led to reduced pigmentation and a change in petal colour from pink to mostly white (Figure 11a) (194).

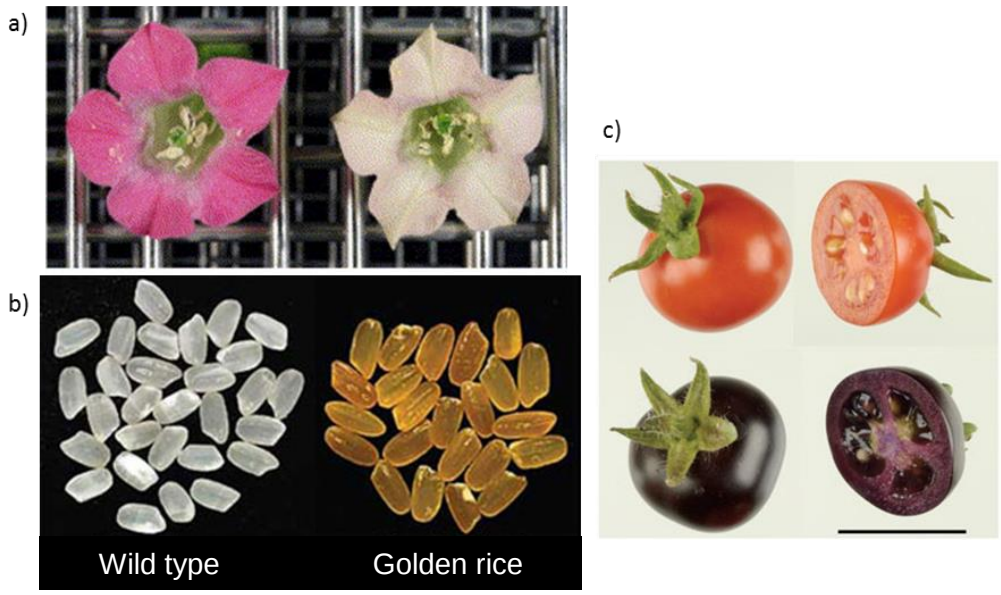


Figure 11 – Examples of altering secondary metabolite yields through bioengineering involving enzymes (a-b) and transcription factors (c). a) The flower phenotype of *Nicotina tabacum* plants. On the left is a wild type flower and on the right an RNA interference (RNAi) chalcone isomerase transgenic that has lower anthocyanin levels leading to reduced red pigmentation. b) Polished wild-type and Golden rice grains. Golden rice contains a T-DNA insert of phytoene synthase (from *Z. mays*), lycopene-β-cyclase (from *Narcissus pseudonarcissus*) and carotene desaturase (from *Erwinia uredovora*) leading to an accumulation of carotenoids. The increase in carotenoids, including β-carotene which is a source of vitamin A, leads to the rice turning a gold colour. c) Whole and cross-section of ripe wild-type (top) and Del/Ros1 (bottom) tomato fruit, scale bar (2 cm). The overexpression of the Del (bHLH) and Ros1 (MYB) transcription factors leads to an increase in anthocyanins which can be seen in the purple colouration of the fruit flesh and the peel. Abbreviations Del = Delila, Ros= Rosea.

Metabolic engineering can also be employed to introduce new pathways into a species or produce unnatural NPs. A classic example of genetically engineering a plant to produce a non-endogenous NP is the production of β-carotene in rice (Golden rice) (Figure 11b). Extensive

research conducted by multiple labs has culminated in β-carotene yields of up to 37 µg/g; achieved by expressing three enzymes phytoene synthase (from *Z. mays*), lycopene-β-cyclase (from *Narcissus pseudonarcissus*) and carotene desaturase (from *Erwinia uredovora*) (195–197). This β-carotene has been shown to be a good source of vitamin A and it is hoped that Golden rice will combat the serious issue of malnutrition in the developing world (198, 199). Work in *C. roseus* has been successful in creating novel NPs containing carbon–halogen bonds improving the compound's bioavailability. The expression of PyrH and RebH prokaryotic halogenases (and their corresponding reductases) generated cholorinated tryptophan at the 5 and 7 positions respectively which were subsequently incorporated in TIA metabolism (200, 201).

Metabolic engineering efforts targeting enzymes that are limiting the flux through a biosynthetic pathway have been successful but rely on correctly identifying the rate determining steps and the end product yields might not be improved as the pathway can contain multiple bottlenecks (202). An alternative approach is to employ TFs that are capable of regulating multiple or all sections of a biosynthetic pathway. CrWRKY1 is able to up-regulate key TIA biosynthetic genes and when overexpressed in hairy root cells increased serpentine accumulation 3-fold (147). The expression of one TF may not be sufficient to activate an entire pathway, however, as is the case with ORCA3, which when overexpressed in *C. roseus* cell cultures does not increase TIA production (176, 203).

The involvement of multiple TFs may be required to increase an entire pathway, for example the overexpression of the bHLH and MYB TFs Delila (Del) and Rosea1 (Ros1) from *Antirrhinum majus* (snapdragon) in tomato led to high levels of anthocyanins in the peel and the flesh (Figure 11c) (204). The total anthocyanin levels in the Del/Ros tomatoes are much greater than when the biosynthetic gene CHI was overexpressed (2.83 mg per g of fresh weight compared to 300 µg per g of fresh weight), because overexpressing CHI only increased anthocynanin production in the peel which represents 5% of the total fruit (192, 204). Overexpression of the maize TFs R and C1 can increase anthocyanin pigmentation in *A. thaliana* and *N. tabacum* (205) and can increase isoflavone levels in *Glycine max* when co-transformed with a construct suppressing a competing anthocyanin biosynthetic enzyme flavanone 3-hydroxylase (206). The key regulators of the paclitaxel biosynthetic pathway need to be identified if they are to be employed in metabolic engineering efforts to increase paclitaxel yields.

1.6 Employing deep sequencing technologies to analyse key transcriptional regulators

As non-model organisms *Taxus spp.* do not have an available sequenced genome. Deep sequencing technologies such as Roche454 and Illumina sequencing, however, provide a cost effective platform to perform transcriptomic studies (207). These technologies can provide a global insight into the transcriptional reprogramming that occurs after MeJA elicitation, help identify missing biosynthetic enzymes and elucidate transcriptional regulators. Deep sequencing technologies have been used in *Ophiorriza pumila* to identify putative enzymes in camptothecin

and anthraquinone biosynthesis (208) and transcript profiling studies in *T. baccata* have successfully in identified a missing paclitaxel biosynthetic enzyme β-phenylalanine-CoA (44). Transcript profiling studies investigating transcriptional reprogramming after MeJA elicitation have identified key regulators in NP metabolism including: ORC1 in *N. tabacum* which regulated nicotine biosynthesis (112, 170); SlMYC1 and SlWRKY73 in *Solanum lycopersicum* that activate terpene synthesis (209); and ORA47 in *A. thaliana* which is a transcriptional activator of the MeJA signalling cascade (114).

There have been numerous efforts at sequencing *Taxus spp.*, some of which explore the transcriptional response after MeJA elicitation and these are summarised in Table 7. Deep sequencing technologies have already been employed to identify regulators of paclitaxel biosynthesis. In 2015 Lenka *et al.* identified two bHLH TFs TcJAMYC2 and TcJAMYC4, from an as yet unpublished 454 GS FLX *Taxus* transcriptome sequence dataset, that were induced by MeJA and can negatively regulate six and five paclitaxel biosynthetic promoters respectively (TcJAMYC2 represses *T5αH* and both TFs repress *DBAT, DBBT, PAM, BAPT* and *DBTNBT*) (184).

Species of Taxus	Explant	MeJA elicitation	Profiling method	Ref
T. chinensis	Cultured suspension cells	Yes	Illumina deep sequencing	(210)
T. chinensis	Cultured suspension cells	Yes	Illumina deep sequencing	(211)
T. chinensis	Cultured suspension cells	Yes	Random Sanger sequencing of cDNA library	(212)
T. cuspidata	Cultured suspension cells	Yes	SSH cDNA library	(213)
T. cuspidata	CMC suspension culture	No	454 deep sequencing	(64)
T. cuspidata	Needles	No	454 deep sequencing	(214)
T. mairei	Roots, leaves, stems	No	Illumina deep sequencing	(215)
T. mairei	Leaves	No	Illumina deep sequencing	(216)
T. media	Cultured suspension cells	Yes	Illumina deep sequencing	(217)
T. baccata	Cultured suspension cells	Yes	cDNA-AFLP	(218)

Table 7 – A table detailing the sequencing efforts already published in *Taxus spp.* The species, explant tested, whether the response to methyl jasmonate (MeJA) elicitation was investigated, the transcriptomic profiling method used and the associated reference are stated. Research has been conducted on suspension cultures and across different parts of the plant using deep sequencing technologies and other profiling methods, such as cDNA libraries. The table is modified from Cusido *et al.* (219). Abbreviations SSH = a PCR based suppression subtractive hybridization, cDNA-AFLP = complementary DNA- amplified fragment length polymorphism.

Three other *Taxus* TFs have been reported: 1) in 2013 Li *et al.* isolated the cDNA of TcWRKY1 from *T. chinensis* using yeast-one-hybrid with important regulatory elements of the *DBAT* promoter as bait. This TF was shown to be MeJA inducible and activate DBAT expression (183); 2) Walker *et al.* in 2015 also identified TcMYC1 using degenerate primers based on *A. thaliana* MYC2 and *S. tuberosum* JAMYC10 sequences. TcMYC1 was found to be MeJA inducible and a repressor of *DBAT*, *BAPT* and *DBTNBT* promoters (184). 3) In 2009 Dai *et al.* isolated cDNA of a *T. cuspidata* AP2 TF using three tandem copies of the JERE motif from strictosidine synthase promoter (*STR*) in *C. roseus* as bait in a yeast-one-hybrid experiment. TcAP2 was shown to be responsive to a number of environmental responses, including MeJA and SA, but its transactivation ability has yet to be investigated (220).

1.7 Conclusions

The chemical diversity of PNPs makes them a rich source of pharmaceutical products, especially anti-bacterial and anti-cancer drugs. Among these paclitaxel is a key anti-cancer drug, originally isolated from the bark of *Taxus spp*., whose medicinal uses are continually expanding. This increasing demand is an issue as paclitaxel is found in very low concentrations within the plant. PCC is seen as an attractive production route compared to natural harvest, total and semi synthesis. CMC culture provides a good platform from which to increase drug production as they possess superior growth properties on an industrial scale compared to typical dedifferentiated cell culture.

Elicitors, such as MeJA, can up-regulate paclitaxel production in PCC however their effect is only transient. Bioengineering involving TFs has been shown to successfully increase NP yields. Deep sequencing technologies, such as Roche454 and Solexa sequencing, can be used to identify key regulators of the paclitaxel biosynthetic pathway and this knowledge can then be exploited in metabolic engineering efforts to increase yield of paclitaxel in PCC.

1.8 Objectives of the project

My research has been conducted as part of a larger project that has been ongoing since 2006. Numerous past and present PhD students (past: Zejun Yan, Rabia Amir, Eunjung Kwon and Thomas Waibel and present: Marisol Ochoa Villarreal) and visiting academics from Jiangsu Normal University in China have contributed to the project over the years. The work was also completed in collaboration with our South Korean industrial partner Unhwa who provided us with CMCs from three different *Taxus spp.* Establishing the transcriptome of *T. cuspidata* CMCs after MeJA elicitation and the identification of 19 candidate TFs that were highly up-regulated at an early time point after MeJA elicitation (0.5 h) was conducted by Waibel, Zejun, Amir and Kwon and is discussed in further detail in section 2.1.1 and section 2.3 respectively.

The goal of the overall project is to identify important regulators of the paclitaxel biosynthetic pathway that are involved in the MeJA elicitation response in *T. cuspidata* CMCs, with the final objective being to increase paclitaxel production in CMC culture. Once TFs have been identified and characterised as regulators the approach will be to transform them into *Taxus spp.* CMCs and observe any changes to metabolite production that will hopefully increase paclitaxel production yields in culture.

The specific aim of my PhD project is to investigate whether candidate TFs are capable of regulating the expression of known promoters of the paclitaxel biosynthetic pathway. The ability of the 19 candidate TFs to interact with the 10 known promoters of the the paclitaxel biosynthetic pathway were investigated *in vivo* using plant transient assays and *in vitro* using gel shift assays. Possible combinatorial control of the paclitaxel biosynthetic pathway by MYB and bHLH TFs was explored *in vivo* using yeast-two-hybrid and plant transient assays. Possible post-translational control of bHLH TFs by identified *Taxus* JAZ proteins was explored *in vivo* also using yeast-two-

hybrid assays. Once regulators of the paclitaxel biosynthetic pathway were identified attempts were made to establish a particle bombardment-based assay to transform *Taxus spp.* CMCs. This was conducted in order to attempt to verify the initial working hypothesis that these TFs can increase paclitaxel production in PCC which could be exploited industrially.

Chapter 2 - Analysis of *Taxus cuspidata* CMC gene expression after methyl jasmonate elicitation

This work was completed in collaboration with Thomas Waibel, Rabia Amir and Zejun Yan (221–223).

2.1 Introduction

JA and its conjugates (e.g. MeJA) are small signalling molecules ubiquitous in the plant kingdom. Their production is triggered by environmental stresses, such as pathogen attack, and leads to massive reprogramming of gene expression affecting a variety of processes including growth, development and secondary metabolism (114). MeJA is a well-known elicitor of *Taxus spp.* cell culture and can increase paclitaxel production in CMCs by 14 000% (64, 92). Genome-wide transcript profiling studies have found that genes involved in specific biosynthetic pathways are tightly co-regulated and gene induction occurs early (1–4 h) after elicitation, indicating that the activation is directly caused by MeJA (85). MeJA activates transcriptional cascades which are important in co-ordinating the plant's stress response.

High throughput screening technologies, such as Roche454 and Solexa/Illumina sequencing, provide a rapid cost effective platform to study changes in gene expression. Several groups have used these technologies to explore *Taxus spp.* transcriptomes, however no one has previously investigated the effect of MeJA elicitation on CMCs (Table 7).

2.1.1 Establishing the *Taxus cuspidata* CMC transcriptome

The *T. cuspidata* CMC transcriptome was established using Roche 454 sequencing while Illumina sequencing was employed to compare the molecular signatures of CMCs in response to MeJA (100 µM) at three time points (0.5, 2 and 12h) after elicitation to a mock treatment with DMSO (0 h) in biological triplicate (Figure 12). Sample preparation and RNA extraction were completed by Waibel and the remaining protocol was completed by Florian Halbritter at GenePool (Edinburgh, UK). Mapping and annotation of the 454 contigs was completed by Halbritter using mapping software MAQ v. 6.0.8 . Statistical analysis was performed in R using edgeR (224) by Simon Tomlison, Genepool (Edinburgh, UK) to identify contigs with differential expression at least one time point post elicitation compared to 0 h. EdgeR employs an overdispersed Poisson distribution to model the read count, with the degree of overdispersion moderated using an empirical Bayes procedure. Differential expression was assessed using a modified version of Fisher's exact test with the *p*-values adjusted for the false discovery rate of ≤0.05. 1646 contigs were identified that had differential expression at at least one time point tested after MeJA elicitation (Figure 12). The integrity of the biological replicates was established using hierarchical cluster analysis which showed that all except one of the replicates (#2 2h) was closely mapped together (Figure 13).

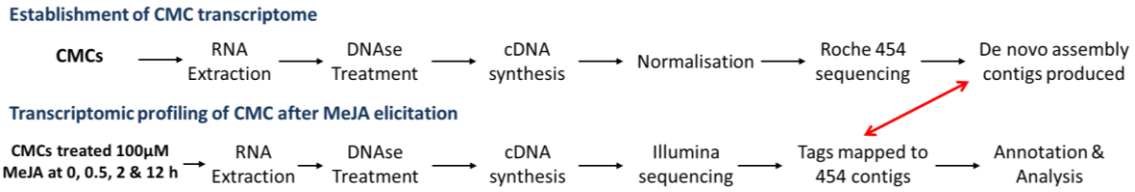


Figure 12 – Schematic illustration of the sequencing performed on *Taxus cuspidata* cambial meristematic cell (CMC) culture to establish their transcriptome and response at three time points (0.5, 2 and 12 hours (h)) after 100µM methyl jasmonate (MeJA) elicitation. For both sequencing experiments samples were prepared by extracting RNA, subjecting it to DNase treatment and subsequent cDNA synthesis. When the CMC transcriptome was established using Roche 454 sequencing an additional cDNA normalisation step was required. Roche 454 produced contigs which were assembled de novo to produce the transcriptome of *T. cuspidata* CMCs. The transcript profiling of CMCs after MeJA elicitation employed Illumina sequencing and the tags produced were mapped back (red arrow) to the transcriptome produced by Roche 454 sequencing. These were subsequently annotated using BLASTX (Basic local alignment search tool) (225) and analysed using edgeR software (224) to identify differential expression at at least one time point after MeJA elicitation compared to 0 h. EdgeR used a modified Fisher's exact test to assess differential expression with *p*-values adjusted for the false discovery rate of ≤0.05.

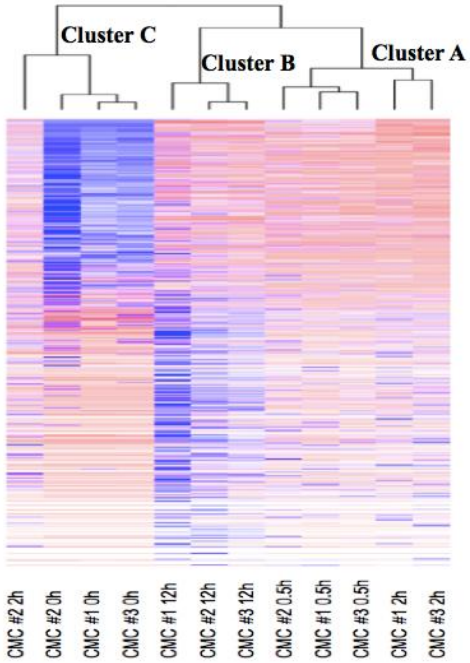


Figure 13 – The heatmap and dendrogram of the 1646 differentially expressed genes that were differential expressed at at least one time point after methyl jasmonate elicitation, tested with a false discovery rate of ≤0.05. Samples are clustered based on their gene expression pattern and form three clusters. Red indicates gene up-regulation, blue indicates gene down-regulation. All three replicates of time point 0 hours (h) are in cluster C along with the anomalous 2 h #2 replicate. All three replicates of 0.5 h and two replicates #1 and #3 of 2 h are in Cluster A and all three 12h replicates are in Cluster B. This shows the integrity of the biological replicates was satisfactory except for the #2 2 h.

2.2 Analysis of MeJA responsive *T. cuspidata* CMC transcriptome

Of the differential expressed genes (DEGs) identified 941 were significantly up-regulated, while 713 were down-regulated in response to MeJA elicitation. The top 10 up-regulated genes at each time point include at least four paclitaxel biosynthetic genes; with T7βH, DBTNBT and T13αH represented at every time point. While the top down-regulated genes included dehydrin and glutathione-transferase proteins.

MeJA activates a transcriptional cascade which can be seen in the data as the number of DEGs increased from 722 at 0.5 h to 1084 at 12 h (Figure 14). The initial changes in gene

expression (0.5 h) are more inductive, while at the later time point there is a 3-fold increase in the number of down-regulated DEGs. The number of DEGs is slightly reduced at 2 h, suggesting a possible lag after the initial wave triggered by MeJA. The later wave in gene expression shows positive and negative regulation showing the tight control that the plant has over its MeJA response, which is probably controlled by the initial wave.

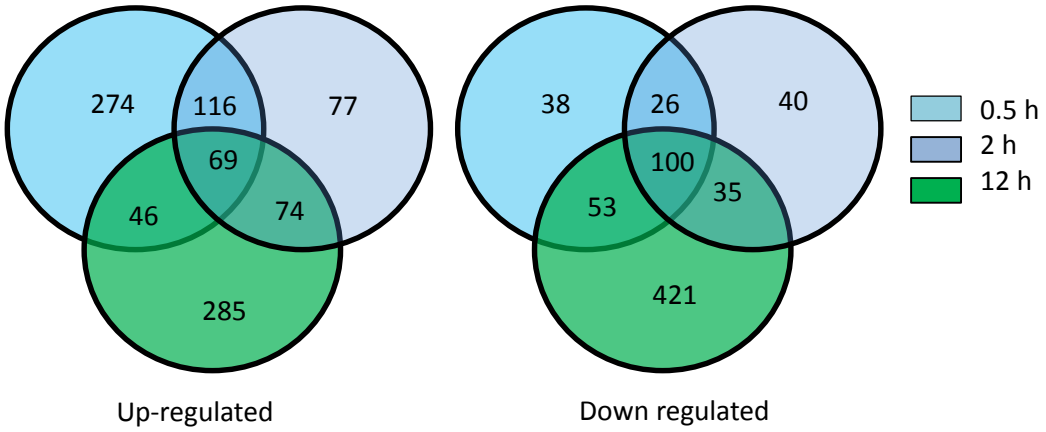


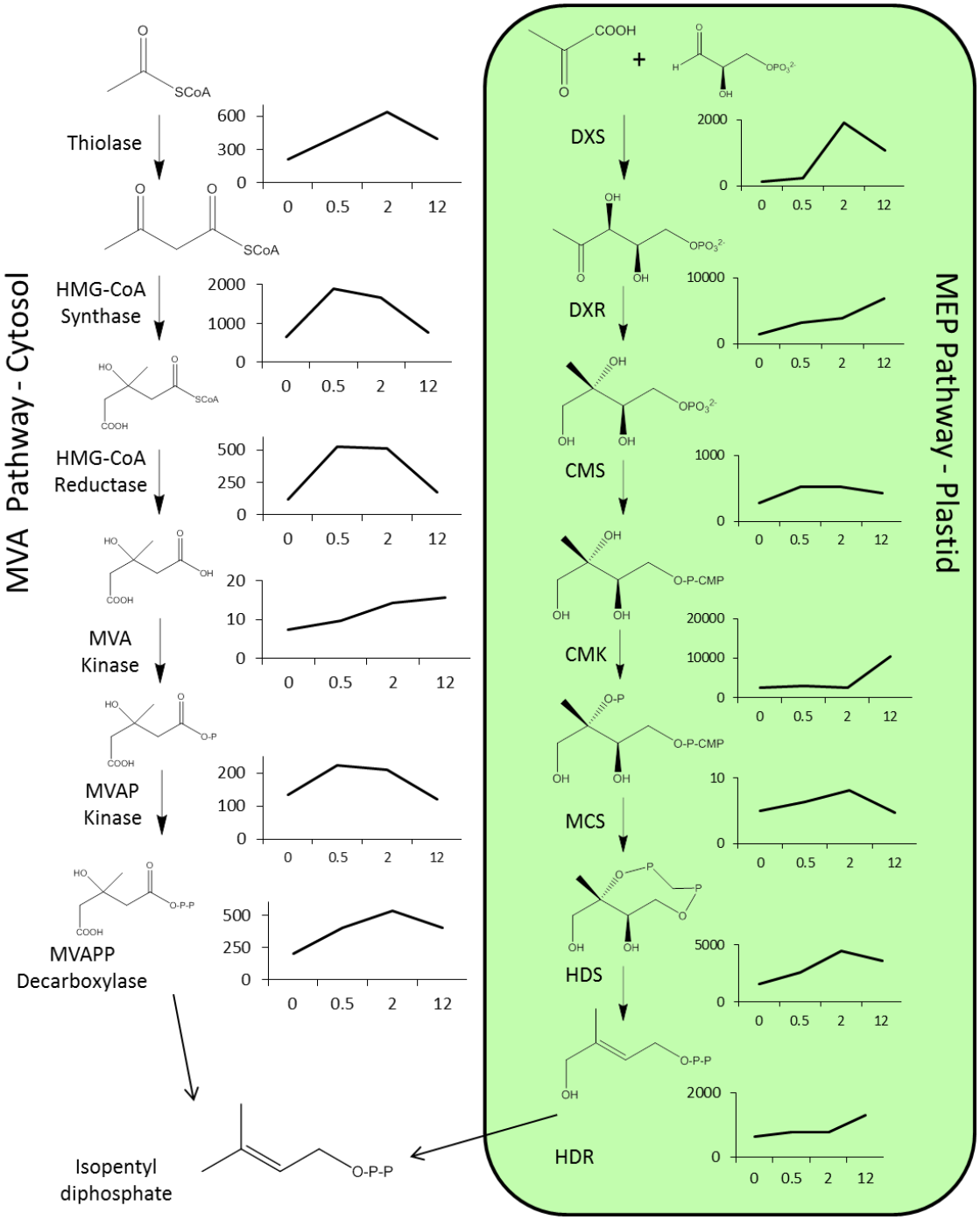
Figure 14 – Venn diagram showing the number of differentially expressed genes (DEGs) at three time points 0.5 hour (h), 2h and 12h after methyl jasmonate elicitation that are either up- and down-regulated. The total number of DEGs increased over time with the majority of DEGs at an early time point (0.5 h) being up-regulated while the number of down-regulated DEGs significantly increasing at 12 h.

2.2.1 Analysis of terpenoid biosynthesis

2.2.1.1 Isoprenoid precursor synthesis

The production of paclitaxel in *T. cuspidata* CMCs requires the formation of the isoprenoid precursors IPP and DMAPP by either the cytosolic MVA pathway or the plastidial MEP pathway. All the enzymes of both pathways were identified in the sequencing data. In the MVA pathway the gene expression increased rapidly after MeJA addition, however their expression levels started to reduce by 12 h (Figure 15). This suggests there is an increase in the production of IPP after MeJA addition but the MVA pathway starts to be negatively regulated by 12 h. The expression of the genes in the MEP also increased but compared to the MVA pathway the expression of three genes – DXR, CMK and HDR – were still increasing at 12 h (Figure 15). MeJA elicitation caused an increase in nearly all the genes in both pathways, but the peak in gene expression differs between enzymes.

The IPP and DMAPP produced by the MEP and MVA pathway undergo multiple rounds of condensation to produce the precursor GGPP, which is required for the first committed step of the paclitaxel biosynthetic pathway (Figure 6). The expression of all the enzymes involved in this process increased after MeJA elicitation. The peak in gene expression occurs later for the enzymes lower down the biosynthetic pathway (Figure 16). The first steps involving isopentenyl-diphosphate isomerase (IPS) and geranyl diphosphate synthase (GPS) both had their highest expression at 0.5 h, while farnesyl diphosphate synthase (FPS) peak expression was at 2 h and geranylgeranyl diphosphate synthase (GGPPS) expression increased up to 12 h.



Enzyme abbreviations

MEP	DXS	1-deoxy-D-xylulose 5-phosphate (DXP) synthase
	DXR	DXP reductoisomerase
	CMS	1,4-diphosphocytidyl-2*C*-methyl-D-erythritol(CDP-ME) synthase
	CMK	CDP-ME kinase
	MCS	2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase
	HDS	4-hydroxy-3-methylbut-2-enyl diphosphate (HMBPP) synthase
	HDR	HMBPP reductase
MVA	HMG-CoA	3-hydroxy-3-methylglutaryl-CoA
	MVA	Mevalonate
	MVAP	Phosphomevalonate
	MVAPP	Diphosphomevalonate

Figure 15 – The cytosolic mevalonic acid (MVA) pathway and the plastidial non-mevalonate pathway (MEP), showing the change in gene expression after methyl jasmonate (MeJA) elicitation at three time points (0.5, 2 and 12 hour (h)). The arrow show the transformation performed by each enzyme and their abbreviations are stated in the table. The graphs adjacent to each enzyme show the gene expression level (reads per minute) vs. time post MeJA elicitation (h). All the enzymes in the MEP and MVA have increased expression after MeJA, however the expression of the majority

of enzymes in the MVA pathway has started to reduce by 12 h. In contrast the expression of three genes in the MEP DXR, CMK and HDR were still increasing at 12 h.

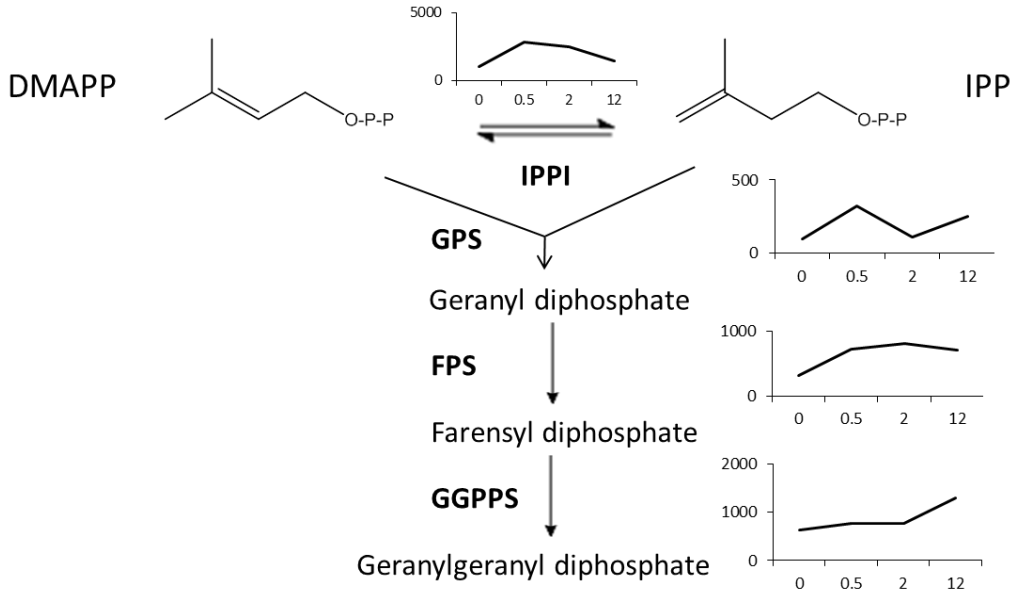


Figure 16 – The change in gene expression of enzymes involved in the production of geranylgeranyl diphosphate (GGPP) at three time points (0.5, 2 and 12 hours (h)) after methyl jasmonate (MeJA) elicitation. The isoprenoid precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) undergo rounds of condensation catalysed by the enzymes geranyl diphoshate synthase (GPS), farensyl diphoshate synthase (FPS) and GGPP synthase (GGPPS). IPP can be isomerised into DMAPP by isopentenyl pyrophosphate isomerase (IPPI). The graphs adjacent to each enzymatic transformation show the gene expression level (reads per minute) vs. time post MeJA elicitation (h).

Overall there is an increase in gene expression in all the genes of the early terpenoid pathway after MeJA elicitation providing a greater pool of precursors that can be used to produce paclitaxel. However none of these genes (except GGPPS) were found in the 1646 DEGs suggesting that the pool of precursors does not substantially increase and that the increase in paclitaxel after MeJA addition is due to changes in the flux of competing terpenoid pathways.

2.2.1.2 Paclitaxel biosynthesis

In comparison, the genes of the paclitaxel biosynthetic pathway were all found to be significantly up-regulated at at least one time point after MeJA elicitation. The paclitaxel biosynthetic pathway is believed to contain 19 steps, with only 14 of these enzymes currently characterised (discussed in 1.2.4 Paclitaxel biosynthetic pathway). All the characterised enzymes, except CoA-ligase were identified in our DEGs along with an unknown “taxoid-*O*-acetyl transferase” (T-O-AT). The expression of all these enzymes increases after MeJA addition, but the highest logFC value was observed at different time points. Most of the early pathway genes (TASY, T5αH TDAT, T10βH and T13αH) have the highest logFC at 2 h, while two of the late pathway genes (DBBT and DBTNBT) continued to increase their gene expression up to 12 h. BAPT is thought to be a possible rate limiting step in the pathway (47) and in our data its expression increased after MeJA elicitation at 0.5 h but then reduced slightly over time, which is different to all the other enzymes. This may explain why BAPT has been observed as a limiting enzyme because its expression was not induced for as long as the other enzymes in the pathway.

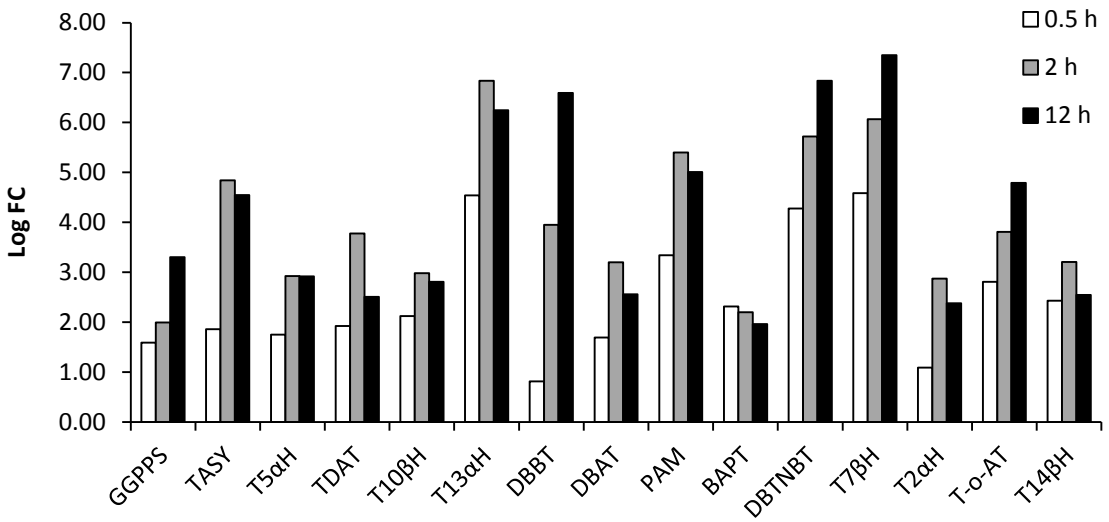


Figure 17 – The log fold change (LogFC) in gene expression of the paclitaxel biosynthetic enzymes at three time points (0.5, 2 and 12 hour (h)) after methyl jasmonate (MeJA) elicitation. The first 11 enzymes are stated in the order they occur in the paclitaxel biosynthetic pathway (see Figure 6 for pathway and **Table 2** for enzyme abbreviations) and the remaining four enzymes are not found in the linear paclitaxel pathway but are part of the branched pathway which produces a range of taxanes within *Taxus spp*. The LogFC values for all the enzymes increases after MeJA, however the peak in gene expression alters between the genes.

Xylosyltransferases can compete with the paclitaxel biosynthesis shunting intermediates out of the pathway by adding a xylsoyl group to form, for example, 10-deacetylpaclitaxel and 10-deacetylcephalomannine. Only one xylosyltransferase was identified in our DEGs and this was significantly down-regulated at 12 h. This implys that during MeJA elicitation intermediates are not funnelled out of the pathway which concurs with data published by Sun *et al.* (217).

2.2.2 Analysis of genes involved in hormone signalling pathways

2.2.2.1 Jasmonate signalling pathway

MeJA can induce gene expression in the JA biosynthesis and signalling pathway, producing a positive feedback loop (85). All the genes known to be involved in JA biosynthesis (see Figure 9) were identified in the transcriptome, however only four were significantly up-regulated after MeJA addition – the JA biosynthetic enzymes LOX, AOS and OPDR and the MeJA esterase which converts MeJA to JA (Figure 18). The signalling components COI, SKP1, TPL, NINJA, MYC2 and JAZ proteins (Figure 10) were identified in the transcriptome showing that *T. cuspidata* has the same core JA module found in other species such as *A. thaliana, N. tabacum* and *C. roseus* (100).

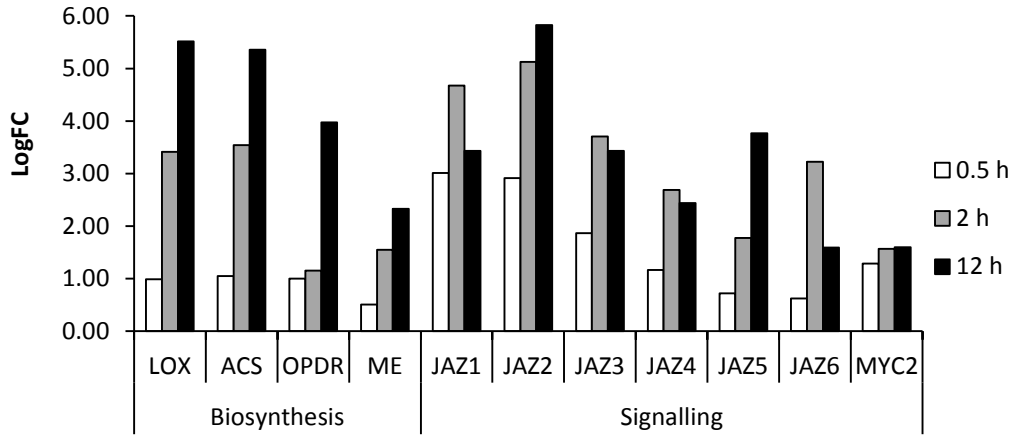


Figure 18 – The log fold change (LogFC) in gene expression of the jasmonic acid (JA) biosynthetic enzymes and signalling components at three time points (0.5, 2 and 12 hour (h)) after methyl jasmonate (MeJA) elicitation. All the genes known to be involved in JA biosynthesis and in the JA core signalling module were identified in the transcriptomic data (see Figure 9 and Figure 10), however only four biosynthetic genes (LOX; lipoxygenase allene, AOS; oxidase synthase, OPDR; 12-oxophytodienoate reductase, ME: MeJA esterase) were found to be significantly up-regulated; and of the signalling components only JAZ proteins and MYC2 (MYC2 is transcription Factor bHLH08058, see section 2.4) were identified in the 1646 significantly differential expressed genes.

Six JAZ proteins were identified that were significantly up-regulated after MeJA elicitation, along with the MYC2 homology bHLH08058 (Figure 18). These findings differ slightly from Li *et al.* who found that the expression of additional JA biosynthesis genes (phospholipase D, DAD1 and AOC) were also significantly increased after MeJA addition (210). The self-activation of JA biosynthesis is well known in the literature (114) and all the JA biosynthetic enzymes identified in our data set were up-regulated after MeJA however the increase was not always significant. qRT-PCR data obtained by Li *et al.* showed higher mRNA levels than their RNA-seq data (210) therefore future qRT-PCR data might confirm a significant increase in all the JA biosynthesis enzymes as observed in other reports.

2.2.2.2 Ethylene

Jasmonates can interact and crosstalk with other plant hormones, such as ethylene, that lead to changes in secondary metabolite production. Ethylene is known to be produced following MeJA elicitation and three ethylene biosynthesis genes *S*-adenosylmethionine-methyltransferase (SAM), aminocyclopropane-1-carboxylic acid (ACC) synthase and ACC oxidase were identified as significantly upregulated in our 1646 DEGs (Figure 19). The role of ethylene in paclitaxel production is unclear as application of ethylene inhibitors to MeJA elicited *Taxus spp.* cultures increased paclitaxel production suggesting it acts as a suppressor (226). However the addition of ethylene at low concentrations in conjunction with MeJA increased production of paclitaxel in *T. cuspidate* cell cultures (90). The increase in ethylene biosynthetic enzymes is likely to have an unknown regulatory effect on paclitaxel production, therefore future research creating RNAi lines of ethylene biosynthetic genes may lead to increased paclitaxel production.

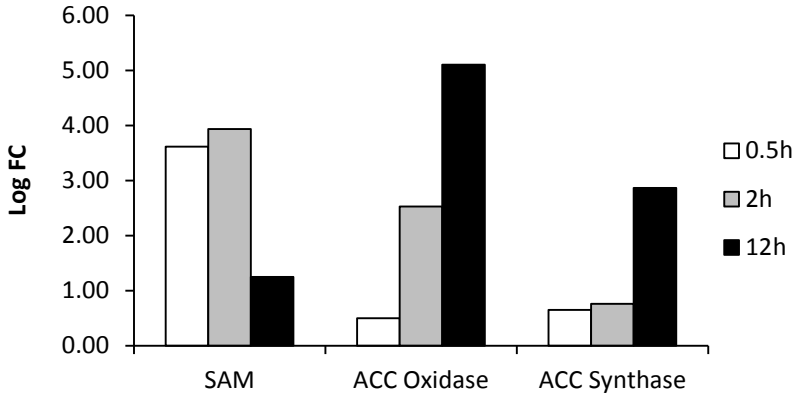


Figure 19 – The log fold change (LogFC) in gene expression of the ethylene biosynthetic enzymes at three time points (0.5, 2 and 12 hour (h)) after methyl jasmonate (MeJA) elicitation. The biosynthetic enzymes are shown in the order that they occur in the biosynthetic pathway and their abbreviations are: *S*-adenosylmethionine-methyltransferase (SAM), aminocyclopropane-1-carboxylic acid (ACC) synthase and ACC oxidase. After MeJA elicitation the expression of the ethylene biosynthetic genes increases, with ACC oxidase and synthase expression highest at 12 h, while SAM expression is lowest at the same time point.

2.2.2.3 Auxin

The interaction between jasmonates and auxin is poorly understood but they seem to act antagonistically (227). Auxin and jasmonates are linked via shared co-repressors such as TPL, which interacts with Auxin/IAAs and the NINJA-JAZ complex (106), and the action of auxin response factors (ARFs). Two ARFs were identified in the DEGs and both were significantly down-regulated after MeJA elicitation, ARF07285 at 0.5 h and ARF21521 at 12 h (Figure 20). Genes denoted auxin induced were also significantly down-regulated after MeJA addition along with two out of three auxin transporters. Two auxin related biosynthetic genes, IAA-amido synthetase and IAA amino acid hydrolase, were also identified in the DEGs and were significantly down-regulated (Figure 20). This data supports the theory that MeJA is capable of negatively regulating auxin responses.

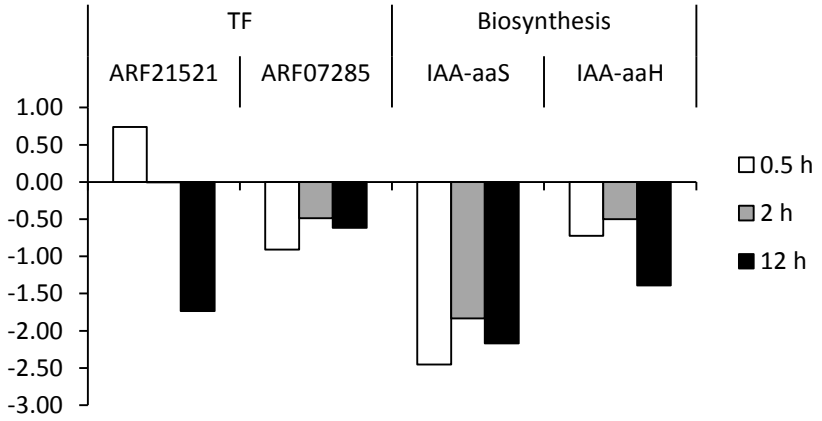


Figure 20 – The log fold change (LogFC) in gene expression of auxin response factors (ARFs) and two auxin biosynthetic genes, IAA-amido synthetase (IAA-aaS), IAA amino acid hydrolase (IAA-aaH) at three time points (0.5, 2 and 12 hour (h)) after methyl jasmonate (MeJA) elicitation. The addition of MeJA negatively regulates some auxin responses.

2.2.2.4 Conclusion

It is important to consider how other hormone signalling pathways may interact with the MeJA response. Jasmonate signalling does not happen in isolation and crosstalk occurs with numerous hormones. A greater understanding of how these pathways might act synergistically or antagonise each other could help to improve paclitaxel production. For example, salicylic acid (SA) is known to act as an antagonist against jasmonates (228) however a combination of ultrasound and SA can increase paclitaxel production in *T. baccata* cell culture (229).

2.3 Identification of methyl jasmonate induced transcription factors

As previously discussed JA elicitation triggers extensive genetic reprogramming and TFs are important in tightly coordinating the cell's response (section 1.4 Transcription factors). Work to identify putative TFs in the *T. cuspidata* CMC transcriptome with significantly altered expression after MeJA elicitation was conduct by Zejun, Amir and Kwon (222, 223). Within the 1646 DEGs 78 were identified as TFs, characterised into 19 different gene families. The largest family was the AP2 (ERF) group, followed by MYB, bHLH, C3H and NAC TF families (Figure 21a). Of the TFs identified, 50 were up-regulated by MeJA elicitation, including a high percentage of the MYB, NAC, LOB, GRAS and WRKY TFs. Whereas of the 29 TFs that were down-regulated the AP2, bZIP and CCAAT TF families were highly represented (Figure 21b).

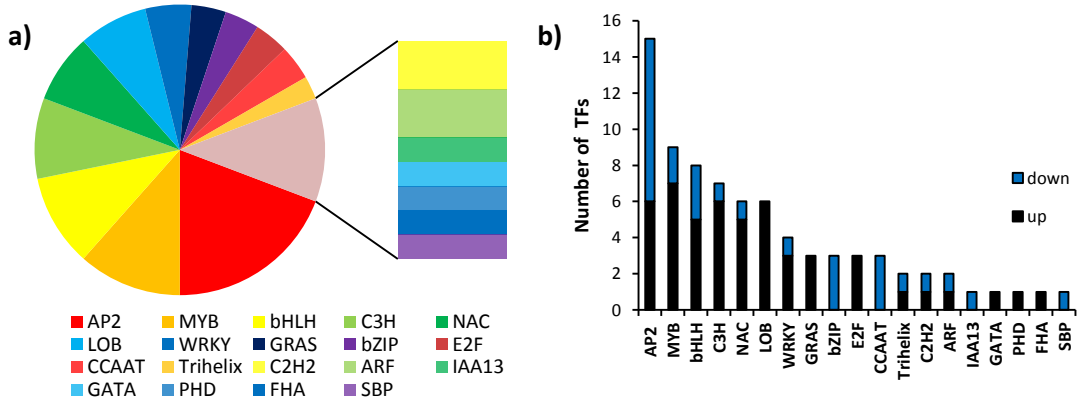


Figure 21 – Summary of the transcription factors (TFs) identified in the 1646 differential expressed genes (DEGs). a) A pie chart showing the TF families represented in the DEGs. The three major families include AP2, MYB and bHLH; b) the number of TFs up- and down-regulated after methyl jasmonate elicitation in each TF family. There are a number of families that are well represented in the up-regulated TFs including MYB, NAC and WRKY TFs, while most of the down-regulated TFs are in the AP2 family.

2.3.1 Identification of TFs highly induced at 0.5 h after MeJA elicitation

MeJA elicitation increases the expression of paclitaxel biosynthetic genes (2.2.1 Analysis of terpenoid biosynthesis) and consequently the amount of paclitaxel produced. This project is interested in identifying TFs that might be involved in the up-regulation of paclitaxel production after the addition of MeJA. Candidate TFs were chosen if they were highly up-regulated at an early time point (0.5 h) after MeJA elicitation; with the aim of identifying an overall regulator. The candidates also had high *e* values ($<e^{-4}$) using BLASTX analysis (225) and were in TF families previously reported to be involved in secondary metabolite regulation (see 1.4.6 JA induced TFs

involved in secondary metabolism). In total 19 TFs were identified; 6 AP2, 4 MYB, 5 NAC, 2 bHLH and 2 WRKY proteins (Table 8); work completed by Amir, Kwon and Zejun (222, 223).

Two of these TFs have already been reported in the literature. WRKY09595 is identical to TcWRKY1 published by Li *et al.* and bHLH08058 is a truncated version of the TcJAMYC4 published recently by Lenka *et al.* (discussed further in 2.4 The open reading frame of the bHLH08058 transcription factor)(183, 184).

Transcription factor family				
AP2	**MYB**	**NAC**	**bHLH**	**WRKY**
AP2 03304	MYB 12379	NAC 06771	bHLH 11748	WRKY09595
AP2 04485	MYB 10385	NAC 09658	bHLH 08058	WRKY 19284
AP2 07245	MYB 10855	NAC 00172		
AP2 01431	MYB 15401	NAC 05638		
AP2 00499		NAC 08447		
AP2 22386				

Table 8 – A table summarising the 19 transcription factors identified by Amir and Zejun (222, 223) **that were significantly up-regulated at 0.5 hours after methyl jasmonate elicitation and a member of a TF family previously reported to be involved in the regulation secondary metabolism.**

The induction in gene expression at an early time point after MeJA elicitation was confirmed by RT-PCR (completed by Amir(223)). All the candidates are highly induced at 0.5 h, but TFs such as both WRKY's increase their expression up to 12 h, while the logFC values for TFs MYB10385 and MYB10855 decrease overtime (Figure 22a). The top five TFs with the highest logFC in gene expression changed over time; at 0.5 h MYB10385, MYB10855 were highly induced while AP200499 and AP222386 were highly expressed at both 0.5 h and 2 h. AP2 04485 and MYB15401 have high gene expression from 2–12 h and logFC of NAC06771, WRKY09595 and WRKY19284 were highest at 12 h (Figure 22b-c). This data shows that the role of different TFs alters overtime. MYB10855 and MYB10385 are more likely to be connected with the initial more inductive transcriptional cascade, while the WRKY TFs may be connected with the second wave of gene expression involved in regulating the gene expression after MeJA.

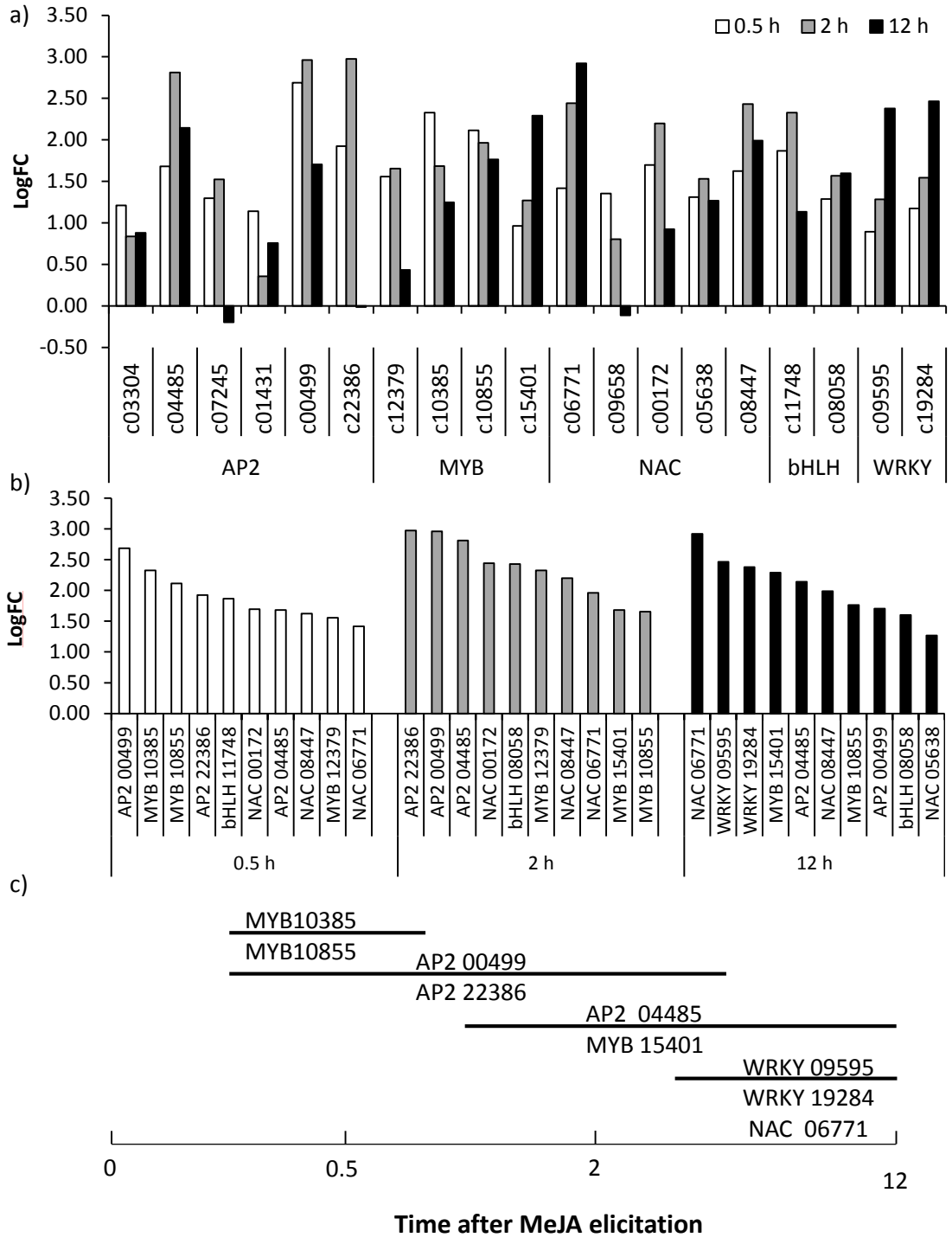


Figure 22 – Summary of induction in gene expression of the 19 candidate transcription factors (TFs) after MeJA elicitation. a) The log fold change (LogFC) in gene expression of the 19 candidate TFs at three time points (0.5, 2 and 12 hour (h)) after methyl jasmonate (MeJA) elicitation. The TF name is based on its family and the original Roche 454 contig number; b) The TFs with the 10 highest LogFC values in genes expression after MeJA addition at each time point (0.5, 2 and 12 h); c) a schematic representation of the time points at which some of the most highly induced TF are expressed. All the TFs are highly induced at an early time point after MeJA elicitation however the absolute LogFC values and change in expression over time differ. MYB10385 and MYB10855 have the highest LogFC values at 0.5 h, while WRKY09595, WRKY19284 and NAC06771 have their highest LogFC value at 12 h. This temporal separation in gene expression suggests that the TFs may have differing roles in gene regulation.

2.4 The open reading frame of the bHLH08058 transcription factor

The open reading frames (ORFs) of the 19 candidate TFs identified by Yan, Amir and Kwon found bHLH08058 to be 225 amino acids in length. In April 2015 Lenka *et al.* published a paper which included the bHLH TF TcJAMYC4 that was an elongated form of bHLH08058 (184). Subsequent analysis of the original contig showed that this elongated reading frame could be there if a mistake was made during Roche454 sequencing and a guanine base was inserted at position 650 (highlight by a red box in Figure 23). 5' and 3' race PCR were previously carried out by Amir to establish the full ORF of the protein. This work identified an extra thymine in the original contig that was not found in subsequent Sanger sequencing (highlighted by the yellow box in Figure 23) and shorten the C-terminal end of the protein. Mistakes can occur during sequencing and assembly of the Roche454 contigs.

