THE UNIVERSITY
of EDINBURGH

Identification of transcription factors controlling the expression of paclitaxel biosynthesis genes in cambial meristematic cells of *Taxus cuspidata*

Zejun Yan

Doctor of Philosophy

The University of Edinburgh

2013

Abstract

Paclitaxel is an antitumor diterpene from *Taxus* spp. that binds tubulin, stabilizes microtubules and induces apoptosis in dividing human cells. It was originally isolated from the bark of *Taxus brevifolia* and approved for clinic uses by the FDA in 1992. Because of its excellent activity in treatment of various cancers, a significant supply shortage has been created by the enormous demand for this natural product. Thus, researchers have been focusing on the development of effective ways to increase the production of paclitaxel and related bioactive molecules. This shortage was initially solved by over-harvesting of *T. brevifolia* bark; however, it is not an environment-friendly, effective and sustainable way to supply paclitaxel. A semi-synthetic route was then developed to convert the more readily available and renewable 10-deacetylbacatin III into paclitaxel. As an alternative, plant cell cultures have been employed to commercially produce paclitaxel and it is a more environment-friendly and sustainable route to end the supply crisis. However, problems associated with plant cell culturing at an industrial scale, such as cell aggregation and variability in yield, significantly affect paclitaxel production. Therefore, a discovery of a better-performing *Taxus* cell line might be a solution to overcome these culturing-associated problems.

A cambial meristematic cell (CMC) line of *Taxus cuspidata* has been isolated, cultured and demonstrated to be a cost-effective and environmentally friendly platform for the sustainable production of paclitaxel (Lee et al. 2010). Compared to dedifferentiated cell (DDC) lines, CMC lines are undifferentiated cells and proved to have stem cell-like properties. When cultured at an industrial scale, this cell line

contains much smaller cell aggregates with many cells appearing as singletons, the biomass of which is still increasing after 22-month culturing, and has much greater paclitaxel production after elicitation (Lee et al. 2010).

In my project, we aimed to identify the transcription factors (TFs) that regulate the expression of paclitaxel biosynthesis genes. We performed Illumina Solexa sequencing on cDNA libraries derived from methyl jasmonate (MeJA)-elicitated CMCs to digitally profile gene expression. Analysis of differentially expressed gene (DEG) abundance led to the discovery of 19 putative TFs and bioinformatic analysis further showed that these 19 TFs belong to 5 different TF families. Further, the DNA binding motifs associated with these TFs can be found in the promoters of the two early, *taxadiene synthase (TASY)* and *taxadiene 5α hydroxylase* (*T5αH*), and three late, *10-deacetylbaccatin III-10-O-acetyltransferase* (*DBAT)*, *phenylpropanoyltransferase* (*PAM*) and *3'-N-debenzoyl-2-deoxytaxol-N-benzoyltransferase* (*DBTNBT*), paclitaxel biosynthesis pathway genes. Then, yeast one-hybrid analysis, gel shifting assays and plant transient expression assays (TEA) were employed to assay TFs that interact with these promoters. Although Y1H screening did not show any convincing TF-promoter interactions, the attempted plant transient expression assay in the leaves of *Nicotiana benthamiana* might be a more suitable system to screen the positive regulators. Finally, the elucidation of a TF regulatory network that controls paclitaxel biosynthesis will guide the rational engineering of CMCs to ultimately increase yields of this important pharmaceutical.

Declaration

I hereby declare that the work presented here is my own and has not been submitted in any form for any degree at this or any other university

Zejun Yan

Table of Contents

Acknowledgements

First of all, I would like to thank my parents and my brother Zemin Yan for funding my PhD and supporting me.

I would like to thank my first and second supervisors Prof. Gary J. Loake and Dr. Chris French. My thank goes also to all past and present lab members.

For providing the plant material utilised in my project I would like to thank the Biotech Company Unhwa in South Korea.

For providing the wheat germ extracts utilised for protein synthesis I would like to thank Dr. Yasuomi Tada at Kagawa University, Japan.

For advice and guidance regarding the general experimental techniques I would like to thank Dr. Xu. X, Dr. Manda, Dr. Steven H. Spoel and Dr. Byung-wook.

For advice and guidance regarding the *Agrobacterium*-mediated infiltration of *N. benthamiana* I would like to thank the members of the Oparka lab.

I also would like to thank my colleagues Eunjung Kwon, Rabia Amir and Suzy Howat who have been working with me on this project.

Finally, I would like to thank my friends Yan Zhou, Yuhua Wang, Jun Wei, Gaoyuan Hu and Xuan Wang who supported me throughout my PhD study at Edinburgh University.

Publication arising from this work

Lee EK, Jin YW, Park JH, Yoo YM, Hong SM, Amir R, Yan Z, Kwon E, Elfick A, Tomlinson S, Halbritter F, Waibel T, Yun BW, Loake GJ (2010) Cultured cambial meristematic cells as a source of plant natural products. Nat biotechnol 28: 1213-1217

Yun UW, Yan Z, Amir R, Hong S, Jin YW, Lee EK, Loake GJ (2012) Plant natural products: history, limitations and the potential of cambial meristematic cells. Biotechnol Genet Eng Rev 28: 47-59

Abbreviations

10- DAB 10-deacetyl-baccatin III
13-AOS 13-allene oxide synthase
13-HPOT (13S)-hydroperoxyoctadecadienoic acid
13-LOX 13-lipoxygenase
3-AT 3-amino-1,2,4-triazole
35Sm *35S* minimal promoter
ABC ATP-binding cassette
ACX1 acyl-CoA oxidase
ADH1 alcohol dehydrogenases 1
Amp^R ampicillin resistance
ANR anthocyanin reductase
AOC allene oxide cyclase
AP2/ERF apetala 2/ethylene-responsive element binding factor
BAPT taxoid C13-O-phenyl-propanoyltransferase
bHLH basic helix-loop-helix
BL-SOM batch-learning self-organizing map
BMS Bristol-Myers Squibb
CaMV*35S* cauliflower mosaic virus *35S* promoter
ChIP chromatin immunoprecipitation
ChIP-chip chromatin immunoprecipitation coupled with microarray
ChIP-seq chromatin immunoprecipitation coupled with Next-generation sequencing
CMCs cambial meristematic cells
Cm^R chloramphenicol resistance
CNS central nervous system
CPT camptothecin
CSC cell suspension culture
CSDs conserved signature domains
CUC2 cup-shaped cotyledon2
DBAT 10-deacetylbaccatin 10β-*O*-acetyltransferase