Primers were designed based on the TcMYC4 nucleotide sequence (JX519290.1) and a PCR fragment of the correct size was cloned into pDONR221. However when multiple colonies were sequenced using Sanger sequencing, they all showed that the inserted guanine base suggested in TcMYC4 was not there. This lack of an insertion truncated the protein at only 45 residues because a stop codon was brought into frame. However further analysis of the whole contig showed that a longer ORF exists if the inserted thymine residue at 928 (yellow box in Figure 23) is removed. This ORF is 1611 base pairs (bp) long producing a 536 amino acids long that aligns with the majority of TcMYC4 starting at Asp32 (red arrow Figure 24).

This elongated ORF (bHLH08058-FE) was difficult to amplify but was successfully cloned into pDONR221 and subsequently into p2GW7,0. The ORF of the constructs was confirmed by Sanger sequencing at the pDONR and pDEST stage. It seems likely that this extended ORF is correct because when the protein sequence was aligned with other published MYC2 TFs from *A. thaliana*, *C. roseus* and *N. tabacum* and a candidate *Pinus tadea* MYC2 homolog, the alignment in the extended N-terminal region was high; and this region has been shown to be very important for interactions with JAZ proteins (230). The *P. tadea* MYC2 homolog was identified in the *P. taeda* Transcriptome Assembly v1.0 available at the Dendrome forest tree genome database (http://dendrome.ucdavis.edu). Using the AtMYC2 sequence as a query in BLASTn (225) the top candidate produced 5A_I7_OT_comp33322_c0_seq3 had an *e* value of $4e^{-35}$.

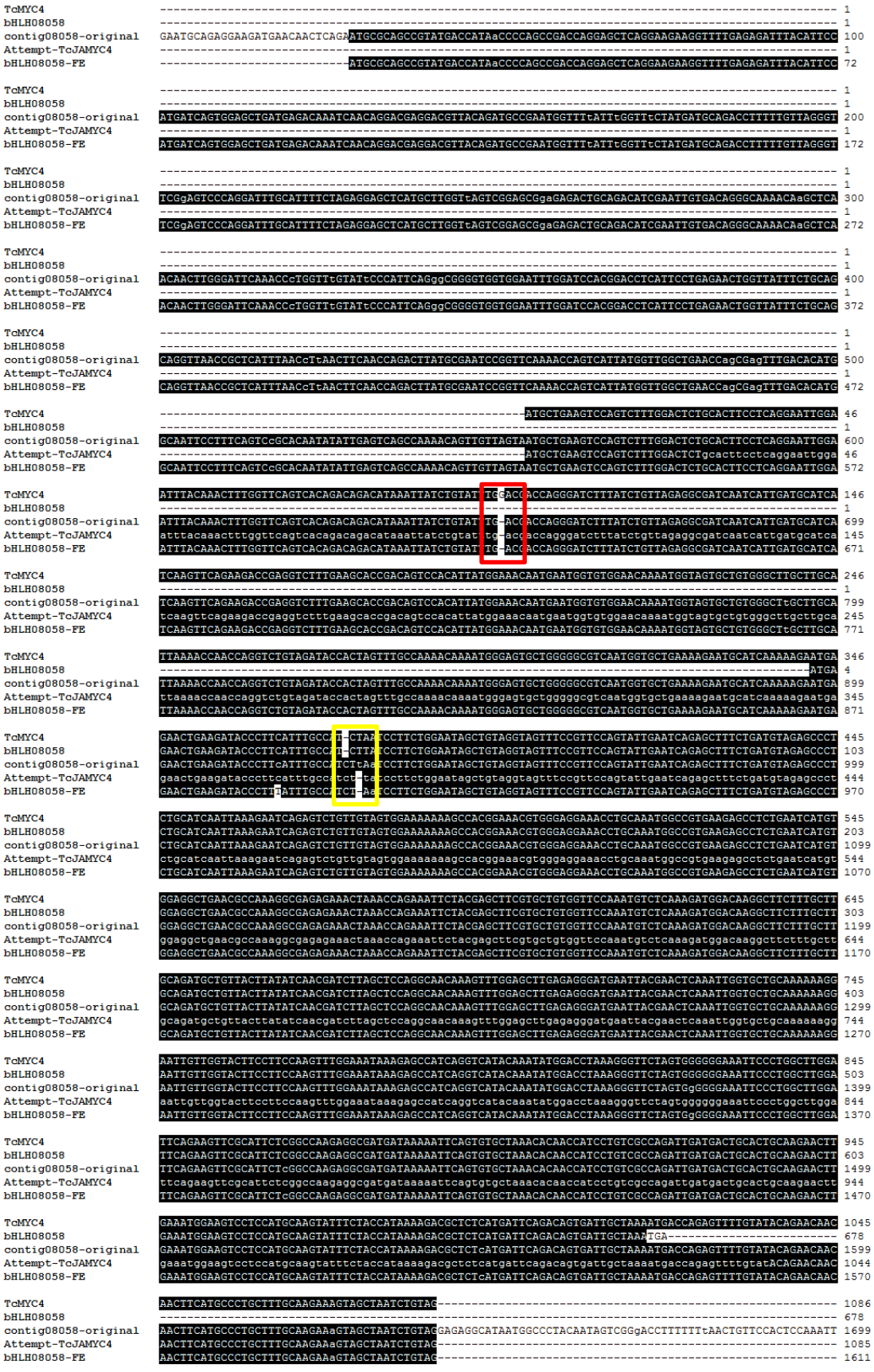


Figure 23 – The nucleotide alignment of the original 08058 contig with the open reading frame (ORF) identified by Yan, Amir and Kwon (bHLH08058) (222, 223)**, the Walker TcMYC4 ORF** (184)**, our attempt at cloning the TcMYC4 ORF and the further extended version of bHLH08058 (bHLH08058-FE).** The alignment was produced using ClustalOmega

(231, 232) and the image produced using BioEdit v7.2.5 (233). The red box highlights an inserted guanine base in the TcMYC4 sequence which none of the other sequences contain. The yellow box highlights an inserted thymine residue in the original contig08058 which was proven to be an error through multiple Sanger sequencing experiments completed by Amir (223) and was not present in all the other sequences. The bHLH08058-FE ORF is significantly longer in the N-terminal region (867 base pairs (bp)) compared to the original bHLH08058 ORF and is 69 bp longer at the C terminal. Attempts to amplify the published TcMYC4 ORF were unsuccessful as the highlighted inserted guanine base was not found.

Figure 24 – The protein alignment of the original bHLH08058 (bHLH08058-original) identified by Yan, Amir and Kwon (222, 223)**, the Walker TcMYC4** (184) **and the further extended version of bHLH08058 (bHLH08058-FE).** The alignment was produced using ClustalOmega (231, 232) and the image produced using BioEdit v7.2.5 (233). The arrow indicates where the sequence similarity between TcMYC4 and bHLH08058-FE starts. Before this region there is very little alignment between the two proteins. bHLH08058-original has high sequence similarity with the other two proteins but is significantly truncated at the N-terminus and has a 22 amino acid truncation at the C-terminus.

Unfortunately due the fact that the TcMYC4 sequence was only published in April 2015 the majority of the testing in this thesis occurred with the truncated form of the protein. The original 675 bp ORF was used during testing for possible interactions with the 10 paclitaxel biosynthetic promoters *in vivo* using plant transient assays, during exploration of combinatorial control of the paclitaxel biosynthetic pathway in concert with MYB TFs *in vivo* using yeast-two-hybrid and plant transient assays and to explore possible post-translational control with *Taxus* JAZ proteins *in vivo* using yeast-two-hybrid assays. Due to the difficulties in correctly identifying and amplifying the 1611 bp fragment (bHLH08058-FE) the majority of testing could not be repeated in the available time. Individual regulation of the paclitaxel biosynthetic promoters was tested but the other experiments need to be repeated in the future with the elongated TF.

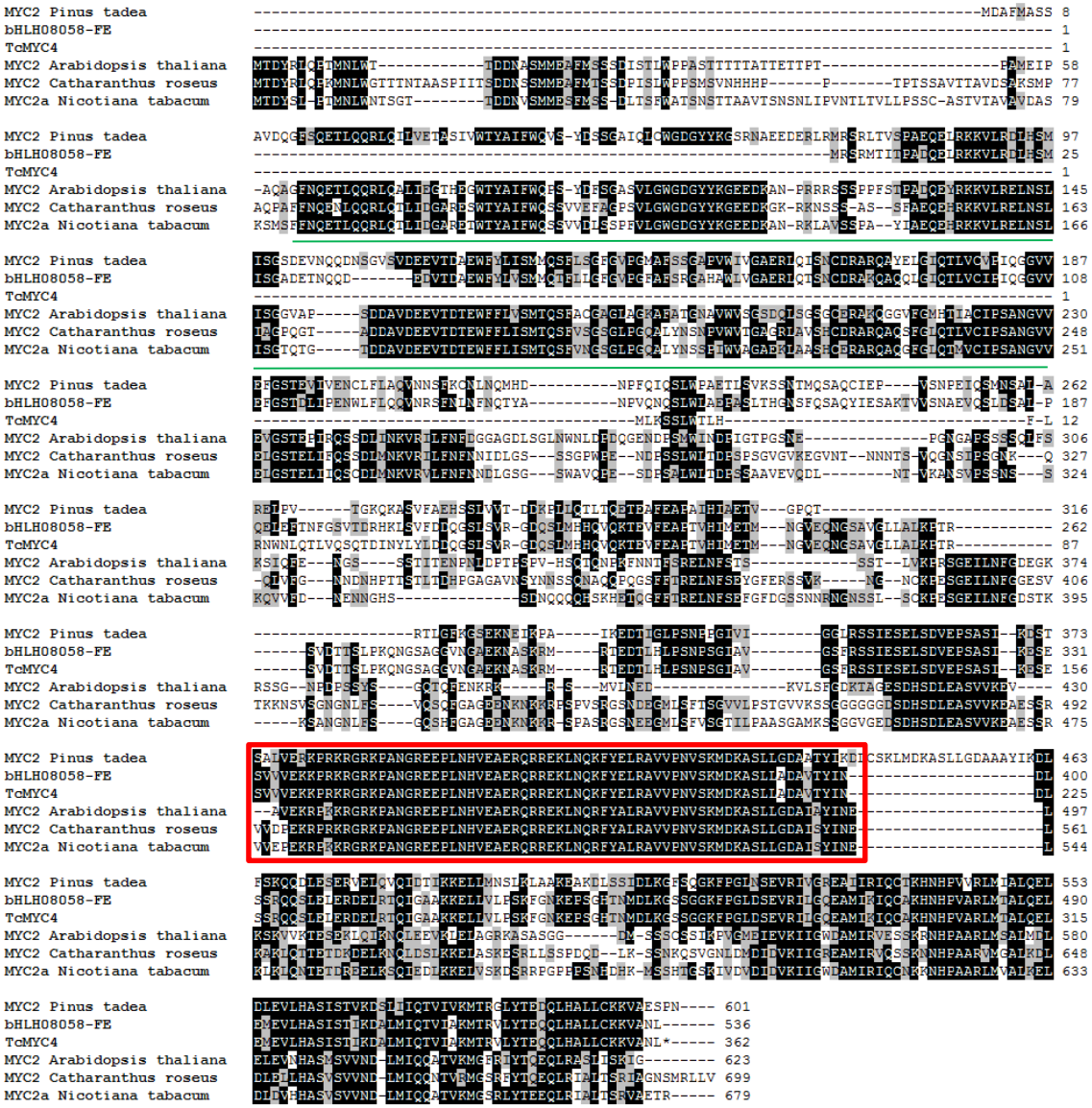


Figure 25 – The protein alignment of the further extended version of bHLH08058 (bHLH08058-FE), the Walker TcMYC4 (184)**, published examples of MYC2 in *Arabidopsis thaliana, Catharanthus roseus* and *Nicotiana tabacum* and an identified MYC2 homolog in *Pinus tadea*.** The alignment was produced using ClustalOmega (231, 232) and the image produced using BioEdit v7.2.5 (233). All the transcription factors have a high level of similarity in the C terminal bHLH domain highlighted by the red box. bHLH08058-FE also has high level of similarity at the N-terminus with MYC2 homologs in other species, whereas TcMYC4 is missing this region that has been shown to be important in JAZ protein interaction (230) (highlighted with a green line).

2.5 Investigating whether the 19 candidate TFs are primary response genes

Primary response genes are defined as those that are able to respond to a signal without *de novo* protein synthesis (174). Cycloheximide (CHX) is an inhibitor of eukaryotic translation and is a well-known reagent used to inhibit protein synthesis. It blocks the elongation phase of translation, preventing peptidyl transfer from tRNA to the ribosome. CHX was therefore used to determine whether the 19 candidate TFs were primary response genes. The induction of the 19 TFs, except MYB12379, at an early time point after MeJA elicitation was not inhibited by the addition of CHX, showing that they are primary response genes (Figure 26). Identifying a secondary response gene as a control proved slightly problematic as many of the controls used in

other papers could not be identified in the *T. cuspidata* CMC transcriptome. LOX2 (lipoxygenase 2), a JA biosynthetic enzyme, was shown by Chung *et al.* in Northern blot experiments to be a secondary response gene with significantly reduced levels of mRNA transcript after CHX addition. Our results used LOX2 as a relaxed control to show that the induction in gene expression after MeJA elicitation was delayed by the addition of CHX.

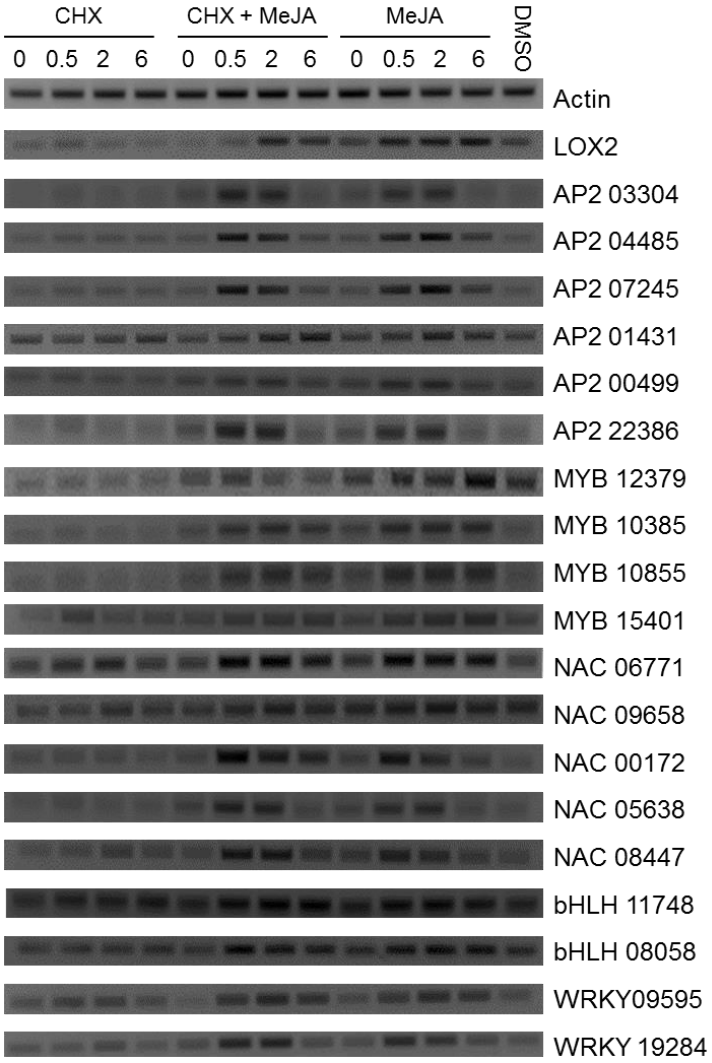


Figure 26 – The effect of cycloheximide (CHX) treatment on jasmonate acid (JA) responsive transcription factors (TFs) in *Taxus cuspidata* cambial meristematic cell (CMC) culture. T. cuspidata CMC culture was treated 5 days post sub-culture with either a mock control (0.2% DMSO), 100 µM methyl jasmonate (MeJA), 100 µM CHX, or a combination of MeJA and CHX (both 100 µM). Cultures were pre-treated with CHX for 1.5 h before elicitation with MeJA. 2ml of culture was collected at the indicated times post addition of MeJA. RNA was extracted using a CTAB based method, converted into cDNA and used in RT-PCR analysis with appropriate primers (see section 8.4 for further details). The house keeping gene actin and a known secondary response gene LOX2 (lipoxygenase 2, a JA biosynthetic enzyme) were used as controls. Actin expression was high under all conditions tested and the activation of LOX2 expression by MeJA was delayed after CHX addition. The RT-PCR analysis showed that all the 19 candidate TFs, with the exception of MYB12379, are primary response genes as their expression was not affected by the addition of CHX.

2.6 Prediction of the cognate binding sites for the 19 candidate TFs in the paclitaxel biosynthetic promoters

To regulate gene expression TFs bind to specific motifs within promoters. The known binding sites (BSs) of the 5 TF families identified as highly up-regulated at an early time point (0.5

h) after MeJA elicitation are summarised in section 1.4 (Table 5*).* The putative BSs for AP2, MYB, NAC, bHLH and WRKY were identified in the 10 paclitaxel promoters available on the NCBI with the aid of the PLACE (Plant *cis*-acting regulatory DNA elements) tool (234) and are summarised in Table 9. The location of these BSs is represented in Figure 27. Every promoter has a putative cognate BS for the 5 TF families under investigation, except for *PAM* which does not contain a W-box. This analysis showed that there are a large number of potential BSs that the 19 candidate TFs could interact with in the promoters of the paclitaxel biosynthetic pathway genes.

TF Family	DNA motif	Frequency in Promoter									
		TASY	*T5αH*	*T13αH*	*TDAT*	*T10ϐH*	*DBBT*	*DBAT*	*PAM*	*BAPT*	*DBTNBT*
AP2	GCC Box - GCCGCC/ $GCCN_3GGC$	1	0	0	1	0	1	0	0	0	1
	DRE element - GCCGAC	0	0	0	0	0	0	0	0	0	0
	JERE motif CTCTTAGACCGC CTTCTT	3	3	0	2	1	0	0	0	1	10
	RAV - CAACA/CACCTG	0	0	0	2	0	0	0	0	0	1
MYB	CNGTT(A/G) consensus	3	1	4	3	0	2	1	0	5	3
	ACC(A/T)A(A/C) consensus	1	0	3	5	3	1	9	1	3	2
	TTAGGG motif	0	1	0	0	1	0	0	0	1	0
	TATCCA motif	2	1	0	4	1	0	0	0	3	1
	GATA motif	16	9	10	11	8	7	10	3	14	12
NAC	CGT[G/A] motif	11	11	9	16	13	6	6	3	3	11
bHLH	E-box -CANNTG	7	7	10	9	6	6	8	3	5	9
WRKY	W- box TGAC(C/T)	12	4	6	10	4	5	10	0	11	3

Table 9 – Summary of the binding site (BS) analysis of 10 paclitaxel biosynthetic promoters. BSs were identified with the aid of the PLACE (Plant *cis*-acting regulatory DNA elements) tool (234)**.** See Table 2 for paclitaxel biosynthetic promoter abbreviations and Table 5 for further information on the BSs. The name of the BS and the consensus motif are stated for each transcription factor (TF) family. The number of times a binding site occurs in a promoter in the 5' and 3' direction is stated. The paclitaxel promoters are rich in cognate BSs of the 19 candidate TFs, therefore we can hypothesised that these TFs may regulate the paclitaxel biosynthetic pathway.

The sequence surrounding the recognised binding motif is important in determining binding specificity, because it is unlikely that a TF will bind to all of its identified cognate BSs. There is little species specific information about BSs in *Taxus spp.* as only two papers have explored the binding specificity of *Taxus* TFs (183, 184). The PLACE database contains *cis*–acting elements from published reports in all vascular plants; consequently BSs might have been erroneously identified based on species that are evolutionarily distinct to *Taxus.*

The well-established binding motif of AP2 and ERF TFs, the GCC-box motif, is not well represented in the paclitaxel promoters with only two promoters *TASY* and *TDAT* containing the GCCGCC motif and no promoters contain the DRE core sequence (Table 9) (115, 116, 131). Just over half of the promoters contain part of the jasmonate- and elicitor-responsive element (JERE) motif identified as the region ORCA3 binds to in the promoter of the TIA biosynthetic gene strictosidine synthase (176). The gene expression of paclitaxel genes is induced by MeJA

elicitation therefore it correlates that the BSs in the paclitaxel biosynthetic promoters are likely to be JA responsive.

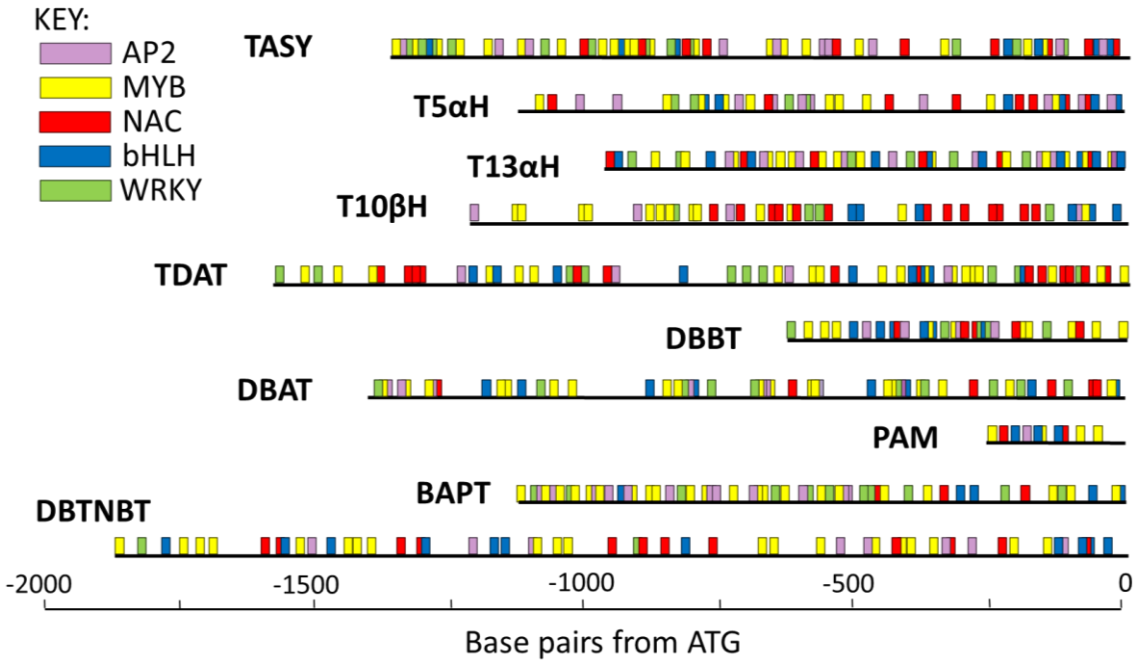


Figure 27 – A schematic representation of the binding sites (BSs) for the AP2, MYB, NAC, bHLH and WRKY transcription factor (TF) families in 10 paclitaxel biosynthetic promoters. See Table 2 for paclitaxel biosynthetic promoter abbreviations and Table 9 for the number of BSs in each promoter. BSs are indicated by boxes, colour coded based on TF family, AP2 purple, MYB yellow, NAC red, bHLH blue and WRKY green. The image shows the length of each promoter and location of BSs relative to the start codon (ATG). The paclitaxel promoters are rich in cognate BSs for the 19 candidate TFs and are located throughout the promoter.

The MYB TF family is highly expanded in plants compared to animals and its members do not always share binding specificity based on their structural similarity. The DNA BSs of 87 plant MYB proteins have been reported in the literature and these span a number of different motifs, therefore complicating the prediction of possible binding motifs. Many R2R3–MYB TFs recognise AC elements which are enriched in adenosine and cytosine residues and can act as activators and repressors at these sites (123). All of the promoters contain the GATA motif, while 90% contain the ACC(A/T)A(A/C) motif and 80% the CNGTT(A/G). There are likely to be more unidentified BSs as these motifs are not well defined. The BS of NAC, bHLH and WRKY TFs are more defined, binding to CGT[G/A], E-box and W-box respectively; and there are a large number of these were identified in every promoter – except a W-box in *PAM*. The large number of cognate BSs identified in the 10 paclitaxel promoters strengthens the hypothesis that the 19 candidate TFs may be involved in directly regulating paclitaxel production.

2.7 Analysis of possible post-translational regulation of TFs

JA can also regulate gene expression post-translationally using phosphorylation cascades. In *N. tabacum* the activity of ORC1 and MYC-like proteins is increased by JAM1, a JA stimulated mitogen-activated protein kinase kinase (MAPKK). Mitogen-activated protein kinase (MAPK) cascades are also implicated in regulating WRKY33 expression in *A. thaliana* (99, 100), NtMYB2

in *N. tabacum* and PtMYB4 in *P. taeda* (125). A MAPK and MAPKK were identified in DEGs that were significantly up-regulated after MeJA elicitation (Figure 28). The highest logFC for these proteins is at an early time point (0.5 h) after MeJA elicitation suggesting their possible involvement in regulating the 19 candidate TFs.

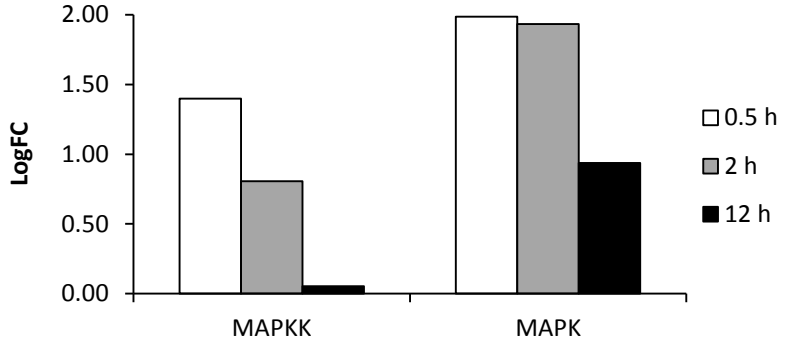


Figure 28 - The log fold change (LogFC) in gene expression of mitogen-activated protein kinase kinase (MAPKK) and mitogen-activated protein kinase (MAPK) at three time points 0.5, 2 and 12 hour (h) after methyl jasmonate (MeJA) elicitation. These kinases are activated at an early time point after MeJA elicitation and are involved in phosphorylation cascades that affect protein activity. It can therefore be hypothesised that these kinases might regulate the function of the 19 candidate transcription factors along with other possible post-translational controls.

2.6 Discussion

MeJA is known to trigger a transcriptional cascade in plants (114). The effect of MeJA on the *T. cuspidata* transcriptome was investigated by collecting cells at 3 time points (0.5, 2 and 12 h) after MeJA elicitation and analysing their transcriptomic profile using Illumina sequencing. 1646 contigs were identified as differential expressed at least one time point tested compared to 0 h using EdgeR software, which used a modified version of Fisher's exact test with the *p*-values adjusted for the false discovery rate of ≤0.05 (Figure 12).

MeJA triggers a transcriptional cascade in *T. cuspidata* CMCs with the number of DEGs increasing between 0.5 h and 12 h, as observed in *A. thaliana* (114). A greater number of genes were differentially expressed at 0.5 h in *T. cuspidata* compared to *A. thaliana* suggesting the CMC response to MeJA is more complex, propably because *Taxus spp.* produce a wider variety of secondary metabolites. In *A. thaliana* at an early time point after MeJA elicitation the initial wave was purely inductive, whereas the *T. cuspidata* DEGs were predominately induced but a large number of down-regulated genes were also identified. The transcriptional cascade leads to a significant increase in the number of down-regulated genes at 12 h and in both species 60% of genes that were only differentially expressed at a late time point were down-regulated. This suggests that the plant is switching off pathways that are not required for defence and channelling energy into others.

Paclitaxel as a terpenoid requires the production of the isoprenoid precursors IPP and DMAPP. Analysis of the MVA and MEP pathways showed that transcript levels increased for all the genes involved but not significantly; whereas all the genes of the paclitaxel biosynthesis pathway were significantly up-regulated after MeJA elicitation. There may be a small increase in the pool of precursors available but the increase in paclitaxel yields is therefore more likely to be

due to the shunting of IPP, DMAPP and other terpenoid precursors away from competing pathways. Metabolic reprogramming requires large amounts of energy but of the 8 tricarboxylic acid (TCA) cycle enzymes (all of which were identified in the CMC transcriptome) only two, citrate synthase and malate dehydrogenase, had significantly increased expression after MeJA. This conflicts with data found by Sun *et al.* in *T. media* where TCA cycle genes were decreased in abundance after MeJA addition but concurs with results in *A. thaliana* where 5 TCA related genes were up-regulated (114, 217). An overall study exploring the changes in primary and secondary metabolism after MeJA elicitation would help elucidate the flow of energy and precursors. MAPMAN software provides this platform in *A. thaliana*, however as a non-model organism without a sequenced genome it is currently very difficult to perform this analysis in *Taxus spp*.

The crosstalk between different hormone signalling pathways is important in the plants response to stimuli. MeJA initiates a large transcriptional cascade but this does not occur in isolation. Addition of hormones or hormone inhibitors can alter the levels of paclitaxel production. The addition of 10 ppm (part per million) of ethylene increases paclitaxel production but the presence of 50 ppm inhibits synthesis (90). Three ethylene biosynthesis genes were identified as up-regulated after MeJA elicitation and the effect that this increase in hormone production might have on paclitaxel levels is unknown. Knowledge of the mechanisms employed by different hormones to act synergistically or antagonistically with MeJA could be exploited to increase paclitaxel production.

The antagonistic action of auxin and MeJA can be seen in the significant down regulation of ARFs, auxin transporters and two auxin related biosynthesis genes. The auxin-like phytohormones 2,4-D and picloram are components of the *Taxus* culture media even though auxin down-regulates paclitaxel production. These chemicals are added to the media to improve cell growth as MeJA retards cell culture growth (217); and in tobacco BY-2 cells MeJA has been shown to repress cell cycle genes, in particular those involved in metaphase (235). A balance therefore needs to be struck between the highest possible yields of paclitaxel and the growth rate of the culture and crosstalk between MeJA and auxin is likely to play an important role.

Gibberellic acid (GA) is also involved in regulating growth responses and a component of the *Taxus* cell medium required for growth. 3 GRAS TFs were identified in the DEGs and were all up-regulated after MeJA addition, with GRAS21372 having the highest logFC value at 0.5 h. Two GA biosynthesis genes were also significantly up-regulated after MeJA elicitation. In *A. thaliana* GA and MeJA were found to act synergistically in trichrome development (236). However, GA signalling is known to antagonise JA signalling as DELLA proteins can interact with JAZ preventing MYC2 repression, but the DELLA proteins are degraded in the presence of GA (237). It is therefore important to establish how these hormone pathways crosstalk so that the knowledge can be exploited in bioengineering efforts to increase paclitaxel production in CMCs – perhaps by knocking out antagonistic pathways – but at the same time retaining their superior growth properties, as the hormones found to work against JA are mostly involved in plant growth.

The core JA signalling module was identified in the transcriptomic data as previously observed by Li *et al.* (210). This shows that the JA elicitation pathway is conserved across angiosperms and gymnosperms and that research from other species can be used to find regulators of the paclitaxel biosynthetic pathway. Known regulators have been employed by other groups to find *Taxus* TFs. Three TcJAMYCs were identified using AtMYC2 degenerate primers and TcAP2 was found using the JERE element (which ORCA3 binds with) as bait in yeast-1-hybrid experiments (184, 220). Caution should be taken, however, when employing reverse screen approaches as TFs involved in secondary metabolism are highly species specific (100). MYC2 is well known to act as a transcriptional activator in *A. thaliana* and when TcJAMYC1 was first published in 2009 it was shown to act as an activator, but the paper was subsequently retracted. In 2015 TcJAMYC1 was republished but was now shown to act as a repressor which correlates with our experimentation (Figure 46). This shows the dangers of bias that can occur when looking for homologs of known JA responsive TFs. The identification of our 19 candidate TFs tries to avoid some of this bias by choosing TFs that were significantly up-regulated at 0.5 h. However TF were chosen from families that have been previously reported to be involved in regulating secondary metabolism. TFs with high logFC after MeJA addition include GRAS and LOB TFs, whose families have been linked with plant growth and development (111, 238). However, first concentrating on TFs from the families AP2, MYB, NAC, bHLH and WRKY that have already been connected with secondary metabolism increases the probability of identifying an overall regulator of the paclitaxel biosynthetic pathway (100).

The importance of identifying the full ORF of the TFs of interest was exemplified by the issues confronted in determining the bHLH08058 ORF. Original research by Amir, Yan and Kwon identified bHLH08058 as a protein containing only 225 amino acids. However due to the publication of a longer version, TcMYC4 (361 amino acids), further investigation was conducted which identified an even longer version of the TF, 536 amino acids in length (bHLH08058-FE). The original analysis identified bHLH08058 as a MYC2 homology, therefore the fact that the protein was significantly shorter than those in the literature and was missing key domains shown to be important in JAZ interactions could have suggested that this is not the full ORF (230). This work does show the difficulties of working in an organism without a sequenced genome as the transcriptome has to be assembled *de novo* and errors can occur during the sequencing process. Roche 454 has been shown to have a high error rate in homopolymer regions. These are regions of three or more identical DNA bases that lead to increased light emission making it difficult to correctly determine the DNA sequence. Long homopolymers result in either insertions or deletions errors in the DNA sequence that affect the reading frame (239, 240).

The 19 TFs, except MYB12379, were shown to be primary response genes, as their induction in gene expression after MeJA elicitation was not inhibited by the protein synthesis inhibitor CHX. This concurs with results of other MeJA inducible TFs involved in regulating plant secondary metabolism, for example the induction of ORCA3 and CrMYC2 in *C. roseus* (174) and

JAZ and MYC2 in *A. thaliana* (241) were not reduced by the addition of CHX. This helps to strength our hypothesis that these TFs are overall regulators of the pathway as they are induced at an early time point after MeJA elicitation and do not require *de novo* protein synthesis to respond to MeJA.

A large number of predicted cognate BS for the 5 TF families were identified in 10 paclitaxel biosynthetic promoters. This analysis strengthens the hypothesis that these TFs may regulate paclitaxel biosynthetic genes. Some of the BSs are more difficult to predict, for example MYB TFs can bind to a large variety of different nucleotide sequences. An *in vivo* study is required to screen for interactions between the 10 paclitaxel biosynthetic promoters and the 19 candidate TFs. The paclitaxel biosynthetic pathway is believed to contain 19 steps, 14 of which have been characterised however, the promoter regions for only 10 genes have been identified so far (26, 47). These known promoters are spread across the pathway with equal number in the early and late stages of the pathway; therefore providing, if not a complete picture of the pathway, a well-represented one.

Chapter 3 - An *in vivo Arabidopsis thaliana* transient expression assay to investigate the regulation of *Taxus cuspidata* CMC promoters by the 19 candidate TFs

3.1 Introduction

Transcription factors (TFs) can regulate a vast array of different cellular processes through binding specific DNA motifs. Analysis of the paclitaxel biosynthetic promoters found that they were rich in the cognate binding sites of the candidate TFs (Figure 27 and Table 9). An *in vivo* system was required to screen the 19 candidate TFs for possible interactions with the 10 available paclitaxel biosynthetic promoters. *Taxus spp.* transformation is currently difficult with few examples of successful transformation reported in the literature; therefore a transient expression assay in *A. thaliana* protoplasts was used to perform the screen.

3.1.1 *Arabidopsis thaliana* protoplast transient expression assay (TEA)

3.1.1.1 Protoplasts

Protoplasts are produced by removing the plant cell wall using a cocktail of enzymes including cellulase and macerozyme. Removing the barrier of the plant cell wall permits DNA to be transfected into the cell using techniques such as polyethylene glycol (PEG)-mediated, electroporation and microinjection (242, 243). Protoplasts were originally isolated from tomato seedlings 55 years ago by Cocking, with the first transfection achieved 10 years later by Aoki and Takebe using tobacco mesophyll protoplasts (244, 245). Since then protoplasts have been successfully used to observe an array of cellular processes such as photosynthesis, respiration, cell division, calcium signalling and the regulation of ion channels in a variety of plant species (243). They have also been used to dissect the function of *cis*-elements and TFs in numerous pathways including the JA mediated *A. thaliana* TFs ORA47 & MYC2 (114) and the pathogen induced Pti4, Pti5 and Pti6 from tomato (130).

The isolation of protoplasts from a large range of species and tissues has been published, however *A. thaliana* was chosen due to the existence of well-established protocols in the literature and the ease of growing large quantities of plant material. The isolation of *Taxus* protoplasts has been published (246, 247), but to date successful transfection has been elusive. Work previously conducted by a previous group member (Waibel (221)) attempted to produce a protocol, however the transfection frequency was insufficient for further experimentation (221). *A. thaliana* mesophyll protoplasts have been used by numerous groups to investigate TFs from non-model species (243, 248), which was why it was chosen as the model system.

3.1.1.2 Advantages and Limitations of a TEA

The advantages of using a TEA include that it can be used for high throughput screening

of closely related gene families. The transient nature of the experiments mitigates the need for mutants to be created or the environment to be kept sterile. Mesophyll protoplasts are genetically stable and the process for obtaining a homogenous culture is quick, with numerous comprehensive protocols available (242, 249). The system also permits co-transfection of multiple plasmids and observations can be made within 24 h (243). The drawbacks are that testing will be completed in *A. thaliana* rather than *Taxus spp.* There could be co-factors or post-translational modifications that are not present in *A. thaliana* which could lead to false positive and negative results. The transfection efficiency can be variable and over 50% efficiency is required for reproducible data, therefore experiments must be completed with numerous replicates. Establishing the assay is also time consuming as there are many factors requiring optimisation (Table 10) (250).

Conditions	
Plant Growth Conditions	Day length Age
PEG	Concentration (%) Transfection time
Transfection	Incubation length Methodology Protoplast/DNA ratio
DNA	Concentration Purity Ratio of constructs
Protoplast	Isolation method Concentration

Table 10 – A summary of the experimental conditions that need to be investigated to produce a reproducible *Arabidopsis thaliana* protoplast transient expression assay protocol. PEG is polyethylene glycol which is the solution that permeabilizes the membrane allowing DNA transfection. The large number of conditions that require optimisation can make the assay time consuming to establish.

3.2 Establishing a TEA in *Arabidopsis thaliana* protoplasts

3.2.1 Experimental Design

To investigate the possible interaction between the 19 candidate TFs and the 10 paclitaxel biosynthetic promoters three vectors were required: an effector with the TF under a constitutive promoter, a reporter with the promoter of interest driving a reporter gene such as green fluorescent protein (GFP), β-glucuronidase (GUS) or luciferase and an internal control with a reporter gene under a constitutive promoter (Figure 29). The internal control is vital, due to the variation in transfection efficiency the reporter data must be normalised to the internal standard.

Luciferase was chosen as the reporter gene due to its high sensitivity and the ability to quickly obtain quantitative data using non-hazardous reagents (250). Plants have little or no endogenous luciferase production reducing the possible background, whereas they do contain enzymes that can mimic GUS activity (251). A dual luciferase assay (Promega®) was chosen using firefly (*Photinus pyralis)* luciferase in the reporter plasmid and *Renilla (Renilla reniformis)* luciferase in the internal control (Rluc) (252). It is possible to discriminate between the two luciferases due to their distinct evolutionary origin and dissimilar enzyme structure. Examples of the dual luciferase assay being used successfully with *A. thaliana* protoplasts to investigate TFs including Pti4, Pti5

and Pti6 (130) and ABI4 (253).

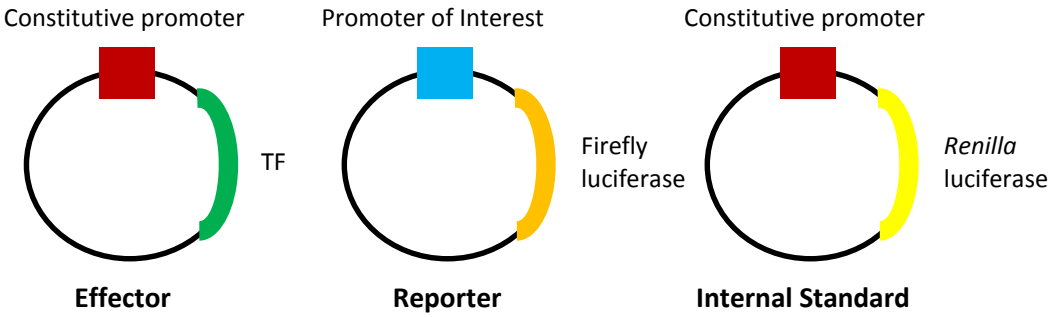


Figure 29 – The design of the three constructs required for the *Arabidopsis thaliana* protoplast transient expression assay. The effector construct has the transcription factor of interest (TF) under the control of a constitutive promoter, such as p35S. The second control is the reporter plasmid which has the promoter under investigation driving the reporter gene firefly luciferase. When transfected into protoplasts the effector will produce the TF and if it can interact with the promoter it will drive the expression of firefly luciferase. The third construct is an internal standard with the reporter gene *Renilla* luciferase under the control of a constitutive promoter, such as p35S. The internal standard acts as control against the variable transfection efficiency and the firefly luciferase values obtained are normalised to the internal control.

The vectors employed in this assay need to be small, < 6 kb if possible, because they yield higher DNA concentrations in *E. coli* and produce a larger signal compared to binary T-DNA vectors (170). The high copy number Gateway™ compatible plant expression vector p2GW7,0 (254) was chosen for the effector and internal standard constructs, while pGWlucB was used for the reporter plasmid (255) (Appendix 1). The known interaction of the *A. thaliana* AP2 TF ORA47 with the promoter of the JA biosynthetic gene lipoxygenase 3 (*LOX3*) was used as a positive control to establish the assay (114).

3.2.2 Optimization of *A. thaliana* Protoplast Isolation Protocol

Two methods, the Sheen protocol (249) and tape sandwich method (242) were evaluated for isolating protoplasts (see section 8.6.2 for methodology). The yield and quality of the protoplasts isolated from plants grown under long day conditions were assessed using a haemocytometer (Table 11). The tape sandwich method was far superior at producing high quality yields of protoplasts with little debris.

Age (weeks)	Total number of protoplasts	
	Sheen	Tape sandwich
5	4.7×10^5	6.1×10^5
4	5.5×10^5	1.4×10^6
3	3.8×10^5	9.9×10^5

Table 11 – The comparison of protoplast yield between the Sheen (249) **and tape sandwich** (242) **protoplast isolation protocols (see section 8.6 for methodology) with long day plants of varying age.** The tape sandwich method produced higher yields of protoplasts compared to the Sheen protocol at all ages tested.

The production of protoplasts using the tape sandwich method was then investigated using plants of different ages and growth conditions (long *vs* short day) (Table 12). Traditionally short day conditions have been used to isolate protoplasts for transfection however recent research has

suggested plants grown under long day conditions are also suitable for protoplast transfection (242). Both conditions produced healthy protoplasts with the optimal age for long day plants of 3–4 weeks, whereas short day plants required 5–6 weeks. However due to issues with the long day conditions including aphid and powdery mildew infections, short day conditions were used for all further experimentation. Protoplasts isolated from stressed plants often experience low transfection efficiencies (249).

Long Day		Short Day	
Age (weeks)	**Protoplasts per gram**	**Age (weeks)**	**Protoplasts per gram**
5	5.17×10^6	7	5.92×10^6
4	8.84×10^6	6	7.32×10^6
3	7.44×10^6	5	6.63×10^6
		4	1.03×10^7

Table 12 – The evaluation of protoplast yield using the tape sandwich method (242) **under long and short day conditions with plants of varying age (see section 8.6 for methodology).** The yield of protoplasts produced does not differ significantly between growth conditions. The optimum yield of protoplasts isolated from long day plants was four weeks and six weeks for short day plants.

3.2.3 Optimization of *A. thaliana* Protoplast Transfection Protocol

Initially two protocols (A and B), see section 8.6.3 for methodology, were investigated for transfecting protoplasts, however neither was successful. Numerous conditions were explored to solve the problem including: plant age, incubation time, PEG % and transfection time. The survival of the protoplasts was monitored under a variety of PEG transfection conditions and it was concluded that they were surviving the transfection. However after 6 h in the incubation solution (WI) in method A all the protoplasts had burst, whereas protoplasts left in the incubation solution from method B (Modified W5) were still viable. The difference between these solutions is that WI has a high mannitol concentration, which could be placing an osmotic stress on the protoplasts. Modified W5 solution was used in subsequent transfections to ensure protoplast survival.

A further literature search led to the testing of method C, which suggested that the pH of the PEG solution needed to be altered to 8–9 and this change led to successful transfection (Figure 30a). A possible reason for this observation is that as PEG ages it is oxidised by the air and becomes more acidic, increasing its ionic strength (256). The pK_a of ethylene glycol is 14.5 therefore you would expect PEG to be a weak base, however the pH of the PEG solution was 3. It is possible that the increased acidity makes the PEG positively charged reducing its ability to interact with the membrane. The mechanism by which PEG permits the entry of DNA into a cell is unknown, making speculation about the specific effect of the change in pH difficult.

A number of conditions were then optimized including: PEG transfection time (5 min), PEG pH (pH 8) and protoplast incubation conditions (light) (Figure 30b-d). The source of calcium in the PEG solution ($CaNO_3$ or $CaCl_2$) and the age of the plant material had little effect on the transfection efficiency, provided the age of the plant material was within the range of 4–6 weeks.

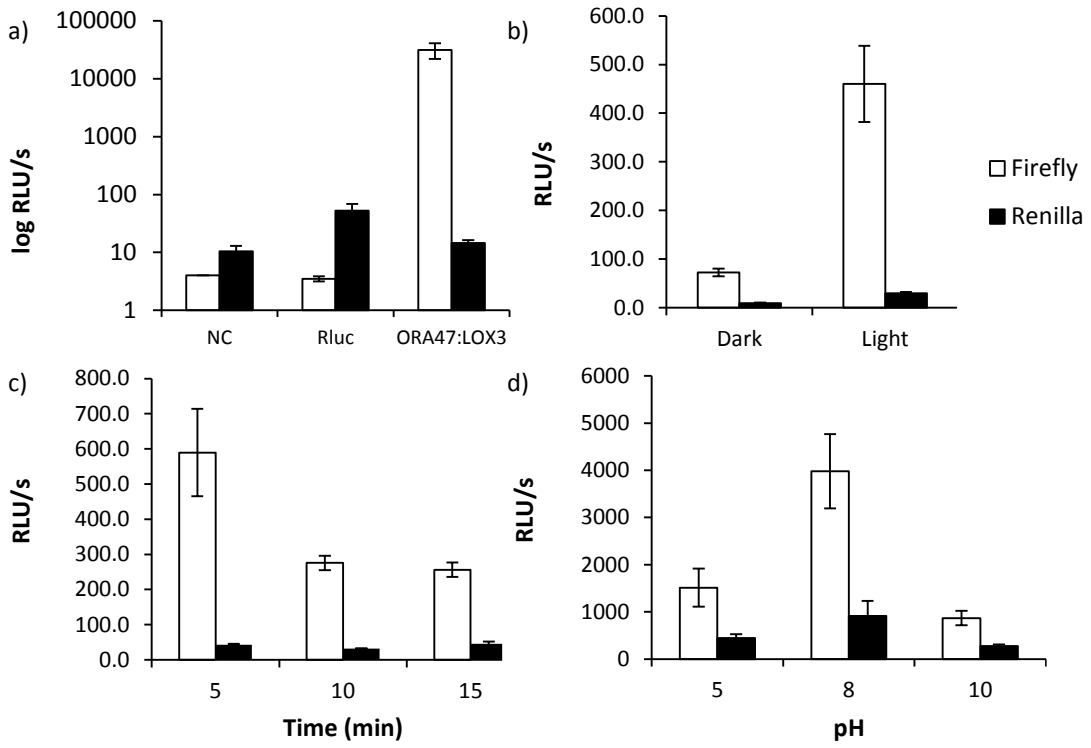


Figure 30 – The optimisation of transfection conditions of the *Arabidopsis thaliana* protoplasts transient expression assay (TEA). a) The successful transfection of *A. thaliana* protoplasts using Method C (see section 8.6.3 for methodology). NC is the negative control where no DNA was transfected, Rluc was transfection with the internal standard containing *Renilla* luciferase, ORA47:LOX3 is the known interaction of *A. thaliana* AP2 TF ORA47 with the jasmonic acid biosynthesis gene promoter lipoxygenase 3 (*LOX3*) (114). Transfection using Method C was successful as significant firefly luciferase was observed when LOX3 and ORA47 were co-transfected compared to NC. The ORA47:LOX3 interaction was used to optimise the transfection conditions testing b) incubation conditions, c) Polyethylene glycol (PEG) transfection time and d) PEG pH. The best conditions identified were light incubation conditions, a PEG incubation time of 5 minutes and a PEG solution with a pH of 8. RLU stands for relative luminescence units with the different values for firefly and *Renilla* luminescence shown. Error bars represent standard deviation, with $n \geq 3$.

The observed *Renilla* relative luminescence unit (RLU) values were far reduced compared to Firefly and could sometimes be too low to allow for further analysis. Attempts to increase the internal control by using a higher amount of DNA led to an increase in *Renilla* values (Figure 31a), however if the total amount of DNA was increased beyond 20 µg the firefly expression was reduced (Figure 31b). A number of different ratios between the 3 constructs were also tested to try and improve the levels of firefly and *Renilla* expression (Figure 31c), however the results were not ideal. Originally the internal control (Rluc) had been cloned into the vector p2GW7,0 using traditional cloning with restriction enzymes *Spe*I and *Sac*II; but a new control was designed using Gateway™ cloning which introduced a 20 bp sequence up-stream of the ATG that contains a Shine-Dalgarno and Kozak sequence. The Kozak sequence is a ribosome binding site that lies close to the ATG in eukaryotes improving translation initiation (257). This led to significantly improved *Renilla* expression in protoplasts and further ratio testing found that a ratio of 4:5:5 (effector:reporter:control) produced the best results (Figure 32a).

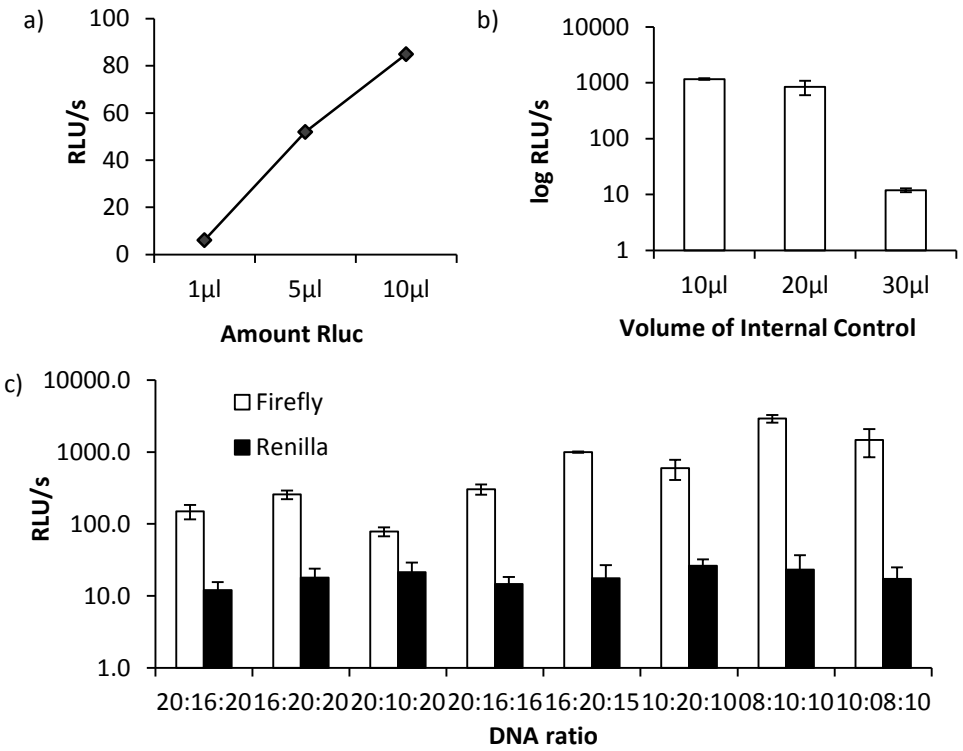


Figure 31 – Optimisation of the amount and ratio of DNA constructs used in *Arabidopsis thaliana* protoplasts transient expression assay (TEA). a) The addition of increasing amounts of the internal control (Rluc) increases *Renilla* expression, but b) increasing the amount of Rluc in each transfection beyond 20 µg reduced firefly expression. a) shows the change in *Renilla* expression, while b) shows the change in Firefly expression with the addition of different amount of Rluc (1µg/µl). c) The effect of different DNA ratios between the three constructs, transcription factor:promoter:Rluc, on observed luminescence. The ratios are expressed as µg DNA Values are in relative luminescent units (RLU), error bars represent standard deviation, n≥3.

A final important factor that dramatically affects transfection efficiency is the purity of the DNA (249). If the DNA is not highly pure minimal expression was observed (Figure 32b), therefore it is essential that the DNA extracted is extremely clean. After extensive optimization a robust protocol, Method D, was used to test for the interactions between the 19 candidate TFs and the 10 paclitaxel promoters. The DNA purity and pH of the PEG solution were identified as factors critical in achieving high transfection efficiency.

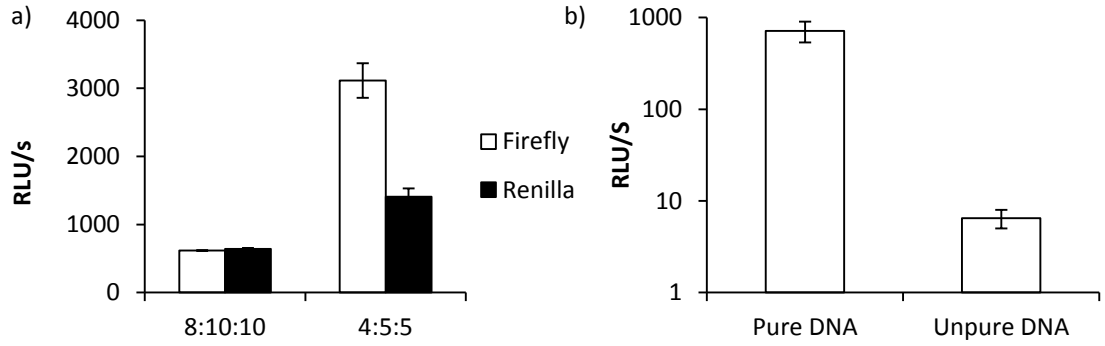


Figure 32 – Further optimisation of transfection efficiency testing a) DNA ratios and b) DNA purity. The DNA ratios between the three constructs, transcription factor:promoter: internal standard, with numbers indicating µg DNA. Values are in relative luminescent units (RLU), error bars represent standard deviation, *n*≥3. A reduction in the amount of DNA used in each transfection produced higher firefly and *Renilla* values. The purity of the DNA is important, with luminescence being lost if the DNA is unpure.