DBD	DNA binding domain
DBTNBT	3'-*N*-debenzoyl-2'-deoxytaxol *N*-benzoyltransferase
DDCs	dedifferentiated cells
DEGs	differentially expressed genes
DIMBOA	2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one
DMAPP	dimethylallyl diphosphate
DRE	dehydration-responsive element
DREB	dehydration-responsive element-binding protein
EAR	ethylene responsive factor-associated amphiphilic repression motif
EMSA	electrophoretic mobility-shift assay
ERE	ethylene-responsive element
EREBPs	ethylene-responsive element binding proteins
eY1H	enhanced Y1H
FAD	Food and Drug Administration
FDR	false discovery rate
FDR	false discovery rate
FISH	fluorescence in situ hybridisation
GAL4DBD	Gal4 DNA binding domain
GGPP	geranylgeranyl diphosphate
GGPPS	geranylgeranyl diphosphate synthase
GLUTs	glucose transporters
GO	gene ontology
GRN	gene regulation network
GUS	β-glucuronidase
H3K27	H3 lysine 27
HIS3	Imidazoleglycerol-phosphate dehydratase
HTH	helix-turn-helix
Hyg^R	hygromycin resistance
InsP5	inositol pentakisphosphate
IPP	isopentenyl diphosphate
JAR1	jasmonate resistant 1
JAZ	jasmonate-ZIM domain

JERE	jasmonate and elicitor responsive element
Kan^R	kanamycin resistant
KAT	L-3-ketoacyl CoA thiolase
kDa	kilodalton
MAPK	mitogen-activated protein kinase
MATE	multidrug and toxic compound extrusion
MeJA	methyl jasmonic acid
MFP	multifunctional protein
MRP	multidrug resistance-associated protein
MS	mass spectrometry
MSC	multiple cloning site
MYBBS	MYB binding site
NAC	NAM, ATAF1, 2, CUC2
NACBS	NAC binding site
NCBI	national centre for biotechnology information
NCEs	new chemical entities
NINJA	novel interactor of JAZ
NLS	nuclear localisation signal
NMR	nuclear magnetic resonance
NST1/3	NAC secondary wall thickening promoting factors1/3
OE	overexpression
OPDA	cis (+)-12-oxophytodienoic acid
OPR3	OPDA reductase 3
ORCA	octadecanoid responsive *C. roseus* AP2
ORF	open reading frame
PAM	phenylalanine aminomutase
PAs	proanthocyanidins
pDES	destination vector
pDONR	donor vector
pENTR	entry clone
PMT	putrescine *N*-methyltransferase
PTEA	plant transient expression assay

PTM	post-translational modification
RIPs	ribosome-inactivating proteins
RNA-seq	RNA sequencing
RT-PCR	reverse transcription-PCR
SCF^{COI1}	SKP1, Cullin, F-box protein E3 ubiquitin ligase
SCs	somatic cells
SIB	sigma factor binding protein
Sp^{R}	spectinomycin resistance
STR	strictosidine synthase
T10βH	taxane 10β-hydroxylase
T13αH	taxane 13α-hydroxylase
T2αH	taxoid 2α-hydroxylase
T5AT	taxa-4-(20),11(12)-dien-5α-ol-*O*-acetyl transferase
T5αH	taxadiene-5α-hydroxylase
T7βH	taxoid 7β-hydroxylase
TASY	taxadiene synthase
TBT	2α-hydroxytaxane 2-O-benoyltransferase
TDC	tryptophan Decarboxylase
TEA	transient expression assay
TFs	transcription factors
TIA	terpenoid indole alkaloid
TPL	TOPLESS
TPSs	terpene synthase
X-gluc	5-bromo-4-chloro-3-indolyl glucuronide
Y1H	yeast one-hybrid
Y2H	yeast two-hybrid
ZIM	zinc-finger inflorescence meristem

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Chapter 9

List of Tables

Chapter 1: Introduction

1.1 Natural Products of Plant Origin

Historically, humans have been utilizing plants as endless sources of medicines for thousands of years. Approximately 25% of all drugs prescribed today are isolated from plants (Schmidt et al. 2007). Plants naturally produce a startling array of natural products, which are not required for their growth and development, but rather essential for plants to interact with their surrounding environment. These so-called plant secondary metabolites have been playing an important role in treatment of diseases ranging in effects from antimalarials (artemisinin), antiparasitics (quinine), anticholinergics (atropine) to antineoplastics (vinblastine and vincristine). In particular, plant natural products have made a great contribution to the area of cancer treatment. Approximately 155 anticancer drugs have been developed since 1940, about half of which are either natural products or direct derivations thereof (Schmidt et al. 2007). Despite the recent techniques used in the field of drug discovery by pharmaceutical companies, such as molecular modelling, combinatorial chemistry and other synthetic chemistry, plant natural products still remain an important source of new drugs, new drug leads and new chemical entities (NCEs)(Newman et al. 2000). Firstly, because plant natural products form a large reservoir of natural chemical diversity. It has been estimated that approximately 50,000 metabolites have been identified in plants; in addition, it is predicted that the final number will exceed 200,000 (Hounsome et al. 2008). Such a rich diversity partially results from an evolutionary process driven by natural selection; as a consequence, plants acquire improved defence against attacks from microbes, insect and animals (Dixon 2001). Secondly, natural products have naturally built-in chirality, and thus uniquely bind to their targets, including complex proteins and other three-dimensional biological receptors (Kingston 2009). Consequently, they might perform better as drugs than artificially synthesised compounds produced by combinatorial chemistry in the treatment of diseases (Yun et al. 2012), because they can treat diseases more specifically and have fewer side effects.