3.3 Testing for interactions between the 19 candidate TFs and 10 paclitaxel promoters

3.3.1 Cloning TFs and promoters for the TEA assay

The candidate TFs were cloned into the pENTRY vector pDONR221 and subsequently into the pDEST vector p2GW7,0 (see Table 21 for primers and Appendix 1 for vector map). The sequences were confirmed using Sanger sequencing (Genepool, Edinburgh UK). 10 promoter sequences were avaliable on the NCBI from a variety of *Taxus spp.* and these were cloned into pDONR221 and then into pGWlucB (see Table 22 for primers and Appendix 1 for vector map). Certain seqeunces, such as *T10βH*, *T5αH* and *BAPT* required truncation at the 5' and 3' end before they were successfully cloned from *T. cusipidata* CMC gDNA. The promoter *DBTNBT* was broken into two fragments due to difficulties encountered when attempting to amplify the full length; *DBTNBT-2* is the first 950 bp from the ATG and *DBTNBT-1* is the remaining 840 bp. Table 13 summarises the identity between the seqeunces available on the NCBI and those amplified from *T. cuspidata* CMC gDNA. There is a high percentage identity between the nucleotide sequences with most alterations in the sequence not affecting the BSs. The promoter with the lowest sequence identity is *BAPT*, but the NCBI seqeunce is connected to the Nims *et al* paper (258) which was retracted, therefore the differeneces between the sequences may be due to poor sequencing conducted in this paper.

Promoter	Source	Length (bp)	Identity (%)	Notes
TASY	**NCBI**	**1433**	**99.9**	**No truncation**
	CMC	1434		Inserted C – does not affect BS
T5αH	NCBI	1215	97.9	Truncated 21 bp 5' and 20 bp 3' end
	CMC	1166		6 differences between 747-855, introduces one Ebox
T13αH	NCBI	934	97.1	Truncated 33 bp 5' and 36 bp 3' end.
	CMC	893		60bp insert at -78bp, changes do not effect BS
TDAT	NCBI	1611	99.6	No truncation
	CMC	1617		Changes do not affect BS
T10βH	NCBI	1222	94.5	Truncated 79 bp 5' and 23 bp 3' end
	CMC	1113		Region 717-891 numerous differences that do affect BS
DBBT	NCBI	759	98.1	Truncated 53 bp 5'end
	CMC	618		One change effects a BS but doesn't remove
DBAT	NCBI	1408	99.8	No truncation
	CMC	1409		Inserted base in front of NAC BS
PAM	NCBI	257	99.2	No truncation
	CMC	256		Changes do not effect BS
BAPT	NCBI	1239	91.8	Truncated 62 bp 3' end
	CMC	1187		Numerous differences that do affect BS, leading to introduction and loss of BS
DBTNBT	NCBI	1898	97.4	Truncated 100 bp 5' end
	CMC	1791		Numerous differences that do affect some BS

Table 13 – Comparison of promoter sequences amplified from *Taxus cuspidata* CMCs genomic DNA and the sequence available on the NCBI. The length of the promoter amplified and the NCBI sequence are stated with the percentage identify between the two sequences. Most sequences are above 97% except *T10βH* and *BAPT*. Any truncations that were required to amplify the promoter sequences and if the differences between the sequences affect a binding site (BS) are stated.

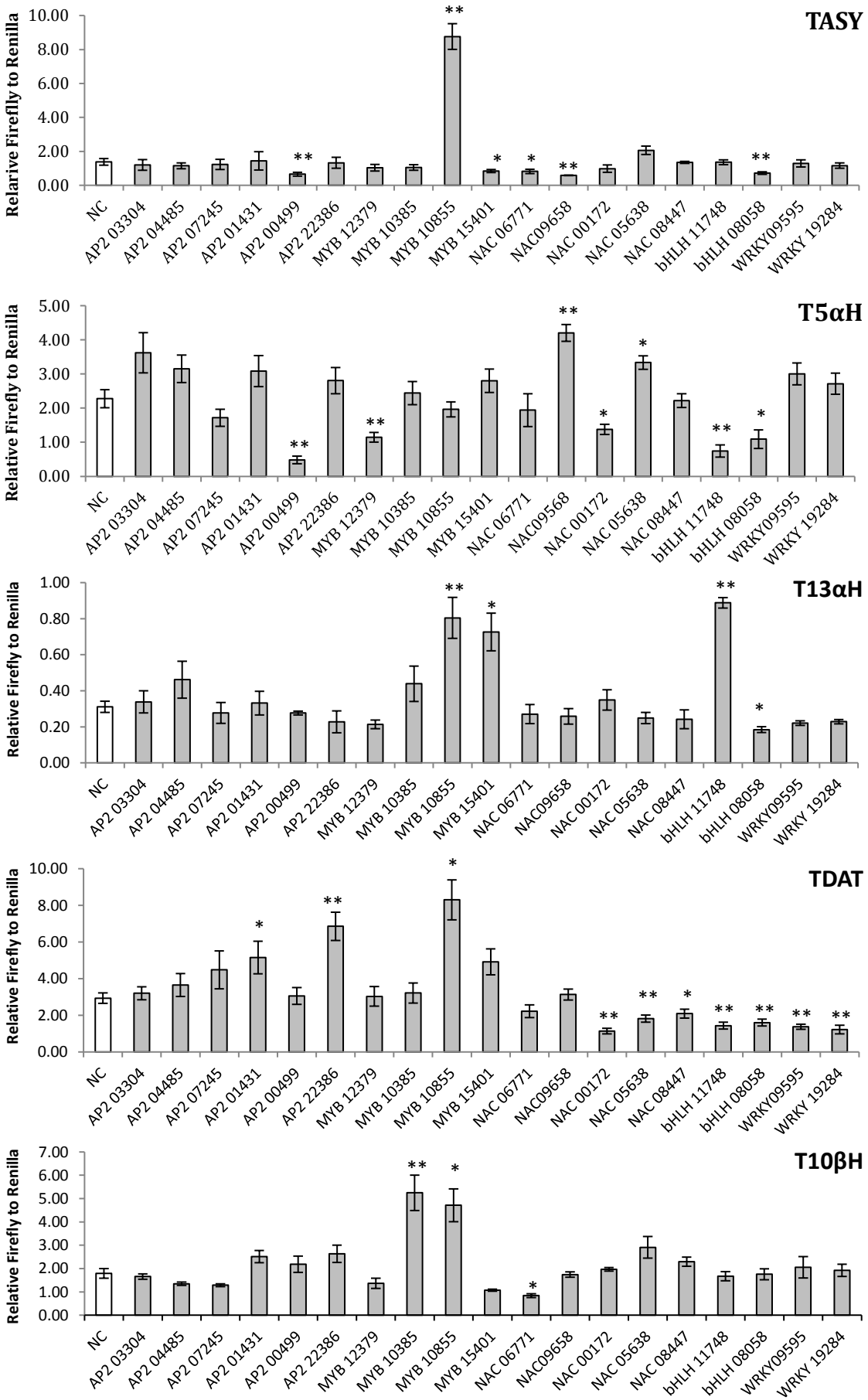
3.3.2 Analysis of *in vivo* TEA screen

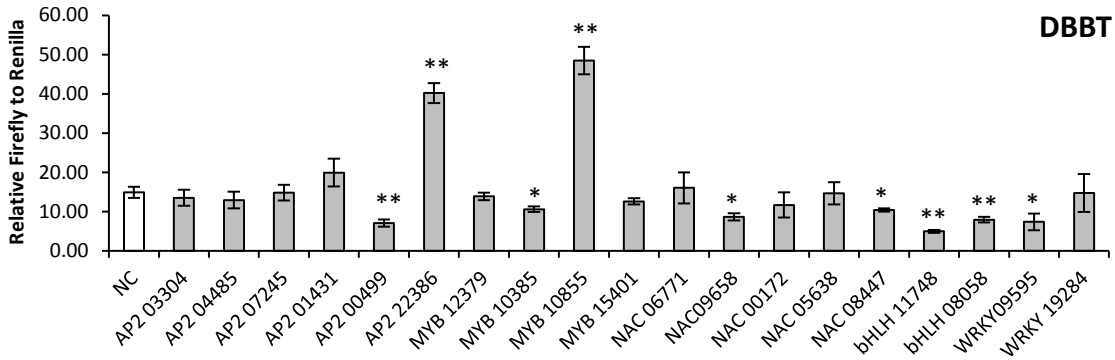
Interactions between the 19 TFs and the 10 paclitaxel biosynthetic promoters were tested using an optimized *A. thaliana* protoplast TEA in at least biological quadruplet. Significant interactions from the negative control were established using a two tailed Student *t*-test, assuming unequal variance with *p* values 0.05 and 0.01. The promoters have different basal levels of expression in *A. thaliana* and their activation and repression by the 19 TFs can be observed in Figure 33 and summarised in Table 14.

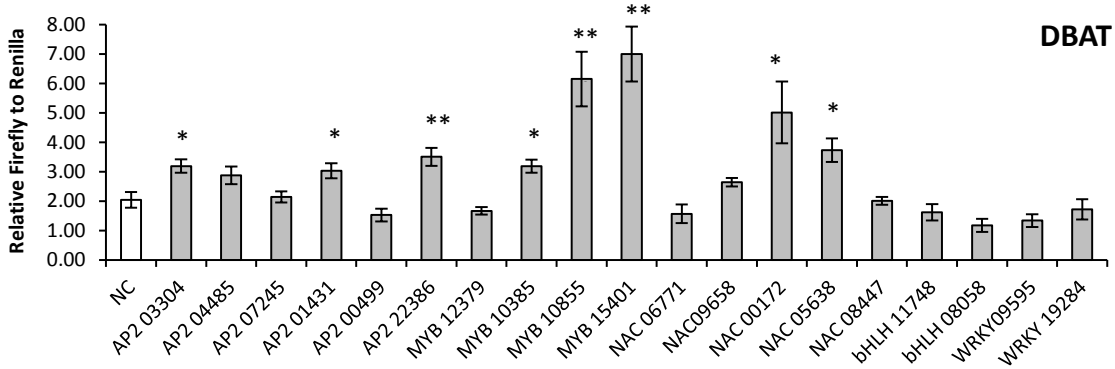
Transcription factor	Promoter											
	TASY	*T5aH*	*T13aH*	*TDAT*	*T10bH*	*DBBT*	*DBAT*	*PAM*	*BAPT*	*DBT-1*	*DBT-2*	*DBTNBT*
AP2 03304												
AP2 04485												
AP2 07245												
AP2 01431												
AP2 00499												
AP2 22386												
MYB12379												
MYB10385												
MYB10855												
MYB15401												
NAC06771												
NAC09658												
NAC00172												
NAC05638												
NAC08447												
bHLH11748												
bHLH08058												
WRKY09595												
WRKY19284												

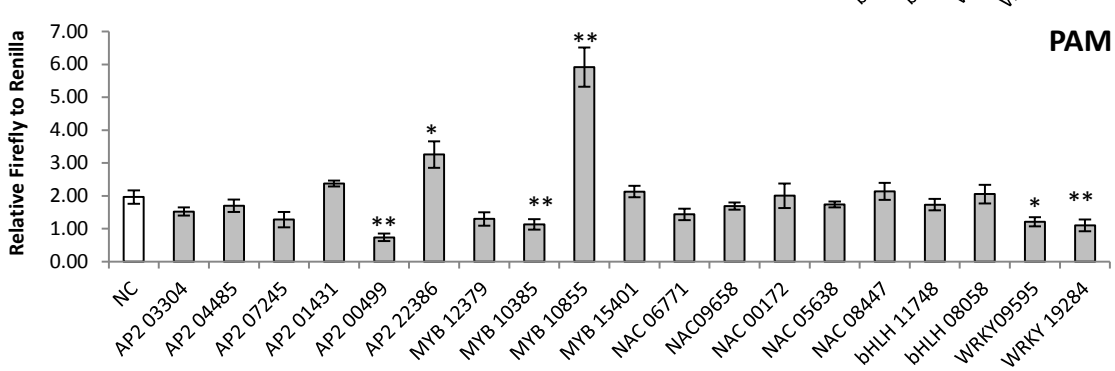
Table 14 – A table summarising the *Arabidopsis thaliana* transient expression assay screen testing for interaction between the 19 candidate transcription factors and the 10 paclitaxel biosynthetic promoters, (see Table 2 for promoter abbreviations). *DBT-1* and *DBT-2* are two fragments of the *DBTNBT* promoter and the *DBTNBT* column shows the summary of both fragments. Purple indicates significant activation of the promoter, blue indicates significant repression of the promoter and white signifies no significant interaction. Significance to the negative control for each promoter was determined using a Student *t*-test, $n \geq 4$, p value 0.05.

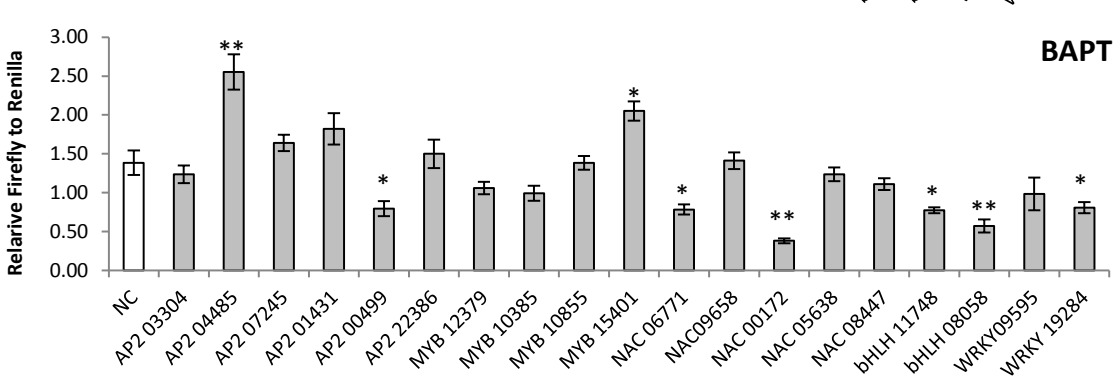
Key	
Positive interaction	
Negative interaction	
No significant interaction	

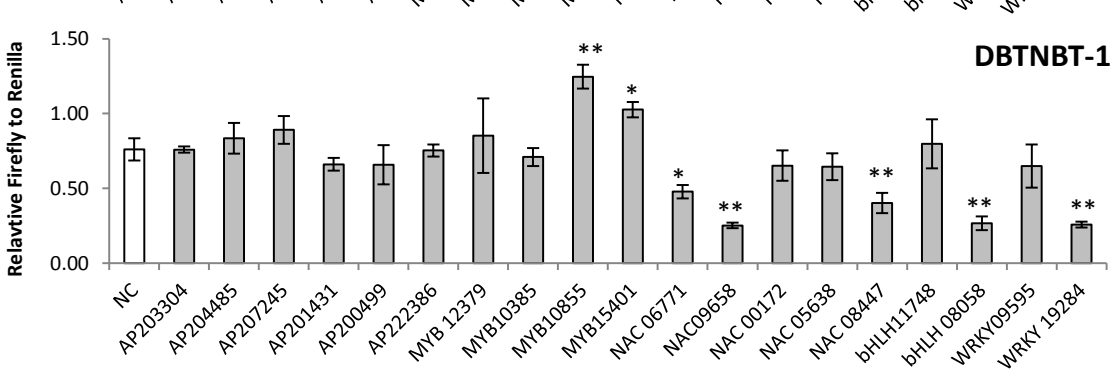

TASY
Relarive Firefly to Renilla
10.00
8.00
6.00
4.00
2.00
0.00
NC
AP2 03304
AP2 04485
AP2 07245
AP2 01431
AP2 00499
AP2 22386
MYB 12379
MYB 10385
MYB 10855
MYB 15401
NAC 06771
NAC09658
NAC 00172
NAC 05638
NAC 08447
bHLH 11748
bHLH 08058
WRKY09595
WRKY 19284
T5αH
Relative Firefly to Renilla
5.00
4.00
3.00
2.00
1.00
0.00
NAC09568
T13αH
1.00
0.80
0.60
0.40
0.20
TDAT
T10βH
7.00
6.00
5.00


DBBT
Relative Firefly to Renilla
60.00
50.00
40.00
30.00
20.00
10.00
0.00
NC
AP2 03304
AP2 04485
AP2 07245
AP2 01431
AP2 00499
AP2 22386
MYB 12379
MYB 10385
MYB 10855
MYB 15401
NAC 06771
NAC09658
NAC 00172
NAC 05638
NAC 08447
bHLH 11748
bHLH 08058
WRKY09595
WRKY 19284


DBAT
Relative Firefly to Renilla
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
NC
AP2 03304
AP2 04485
AP2 07245
AP2 01431
AP2 00499
AP2 22386
MYB 12379
MYB 10385
MYB 10855
MYB 15401
NAC 06771
NAC09658
NAC 00172
NAC 05638
NAC 08447
bHLH 11748
bHLH 08058
WRKY09595
WRKY 19284


PAM
Relative Firefly to Renilla
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
NC
AP2 03304
AP2 04485
AP2 07245
AP2 01431
AP2 00499
AP2 22386
MYB 12379
MYB 10385
MYB 10855
MYB 15401
NAC 06771
NAC09658
NAC 00172
NAC 05638
NAC 08447
bHLH 11748
bHLH 08058
WRKY09595
WRKY 19284


BAPT
Relarive Firefly to Renilla
3.00
2.50
2.00
1.50
1.00
0.50
0.00
NC
AP2 03304
AP2 04485
AP2 07245
AP2 01431
AP2 00499
AP2 22386
MYB 12379
MYB 10385
MYB 10855
MYB 15401
NAC 06771
NAC09658
NAC 00172
NAC 05638
NAC 08447
bHLH 11748
bHLH 08058
WRKY09595
WRKY 19284


DBTNBT-1
Relavtive Firefly to Renilla
1.50
1.00
0.50
0.00
NC
AP203304
AP204485
AP207245
AP201431
AP200499
AP222386
MYB 12379
MYB10385
MYB10855
MYB15401
NAC 06771
NAC09658
NAC 00172
NAC 05638
NAC 08447
bHLH11748
bHLH 08058
WRKY09595
WRKY 19284

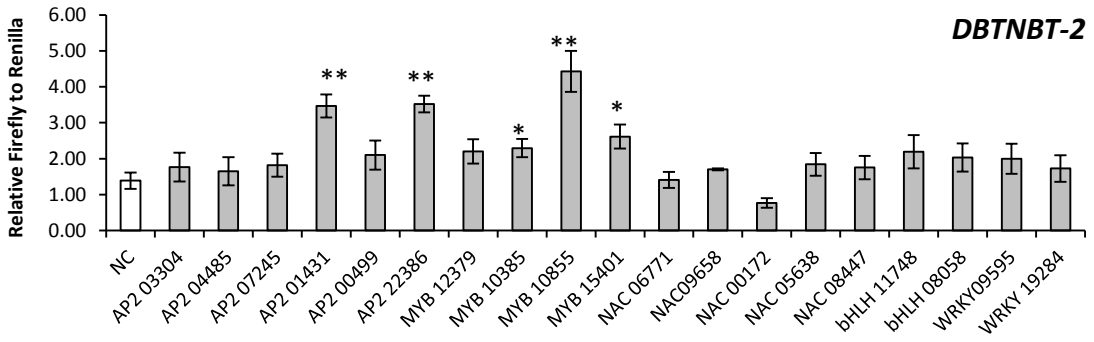


Figure 33 – The summary of the *Arabidopsis thaliana* transient expression assay screen testing for interaction between the 19 candidate transcription factors and the 10 paclitaxel biosynthetic promoters, (see Table 2 for promoter abbreviations). Error bars represent standard error, *n*≥ 4, significance was determined compared to the negative control (NC) in each promoter (highlighted in white) using a Student *t*-test, * *p* value 0.05, ** *p* value 0.01. *DBT-1* and *DBT-2* are the two fragments of the *DBTNBT*, with *DBT-2* being the first 950 bp from the start codon and *DBT-1* is the remaining 840 bp of the promoter.

3.3.3 Promoter analysis of the TEA results

Every promoter was found to interact with at least three TFs and could be both activated and repressed. Overall more negative interactions were identified (42 compared to 29), but no consensus was identified between the ratio of positive and negative interactions and the location of the gene within the pathway (Figure 34). Promoters that had a greater number of positive interactions included the early pathway genes *T13αH* and *T10βH* and the late pathway gene *DBAT*. Promoters that were found to interact with a greater number of repressors also included both early (*TASY, T5αH* and *TDAT*) and late (*PAM, DBBT* and *BAPT)* pathway genes. The presence of both activators and repressors shows that although the paclitaxel biosynthetic genes are up-regulated after MeJA elicitation, the increase is under a strict control to ensure energy is not wasted in the plants defence response.

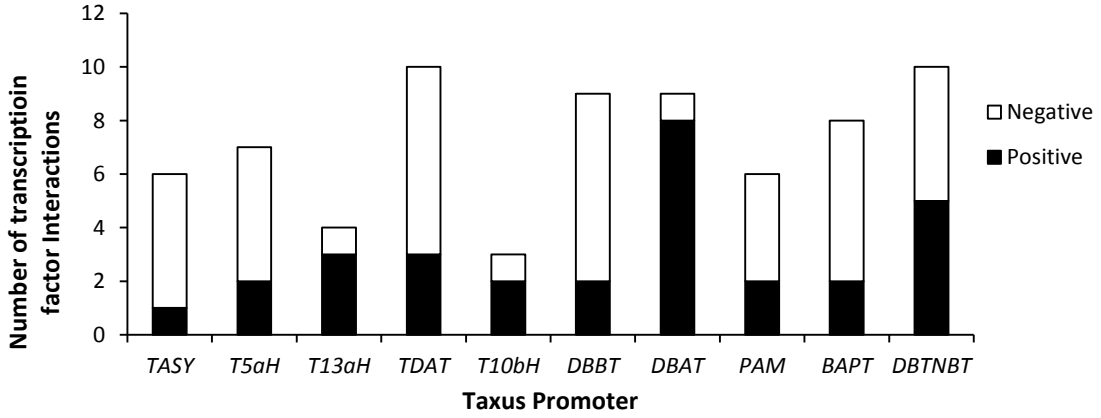


Figure 34 – The number of transcription factors (TFs) that were able to significantly positively or negatively regulate each paclitaxel promoter. See Table 2 for promoter abbreviations. Interactions were identified in an *Arabidopsis thaliana* transient expression assay testing for interaction between the 19 candidate TFs and the 10 paclitaxel biosynthetic promoters. Significance was determined by comparing the addition of a TF to the negative control (empty vector) for each promoter using a Student *t*-test, *n*≥ 4, *p* value 0.05.

The strongest interaction identified by the TEA screen was between *TASY* and MYB10855 (Figure 35). This strong activation of the first committed step of the paclitaxel pathway may

increase the flow of the precursor GGPP into the pathway after MeJA elicitation. However, *TASY* was also shown to be repressed by five TFs (Figure 35), suggesting that although TASY is not thought to be rate limiting (259), its expression is still tightly controlled. Transcript profiling analysis showed that T13αH was highly expressed after MeJA elicitation compared to the other hydroxylases (Figure 17). This increase could be attributed to the identification of three TFs that activated promoter expression and only one TF (bHLH08058) that could repress its activity (Figure 35). In *T. baccata* cell culture MeJA elicitation led to high accumulation of baccatin III suggesting that *DBAT* is not a rate limiting step in the pathway (47). DBAT was found to have high logFC values after MeJA elicitation in *T. cuspidata* CMCs (Figure 17), which could be linked to the identification of eight TFs that were able to activate *DBAT* activity.

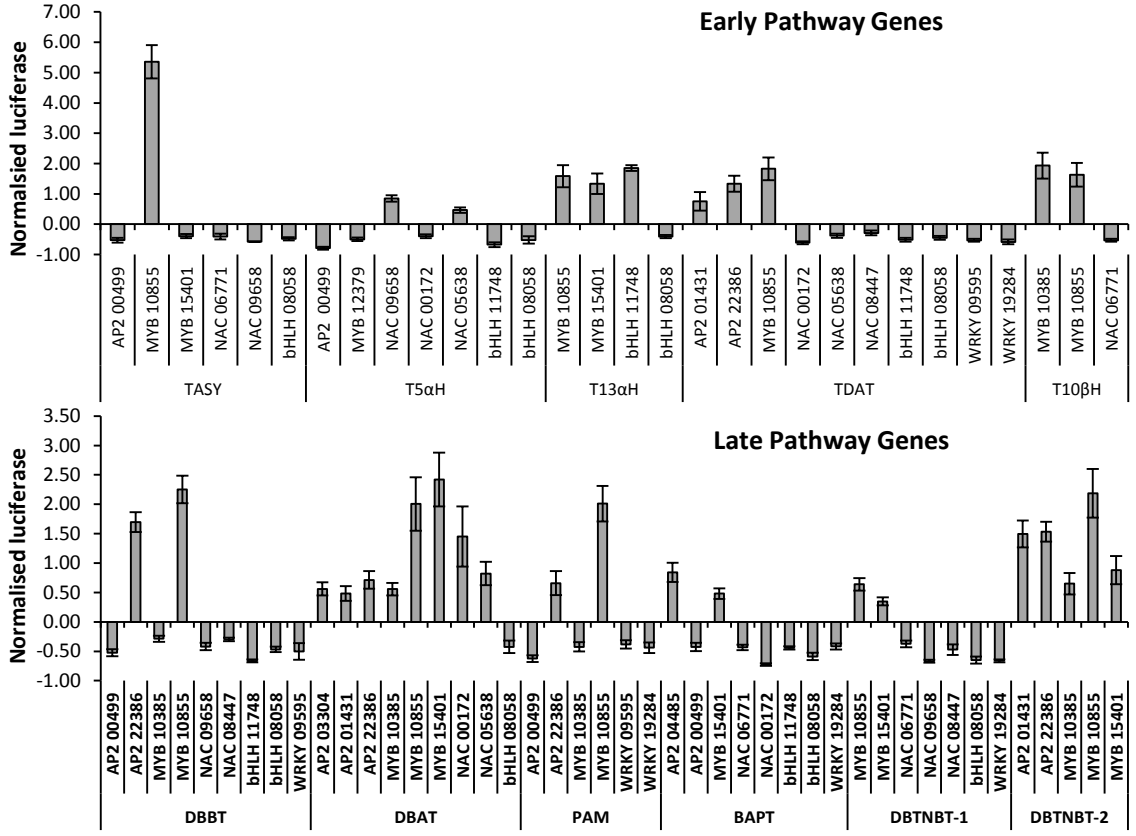


Figure 35 – The normalised significant interactions identified in the *Arabidopsis thaliana* transient expression assay (TEA) screen between the 19 candidate transcription factors (TFs) and the 10 paclitaxel biosynthetic promoters. See Table 2 for promoter abbreviations. Interactions are stated in normalised luciferase, the values obtained in the TEA were normalised to the negative control (NC) for each promoter and then the NC was altered to zero. Interactions are broken down based on the location of the enzymes in the biosynthetic pathway. Significance was determined by comparing the addition of a TF to the NC for each promoter (addition of empty vector), using a Student *t*-test, $n \geq 4$, p value 0.05.

The later steps of the biosynthetic pathway BAPT and DBTNBT have been suggested to act as a bottleneck, with BAPT thought to play a key role in controlling the flux of intermediates through the paclitaxel pathway (47). The limiting effect of BAPT could be related to identification of six TFs that repress its promoter (Figure 35) and the lack of interaction with MYB10855 shown to act as a strong activator on 80% of the paclitaxel promoters (Figure 39). *DBTNBT* was tested in two fragments; only promoter activation was identified in the first 950 bases (*DBTNBT-2*), while

DBTNBT-1 was found to be repressed by five TFs. DBTNBT is highly up-regulated upon MeJA elicitation but an equal number of TFs were able to activate and repress the promoter. A correlation could therefore not be established between the number of activators and repressors a promoter interacted with and its expression after MeJA elicitation. All the promoters bound with multiple TFs, therefore further research is needed to establish which interactions are more dominant, as not all the TFs can regulate the promoter simultaneously.

3.3.4 Transcription factor analysis of the TEA results

A member from each TF family tested was found to interact with at least one of the paclitaxel promoters. Overall the AP2 and MYB TFs were activators of gene expression, while the activity mostly observed with the NAC, bHLH and WRKY families was repression (Figure 36a). A higher number of positive interactions identified in the AP2 and MYB families were found in the late pathway genes, while the negative regulatory activity of the NAC and bHLH families was spread evenly across the paclitaxel biosynthetic pathway (Figure 36b).

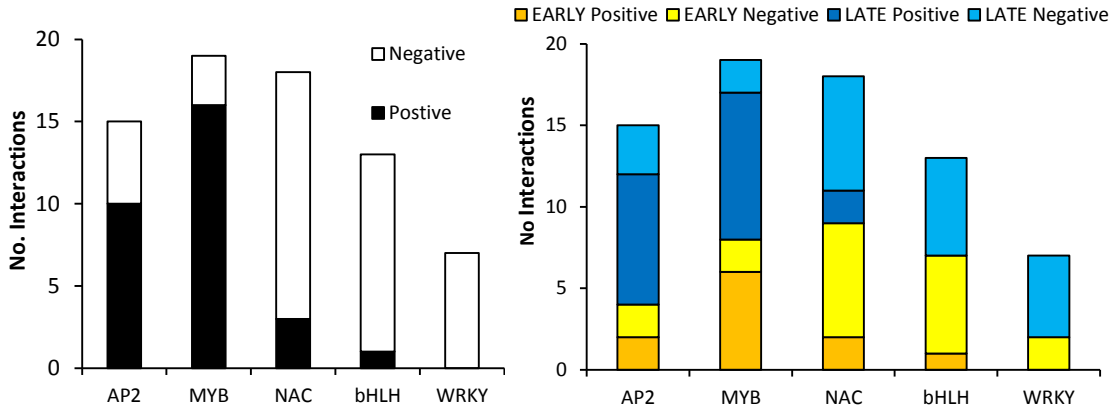


Figure 36 – The number of significant interactions identified in the *Arabidopsis thaliana* transient expression assay screen for each transcription factor (TF) family. a) The number of positive and negative interactions for each TF family is stated. b) The interactions are broken down into positive and negative interactions for the early and late steps of the paclitaxel biosynthetic pathway for each TF family. Significance was determined by comparing the addition of a TF to the negative control (empty vector) for each promoter using a Student *t*-test, $n \geq 4$, p value 0.05.

Only one TF, AP207245, did not interact with any of the promoters tested and three TFs (AP203304, AP204485 & MYB12379) interacted with only one promoter. The remaining 15 TFs were capable of binding with multiple promoters (Table 14) and two of these TFs could interact with 8 out of the 10 promoters tested, MYB10855 a positive regulator and bHLH08058 a negative regulator. These two TFs were each able to regulate the expression of 80% of the paclitaxel pathway, therefore they are top candidates to be used in bioengineering efforts to increase paclitaxel production in *Taxus* CMCs. Six TFs across three families (MYB, NAC and bHLH) acted as dual regulators, with their regulatory activity being promoter dependent. The strength of the interactions also varies, with MYB TFs producing some of the strongest interactions. Eight out of the top ten positive interactions were from the MYB family, with the majority of the strongest interactions occurring in the later part of the biosynthetic pathway. While there was no obvious trend in negative interactions, with the top ten interactions comprising TFs from three families;

NAC, bHLH and AP2. A large number of interactions were identified in the TEA screen, but their biological relevance needs to be established as stronger interactions are likely to predominately regulate the promoter.

3.3.4.1 AP2 TF family

The majority of AP2 TFs tested in this screen had positive interactions with promoters of late pathway genes, except for AP200499 which acted as a repressor for 5 out of the 10 promoters (Figure 37). None of the TcAP2 activators contains the known activation domain EDLL (119) however, activation domains are currently poorly characterised in the AP2 family. AP200499 on the other hand contains the well-established ERF associated amphiphilic repression (EAR) motif (Figure 38a), see section 3.3.4.1.1 Analysis of the AP200499 for further discussion (260).

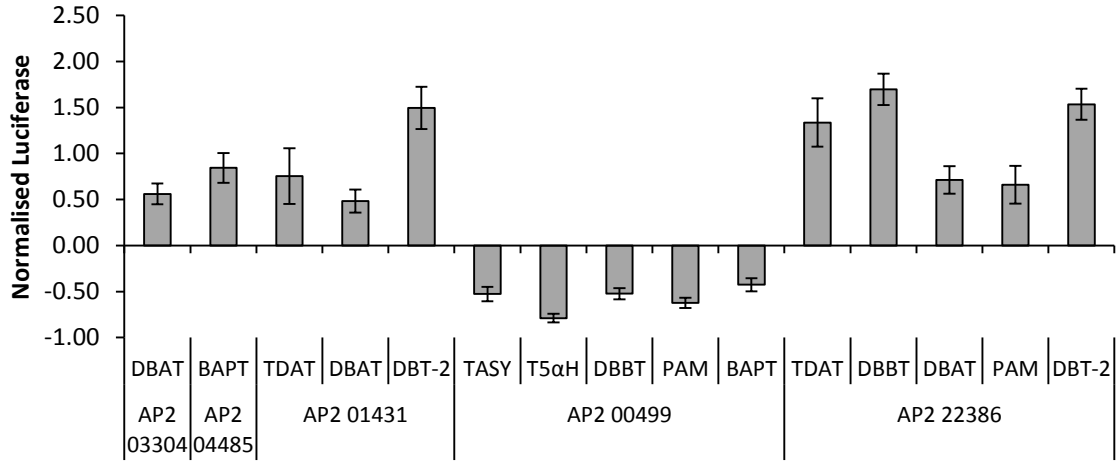


Figure 37 – The normalised significant interactions identified in the *Arabidopsis thaliana* transient expression assay (TEA) screen for transcription factors (TF) in the AP2 family with the 10 paclitaxel biosynthetic promoters. See Table 2 for promoter abbreviations. Interactions are stated in normalised luciferase, the values obtained in the TEA were normalised to the negative control (NC) for each promoter and the NC was then altered to zero. Significance was determined by comparing the addition of a TF to the NC for each promoter (addition of empty vector), using a Student *t*-test, $n \geq 4$, p value 0.05. The majority of the interactions identified in the AP2 family are positive, except with AP200499 which acted as a repressor.

AP207245, which does not bind to any of the promoters tested, is evolutionarily distinct from the rest of the candidate TFs and has little similarity to any known sequence published; BLASTp (225) analysis produced no hits with an *e* value lower than 0.2. None of the TcAP2s have homology with the well characterised ORCA proteins which regulate anthocyanin biosynthesis, or the *Artemisia annua* ERFs that regulate artemisinin biosynthesis. A possible reason for the low similarity between the TFs currently linked to JA reprogramming of secondary metabolite synthesis and the TcAP2s is that they are from angiosperms while *Taxus* is a gymnosperm. Given the distant separation of the two linages of approximately 300 million years the genes are evolutionarily distant from each other (181).

3.3.4.1.1 Analysis of the AP200499 EAR motif

The EAR motif is an active repression motif defined by the consensus patterns LxLxL or DLNxxP (260). Figure 38a shows the alignment of the EAR domain in AP200499 with known

repressors from *A. thaliana*, rice and tobacco. All the proteins contain the DLNxxP motif, which is essential for repression (120). The AP200499 repression activity was investigated using three mutants (Figure 38b). One mutant (Trun) had the C-terminus (where the EAR motif is located) truncated and two other mutants were created using site-directed mutagenesis. L161 has a lysine residue within the EAR motif mutated to alanine, while L167 has a lysine residue outside the motif mutated to an alanine. These mutants were then tested against promoters that had previously been identified as repressed by AP200499; *TASY*, *T5αH*, *DBBT*, *BAPT* and *PAM* (Figure 38c-g).

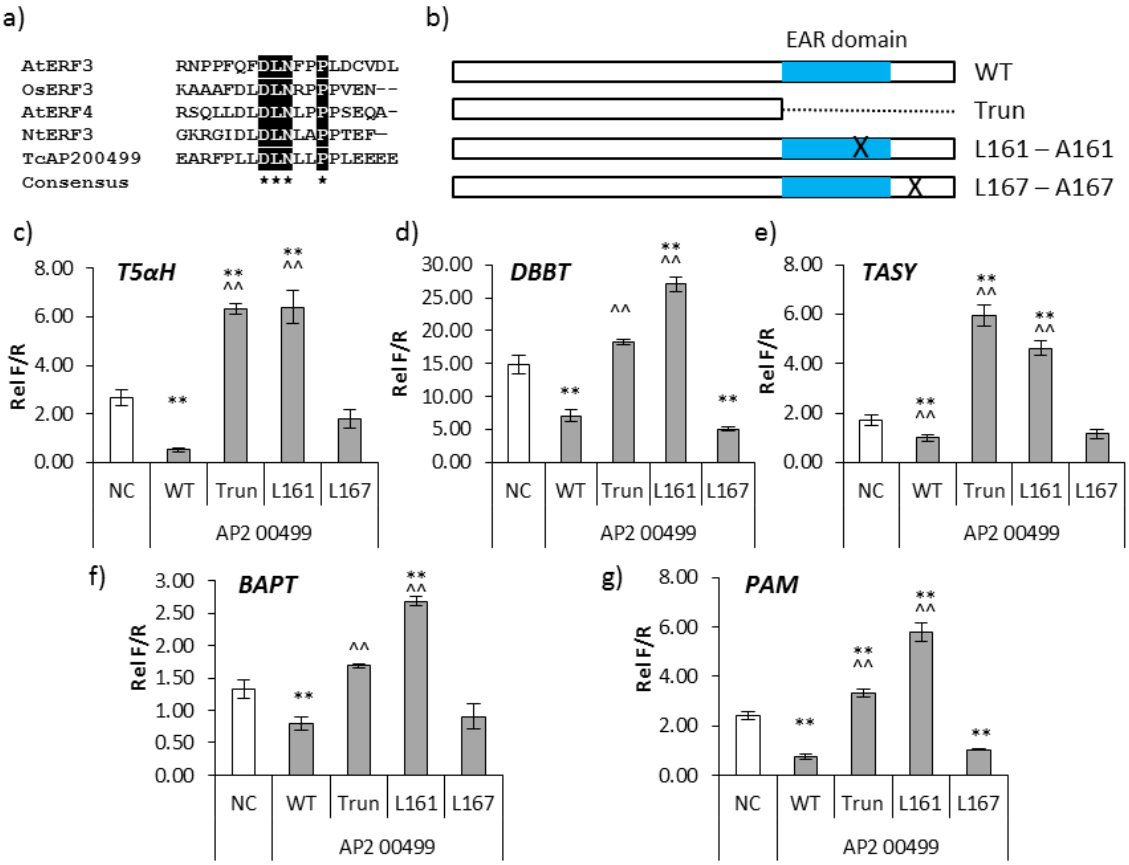


Figure 38 – Analysis of the ERF associated amphiphilic repression (EAR) domain in AP200499. a) The sequence alignment of the EAR domain of AP200499 with known repressors from *Arabidopsis thaliana*, rice and tobacco. The alignment was produced using ClustalOmega (231, 232) and the image produced using BioEdit v7.2.5 (233). All five proteins contain the EAR consensus pattern DLNxxP (260). b) A schematic of the AP200499 mutants created, with the blue box denoting the location of the EAR domain. Three mutants were created, 1) with the C-terminus truncated to remove the EAR domain, 2) a lysine within the EAR domain was mutated to an alanine and 3) a lysine outside the EAR domain was mutated to an alanine. c) The summary of the *A. thaliana* transient expression assay testing for interaction between the paclitaxel biosynthetic promoters *TASY*, *T5αH*, *DBBT*, *BAPT* and *PAM* (see Table 2 for promoter abbreviations) and the four versions of AP200499 presented in b) - the wild type (WT), the C-terminal truncated mutant (Trun), the mutation within the EAR domain (L161) and outside the EAR domain (L167). Error bars represent standard error, $n \geq 4$, Rel F/R stands for relative firefly to *Renilla* luciferase. Significance was determined either by comparing results to the negative control (NC) in each promoter (highlighted in white) or to the WT AP200499 using a Student *t*-test, ** *p* value 0.01 were compared to NC, ^^ *p* value 0.01 compared to WT. Mutation of the EAR domain led to the loss of the repression activity observed in the WT in all promoters tested.

For the five promoters tested, truncation of the N-terminal region of AP200499 led to loss of repression and even activation in three promoters (*T5αH*, *TASY* and *PAM*). Mutation of the lysine residue inside the EAR motif led to loss of repression and significant activation in all promoters tested. Mutating the lysine outside of the EAR domain did not lead to a significant change

compared to WT, however a significant difference with the negative control was sometimes lost (Figure 38c–g). This data confirms that the EAR domain in AP200499 is an active repressor domain.

3.3.4.2 MYB TF family

All of the MYB TFs tested were capable of interacting with at least one of the promoters under investigation, however the most interesting was MYB10855 which could strongly activate 8 of the 10 promoters tested (Figure 39). MYB10855 has the third highest induction in gene expression at an early time point (0.5 h) after MeJA elicitation and is evolutionally distinct from the other TcMYBs and previously identified JA responsive MYBs involved in secondary metabolism. Blastp (225) analysis of the *A. thaliana* genome identified LOF2 as MYB10855 closest homolog with an *e*-value of $2e^{-58}$, which is involved in axillary meristem regulation and lateral organ formation (125, 261). The sequence similarity is not high outside the MYB domain between MYB10855 and LOF2, however the C-terminal domains of MYBs are highly variable and *T. cuspidata* and *A. thaliana* are not closely related species (122). CMCs can produce higher levels of paclitaxel compared to DDCs and this could be linked to the high expression of a TF involved in meristem initiation. MYB10855 is therefore a good candidate to try and improve paclitaxel production in CMCs because of its ability to activate most of the paclitaxel biosynthetic pathway.

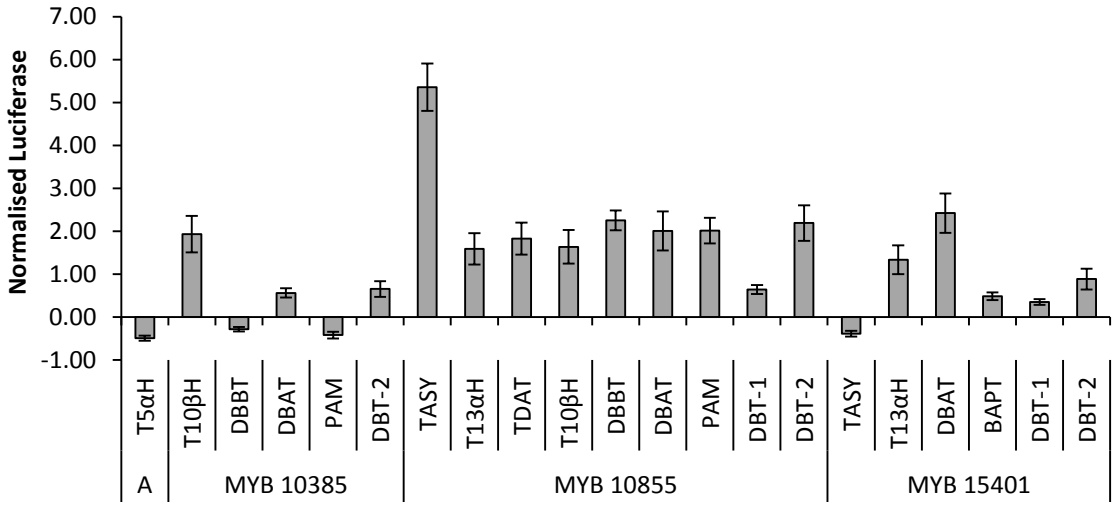


Figure 39 – The normalised significant interactions identified in the *Arabidopsis thaliana* transient expression assay (TEA) screen for transcription factors (TFs) in the MYB family with the 10 paclitaxel biosynthetic promoters. See Table 2 for promoter abbreviations. Interactions are stated in normalised luciferase, the values obtained in the TEA were normalised to the negative control (NC) for each promoter and the NC was then altered to zero. Significance was determined by comparing the addition of a TF to the NC for each promoter (addition of empty vector), using a Student *t*-test, $n \geq 4$, p value 0.05. The MYB TFs mostly activate late pathway promoters and MYB10855 was able to activate 80% of the paclitaxel promoters tested.

Two MYBs, MYB10385 and MYB15401, were shown to work as dual regulators, acting as activators or repressors in a promoter dependent fashion. Dual mode TFs can act in different ways: a) they can be promoter dependent – activating one promoter and repressing another (e.g. CysB) or b) *cis*-element dependent – binding to different *cis*-elements within a promoter leading to altered activity (e.g. AraC) (190). In animals c-MYB is capable of inducing transcription positively and

negatively (262) and in cotton a R2R3-MYB has been identified that has dual control in the phenylpropanoid biosynthetic pathway. There is no correlation between the two dual regulators over which promoters or sections of the pathway are activated or repressed. Only one weak negative interaction was observed for MYB12379. A possible reason for this observation could be that one of its closest *A. thaliana* homologs is WEREWOLF (WER), whose activity is dependent on bHLH TFs GL3 and EGL3. Sequence analysis suggested that MYB12379 contains the bHLH interaction motif (Figure 50) and possible combinational action is explored in Chapter 4.

3.3.4.3 NAC TF family

All of the NAC TFs tested were capable of binding with either 3 or 4 promoters. Most of the interactions identified were negative, with three of the TFs capable of acting as dual regulators (Figure 40).

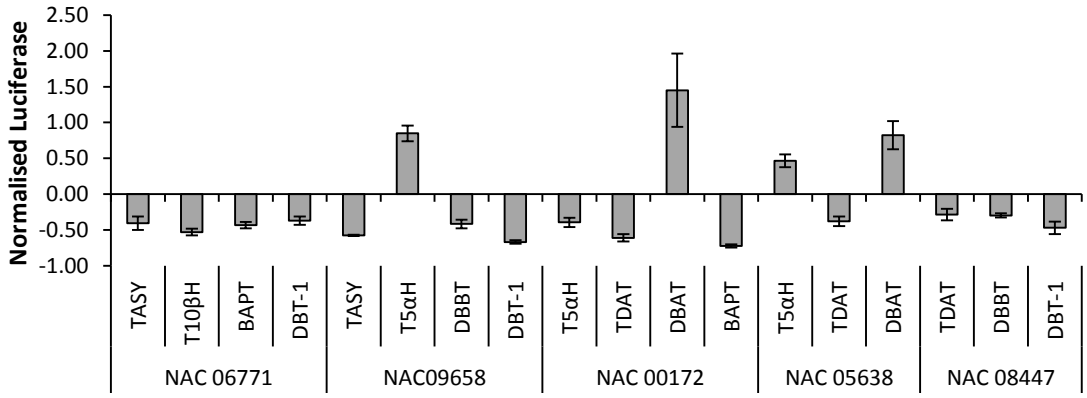


Figure 40 – The normalised significant interactions identified in the *Arabidopsis thaliana* transient expression assay (TEA) screen for transcription factors (TFs) in the NAC family with the 10 paclitaxel biosynthetic promoters. See Table 2 for promoter abbreviations. Interactions are stated in normalised luciferase, the values obtained in the TEA were normalised to the negative control (NC) for each promoter and the NC was then altered to zero. Significance was determined by comparing the addition of a TF to the NC for each promoter (addition of empty vector), using a Student *t*-test, $n \geq 4$, p value 0.05. The majority of the regulatory activity identified with the NAC TFs was negative.

The NAC repression domain (NARD), identified in GmNAC20, has been suggested to be a conserved active repression domain in NAC proteins. This 35 amino acid region is 37.6% hydrophobic, which is characteristic of other repression domains, and contains the LVFY motif that has been shown to be partially required for repression activity (153). All of the NAC TFs, except NAC08447, contain the NARD domain which helps to explain their repression activity; however well-known transcriptional activators ANAC019, ANAC055 and ANAC072 also contain the LVFY sequence (Figure 41)(150, 152). It has been suggested that the interaction between the NARD and activation domain controls the regulatory activity of NAC TFs. If the NARD domain is stronger than the activation domain then the TF acts as a repressor and *vice versa* for activators. However, if the domains are of equal strength the regulatory activity depends on situational cues from the plant (153). For example, ANAC019 functions as a dual regulator for two biosynthetic enzymes in SA metabolism activating *S*-adenosylmethionine-dependent methyltransferase and repressing isochorismate synthase (149). Three of the TFs NAC05638, NAC00171 and NAC09658 can also act as dual regulators, showing the presence of the LVFY does not automatically make the protein

a repressor.

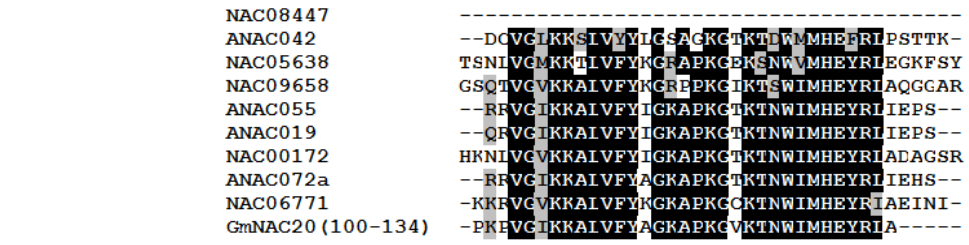


Figure 41 – The protein alignment of the NARD domain of *Taxus cuspidata* NACs and *Arabidopsis thaliana* NAC transcription factors ANAC019, ANAC042, ANAC055 and ANAC072a (143) and the NARD domain from soybean, GmNAC20 (153). Alignment was completed using Clustal Omega (231, 232) and image produced using BioEdit v7.2.5 (233). The NARD domain was reported to act as the repression domain in GmNAC30 and was identified in all the TcNACs except NAC08447; but it was also found in the known transcriptional activators ANAC019, ANAC055 and ANAC072a.

NAC05638 is related to AtCUC1 (CUP SHAPED COTYLEDON) and AtCUC2, with blastp *e* values of $1e^{-78}$ and $2e^{-74}$ respectively, which control embryonic meristem formation and cotyledon separation (261, 263). Alignment of these proteins showed a high level of consensus in the NAC domain and the V motif (Figure 42b). Compared to DDCs the gene expression of NAC05638 in CMCs was significantly increased after MeJA elicitation and in the TEA screen NAC05638 was found to activate *T5αH* and *DBAT* while repressing *TDAT*. CUC1 and CUC2 are involved in meristem control therefore the increase of this meristem related TF could be connected with the higher yield of paclitaxel produced in CMCs.

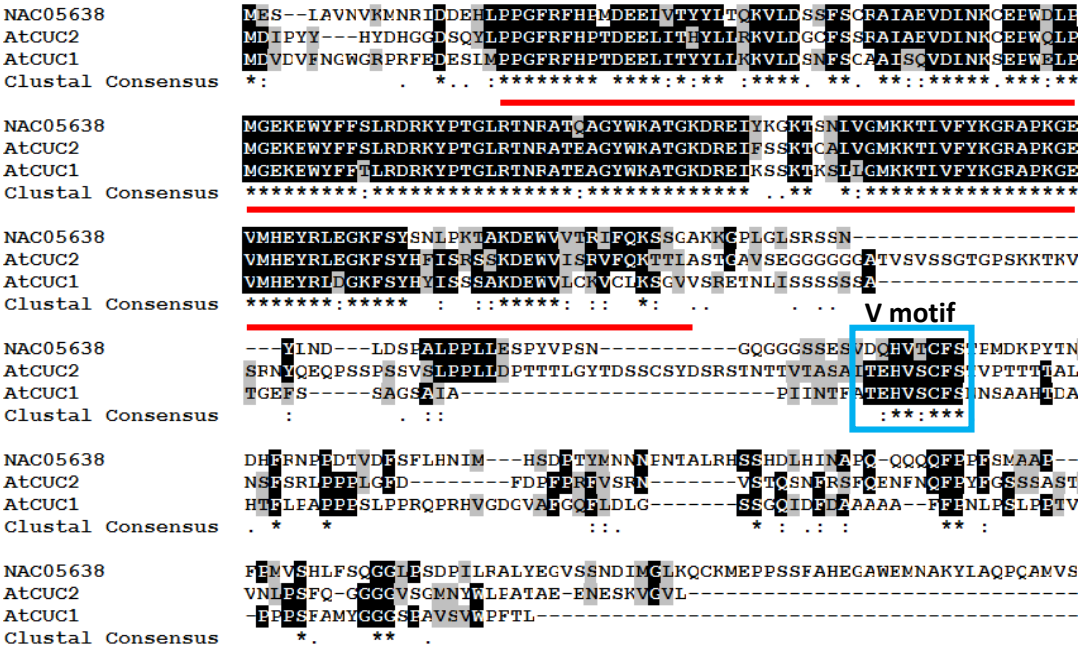


Figure 42 – The protein alignment of NAC05638 with related *Arabidopsis thaliana* NAC transcription factors (TFs) CUC1 and CUC2. The high level of similarity between the TFs in the NAC domain is highlighted by the red line and the partial conservation of the short sequence TEHVSCFS, known as the V motif, is highlight by the blue box. The alignment was produced using ClustalOmega (231, 232) and the image produced using BioEdit v7.2.5 (233).

The data for NAC08447 may not be reliable because previous analysis only identified a truncated version of the protein (performed by Amir (223) and Zejun (222)). The NAC08447 protein tested was only 80 amino acids long and was missing a number of well conserved NAC domains.

However further analysis shows that there is a much longer reading frame in contig08447 that has high similarity to the other TcNACs in the N terminus (Figure 43). The results for this protein are therefore flawed as a significantly truncated protein was tested. NAC09658 may also not be the full transcript as compared to other NAC proteins it seems to be missing part of the N terminus containing the RDRKYP and GWKAT sequences (Figure 43). 5' and 3' RACE PCR needs to be performed on these TFs in the future to establish the full sequence of these TFs.