1.2 Plant-derived antineoplastics

Cancer has been one of the top causes of death worldwide. According to Global Cancer Facts and Figures (2nd edition) by American Cancer Society, cancer is the leading and second leading cause of death in developed and developing countries, respectively. So far, there have been a few plant-derived anticancer drugs approved for clinic uses. Vinca alkaloids, which were discovered by Robert Noble and Charles Beer in 1950's, are isolated from the Madagascar periwinkle (*Catharanthus roseus*) and were the first approved drugs for cancer treatment (Newman et al. 2000). Two of the best known indole alkaloids in clinic use are vinblastine and vincristine (Noble 1990). Camptothecin (CPT), a cytotoxic quinoline alkaloid, which was isolated from the Chinese ornamental tree *Camptotheca acuminata*, is also proven to be an anticancer drug. However, because of its severe bladder toxicity, the modified CPTs, such as topotecan (Hycampin®) and irinotecan (Camptosar®), are approved for clinic use in the treatment of ovarian and small cell lung cancers and colorectal cancers in humans, respectively (Creemers et al. 1996; Fuchs et al. 2006; Newman et al. 2000). Furthermore, homoharringtonine (Omapro®) isolated from the Chinese tree *Cephalotaxus harringtonia* is also a plant-derived anticancer agent for the treatment of acute myelogenous leukemia and chronic myelogenous leukemia (Kantarjian et al. 1996). Finally, Paclitaxel (Taxol®), which was initially isolated from the bark of *Taxus brevifolia,* is a complex diterpene (Wani et al. 1971). Its clinic use was approved in the early 1990. Paclitaxel is a broad-spectrum anticancer agent, because it is significantly active against various cancers, including ovarian cancer, advanced breast cancer, small and non-small cell lung cancer (Rowinsky et al. 1992). Particularly, because of its efficacy and tolerability and the lack of cross-resistance with anthracycline, paclitaxel is an important drug in the treatment of breast cancer (Perez 1998).

1.3 Plant Secondary Metabolites

Plant secondary metabolites are a vast and diverse assortment of organic compounds that are not directly required for plant growth and development. Unlike plant primary metabolites, which are essential for nutrition and metabolic processes inside the plant, plant secondary metabolites participate in interactions between the plant and its surrounding environment, such as interactions between plants and pathogens. Based on their biosynthetic origin, plant secondary metabolites can be grouped into three classes: the terpenoids, the alkaloids, and the phenylpropanoids and allied phenolic compounds. Terpenoids, which are derived from the five-carbon precursor isopentenyl diphosphate (IPP), are a class of more than 25,000 secondary metabolites. Alkaloids, a group of more than 12,000 known organic compounds, are characterised by containing one or more nitrogen atoms. They are principally synthesised from amino acids. Finally, the group of phenylpropanoids and allied phenolic compounds contains about 8,000 or so phenolic compounds, and are synthesised through either of the two metabolic pathways: shikimic acid pathway or the malonate/ acetate pathway (Buchanan 2000).

1.3.1 Plant terpenoids

Terpenoids are made up of a diverse array of structurally varied compounds, the building blocks of which are five-carbon units referred as isopentenyl diphosphate (IPP). All members of terpenoids are biosynthesised by repetitive fusion of the branched IPP units based on an isopentane skeleton. Therefore, terpenoids are classified based on the number of IPP units they contain in their molecular structures. The smallest terpenoids contain a single IPP (C_5) unit; they are named hemiterpenes as a group. Isoprene is the best known example of hemiterpene, which is a volatile product released from photosynthetically active tissues of plants. Terpenoids that contain two IPP units (C_{10}) are called monoterpenes, from which the subsequent nomenclature is derived. They were the first terpenoids isolated from terpentine in 1850s. Monoterpenes are components of volatile essence present in the essential oils secreted by herbs, and comprise as much as 5% of plant dry weight. Furthermore,

diterpenes, which have 20 carbons (4 IPP units), consist of a variety of C_{20}-containing organic compounds ranging from phytol, the gibberellin hormones, phytoalexins and paclitaxel. Triterpenes (C_{30}) also contain a diverse array of structurally and functionally different types of molecules, such as the brassinosteroids (plant hormones), phytoalexins, toxins and components of surface waxes (oleanolic acid of grapes). Tetraterpenes (C_{40}) that contain eight IPP units consists of the carotenoid accessory pigments, which are functionally essential components of photosynthesis. Finally, the polyterpenes, which contain more than 8 IPP units, consist of the prenylated quinone electron carriers (plastoquinone), long-chain polyprenols (dolichol) and long polymers (rubber) (Buchanan 2000).

The biosynthesis of the fundamental precursor IPP is the starting point of terpenoid biosynthesis. Three molecules of acetyl-CoA are required for synthesis of each IPP in the acetate/mevalonate pathway in the cytosol. Fig 1.1 shows the basic enzymology of IPP biosynthesis. Then, a series of prenyl diphosphate homologs, which serve as the immediate precursors of the different classes of terpenoids, are formed through repetitive additions of IPP. The formation of terpenoid skeletons from immediate prenyl diphosphates is catalysed by specific terpenoid synthases located in the plastid. The family of genes encoding plant terpene synthases (TPSs) is highly diversified throughout the plant kingdom. *TPS* genes can be grouped into seven clades (Fig 1.2), with some types of genes showing their lineage-specificity (Chen et al. 2011). Finally, the secondary enzymatic modification to the skeletons gives rise to the structurally and functionally different terpenoids (Buchanan 2000).