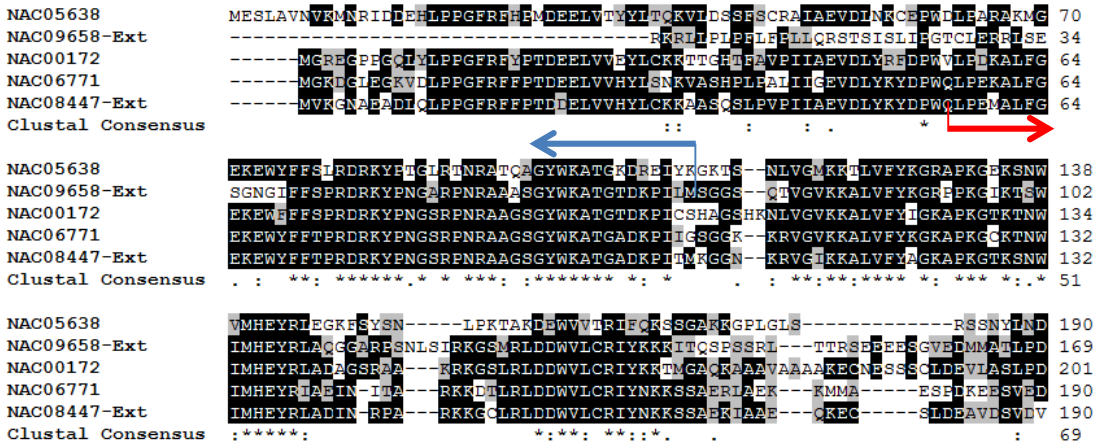


Figure 43 – The alignment of the NAC domains of *Taxus cuspidata* NAC transcription factors (TFs). The red arrow indicates where an elongated form of NAC08447 could extend in the C terminal direction. The blue arrow indicates the extension of NAC09658 in the N terminal direction. The extended versions of these two proteins was proposed by analysing the original Roche454 contigs. The extended regions align well with the NAC domains of the other TFs suggesting that the open reading frames identified by Zejun (222) and Amir (223) were not complete. The alignment was produced using ClustalOmega (231, 232) and the image produced using BioEdit v7.2.5 (233).

3.3.4.4 bHLH TF family

Except for one positive interaction between bHLH11748 and *T13αH* all the interactions identified by the TcbHLHs are negative.

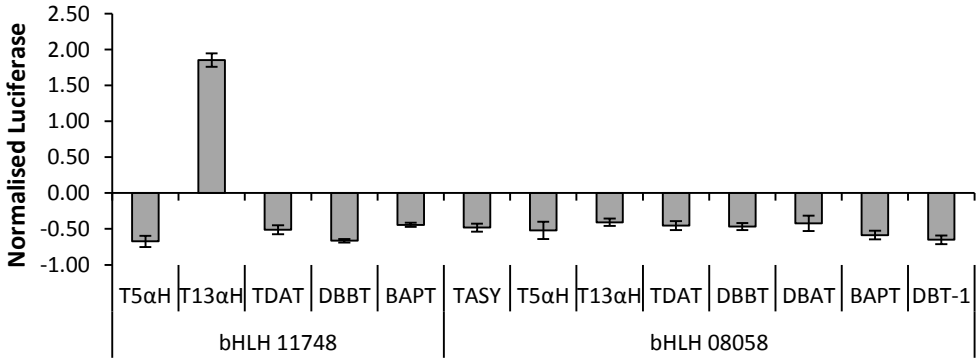


Figure 44 – The normalised significant interactions identified in the *Arabidopsis thaliana* transient expression assay (TEA) screen for transcription factors (TFs) in the bHLH family with the 10 paclitaxel biosynthetic promoters. See Table 2 for promoter abbreviations. Interactions are stated in normalised luciferase, the values obtained in the TEA were normalised to the negative control (NC) for each promoter and the NC was then altered to zero. Significance was determined by comparing the addition of a TF to the NC for each promoter (addition of empty vector), using a Student t-test, n≥ 4, p value 0.05. Except for a positive interaction between bHLH11748 and *T13αH*, the bHLH TFs acted as repressors of the paclitaxel biosynthetic pathway.

Phylogenetic analysis of the TcbHLHs with bHLHs known to regulate secondary

metabolism in response to MeJA showed that bHLH08058 was closely related to MYC2, while bHLH11748 is more closely related to the *A. thaliana* TFs GL3, EGL3 and TT8 that work co-operatively with MYB TFs to regulate anthocyanin biosynthesis (100, 264) (see Chapter 4 for further details). Previous analysis conducted in section 2.4 found that the ORF found by Amir and Zejun was a truncated form and an elongated version (bHLH08058-FE) was identified and successfully cloned. Preliminary investigations found that bHLH08058-FE expanded the TF's repressor activity to include *T10βH* and *PAM* (Figure 46a). It appears that extending the protein enhanced its repression activity allowing it to repress the entire pathway, however due to time constraints full testing has not yet been completed, so statistically significance can not be attached to the result.

The work completed by Lenka *et al.* with TcJAMYC4 using a *T. cuspidata* particle bombardment assay only identified repression in five of the seven promoters tested. (Figure 46b-c). The difference between the data sets could be due to co-factors in *Arabidopsis* or *Taxus* that might interact with the TF altering its regulatory activity. The protocol employed by Lenka *et al.* the measurement to be taken 48 h after bombardment, therefore endogenous regulatory mechanisms were likely to affect the results, whereas the *A. thaliana* TEA was not affected by these. Attempts to amplify the TcJAMYC4 ORF were unsuccessful as an extra guanine was not present (Figure 23). The change in the reading frame at the N-terminus caused by this insertion could also account for the differences observed between the data sets.

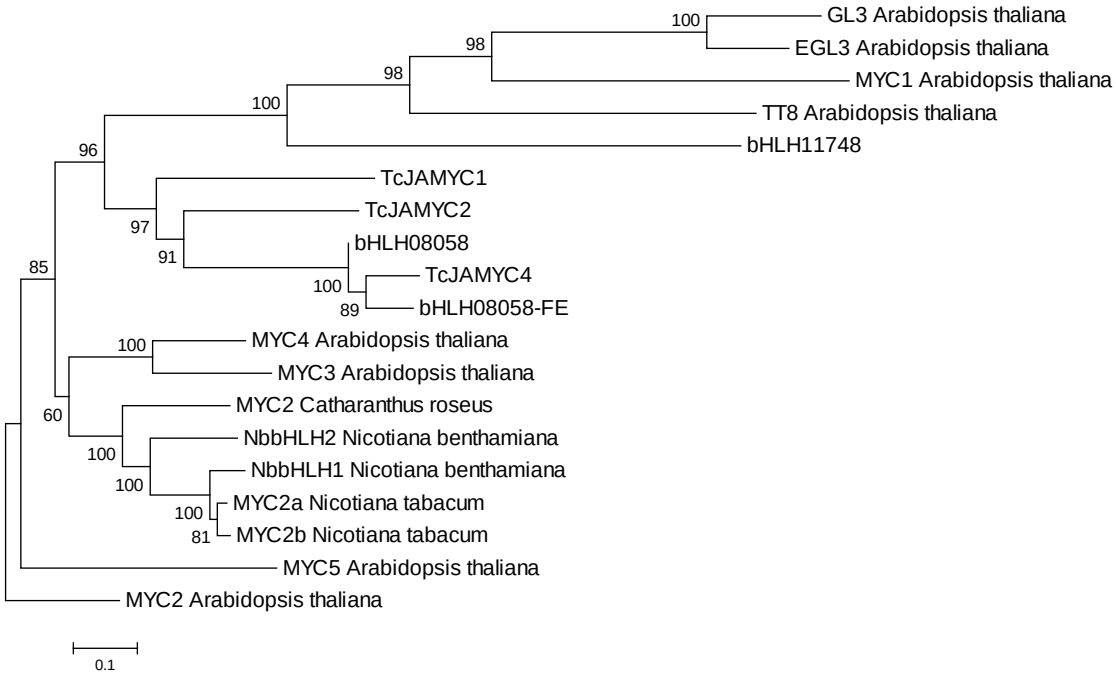


Figure 45 – The phylogenetic analysis of *Taxus cuspidata* CMC bHLHs with bHLHs known to regulate of secondary plant metabolism in response to methyl jasmonate (100) and currently reported *T. cuspidata* MYC2-like TF (TcJAMYCs) (184). The alignment was conducted using Muscle (265) and the rooted neighbour joining tree produced using MEGA5, with a bootstrap value of 1000 and rooted to *Arabidopsis thaliana* MYC2 (266). bHLH08058 is closely related to MYC2-like transcription factors, while bHLH11748 is more closely related to EGL3, GL3 and TT8.

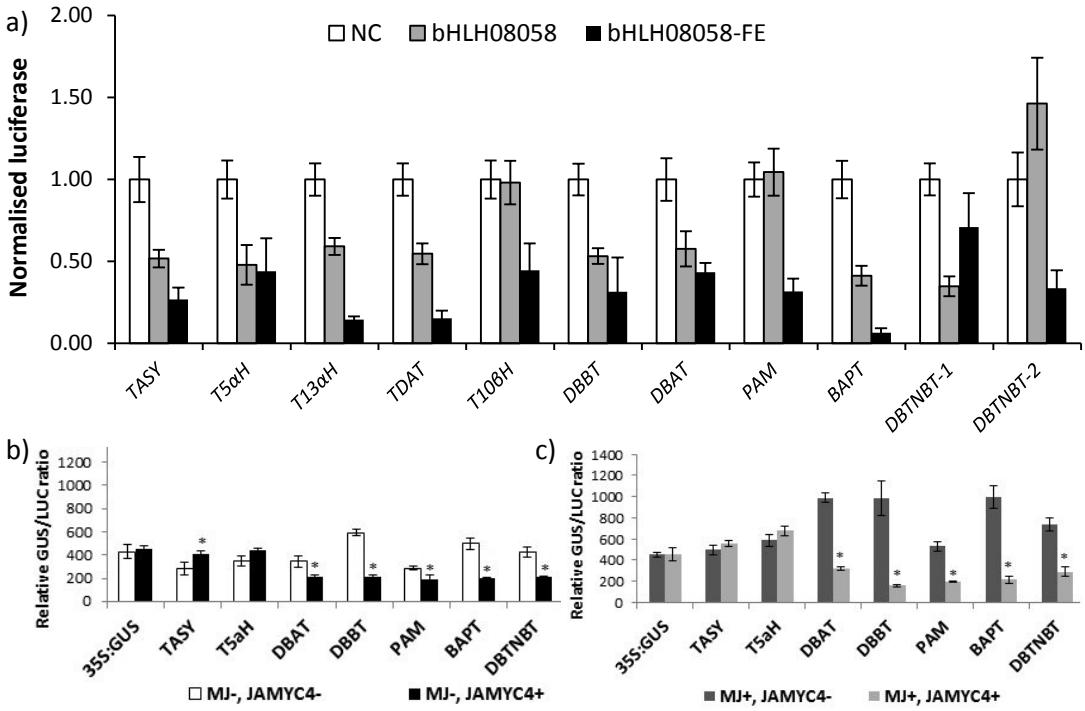


Figure 46 – Comparison of regulatory activity of bHLH08058, bHLH08058-FE (an elongated version) and JAMYC4 reported by Lenka *et al.* 2015 (184)**.** a) Comparison of the *Arabidopsis thaliana* transient expression assay (TEA) results for bHLH08058, bHLH08058-FE and the negative control (NC) for with 10 paclitaxel biosynthetic promoters. See Table 2 for promoter abbreviation, error bars represent standard error, $n \geq 3$. Significance could not be determined as the bHLH08058-FE results were only preliminary, but elongating the transcription factor appears to increase the repression activity of the protein allowing it to negatively regulate the entire paclitaxel biosynthetic pathway. b) and c) are the results of the *Taxus cuspidata* particle bombardment assay reported by Lenka *et al* for TcJAMYC4. b) is without methyl jasmonate treatment (MJ) and c) is with MJ treatment. JAMYC4 is a truncated form of bHLH08058-FE (see Figure 24 for further details). Lenka *et al.*'s results differ from our TEA data as JAMYC4 only repressed five of the seven paclitaxel biosynthetic promoters tested.

3.3.4.5 WRKY TF family

In the TEA screen the two WRKY's were found to repress 5 promoters mostly located in the late pathway genes (Figure 47). In 2012 Li *et al.* published the activation of the DBAT promoter by TcWRKY1 (which is identical to WRKY09595) however this interaction was not identified in the TEA screen (183). Both Li and a previous researcher on the project (Amir) found that WRKY09595 could interact with a W-box in the DBAT promoter (183, 223). However, transient *Taxus* bombardment experiments completed by Li showing DBAT activation did not contain an internal control, therefore the data was not normalised and not independent of the transfection efficiency. It is consequently difficult to make any conclusions from their data and compare it to our TEA results.

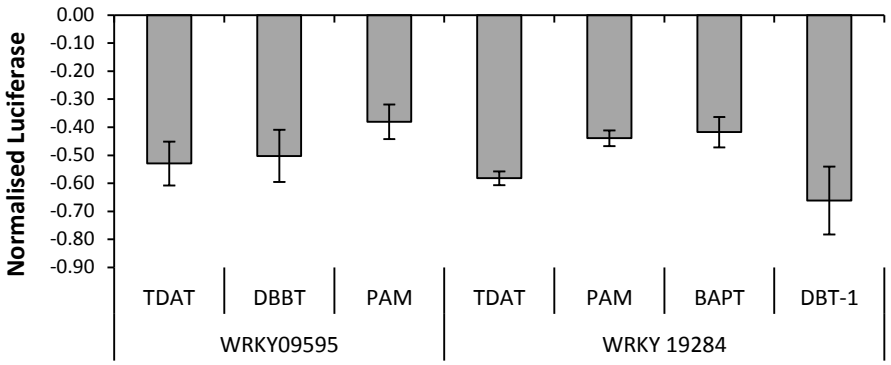


Figure 47 – The normalised significant interactions identified in the *Arabidopsis thaliana* transient expression assay (TEA) screen for transcription factors (TFs) in the WRKY family with the 10 paclitaxel biosynthetic promoters. See Table 2 for promoter abbreviations. Interactions are stated in normalised luciferase, the values obtained in the TEA were normalised to the negative control (NC) for each promoter and the NC was then altered to zero**.** Significance was determined by comparing the addition of a TF to the NC control for each promoter (addition of empty vector), using a Student *t*-test, $n \geq 4$, *p* value 0.05. The WRKY TFs were found to repress five paclitaxel promoters.

A possible reason that an interaction was detected *in vitro* between *DBAT* and WRKY09595 but not in TEA could be that cofactors are required for binding or the protein requires post-translational modifications to be active. In *A. thaliana* SIB1 and SIB2 (sigma factor binding proteins) activate WRKY33 in response to the necrotrophic fungus *Botrytis cincera*; and in *Nicotinia benthamiana* infection by the tobacco mosaic virus activates a SA induced protein kinase (SIKP) that phosphorylates WRKY1, resulting in increased binding activity (267).

MeJA elicitation can activate post-translational modifications, such as phosphorylation as discussed in section 2.7. To observe whether the addition of MeJA affects the regulatory activity of the TcWRKYs transfected protoplasts were treated with 100μM MeJA during the incubation period. MeJA was able to alter the regulatory activity of both TcWRKYs. WRKY09595 had increased repression activity in four promoters (*TASY*, *DBAT*, *BAPT* and *DBTNBT-1*), while WRKY19284 had additional activity as both a repressor and an activator. With the *DBAT* and *BAPT* promoters the addition of MeJA lead to repression by WRKY19284; while WRKY19284 functioned as an activator with *T10βH* and *DBTNBT-2* (Figure 48). WRKY33 activity has been shown to be regulated by phosphorylation at its N terminus in a five serine cluster (165). Both WRKY09595 and WRKY19284 have clusters of 4 and 5 serine residues, respectively, at similar locations within their N terminus that have a high potential to be phosphorylated (predicted using NetPhos2.0 (268)). This data shows that TcWRKYs are affected by situational cues within the plant that can alter their regulatory activity.

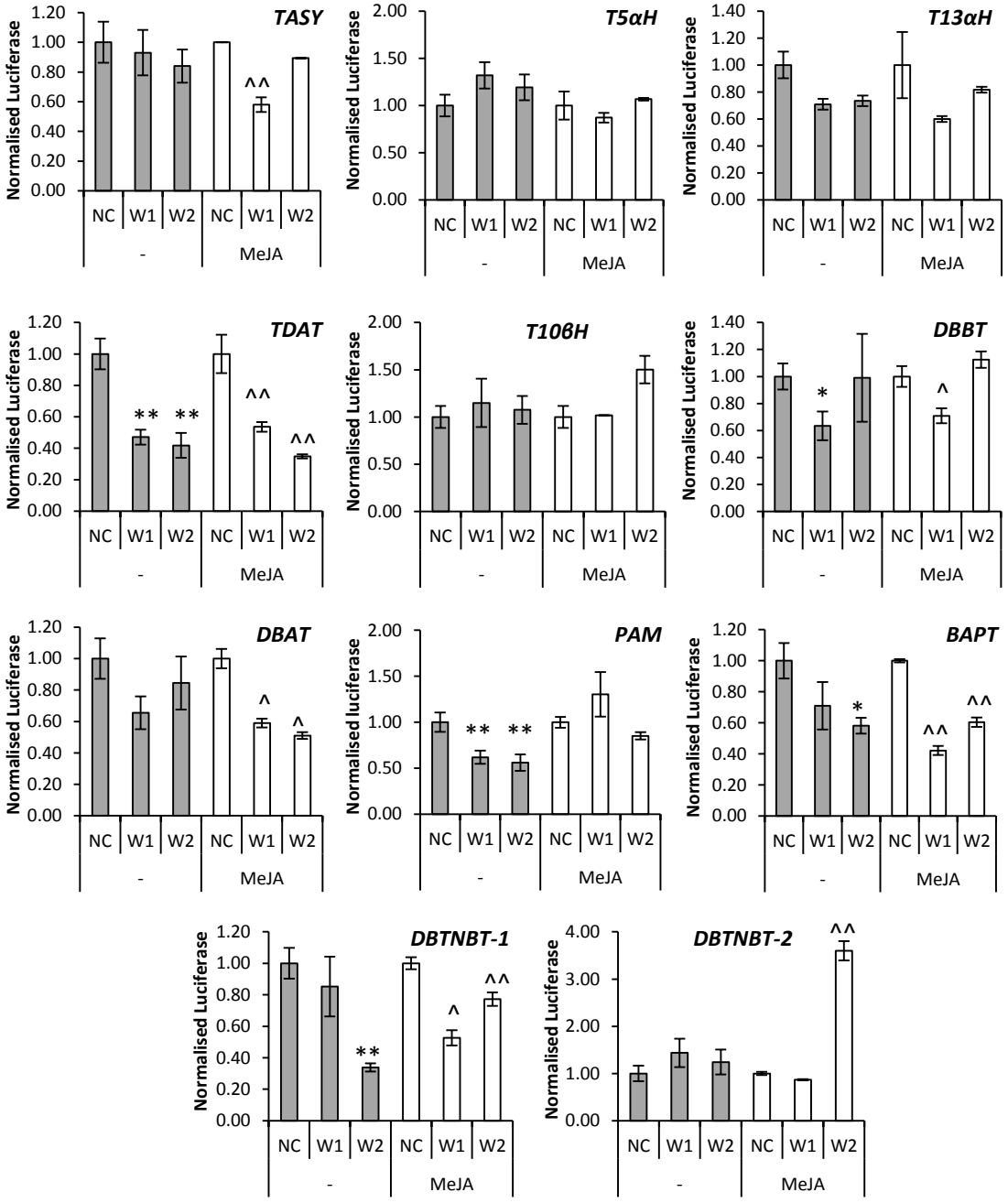


Figure 48 – Analysis of methyl jasmonate (MeJA) addition on TcWRKY regulatory activity. The summary of the *Arabidopsis thaliana* transient expression assay testing of WRKY09595 (W1) and WRKY19284 (W2) with the 10 paclitaxel biosynthetic promoters (see Table 2 for promoter abbreviations) with and without MeJA treatment (100 µM). The error bars represent standard error, $n \geq 3$, significance was determined compared to the negative control (NC) under either normal conditions or MeJA addition using a Student *t*-test, for normal conditions * *p* value 0.05, ** *p* value 0.01 and with MeJA addition ^ *p* value 0.05, ^^ *p* value 0.01.

The exact mechanism by which TcWRKY activity is altered cannot be concluded from this study but future experiments using post-translational inhibitors could elucidate the mechanism. For example, protoplasts could be treated with MeJA and staurosporine, a broad spectrum kinase inhibitor. Staurosporine was able to block the activation of the defence gene GST1 and WRKY39 by flg22 in *fls2 A. thaliana* protoplasts (269). However caution should be taken when interpreting this data as MeJA may activate different systems in *A. thaliana* compared to *Taxus spp.* that may erroneously alter the regulatory activity of the TFs.

3.5 Discussion

An optimised *A. thaliana* protoplasts transient assay (TEA) using the highly sensitive dual luciferase assay was used to perform a screen to identify interactions between the 19 candidate TFs and 10 paclitaxel biosynthetic promoters. During optimization of the transfection protocol the DNA purity and pH of the PEG solution were identified as important factors affecting the transformation efficiency. This system was used to perform in excess of 800 reactions and 72 significant interactions were observed leading to both an increase and decrease in promoter activity.

Every promoter under investigation was found to interact with at least 3 TFs and could be both activated and repressed. This suggests that when the expression of the paclitaxel biosynthetic pathway is increased after MeJA elicitation it is still tightly regulated. The comparative expression of the biosynthetic genes after MeJA addition does not correlate with the number of activators and repressors found to interact with respective promoters. TFs from the AP2 and MYB families were primarily activators of late biosynthetic genes, which are speculated to be rate limiting steps in the pathway; while NAC, bHLH and WRKY TFs were found to function as repressors across the entire pathway. TFs that produced the strongest interactions and regulated the largest proportion of the pathway, such as MYB10855 and bHLH08058, are the best candidates to use in bioengineering efforts to increase paclitaxel production in *Taxus spp* CMCs.

AP2 TFs have been reported to activate JA responsive pathways. The ORCA proteins, for example, control TIA production in *A. annua*, however the TcAP2s are not closely related (99, 176). The functionality of the EAR domain present in AP200499 was confirmed using truncated and site directed mutagenesis mutants. Repression was eliminated when the mutations occurred within the EAR domain. AP200499 is likely to be involved in a negative feedback loop working as a transcriptional repressor to control the paclitaxel biosynthesis and may function in a similar way to ZCT1, ZCT2 and ZCT3 that regulate alkaloid production in *C. roseus* (177).

MYB10855 was identified as a possible overall regulator of the paclitaxel biosynthetic pathway strongly activating 8 out the 10 promoters tested. Research searching for master switches that mimic the JA response has currently been unsuccessful. ORCA3 is a classic example of a TF that can up-regulate part of a biosynthetic pathway. It can up-regulate four TIA biosynthetic genes but does not activate two other genes in the pathway, therefore overexpression of ORCA3 did not lead to an increase in TIA production (176). MYB10855 is able to interact with a significant number of promoters but it cannot activate *BAPT* or *T5αH*. BAPT has been proposed to be a rate limiting step of the pathway (219); however this may not be true in CMCs because BAPT expression is elevated compared to DDCs (221). The concerted activation of the entire pathway is therefore likely to require a combination of TFs.

MYB10855 is evolutionarily distinct compared to the other TcMYBs and previously identified JA responsive MYBs. MYB10855's closest *A. thaliana* homology is LOF2 which is involved in meristem regulation (125, 261), while NAC05638 was also found to be related to CUC1 and CUC2

which controls meristem formation (261, 263). Secondary metabolism is often spatially regulated e.g. arteminsin biosynthesis occurs exclusively in the trichromes after MeJA elicitation (100) and paclitaxel is produced in the bark of *Taxus spp*. The role of TFs in regulating this specificity is unknown, but higher expression of TFs linked with meristem formation might be connected with the increased yields of paclitaxel in CMCs. NAC05638 is significantly down-regulated in DDCs after MeJA elicitation and the logFC of MYB10855 in DDCs is half that of CMCs. LOB TFs are also specifically expressed at organ boundaries and 6 were identified in the *T. cuspidata* transcriptome, four of which were significantly up-regulated at 0.5 h after MeJA elicitation in CMCs. In DDCs only one of these TFs was up-regulated, therefore it may be interesting to test candidates from this family to observe if they can alter expression of the paclitaxel biosynthetic promoters. A greater insight into how CMCs produce higher levels of paclitaxel could help in engineering efforts to further improve production.

One of the limitations of using *A. thaliana* as a system is that cofactors that might be required for activity that are not present and endogenous signalling components may erroneously activate the *Taxus* TFs. It would have been implausible, however, to conduct this screen using a *Taxus* based assay, such as particle bombardment, and previous groups employing this system have only concentrated on one or a few TFs (183, 184). The screen was not completed with MeJA as most reports of TEAs identifying regulators in the literature (such as the identification of major JA regulators MYC2 and ORA47 (114) did not add exogenous elicitors. The regulatory activity of the WRKY TFs was altered after MeJA treatment suggesting that MeJA activated endogenous signalling pathways that could have led to post-translational modifications, such as phosphorylation. Further testing with post-translational inhibitors is required to identify the mechanism for the change in regulatory activity.

Out of the 19 candidate TFs identified in this study only two have been reported previously, WRKY09595 by Li *et al.* (183) and a version of bHLH08058/bHLH08058-FE (TcJAMYC4) by the Roberts group (184). The TEA data for WRKY09595 does not concur with that of Li *et al.*, however their work did not include a vital control in their *in vivo* expression experiments making it difficult to draw any conclusions between the results. Both the Roberts group and our data identified bHLH08058 as a negative regulator. However, our TEA results found it to repress 8 promoters and preliminary results with bHLH08058-FE showed it repressed the entire pathway, while TcJAMYC4 was found to only repress five paclitaxel promoters. Other than the difference in length between the proteins (which could have caused the differences in regulatory activity) our results were conducted in *A. thaliana* while their experiments were completed in *Taxus* with the addition of MeJA. It is important to confirm the interaction observed in our TEA in *Taxus* to observe if this alters the regulatory activity of the TF.

bHLH08058 is a homolog of MYC2 best known to act as a positive regulator, for example in flavonoid biosynthesis in *A. thaliana*. However, our results and the Roberts data both found bHLH08058 to act as a negative regulator. MYC2 can activate other TFs, such as ORCA3 in *C.*

roseus thereby indirectly up-regulating part of the TIA pathway (100); therefore it is possible that bHLH08058 might regulate other TFs that activate the paclitaxel biosynthetic pathway. MYC2 has been shown to act as a negative control on tryptophan derived indole glucosinolate synthesis (100), so this is not the first example of a MYC2-like protein acting as a repressor. Negative feedback loops are important to ensure that the plant avoids wasting energy on unnecessary defence responses.

The goal of this project is to identify TFs which can up-regulate the paclitaxel biosynthetic pathway and exploit these TFs to try and improve yields of paclitaxel in CMCs. It is therefore important to consider what the rate limiting steps in the pathway are in order to concentrate research in these areas. In *T. baccata* cell cultures the hydroxylases in the pathway have higher expression than the transferases (25). PAM and TASY were observed to be highly expressed and other studies in *T. candensis* concur that the first committed step of the pathway is not rate limiting (270). There was an accumulation of baccatin III indicating that DBAT was also not a rate determining step, even though its expression was observed to be low. This work suggested that the transferases BAPT and DBTNBT may be limiting the production of paclitaxel (25). In *T. baccata* platelets aerial parts of the plant contained high levels of deacylbaccatin III suggesting that DBAT was limiting the flux of the pathway. The expression of BAPT and DBTNBT increased after 12 months leading to a concomitant increase in paclitaxel (259). Expression of the biosynthetic genes differs between the roots and the aerial parts of the plant and was age dependent. The data available in the literature is not definitive about which steps are rate determining. CMC are spatially different to the work previously conducted in cell cultures and platelets therefore the rate limiting enzymes may be different. A number of studies have suggested BAPT to be rate limiting and in CMCs its expression is up-regulated compared to DDCs, which could be partially linked to the higher production yields (221). MYB10855 can activate 8 promoters but BAPT is not among them, however it can up-regulate DBAT and DBTNBT which have also been suggested to limit the production of paclitaxel. AP204485 and MYB15401 were identified as positive regulators of BAPT and therefore a combination of TFs may be required to produce a concerted increase in paclitaxel production. DBAT in certain studies has been shown to be rate limiting and a large number of activators were identified in the TEA screen. This provides a good set of tools to try and engineer an increase in biosynthetic gene expression in the future.

A majority of the examples of TFs in the literature that are MeJA induced and regulate secondary metabolism were reported in angiosperms. The JA core module has been identified previously in *T. chinensis* (210) and was observed in our transcriptomic data; therefore it is likely that other JA signalling mechanisms are conserved, but the secondary metabolite pathways elicited are species specific (100). Care must therefore be taken when trying to identify homology with known JA responsive TFs as sequence similarity outside characterised domains is low. There is approximately 300 million years evolutionary distance between gymnosperms and angiosperms, with limited sequence information available for gymnosperm TFs (181). This makes our work more

unique because little prior research has been conducted in this area.

The TEA screen identified 29 activators and 42 repressors of the paclitaxel biosynthetic promoters. These interactions need to be confirmed *in planta* but provide a good basis for future work. A number of the TFs can alter large percentages of the pathway but no overall regulator was identified. This screen has provided a wide range of different TFs candidates that can be explored to try and improve paclitaxel production levels on CMCs.

Chapter 4 - Combinatorial analysis of transcription factor regulation

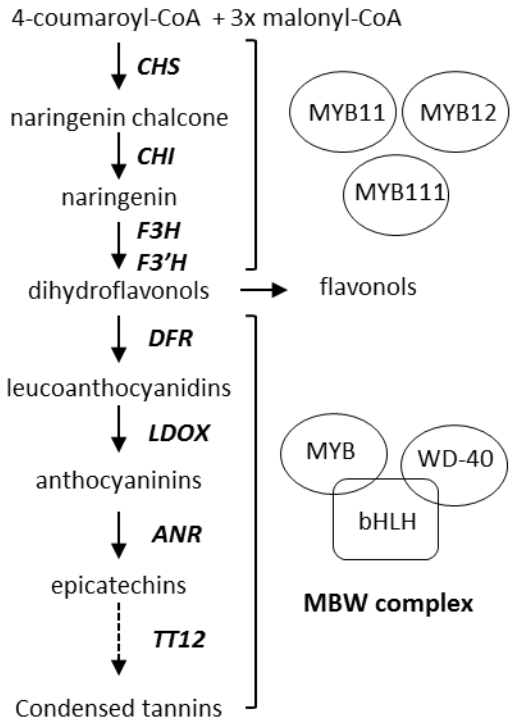
The yeast-two-hybrid work in this chapter was completed in collaboration with visiting academics from Jiangsu Normal University (Jiangsu, China) Cao Xiao, Xing Qin and Ju Xiuyun.

4.1 Introduction

TFs can work individually or in combination to achieve metabolic reprogramming after MeJA elicitation. The combinatorial action of AP2 and bHLH TFs has been demonstrated in *C. roseus* (271) and *N. tabacum* (185), where ORCA3 and MYC2, and their *N. tabacum* homologs, work in concert to regulate TIA and tobacco biosynthesis respectively. JAZ proteins have been shown to interact with numerous targets (99, 105), including the JA responsive TF MYC2, repressing their ability to regulate downstream target genes (107, 108).

4.2 Combinatorial action of MYB and bHLH TFs

One of the best characterised examples of plant combinatorial gene control is the regulation of flavonoid production by particular sub classes of MYB and bHLH TFs, which has been demonstrated in a variety of different species (Figure 49) (122, 264). The regulation occurs *via* a tertiary complex comprising a R2R3-MYB TF, a bHLH TF and a WD-repeat (WD-40) protein (the MBW complex) (272). In *A. thaliana* the early steps of the anthocyanin biosynthetic pathway are regulated by three functionally redundant MYBs, MYB11, MYB12 and MYB111; while the late pathway is regulated by the MBW complex comprised of MYB TFs PAP1, PAP2, TT2 and MYBL2, bHLH TFs GL3, EGL3 and TT8 and the WD-40 protein TTG1 (273). In *Z. mays* the MYB TFs C1 and PL, the bHLH TFs R and B and the WD-40 protein PAC1 have been shown to act in a MBW complex to regulate the anthocyanin pathway (144, 274, 275). The MBW complex has also been implicated in the regulation of anthocyanin production in numerous other species including Petunia, grape and apple and the TFs involved are summarised in Figure 49.



	TF in MBW complex		
Species	**MYB**	**bHLH**	**WDR**
Arabidopsis thaliana	PAP1 PAP2 MYB113 MYB114	GL3 EGL3 TT8	TTG1
Petunia hybrid	AN2 AN4	AN1	AN11
Zea mays	C1 PL	R B	PAC1
Apple	MYB1 MYB10 MYBA	bHLH3 bHLH33	
Grape	MYBA1 MYBA2 MYBA5a MYBA5b	MYC1 MYCA1	WDR1 WDR2

Figure 49 – The combinatorial action of the MYB-bHLH-WD-40 (MBW) complex. The flavonoid biosynthetic pathway in *Arabidopsis thaliana* with the known regulators of early and late biosynthetic genes. The pathway starts with the general phenylpropanoid metabolism that produces the precursors 4-coumaroyl-CoA and malonyl-CoA. These precursors undergo numerous transformations by the early biosynthetic genes (EBGs) chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H) and flavanone 3'-hydroxylase (F3'H) to produce dihydroflavanols. These can be used to produce flavanols or can be further transformed by the late biosynthetic genes (LBGs) dihydroflavonol 4-reductase (DFR), leucoanthocyanidin dioxygenase (LDOX), anthocyanidin reductase (ANR) and Transparent testa 12 (TT12) to form anthocyanins or condensed tannins. The biosynthetic pathway is adapted from Li 2014 (273). The EBGs are regulated by three functionally redundant MYB TFs MYB11, MYB12 and MYB111; while the LBGs are regulated by the R2R3-MYB/bHLH/WD40 (MBW) complex. The components of the MBW complex that regulate anthocyanin biosynthesis in *A. thaliana*, Petunia, maize, apple and grape are summarised in the table (264).

4.2.1 Prediction of TcMYB and TcbHLH interactions

Comparison of our four TcMYBs with the known bHLH interaction motif [DE]Lx$_2$[RK]x$_3$Lx$_6$Lx$_3$R, which is located in the R3 repeat of MYB TF (Figure 50) and conserved across angiosperms and gymnosperms (122), only identified the motif in MYB12379. The closest *A. thaliana* homolog of MYB12379 is WER which is known to act in combination with GL3 and EGL3 to regulate epidermal cell development (135, 276). There is no defined interaction motif in bHLH TFs and blast analysis shows that neither of the TcbHLHs is closely related to GL3, EGL3 and TT8 (Figure 45) therefore it is hard to predict if the proteins will interact with MYB TFs.

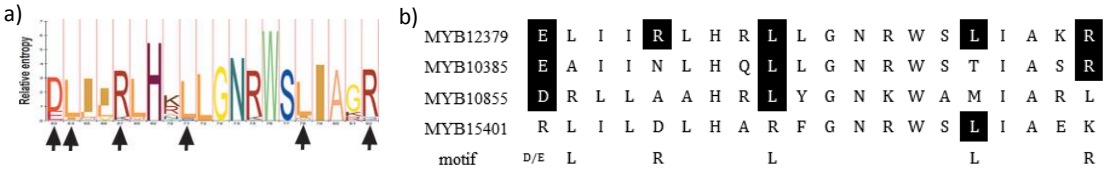


Figure 50 – The identification of bHLH interaction motif in the four *Taxus cuspidata* MYB transcription factors (TFs). a) The sequence logo of the interaction motif found in the R3 repeat of R2R3-MYB proteins, arrows indicate conserved residues (sequence logo adapted from Feller *et al*. 2011 (122)). b) The alignment of the R3 repeat of the four MYB TFs in the 19 candidate TFs with the residues that are conserved with the bHLH interaction motif [DE]Lx$_2$[RK]x$_3$Lx$_6$Lx$_3$R

highlighted in black. MYB12379 is the only TcMYB that contains all of the residues and was therefore predicted as the most likely to interact with bHLH TFs.

Yeast-two-hybrid (Y2H) and protoplast TEAs were completed concurrently to observe if the TcMYBs and TcbHLHs could interact and whether this resulted in a change in promoter activity.

4.3 Yeast two hybrid

4.3.1 Introduction

Yeast two hybrid (Y2H) is a powerful tool used to detect protein-protein interactions in living *S. cerevisiae* cells. This *in vivo* technique utilizes the modular structure of the galactose dependent transcriptional activator Gal4 in yeast. The two genes under investigation are cloned in either a bait or a prey vector and transformed into an appropriate yeast strain. The bait gene (X) is expressed as a fusion protein with the Gal4 DNA binding domain (BD) attached, while the prey gene (Y) is expressed as a fusion protein with Gal4 activation domain (AD). If the two proteins interact then the BD and AD are brought proximal to each other and this leads to transcriptional activation of the reporter genes (Figure 51).

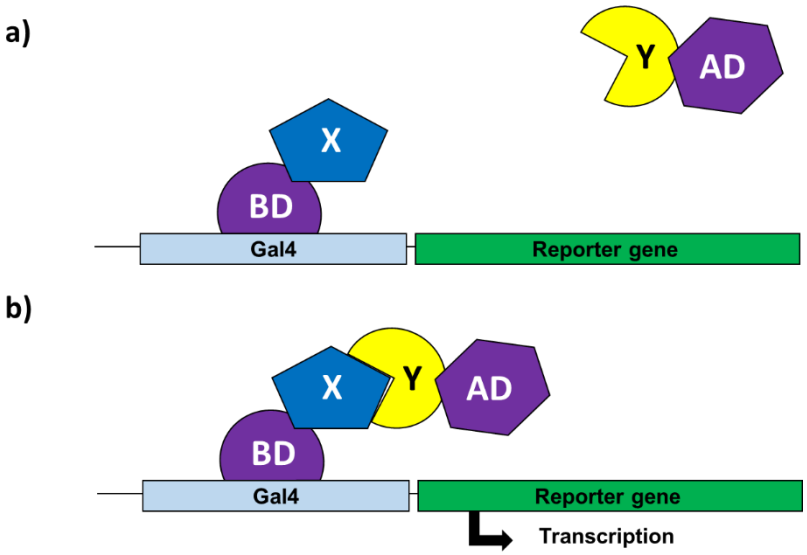


Figure 51 – The yeast two hybrid system. The bait gene X is fused to the Gal4 binding domain (BD) while the prey gene Y is fused to the Gal4 activation domain (AD). The BD binds to the activator sequence in the Gal4 promoter. If the X and Y proteins interact this brings the AD and BD together allowing transcriptional activation of the reporter gene.

The advantages of this system are that it is easy to implement, inexpensive and *in vivo* therefore the proteins are more likely to be in their native conformation. Multiple reporter genes can be employed to increase the stringency of the assay, however this penalises the detection of weak interactions. Limitations of the technique include that the interaction must occur in the nucleus, the system produces a large number of false positives and that relevant cofactors, that can introduce post-translation modifications, may not be present yeast (277).

4.3.2 Results of Y2H assay

4.3.2.1 Testing for interaction between TcMYBs and TcbHLHs

A screen for possible interactions between the TcMYB and TcbHLH TFs was conducted

using Y2H assays. The four MYB and two bHLH TFs previously cloned into pDONR221 were cloned into the Gateway™ bait vector pDEST32 and prey vector pDEST22 respectively (see Appendix 1 for vector maps). The resulting constructs were transformed in the yeast strain AH109 and plated out onto selective (SC-L-W-H) and non-selective media (SC-L-W). No interaction was observed between MYB12379 and the TcbHLHs even though a bHLH interaction motif had been identified in the MYB TF. The only interaction observed in the screen was between MYB10855 and bHLH11748 but the interaction was not strong as addition of 0.5mM 3-amino-1,2,4-triazole (3AT) inhibited the interaction (Figure 52).

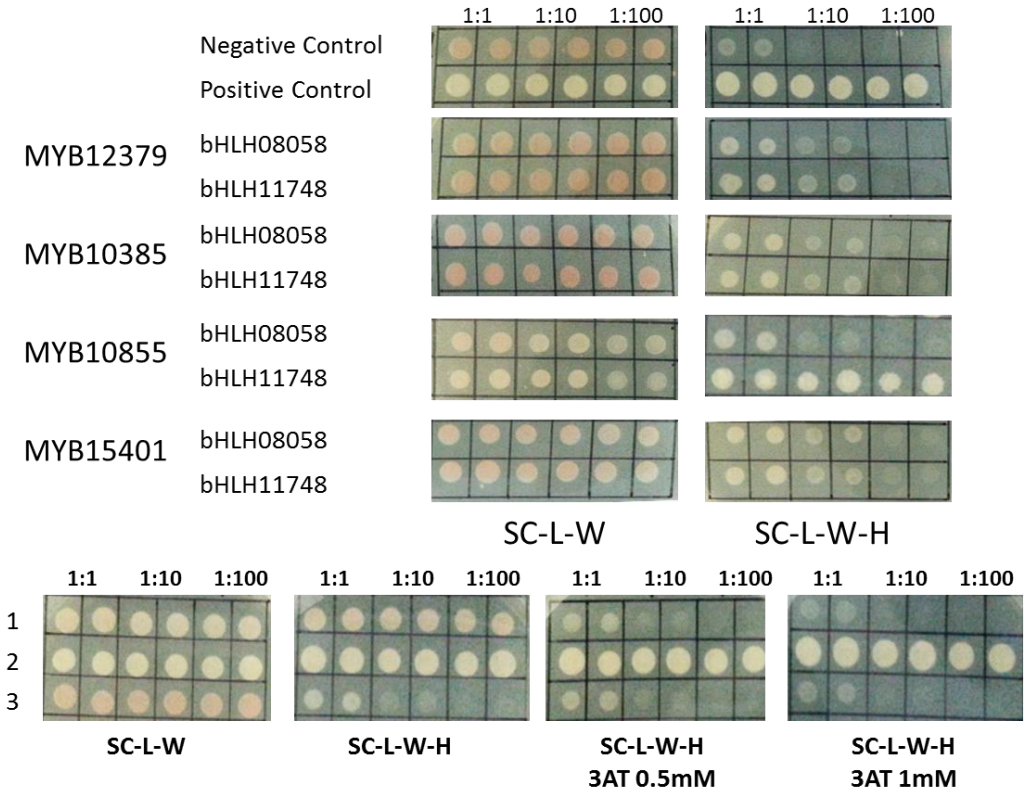


Figure 52 – Testing for possible interactions between *Taxus cuspidata* MYB and bHLH TFs using yeast-two-hybrid assays. pDONR221-MYB constructs were cloned into the bait Gateway™ vector pDEST32 and the pDONR221-bHLH constructs cloned into the prey Gateway™ vector pDEST22. The yeast strain AH109 was transformed with the stated combination of TFs. The top panel shows the screen for interactions between the TcMYBs and TcbHLHs. The negative control is yeast transformed with the empty vectors pDEST32 and pDEST22 and the positive control is the known interaction between *A. thaliana* COI1 and JAZ. The transformations were plated out on selective media deficient in leucine, tryptophan and histidine (SC-L-W-H) and non-selective media deficient in leucine and histidine (SC-L-W). The ratios show the dilution of an overnight culture of the transformed yeast which was subsequently spotted out onto the media. The only interaction observed was between MYB10855 and bHLH11748. The lower panel shows the loss of the MYB10855-bHLH11748 interaction with the addition of increasing concentrations of 3-amino-1,2,4-triazole (3AT). Lane 1 = MYB10855 and bHLH11748, 2 = Positive control, 3 = Negative control. The MYB10855-bHLH11748 interaction was abolished after the addition of only 0.5mM 3AT, showing that it is very weak.

4.3.2.2 Testing for interaction between TcbHLH and JAZ proteins

The bHLH TF MYC2 is a well-known target of JAZ proteins (107, 108). Six JAZ proteins were identified as significantly up-regulated after MeJA elicitation (Figure 18). Out of the six candidate JAZ proteins four were found to contain the conserved TIFY motif, required for dimerization and the CO, CO-like, TOC1 (CCT) -2 domain involved in protein-protein interactions

(105). The three JAZ proteins that were most highly up-regulated after MeJA elicitation were chosen for testing in Y2H assays to observe for possible interaction with TcbHLHs. bHLH08058 is a MYC2 like TF therefore it can be hypothesised that it is likely to interact with the JAZ proteins. No interactions were identified (Figure 53) however, and a possible reason is that JAZ proteins only interact with bHLHs from the subgroup III and bHLH11748 is not in this clade.

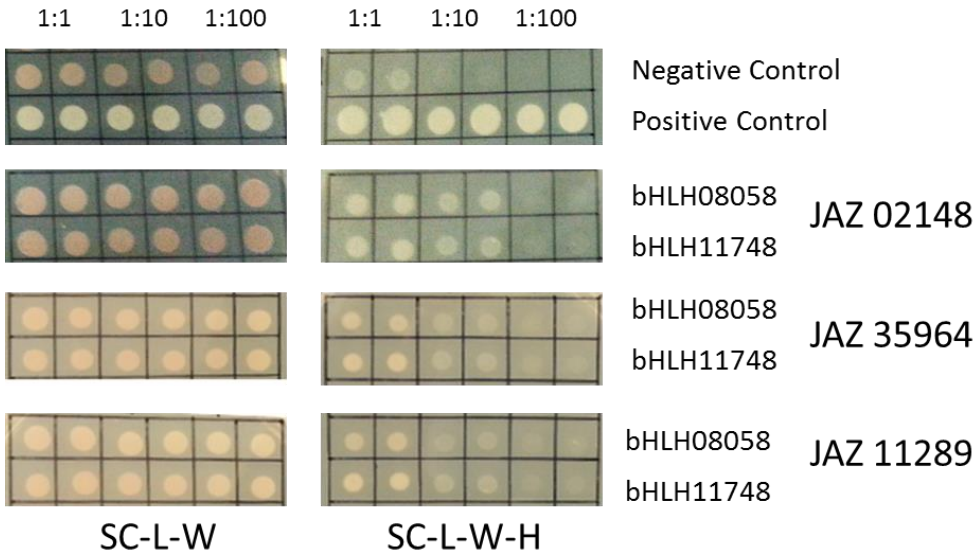


Figure 53 – Testing for possible interactions between three *Taxus cuspidata* JAZ proteins and two TcbHLH TFs using yeast-two-hybrid assays. The yeast strain AH109 was transformed with the stated combination of proteins. JAZ proteins were cloned into bait Gateway™ vector pDEST32 and the bHLH TFs were cloned into the prey Gateway™ vector pDEST22. The negative control is yeast transformed with the empty vectors pDEST32 and pDEST22 and the positive control is the known interaction between *A. thaliana* JAZ1 and MYC2. The transformations were plated out on selective media deficient in leucine, tryptophan and histidine (SC-L-W-H) and non-selective media deficient in leucine and histidine SC-L-W. The ratios show the dilution of an overnight culture of the transformed yeast spotted out onto the media. No interactions were observed between the JAZ proteins and the TcbHLHs.

One of the possible reasons that few interactions were identified in the Y2H assays is that the bHLH08058 used in the experiments was a truncated form of the TF (see 2.4 The open reading frame of the bHLH08058 transcription factor) which does not contain the JAZ interacting domain (JID) at its N-terminus, shown in *A. thaliana* to be required for JAZ protein interaction (Figure 10) (230). The extended version of bHLH08058 does contain some of the JID (Figure 25) but due to time restraints (as the ORF was only successfully cloned in July 2015) Y2H assays have not yet been repeated with bHLH08058-FE. All the experiments performed above were with the truncated form of the TF which is possible reason why no interactions were identified.

4.4 Results of the combinatorial TEA

Verification that the TEA system could be used for combinatorial analysis was established using the positive control of PAP1 and EGL3 with the flavonoid dihydroflavonol 4-reductase (*DFR*) promoter. Individually PAP1 and EGL3 cannot up-regulate *DFR* however in combination they were able to significantly increase gene expression (278) (Figure 54). The combinatorial results were compared to the original TEA data. This is not a perfect comparison as 4 µg of each TF was added in the individual experiments while 2 µg of each TF was added in the combinatorial assays. If the

addition of a second TF had no effect there still may be a difference as 2 µg less of the TF is available. However to redo the TEA using 2 µg of empty vector and 2 µg of TF would require an extra 260 assays and this was too expensive and time consuming to complete.

With the exception of *T10βH*, the addition of the TcbHLHs did not lead to a significant increase in promoter activity as observed in control of PAP1 and eGL3 activating *DFR* (Figure 55). Individually MYB12379, bHLH11748 and bHLH08058 were not able to alter *T10βH* activity, but when used in combination they were able to significantly activate *T10βH* activity.

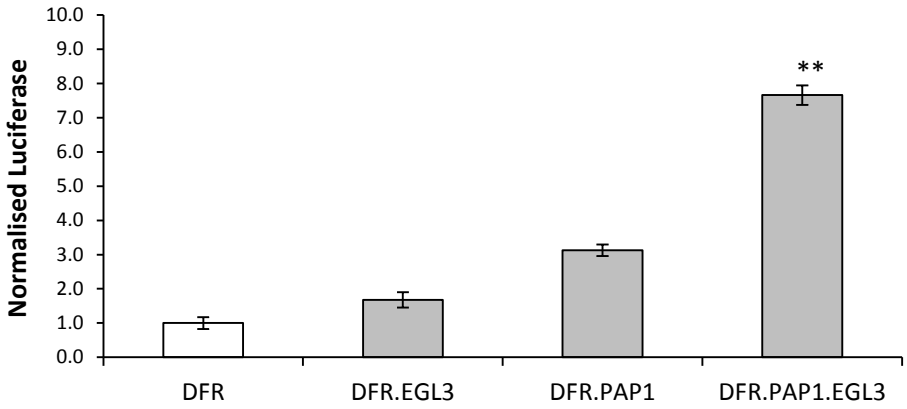


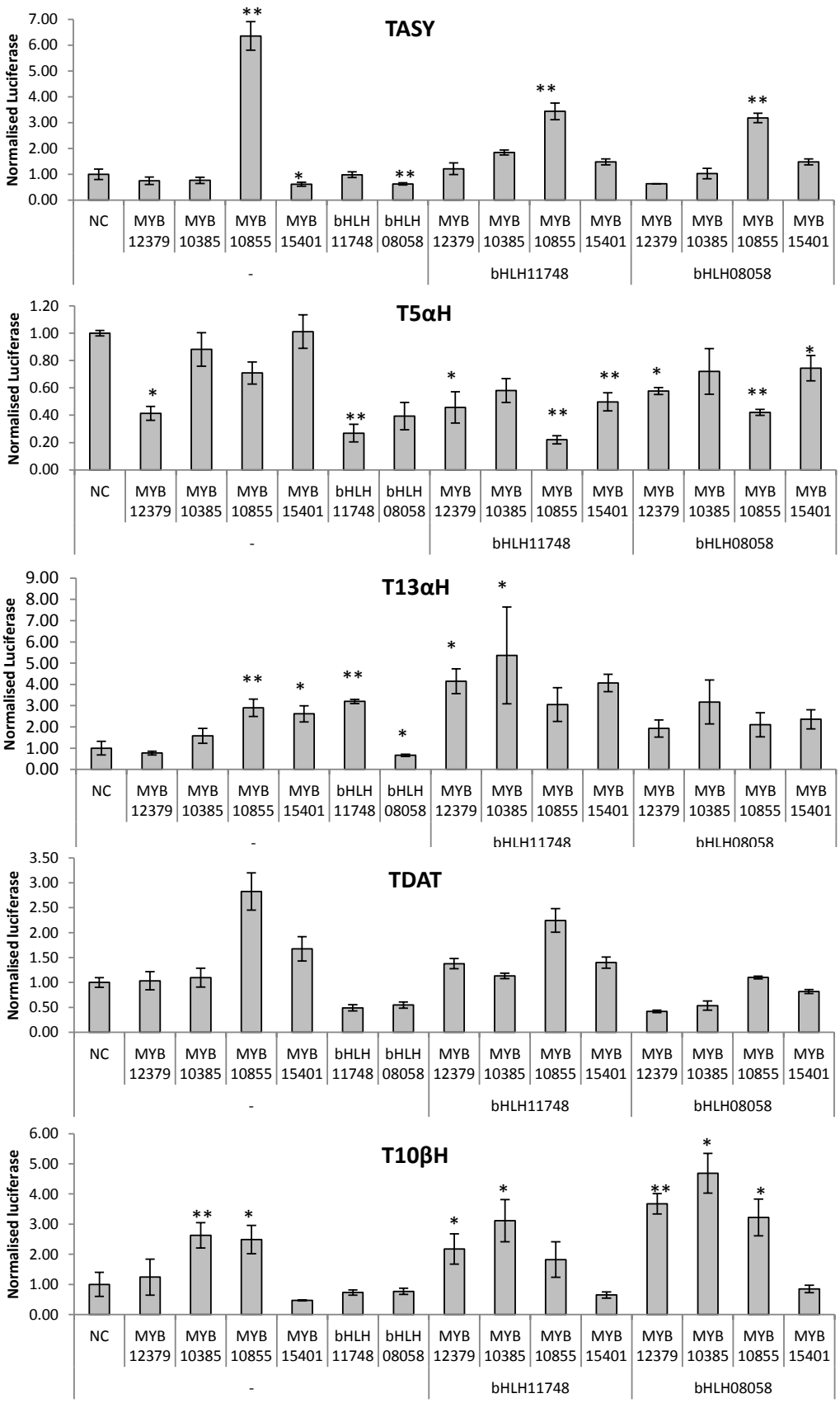
Figure 54 – Results of the combinatorial *Arabidopis thaliana* protoplast transient expression assay (TEA) with the known combinatorial regulators of the flavonoid biosynthesis gene dihydroflavonol 4-reductase (DFR), PAP1 and EGL3. This combination was used to show that the TEA can be employed to test for combinatorial regulation of the paclitaxel biosynthetic promoters. Individually EGL3 and PAP1 do not activate *DFR* expression but when used in combination they significantly activate *DFR* expression. Significance was determined using a Student *t*-test, $n \geq 3$, ** *p*-value $\leq 1x10^{-8}$, error bars denote standard error. The graph shows normalised luciferase (the luciferase values obtained in the TEA normalised to DFR with empty vector) against *DFR* with or without the addition of EGL3 and PAP1 individually or in combination.