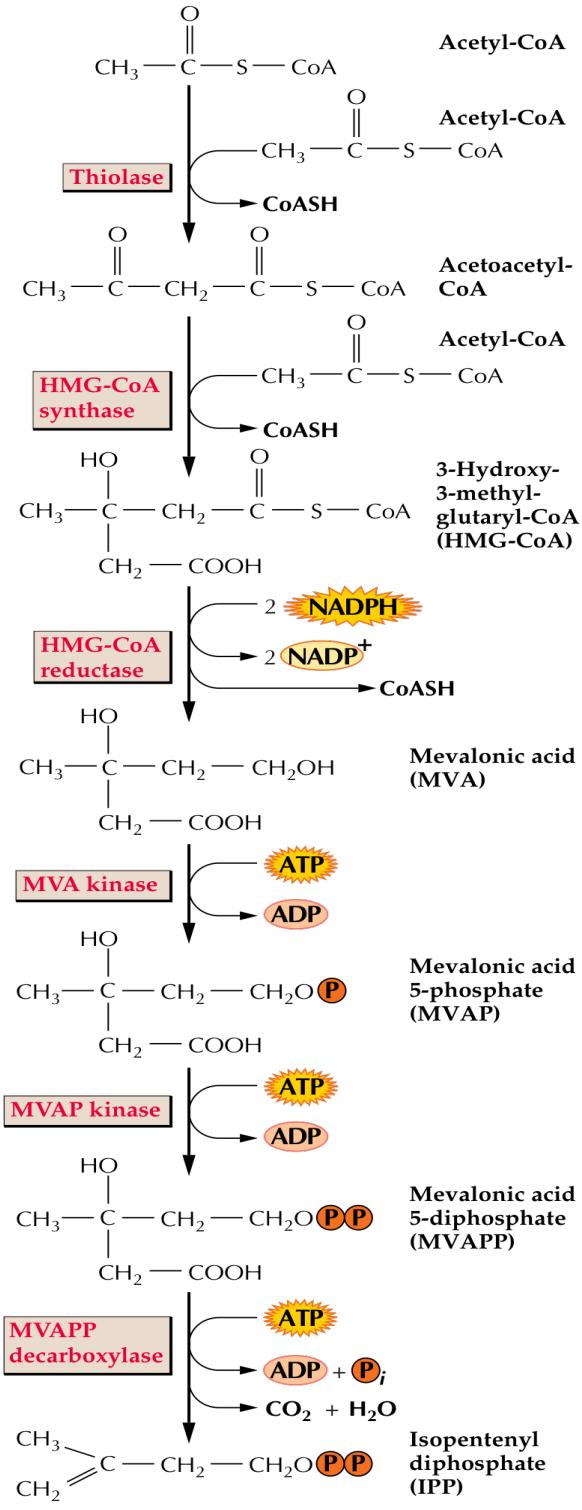


Fig 1.1 (Buchanan 2000) The biosynthesis pathway of isopentenyl diphosphate (IPP) through acetate/mevalonate pathway, the basic building block of terpenoids, in cytosol.

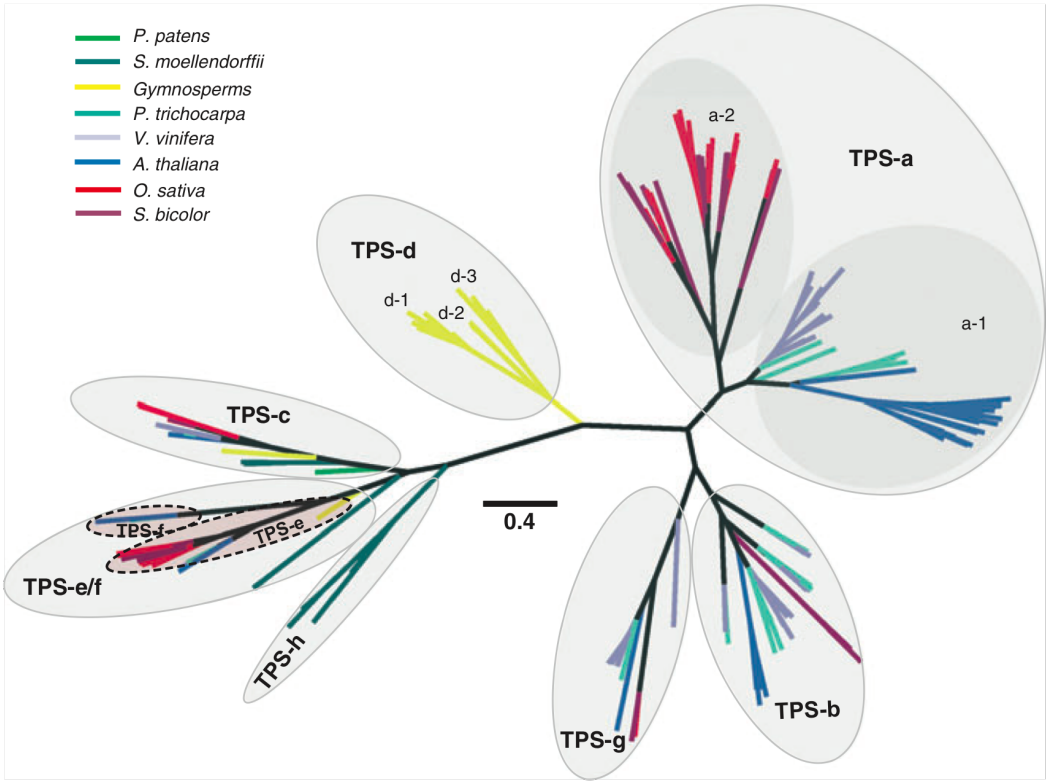


Fig 1.2 (Chen et al. 2011) Phylogeny of putative TPSs from seven sequenced plant genomes and TPSs from gymnosperms. Based on the phylogeny and functions of known TPSs, the TPSs family can be classified into seven subfamilies. Whereas the three subfamilies that include TPS-a, b and g are angiosperm-specific, TPS-d is gymnosperm-specific. TPS-a is further divided into two groups. TPS-a-1 is dicot-specific and a-2 is monocot-specific. TPS-c is most conserved among land plants. In contrast, TPS-e/f is most conserved among vascular plants. The latter emerged from two separate TPS-e and TPS-f subfamilies. Finally, the TPS-h is *Selaginella moellendorffii*-specific. 0.4 is the scale of branch length.

1.3.2 Paclitaxel

1.3.2.1 History

In 1964, yew extract from the bark of a Pacific Yew tree (*Taxus brevifolia*) was first shown to inhibit the growth of cancer cells. In 1967, Monroe Wall confirmed that it was paclitaxel present in the extract that showed an anticancer activity. It was approved for clinic uses by the Food and Drug Administration (FAD) in 1992. The synthesis of paclitaxel began by extracting paclitaxel from yew trees in 1971. However, plant extraction is neither environment friendly nor economically practical. It would have taken about 360,000 yew trees to treat all the patients of ovarian cancer in USA for one year. In addition, it takes hundreds of years for a yew tree to yield a good amount of paclitaxel. In 1980, 10-deacetyl-baccatinIII (10-DAB), the precursor of paclitaxel, was isolated from the needles of the European yew tree (*Taxus baccata*) by a French group. Pierre Potier and Andrew Greene successfully converted 10-DAB into paclitaxel. Although the yield was little, it showed a better way to synthesise paclitaxel. Since the production of taxusin (a paclitaxel-related compound that occurs naturally in yew) by Robert Holton in 1988, pharmaceutical companies started funding research towards the chemical synthesis of paclitaxel. In 1992, Bristol-Myers Squibb (BMS) patented a new semisynthesis set up by Holton, the metal alkaloid process. 10-DAB was the source in the process and supplied by extraction of renewable yew needle. Hence, the need to supply yew bark ended at the end of 1993. At the same time, Gary Strobel with his colleagues isolated a fungus from the bark of the Pacific yew in North West Montana named Taxo -myces andreanae. After analysis of the product produced by the fungus culture, a trace amount of paclitaxel was found. Although the production of paclitaxel by this fungus is very low, and therefore not competitive, it is still promising. Finally, Rainer Zocher at TU Berlin started isolating enzymes that catalyse the reactions needed for paclitaxel production from 10-DAB. As a result, 10-DAB acetyl transferase was successfully isolated and its expression in *E. coli* could lead to a toxic-free production of paclitaxel from 10-DAB (Renneberg 2007).

1.3.2.2 Chemistry and biological role of paclitaxel ($C_{47}H_{51}NO_{14}$)

Paclitaxel is a natural occurring diterpene with a complex molecular structure (Fig 1.3). It is based on a C_{20}-terpenoid and belongs to the taxane family. It has a core ring structure, known as the taxane ring cluster, which is made up of three rings. The carbon-13 (C_{13}) side chain is in extended conformation, which places the benzamide group farthest from the core. There are benzoate and acetate groups present on C_2 and C_{10}, respectively. In addition, the C_2 benzoate and C_{10} acetate groups extend away from the taxane ring cluster in opposite direction. Furthermore, the acetate group on C_4 points away from the C_2 benzoate (Mastropaolo et al. 1995).

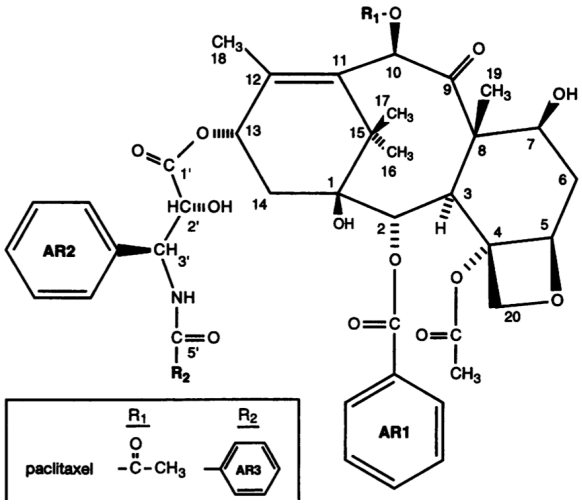


Fig 1.3 (Mastropaolo et al. 1995) Molecular structure of paclitaxel. It contains a core ring structure and a side chain at the position of carbon-13.

Paclitaxel inhibits cancer cell growth by promoting the assembly of microtubules. Microtubules are thought to be "the most strategic subcellular targets of anticancer chemotherapeutics" (Rowinsky 1991), as they play important roles in many aspects

of cellular biological processes, such as shaping the cell, positioning and reorganising the organelles and moving cells. Additionally, microtubules are absolutely necessary for cell division (Nicolaou 1994). Microtubules are built up from tubulin dimers. Each dimer consists of two subunits: α and β tubulins, which exist as α-β heterodimers. Paclitaxel binds into a pocket in the second globular domain of β tubulin, which is present on the microtubule's inner surface. The amino acid residues in closest contact with paclitaxel are labelled (Fig 1.4). This physical interaction counteracts the effects of GTP hydrolysis on the other side of the monomer; as a consequence, the depolymerisation of microtubules is inhibited and cell cycle is arrested (Amos and Lowe 1999; Nicolaou 1994; Schiff et al. 1979).

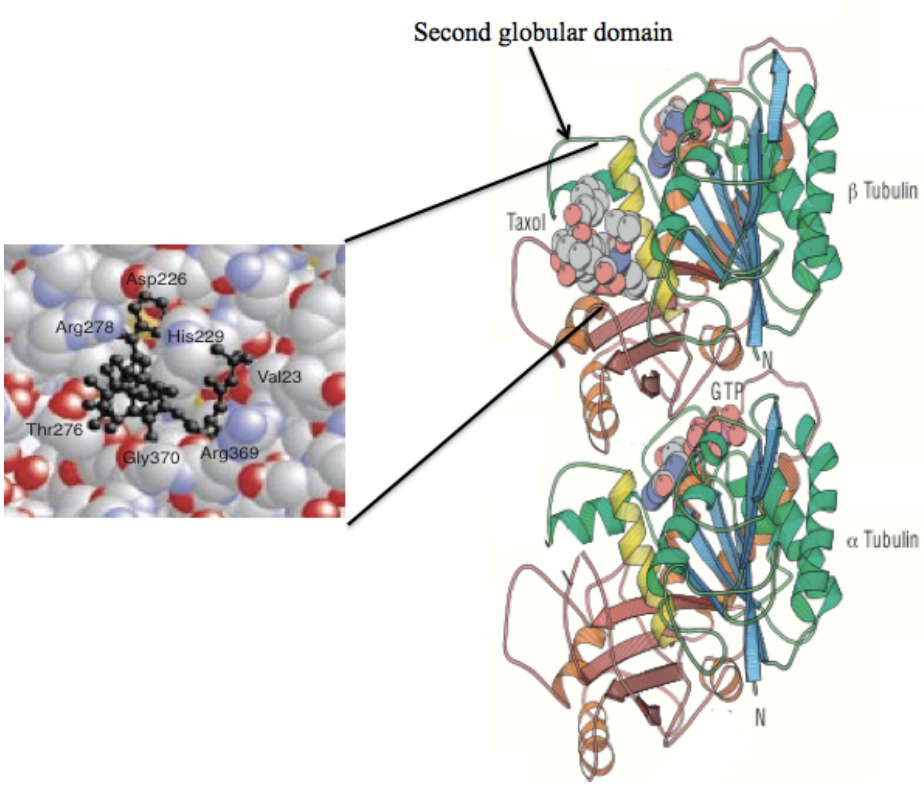


Fig 1.4 [Adapted from (Amos and Lowe 1999)] Right panel: The structure of a tubulin dimer: α and β tubulin. Left panel: the amino acids that are in closest contact with paclitaxel.