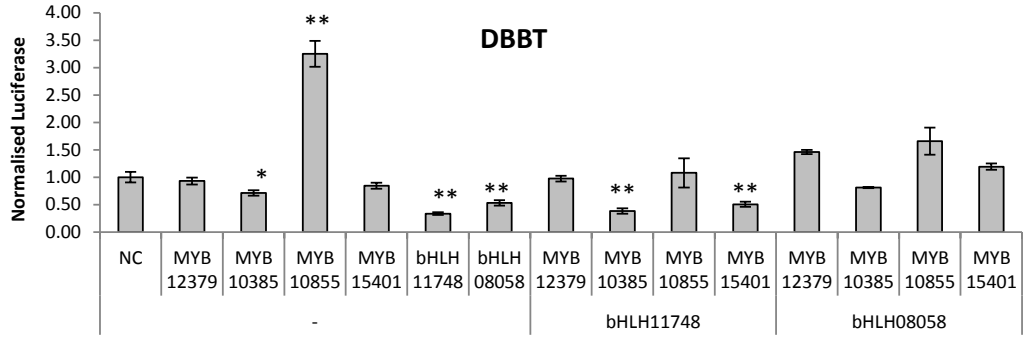
The combinatorial TEA results show that there is a balance between the activation of MYB10855 and the repression of the TcbHLHs which is promoter dependent. MYB10855 can activate 8 promoters 8 paclitaxel promoters (*TASY, T13αH, TDAT, T10βH, DBBT, DBAT, PAM and* DBTNBT) but the addition of bHLH08058 leads to the loss of this activity in *T13αH*, *TDAT*, *DBBT* and *DBTNBT*, while bHLH11748 removes the up-regulation in *T13αH*, *T10βH*, *DBBT*, *PAM*, and *DBTNBT*. Both TcbHLHs were able to block MYB10855 activation of *T13αH, DBBT* and *DBTNBT* and did not abolish the activation of *TASY* and *DBAT*. The repression activity of the TcbHLHs was lost in the majority cases with the addition of TcMYBs, with exceptions including bHLH08058 in combination with all TcMYBs against *DBTNBT-1*. These results show that the regulatory activity of the TFs in isolation can be affected by the addition of another TF.

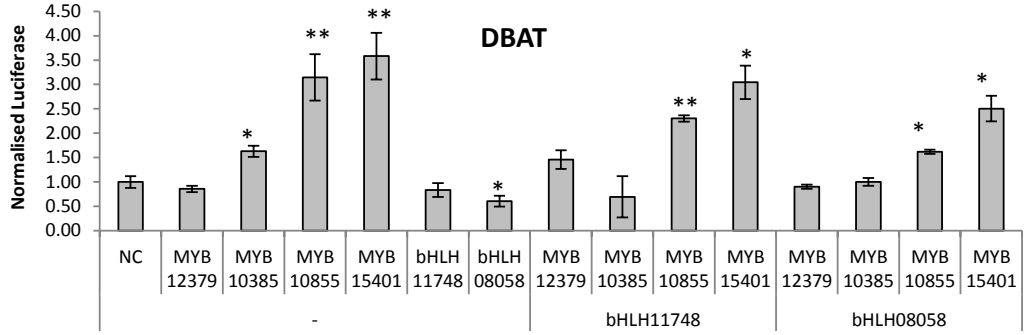
The TFs may be competing for binding sites and stronger interactions are more likely to prevail. There was no consensus, however, between the promoters that retained MYB10855 activity and the respective location of E-boxes and MYB BSs within the promoter (BS identified in section 2.6 see Figure 27). For example both *T10βH* and *PAM* lose MYB10855 activation when bHLH11748 is added but *T10βH* contains an E-box which overlaps with the MYB BS (identified using EMSAs see Chapter 5 for details) while this is not true in *PAM*.

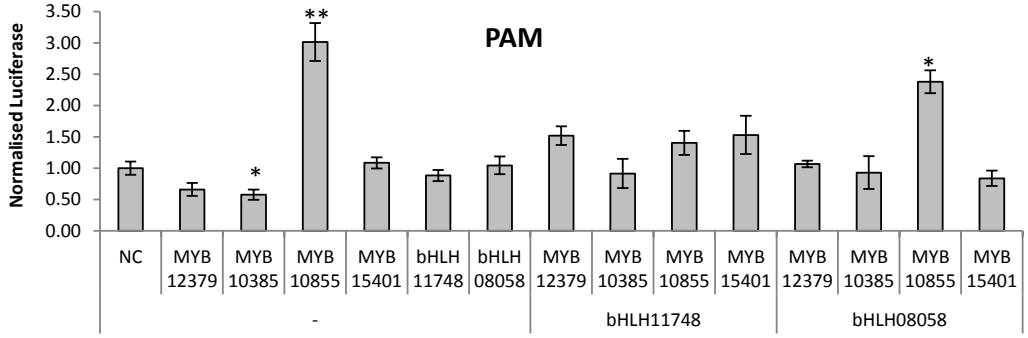
There is also little correlation between the strength of an interaction and its ability to be

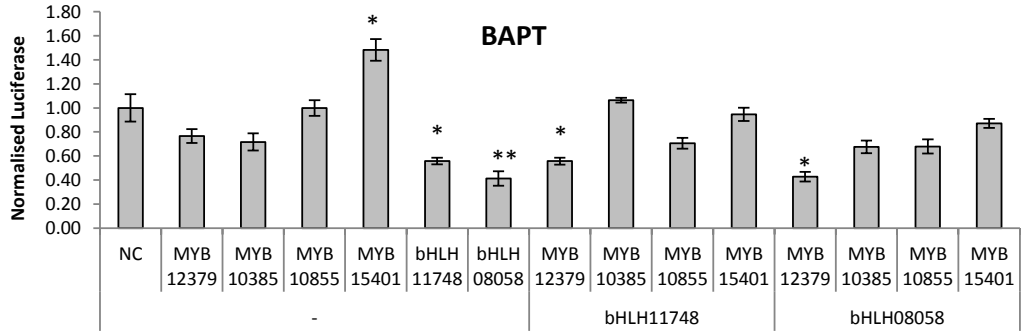
maintained after the addition of second TF. The strong repression of *T5αH* by bHLH11748 was not lost by the addition of TcMYBs (except with MYB10385), but the activation of *DBBT* by MYB10855 which was one of the top three positive interactions was lost after the addition of both TcbHLHs. The change in regulatory activity might be due to the reduction in the quantities of DNA used in the assays. The interaction between MYB10855 and *TASY* is strong and tested in combination with TcbHLHs reductions in promoter activation were observed. Individually bHLH08058 can repress *TASY* while bHLH11748 cannot. The similarity in the reduction of *TASY* activation in the MYB10885-TcbHLH combinations is therefore more likely to be due to the reduced amount of DNA in the transfections rather than suppression of activity by the TcbHLHs.

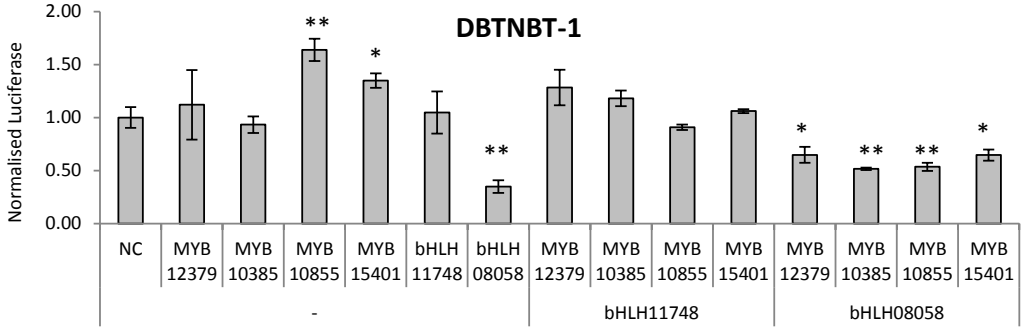

TASY
T5αH
T13αH
TDAT
T10βH
Normalised Luciferase
Normalised luciferase
NC
MYB 12379
MYB 10385
MYB 10855
MYB 15401
bHLH 11748
bHLH 08058
-
bHLH11748
bHLH08058


DBBT
Normalised Luciferase
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
NC
MYB 12379
MYB 10385
MYB 10855
MYB 15401
bHLH 11748
bHLH 08058
MYB 12379
MYB 10385
MYB 10855
MYB 15401
MYB 12379
MYB 10385
MYB 10855
MYB 15401
-
bHLH11748
bHLH08058


DBAT
Normalised Luciferase
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
NC
MYB 12379
MYB 10385
MYB 10855
MYB 15401
bHLH 11748
bHLH 08058
MYB 12379
MYB 10385
MYB 10855
MYB 15401
MYB 12379
MYB 10385
MYB 10855
MYB 15401
-
bHLH11748
bHLH08058


PAM
Normalised Luciferase
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
NC
MYB 12379
MYB 10385
MYB 10855
MYB 15401
bHLH 11748
bHLH 08058
MYB 12379
MYB 10385
MYB 10855
MYB 15401
MYB 12379
MYB 10385
MYB 10855
MYB 15401
-
bHLH11748
bHLH08058


BAPT
Normalised Luciferase
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
NC
MYB 12379
MYB 10385
MYB 10855
MYB 15401
bHLH 11748
bHLH 08058
MYB 12379
MYB 10385
MYB 10855
MYB 15401
MYB 12379
MYB 10385
MYB 10855
MYB 15401
-
bHLH11748
bHLH08058


DBTNBT-1
Normalised Luciferase
2.00
1.50
1.00
0.50
0.00
NC
MYB 12379
MYB 10385
MYB 10855
MYB 15401
bHLH 11748
bHLH 08058
MYB 12379
MYB 10385
MYB 10855
MYB 15401
MYB 12379
MYB 10385
MYB 10855
MYB 15401
-
bHLH11748
bHLH08058

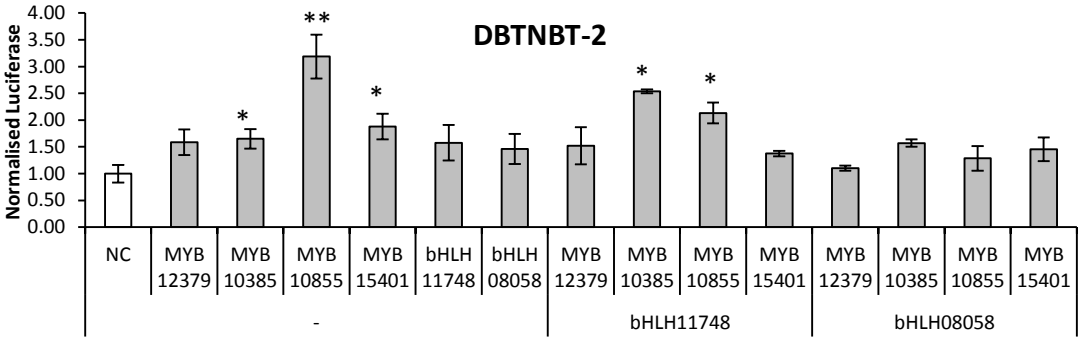


Figure 55 – Comparison of the individual and combinatorial analysis of the four MYBs and two bHLHs in the 19 candidate transcription factor (TFs) against 10 paclitaxel promoters in an *Arabidopsis thaliana* protoplast transient expression assay (TEA). See Table 2 for paclitaxel promoters abbreviations. Significance was determined using a Student *t*-test, $n \geq 3$, * p value 0.05, ** p value 0.01, error bars denote standard error. The graphs show normalised luciferase (the luciferase values obtained in the TEA normalised to negative control (NC). The NC is the promoter transfected with the empty p2GW7,0 vector and highlighted in white) against 10 paclitaxel promoters with or without the addition of TcMYBs (MYB12379, MYB10385, MYB10855 and MYB15401) and TcbHLHs (bHLH08058 and bHLH11748) individually or in combination.

4.5 Discussion

TFs can work cooperatively to achieve metabolic reprogramming and the combinatorial regulation of flavanoid biosynthesis by the MBW complex has been published in numerous species (Figure 49) (264, 273). It is therefore surprising that few interactions between TcMYBs and TcbHLHs were identified in the Y2H and transient expression assays. In Y2H assays only one interaction was observed, between MYB10855 and bHLH11748, which was very weak (Figure 52); and in the TEAs the only significant increase in promoter activity was observed in *T10βH*, where MYB12379-bHLH11748 and MYB12379-bHLH08058 combinations both led to a significant increase in *T10βH* activity (Figure 55). The TEA data showed that the presence of competing TFs can alter the regulatory activity seen in isolation. This is important because *in planta* MeJA causing a transcriptional cascade activating a wide range of different TFs that work as part of a network to control metabolic reprogramming. The strength of the interaction and the relative location of the BSs are likely to be important factors in determining which TFs preferentially regulated gene expression, however no correlation in these factors was observed.

A possible reason that so few interactions were observed could be that WD40 proteins might be required for bHLH-MYB regulatory activity. Six WD-40 proteins were identified in the 1646 DEGs, three of which were up-regulated after MeJA elicitation with their highest logFC values of 0.5 h (Figure 56). Amplification of their ORFs proved problematic however, which is why they were not included during testing. PAP1 and EGL3 form a complex with WD-40 protein TTG1 to regulate anthocyanin production in *A. thaliana* (278). However in the TEA PAP1 and EGL3 were able to activate *DFR* expression without the addition of TTG1; therefore the addition of WD-40 proteins may not be required to regulate gene expression *in vivo*. WD-40 proteins are involved in chromatin remodelling and as the plasmids used in the TEA are not associated with chromatin this may

explain why activation of the *DFR* promoter in protoplasts was not TTG1 dependent; suggesting that the lack of WD-40 proteins in our combinatorial assay is unlikely to be the reason for the lack of observed interactions.

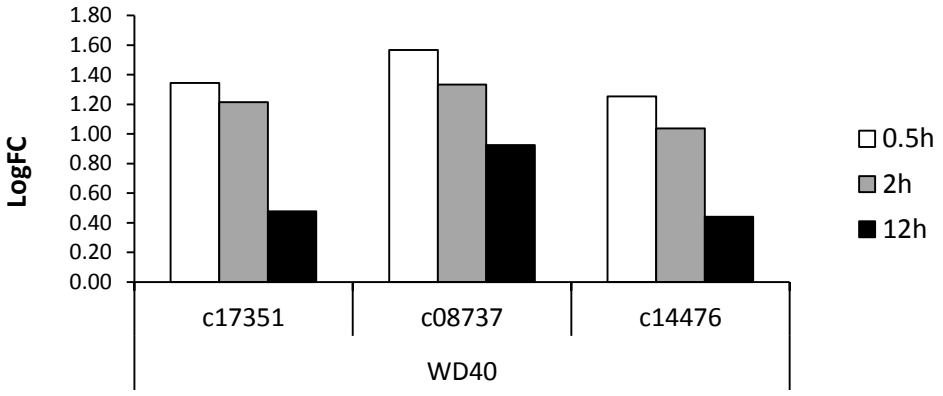


Figure 56 – The log fold change (LogFC) in gene expression of three WD-40 proteins identified in the significantly differently expressed genes (DEGs) at three time points (0.5, 2 and 12 hour (h)) after methyl jasmonate (MeJA) elicitation. WD-40 proteins are known to work in combination with MYB and bHLH TFs to regulate flavonoid biosynthesis. Three WD-40 proteins were significantly up-regulated after MeJA elicitation and their highest logFC value are at 0.5 h.

MYB12379 was predicted to interact with the bHLHs because it contains the known bHLH interaction motif in its R3 repeat (122). The only promoter that showed any combinatorial activation was *T10βH,* where MYB12379 in combination with both TcbHLH's was able to significantly activate the promoter. A reason for the scarcity of combinatorial interactions identified could be that neither of the TcbHLHs are close homologs of the bHLH sub-family known to be involved in MYB combinatorial control. bHLH08058 is a MYC2-like TF while the closest *A. thaliana* homolog of bHLH11748 is ICE2 (SCREAM2) which participates in the cold acclimation pathway (279). ICE2 is in the bHLH sub group IIIb while GL3, EGL3 and TT8 are members of the sub group IIIf.

Species	Transcription factor	454 contig	Score bits	E value
Arabidopsis thaliana	GL3	contig32074	127	$2e^{-29}$
		contig23037	92	$4e^{-19}$
	EGL3	contig32074	126	$2e^{-29}$
		contig23037	115	$3e^{-26}$
	TT8	contig23037	129	$2e^{-30}$
		contig12598	124	$12e^{-28}$
Pinus taeda	6A_I22	contig32074	109	$8e^{-25}$
		contig23037	98	$3e^{-21}$

Table 15 – A table summarising the tblastn (225) **analysis *Arabidopsis thaliana* GL3, EGL3 and TT8 and identified *Pinus taeda* homolog 6A_I22 against the *Taxus cuspidata* CMC transcriptome.** The top two contigs identified with their respective score in bits and e-values are stated. The analysis identified three contigs that could be possible *T. cuspidata* homologs.

BLASTp analysis was employed to identify GL3, EGL3 and TT8 homologs in the *P. taeda* transcriptome (available at http://dendrome.ucdavis.edu/). The sequence 6A_I22_NT_comp18236_c0_seq1 (6A_I22) was the top result for all three TFs with *e*-values of $9e^{-44}$, $3e^{-48}$ and $3e^{-45}$ respectively. *P. taeda* 6A_I22 and *A. thaliana* GL3, EGL3 and TT8 were used in tblastn (225) analysis to identify possible candidates in the *T. cuspidata* CMC transcriptome. The analysis produced three contigs, contig32074, contig23037 and contig12598 (Table 15), none of

which are in the 1646 DEGs. Contig32074 and contig23037 are short at 328 bp and 461 bp respectively; therefore it is unlikely that the ORF identified in these contigs is the full protein. Contig12598 is longer at 963 bp and an identified ORF aligns well with the C-terminal region of TT8 but the protein probably extends in the N-terminal direction (Figure 57). If this bHLH TF is to be tested for possible combinatorial regulation with MYB12379 the complete ORF needs to be identified.

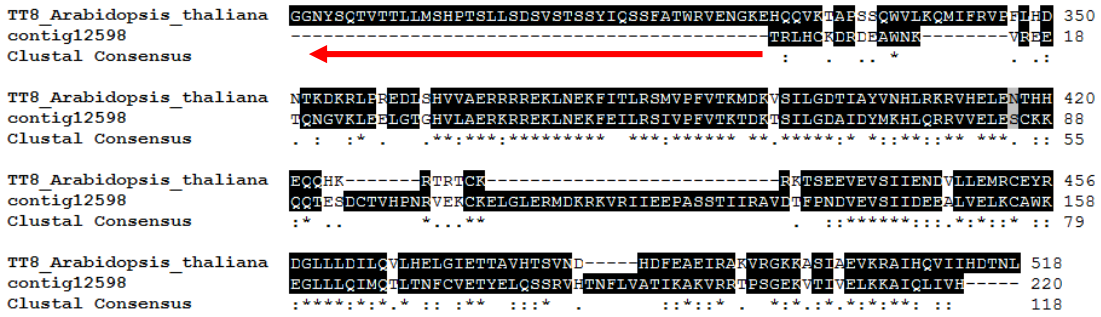


Figure 57 – The protein alignment of an open reading frame (ORF) identified in contig12598 with the *Arabidopsis thaliana* TF TT8. The alignment was produced using ClustalOmega (231, 232) and the image constructed using BioEdit v7.2.5 (233). The two proteins have similarity in the C-terminal region of TT8; however the ORF of contig12598 is only 220 amino acids long and is therefore likely to extend into the N-terminal direction – indicated by the red arrow.

bHLH TFs, such as MYC2, GL3, EGL3 and TT8 in *A. thaliana*, are known targets of JAZ proteins and MYC2 has been observed to bind to all of the 12 recognised JAZ repressors (99, 105). It is surprising that neither of the TcbHLHs were able to interact with the JAZ proteins; however the JID is not present in the TcbHLHs possibly explaining why interaction wasn't observed (230). The elongated version of bHLH08058 (bHLH08058-FE) does have some conservation of the JID domain but due to time restraints the Y2H assay testing with this TF has not been completed. Future Y2H assays testing bHLH08058-FE may yet identify interactions with JAZ proteins. One of the closest homologs to bHLH11748 ICE1 is able to interact with two JAZ proteins (JAZ1 and JAZ4) through the C terminus leading to repression of ICE1 function (280). This is different to the JID which is located at N- terminus, showing that there are a variety of different sequences can interact with JAZs and the location and specificity of the TcJAZs may be different to those in *A. thaliana.*

The Y2H assay may have produced false negatives due to limitations in the methodology. The fused yeast reporter proteins can cause steric hindrance which obstructs the interaction. Some post-translational modifications, such as tyrosine phosphorylation, are present in higher eukaryotes but are absent *S. cerevisiae* and these maybe required for interaction to occur (277, 281). If the interaction is not symmetric then observation of the interaction may depend on whether a protein is used as prey or bait. False negatives cause major problems with reproducibility, in two independent large scale Y2H screens of the yeast interactome the data had only 30% overlap (277). JAZ35964 and bHLH08058 were found to interact in initial investigations but the result could not be reproduced. However many studies have employed Y2H to elucidate JAZ binding therefore the lack of observed binding might be because the proteins do not interact (163).

JAZ proteins are known to interact with a range of proteins including co-repressors TPL and NINJA (106), other JAZ proteins (107) and MYB TFs (163, 282). For example MYB21 and MYB24 involved in male fertility interact with JAZ1, 8 and 11 and this is speculated to attenuate MYB functionality (282). Future research could be conducted investigating the probable interactions between these proteins and the identified JAZs. There are other combinations of TFs that could also be interesting to test such as TPL and AP200499 as TPL is known to interact with EAR motifs (100).

Chapter 5 – *In vitro* electromobility shift assay experimentation to determine the binding site specificity of the 19 candidate transcription factors.

This work was completed in collaboration with Marisol Ochoa Villarreal.

5.1 Introduction

The electromobility shift assay (EMSA) also known as the gel retardation or band shift assay is a rapid and sensitive means of detecting DNA-protein interactions (283). The assay is based on the simple rationale that complexes of differing size migrate at different rates through a non-denaturing polyacylamide gel matrix. A DNA-protein complex has a greater molecular weight compared to free DNA meaning it migrates slower through the gel, leading to band retardation (284). A wide range of DNA and protein sizes can be accommodated. DNA can be labelled with radioactivity or a biological moiety, such as biotin, and proteins can either be purified or used in a crude extract (285, 286). There are limitations to the assay for example; the sample is not at chemical equilibrium therefore complexes which rapidly dissociate are hard to observe: and the assay does not provide information on the exact location of the binding or the identity of the proteins present in the complex (285). The sensitivity and well-established use of EMSA in identifying TF-*cis*-element interactions was the reason it was chosen to determine the binding specificity site of candidate TFs (287).

5.2 Experimental Design

The aim of these experiments was to screen the 19 candidate TFs against the 10 paclitaxel biosynthetic promoters *in vitro* to identify any interactions. This data could then be compared to the TEA results (Figure 33) to see if they complement and provide information on the exact location of the BS within the promoter.

5.2.1 Selection of *cis*-elements for DNA probes

Short DNA probes of 20 bp were chosen to try and establish the BS of the candidate TFs. The BS analysis conducted in section 2.6 identified the putative binding motifs for each TF family and synthetic oligonucleotides were designed to contain the proposed BS and the surrounding nucleotide sequence. The first three or four BSs in a promoter (those closest to the start codon) were chosen for initial testing and if no interaction was observed BSs further away from the start codon were subsequently investigated. Small synthetic DNA probes produce high electrophoretic resolution and are cheap and easy to synthesise.

There are issues, however, with using short oligonucleotides. The BS is always close to a molecular end, which can result is anomalous binding due to structural or electrostatic end-effects (285). The BSs were identified based on previously published motifs and there is little species

specific information, therefore if the TF binds to a novel BS it will not be identified. However work completed by a previous group member found work with long fragments to be problematic (Amir (223)). The DNA probes were end-labelled with radioactive ^{32}P because of its high sensitivity ($\leq 10^{-18}$ mol can be observed) and the addition of a phosphate does not produce any artificial features that may interfere with binding (285).

5.2.2 Protein Production

Initial EMSA testing was completed in collaboration with Amir using an *in vitro* transcription/translation system using wheat germ extract provided by Yasuomi Tada (223). Issues with the supply of the wheat germ extract led to proteins being synthesised using a commercial kit (Promega); however protein yields were significantly reduced. All but one TF, AP207245, was successfully produced with this system but the concentration of the proteins was very low. This led to weak EMSAs and a high background due to the large number of additional proteins present in the wheat germ extract.

An *in vivo E. coli* production system was employed to increase protein production levels. The TFs were cloned into pDONR221 and subsequently into the Gateway™ GST-tag destination vectors pDEST15 and pDEST24 (see Table 25 for primers and Appendix 1 for vector maps). The destination vectors add a GST tag N- and C-terminally respectively and the choice of vector was decided based on the location of the DNA binding domain. The GST tag was placed at the opposite end of the protein to the DNA binding domain to reduce the possibility of the tag interfering with binding, therefore AP2 TFs were cloned into pDEST15 and bHLH, MYB, NAC and WRKY TFs were cloned into pDEST24. Optimization of *E. coli* protein production conditions was performed by Ochoa Villarreal testing a variety of IPTG induction variables. The TFs were successfully produced in *E. coli* but multiple bands were observed in the western blot (Figure 58a). The additional bands could be reduced *via* purification using glutathione agarose beads 4B (Figure 58b). EMSAs conducted using purified protein produced cleaner blots with reduced background.

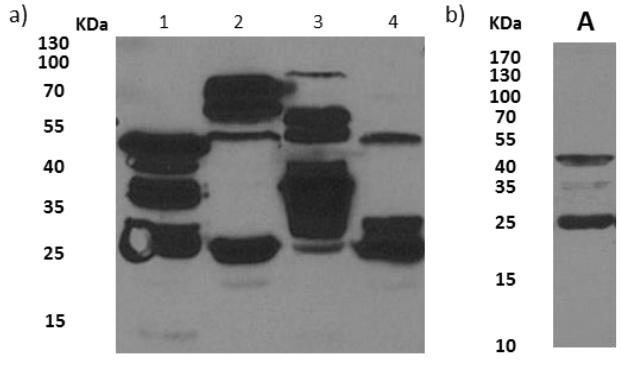


Figure 58 – Protein production and purification of *Taxus* transcription factors (TFs) using an *in vivo E. coli* expression system. a) Western blot of four *Taxus* TFs produced in *E. coli* without subsequent purification, Lanes 1-4 AP200499, MYB10385, MYB10855, MYB15401, antibody anti-GST-HRP (5000:1), blot exposure time 1 min. The size of the TFs with a GST tag are AP200499 – 45kDa, MYB10385 – 60kDA, MYB10855 – 75kDa and MYB15401 – 57kDa. The TFs were successfully synthesised but multiple bands were observed in the Western blot. The band at 25kDa is GST and the other bands are likely to be partially degraded TFs. b) Western blot of AP200499 produced in *E. coli* subsequently

purified with glutathione agarose beads 4B (GE healthcare), antibody anti-GST-HRP (5000:1), blot exposure time 1 min. The purification successfully removed the addition bands. Only two bands are observed the upper of which is AP200499 and the lower is free GST.

5.2.3 Modifying the experimental design

During experimentation it was concluded that the initial plan of completing a full screen of the 19 TFs against the 10 paclitaxel biosynthetic promoters was unfeasible therefore, three candidates were chosen for further study MYB10855, MYB10385 and AP200499. These three TFs have the highest logFC at 0.5 h after MeJA elicitation (Figure 59a) and the TEA results showed that MYB10855 was a positive regulator of 8 of 10 promoters tested, while AP200499 was found to be a negative regulator of five promoters (Figure 59b).

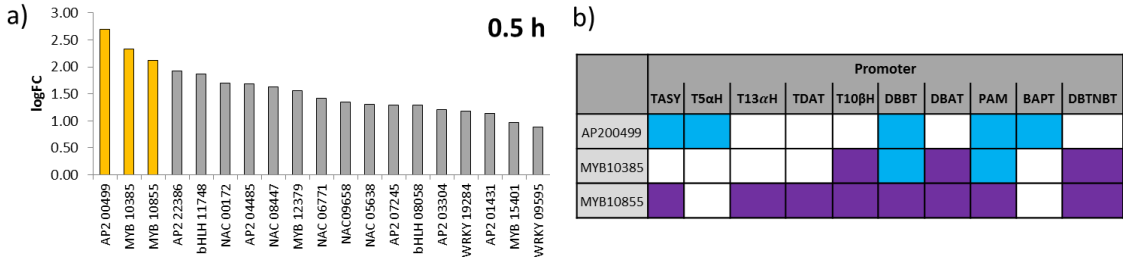


Figure 59 – Justification for choosing AP200499, MYB10385 and MYB10855 as candidates for EMSA analysis. a) The log fold change (logFC) of the 19 candidate transcription factors (TFs) at 0.5 hour (h) after methyl jasmonate elicitation. AP200499, MYB10385 and MYB10855 have the highest logFC values at this time point and are highlighted in yellow. b) The summary of the *A. thaliana* protoplast transient expression assay (TEA) results for AP200499, MYB10385 and MYB10855 against the 10 paclitaxel promoters (see section 3.3.2 Analysis of *in vivo* TEA for further details and Table 2 for promoter abbreviations). Boxes highlighted in purple indicated significant up-regulation, blue indicates significant down-regulation and white indicates no significant interaction. Significance was determined using a Student *t*-test, $n \geq 4$, p-value ≤ 0.05. AP200499 negatively regulates 5 promoters, MYB10855 activates 8 promoters and MYB10385 is a dual regulator of 5 promoters.

5.3 EMSAs investigating the binding specificity of MYB10385

EMSAs investigating the binding specificity of MYB10385 identified binding in two of the paclitaxel biosynthetic promoters, *TASY* and *DBTNBT*. The binding was very weak, however, as very low concentrations of the unlabelled competitor probe were able to abolish the interactions (Figure 60a). This is in contrast to testing completed by Ochoa Villarreal using MYB10855 where interactions were only lost after a 500-1000 excess of the competitors probe was added (Appendix 2a). The identified consensus binding motif for MYB10385 contains an ANC sequences which is found within the known ACC(A/T)A(A/C) MYB binding motif (Figure 60c) and is similar to that identified for MYB10855 (Figure 63b).

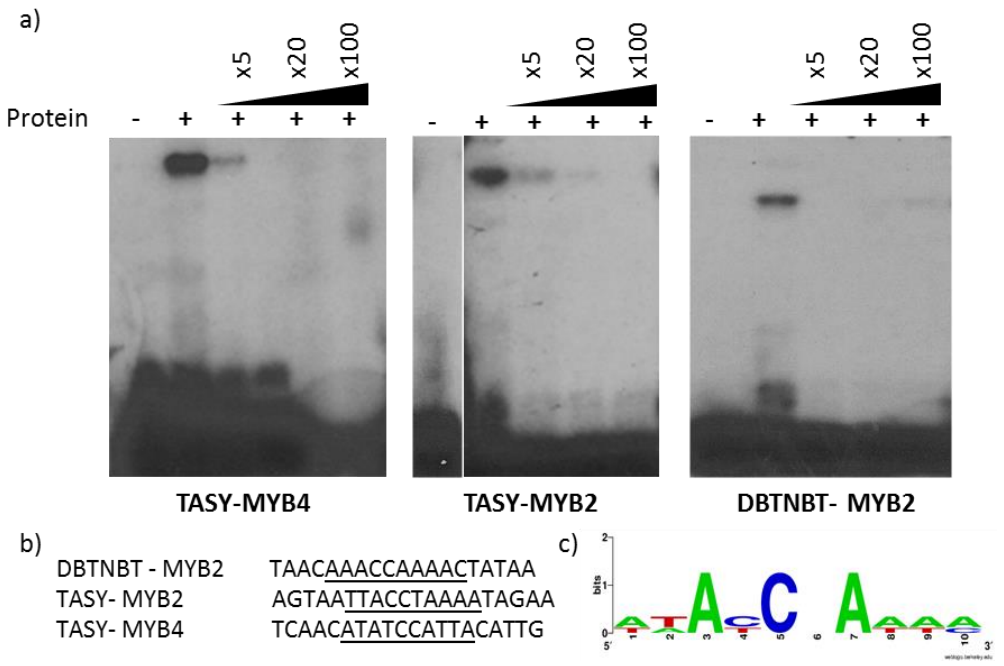


Figure 60 – Summary of the interactions identified between MYB10385 and the paclitaxel biosynthetic promoters using electromobility shift assays (EMSAs). a) EMSAs were completed with GST tagged protein purified with glutathione agarose beads 4B (GE healthcare) and ^{32}P labelled 20 bp DNA probes, blots were exposed at -80°C for 24-72 hours. Whether the protein (MYB103855) was loaded in each well and the addition of increasing amounts of unlabelled competitor probe is stated above the blot. The name of the probe under investigation is stated below the blot. Interactions were identified in *TASY* and *DBTNBT* however, the interactions identified are weak because binding was lost after the addition of very low quantities (5x excess) of unlabelled probe. b) The 5' DNA sequences of the probes which MYB10385 was identified as binding with in the EMSAs and c) the binding consensus sequence created from the central nucleotides (underlined) within these probes. Sequence logo was generated with WebLogo (288). The consensus sequence contains the sequence AnC.

5.4 EMSAs investigating the binding specificity of AP200499

EMSAs investigating the binding of AP200499 with the paclitaxel promoters identified strong binding in *TASY*. The interaction was not abolished after 100 fold excess of the competitor probe was added and additional analysis by Ochoa Villarreal confirmed that the interaction was only lost after addition of 1000x competitor probe. The probe (TASY-AP2-2) contains the well-characterised GCCGCC motif that is the BS for numerous AP2 TFs such as NtERF1-4 (128) and Pti4, Pti5 and Pti6 (130).

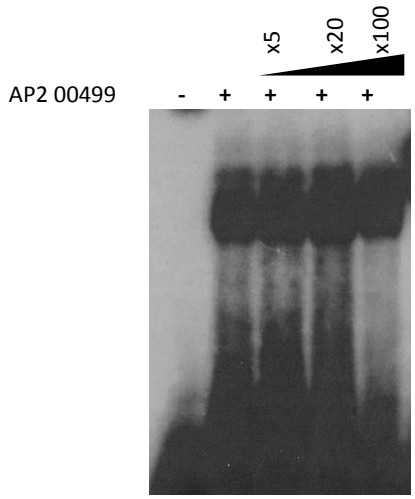


Figure 61 – The binding of AP200499 with the TASY AP2-2 probe in an electromobility shift assay (EMSA). a) EMSAs were completed with GST tagged protein purified with glutathione agarose beads 4B (GE healthcare) and ^{32}P labelled 20 bp DNA probes, blots were exposed at -80°C for 24-72 hours. Whether AP200499 was loaded in each well and the addition of increasing amounts of unlabelled competitor probe is stated above the blot and the name of the probe under investigation is stated below the blot. AP200499 interacts with strongly with the TASY AP2-2 probe as the binding was not lost by the addition of increasing amounts of unlabelled competitor.

5.5 Discussion

5.5.1 Analysis of identified MYB10385 and MYB10855 binding sites

The TEA identified 5 promoters which had significantly altered activity with MYB10385, *T10βH*, *DBBT*, *PAM*, *DBAT* and *DBTNBT*. Interactions were only observed in one of these promoters, *DBTNBT*, but MYB10385 was found to bind to two DNA sequences in *TASY* (Figure 60a). An explanation for this difference could be that the protein is capable of binding to multiple sites but the surrounding sequence affects whether the TF can regulate gene expression or more simply that these interactions are too weak to make any firm conclusions.

Testing completed by Ochoa Villarreal focused on MYB10855 because it was shown in the TEA to interact strongly with 8 out the 10 paclitaxel biosynthetic promoters. This work identified probes in all 8 paclitaxel promoters (*TASY, T13αH, TDAT, T10βH, DBBT, DBAT, PAM and DBTNBT)* that MYB10855 bound with (Figure 63a) and the specificity of these interactions was confirmed using competitor assays, except for *TDAT* due to the weakness of the interaction (Appendix 2a).

One of the possible reasons that interactions could be observed with MYB10855 but not MYB10385 is that although the activity observed in the TEA was statistically significant the MYB10855 interactions were much stronger (Figure 62). MYB10385 could also require post-translational modifications to act such as phosphorylation or redox control. In maize P1 contains two cysteines, four residues apart, in the R2-MYB domain that form an intramolecular bond under oxidising conditions which blocks binding (125). These residues are conserved in MYB10385 but not in MYB10855. The MYB protein WER cannot bind to the *CAPRICE* promoter without the addition of 10 mM dithiothreitol (DTT), a reducing agent thought to abolish a disulphide bond within

the protein (135). The binding buffer used in our EMSAs contained only 1mM DTT, therefore higher concentrations may alter the conformational state of MYB10385.

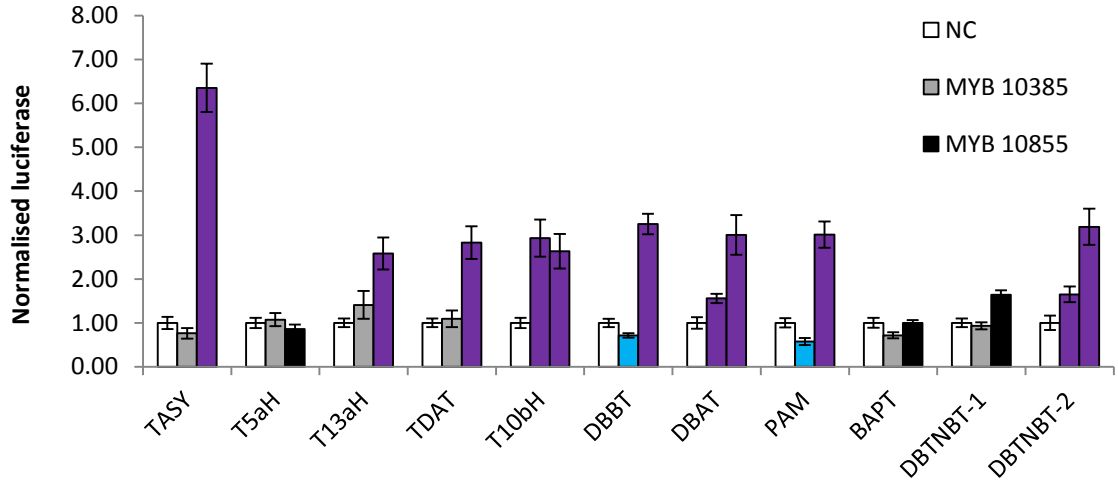


Figure 62 – Comparison of the *Arabidopsis thaliana* protoplast transient assay (TEA) results for MYB10385 and MYB10855. The graph shows normalised luciferase (the luciferase values obtained in the TEA normalised to the negative control (NC) in each promoter) against the paclitaxel biosynthetic promoters (see Table 2 for abbreviations). Bars highlighted in purple indicate significant activation of the promoter, while blue indicates significant repression of the promoter. Significant interactions were determined compared to the NC for each promoter using a Student *t*-test, $n \geq 4$, p-value ≤ 0.05. Error bars represent standard error. The normalised luciferase values for MYB10855 are mostly higher than those of MYB10385.

Initial inspection of the DNA sequences that MYB10855 was found to bind with suggests that it is capable of binding with 3 different MYB BS, ACC(A/T)A(A/C), CNGTTA and GATAA. It is not uncommon for MYB TFs to bind with different sites as their binding is inherently flexible, for example the petunia MYB Ph3 can bind with two MYB BS (T/C)AAC(G/T)G(A/C/T)(A/C/T) and AGTTAAGTTA (123). Many R2R3 MYB TFs bind AC rich elements with depleted guanine content and the DNA sequences identified by EMSA were AC rich, most having a low guanine content (Figure 63a)

The AC region of MYB BSs are relatively short (5-6bp) therefore, further analysis of the AC elements produced a consensus sequence where the first four base were NA(A/C)C (Figure 63b), which is similar to that produced for MYB10385 - ANC (Figure 60c). The second adenosine and fourth cysteine were identified as important residues for DNA binding of the oncoprotein *c*-MYB. The NMR structure of *c*-MYB showed that residues Asn183 and Lys182 interact with the adenosine and cysteine respectively (126, 289) and these two residues are conserved in MYB10855 and MYB10385. Research on AtMYB61 has also shown the importance of the ACC motif with EMSA studies proving that mutations outside of the first three residues have little effect on the binding activity (290). This concurs with our consensus sequences where there is little similarity after the 4th base. The elasticity in the second half of the motif is also seen in Pine where PtMYB1 and PtMYB4 can bind to three AC elements found in the phenylpropanoid biosynthetic gene phenylalanine ammonia-lyase (136, 137).

Promoter	5' Synthetic Oligo	%AC	%G	AC Element
TASY	TAAGCAATAACAAAGCACGA	75	15	TAACAAAG
DBBT	TTTCTGGACAGTTACATTCT	40	15	GACAGTTA
DBTNBT	GCCTTTTACCAACCATATTT	55	5	TACCAACC
PAM	GACGCAACGATAAATTGCAG	60	25	CAACGATA
DBAT	TGTCCTCACCAAAATTAGTT	55	10	CACCAAAA
TDAT	CTTATTAACTGATAATATGG	45	15	TAACTGATA
T10βH	CCAACAAACCAAAAGGGCAT	80	15	AACCAAAA
T13αH	ATTATTCACCAACAATTGTA	60	5	CACCAACA

b)

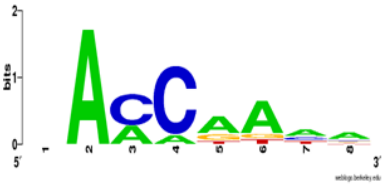


Figure 63 – Summary and analysis of MYB10855 binding sites in eight paclitaxel biosynthetic promoters. a) A table containing the 5' nucleotide sequence of the DNA probes which MYB10855 was identified as binding with in electromobility shift assays. The binding motif in each probe is highlighted in colour and underlined, red denotes ACC(A/T)A(A/C) motifs, purple CNGTTTA motifs and green GATAA motifs. The % adenine and cytosine, % guanine content and AC rich element for each probe are shown. See Table 2 for promoter abbrevaitions. b) The sequence logo showing the consensus sequence of the AC element. Sequence logos generated with WebLogo (288). MYB10855 binds to AC rich elements in 8 promoters and the consensus sequence generated from these elements contains the sequence A(A/C)C.

In the TEA MYB10855 was found not to alter the activity of *BAPT* or *T5αH*. No ACC(A/T)A(A/C) BSs were identified in *T5αH* but 3 were identified in *BAPT* (Figure 64), therefore the presence of AC rich elements in a promoter does not directly predict the ability of a TF to interact. One of the ACC motifs in *BAPT* (caccaaca) is very similar to the DNA sequences found to interact with MYB10855 *in vitro.* Testing completed by Ochoa Villarreal found that MYB10855 did not interact with this motif, suggesting the surrounding sequence is important in controlling TF regulatory activity.

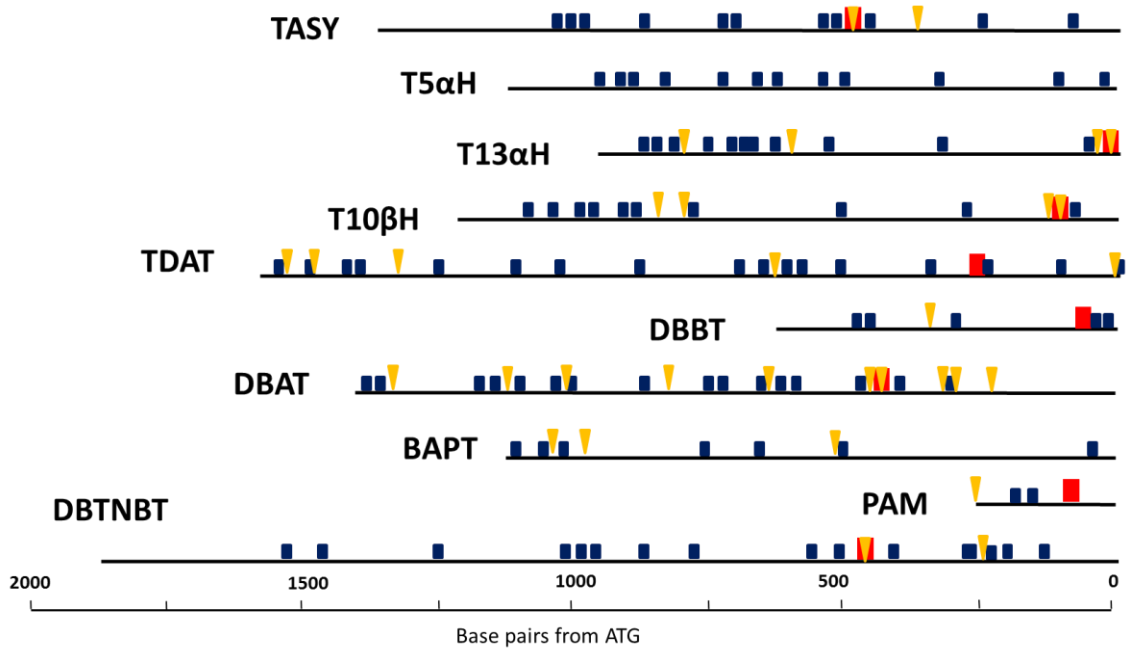


Figure 64 – The location of the MYB binding sites (BSs) in the 10 paclitaxel biosynthetic promoters. The image shows the representative length of each promoter and location of MYB BSs relative to the start codon (ATG). Yellow triangles indicates ACC(A/T)A(A/C) motif and blue boxes ACC rich regions containing consensus sequence A(A/C)C(A/C)(A/C)(A/C). The red boxes indicates the location of the MYB10855 BS identified by EMSAs in 8 promoters (*TASY, T13αH, TDAT, T10ϐH, DBBT, DBAT, PAM* and *DBTNBT*). See Table 2 for paclitaxel promoter abbreviations. The paclitaxel promoters are rich in cognate MYB BSs located throughout the promoter, however MYB10855 was found to only interact with certain ACC rich elements.

MYB BSs are generally located 500 bp from the start codon (123). The location of the identified BSs was sometimes much closer to the start codon, for example in *T13αH* MYB10855 was found to interact with MYB1 only 8 bp into the promoter (Figure 64). The original identification of the BSs used a number of consensus sequences including ACC(A/T)A(A/C), however further analysis of the promoter using the consensus sequence produced from the current EMSA results (Figure 63b), NA(A/C)C, identified another 3 possible AC rich binding elements located 340, 335 and 736 bp away from the ATG. This demonstrates some issues with the experimental design. Firstly BSs were identified using previous published motifs, which could be slightly altered in *Taxus spp.* and secondly if a BS was identified within the first three or four predicted BSs other possible motifs further along the promoter were not tested.

5.5.2 Analysis of identified AP200499 binding sites

AP200499 was found to interact strongly with the AP2-2 probe in *TASY* which contains the GCC binding motif. AP200499 was found to repress five paclitaxel promoters in the TEA, however further experimentation by Ochoa Villarreal only identified significant binding in *DBBT* and weak binding in *T5αH* (see Appendix 2b). The DNA sequences AP200499 interacted with contain the consensus sequence GCCG (Figure 65), however this motif was also present in the AP2-2 probe in *BAPT* and AP200499 did not interact with it. GCCGCC is a well-known motif for AP2 interactions (115) and the consensus identified retains the first four bases.

a)

Promoter	5' Synthetic oligo
TASY	TGCGCCTGCCGCCTGACTTA
DBBT	GGGAAGGCCGGCCCGGCATG
T5αH	TAGTGTGATGTGCCGAAAGG

b)

Figure 65 – Summary AP200499 binding motifs and consensus sequence in *TASY*, *DBBT* and *T5αH*. a) A table containing the 5' nucleotide sequence of the DNA probes which AP200499 was identified as binding with in the EMSAs. The binding motif in each probe is highlighted and underlined in green; b) the consensus sequence of the three binding motifs is the of GCCG which is a truncated version of the GCCGCC motif. Sequence logo generated with WebLogo (288).

5.5.3 Analysis of experimental design

The current EMSA analysis of the paclitaxel biosynthetic promoters has established with some confidence that the consensus sequence for MYB10855 is A(A/C)C, but as yet has not elucidated the binding motifs of MYB10385 and AP200499 in all the promoters that they were shown to interact with in the TEA. One of the limitations of the EMSA is that it does not provide direct information on the exact location of the binding site (287). Using small oligonucleotides reduces potential non-specific binding and increases electrophoretic resolution (285), however correctly predicting the BSs to analyse may have led to the difficulties in establishing the binding motif. There is little species specific information about TF binding in *Taxus*, therefore the identification of the putative BSs (5.2.1 Selection of *cis*-elements for DNA probes) may have missed novel or altered binding motifs. The well characterised motif GCC motif (115), for example, can alter between promoters and species (Table 16) (109). ORA59 binds to the established consensus sequence of GCCGCC (131) in *PDF2.1*, but the TFs ORC1 and ORCA3 activate

expression *via* slightly modified versions of the motif. ORC1 is capable of activating a number of nicotine biosynthetic promoters that contain slightly different GCC motifs that also require the presence of a G box to observe transactivation (185). Three consecutive sequences in *STR* were identified as crucial in the JERE motif for ORCA3 activity however only M3 has a recognisable GCC motif (Table 16) (176).

TF	Species	Promoter	Motif Information		Ref
ORC1	*Nicotiana tabacum*	Nicotine Biosynthesis	GCC motif	G Box	(185)
		PMT1	TGCGCCC	CACGTT	
		PMT2	TCCGCCC	CACGTT	
		QPRT2	CCAGCCA	AACGTG	
		ADC1	TCGGCCT	CACGTT	
		ADC2	TCGGCCT	CACGTT	
ORA59	*Arabidopsis thaliana*	Defence	GCC motif		(131)
		PDF1.2	GCCGCC x 2		
ORCA3	*Catharanthus roseus*	TIA Biosynthesis	JERE Motif		(133)
		STR	M2 M3 M4		
			CTCTTA-GACCGC-CTTCTT		

Table 16 – A table summarising GCC like binding motifs in three different promoters and species. A GCC motif and G-box are required for ORC1 activation of nicotine biosynthetic genes. ORA59 modulates PDF1.2 expression through two functionally equivalent GCCGCC motifs and three fragments of the JA responsive JERE motif (M2/M3/M4) are required for *STR* activation by ORCA3. Abbreviations: PMT, putrescine N-methyltransferase, QPRT, quinolinate phosphoribosyltransferase ADC, arginine decarboxylase, PLANT DEFENSIN1.2, PDF1.2, TIA, terpenoid indole alkaloid, STR, strictosidine synthase.

A different approach might therefore be required to establish the exact location of TF binding in paclitaxel biosynthetic promoters. DNA foot printing EMSAs or promoter deletion and mutation analysis in transactivation assays could be employed to establish the binds site as used by Fits *et al.* for ORCA3 and Boer *et al.* for ORC1 (133, 185). This approach would also help to clarify the MYB BSs as their well-known binding flexibility makes selecting specific fragments difficult. Another methodology that could be employed is cyclic amplification and selection of targets (CASTing) (123). This involves mixing the desired protein with a pool of random oligonucleotides that undergo multiple rounds of PCR to enrich the pool with nucleotides that contain the binding motif (291). This protocol has been successfully used to identify the binding affinity of AtMYB61 (290), ANAC019 and ANAC092 (142).

The use of purified recombinant proteins produced by *E. coli* was far superior to work completed using wheat germ extract or crude *E. coli* extracts, producing significantly cleaner blots. Holden *et al.* however stated that if recombinant proteins were employed it is important to consider post-translational modifications that may necessary for activity (286). It has already been established that a MAPK signalling cascade is up-regulated after MeJA elicitation and there are many examples in the literature of phosphorylation altering TF activity; for example NtMYB2 is positively regulated by by a cyclin dependent kinase (292) and ORC1 activity is up-regulated by MAPKK JAM1(100).

Prediction of the possible phosphorylation sites of in AP200499, MYB10385 and MYB10855 was conducted using NetPhos 2.0 (268) and GPS 3.0 (293). Nine serine and one threonine possible phosphorylation sites were predicted in AP200499 which were located in coil

ordered regions of the protein. MYB10385 and MYB10855 were predicted to have a far greater number of phosphorylation sites at serine, threonine and tyrosine residues – MYB10385 Ser: 24, Thr: 4, Tyr: 5 and MYB10855 Ser: 21, Thr: 5 Tyr: 6 (Table 17). The distribution of these residues is predicted to occur mostly in the coil region distributed across both ordered and disordered regions. Phosphorylation is not always required for *in vitro* binding as PtMYB4 can bind to ACC elements in EMSAs, but the lack of post-translational modification does alter transactivation in yeast (294).

The current EMSA studies have shed light on the binding affinity of MYB10855 but further testing needs to be completed to confirm the exact binding specificity of the binding, while significantly more research is required to identify the binding conditions and motifs for MYB10385 and AP200499.

Chapter 6 – Attempts to establish a transient particle bombardment assay in *Taxus* CMCs

6.1 Introduction

The interactions observed in the *in vivo* TEA and in the *in vitro* EMSA ideally need to be confirmed in *Taxus spp*. There are two main ways to transform PCC; 1) *Agrobacterium* transformation and 2) particle bombardment. Our Korean collaborators (Unhwa) are currently working to produce a protocol to stably transform *Taxus spp.* CMCs using *Agrobacterium*, with the aim of creating overexpression lines with our top candidate TFs, such as MYB10855. Concurrently, we have been working to try and establish a transient particle bombardment assay in *Taxus spp*. CMCs to confirm the interactions observed *in vitro* and *in vivo*.

6.1.1 Particle bombardment

Particle bombardment is a physical method of gene delivery that was first developed in the late 1980s by Sandford *et al.* to overcome the barriers present in transforming plant species (295, 296). One of the most widely used systems for microprojectile bombardment is the Biolistic® PDS-1000/He produced by BioRad laboratories.