1.3.2.3 Paclitaxel biosynthetic pathway

In order to safeguard against potential shortages, as well as bring down the cost of cancer treatment, the production of a larger amount of paclitaxel would be required. This could be achieved by either semi-synthetic approaches or by genetically engineered biosynthetic methods. Nevertheless, it would be helpful to elucidate the complete paclitaxel biosynthetic pathway in *Taxus* spp.

The first step of the paclitaxel biosynthetic pathway is the cyclisation of geranylgeranyl diphosphate (GGPP). GGPP is a universal diterpenoid progenitor, which also serves as starting materials for many other biosynthetic pathways (Guerra-Bubb et al. 2012). GGPP is a diterpene and formed by condensing three IPP units with dimethylallyl diphosphate (DMAPP). This takes place in plant plastids and geranylgeranyl diphosphate synthase (GGPPS) is responsible for this condensation reaction (Lange and Ahkami 2013). Its cDNA was cloned from a methyl jasmonate (MeJA)-induced *Taxus canadensis* gene library. The recombinant protein was expressed in yeast, and radiochromatographic analysis confirmed its catalytic activity. It supplies the precursor for paclitaxel synthesis in *Taxus canadensis* suspension cell cultures (Hefner et al. 1998). The enzyme taxadiene synthase (TASY), is responsible for converting the linear GGPP to the parent hydrocarbon taxa-4 (5),11(12)-diene. Along with the production of the primary product (taxa-4 (5),11(12)-diene) of TASY, the production of taxa-4 (20),11(12)-diene and taxa-3 (4),11(12)-diene are only 5% and less then 1% of the total products, respectively. This is the first committed step of paclitaxel synthesis, which is a slow, but not the rate-limiting step of the pathway (Hezari et al. 1997). The cDNA encoding TASY was cloned from a cDNA library constructed from poly (A)+ RNA extracted from the Pacific yew (*Taxus brevifolia*) stems by a homology-based PCR cloning method (Wildung and Croteau 1996). The recombinant protein of TASY has been functionally expressed in *E. coli* and shown to resemble the native enzyme in kinetic properties (Williams et al. 2000). Furthermore, the X-ray crystal structure of a truncated version of TASY (TXS) was elucidated by Koksal and coworkers (Fig 1.5) (Koksal et al. 2011). The first hydroxylation reaction in the paclitaxel pathway is the conversion of both taxa-4 (5), 11(12)-diene and taxa-4 (20), 11(12)-diene into taxa-4

(20),11(12)-dien-5α-ol by taxadiene-5α-hydroxylase (T5αH) (Jennewein et al. 2004a). Following this hydroxylation, the next step in the pathway is the acetylation reaction at C5, in which taxa-4 (20), 11(12)-dien-5α-ol is converted into taxa-4 (20), 11(12)-dien-5α-acetate. The corresponding enzyme is taxa-4 (20), 11(12)-dien-5α-ol-*O*-acetyl transferase (TAT). Its cDNA was cloned from induced *Taxus canadensis* and expressed in *E. coli* by the Croteau lab (Walker et al. 2000; Wheeler et al. 2001). Two more predicted hydroxylation steps occur in the pathway: 13α- and 10β-hydroxylations. The two corresponding enzymes, taxane 13α-hydroxylase and taxane 10β-hydroxylase, have been fully characterised (Jennewein et al. 2001; Kaspera and Croteau 2006). Although taxane 13α-hydroxylase (T13αH) shares 63% identity with taxane 10β-hydroxylase (T10βH) in amino acid sequence, the former prefers taxa-4 (20),11(12)-dien-5α-ol as it substrate, the latter one uses taxa-4 (20),11(12)-dien-5α-acetate as the substrate for the production of taxa-4 (20),11(12)-dien-5α, 13α-ol and taxa-4 (20),11(12)-dien-5α-yl-acetate-10β-diol, respectively (Jennewein et al. 2001). However, the formation of 10-deacetylbaccatin III (10-DAB III) from these two hydroxylated products in the pathway remains unknown. Furthermore, 10-deacetylbaccatin 10β-*O*-acetyltransferase (DBAT) was found to catalyse acetylation of 10-DAB III at C10 into baccatin III. The cDNA of *DBAT* was amplified from a cDNA library constructed from mRNA of methyl jasmonate-induced *Taxus* cells by a homology-based PCR cloning strategy. The recombinant protein was expressed in *E. coli*, and the analysis by ^{1}H-nuclear magnetic resonance (NMR) and mass-spectrometry (MS) verified its function (Walker and Croteau 2000). The final steps in the pathway include C13 side-chain construction and implementation. In the process of C13 construction, phenylalanine aminomutase (PAM) first converts *2S*-α-phenylalanine to *3R*-β-phenylalanine, which is further linked with one molecule of CoA by amino phenylpropanoyl CoA ligase in the presence of ATP and CoA, resulting in β-phenylalanoyl CoA (Fig 1.6b). The cloning, heterologous expression and functional characterisation of a PAM-like enzyme from a *Taxus cuspidata* cDNA library in *E. coli* were accomplished by the Croteau group (Walker et al. 2004). The constructed C13 side-chain, the acyl donor, is then fused to Baccatin III by taxoid C13-*O*-phenyl-propanoyltransferase (BAPT), forming N-debenzoyl-2'-deoxytaxol. The function of BAPT was also verified by the Croteau lab (Walker et

al. 2002a). The final acylation step in paclitaxel pathway is catalysed by 3'-*N*-debenzoyl-2'-deoxytaxol *N*-benzoyltransferase (DBTNBT), forming 2'-deoxytaxol from the coupling of the substrate *N*-debenzoyl-(3'*RS*)-2'-deoxytaxol with benzoyl-CoA (Walker et al. 2002b). Finally, the benzamidation of 2'-deoxytaxol gives rise to paclitaxel and marks the end of the paclitaxel biosynthetic pathway (Croteau et al. 2006; Guerra-Bubb et al. 2012).

Overall, the cyclisation of GGPP to form the taxane skeleton, 8 P450-mediated oxygenations, three CoA-dependent acyl/aroyl transfers, oxetane ring formation to yield Baccatin III and five additional steps for the functional C13 side-chain construction and implementation constitute the biosynthetic pathway of paclitaxel in *Taxus* spp. (Fig 1.6). Furthermore, a targeted molecular approach has been the major method used to identify the structural genes involved in a biosynthetic pathway by different groups. For example, a homology-based PCR cloning strategy was used by the Croteau group to identify structural genes encoding enzymes functionally associated with cyclisation, oxygenation and acylation of GGPP to yield paclitaxel from a cDNA library generated from mRNA of MeJA-induced *Taxus* cells (Croteau et al. 2006; Kaspera and Croteau 2006). Table 1 summarises some functionally important genes involved in the paclitaxel pathway, which have been cloned, heterologously expressed and characterised so far through targeted molecular approaches.

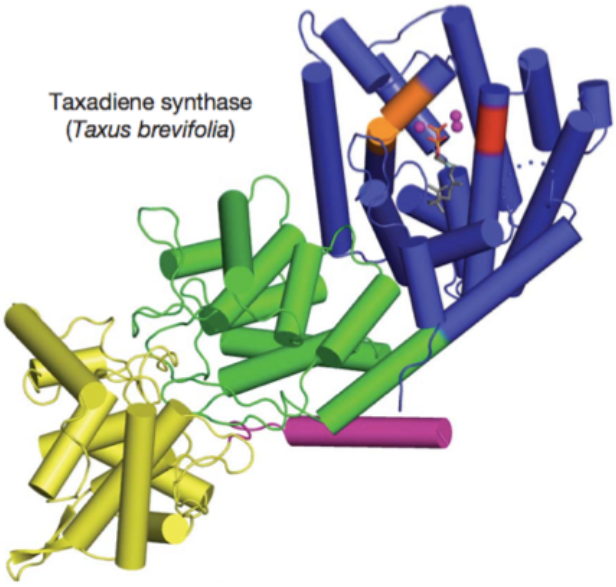


Fig 1.5 (Koksal et al. 2011) X-ray crystal structure of a truncated version of taxadiene synthase of *Taxus brevifolia*.