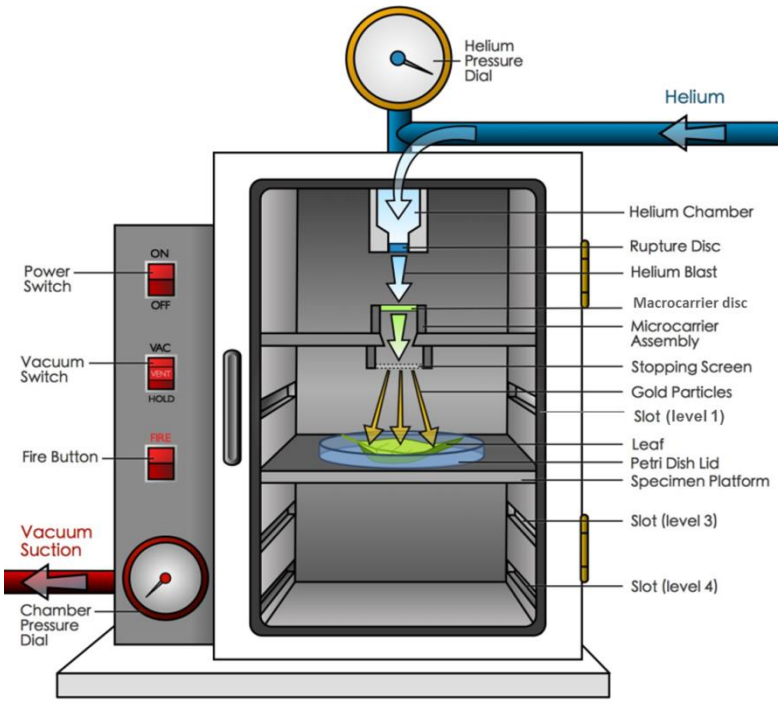


Figure 66 – A schematic of the Biolistic® PDS-1000/He system. The bombardment occurs under a vacuum and the speed at which the vacuum is created and released as well as the final vacuum pressure in Hg/mm is controlled by switches and dials on the left side of the instrument. Helium enters the system and is pressurised in the helium chamber. Rupture disks of differing pressure can be used to alter the helium pressure. When the helium reaches the required pressure – the pressure can be monitored using the helium pressure dial – the rupture disk brakes releasing a blast of helium at the microcarrier assembly. The microcarrier assembly consist of DNA coated microcarriers loaded onto a macrocarrier disk. The blast of helium propels the macrocarrier through the microcarrier assembly until it is hits the stopping screen. The microcarriers continue travelling past the stopping screen and penetrate the target cells

below (the example in the image is a leaf). The distance between the stopping screen and the target cells can be altered to four distances 3cm, 6cm, 9cm and 12cm which are respectively slot 1, the position of the specimen platform on the image, slot 3 and slot 4. Image adapted from http://i.imgur.com/DoXu3tn.jpg.

This system employs compressed helium to accelerate DNA coated metal particles (microcarriers) towards target cells. Helium is pressurised in a chamber until it reaches the required pressure at which point it is released by the rupture disk. This produces a blast of helium that propels the macrocarrier sheet, loaded with microcarriers, through the microcarrier assembly. The macrocarrier then hits the stopping screen but the microcarriers continue to travel and penetrate the target cells below (Figure 66). If the mircocarrier successfully lodges inside the cell the DNA elutes off the metal particle and is expressed. The system has a number of well-defined parameters that require optimization including; a) composition and size of microcarrier, b) vacuum pressure, c) helium pressure and d) distance between stopping screen and target cells.

There are few examples in the literature of *Taxus spp.* transformation however a protocol for particle bombardment has been published by the Roberts group (297) which was subsequently employed to show TcJAMYC activity in *T. cuspidata* (184). This protocol was used as a basis to establish a *Taxus spp.* CMC particle bombardment protocol.

6.2 Establishing parameters of particle bombardment assay

Since the original isolation of *T. cuspidata* CMCs (64) Unhwa has successfully isolated CMCs from two other *Taxus spp., T. baccata* and *T. media*. *T. baccata* is a high paclitaxel producing species with better growth properties compared to *T. cuspidata* including, higher growth rates and smaller aggregates. Attempts to produce a *Taxus spp.* based particle bombardment assay were therefore initiated using *T. baccata* CMCs.

Initial experiments were performed using luciferase as a reporter because plants lack any native luciferase activity and it would mitigate the need for further cloning. Green fluorescent protein (GFP) was not chosen initially as *Taxus* cells have endogenous fluorescence in their cell wall (Figure 69). The luciferase assay is a destructive process but this is not an issue as the aim is to produce a transient expression system. A protocol for transferring CMCs from liquid to solid media was established with the cell volume and solid media compositions among the parameters tested to ensure the cells were not stressed and could be easily collected for luminescence quantification.

Gold with a particle size of 1.6 µm was chosen as the metal for the microcarrier rather than tungsten as the Roberts group found the transformation efficiency was significantly lower with tungsten. Normally a particle size of 1.0 µm is recommended for plant cell culture work however Vongpaseuth *et al.* and previous work conducted in our lab produced higher transformation efficiencies with microcarriers of 1.6 µm (221, 297). However the suggested reason for this observation with the *T. cuspidata* cell line P991 was that *Taxus* cells are larger than other plant cell cultures, but CMCs are smaller than traditional DDCs (64) therefore testing with smaller microcarriers might be required.

6.2.1 Particle bombardment of onion cells

Onion cells, which are easier to transform, were used to confirm that the microcarriers were being successfully coated in DNA and the correct operation of the machine. Cells were bombarded with a construct containing GFP under the p35S constitutive promoter. Figure 67 confirms that these steps were performed successfully as GFP expression was observed in the onion cells.

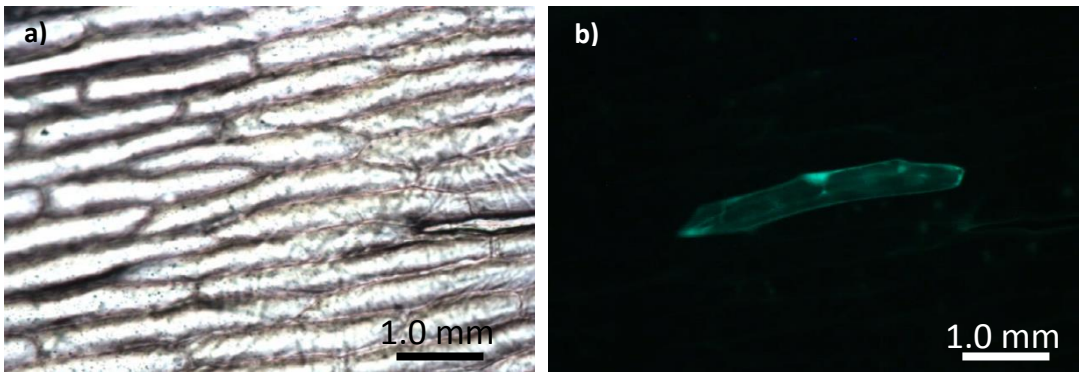


Figure 67 – An example of successful particle bombardment of an onion cell. Visualisation of the cell was performed at 10x magnification under a) white light and b) blue light to produce a: a) bright field and b) fluorescent image. The scale bars represent 1.0 mm. The transformation was successful because GFP expression can be observed throughout the cell, in particular in the nucleus located at the central top position in the cell.

6.2.2 Particle bombardment of T. baccata CMCs using luciferase as a reporter

Experimentation was completed using the Rluc construct (*Renilla* luciferase under the constitutive p35S promoter) testing three different stopping distances and two rupture disk pressures. Multiple bombardments were also investigated as this has been shown to increase transformation efficiency (Figure 68a). None of the tested conditions led to a significant transformation efficiency. Shorter distances and higher pressures have been shown to increase transformation rates (297), however testing at 1350 psi did not produce higher results and led to the cells exploding out of the dish. Further testing of vacuum pressures and DNA concentrations was also completed but did not led to an increase in observed luminescence (Figure 68b). Microcarriers at 1.0 µm were also tested but showed little difference.

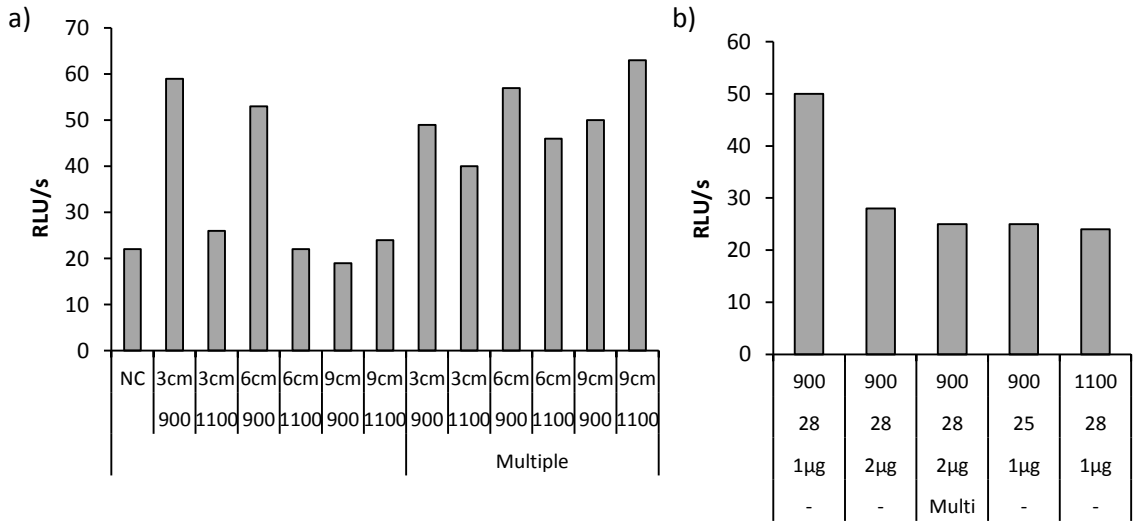


Figure 68 – Particle bombardment of *Taxus baccata* CMCs with luciferase as the reporter gene. The cells are bombarded with a construct containing *Renilla* luciferase under the constitutive p35S promoter. a) Testing at two rupture disk pressures 900 and 1100 psi and three stopping distances 3, 6 and 9 cm and multiple bombardments (multiple) which involved bombarding the same sample three times. NC= negative control which are cells bombarded with no DNA; b) Further experimentation of particle bombardment conditions. Parameters altered in the order they appear on the graph are rupture disk pressure (psi), vacuum pressure (mm/Hg), DNA concentration (µg) and multiple bombardment (Multi) - three multiple bombardments on the same sample. The luciferase is quantified in relative luminescent units (RLU). The results show that the transformation of *T. baccata* CMCs was unsuccessful as no combination of conditions lead to high levels of *Renilla* luminescence.

6.2.3 Particle bombardment using GFP as a reporter

Particle bombardment was subsequently performed with GFP as a reporter to try and visualise the number of transformed cells. There is a lot of endogenous GFP expression in CMCs, especially in the plant cell wall, which made evaluating the transformation frequency difficult (Figure 69a). *T. baccata* CMCs were transformed with a construct containing GFP under a constitutive p35S promoter. If cells were successfully transformed GFP expression was observed throughout the cell (Figure 69b), however the transformation efficiency was very low.

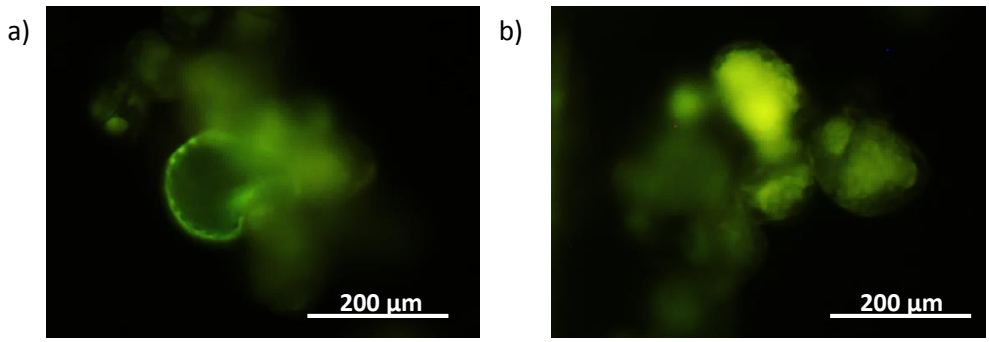


Figure 69 – Visualisation of GFP expression in *Taxus baccata* CMCs. Images showing a) the autofluorescence of the *T. baccata* cell wall and b) *T. baccata* CMCs successfully transformed with p35S:GFP construct expression GFP throughout the cell. Visualisation of the cells was performed at 40x magnification under blue light to produce fluorescent images. The scale bars represent 200 µm.

An issue with particle bombardment is that the process stresses the cells which leads to the production of polyphenols that turn the cells red (Figure 70a). Stress can increase the background GFP production making it difficult to distinguish cells that have been transformed. The cell in Figure 70b looks to be transformed as it is producing GFP but from the bright field image it can be seen that one of the cells is dying as the plasma membrane has come away from the cell wall (Figure 70c).

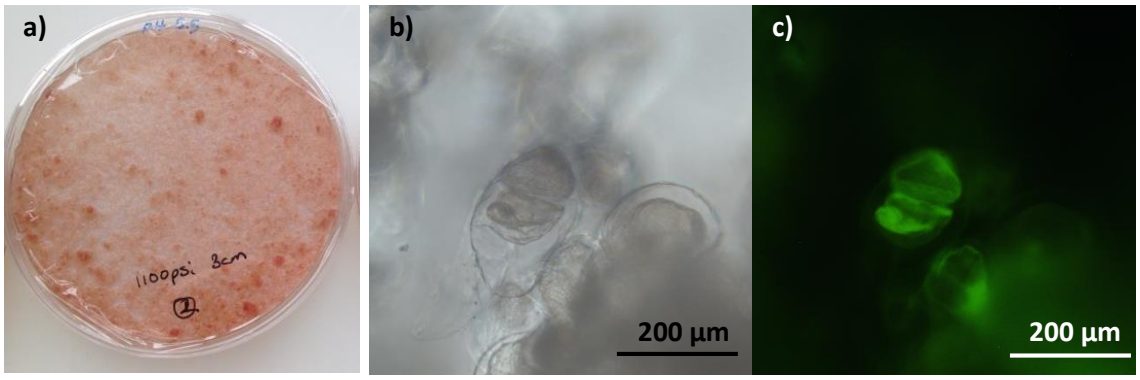


Figure 70 – Visualisation of the stress caused by particle bombardment on *Taxus baccata* CMCs. a) *T. baccata* CMCs on solid media 2 days after bombardment. The red colour of the cells is due to polyphenols produced as part of the stress response. The visualisation of *T. baccata* CMCs transformed with p35S:GFP construct at 40x magnification under b) white light and c) blue light to produce a: b) bright field and c) fluorescent image. The scale bars represent 200 µm. c) Two cells are identified as producing GFP under blue light but c) the bright field image shows that the lower of the two cells is dying as the cell membrane has come away from the cell wall.

6.3 Discussion

Work to try and establish a particle bombardment assay for *T. baccata* CMCs has not been successful to date, but if possible it would be helpful to confirm the interactions observed in *A. thaliana* protoplast TEA. The onion bombardment experiments suggested that the DNA was being coated correctly and the GFP observations implied that some CMCs were being transformed but the efficiency was negligible. Research could be conducted to increase the efficiency of coating the microcarrier with DNA using the onion bombardment as a reference. In the 2015 Walker paper where particle bombardment was used to observe the activity of TcJAMYCs the cells were spread into a circle with a diameter of 3 cm and pressed into the agar to immobilise them; which is different to the methodology used in the original article by Vongpaseuth *et al.* An issue experienced during bombardments was that using higher pressures and shorter stopping distances, which had been previously suggested to improve the transformation efficiency, led to cells exploding out of the dish.

Different reporters could be explored to visualise the bombardment efficiency. An example is β-glucuronidase (GUS) which has been used to identify *T. cuspidata* cells transformed by *Agrobacterium spp.* (298) but *Taxus* cells do have native GUS activity that needs to be reduced by increased the pH of the histochemical buffer before visualisation. Another alternative reporter gene is the red fluorescent protein DsRed from *Discosoma spp.* which has already been proven to be expressed in *T. cuspidata* cell lines P991 and CO93D (297). Luciferase is still an option as it has been successfully used to quantify gene expression in *Taxus* and provides quantitative data (184). More work, therefore, needs to be conducted to try and establish the best reporter gene to use in a transient expression protocol in *Taxus spp.* CMCs.

In the original article on *Taxus* particle bombardment the cells were transferred onto solid media 12- 24 h before bombardment (297), therefore this was the methodology used in our experimentation. However in a later paper it was stated that bombardment was undertaken 2 hours after transferring the cells onto solid medium (184). This change may reduce the background GFP expression produced by the stress of transferring the cells from a liquid to a solid environment.

Attempts to visualise the transformation earlier than 2 days might also produce better results as it would reduce the length of time that the cells are stressed. There are numerous areas that need to be investigated to establish a reproducible *Taxus* particle bombardment transient assay.

Chapter 7 – General Discussion

Paclitaxel is a key anticancer drug originally isolated from the bark of *Taxus brevifolia*. Due to its low concentration within *Taxus spp.* and high chemical complexity, PCC is seen as an attractive production route. CMC culture has superior growth properties compared to traditional DDC culture and is therefore a good platform from which to improve paclitaxel production. MeJA is a well-known elicitor of PNPs and is employed commercially to increase paclitaxel production in PCC. The identification of putative regulators involved in the up-regulation of paclitaxel production after MeJA elicitation could be exploited in bioengineering efforts to increase paclitaxel yields.

7.1 MeJA-mediated transcriptional reprogramming

The transcriptional cascade triggered in *T. cuspidata* CMCs after MeJA elicitation was observed using deep sequencing technologies Roche454 and Illumina sequencing. MeJA triggers an early (0.5 h) mostly inductive wave in gene expression in CMCs, followed by a second later wave (12 h) of genes that are up- and down-regulated (40 and 60% respectively) (Figure 14), showing that CMCs have tight regulation over their response to MeJA. The extensive metabolic reprogramming triggered by MeJA led to all the enzymes in the paclitaxel biosynthetic pathway being significantly up-regulated (Figure 17). The enzymes involved in isoprenoid precursor production on the other hand were not significantly up-regulated after MeJA elicitation (Figure 15), suggesting that the increase in paclitaxel yields is probably due to the shunting of intermediates out of competing pathways rather than an increased pool of isoprenoid precursors.

The transcript profiling studies of MeJA treated *T. cuspidata* CMCs allowed the exploration of possible agonist and antagonist effects of hormone crosstalk. Auxin and GA are important growth hormones that are likely to act antagonistically against MeJA, which as a stress signal prioritises defence over cell growth. Auxin ARFs, biosynthetic enzymes and transporters were down-regulated after MeJA elicitation (Figure 20), but GA GRAS TFs and biosynthetic genes were up-regulated. GA and 2,4-D (a synthetic auxin analogue) are components of *T. cuspidata* and *T. baccata* media respectively (Table 20). The identification of ARF and GRAS TFs that are MeJA regulated could be important in creating a balance between growth and paclitaxel production to produce high yielding lines that retain the superior growth properties of CMCs.

Analysis of the *T. cuspidata* CMC transcriptome identified all the components of the core JA signalling module found in many other plant species (100). The regulation of secondary metabolism is species specific but as the core module is conserved across angiosperms and gymnosperms research conducted in *A. thaliana* could be translated into *Taxus spp.* to try and improve paclitaxel yields. Perturbing the expression of JAZ proteins might lead to improved levels of paclitaxel production, as they act as a transcriptional repressors (107, 108). JAZ proteins can interact with a range of TFs, for example in *A. thaliana* they can bind to components of the MBW complex (MYB75, GL1, GL3 and EGL3) that regulate anthocyanin production (163). JAZ proteins were found to not interact with our MYC2 like TF bHLH08058 in Y2H experiments (Figure 53), but

this result was likely due to fact that the protein tested is a truncation version (Figure 24). Future experimentation with the recently isolated elongated form of the TF (bHLH08058-FE) may yield a different result. There is, however, a large amount of functional redundancy in the JAZ protein family with 12 JAZ proteins characterised in *A. thaliana* (105) and six JAZ proteins identified in our 1646 DEGs so far, which might complicate bioengineering efforts.

7.2 Regulators of the paclitaxel biosynthetic pathway

Research completed by previous members of the project (Yan (222), Amir (223)) identified 19 putative regulators of the paclitaxel biosynthetic pathway, which were highly up-regulated at an early (0.5 h) time point after MeJA elicitation. The TFs were from families that have been previously shown to be involved in the regulation of secondary metabolism and the aim was to identify an overall regulator of the paclitaxel biosynthetic pathway. Analysis of the available paclitaxel biosynthetic promoters showed that they were rich in the cognate binding sites of the candidate TFs. The possible regulation of the paclitaxel promoters by the 19 candidate TFs was screened using an *A. thaliana* protoplast TEA.

The results of this screen found that every promoter was regulated by at least three TFs and promoter activity could be activated or repressed (Figure 33). The presence of activators and repressors in the TFs induced at an early time point shows that the MeJA induced increase in paclitaxel is tightly regulated. AP2 and MYB TFs were found to mostly up-regulate the late pathway genes, thought to be the location of the rate limiting step (259). The exception to this was AP200499 which was shown to have a functional EAR domain (120) in site-directed mutagenesis experiments and could repress 5 paclitaxel promoters (Figure 38).

NAC, bHLH and WRKY TFs were mostly down-regulators of promoter activity, affecting genes across the whole pathway. The activity of the *Taxus* MYC2 homolog bHLH08058 differs from MYC2 TFs in other species. MYC2 is best known to act as a positive regulator of secondary metabolism, for example in *A. thaliana At*MYC2 positively regulates flavonoid biosynthesis (162), in *N. benthamiana Nt*bHLH1 and *Nt*bHLH1 activate nicotine biosynthesis (168) and the *C. roseus* homolog *Cr*MYC2 regulates TIA biosynthesis through activation of ORCA3 (173, 174). The TEA results show that bHLH08058 and bHLH08058-FE act as repressors on the paclitaxel biosynthetic pathway (Figure 46). The data partially concurs with that published by Lenka *et al* which identified three MYC2 like TFs as negative regulators of the paclitaxel biosynthetic pathway (184). Creation of RNAi or CRISPR/Cas9 bHLH08058-FE lines of *Taxus* CMCs may led to increased yields of paclitaxel, as the TF was shown to act as a repressor of the entire pathway.

The TEA results identified MYB10855 as a good candidate for bioengineering efforts to increase paclitaxel production. MYB10855 strongly activates eight paclitaxel biosynthetic promoters –*TASY, T13aH, TDAT, T10βH, DBBT, DBAT, PAM, DBTNBT* – which are spread across the whole paclitaxel biosynthetic pathway. MYB10855 is able to activate *DBTNBT* one of the two enzymes proposed to act as a bottleneck on the biosynthetic pathway but not the other, *BAPT* (219, 259). Overexpression of MYB10855 in *Taxus spp.* CMCs might therefore not lead to an

increase in paclitaxel yields, because only part of the pathway is activated. This was observed with ORCA3 which can activate four TIA biosynthetic enzymes but not the complete pathway; hence when ORCA3 was overexpressed in *C. roseus* cell culture TIA levels did not increase (176). The lack of interaction between MYB10855 and *BAPT*, however, may not be an issue in CMCs. Work completed by Waibel (221) showed that *BAPT* expression was significantly up-regulated in CMCs compared to DDCs, therefore in CMCs *BAPT* may not act as a bottleneck as was found in *T. baccata* platelets (47, 219, 259). Multiple TFs, such as MYB10855 and MYB15401, might be required to upregulate the entire paclitaxel biosynthetic pathway.

7.2.1 Combinatorial regulation

A combination of TFs might be required to increase paclitaxel production in *Taxus spp.* CMC culture, as was the case for increasing anthocyanin production in tomatoes with the TFs Del and Ros1 (Figure 11c) (204). MYB and bHLH TFs are well known to interact in a MBW complex to regulate anthocyanin production in numerous plant species (264, 273). MYB12379 contains the bHLH interaction motif $[DE]Lx_2[RK]x_3Lx_6Lx_3R$ (Figure 50b) (122) but with the exception of *T10βH* the addition of the two TcbHLHs did not lead to a significant increase in promoter activity, as observed in the positive control of PAP1 and EGL3 activating *DFR* expression (278). A wider screen of *T. cuspidata* bHLH may yet produce combinations that regulate the paclitaxel pathway, as the bHLHs tested had little homology to the known combinatorial regulators GL3, EGL3 and TT8 (100, 264).

The regulation of the TFs identified as modulators of paclitaxel biosynthetic enzymes is an important factor that needs to be considered. The 19 TFs will work as part of a network to produce a co-ordinated response to MeJA. Their activity could be regulated by combinatorial interactions with other proteins or by post-translational modifications. The investigations into MYB and bHLH combinatorial control showed that the activity observed by a TF in isolation can be altered by the addition of another TF. *T13αH, TASY* and *DBAT* were all shown to be activated by MYB10855 and repressed by bHLH08058. When combinatorial testing was conducted the MYB10855 activation of *T13αH* expression was supressed by the addition of bHLH08058 but in *TASY* and *DBAT* the significant up-regulation of the promoter was retained, although with slightly weaker activation.

JAZ proteins have been shown to interact with a wide variety of different TFs in *A. thaliana*, regulating their activity. The interaction of JAZ proteins with MYC2 repressing their ability to activate flavonoid biosynthesis is well established (162). However Y2H experiments did not identify a similar interaction between the Taxus MYC2 homolog bHLH08058 and identified JAZ proteins. The results of this experiment are flawed, however, because testing was completed with a truncated form of the protein lacking the JAZ interaction sites (230). JAZ proteins have also been shown to regulate numerous TFs including MYBs therefore, further experimentation conducted with the elongated form of bHLH08058 (bHLH08058-FE), the TcMYBs and the identified JAZ proteins may yet yield interactions, indicating that the JAZ proteins regulate the activity of candidate TFs.

7.2.2 Post-translational modifications

The regulatory activity of TFs can also be controlled at the post-translational level and a MAPK signalling cascade was identified as significantly up-regulated after MeJA elicitation. Treating the TEAs with 100 µM MeJA altered the regulatory activity of WRKY09595 and WRKY19284 (Figure 48), which both contain clusters of 4 and 5 serine residues at their N-terminus similar to those in WRKY33, whose activity is regulated by MPK3 and MPK6 in *A. thaliana* (165, 188). However, direct evidence is still required to implicate phosphorylation as the basis for the changes in regulatory activity.

A wide variety of post-translational modifications are known to modulate protein activity (186). Analysis of the four top candidate TFs, MYB10385, MYB10855, AP200499 and bHLH08058-FE using the cuckoo workshop GPS software (293, 299–302) and NetPhos 2.0 (268) to predict numerous possible phosphorylation, *S*-nitrosylation, sumolyation, palmitoylation and tyrosine nitration sites was undertaken (Table 17) . This data suggests that the activity of TFs could be modulated by numerous post-translational modifications. However, these predictions need to be viewed with caution as the majority of the work that these algorithms are based on was not conducted in plants, and without a crystal structure of the TF or a close homolog it is difficult to predict if a residue will be solvent exposed. Regulation of TF activity could be explored by treating transfected protoplasts with post-translational modification donors or inhibitors, such the broad spectrum kinase inhibitor staurosporine and the nitric oxide scavenger 2-4-carboxyphenyl-4,4,5,5-tetramethylimidazoline-1-oxyl-3-oxide (cPTIO).

	Phosphorylation			SNO	SUMO		Palmitoylation	Tyrosine Nitration
	Ser	Thr	Tyr		Sumolyation	SUMO interaction		
AP200499	9	1	0	0	0	0	0	2
MYB10385	24	4	5	1	0	1	2	2
MYB10855	21	5	6	2	0	0	2	4
bHLH08058	18	6	3	0	1	1	1	5
bHLH08058-A	23	6	5	0	2	3	1	5

Table 17 – A summary of the post-translational modifications predicted in MYB10385, MYB10855, AP200499 and bHLH08058-FE. The number of possible phosphorylation, *S*-nitrosylation (SNO) sumolyation, palmitoylation and tyrosine nitration sites were predicted using software: NetPhos 2.0 (268), GPS-SNO 1.0 (299), GPS -SUMO 1.0 (300), CCS-PALM 2.0 (301) and GPS-YNO2 1.0 respectively (302). The table states the number of sites predicted by the relevant software in each transcription factor (TF). Phosphorylation is broken down into the three possible residues that can be phosphorylated serine (Ser), threonine (Thr) and tryrosine (Tyr). Whether the protein is sumoylated or interacts with SUMO (small ubiquitin -like modifier) is stated. A number of post-translational modification have been predicted that may modulate TF activity including: a large number of phosphorylation sites in all four TFs; three *S*-nitrosylation sites in MYB10385 and MYB10855; sumoylation sites in AP200499 and bHLH08058; palmitoylation in all TFs except AP200499 and tyrosine nitration sites in all the four TFs.

7.3 Identifying the binding specificity of the TFs

One of the original aims of the project was to undertake an *in vitro* screen of the 19 TFs against the 10 paclitaxel biosynthetic promoters using EMSAs. The design of this experiment was however unrealistic due to the high labour requirement and extensive optimisation required to test

just a single protein. The identification of the potential BSs also lead to issues during the experimentation. A large number of cognate BSs were identified for the 19 candidate TFs in the paclitaxel biosynthetic promoters, however, the predictions are mostly based on work conducted in angiosperms and there is little gymnosperm specific information. BSs can vary between TFs, promoters and species and this is especially true for MYB TFs, which have an inherent flexibility in their binding specificity. The design of the EMSA studies was therefore slightly flawed as novel binding sites could not be identified. In hindsight a better methodology may have been to employ CASTing which is useful for identifying the binding specificity in TF families where little sequence information is available, such as for NAC TFs ANAC019 and ANAC092 (142), or for TF families, such as MYBs, where the BS is flexible, for example AtMYB61(290). Promoter deletion or truncation analysis could also have been used to identify the location of important regulatory elements within the promoter, an approach used to successfully identify the JERE element of the *STR* promoter that binds ORCA3 (133).

7.4 Transformation of *Taxus* CMCs

Unfortunately during this project a *Taxus spp.* CMC particle bombardment transformation system was not successfully established. Work conducted by our South Korean industrial partners, Unhwa, using *Agrobacterium* transformation has also not been successful in creating a reproducible protocol to date. This is a task that needs to be completed to accomplish the overall aim of the project to bioengineer *Taxus spp.* CMCs to improve paclitaxel production and confirm the results of the *A. thaliana* TEA screen in *Taxus spp*. A large number of potential TFs have been identified that could increase paclitaxel yields but without a transformation system this knowledge cannot be exploited.

7.5 Future work

There are many areas of the project that warrant further research including:

1. Employing Y2H assays to identify possible regulators of the 19 TFs. This includes testing for JAZ protein interactions with the elongated form of bHLH08058 and MYB TFs. A possible interaction between AP200499 and TPL could also be explored as TPL is known to interact with proteins that contain an EAR domain.

2. Exploring whether the addition of MeJA can alter the regulatory activity of the remaining 17 TFs in *A. thaliana* TEAs as observed with TcWRKYs; and if so elucidating the exact mechanism by which their activity is altered. This could be achieved by treating the TEAs with post-translational inhibitors such as the broad spectrum kinase inhibitor staurosporine and the nitric oxide scavenger cPTIO.

3. Identifying the binding specificity of key candidate TFs such as MYB10855 and bHLH08058 using CASTing (291) or promoter deletion analysis. Possible post-translational modification could also be explored by treating the purified protein with kinases to phosphorylate

the protein and thioredoxins to blunt redox modifications.

4. Isolation of the promoter regions of the 19 TFs using thermal asymmetric interlaced (TAIL) PCR (303) could be undertaken to provide information on how the TFs are regulated. TFs are well known to auto-regulate creating positive feedback loops and TFs can activate the expression of partner TFs; for example MYC2 is known to activate expression ORCA2 and ORCA3 in *C. roseus* (174) and ORC1 in *N. tabacum* (185), thereby indirectly regulating alkaloid production in both species. The promoter regions could be tested in *A. thaliana* TEAs to observe any auto-regulatory activity or regulation by other TFs. This information would then be used to construct a hierarchical network of how the TFs modulate the MeJA response.

5. The identification of the regulatory elements of the precursor pathway would be useful to understand how the whole terpenoid pathway is regulated. The transcriptomic analysis suggested that the MVA and MEP pathways were not significantly up-regulated after MeJA elicitation. If regulators of the pathway could be identified and manipulated it might increase the pool of IPP and DMAPP thereby possibly increasing the flux of precursors into the paclitaxel biosynthetic pathway. The candidate genes of the MVA and MEP pathway have already been identified therefore TAIL-PCR could be employed to identify the promoter region.

6. Current information about the rate limiting steps of the paclitaxel biosynthetic pathway was conducted mostly in *T. baccata* using platelets and cell culture. The expression of the paclitaxel biosynthetic genes has already been shown to differ between CMCs and DDCs (Waibel (221)); therefore research into the rate limiting steps in *Taxus* CMCs might identify different enzymes that need to be targeted to increase paclitaxel production. This work would require metabolic and genetic time course analysis of *Taxus spp.* CMCs treated with MeJA. The taxane quantity and composition could be quantified with liquid chromatography- mass spectrometry (LC-MS) and the transcript levels monitored by q-RT-PCR.

7. The identification of missing genes in the paclitaxel biosynthetic pathway is important work that needs to be completed. The missing enzymes could be identified using Cluster analysis of the *T. cuspidata* CMC transcriptome with the aim of identifying transcripts that have similar expression profiles to the known paclitaxel biosynthetic genes. Another option is *in silico* analysis (searching for candidates using known enzymes that have similar catalytic function) which was recently employed by Onrubia *et al.* to identify the missing enzyme β-phenylalanine-CoA ligase (44).

8. A *Taxus* transient and stable transfection system needs to be established for *Taxus spp.* CMCs. This is important to confirm the TF regulatory activity observed in the *A. thaliana* TEAs and to create *Taxus spp.* CMC lines that have the top candidate TFs either overexpressed e.g. MYB10855 or suppressed using RNAi or CRISPR/Cas9 e.g. bHLH08058-FE line. Further optimisation of the particle bombardment assay may yet yield a transient protocol and the recent publication by Lenka *et al.* 2015 (184) employing particle bombardment to identify the activity TcJAMYCs demonstrates that a sufficient transfection efficiency can be achieved. Exploration of

Taxus spp. protoplast transfection could also yield a viable transient protocol. *Agrobacterium* transformation is likely to be the best candidate for producing stable transformants and there is an example of this in the literature, but it can take over 6 months to grow sufficient calluses to grow in suspension culture (298). Once overexpression and suppression CMC lines have been established any change in taxane production can be quantified using LC-MS.

7.6 Summary

This project has been successful in identifying MeJA induced regulators of the paclitaxel biosynthetic pathway. Analysis of the MeJA elicited *T. cuspidata* CMC transcriptome found possible crosstalk between hormone signalling pathways and identified 19 candidate TFs. These were shown to regulate the paclitaxel biosynthetic promoters *in vivo* in a TEA screen, with two TFs identified that could regulate 80% of the pathway, MYB10855 a positive regulator and bHLH08058 a negative regulator. The possibility of combinatorial control of the paclitaxel biosynthetic pathway by MYB and bHLH TFs was also explored, but few significant changes in activity were identified. The possibility of TF activity being post-translationally controlled was investigated *in vivo* using Y2H assays with bHLH TFs and JAZ proteins, and *in vivo* using MeJA treated TEAs with WRKY TFs. The regulatory activity of the WRKY TFs was altered by the addition of MeJA but the mechanism by which it occurred still needs to be elucidated. An *in vitro* screen for possible interaction between the 19 TFs and the 10 paclitaxel biosynthetic promoters using EMSAs was realised to be impractical, but the binding specificity of MYB10855 was established. Even though the project was successful in identifying regulators of the paclitaxel biosynthetic pathway much more work needs to be conducted to accomplish the overall aim of the project, which is to produce a *Taxus* CMC line with superior paclitaxel yields.

Chapter 8 – Material and methods

8.1 General equipment and Materials

Unless otherwise stated all chemicals used were supplied by Sigma Aldrich at analytical grade (≥ 98 %) and all restriction enzymes were from New England Biolabs (NEB). All primer sequences are written 5'- 3' and were supplied by Life technologies. All chemical used in plant cell culture were bought from Duchefa.

PCR was completed using PTC-200 Peltier Thermal cycler (MJ Research) or G-Storm 482 Thermal Cycler. Centrifugation was done using Micormax centrifuge (Thermo), Accuspin Micro17 (FisherSci), Superspeed (Du pont) and Evolution (Sorvall). Luminometer results were obtained on a SpectraMax® M Series Multi-Mode Microplate Reader (Molecular Devices). Absorbance values recorded on GeneQuant-1300 (GE healthcare) and Nanodrop-1000. Magnification was conducted using a Nikon eclipse E600 microscopes using a Leica DFC 425C camera and blue light produced by a Nikon super high pressure mercury lamp.

Standard molecular protocols were followed unless stated from Sandbrook and Russel Molecular Cloning: a laboratory manual (304) and H_20 denotes sterilised double distilled water.

8.2 Plant material

8.2.1 Arabidopsis thaliana

Arabidopsis thaliana seeds of the ecotype Columbia (Col-0) were placed on soil composed of peat moss, vermiculite and sand at a ratio of 4:1:1 respectively and stratified for 48-72 hours (h) at 4°C in the dark. The seeds were then transferred to growth rooms under long (21°C, 16 h, 65% humidity and 100 $\mu mol.m^{-2}.s^{-1}$) and short (21°C, 8 h, 65% humidity and 100 $\mu mol.m^{-2}.s^{-1}$) day photoperiod conditions.

8.2.2. Maintenance of *Taxus spp*. CMCs

Taxus baccata, *cuspidata* and *media* CMCs were obtained from the South Korean biotechnology company Unhwa (64). The cells were grown in liquid media at 100 revolutions per minute (rpm) at 25 °C in the dark and were sub cultured every 14 days (see Table 20 for media composition). Solid media for *T. cuspidata* and *T. baccata* was prepared by the addition of 2 % gelrite and 0.05 % charcoal to the culture media and cells were grown on solid medium at 25 °C and sub cultured every 14 days.

8.3 Transcriptomic Profiling of *Taxus cuspidata* CMCs

Elicitation was performed 5 days post sub-culture with 100μM MeJA. Control (unelicited) cells were subject to mock elicitation by adding equal volume of ethanol. For transcriptomic analysis the cultures were sampled at 0.5, 2 and 12 h time points after elicitation in biological triplicate.

RNA was extracted using RNeasy plant RNA kit (Qiagen) following manufacturer's instructions. The *T. cuspidata* transcriptome was established using Roche454 sequencing using Edge R to map and annotate the contigs (224). Global gene expression after MeJA elicitation was performed using Illumina Solexa sequencing and differential expression was assessed using a modified Fisher's exact test. A false discovery rate (FDR) of ≤ 0.05 was used to detect differentially expressed contigs (*n* = 1,229) (64).

8.4 Gene expression analysis after treatment with MeJA and CHX

8.4.1 Treatment of *T. cuspidata* CMCs

T. cuspidata CMCs 5 days post sub-culture were pre-treated with 100 µM CHX for 1.5 h and then elicited with 100 µM MeJA, with DMSO (0.2 % v/v) used as a mock control (241). Cells were harvested at various time points after MeJA elicitation, the liquid media removed and then samples were frozen in liquid nitrogen and stored at -80 °C.

8.4.2 RNA extraction

RNA was extracted using an adapted CTAB method (305, 306). Briefly, 200 mg of frozen sample was powered in liquid nitrogen using a mortar and pestle. The sample was then added to 900 µl of pre-warmed (65 °C) extraction buffer, thoroughly vortexed and incubated at 65 °C for 10 minutes (min). An equal volume of chloroform: isoamyl alcohol (24:1 v/v) was added and the tube inverted vigorously and centrifuged at 11,000 *g* for 10 min at 4 °C.

The aqueous layer was recovered and a second extraction with chloroform:isoamyl alcohol performed. The supernatant was transferred to new a microcentrifuge tube and the RNA precipitated with LiCl (3M final concentration). The mixture was left overnight at 4 °C and the RNA pelleted after centrifugation at 17,000 *g* for 20 min at 4 °C.

The pellet was resuspended in 500 µl of SSTE buffer (pre-heated at 65 °C) and an equal volume of chloroform:isoamyl alcohol added. The mixture was then centrifuged at 17,000 *g* for 10 min at 4°C and the supernatant transferred to a new microcentrifuge tube. The RNA was precipitated with 0.7 volume of cold isopropanol and immediately centrifuged at 17,000 *g* for 20 min at 4 °C. The pellet was washed with 70 % ethanol, dried and resuspended in DEPC–water.

8.4.3 RT-PCR analysis

cDNA was synthesised using an Omniscript reverse transcriptase kit (Qiagen) and subsequent real time (RT)- PCR reactions performed using Crimson Taq (NEB). The primer sequences employed are listed in Table 23.

8.5 Cloning

8.5.1.1 Preparation of *E. coli* using KCM method

An overnight culture of *E. coli* was diluted 1/100 in LB then grown at 37 °C until an OD_{600} 0.4-0.6 was reached. Cells were pre-chilled in an ice bath for 2 min then centrifuged in sterile pre-

chilled bottle at 5000 rpm for 10 min at 4 °C. The supernatant was discarded and the resulting pellet resuspended in a 1/10 volume of ice cold TSS. The cells were aliquoted then snap frozen in liquid nitrogen and stored at -80 °C.

8.5.1.2 Transformation of *E. coli* using KCM method

Cells were slowly thawed on ice and then 100 µl was added to a tube containing 20 µl 5X KCM, DNA and ddH_2O to a total of 100 µl and incubated on ice for 20 min. The cells were transformed using heat shock of 37 °C for 5 min and then rested on ice for 2 min. 500 µl of pre-warmed LB was added to the tube and incubated at 37 °C for 40-60 min. The transformation was plated out onto selective media and incubated overnight at 37 °C.

8.5.1.3 Identification of Positive Transformants

Positive transformants were identified using colony PCR using appropriate primers and Taq polymerase, for example pDONR221 based cloning required M13 primers to detect positive colonies (Table 28). The nucleotide sequence of the insert was confirmed using Sanger sequencing completed by Genepool, Edinburgh UK.

8.5.2 Extraction of genomic DNA from *Taxus spp.*

100 mg of frozen sample was powered in liquid nitrogen using a chilled mortar and pestle, added to extraction buffer and incubated for 30 min at 65 °C in a shaking water bath. An equal volume of chloroform/isoamylalcohol (24:1) was then added, the sample mixed and centrifuged at 10 000 *g* for 10 min at 4 °C. The aqueous phase was transferred to the new microcentrifuge tube and the chloroform/isoamylalcohol extraction repeated. The DNA was then precipitated using 1.5 volume of isopropanol and incubated at 20 °C for at least 1 h. The DNA was then recovered by centrifuging the sample for 30 min at 17, 000 *g* at 4 °C and the pellet washed with 70 %. The pellet was then air dried and resuspended in 50 µl of TE by incubated at 65 °C for 5 min.

8.5.3 Cloning of p2GW7,0-Rennila luciferase (p2GW7,0-Rluc)

Renilla luciferase was amplified from the pRL-TK vector (Promega) using primers Rluc-*Spe*I and Rluc-*Sac*II (Table 28) and Phusion polymerase (NEB). The purified PCR product and p2GW7,0 vector(254) were separately digested with restriction enzymes *Spe*I and *Sac*II (NEB), under conditions specified by manufacture. The digested products were mixed together in a vector: insert ratio of 1:3 and ligated using T4 DNA ligase (NEB). Once ligation was complete the products were transformed into *E. coli* DH5*α*.

8.5.4 Gateway™ cloning

TFs were amplified from cDNA and promoters from gDNA using Phusion polymerase (NEB). Gateway™ cloning technology (Invitrogen) is based on the site-specific recombination properties of bacteriophage lambda (307) and was used to clone the TFs and promoters into destination vectors p2GW7,0 (254) and pGWlucB (255) respectively (see Table 21 and Table 22

for primers and Appendix 1 for vector maps). Renilla luciferase was also cloned into p2GW7,0. TFs were cloned into expression vectors pDEST15 and pDEST24 depending on the location of their binding site, producing either an N terminal or C terminal tagged protein respectively (see Table 25 for primers and Appendix 1 for vector maps). bHLH TFs were cloned into prey vector pDEST22, MYB TFs were cloned into the bait vector pDEST32 and JAZ proteins were cloned into pDONR221 and then pDEST32 (see Table 27 for JAZ primers).

8.5.4.1 BP reaction

Primers were designed with *att*B flanking sequences to produce an *att*B–PCR product, which was amplified using Phusion Polymerse (NEB) and then gel purified using GeneJET gel extraction kit (Life technologies). pDONR221 (3 µl), *att*B –PCR product (3 µl) and BP Clonase II (Invitrogen) (1.5 µl) were mixed and incubated at 25°C for 3 h - overnight. Proteinase K (1 µl) was added to terminate the reaction and the mixture incubated at 37°C for 10 min. 1µl of the resulting mixture was transformed in *E. coli* DH5α.

8.5.4.2 LR reaction

pENTRY (3 µl), p2GW7.0 (1 µl) and LR Clonase II (Invitrogen) (1µl) were mixed well and then incubated at 25°C between 3 h - overnight. Proteinase K (1 µl) was added to terminate the reaction and the mixture incubated at 37°C for 10 min. 1µl of the resulting mixture was then transformed in *E. coli* DH5α.

8.5.5 Site directed mutagenesis of AP200499

The truncated form of AP200499 was created by designing primers to amplify AP200499 with the C-terminal region missing and using Gateway™ technology to clone the fragment into pDONR221 and subsequently into p2GW7,0. Site-directed mutagenesis reactions were performed following a protocol from Sandbrook and Russel (304) to produce mutations within the EAR domain, L161, and outside the EAR domain, L167. The primers used for cloning are in Table 24.

8.6 Arabidopsis protoplast transient assay

8.6.1 DNA isolation and purification

DNA was extracted using with either a GeneJET maxiprep kit (Fermentas) or QIAfilter Maxiprep Kit (Qiagen). The integrity and purity of the DNA were ascertained by electrophoresis on a 1% agarose gel, the $A_{260\ nm}/A_{280\ nm}$ ratio and the $A_{260\ nm}/A_{230\ nm}$ ratio. The $A_{260\ nm}/A_{280\ nm}$ ratio was in the range of 1.8-2.0 and the $A_{260\ nm}/A_{230\ nm}$ ratio was in the range of 2.0-2.2. The concentration was calculated using spectrophotometer at an absorbance of 260 nm.

8.6.2 Protoplast Isolation

8.6.2.1 Tape Sandwich method

Protoplasts were isolated from *Arabidopsis thaliana* mesophyll tissue following the Tape

Sandwich protocol (242). Leaves were collected from 3 to 5 week old plants grown under optimal and low light conditions. The upper epidermal surface was secured to autoclave tape, while magic tape was affixed to the lower epidermal surface and gently pressed down. The magic tape was then carefully pulled away removing the lower epidermal layer. The autoclave tape was placed with the leaves facing down in a petri dish containing 10 ml of enzyme solution. The leaves were gently shaken at 50 rpm in the light for approximately 40 min until the protoplasts were released into solution.

Protoplasts were centrifuged in a round bottom tube at 100 *g* for 3 min. The enzyme solution removed and the protoplasts washed twice in ice chilled modified W5 solution and then incubated on ice for 30 min. During this period the quality and quantity of the protoplasts was calculated using a haemocytometer. The protoplasts were centrifuged at 100 *g* for 3 min and resuspended in MMG solution to a final concentration of 2 x 10^4 – 5 x 10^5 protoplasts (pps)/ml.

8.6.2.2 Sheen Method

Protoplasts were isolated from *Arabidopsis thaliana* mesophyll tissue following the Sheen protocol (249). Leaves were collected from 3 to 5 week old plants grown under optimal and low light conditions. 0.5-1 mm leaf strips were cut using a razor blade and immediately transferred into enzyme solution dipping both sides and ensuring the strips were completely submerged. The strips were put under vacuum desiccation for 30 min in the dark and then left for a further 3 h at room temperature in the dark. An equal volume of W5 solution was added and solution filtered through 75 µm nylon mesh. The resulting solution was centrifuged at 100 *g* for 2 min, the supernatant discarded and the resulting pellet resuspended in W5 solution to produce a final concentration 2 x 10^5 pps/ml. The protoplasts were rested on ice for 30 min and during this time the quality and quantity of the protoplasts was observed using a haemocytometer. The solution was centrifuged at 100 *g* for 2 min, the supernatant discarded and the resulting pellet resuspended in MMG solution to a concentration of 2 x 10^4 – 5 x 10^5 pps/ml.

8.6.3 Transfection of *Arabidopsis* protoplasts

8.6.3.1 Method A- Sheen

Protoplasts were transfected following the Sheen protocol (249).100 µl of protoplasts (2 x 10^4 – 5 x 10^5 pps/ml) were added to 10 µl of DNA (20 µg) and gently mixed. Freshly prepared polyethylene glycol (PEG)/calcium solution (110 µl) was added and the transfection mixture incubated at room temperature for 5-30 min. The solution was then diluted with 420 µl W5 solution and mixed by gently rocking. The solution was centrifuged at 100 *g* for 2 min, supernatant discarded and protoplasts resuspended in 1 ml WI solution and incubated in 6-well tissue culture plate, coated with 5% BSA, for 2-16 h at room temperature. Protoplasts were harvested by centrifugation at 100 *g* for 2 min, the supernatant removed and the samples frozen on dry ice and stored at 80 °C.

8.6.3.2 Method B - Tape Sandwich

Protoplasts were transfected following the Tape Sandwich protocol(242). 200 µl protoplasts ($2 \times 10^4 – 5 \times 10^5$ pps/ml) was mixed with 15 µl DNA (20-30 µg) at room temperature. An equal volume of freshly prepared PEG/calcium solution was added, mixed and then incubated at room temperature for 5 min. After incubation the mixture was diluted with 3 ml of modified W5 solution and centrifuged at 100 *g* for 1 min. The wash step was repeated twice and the final pellet resuspended in 1 ml of W5 and incubated in 6 well plated coated with 1% BSA at room temperature for 16 hr in the light. Protoplasts were harvested by centrifugation at 100 *g* for 2 min, the supernatant removed and the samples frozen on dry ice and stored at 80 °C.

8.6.3.3 Method C - Negrutiu

Protoplasts were transfected following the Negrutiu protocol (308). 250 µl of protoplasts ($0.5- 1 \times 10^6$ pps/ml) were added to 20µl of DNA (20µg) and gently mixed. Freshly prepared PEG/calcium solution adjusted to pH 8-9 was very slowly added to the mixture and gently mixed. The transfection mixture was incubated at room temperature for 15-30 min then 10ml of W5 solution was carefully added and subsequently centrifuged at 100 *g* for 5 min at 4°C. The resulting pellet was resuspended in 1.5ml of W5 solution and incubated in 6-well plated coated with 1% BSA at room temperature for 20 h in the dark. Protoplasts were harvested by centrifugation at 100 *g* for 2 min, the supernatant removed and the samples frozen on dry ice and stored at 80 °C.

8.6.3.4 Method D – Optimised Protocol

Protoplasts were transfected following an optimised version of the Negrutiu protocol (308). 250 µl of protoplasts (5×10^5 pps /ml) were added to 14 µl of DNA (1 µg/µl) in ratio of 4:5:5 (effector:reporter:internal control). Freshly prepared 40% PEG/calcium solution adjusted to pH 8-9 was added, mixed gently and incubated for 5 min at room temperature. 5 ml of Mod W5 solution was slowly added to the solution and then centrifuged at 100 *g* for 5 min. The supernatant was removed and the resulting pellet resuspended in 1.4ml of modified W5 solution and incubated in 6-well plated coated with 1% BSA for 20 h in the light at 25 °C. Protoplasts were harvested by centrifugation at 100 *g* for 2 min, the supernatant removed and the samples frozen in liquid nitrogen and stored at 80 °C.

8.6.4 Dual Luciferase Assay

The Dual Luciferase Assay was conducted using Dual-Luciferase Reporter Assay System (Promega®) following manufacturer's instructions. Briefly, 100 µl of protoplast lysis buffer was added to frozen protoplasts, vortexed for 2 seconds (s) and left on ice for 5 min. The solution was then centrifuged at 1,000 *g* for 2 min. The luminometer was set up to record a 10 s measurement for each assay and 100 µl of LARII was pre-dispensed into an opaque luminometer plate. 20 µl of lysate was added, mixed via pipetting and the reading taken immediately (firefly measurement). 100 µl of Stop and Glo reagent was then added and the second measurement was taken (*Renilla*

measurement).

8.7 Recombinant protein production

8.7.1 SDS-PAGE gel and Western Blot

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot were employed to observe protein production and were performed according to standard laboratory procedures (304). Proteins produced *via* the *in vitro* transcription/translation system were detected using anti-FLAG as the primary antibody (1: 2000) and anti-mouse-HRP conjugate (1:5000) as the secondary antibody. The detection of proteins expressed in *E. coli* used the anti-GST-HRP conjugate antibody (1:5000).

8.7.2 Recombinant protein production in *E. coli*

TFs were cloned into the expression vectors pDEST15 or pDEST24 and transformed into the *E. coli* expression strain BL21 using the KCM method. Successful colonies were checked using colony PCR and a single one grown in 5 ml of LB overnight at 37°C. The overnight culture was diluted 1/100 in LB and grown in a shaker at 37°C until and an OD_{600} of 0.6-0.7, when IPTG (final concentration 0.1 mM) was used to induce recombinant protein production. The cells were then grown for a further 6 h at 37 °C and then harvested by centrifugation for 5 min at 5000 rpm at 4 °C and stored at -80 °C. The frozen pellet was resuspended on ice using 1/10 volume of freshly made *E. coli* lysis buffer and an aliquot analysed using SDS-PAGE and Western blot.

8.7.3 Purification of recombinant proteins

GST-tagged recombinant proteins were purified using Glutathione Agarose beads 4B (Invitrogen®) following manufacturer's instructions. Briefly, the lysate of the overexpressed proteins was incubated with the agarose beads on a roller for 30 min at 4°C. The beads were then washed with PBS buffer to remove non bound proteins. Proteins were eluted off the beads with elution buffer (20 mM glutathione, 50mM NaCl) incubated for 30 min on the roller for 30 min at 4°C.

8.8 EMSA

8.8.1 Annealing of DNA probes

20 bp DNA oligoes were designed to contain the putative binding site and surrounding sequence and are summarised in Table 26. The oligoes were annealed using 10 µM of forward and reverse strand in a PCR thermocycler using the conditions 95 °C - 10 min, 60 °C – 30 min, 28 °C – 30 min and 18 °C – 30 min.

8.8.2 Labelling DNA probes

100 ng of each oligo was labelled with γ-^{32}P-ATP using T4 polynucleotide kinase for 1 h at 37 °C. The reaction was then purified using sephadex G-50 column (GE Healthcare) to eliminate excess nucleotides and stored at -20 °C.

8.8.3 Binding Reaction

5 µl of purified protein or extract was added to 1x binding buffer containing poly d(I-C) (0.5 µg/µl) and pre-incubated for 20 min at 4 °C. 1 µl of the labelled probe was then added subsequently incubated for 20 min at room temperature. For competition experiments an increasing molar excess of non-radioactive competitor was added to the mixture.

The binding reaction was run out on a 6% non-denaturing polyacrylamide gel pre-run for 30 min at 20 V to eliminate excess APS. Gel composition; acrylamide/bis (29:1) (0.2%), TBE (0.15x), APS (3.33%) and TEMED (0.1%) and 0.25x TBE was used as the electrophoretic buffer. The reaction was run for 80 min at 120V and the gel transferred to Whatman™ 3MM filter paper and dried for 2 h at 80 °C under vacuum. The gel was exposed using X-ray film overnight at -80 °C.

8.9 Yeast-two-hybrid analysis

pDEST32-MYB/pDEST22-bHLH and pDEST-JAZ/pDEST22-bHLH combinations were transformed into yeast strain AH109. The yeast was transformed by harvesting 1 ml of overnight culture by centrifugation (grown at 30 °C, 200 rpm), washing the resulting pellet twice with water and resuspending in 1x LiAc/TE buffer (1xTE, 100mMLiAc). This culture was incubated at 30 °C for 30 min, harvested by centrifugation and resuspended in 500 µl reaction buffer. 5 µl of each plasmid was added and the mixture incubated at 30 °C for 30 min and then heat shocked at 42 °C for 15 min. The cells were harvested by centrifugation and resuspended in 1x TE buffer and plated out on selective media deficient in leucine, tryptophan and histidine (SC-L-W-H). Resulting colonies were grown overnight in YPDA then diluted in H_2O (1:10, 1:100 and 1:1000) and 5 µl spotted onto selective media (SC-L-W-H) and non-selective media deficient in leucine and histidine (SC-L-W). The strength of the interactions was test by the addition of increasing concentrations of 3-amino-1,2,4-triazole (3AT).

8.10 Solutions

Solution	Final Concentration	Reagent
W5 Solution	2mM	MES pH 5.7
	154mM	NaCl
	125mM	$CaCl_2$
	5mM	KCl
Modified W5	2mM	MES pH 5.7
	154mM	NaCl
	125mM	$CaCl_2$
	5mM	KCl
	5mM	Glucose
MMG Solution	4mM	MES pH 5.7
	0.4M	Mannitol
	15mM	$MgCl_2$
WI solution	4mM	MES pH 5.7
	0.5M	Mannitol
	20mM	KCl
Protoplast Enzyme solution Sheen	20mM	MES pH 5.7
	0.4M	Mannitol
	20mM	KCl
	10mM	$CaCl_2$
	0.1% wt/v	BSA
	1.5%	Cellulase
	0.4 %	Macerozme
Protoplast Enzyme solution Tape Sandwich	20mM	MES pH 5.7
	0.4M	Mannitol
	20mM	KCl
	10mM	$CaCl_2$
	0.1% wt/v	BSA
	1% wt/v	Cellulase
	0.25%	Macerozme
PEG/Calcium transfection solution	20-40 wt/v	PEG
	0.1M	$CaCl_2$ / $Ca(NO_3)_2$
	0.2M	Mannitol
5 x TBE	1.1M	Tris-Cl
	900mM	Boric acid
	25mM	EDTA pH 8.5

1 x TAE	40mM	Tris-Cl
	1mM	EDTA pH 8.5
	20mM	Glacial acetic acid
TE Buffer	10mM	Tris-Cl pH 8.4
	1mM	EDTA
5x KCM	0.5M	KCl
	0.15M	$MgCl_2$
	0.25M	$CaCl_2$
TSS	73ml	LB
	10%	PEG 3350
	5%	DMSO
	20mM	$MgSO_4$ pH 6.5
	ddH_2O 100ml	
***E. coli* Lysis Buffer**	25 mM	Tris-HCl of pH 7.5
	100mM	NaCl
	0.1%	Triton
	1mg/ml	Lysozyme
	1x	proteinase inhibitor
	1mM	DTT
EMSA Binding Buffer	20mM	HEPES pH 7.9
	1mM	DTT
	50mM	KCl
	5 %	Glycerol
	5 mM	$MgCl_2$
	50 µg/µl	BSA
RNA Extraction buffer for *Taxus* CMCs	2 %	CTAB
	2.5 %	PVP-40
	2 M	NaCl
	100 mM	Tris-HCl pH 8.0
	25 mM	EDTA pH 8.0
	2 %	β-mercaptoethanol
gDNA Extraction Buffer for *Taxus* CMCs	100mM	Tris-HCl pH 8.0
	25mM	EDTA
	2M	NaCl
	2%	CTAB
	500 µl/l	Spermidine
	2%	PVP
	5%	PVPP
	2 %	β-mercaptoethanol

SSTE	10 mM	Tris–HCl pH 8.0
	1 mM	EDTA pH 8.0
	1 %	SDS
	1 mM	NaCl
Y2H reaction buffer	100 mM	LiAC
	33.3 %	PEG
	20 µg	Salmon sperm DNA

Table 18 – The composition of solutions, showing final concentrations and where appropriate pH requirements.

8.11 Growth Media

Media	Concentration (wt/v)	Reagent
Bacterial growth media	1 %	Tryptone
- LB media	0.5 %	Yeast extract
	1 %	Sodium chloride
	2 %	Agar (for solid media)
Yeast Growth Media	1 %	Bacto yeast extract
- YPD media	2 %	Bacto peptone
	2 %	Glucose
	2 %	Agar (for solid media)
- Media for Y2H	1 %	Bacto yeast extract
	2 %	Glucose
	0.06 %	Amino acid supplement L-W-H (Selective)
	0.06 %	Amino acid supplement L-W (Non-selective)
	2%	Agar (for solid media)

Table 19 – The composition of media for bacterial and yeast.

8.11.1 Taxus culture media

Species	Final Conc^{n} (mg.L)	Reagent	Sterilisation
Taxus Media	1012	Potassium nitrate	Autoclave
	121.56	Magnesium sulphate heptahydrate	
	10	Manganese sulphate hydrate	
	2	Zinc sulphate hepthydrate	
	0.025	Copper sulphate pentahydrate	
	113.23	Calcium dihydrate	
	0.75	Potassium Iodide	
	0.025	Cobalt dichloride hexahydrate	
	130.44	Sodium phosphate monobasic dehydrate	
	3	Boric acid	
	0.25	Sodium molybdate dehydrate	
	36.7	FeNa – EDTA	
	200	Myo-inositol	
	20	Thiamine-HCL	
	2	Nicotinic acid	
	2	Pyridoxine-HCL	
	133	Aspartic acid	
	175	Arginine	
	75	Glycine	
	115	Proline	
	1	Picloram	
	pH 5.5		

	0.1	Gibberellic acid 3	Filter
	100	Ascorbic acid	
	150	Citric acid	
	2%	Sucrose	
Taxus cuspidata	1012	Potassium nitrate	Autoclave
	121.56	Magnesium sulphate heptahydrate	
	10	Manganese sulphate hydrate	
	2	Zinc sulphate hepthydrate	
	0.025	Copper sulphate pentahydrate	
	113.23	Calcium dihydrate	
	0.75	Potassium Iodide	
	0.025	Cobalt dichloride hexahydrate	
	130.44	Sodium phosphate monobasic dihydrate	
	3	Boric acid	
	0.25	Sodium molybdate dihydrate	
	36.7	FeNa - EDTA	
	450	Myo-inositol	
	20	Thiamine-HCL	
	2	Nicotinic acid	
	2	Pyridoxine-HCL	
	133	Aspartic acid	
	175	Arginine	
	75	Glycine	
	115	Proline	
	1	Picloram	
	pH 5.5		
	500	Casein hydrolysate	Filter
	0.5	Gibberellic acid 3	
	100	Ascorbic acid	
	150	Citric acid	
	3%	Sucrose	
Taxus baccata	2500	Potassium nitrate	Autoclave
	130.44	Ammonium sulphate	
	121.56	Magnesium sulphate heptahydrate	
	10	Manganese sulphate hydrate	
	2	Zinc sulphate hepthydrate	
	0.025	Copper sulphate pentahydrate	
	113.23	Calcium dihydrate	
	0.75	Potassium Iodide	
	0.025	Cobalt dichloride hexahydrate	
	130.44	Sodium phosphate monobasic dihydrate	
	3	Boric acid	
	0.25	Sodium molybdate dihydrate	
	36.7	FeNa - EDTA	
	100	Myo-inositol	
	10	Thiamine-HCL	
	1	Nicotinic acid	
	1	Pyridoxine-HCL	
	3	2,4-D	
	0.5	Kinetin	
	1500	Polyvinyl pyrrolidone (PVP10)	
	3%	Sucrose	
	pH 5.6		

Table 20 – The composition of Taxus baccata, media and cuspidata liquid culture media. Table give the final concentration of each compound, the pH of the solution altered with sodium hydroxide and the method of sterilisation.

8.12 Primers

8.12.1 Gateway™ cloning of 19 candidate TFs for TEA

TF	Forward	Reverse
AP2 03304	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGCTTCCATGACGAAGGGATTATT	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGC ATCCAAGTGTAGGGTGGTG
AP2 04485	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGAAATATCATGCAAAGGACAGAGCG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGTA TTCCTCAAATAGTTCCGACTCCCACA
AP2 07245	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGCGCGTCGTCGCAGTTGGGTAA	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGCC ATAGGTCTTCATCCTTAAT
AP2 01431	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGAGGATCACCAGCATATGAGAG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAAA TGTGCTTCTCGTGAAGGAT
AP2 00499	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGCGTCGACAATGGAAAGGTTAG	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATTG CCCTTCTTTTTCTTCTTCCTCC
AP2 22386	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GCCGGCCAGGAAGAGGATCCAAA	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAGA GAAATGATCAGAAGTAGGAGAAGGG
MYB 12379	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGCTCCAATCCCCAGGCATTCTT	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATAG ACTTTCCCAGAAGTCACATAAATCTG
MYB 10385	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGGGCGAGCTCCGTGCTGTGATA	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACAA CGTAGGCGATGATGCGGCT
MYB 10855	GGGGACAAGTTTGTACAAAAAAGCAGGCTTC ATGTGTCAAACACAAGAGGAAGGAG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCA ACGAAGTCCTTCTGTGGTACC
MYB 15401	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GAGCTCCAAGTGGTTGAATTTGA	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATAT TGGCTGTTGCATTGGAAT
NAC 06771	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGGTAAAGATGGTCTAGAAGGGAAGG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACTG GAAAATCATATCAGACGAGTGG
NAC 09658	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GAGTGCACATAGTAGAACAAAGACAAC	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAATT GAGGGTAATACTTGTTGTTGTTGTT
NAC 00172	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGGGAGAGAAGGACCTCCCGGAC	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTG TGGTCTAGGTTGCAGATAG
NAC 05638	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGAGTCCCTGGCGGTGAATGTGA	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAATA GCCAGACCACAAACAGTCTGGC
NAC 08447	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GTGTCAAACACAAGAGGAAGGAGCC	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATGT AGTTCCCACTCCCAGAAAATCA
bHLH11748	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGAGTTGGCAGGTGACCAGGGTT	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACTC ACTCAAAATGGTTGCCCCTAAATTC
bHLH08058	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GAGAACTGAAGATACCCTTCATT	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATTT AGCAATCACTGTCTGAATC
WRK09595	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GTTCATGATAAACCTCATG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATGA ATCTTGAGTCAGAAACCCTTG
WRK19284	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GGACCAGCACAAAATTGAGCTTTTG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTTT GGGAATAACTCTTGTGTATTTCTGATAAAA

Table 21 – The nucleotide sequences of the primers used to clone 19 candidate TFs into pDONR221 and then subsequently into p2GW7,0 for use in the TEA.

8.12.2 Gateway™ cloning of paclitaxel promoters for TEA

Promoter	Forward	Reverse
TASY	GGGGACAAGTTTGTACAAAAAAGCAGGCTAACTCG CAATAGCTAGGACATCTT	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCTGCA GAGAGGCAGGG
T5αH	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGGATC GAGATGTCAAGAA	GGGGACCACTTTGTACAAGAAAGCTGGGTATAATAC TTTCTCGGCCATCA
T13bH	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGCC TAAATCTCACGTATGT	GGGGACCACTTTGTACAAGAAAGCTGGGTGGAAAGG AGTAAAGGGGTTA
TDAT	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGGTGA TATATATCAGGATCGAG	GGGGACCACTTTGTACAAGAAAGCTGGGTTGGTCGA AATACTGATAAAAGAG
T10bH	GGGGACAAGTTTGTACAAAAAAGCAGGCTAACCAG CTGTTATGCCA	GGGGACCACTTTGTACAAGAAAGCTGGGTTTGGAGC AGGTGGGT
DBBT	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGTCAA ATTTAACATCGCTGTA	GGGGACCACTTTGTACAAGAAAGCTGGGTTGTAGAT ATTTGGACTCTCTTCTCCT
DBAT	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCCAC AAAACTAGACAAGTCAT	GGGGACCACTTTGTACAAGAAAGCTGGGTAAAGCTG GGTCTGTTGAGC
PAM	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTACCT AAACAGACAAGACAACG	GGGGACCACTTTGTACAAGAAAGCTGGGTTGCAGAG CAAGGAAAATAAA
BAPT	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAATCC CTAACACACACATGCATAC	GGGGACCACTTTGTACAAGAAAGCTGGGTGGAATTG AGCAGCTGAATAATTTTTTAT
DBTNBT-1	GGGGACAAGTTTGTACAAAAAAGCAGGCTGTTATA TCGTCGAGGCTTGC	GGGGACCACTTTGTACAAGAAAGCTGGGTCCCAATG ATCCACGAGG
DBTNBT-2	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCTCGT GGATCATTGGG	GGGGACCACTTTGTACAAGAAAGCTGGGTGACTGGA TCAAAGATGAAACGat

Table 22 – The nucleotide sequences of the primers used to clone the 10 paclitaxel promoters into pDONR221 and then subsequently into pGWlucB in the TEA. DBTNBT broken into two fragments fragment 2 is the first 900 bp of the promoter and fragment 1 is the remaining 900 bp.

8.12.3 RT-PCR of 19 candidate TFs

TF	Forward	Reverse
AP2 03304	CCTGCTGAATCCGAACATCT	TCACAACCGCTATGGACTTG
AP2 04485	ACAACTGGAATTGGCCCTCT	CGTTTGCAGAGGAAAGGTGT
AP2 07245	TCGCAGTCTCGTTGAAGAAA	TCAGTCAGTCCACAGCGAGT
AP2 01431	GAAGAAAAAGTGCGCAGTCC	TTTGGGCTTCAAATTGCTTC
AP2 00499	TATTCTTCGCACGCTTCTCC	CGTCTCGGTCTTCTTGGAAA
AP2 22386	CCAAACCTCTCTGGAAAGCA	GGGCTGAATCCACAACTCAT
MYB 12379	CAACACAACTCGAAGCAGTCA	ACCGTGTAGTCATGGGGTTC
MYB 10385	TTTCCCTCAACAATGGAGGA	CGTTCACCCCCAGTTTAATG
MYB 10855	TTTCTGGGAGTGGGAACTACA	CCATGAAACAATCCCTCCAT
MYB 15401	CATCCCAATATTCCGACGAC	GGTCTGCCGTGGACTTAGAA
NAC 06771	AATGTTGATGGCGTTCACAG	TGCACGTTGTAGAGGCTGTC
NAC09658	CAACAACAATGTGCCAACCT	TTGAACGTGAGCTTCTGTGG
NAC 00172	CACAATTGGGATTTGGGTATG	GAAATGCACTGTCGAACACC
NAC 05638	CCCTTTCTCCCAGGTTCTTT	GGGAAGCAAAATGGGTCTAA
NAC 08447	CGGATTCAGCTCCATTCTTG	ACAGCCAAGGAGCAGGACTA
bHLH 11748	AGCAGTAAGCAACGATGCAG	CTCAAAATGGTTGCCCCTAA
bHLH 08058	GATGGTTTTCTCCATTGTTGC	CATAAGCCTTTGCCTTTTGC
WRKY 09595	CCCATGGATCTGGTTGGTTA	TGTCTTCGACCCGCATAAAT
WRKY 19284	AGGGCTCAGATGGTTACAGC	CAGCCAAGGCTGCAGTATAA

Table 23 – The nucleotide sequences of the primers used in RT-PCR of 19 candidate TFs.

8.12.4 RT-PCR of paclitaxel biosynthetic enzymes

Primer Name	Sequence
AP200499-END-For	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAGGAGATAGAACCATGGCGTCGACA ATGGAAA
AP200499-END-Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAAGCGGGAATCTTGCCTCT
AP200499-L161	F - CAGAGGCGGGAGAAGATTGGCATCTAAAAGCGGGAATCTT R – AAGATTCCCGCTTTTAGATGCCAATCTTCTCCCGCCTCTG
AP200499-L167	F - TCTTTTTCTTCTTCCTCCGCAGGCGGGAGAAGATTGAGAT R - ATCTCAATCTTCTCCCGCCTGCGGAGGAAGAAGAAAAAGA

Table 24 – The nucleotide sequences of the primers used to create mutations in the EAR motif of AP200499.

8.12.5 Gateway™ cloning of 19 candidates for Protein Expression

TF	Tag	Forward	Reverse
AP2 03304	N-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGC TTCCATGACGAAGGGATTATT	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAGG CATCCAAGTGTAGGGTGGTG
AP2 04485	N-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGA AATATCATGCAAAGGACAGAGCG	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAGTAT TCCTCAAATAGTTCCGACTCCCACA
AP2 07245	N-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGC GCGTCGTCGCAGTTGGGTAA	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAGCC ATAGGTCTTCATCCTTAAT
AP2 01431	N-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGA GGATCACCAGCATATGAGAGTC	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAAAT GTGCTTCTCGTGAAGGATAC
AP2 00499	N-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGC GTCGACAATGGAAAGGTTAGAA	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATTGC CCTTCTTTTTCTTCTTCCTCCTC
AP2 22386	N-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCC GGCCAGGAAGAGGATCCAAACC	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAGA GAAATGATCAGAAGTAGGAGA
MYB 12379	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGC TCCAATCCCCAGGCATTCTTTC	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGATAG ACTTTCCCAGAAGTCACATAAATCTG
MYB 10385	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGG GCGAGCTCCGTGCTGTGATAAA	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGACAA CGTAGGCGATGATGCGGCTTG
MYB 10855	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTG TCAAACACAAGAGGAAGGAG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGAACG AAGTCCTTCTGTGGTACC
MYB 15401	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAG CTCCAAGTGGTTGAATTTGAAG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGAGTA TTGGCTGTTGCATTGGAATTG
NAC 06771	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGG TAAAGATGGTCTAGAAG	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTGGAA AATCATATCAGACG
NAC09658	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTC CGGTGGCTCAC	GGGGACCACTTTGTACAAGAAAGCTGGGTCATTGAG GGTAATACTTGTTGTTGTT
NAC 00172	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGG GAGAGAAGGACCT	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTGTGG TCTAGGTTGCAGATA
NAC 05638	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGA GTCCCTGGCGG	GGGGACCACTTTGTACAAGAAAGCTGGGTCATAGCC AGACCACAAACAGTC
NAC 08447	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGT GAAAGGAAATGCC	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGTTACA AATGATAAATTTTCATATAAAC
bHLH 11748	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGA GTTGGCAGGTGACCAGGGTTTG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGACTC ACTCAAAATGGTTGCCCCTAA
bHLH 08058	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAG AACTGAAGATACCCTTCATTTG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGATTTA GCAATCACTGTCTGAATCAT

WRKY 09595	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTT CATGATAAACCTCATGTCCAGT	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGAATCT TGAGTCAGAAACCCTTGAGG
WRKY 19284	C-terminal	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGA CCAGCACAAAATTGAGCTTTTG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGATTTT GGGAATAACTCTTGTGTATT

Table 25 – The nucleotide sequences of the primers used to clone 19 candidate TFs into pDONR221 and then subsequently into the Gateway™ GST-tag destination vectors pDEST15 (N-terminal) and pDEST24 (C-terminal).

8.12.6 DNA probes for EMSA

	Protein tested in EMSA			
	MYB10385 and MYB10855		AP2 00499	
Promoter	Name	Sequence	Name	Sequence
TASY	MYB-1-F	AAACACGTGATATGCGCCTG	AP2-1-F	CAGCATTTGATTTGGCATTT
	MYB-1-R	CAGGCGCATATCACGTGTTT	AP2-1-R	AAATGCCAAATCAAATGCTG
	MYB-2-F	AGTAATTACCTAAAATAGAA	AP2-2-F	TGCGCCTGCCGCCTGACTTA
	MYB2-R	TTCTATTTTAGGTAATTACT	AP2-2-R	TAAGTCAGGCGGCAGGCGCA
	MYB-3-F	TAAGCAATAACAAAGCACGA	AP2-3-F	ATACCTCTCAACACAAAACT
	MYB-3-R	TCGTGCTTTGTTATTGCTTA	AP2-3-R	AGTTTTGTGTTGAGAGGTAT
	MYB-4-F	TCAACATATCCATTACATTG	-	
	MYB-4-R	CAATGTAATGGATATGTTGA	-	
T5aH	-		AP2-1-F	GCAGACTTGATTTCACATAT
	-		AP2-1-R	ATATGTGAAATCAAGTCTGC
	-		AP2-2-F	GCAGGCTTGATTTCACGTGG
	-		AP2-2-R	CCACGTGAAATCAAGCCTGC
	-		AP2-3-F	CTCTCTCTGATTTTGTCTTT
	-		AP2-3-R	AAAGACAAAATCAGAGAGAG
	-		AP2-4-F	ATTTTGACAACAACTAGAGG
	-		AP2-4-R	CCTCTAGTTGTTGTCAAAAT
	-		AP2-5-F	AGAGATGATGGCCGAGAAAG
	-		AP2-5-R	CTTTCTCGGCCATCATCTCT
	-		AP2-6-F	AAAAAGGCGTGAGAAGTTAT
	-		AP2-6-R	ATAACTTCTCACGCCTTTTT
	-		AP2-7-F	CCTTTCGGCACATCACACTA
	-		AP2-7-R	TAGTGTGATGTGCCGAAAGG
	-		AP2-8-F	AGGATGCACCCCCAACTATT
	-		AP2-8-R	AATAGTTGGGGGTGCATCCT
	-		AP2-9-F	GAACTTGTCCCCCTACAATC
	-		AP2-9-R	GATTGTAGGGGGACAAGTTC
TDAT	MYB-1-F	CATCCATCCATCCATCATATC	-	
	MYB1-R	GATATGATGGATGGATGGATG	-	
	MYB-2-F	CTTATTAACTGATAATATGG	-	
	MYB-2-R	CCATATTATCAGTTAATAAG	-	
	MYB-3-F	ATAAGGAAGATAGCCTTATT	-	
	MYB-3-R	AATAAGGCTATCTTCCTTAT	-	
	MYB-4-F	TTCAATGTATCCATTGAATT	-	
	MYB-4-R	AATTCAATGGATACATTGAA	-	
	MYB-7-F	AGCTATTTTTAACAAAGAGA	-	

	MYB-7-R	TCTCTTTGTTAAAAATAGCT	-
	MYB-8-F	ATATGGAAAACTAGAAATTC	-
	MYB-8-R	GAATTTCTAGTTTTCCATAT	-
	MYB-9-F	GCTACTATTTTATATCGCAT	-
	MYB-9-R	ATGCGATATAAAATAGTAGC	-
	MYB-10-F	CATCTTTTTTGGTTCTGCCC	-
	MYB-10-R	GGGCAGAACCAAAAAAGATG	-
	MYB-11-F	TCACTTCCTTTGGTGTATTC	-
	MYB-11-R	GAATACACCAAAGGAAGTGA	-
	MYB-12-F	ACCCATTTTGGTTGGAAACA	-
	MYB-12-R	TGTTTCCAACCAAAATGGGT	-
T13αH	MYB-1-F	ATTATTCACCAACAATTGTA	-
	MYB-1-R	TACAATTGTTGGTGAATAAT	-
	MYB-2-F	CAACAGATGATAAGGGAGTG	-
	MYB2-R	CACTCCCTTATCATCTGTTG	-
	MYB-3-F	TGATTTCCCGTTATCAACAG	-
	MYB-3-R	CTGTTGATAACGGGAAATCA	-
	MYB-4-F	CTGAACACGATATAAATGCC	-
	MYB-4-R	GGCATTTATATCGTGTTCAG	-
DBBT	MYB-1-F	AATTCACAAACCATTTCTGC	AP2-1-F GGGAAGGCCGGCCCGGCATG
	MYB-1-R	GCAGAAATGGTTTGTGAATT	AP2-1-R CATGCCGGGCCGGCCTTCCC
	MYB-2-F	TTTCTGGACAGTTACATTCT	AP2-2-F AGGGGGAACGCCTTCGGCAT
	MYB-2-R	AGAATGTAACTGTCCAGAAA	AP2-2-R ATGCCGAAGGCGTTCCCCCT
	MYB-3-F	ATCAAGCTAACAGCGACGTA	-
	MYB-3-R	TACGTCGCTGTTAGCTTGAT	-
	MYB-4-F	ACGTATGATATATCTCCACA	-
	MYB-4-R	TGTGGAGATATATCATACGT	-
	MYB-5-F	TGTTAAAACCAACAACGTGA	-
	MYB-5-R	TCACGTTGTTGGTTTTAACA	-
DBAT	MYB-1-F	AAAAAAAACCAAATAATAAT	-
	MYB-1-R	ATTATTATTTGGTTTTTTTT	-
	MYB-2-F	TGTCCTCACCAAAATTAGTT	-
	MYB2-R	AACTAATTTTGGTGAGGACA	-
	MYB-3-F	AAGACACTGATATCAACATA	-
	MYB-3-R	TATGTTGATATCAGTGTCTT	-
	MYB-4-F	TGTCCTTACCAAAATTAGTT	-
	MYB-4-R	AACTAATTTTGGTAAGGACA	-
	MYB-5-F	TGTCCTTACCAAAATTAGTT	-
	MYB-5-R	AACTAATTTTGGTAAGGACA	-
	MYB-6-F	AAGACACTGATATCAACATA	-
	MYB-6-R	TATGTTGATATCAGTGTCTT	-
	MYB-7-F	TCACATTCTTAGGTGTGTGA	-
	MYB-7-R	TCACACACCTAAGAATGTGA	-
	MYB-8-F	ATTTCCTTTTAGGTATAAAC	-
	MYB-8-R	GTTTATACCTAAAAGGAAAT	-
	MYB-9-F	GTCAATTTTTTGGTTTGATT	-

	MYB-9-R	AATCAAACCAAAAAATTGAC	-	
	MYB-10-F	TTCCTTTTAGGTATAAACAA	-	
	MYB-10-R	TTGTTTATACCTAAAAGGAA	-	
	MYB-11-F	CCTTACCAAAATTAGTTGTT	-	
	MYB-11-R	AACAACTAATTTTGGTAAGG	-	
	MYB-12-F	GTCATTTTTTTGGTCTGATT	-	
	MYB-12-F	AATCAGACCAAAAAAATGAC	-	
PAM	MYB-1-F	GACGCAACGATAAATTGCAG	AP2-1-F	AAAAAATTAAATCATTGAGT
	MYB-1-R	CTGCAATTTATCGTTGCGTC	AP2-1-R	ACTCAATGATTTAATTTTTT
	MYB-2-F	CAGCATGTGATAAATGCAAC	AP2-2-F	GGGGGCCAGAACCATGTGTT
	MYB-2-R	GTTGCATTTATCACATGCTG	AP2-2-R	AACACATGGTTCTGGCCCCC
	MYB-3-F	CTACCTAAACAGACAAGA	AP2-3-F	ACGTCTCGCCCTTCCCTAGA
	MYB-3-R	TCTTGTCTGTTTAGGTAG	AP2-3-R	TCTAGGGAAGGGCGAGACGT
	-		AP2-4-F	ACGTTCCCCGGGGGCCAGAA
	-		AP2-4-R	TTCTGGCCCCCGGGGAACGT
BAPT	-		AP2-1-F	TGAATATGGGGCCTCCATCG
	-		AP2.1-R	CGATGGAGGCCCCATATTCA
	-		AP2.2-F	ATATCCCTTTCGGCATTCAC
	-		AP2.2-R	GTGAATGCCGAAAGGGATAT
	-		AP2.3-F	CTCCAATCCCCCGATGTTTG
	-		AP2.3-R	CAAACATCGGGGGATTGGAG
	-		AP2.4-F	TTCACCCTTGAAAGTCATTT
	-		AP2.4-R	AAATGACTTTCAAGGGTGAA
	-		AP2.5-F	TTCATTTCCCTTTCACTTCT
	-		AP2.5-R	AGAAGTGAAAGGGAAATGAA
	-		AP2.6-F	CACCAACATGATTTGTTTCC
	-		AP2.6-R	GGAAACAAATCATGTTGGTG
	-		AP2.7-F	CTGAATCAAATCAACAATAA
	-		AP2.7-R	TTATTGTTGATTTGATTCAG
	-		AP2.8-F	TACCAAAATCATCTAGATAT
	-		AP2.8-R	ATATCTAGATGATTTTGGTA
DBTNBT	MYB-1-F	TCCACTTAGATATTAAAAAA	-	
	MYB-1-R	TTTTTTAATATCTAAGTGGA	-	
	MYB-2-F	TTAACAAACCAAAACTATAA	-	
	MYB-2-R	TTATAGTTTTGGTTTGTTAA	-	
	MYB-3-F	AAAAAAATGATATCATTCTT	-	
	MYB-3-R	AAGAATGATATCATTTTTTT	-	
	MYB-4-F	TTTTTATGGATAGCTGTAAT	-	
	MYB-4-R	ATTACAGCTATCCATAAAAA	-	
	MYB-5-F	GCCTTTTACCAACCATATTT	-	
	MYB-5-R	AAATATGGTTGGTAAAAGGC	-	

Table 26 – The nucleotide sequence of DNA probes used in EMSA analysis of MYB TFs, MYB10385 & MYB10855 and AP2 00499.

8.11.7 Gateway™ cloning of JAZ proteins

TF	Forward	Reverse
JAZ11289	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAGGA GATAGAACCATGCAGTGGCAACAGCTCATGTCT	GGGGACCACTTTGTACAAGAAAGCTGGGTCTC ACTTGTCAGGTGGACATCCATCA
JAZ 02148	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAGGA GATAGAACCATGGAAGAACATACTCAGTGGAGC	GGGGACCACTTTGTACAAGAAAGCTGGGTCTT AATATTGGTGACGAGTGGAATAT
JAZ35964	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAGGA GATAGAACCATGGAAGCGTTGAGCATTTACA	GGGGACCACTTTGTACAAGAAAGCTGGGTCTT AACAAGATCTTTGTGACGGAAAA

Table 27 – The nucleotide sequences of the primers used to clone three JAZ proteins into pDONR221 and then subsequently into the Gateway™ destination vectors pDEST33.

8.11.8 Additional cloning primers

Name	Sequence
Rluc-SpeI	GGACTAGTATGATCCAGAACAAAGGAAACG
Rluc SacII	TCCCCGCGGTTATTGTTCATT
Rluc *att*B – F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAGGAGATAGAACCATGATCCA GAACAAAGGAAAC
Rluc *att*B – R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTGTTCATTTTTGAGAACTCG
***att*B F**	ACAAGTTTGTACAAAAAAGCAG
***att*B R**	ACCACTTTGTACAAGAAAGCT
M13 F	GTAAAACGACGGCCAG
M13 R	CAGGAAACAGCTATGAC

Table 28 –Primers used during cloning of Renilla luciferase into p2GW7,0 with restriction enzymes & Gateway™ cloning and sequencing primers M13 for pDONR221 and att*B* for destination vectors.

Bibliography

1. Oksman-Caldentey KM, Inzé D (2004) Plant cell factories in the post-genomic era: New ways to produce designer secondary metabolites. *Trends Plant Sci* 9(9):433–440.
2. Staniek A, et al. (2013) Natural products - modifying metabolite pathways in plants. *Biotechnol J* 8(10):1159–1171.
3. Leete E (1969) Biosynthesis of quinine and related alkaloids. *Acc Chem Res* 2(2):59–64.
4. Rates SM (2001) Plants as source of drugs. *Toxicon* 39(5):603–13.
5. Newman D, Cragg G (2012) Natural products as sources of new drugs over the period 1981–2010. *J Nat Prod* 75:311–335.
6. Wilson S a., Roberts SC (2014) Metabolic engineering approaches for production of biochemicals in food and medicinal plants. *Curr Opin Biotechnol* 26:174–182.
7. Cragg GM (1998) Paclitaxel (Taxol): A success story with valuable lessons for natural product drug discovery and development. *Med Res Rev* 18(5):315–331.
8. Wani MC, Taylor HL, Wall ME, Coggon P, McPhail a T (1971) Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. *J Am Chem Soc* 93(9):2325–2327.
9. Schiff PB, Fant J, Horwitz SB (1979) Promotion of microtubule assembly in vitro by taxol. *Nature* 277(665-667).
10. Parness J, Horwitz SB (1981) Taxol binds to polymerized tubulin in vitro. *J Cell Biol* 91(2 I):479–487.
11. Malik S, et al. (2011) Production of the anticancer drug taxol in Taxus baccata suspension cultures: A review. *Process Biochem* 46(1):23–34.
12. Ochoa-Villarreal M, et al. (2015) Cambial meristematic cells: a platform for the production of plant natural products. *N Biotechnol*. doi:10.1016/j.nbt.2015.02.003.
13. Stone GW, et al. (2004) A Polymer-Based, Paclitaxel-Eluting Stent in Patients with Coronary Artery Disease. *N Engl J Med* 350(3):221–231.
14. Shemesh OA, Spira ME (2011) Rescue of neurons from undergoing hallmark tau-induced Alzheimer's disease cell pathologies by the antimitotic drug paclitaxel. *Neurobiol Dis* 43(1):163–75.
15. Tabata H (2006) Production of paclitaxel and the related taxanes by cell suspension cultures of Taxus species. *Curr Drug Targets* 7(4):453–461.
16. Stierle A, Strobel G, Stierle D (1993) Taxol and taxane production by Taxomyces andreanae, an endophytic fungus of Pacific yew. *Science* 260(5105):214–216.
17. Miele M, Mumot AM, Zappa A, Romano P, Ottaggio L (2012) Hazel and other sources of paclitaxel and related compounds. *Phytochem Rev* 11(2-3):211–225.
18. Schiff PB, Horwitz SB (1980) Taxol stabilizes microtubules. 77(3):1561–1565.
19. Leone LM, Roberts SC (2013) Acessing Anti-cancer Natural Products by Plant Cell Culture. *Natural Products and Cancer Drug Discovery*, pp 193–221.
20. Band Horwitz S (1992) Mechanism of action of taxol. *Trends Pharmacol Sci* 13:134–136.
21. Fang W-S, Liang X-T (2005) Recent progress in structure activity relationship and mechanistic studies of taxol analogues. *Mini Rev Med Chem* 5(1):1–12.
22. Kingston DGI (2007) The shape of things to come: Structural and synthetic studies of taxol and related compounds. *Phytochemistry* 68(14):1844–1854.
23. Kingston D (2000) Recent Advances in the Chemistry of Taxol. *J Nat Prod* 63(5):726–734.
24. Ojima I, et al. (1994) Synthesis and Structure-Activity Relationships of Novel Nor-Seco Analogs of Taxol and Taxotere Summary : Novel nor-seco analogs of taxol and taxotere are synthesized from 14-hydroxy-10-deacetylbaccatin III through periodic acid oxidation and NaBHsCN reduc. (18):515–517.
25. Expósito O, et al. (2009) Biotechnological production of taxol and related taxoids: current state and prospects. *Anticancer Agents Med Chem* 9(1):109–121.
26. Howat S, et al. (2014) Paclitaxel: Biosynthesis, production and future prospects. *N Biotechnol* 31(3):242–245.
27. Roberts SC (2007) Production and engineering of terpenoids in plant cell culture. *Nat Chem Biol* 3(7):387–395.
28. Rodríguez-Concepción M (2006) Early steps in isoprenoid biosynthesis: Multilevel regulation of the supply of common precursors in plant cells. *Phytochem Rev* 5(1):1–15.
29. Expósito O, et al. (2009) Effect of taxol feeding on taxol and related taxane production in Taxus baccata suspension cultures. *N Biotechnol* 25(4):252–9.
30. Srinivasan V, Ciddi V, Bringi V, Shuler ML (1996) Metabolic inhibitors, elicitors, and precursors

as tools for probing yield limitation in taxane production by Taxus chinensis cell cultures. *Biotechnol Prog* 12(4):457–65.
31. Eisenreich W, Menhard B, Hylands PJ, Zenk MH, Bacher A (1996) Studies on the biosynthesis of taxol: the taxane carbon skeleton is not of mevalonoid origin. *Proc Natl Acad Sci* 93(13):6431–6436.
32. Cusidó RM, et al. (2007) Source of isopentenyl diphosphate for taxol and baccatin III biosynthesis in cell cultures of Taxus baccata. *Biochem Eng J* 33(2):159–167.
33. Hefner J, Ketchum REB, Croteau R (1998) Cloning and Functional Expression of a cDNA Encoding Geranylgeranyl Diphosphate Synthase from Taxus canadensis and Assessment of the Role of this Prenyltransferase in Cells Induced for Taxol Production. *Arch Biochem Biophys* 360(1):62–74.
34. Koepp AE, et al. (1995) Cyclization of Geranylgeranyl Diphosphate to Taxa-4 (5), 11 (12) -diene Is the Committed Step of Taxol Biosynthesis in Pacific Yew. *J Biol Chem* 270(15):8686–8690.
35. Hefner J, et al. (1996) Cytochrome P450-catalyzed hydroxylation of taxa-4(5),11(12)-diene to taxa-4(20),11(12)-dien-5alpha-ol: the first oxygenation step in taxol biosynthesis. *Chem Biol* 3(6):479–489.
36. Walker K, Schoendorf A, Croteau R (2000) Molecular cloning of a taxa-4(20),11(12)-dien-5alpha-ol-O-acetyl transferase cDNA from Taxus and functional expression in Escherichia coli. *Arch Biochem Biophys* 374(2):371–80.
37. Jennewein S, Rithner CD, Williams RM, Croteau RB (2001) Taxol biosynthesis : Taxane 13α-hydroxylase is a cytochrome P450-dependent monooxygenase. 98(24):13595–13600.
38. Schoendorf A, Rithner CD, Williams RM, Croteau RB (2001) Molecular cloning of a cytochrome P450 taxane 10β-hydroxylase cDNA from Taxus and functional expression in yeast. *PNAS* 98(4):10–15.
39. Jennewein S, Rithner CD, Williams RM, Croteau R (2003) Taxoid metabolism: Taxoid 14β-hydroxylase is a cytochrome P450-dependent monooxygenase. *Arch Biochem Biophys* 413(2):262–270.
40. Chau M, Croteau R (2004) Molecular cloning and characterization of a cytochrome P450 taxoid 2α-hydroxylase involved in Taxol biosynthesis. *Arch Biochem Biophys* 427(1):48–57.
41. Walker K, Croteau R (2000) Taxol biosynthesis : Molecular cloning of a benzoyl-CoA:taxane 2α-O-benzoyltransferase cDNA from Taxus and functional expression in Escherichia coli. 90(5):6–11.
42. Walker K, Croteau R (2000) Molecular cloning of a 10-deacetylbaccatin III-10-O-acetyl transferase cDNA from Taxus and functional expression in Escherichia coli. *PNAS* 97:583–587.
43. Walker KD, Klettke K, Akiyama T, Croteau R (2004) Cloning, heterologous expression, and characterization of a phenylalanine aminomutase involved in taxol biosynthesis. *J Biol Chem* 279(52):53947–53954.
44. Ramírez-Estrada K, et al. (2015) Transcript profiling of jasmonate-elicited Taxus cells reveals a β-phenylalanine-CoA ligase. *Plant Biotechnol J* Available . doi:10.1111/pbi.12359.
45. Walker K, Fujisaki S, Long R, Croteau R (2002) Molecular cloning and heterologous expression of the C-13 phenylpropanoid side chain-CoA acyltransferase that functions in Taxol biosynthesis. *Proc Natl Acad Sci U S A* 99(20):12715–20.
46. Walker K, Long R, Croteau R (2002) The final acylation step in taxol biosynthesis: cloning of the taxoid C13-side-chain N-benzoyltransferase from Taxus. *Proc Natl Acad Sci U S A* 99(14):9166–9171.
47. Onrubia M, et al. (2013) Bioprocessing of plant in vitro systems for the mass production of pharmaceutically important metabolites: paclitaxel and its derivatives. *Curr Med Chem* 20(7):880–91.
48. Guerra-Bubb J, Croteau R, Williams RM (2012) The early stages of taxol biosynthesis: An interim report on the synthesis and identification of early pathway metabolites. *Nat Prod Rep* 29(6):683.
49. Jennewein S, Croteau R (2001) Taxol: Biosynthesis, molecular genetics, and biotechnological applications. *Appl Microbiol Biotechnol* 57(1-2):13–19.
50. Walker K, Croteau R (2001) Taxol biosynthetic genes. *Phytochemistry* 58(1):1–7.
51. Heinig U, Jennewein S (2009) Taxol : A complex diterpenoid natural product with an evolutionarily obscure origin. *African J Biotechnol* 8(8):1370–1385.
52. Shigemori H, Kobayashi J (2004) Biological Activity and Chemistry of Taxoids from the Japanese Yew, Taxus cuspidata. *J Nat Prod* 67(2):245–256.
53. United states of America (1992) *Public Law 102-404 102d Congress An Act to provide for the management of federal lands containing the Pacific yew to ensure a sufficient supply of taxol, a*

cancer-treating drug made from the Pacific yew.
54. Patel RN (1998) Tour de paclitaxel: biocatalysis for semisynthesis. *Annu Rev Microbiol* 52:361–395.
55. Colin M, Guenard D, Gueritte-Voegelein F, Potier P (1989) Taxol derivatives, their preparation and pharmaceutical compositions containing them. Available at: https://www.google.com/patents/US4814470 [Accessed July 8, 2015].
56. Christen A, Gibson DM, Bland J (1991) Production of taxol or taxol-like compounds in cell culture. Available at: http://www.google.co.uk/patents/US5019504 [Accessed July 14, 2015].
57. Holton R a, et al. (1994) First Total Synthesis of Taxol. 1. Functionalization of the B Ring. *J Am Chem Soc* 116(2):1597–1598.
58. Holton R a, et al. (1994) First Total Synthesis of Taxol. 2. Completion of the C and D Rings Robert. *J Am Chem Soc* 116(10):1599–1600.
59. Nicolaou KC, et al. (1994) Total synthesis of taxol. *Nature* 367(6464):630–634.
60. Nicolaou KC, et al. (1995) Total synthesis of Taxol. 1. Retrosynthesis, degradation, and reconstitution. *J Am Chem Soc* 117(2):624–633.
61. Nicolaou KC, et al. (1995) Total synthesis of Taxol. 4. The final stages and completion of the synthesis. *J Am Chem Soc* 117(2):653–659.
62. Holton RA (1993) US 5254703 A- Semi-synthesis of taxane derivatives using metal alkoxides and oxazinones. Available at: http://www.google.com/patents/US5254703 [Accessed July 15, 2015].
63. Bringi V, Kakrade PG, Prince CL, Roach BL (2005) EP001538214A1- Enhanced production of taxanes by cell cultures of taxus species.
64. Lee E-K, et al. (2010) Cultured cambial meristematic cells as a source of plant natural products. *Nat Biotechnol* 28(11):1213–1217.
65. Guo BH, Kai GY, Jin HB, Tang KX (2006) Taxol synthesis. *African J Biotechnol* 5(1):15–20.
66. Holton RA (1989) Method for preparation of taxol using β-lactam. Available at: http://www.freepatentsonline.com/5175315.html [Accessed July 14, 2015].
67. Eisenstein RI, Resnick DS (2001) Going for the big one. *Nat Biotechnol* 19(9):881–2.
68. Ro D-K, et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440(7086):940–943.
69. Yang J, et al. (2012) Bio-isoprene production using exogenous MVA pathway and isoprene synthase in Escherichia coli. *Bioresour Technol* 104(OCTOBER 2011):642–647.
70. Santos CNS, Koffas M, Stephanopoulos G (2011) Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab Eng* 13(4):392–400.
71. Marienhagen J, Bott M (2013) Metabolic engineering of microorganisms for the synthesis of plant natural products. *J Biotechnol* 163(2):166–178.
72. Huang Q, Roessner C a., Croteau R, Scott a. I (2001) Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. *Bioorganic Med Chem* 9(9):2237–2242.
73. Wilson S a., Roberts SC (2012) Recent advances towards development and commercialization of plant cell culture processes for the synthesis of biomolecules. *Plant Biotechnol J* 10(3):249–268.
74. DeJong JM, et al. (2006) Genetic engineering of taxol biosynthetic genes in Saccharomyces cerevisiae. *Biotechnol Bioeng* 93(2):212–224.
75. Jennewein S, et al. (2005) Coexpression in yeast of Taxus cytochrome P450 reductase with cytochrome p450 oxygenases involved in taxol biosynthesis. *Biotechnol Bioeng* 89(5):588–598.
76. Besumbes Ó, et al. (2004) Metabolic engineering of isoprenoid biosynthesis in Arabidopsis for the production of taxadiene, the first committed precursor of taxol. *Biotechnol Bioeng* 88(2):168–175.
77. Kovacs K, et al. (2007) Redirection of carotenoid metabolism for the efficient production of taxadiene [taxa-4(5),11(12)-diene] in transgenic tomato fruit. *Transgenic Res* 16(1):121–126.
78. Anterola A, Shanle E, Perroud PF, Quatrano R (2009) Production of taxa-4(5),11(12)-diene by transgenic Physcomitrella patens. *Transgenic Res* 18(4):655–660.
79. Smetanska I (2008) Production of secondary metabolites using plant cell cultures. *Adv Biochem Eng Biotechnol* 111:187–228.
80. Kolewe ME, Gaurav V, Roberts SC (2008) Pharmaceutically active natural product synthesis and supply via plant cell culture technology. *Mol Pharm* 5(2):243–256.
81. Frense D (2007) Taxanes: Perspectives for biotechnological production. *Appl Microbiol Biotechnol* 73(6):1233–1240.
82. Zhong J-J (2002) Plant cell culture for production of paclitaxel and other taxanes. *J Biosci Bioeng*

94(6):591–599.
83. Namdeo (2007) Plant cell elicitation for production of secondary metabolites: A review. *Pharmacogn Rev* 1(1):69.
84. Zhao J, Davis LC, Verpoorte R (2005) Elicitor signal transduction leading to production of plant secondary metabolites. *Biotechnol Adv* 23(4):283–333.
85. Pauwels L, Inzé D, Goossens A (2009) Jasmonate-inducible gene: what does it mean? *Trends Plant Sci* 14(2):87–91.
86. Onrubia M, et al. (2010) An approach to the molecular mechanism of methyl jasmonate and vanadyl sulphate elicitation in Taxus baccata cell cultures: The role of txs and bapt gene expression. *Biochem Eng J* 53(1):104–111.
87. Wu J, Wang C, Mei X (2001) Stimulation of taxol production and excretion in Taxus spp cell cultures by rare earth chemical lanthanum. *J Biotechnol* 85(1):67–73.
88. Li Y-C, Tao W-Y (2009) Paclitaxel-producing fungal endophyte stimulates the accumulation of taxoids in suspension cultures of Taxus cuspidate. *Sci Hortic (Amsterdam)* 121(1):97–102.
89. Wang Y-D, Yuan Y-J, Wu J-C (2004) Induction studies of methyl jasmonate and salicylic acid on taxane production in suspension cultures of Taxus chinensis var. mairei. *Biochem Eng J* 19(3):259–265.
90. Mirjalili N, Linden JC (1996) Methyl jasmonate induced production of taxol in suspension cultures of Taxus cuspidata: ethylene interaction and induction models. *Biotechnol Prog* 12(1):110–8.
91. Khosroushahi AY, et al. (2006) Improved Taxol production by combination of inducing factors in suspension cell culture of Taxus baccata. *Cell Biol Int* 30(3):262–269.
92. Yukimune Y, Tabata H, Higashi Y, Hara Y (1996) Methyl jasmonate-induced overproduction of paclitaxel and baccatin III in Taxus cell suspension cultures. *Nat Biotechnol* 14(9):1129–32.
93. Bentebibel S, et al. (2005) Effects of immobilization by entrapment in alginate and scale-up on paclitaxel and baccatin III production in cell suspension cultures of Taxus baccata. *Biotechnol Bioeng* 89(6):647–55.
94. Onrubia M, et al. (2013) Coronatine, a more powerful elicitor for inducing taxane biosynthesis in Taxus media cell cultures than methyl jasmonate. *J Plant Physiol* 170(2):211–219.
95. Roberts S, Kolewe M (2010) Plant natural products from cultured multipotent cells. *Nat Biotechnol* 28(11):1175–1176.
96. Baebler Š, et al. (2005) Establishment of cell suspension cultures of yew (Taxus × media Rehd. and assessment of their genomic stability. *Vitr Cell Dev Biol - Plant* 41(3):338–343.
97. Naill MC, Roberts SC (2005) Cell cycle analysis of Taxus suspension cultures at the single cell level as an indicator of culture heterogeneity. *Biotechnol Bioeng* 90(4):491–500.
98. Kombrink E (2012) Chemical and genetic exploration of jasmonate biosynthesis and signaling paths. *Planta* 236(5):1351–1366.
99. Wasternack C, Hause B (2013) Jasmonates: Biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Ann Bot* 111(6):1021–1058.
100. De Geyter N, Gholami A, Goormachtig S, Goossens A (2012) Transcriptional machineries in jasmonate-elicited plant secondary metabolism. *Trends Plant Sci* 17(6):349–359.
101. Wasternack C (2007) Jasmonates: An update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Ann Bot* 100(4):681–697.
102. Turner JG, Ellis C, Devoto A (2002) The jasmonate signal pathway. *Plant Cell* 14 Suppl:S153–S164.
103. Vick B a, Zimmerman DC (1984) Biosynthesis of jasmonic Acid by several plant species. *Plant Physiol* 75(2):458–461.
104. Sheard LB, et al. (2010) Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature* 468(7322):400–405.
105. Pauwels L, Goossens A (2011) The JAZ proteins: a crucial interface in the jasmonate signaling cascade. *Plant Cell* 23(9):3089–3100.
106. Pauwels L, et al. (2010) NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature* 464(7289):788–91.
107. Chini A, et al. (2007) The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* 448(7154):666–671.
108. Thines B, et al. (2007) JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. *Nature* 448(7154):661–665.
109. Memelink J (2009) Regulation of gene expression by jasmonate hormones. *Phytochemistry* 70(13-14):1560–1570.
110. Szemenyei H, Hannon M, Long JA (2008) TOPLESS mediates auxin-dependent transcriptional

repression during Arabidopsis embryogenesis. *Science* 319(5868):1384–6.
111. Hirsch S, Oldroyd GED (2009) GRAS-domain transcription factors that regulate plant development. *Plant Signal Behav* 4(8):698–700.
112. Goossens A, et al. (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proc Natl Acad Sci U S A* 100(14):8595–8600.
113. Rischer H, et al. (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in Catharanthus roseus cells. *Proc Natl Acad Sci U S A* 103(14):5614–5619.
114. Pauwels L, et al. (2008) Mapping methyl jasmonate-mediated transcriptional reprogramming of metabolism and cell cycle progression in cultured Arabidopsis cells. *Proc Natl Acad Sci U S A* 105:1380–1385.
115. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim Biophys Acta - Gene Regul Mech* 1819(2):86–96.
116. Licausi F, Ohme-Takagi M, Perata P (2013) APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: Mediators of stress responses and developmental programs. *New Phytol*:639–649.
117. Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M (1998) A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J* 17(18):5484–96.
118. Thirugnanasambantham K, et al. (2014) Role of Ethylene Response Transcription Factor (ERF) and Its Regulation in Response to Stress Encountered by Plants. *Plant Mol Biol Report*:347–357.
119. Tiwari SB, et al. (2012) The EDLL motif: A potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J* 70(5):855–865.
120. Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M (2001) Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *Plant Cell* 13(8):1959–1968.
121. Klempnauer K-H, Gonda TJ, Michael Bishop J (1982) Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: The architecture of a transduced oncogene. *Cell* 31(2):453–463.
122. Feller A, MacHemer K, Braun EL, Grotewold E (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J* 66(1):94–116.
123. Prouse MB, Campbell MM (2012) The interaction between MYB proteins and their target DNA binding sites. *Biochim Biophys Acta - Gene Regul Mech* 1819(1):67–77.
124. Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H (1987) The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J* 6(12):3553–8.
125. Dubos C, et al. (2010) MYB transcription factors in Arabidopsis. *Trends Plant Sci* 15(10):573–581.
126. Ogata K, et al. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* 79(4):639–648.
127. Du H, et al. (2009) Biochemical and molecular characterization of plant MYB transcription factor family. *Biochemistry (Mosc)* 74(1):1–11.
128. Ohme-Takagi M, Shinshi H (1995) Ethylene-Inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell* 7:173–182.
129. Fujimoto SY, Ohta M, Usui a, Shinshi H, Ohme-Takagi M (2000) Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell* 12(3):393–404.
130. Gu Y-Q, et al. (2002) Tomato transcription factors pti4, pti5, and pti6 activate defense responses when expressed in Arabidopsis. *Plant Cell* 14:817–831.
131. Zarei A, et al. (2011) Two GCC boxes and AP2/ERF-domain transcription factor ORA59 in jasmonate/ethylene-mediated activation of the PDF1.2 promoter in Arabidopsis. *Plant Mol Biol* 75:321–331.
132. Sakuma Y, et al. (2002) DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem Biophys Res Commun* 290(3):998–1009.
133. Van der Fits L, Memelink J (2001) The jasmonate-inducible AP2/ERF-domain transcription factor ORCA3 activates gene expression via interaction with a jasmonate-responsive promoter element. *Plant J* 25(1):43–53.
134. Kagaya Y, Ohmiya K, Hattori T (1999) RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants.

Nucleic Acids Res 27(2):470–478.
135. Koshino-Kimura Y, et al. (2005) Regulation of CAPRICE transcription by MYB proteins for root epidermis differentiation in Arabidopsis. *Plant Cell Physiol* 46(6):817–826.
136. Patzlaff A, et al. (2003) Characterisation of Pt MYB1, an R2R3-MYB from pine xylem. *Plant Mol Biol* 53(4):597–608.
137. Patzlaff A, et al. (2003) Characterisation of a pine MYB that regulates lignification. *Plant J* 36(6):743–754.
138. Grotewold E, Drummond BJ, Bowen B, Peterson T (1994) The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. *Cell* 76(3):543–553.
139. Hwang MG, Chung IK, Kang BG, Cho MH (2001) Sequence-specific binding property of Arabidopsis thaliana telomeric DNA binding protein 1 (AtTBP1). *FEBS Lett* 503(1):35–40.
140. Schaffer R, et al. (1998) The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* 93(7):1219–29.
141. Baranowskij N, Frohberg C, Prat S, Willmitzer L (1994) A novel DNA binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator. *EMBO J* 13(22):5383–92.
142. Olsen AN, Ernst H a., Leggio LL, Skriver K (2005) DNA-binding specificity and molecular functions of NAC transcription factors. *Plant Sci* 169(4):785–797.
143. Tran L-SP, et al. (2004) Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell* 16(9):2481–2498.
144. Kong Q, et al. (2012) PNAS Plus: Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. *Proc Natl Acad Sci* 109(30):E2091–E2097.
145. Zhang HB, Bokowiec MT, Rushton PJ, Han SC, Timko MP (2012) Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple jasmonate-inducible steps in nicotine biosynthesis. *Mol Plant* 5(1):73–84.
146. Ciolkowski I, Wanke D, Birkenbihl RP, Somssich IE (2008) Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. *Plant Mol Biol* 68(1-2):81–92.
147. Suttipanta N, et al. (2011) The Transcription Factor CrWRKY1 Positively Regulates the Terpenoid Indole Alkaloid Biosynthesis in Catharanthus roseus. *Plant Physiol* 157(4):2081–2093.
148. Solano R, Fuertes A, Sánchez-Pulido L, Valencia A, Paz-Ares J (1997) A single residue substitution causes a switch from the dual DNA binding specificity of plant transcription factor MYB.Ph3 to the animal c-MYB specificity. *J Biol Chem* 272(5):2889–2895.
149. Jensen MK, Skriver K (2014) NAC transcription factor gene regulatory and protein-protein interaction networks in plant stress responses and senescence. *IUBMB Life* (13):156–166.
150. Olsen AN, Ernst H a., Leggio LL, Skriver K (2005) NAC transcription factors: Structurally distinct, functionally diverse. *Trends Plant Sci* 10(2):79–87.
151. Ernst HA, Olsen AN, Larsen S, Lo Leggio L (2004) Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. *EMBO Rep* 5(3):297–303.
152. Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) NAC transcription factors in plant abiotic stress responses. *Biochim Biophys Acta - Gene Regul Mech* 1819(2):97–103.
153. Hao YJ, et al. (2010) Plant NAC-type transcription factor proteins contain a NARD domain for repression of transcriptional activation. *Planta* 232(5):1033–1043.
154. Pires N, Dolan L (2010) Early evolution of bHLH proteins in plants. *Plant Signal Behav* 5(7):911–912.
155. Toledo-Ortiz G, Huq E, Quail PH (2003) The Arabidopsis Basic / Helix-Loop-Helix Transcription Factor Family. *Plant Cell* 15(August):1749–1770.
156. Heim M a., et al. (2003) The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. *Mol Biol Evol* 20(5):735–747.
157. Chen L, et al. (2012) The role of WRKY transcription factors in plant abiotic stresses. *Biochim Biophys Acta* 1819(2):120–8.
158. Rushton PJ, Somssich IE, Ringler P, Shen QJ (2010) WRKY transcription factors. *Trends Plant Sci* 15(5):247–58.
159. Yamasaki K, et al. (2005) Solution structure of an Arabidopsis WRKY DNA binding domain. *Plant Cell* 17(3):944–956.

160. Duan M-R, et al. (2007) DNA binding mechanism revealed by high resolution crystal structure of Arabidopsis thaliana WRKY1 protein. *Nucleic Acids Res* 35(4):1145–54.
161. Afrin S, Luo JHZ (2015) JA-mediated transcriptional regulation of secondary metabolism in medicinal plants. *Sci Bull* 60(12):1062–1072.
162. Dombrecht B, et al. (2007) MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. *Plant Cell* 19(7):2225–45.
163. Qi T, et al. (2011) The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in Arabidopsis thaliana. *Plant Cell* 23(5):1795–814.
164. Hirai MY, et al. (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc Natl Acad Sci U S A* 104(15):6478–83.
165. Mao G, et al. (2011) Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in Arabidopsis. *Plant Cell* 23(4):1639–53.
166. Saga H, et al. (2012) Identification and characterization of ANAC042, a transcription factor family gene involved in the regulation of camalexin biosynthesis in Arabidopsis. *Mol Plant Microbe Interact* 25(5):684–96.
167. Gális I, et al. (2006) A novel R2R3 MYB transcription factor NtMYBJS1 is a methyl jasmonate-dependent regulator of phenylpropanoid-conjugate biosynthesis in tobacco. *Plant J* 46(4):573–592.
168. Todd AT, Liu E, Polvi SL, Pammett RT, Page JE (2010) A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in Nicotiana benthamiana. *Plant J* 62(4):589–600.
169. Shoji T, Kajikawa M, Hashimoto T (2010) Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *Plant Cell* 22(10):3390–3409.
170. De Sutter V, et al. (2005) Exploration of jasmonate signalling via automated and standardized transient expression assays in tobacco cells. *Plant J* 44:1065–1076.
171. Onkokesung N, et al. (2012) MYB8 controls inducible phenolamide levels by activating three novel hydroxycinnamoyl-coenzyme A:polyamine transferases in Nicotiana attenuata. *Plant Physiol* 158(1):389–407.
172. Skibbe M, Qu N, Galis I, Baldwin IT (2008) Induced plant defenses in the natural environment: Nicotiana attenuata WRKY3 and WRKY6 coordinate responses to herbivory. *Plant Cell* 20(7):1984–2000.
173. Chatel G, et al. (2003) CrMYC1, a Catharanthus roseus elicitor- and jasmonate-responsive bHLH transcription factor that binds the G-box element of the strictosidine synthase gene promoter. *J Exp Bot* 54(392):2587–2588.
174. Zhang H, et al. (2011) The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in Catharanthus roseus. *Plant J* 67(1):61–71.
175. Li CY, et al. (2013) The ORCA2 transcription factor plays a key role in regulation of the terpenoid indole alkaloid pathway. *BMC Plant Biol* 13(1):155.
176. van der Fits L, Memelink J (2000) ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science* 289(5477):295–297.
177. Pauw B, et al. (2004) Zinc finger proteins act as transcriptional repressors of alkaloid biosynthesis genes in Catharanthus roseus. *J Biol Chem* 279(51):52940–8.
178. Xu Y-H, Wang J-W, Wang S, Wang J-Y, Chen X-Y (2004) Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-delta-cadinene synthase-A. *Plant Physiol* 135(1):507–15.
179. Yu Z-X, et al. (2012) The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in Artemisia annua L. *Mol Plant* 5(2):353–65.
180. Ma D, et al. (2009) Isolation and characterization of AaWRKY1, an Artemisia annua transcription factor that regulates the amorpha-4,11-diene synthase gene, a key gene of artemisinin biosynthesis. *Plant Cell Physiol* 50(12):2146–61.
181. Bedon F, Grima-Pettenati J, Mackay J (2007) Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (Picea glauca). *BMC Plant Biol* 7:17.
182. Bedon F, et al. (2010) Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses. *J Exp Bot* 61(14):3847–64.
183. Li S, Zhang P, Zhang M, Fu C, Yu L (2013) Functional analysis of a WRKY transcription factor involved in transcriptional activation of the DBAT gene in Taxus chinensis. *Plant Biol* 15(1):19–

26.
184. Lenka SK, et al. (2015) Jasmonate-responsive expression of paclitaxel biosynthesis genes in Taxus cuspidata cultured cells is negatively regulated by the bHLH transcription factors TcJAMYC1, TcJAMYC2, and TcJAMYC4. *Front Plant Sci* 6(Article115):1–13.
185. De Boer K, et al. (2011) APETALA2/ETHYLENE RESPONSE FACTOR and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. *Plant J* 66(6):1053–1065.
186. Vom Endt D, Kijne JW, Memelink J (2002) Transcription factors controlling plant secondary metabolism: What regulates the regulators? *Phytochemistry* 61(2):107–114.
187. Gu YQ, Yang C, Thara VK, Zhou J, Martin GB (2000) Pti4 is induced by ethylene and salicylic acid, and its product is phosphorylated by the Pto kinase. *Plant Cell* 12(5):771–86.
188. Ishihama N, Yoshioka H (2012) Post-translational regulation of WRKY transcription factors in plant immunity. *Curr Opin Plant Biol* 15(4):431–437.
189. Heine GF, Hernandez JM, Grotewold E (2004) Two cysteines in plant R2R3 MYB domains participate in REDOX-dependent DNA binding. *J Biol Chem* 279(36):37878–37885.
190. Boyle P, Després C (2010) Dual-function transcription factors and their entourage: unique and unifying themes governing two pathogenesis-related genes. *Plant Signal Behav* 5(6):629–634.
191. Zhang Y, Butelli E, Martin C (2014) Engineering anthocyanin biosynthesis in plants. *Curr Opin Plant Biol* 19:81–90.
192. Muir SR, et al. (2001) Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. *Nat Biotechnol* 19(5):470–474.
193. Nishihara M, Nakatsuka T (2011) Genetic engineering of flavonoid pigments to modify flower color in floricultural plants. *Biotechnol Lett* 33(3):433–441.
194. Nishihara M, Nakatsuka T, Yamamura S (2005) Flavonoid components and flower color change in transgenic tobacco plants by suppression of chalcone isomerase gene. *FEBS Lett* 579(27):6074–8.
195. Beyer P, et al. (2002) Golden Rice: introducing the beta-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. *J Nutr* 132(3):506S–510S.
196. Ye X, et al. (2000) Engineering the Provitamin A (B-Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm. *Science (80-)* 287(5451):303–305.
197. Paine J, et al. (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat Biotechnol* 23(4):482–487.
198. Tang G, et al. (2012) β-Carotene in Golden Rice is as good as β-carotene in oil at providing vitamin A to children. *Am J Clin Nutr* 96(3):658–64.
199. Tang G, Qin J, Dolnikowski GG, Russell RM, Grusak MA (2009) Golden Rice is an effective source of vitamin A. *Am J Clin Nutr* 89(6):1776–83.
200. Winter JM, Tang Y (2012) Synthetic biological approaches to natural product biosynthesis. *Curr Opin Biotechnol* 23(5):736–743.
201. Runguphan W, Qu X, O'Connor SE (2010) Integrating carbon-halogen bond formation into medicinal plant metabolism. *Nature* 468(7322):461–464.
202. Broun P (2004) Transcription factors as tools for metabolic engineering in plants. *Curr Opin Plant Biol* 7(2):202–209.
203. Moses T, Pollier J (2013) Bioengineering of plant (tri) terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro. *New Phytol* 200:27–43.
204. Butelli E, et al. (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat Biotechnol* 26(11):1301–1308.
205. Lloyd AM, Walbot V, Davis RW (1992) Arabidopsis and Nicotiana anthocyanin production activated by maize regulators R and C1. *Science* 258(5089):1773–5.
206. Yu O, et al. (2003) Metabolic engineering to increase isoflavone biosynthesis in soybean seed. *Phytochemistry* 63(7):753–763.
207. Bräutigam a., Gowik U (2010) What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biol* 12(6):831–841.
208. Yamazaki M, et al. (2013) Coupling deep transcriptome analysis with untargeted metabolic profiling in ophiorrhiza pumila to further the understanding of the biosynthesis of the anti-cancer alkaloid camptothecin and anthraquinones. *Plant Cell Physiol* 54(5):686–696.
209. Spyropoulou E a, Haring M a, Schuurink RC (2014) RNA sequencing on Solanum lycopersicum trichomes identifies transcription factors that activate terpene synthase promoters. *BMC Genomics* 15(1):402.
210. Li S, et al. (2012) Transcriptional profile of Taxus chinensis cells in response to methyl jasmonate. *BMC Genomics* 13(1):295.

211. Qiu D, et al. (2009) High throughput sequencing technology reveals that the taxoid elicitor methyl jasmonate regulates microRNA expression in Chinese yew (Taxus chinensis). *Gene* 436(1-2):37–44.
212. Jennewein S, Wildung MR, Chau M, Walker K, Croteau R (2004) Random sequencing of an induced Taxus cell cDNA library for identification of clones involved in Taxol biosynthesis. *Proc Natl Acad Sci U S A* 101(24):9149–9154.
213. Lenka SK, et al. (2012) Identification and expression analysis of methyl jasmonate responsive ESTs in paclitaxel producing Taxus cuspidata suspension culture cells. *BMC Genomics* 13(1):148.
214. Wu Q, et al. (2011) Transcriptome analysis of Taxus cuspidata needles based on 454 pyrosequencing. *Planta Med* 77(4):394–400.
215. Hao DCC, Ge G, Xiao P, Zhang Y, Yang L (2011) The first insight into the tissue specific taxus transcriptome via illumina second generation sequencing. *PLoS One* 6(6):e21220.
216. Hao D-CC, Yang L, Xiao P-GG, Liu M (2012) Identification of Taxus microRNAs and their targets with high-throughput sequencing and degradome analysis. *Physiol Plant* 146(4):388–403.
217. Sun G, et al. (2013) Deep Sequencing Reveals Transcriptome Re-Programming of Taxus × media Cells to the Elicitation with Methyl Jasmonate. *PLoS One* 8(4). doi:10.1371/journal.pone.0062865.
218. Onrubia Ibáñez M (2012) A molecular approach to taxol biosynthesis.
219. Cusido RM, et al. (2014) A rational approach to improving the biotechnological production of taxanes in plant cell cultures of Taxus spp. *Biotechnol Adv*. doi:10.1016/j.biotechadv.2014.03.002.
220. Dai Y, et al. (2009) Isolation and characterization of a novel cDNA encoding methyl jasmonate-responsive transcription factor TcAP2 from taxus cuspidata. *Biotechnol Lett* 31(11):1801–1809.
221. Waibel T (2011) Transcriptional regulation of taxol™ biosynthesis in taxus cuspidate procambium cells. Dissertation (University of Edinburgh). Available at: http://hdl.handle.net/1842/5681.
222. Yan Z (2013) Identification of transcription factors controlling the expression of paclitaxel biosynthesis genes in cambial meristematic cells of Taxus cuspidata. Dissertation (University of Edinburgh).
223. Amir R (2013) Identification and characterization of key regulators of paclitaxel biosynthesis in Taxus cuspidata. Dissertation (University of Edinburgh).
224. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–40.
225. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–402.
226. Zhang C-H, Wu J-Y (2003) Ethylene inhibitors enhance elicitor-induced paclitaxel production in suspension cultures of Taxus spp. cells. *Enzyme Microb Technol* 32(1):71–77.
227. Hentrich M, et al. (2013) The jasmonic acid signaling pathway is linked to auxin homeostasis through the modulation of YUCCA8 and YUCCA9 gene expression. *Plant J* 74(4):626–37.
228. Thaler JS, Humphrey PT, Whiteman NK (2012) Evolution of jasmonate and salicylate signal crosstalk. *Trends Plant Sci* 17(5):260–270.
229. Rezaei A (2011) Stimulation of taxol production by combined salicylic acid elicitation and sonication in Taxus baccata cell culture. 3:193–197.
230. Niu Y, Figueroa P, Browse J (2011) Characterization of JAZ-interacting bHLH transcription factors that regulate jasmonate responses in Arabidopsis. *J Exp Bot* 62(6):2143–2154.
231. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7(1):539.
232. Goujon M, et al. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 38(Web Server issue):W695–9.
233. Hall TA (1999) BioEdit: a user-friendly biological seqeunce alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98.
234. Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27(1):297–300.
235. Swiatek A, Lenjou M, Van Bockstaele D, Inzé D, Van Onckelen H (2002) Differential effect of jasmonic acid and abscisic acid on cell cycle progression in tobacco BY-2 cells. *Plant Physiol* 128(1):201–11.
236. Traw MB, Bergelson J (2003) Interactive effects of jasmonic acid, salicylic acid, and gibberellin on induction of trichomes in Arabidopsis. *Plant Physiol* 133(3):1367–75.
237. Hou X, Lee LYC, Xia K, Yan Y, Yu H (2010) DELLAs modulate jasmonate signaling via

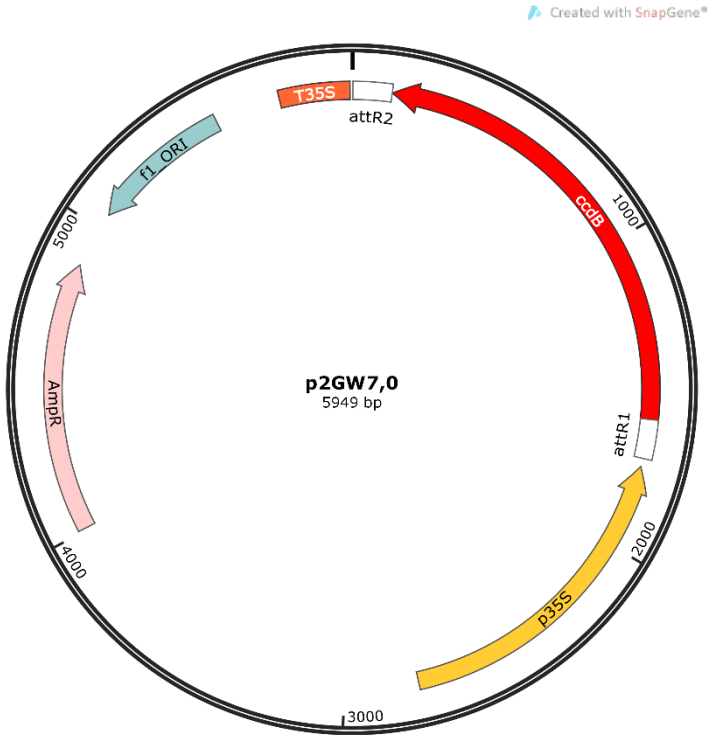
competitive binding to JAZs. *Dev Cell* 19(6):884–94.
238. Husbands A, Bell EM, Shuai B, Smith HMS, Springer PS (2007) Lateral organ boundaries defines a new family of DNA-binding transcription factors and can interact with specific bHLH proteins. *Nucleic Acids Res* 35(19):6663–6671.
239. Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT (2012) Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One* 7(2):e30087.
240. Quince C, et al. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* 6(9):639–41.
241. Chung HS, et al. (2008) Regulation and function of Arabidopsis JASMONATE ZIM-domain genes in response to wounding and herbivory. *Plant Physiol* 146(3):952–964.
242. Wu F-H, et al. (2009) Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods* 5:16.
243. Sheen J (2001) Update on Signal Transduction Signal Transduction in Maize and Arabidopsis Mesophyll Protoplasts 1. *Society* 127(December):1466–1475.
244. Cocking EC (1960) A method for the isolation of plant protoplasts and vacuoles. *Nature* 187:962–963.
245. Takebe I, Aoki S (1969) Infection of tobacco mesophyll protoplasts by tobacco mosaic virus ribonucleic acid. *Virology* 39:439–448.
246. Roberts SC, Naill M, Gibson DM, Shuler ML (2003) A simple method for enhancing paclitaxel release from Taxus canadensis cell suspension cultures utilizing cell wall digesting enzymes. *Plant Cell Rep* 21:1217–1220.
247. Luo JP, Mu Q, Gu YH (1999) Protoplast culture and paclitaxel production by Taxus yunnanensis. *Plant Cell Tissue Organ Cult* 59:25–29.
248. Czemmel S, et al. (2009) The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. *Plant Physiol* 151(3):1513–1530.
249. Yoo S-D, Cho Y-H, Sheen J (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc* 2(7):1565–1572.
250. Sheen J (2002) A Transient Expression Assay Using Arabidopsis Mesophyll Protoplasts A . Protoplast Isolation. *Plant Cell*:1–6.
251. Luehrsen KR, Walbot V (1993) Firefly luciferase as a reporter for plant gene expression studies. *Promega Notes Mag* 44(44):24.
252. Promega (2003) Dual-Luciferase Reporter Assay System. *System*:1–22.
253. Li P, et al. (2014) The ABI4-Induced Arabidopsis ANAC060 Transcription Factor Attenuates ABA Signaling and Renders Seedlings Sugar Insensitive when Present in the Nucleus. *PLoS Genet* 10(3):1–10.
254. Karimi M, Depicker A, Hilson P (2007) Recombinational cloning with plant gateway vectors. *Plant Physiol* 145(December):1144–1154.
255. Del Vecchio I, Zuccotti A, Canneva F, Lenzken SC, Racchi M (2007) Development of the first Gateway firefly luciferase vector and use of reverse transcriptase in FLOE (Fluorescently Labeled Oligonucleotide Extension) reactions. *Plasmid* 58:269–274.
256. Stability PEG, et al. (2012) PEG Stability: A look at pH and Conductivity changes. *Hampt Res*. Available at: hamptonresearch.com/documents/growth_101/27.pdf [Accessed January 16, 2015].
257. Kozak M (1991) Structural Features in Eukaryotic Messenger-Rnas That Modulate the Initiation of Translation. *J Biol Chem* 266(30):19867–19870.
258. Nims E, Vongpaseuth K, Roberts SC, Walker EL (2009) WITHDRAWN: TcJAMYC: A bHLH transcription factor that activates paclitaxel biosynthetic pathway genes in yew. *J Biol Chem*. doi:10.1074/jbc.M109.026195.
259. Onrubia M, et al. (2011) The relationship between TXS, DBAT, BAPT and DBTNBT gene expression and taxane production during the development of Taxus baccata plantlets. *Plant Sci* 181(3):282–287.
260. Kagale S, Rozwadowski K (2011) EAR motif-mediated transcriptional repression in plants: An underlying mechanism for epigenetic regulation of gene expression. *Epigenetics* 6(2):141–146.
261. Lee D-K, Geisler M, Springer PS (2009) LATERAL ORGAN FUSION1 and LATERAL ORGAN FUSION2 function in lateral organ separation and axillary meristem formation in Arabidopsis. *Development* 136(14):2423–2432.
262. Sakura H, et al. (1989) Delineation of three functional domains of the transcriptional activator encoded by the c-myb protooncogene. *Proc Natl Acad Sci U S A* 86(15):5758–5762.
263. Taoka KI, et al. (2004) The NAC domain mediates functional specificity of CUP-SHAPED

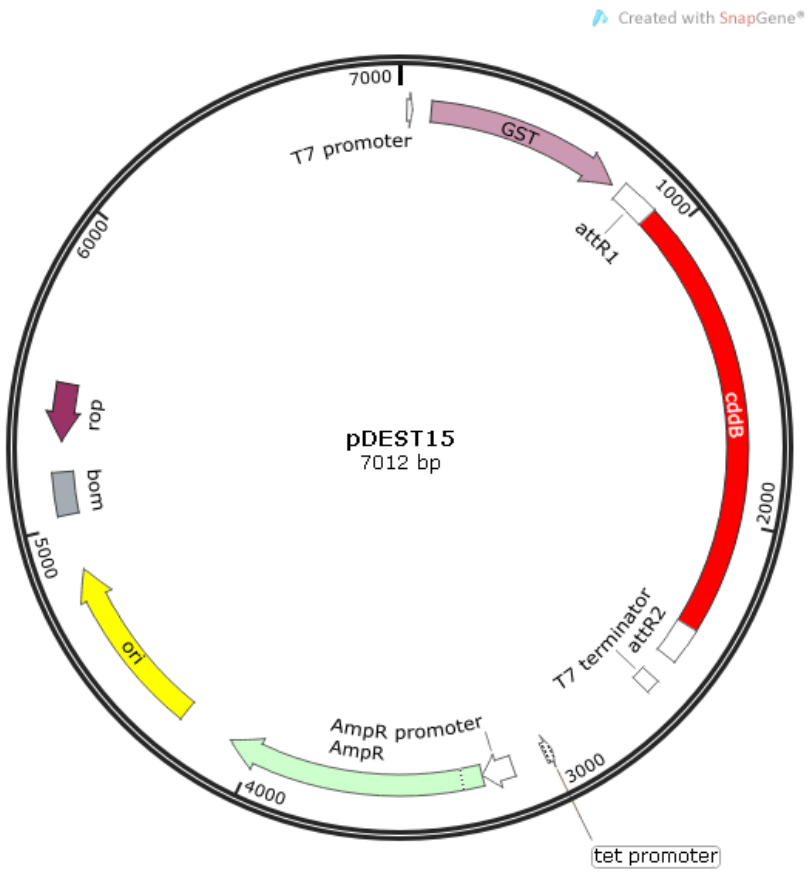
COTYLEDON proteins. *Plant J* 40(4):462–473.
264. Patra B, Schluttenhofer C, Wu Y, Pattanaik S, Yuan L (2013) Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim Biophys Acta - Gene Regul Mech* 1829(11):1236–1247.
265. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
266. Tamura K, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–9.
267. Alves M, et al. (2014) Transcription Factor Functional Protein-Protein Interactions in Plant Defense Responses. *Proteomes* 2(1):85–106.
268. Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 294(5):1351–1362.
269. Asai T, et al. (2002) MAP kinase signalling cascade in Arabidopsis innate immunity. 415(February):977–983.
270. Hezari M, Ketchum RE, Gibson DM, Croteau R (1997) Taxol production and taxadiene synthase activity in Taxus canadensis cell suspension cultures. *Arch Biochem Biophys* 337(2):185–190.
271. Memelink J, Gantet P (2007) Transcription factors involved in terpenoid indole alkaloid biosynthesis in Catharanthus roseus. *Phytochem Rev* 6(2-3):353–362.
272. Hichri I, et al. (2011) Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J Exp Bot* 62(8):2465–2483.
273. Li S (2014) Transcriptional control of flavonoid biosynthesis: Fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signal Behav* 8(12):1–7.
274. Grotewold E, et al. (2000) Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proc Natl Acad Sci U S A* 97(25):13579–84.
275. Carey CC, Strahle JT, Selinger DA, Chandler VL (2004) Mutations in the pale aleurone color1 regulatory gene of the Zea mays anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in Arabidopsis thaliana. *Plant Cell* 16(2):450–64.
276. Lee MM, Schiefelbein J (1999) WEREWOLF, a MYB-related protein in Arabidopsis, is a position-dependent regulator of epidermal cell patterning. *Cell* 99(5):473–483.
277. Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* 10(6):2763–2788.
278. Zimmermann IM, Heim M a., Weisshaar B, Uhrig JF (2004) Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J* 40(1):22–34.
279. Kanaoka MM, et al. (2008) SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to arabidopsis stomatal differentiation. *Plant Cell* 20(7):1775–1785.
280. Hu Y, Jiang L, Wang F, Yu D (2013) Jasmonate regulates the inducer of cbf expression-C-repeat binding factor/DRE binding factor1 cascade and freezing tolerance in Arabidopsis. *Plant Cell* 25(8):2907–24.
281. Stynen B, Tournu H, Tavernier J, Van Dijck P (2012) Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. *Microbiol Mol Biol Rev* 76(2):331–82.
282. Song S, et al. (2011) The Jasmonate-ZIM domain proteins interact with the R2R3-MYB transcription factors MYB21 and MYB24 to affect Jasmonate-regulated stamen development in Arabidopsis. *Plant Cell* 23(3):1000–1013.
283. Michael F. Smith J and SD-G (2001) Electrophoretic mobility shift assay (EMSA) Michael F . Smith , Jr . and Sandrine Delbary-Gossart University of Virginia School of Medicine Division of Gastroenterology and Hepatology Charlottesville , VA 22908. *Methods Mol Med*.
284. Laniel M, Béliveau A, Guérin SL (2001) Electrophoretic Mobility Shift Assays for the Analysis of DNA-Protein Interactions. *DNA-Protein Interactions*, pp 13–31. 2nd Ed.
285. Hellman LM, Fried MG (2007) Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein-Nucleic Acid Interactions. *Nat Protoc* 2(8):1849–1861.
286. Holden NS, Tacon CE (2011) Principles and problems of the electrophoretic mobility shift assay. *J Pharmacol Toxicol Methods* 63(1):7–14.
287. Alves C, Cunha C (2012) Electrophoretic Mobility Shift Assay: Analyzing Protein–Nucleic Acid Interactions. *CdnIntechopenCom*. Available at: http://cdn.intechopen.com/pdfs/34920.pdf.
288. Crooks G, Hon G, Chandonia J, Brenner S (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190.

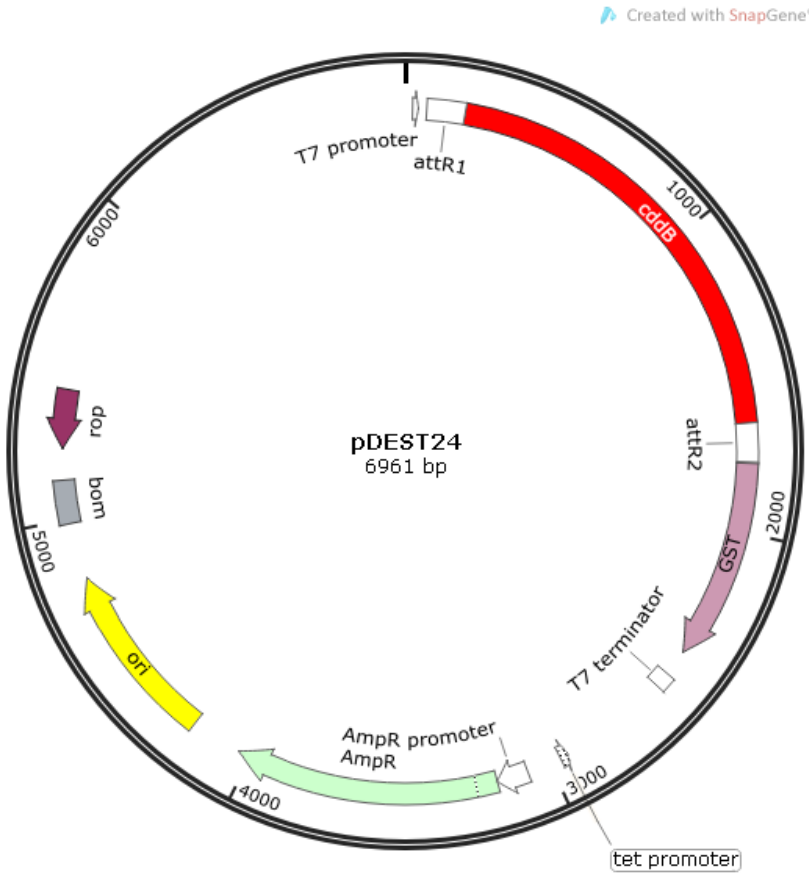
289. Deng QL, Ishii S, Sarai A (1996) Binding site analysis of c-Myb: screening of potential binding sites by using the mutation matrix derived from systematic binding affinity measurements. *Nucleic Acids Res* 24(4):766–74.
290. Prouse MB, Campbell MM (2013) Interactions between the R2R3-MYB Transcription Factor, AtMYB61, and Target DNA Binding Sites. *PLoS One* 8(5):e65132.
291. Wright WE, Binder M, Funk W (1991) Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. *Mol Cell Biol* 11(8):4104–4110.
292. Araki S, Ito M, Soyano T, Nishihama R, Machida Y (2004) Mitotic cyclins stimulate the activity of c-Myb-like factors for transactivation of G2/M phase-specific genes in tobacco. *J Biol Chem* 279(31):32979–32988.
293. Xue Y, et al. (2008) GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol Cell Proteomics* 7(9):1598–608.
294. Morse AM, Whetten RW, Dubos C, Campbell MM (2009) Post-translational modification of an R2R3-MYB transcription factor by a MAP Kinase during xylem development. *New Phytol* 183(4):1001–13.
295. Sanford JC, Klein TM, Wolf ED, Allen N (1987) Delivery of substances into cells and tissues using a particle bombardement process. *Part Sci Technol* 5(1):27–37.
296. Klein TM, Arentzen R, Lewis PA, Fitzpatrick-McElligott S (1992) Transformation of microbes, plants and animals by particle bombardment. *Biotechnology (N Y)* 10(3):286–291.
297. Vongpaseuth K, Nims E, St Amand M, Walker EL, Roberts SC (2007) Development of a particle bombardment-mediated transient transformation system for Taxus spp. cells in culture. *Biotechnol Prog* 23:1180–1185.
298. Ketchum REB, Wherland L, Croteau RB (2007) Stable transformation and long-term maintenance of transgenic Taxus cell suspension cultures. *Plant Cell Rep* 26:1025–1033.
299. Xue Y, et al. (2010) GPS-SNO: computational prediction of protein S-nitrosylation sites with a modified GPS algorithm. *PLoS One* 5(6):e11290.
300. Zhao Q, et al. (2014) GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res* 42(Web Server issue):W325–30.
301. Ren J, et al. (2008) CSS-Palm 2.0: an updated software for palmitoylation sites prediction. *Protein Eng Des Sel* 21(11):639–44.
302. Liu Z, et al. (2011) GPS-YNO2: computational prediction of tyrosine nitration sites in proteins. *Mol Biosyst* 7(4):1197–204.
303. Thanh T, Chi VTQ, Abdullah MP, Omar H, Napis S (2012) Efficiency of ligation-mediated PCR and TAIL-PCR methods for isolation of RbcS promoter sequences from green microalga Ankistrodesmus convolutus. *Mol Biol* 46(1):58–64.
304. Sambrook J, Russel DW (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press). 3rd Ed.
305. Salter MG, Conlon HE (2007) Extraction of plant RNA. *Methods Mol Biol* 362:309–314.
306. Gambino G, Perrone I, Gribaudo I (2008) A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochem Anal* 19(6):520–525.
307. Hartley JL, Temple GF, Brasch MA (2000) DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Res* 10(11):1788–1795.
308. Negrutiu I, Shillito R, Potrykus I, Biasini G, Sala F (1987) Hybrid genes in the analysis of transformation conditions : I. Setting up a simple method for direct gene transfer in plant protoplasts. *Plant Mol Biol* 8(5):363–73.

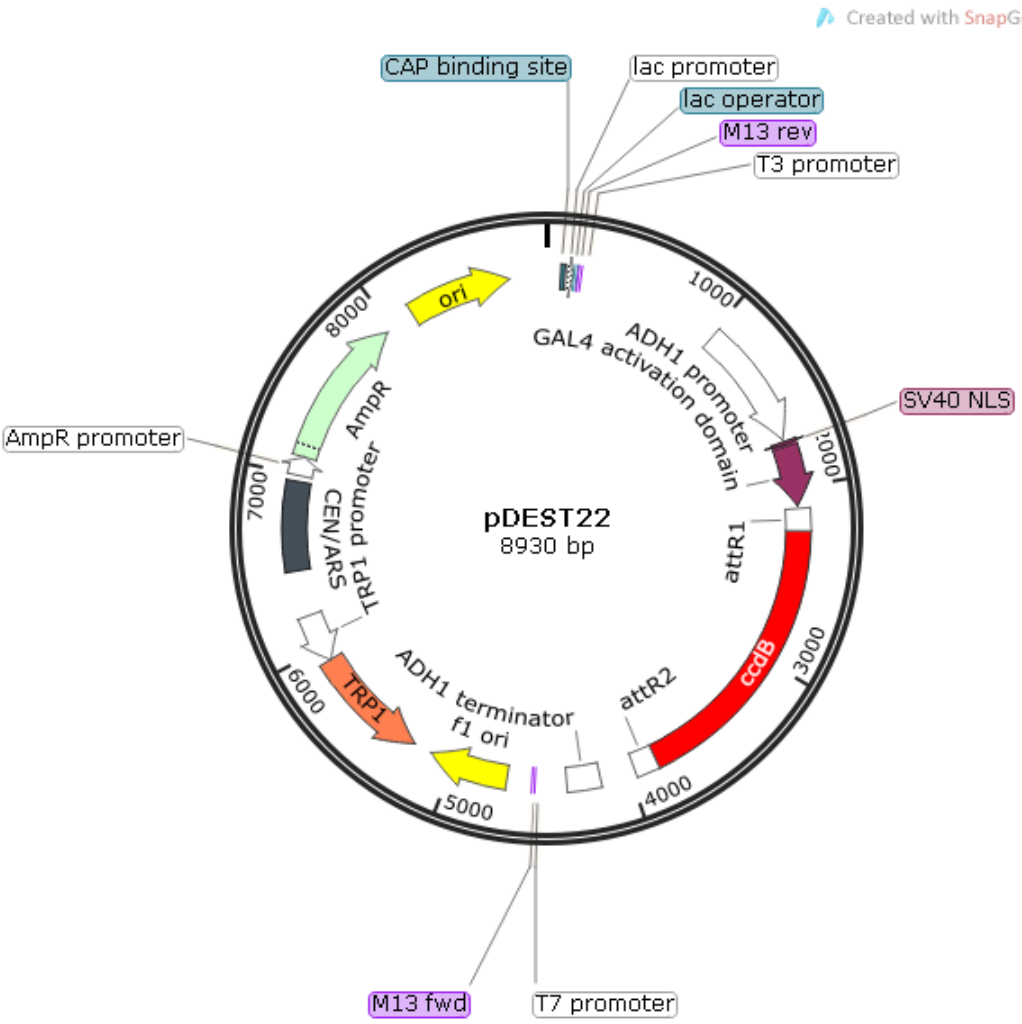
Appendixes

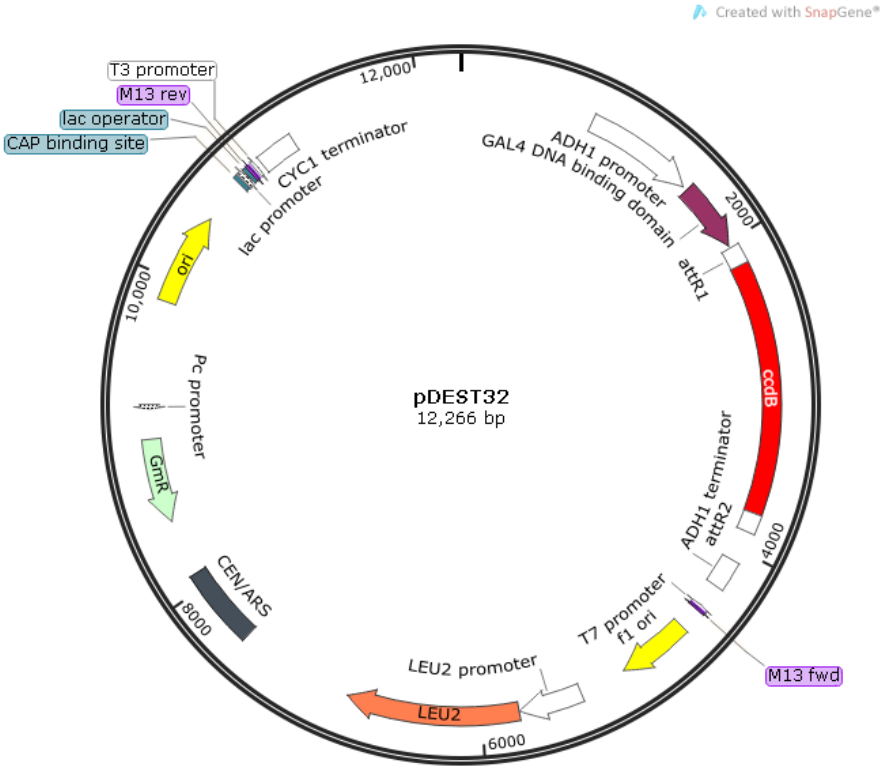
Appendix 1 –Vector maps of pDEST vectors used in thesis. p2GW7,0 and pGWlucB were the pDEST vectors for the transcription factors and paclitaxel promoters, respectively, in the *A. thaliana* protoplast transient assay. pDEST15 and pDEST24 are vectors used in *in vitro E. coli* protein expression, producing N- and C- terminally GST-tag protein respectively. pDEST22 is the prey vector and pDEST32 is bait vector in Y2H.

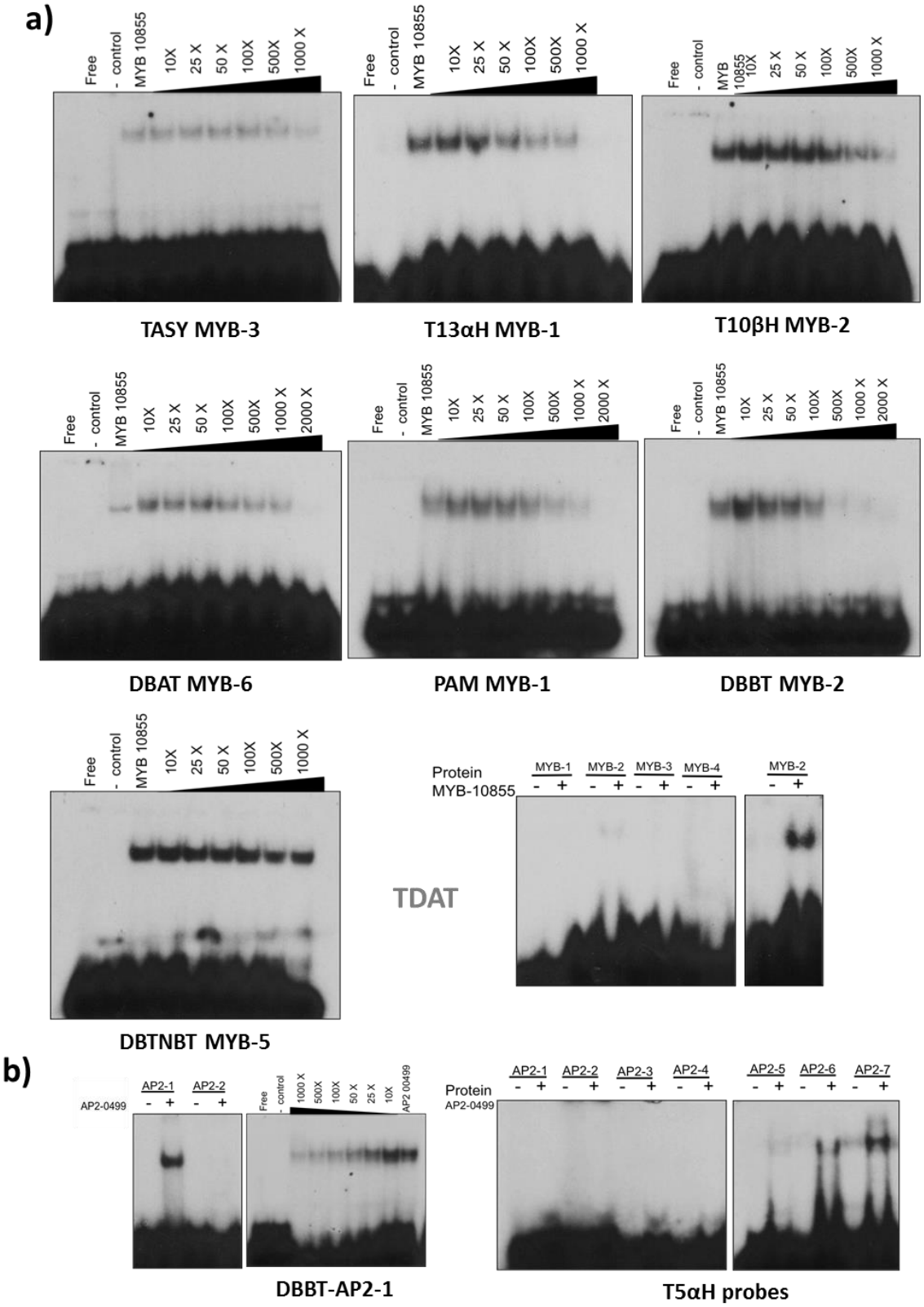



Created with SnapGene®
7000
T7 promoter
GST
1000
attR1
6000
ccdB
rop
bom
pDEST15
7012 bp
2000
5000
T7 terminator
attR2
ori
AmpR promoter
AmpR
3000
4000
tet promoter


Created with SnapGene®
T7 promoter
attR1
1000
ccdB
6000
attR2
rop
pDEST24
6961 bp
GST
bom
2000
5000
T7 terminator
ori
AmpR promoter
AmpR
3000
4000
tet promoter


Created with SnapGene®
CAP binding site
lac promoter
lac operator
M13 rev
T3 promoter
SV40 NLS
AmpR promoter
M13 fwd
T7 promoter
pDEST22
8930 bp
ori
AmpR
AmpR promoter
CEN/ARS
TRP1 promoter
TRP1
ADH1 terminator
f1 ori
attR2
ccdB
attR1
ADH1 promoter
GAL4 activation domain
1000
2000
3000
4000
5000
6000
7000
8000


Created with SnapGene®
T3 promoter
M13 rev
lac operator
CAP binding site
CYC1 terminator
lac promoter
ori
Pc promoter
GmR
CEN/ARS
LEU2 promoter
LEU2
T7 promoter
f1 ori
M13 fwd
ADH1 terminator
attR2
ccdB
attR1
ADH1 promoter
GAL4 DNA binding domain
pDEST32
12,266 bp
2000
4000
6000
8000
10,000
12,000



Appendix 2 - Summary of the electromobility shift assay (EMSA) results for MYB10855 and AP200499 completed by Ochoa Villarreal. a) Binding of MYB10855 to DNA probes in 8 paclitaxel promoters *TASY, T13αH, TDAT, T10ßH, DBBT, DBAT, PAM and DBTNBT* and; b) binding of AP200499 to DNA probes in *T5αH* and *DBBT.* EMSAs were completed with

GST tagged protein purified with glutathione agarose beads 4B (GE healthcare) and ^{32}P labelled 20 bp DNA probes, blots were exposed at -80°C for 24-72 hours. Free lanes are the DNA probe without any protein,– control stands for negative control and involves the addition of GST, the remaining lanes have the TF under investigation with increasing amounts of unlabelled competitor probe as stated and the probe under investigation is stated at the bottom of the blot. Competitor assays involving the addition of increasing amounts of unlabelled probe show the specificity of the interactions but were not completed for MYB10855 with *TDAT* and AP200499 with *T5αH* because the interaction was too weak.

Addendum

Further work was conducted after the submission of the thesis to establish the full ORF of bHLH08058. TAIL-PCR was used to extend the sequence in the 5' direction, which produced a 700bp sequence. This contained a longer open reading frame for the bHLH TF, bHLH08058-A and a 413bp region of the promoter (Figure 71). The protein alignment of bHLH08058-A with MYC2 homologs from *A. thaliana, C. roseus, N. tabacum* and *P. tadea* showed a high similarity in the N-terminal region where the JID is located (Figure 72).

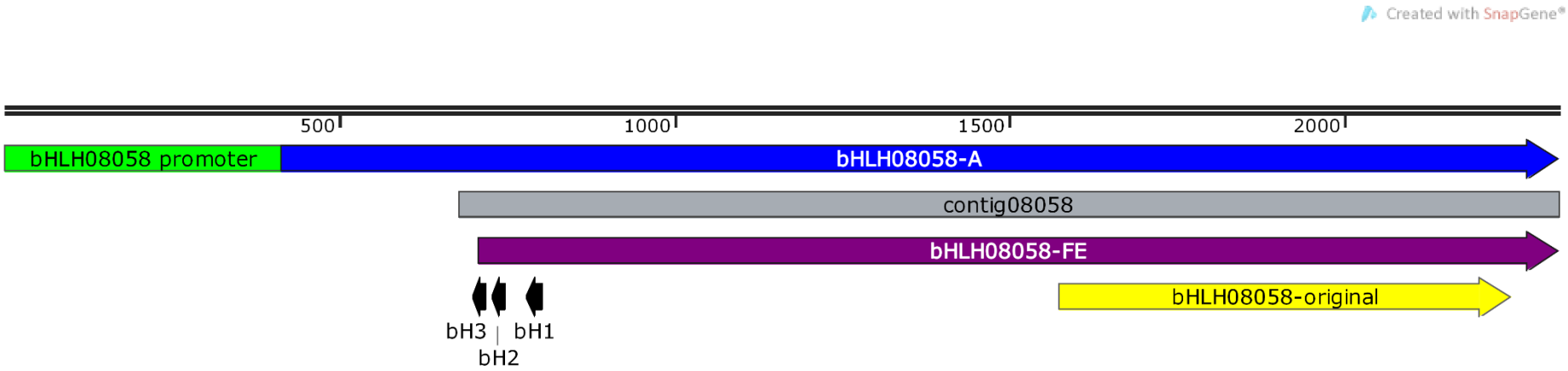


Figure 71 – The different identified open reading frames (ORFs) of bHLH08058. bHLH08058-FE (purple, 1611bp) is the further extended version of bHL08058-original (yellow, 675bp) that could be identified in the Roche454 contig08058 (grey, 1639bp), see figure 23 for further discussion on bHLH08058-FE identification. TAIL- PCR using the bH1, bH2 and bH3 primers (black arrows) were used to extend the sequence in the 5' direction. A longer ORF bHLH08058-A (blue, 1905bp) and a 414bp region of the bHLH08058 promoter was identified. Image created using SnapGene software (from GSL Biotech; available at snapgene.com).

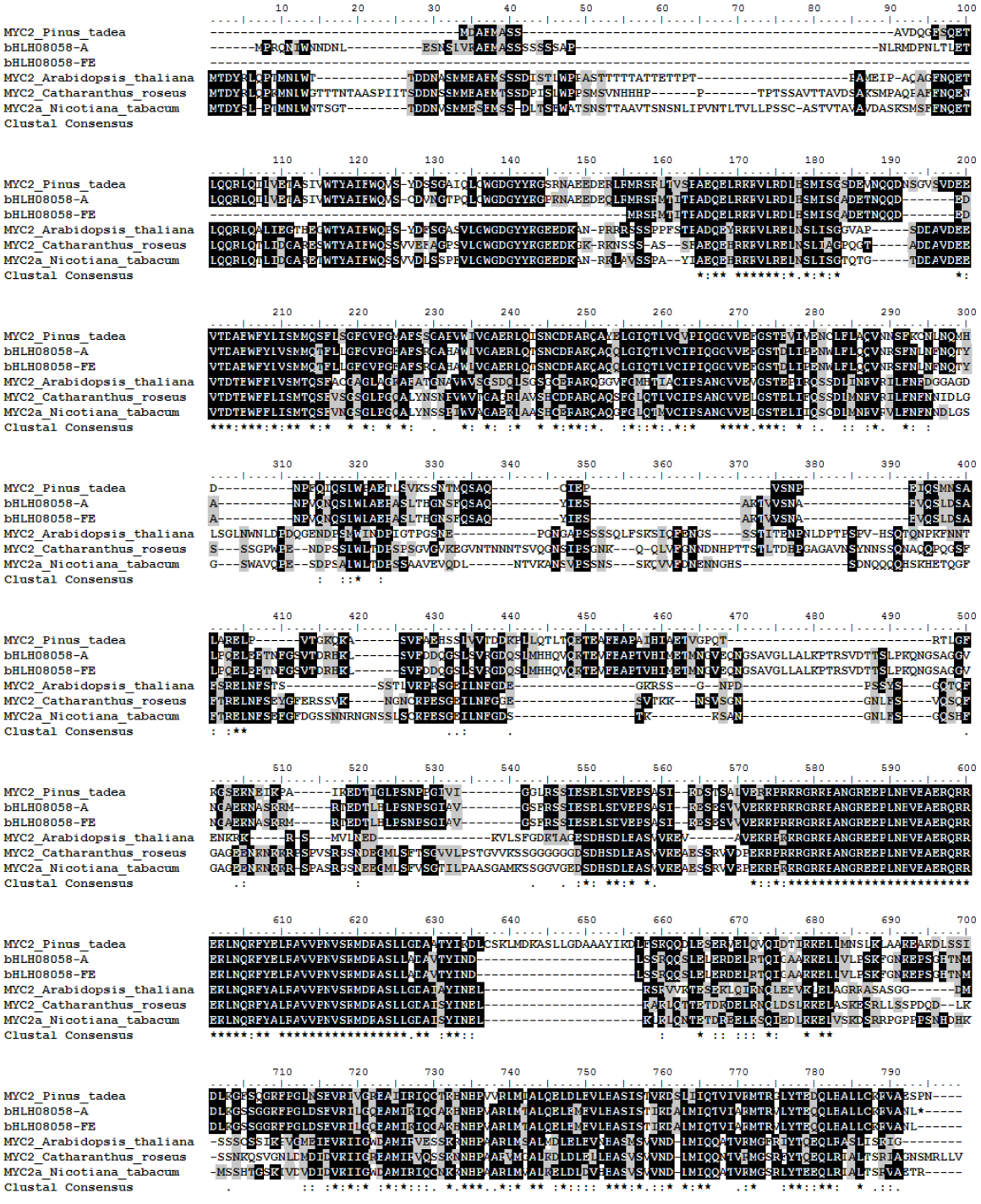


Figure 72 - The protein alignment of the further extended version of bHLH08058 (bHLH08058-FE) and the ORF bHLH08058-A identified by TAIL-PCR, the Walker TcMYC4 (185), published examples of MYC2 in *Arabidopsis thaliana, Catharanthus roseus* and *Nicotiana tabacum* and an identified MYC2 homolog in *Pinus tadea*. The alignment was produced using ClustalOmega (232, 233) and the image produced using BioEdit v7.2.5 (234). There is a high level of sequence alignment between bHLH08058-A and the published examples of MYC2 extending beyond bHLH08058-FE therefore it is highly likely that the ORF identified by TAIL-PCR is correct.