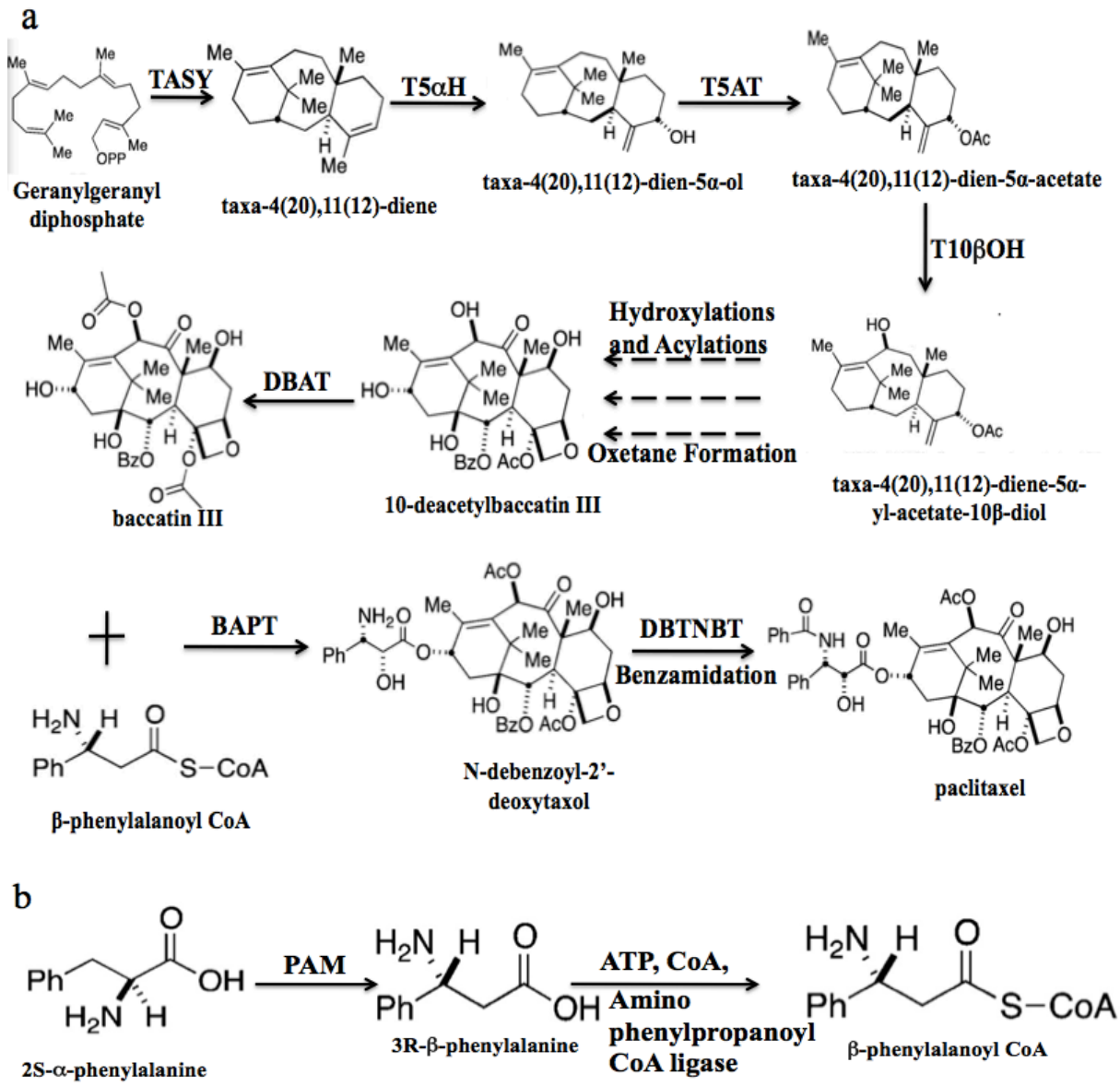


Fig 1.6 (Croteau et al. 2006; Guerra-Bubb et al. 2012; Jennewein et al. 2004b) (a) A simplified paclitaxel biosynthesis pathway. (b) Synthesis of C-13 side chain, β-phenylalanoyl CoA. Dotted lines represent unknown enzymatic steps. The brackets indicate a double bond between two carbon atoms. *TASY*: taxadiene synthase; *T5αH*: taxadiene-5α-hydroxylase; *T5AH*: taxa-4 (20), 11(12)-dien-5α-ol-*O*-acetyl transferase; *T10βH*: taxane 10β-hydroxylase; *DBAT*: 10-deacetylbaccatin 10β-*O*-acetyltransferase; *PAM*: phenylalanine aminomutase; *BAPT*: taxoid C13-O-phenylpropanoytransferase; *DBTNBT*: 3'-*N*-debenzoyl-2'-deoxytaxol *N*-benzoyltransferase.

Table 1 Characterisation of functionally important structural genes involved in the paclitaxel biosynthesis pathway.

Genes	ORF (bp)	Protein size (aa)	cDNA library	Expression host	Reference
GGPPS	1182	393	*T. canadensis*	Yeast	(Hefner et al. 1998)
TASY	2586	862	*T. brevifolia*	*E. coli*	(Wildung and Croteau 1996)
T5αH	1509	502	*Taxus cell*	Yeast and insects	(Jennewein et al. 2004a)
T5AT	1317	439	*Taxus cell*	*E. coli*	(Walker et al. 2000)
TBT	1320	438	*Taxus cell*	*E. coli*	(Walker et al. 2002b)
T7βH	1503	499	*Taxus cell*	Yeast	(Chau et al. 2004)
T2αH	1488	495	*Taxus cell*	Yeast	(Chau and Croteau 2004)
T13αH	1458	484	*Taxus cell*	Yeast	(Jennewein et al. 2001)
T10βH	1494	498	*Taxus cell*	Yeast	(Schoendorf et al. 2001)
DBAT	1320	440	*Taxus cell*	*E. coli*	(Walker and Croteau 2000)
DBTNBT	1323	441	*T. cuspidata*	*E. coli*	(Walker et al. 2002b)
BAPT	1335	445	*T. cuspidata*	*E. coli*	(Walker et al. 2002a)
PAM	2094	698	*T. cuspidata*	*E. coli*	(Walker et al. 2004)

GGPPS: geranylgeranyl diphosphate synthase; *TASY*: taxadiene synthase; *T5αH*: taxadiene-5α-hydroxylase; *DBAT*: 10-deacetylbaccatin 10β-*O*-acetyltransferase; *PAM*: phenylalanine aminomutase; *BAPT*: taxoid C13-O-phenylpropanoytransferase; *DBTNBT*: 3'-*N*-debenzoyl-2'-deoxytaxol *N*-benzoyltransferase; *T10βH*: taxane 10β-hydroxylase; *T13αH*: taxane 13α-hydroxylase; *T2αH*: taxoid 2α-hydroxylase; *T7βH*: taxoid 7β-hydroxylase; *TBT*: 2*α*-hydroxytaxane 2-O-benzoyltransferase. *T5AH*: taxa-4(20),11(12)-dien-5α-ol-*O*-acetyl transferase.

1.3.2.4 Medical applications of paclitaxel

Paclitaxel as a secondary metabolite has no known function in *Taxus* spp. However, it may be a defence compound inhibiting the growth of plant pathogenic fungal hyphae in the plant, because paclitaxel is found to inhibit the growth of several fungi, such as *Phytophthora* and *Pythium species* (Oomycetes) and *Rhizoctonia solani* (Basidiomycetes) (Flores 1994). Historically, it has been an important drug in the treatment of various cancers (Chu et al. 2005; Hammond et al. 2000; Perez 1998; Rowinsky et al. 1992). Evidence indicates that it kills cancer cells through the induction of apoptosis, and the accumulation of H_2O_2 is an early and crucial step in this induced cancer cell death (Alexandre et al. 2006; Wang et al. 2000). Apart from its clinic use in cancer treatment, paclitaxel is proved to function in the mature central nervous system (CNS) by facilitating axon regeneration (Sengottuvel et al. 2011). In addition, it could be an adjuvant candidate in the process of promoting both Th1 and Th2 immune responses induced by ovalbumin in mice (Yuan et al. 2010). Therefore, paclitaxel has a broad-spectrum application in human disease treatment. Furthermore, the delivery of paclitaxel is problematic, because of lack of selective cytotoxicity towards cancer cells; thus, cancer cells as well as normal cells may be affected, and it leads to some side effects. Another major problem that affect wide clinic application of the drug is its poor water solubility. However, the coupling of paclitaxel with delivery-facilitating molecules can ease and enhance the targeted delivery of paclitaxel to its target cancer cells (Gibson et al. 2007) . For example, the glycan conjugation approach enhances the delivery of glycan-based paclitaxel into cancer cells that overexpress glucose transporters (GLUTs) (Lin et al. 2008).

1.3.2.5 Traditional strategies for paclitaxel synthesis

After paclitaxel was approved for its clinic use in cancer treatment, the demand for this anticancer drug dramatically increased. BMS earned $1.5 billion from selling paclitaxel worldwide in 1999 (Malik 2011) and the total market for it was $1 billion per year worldwide (http://www.strategyr.com/Bulk_Paclitaxel_Market_Report.asp).

In particular, paclitaxel is in demand as a drug in the chemotherapy of ovarian cancer and its shortage can have a significant negative impact on the average cost of primary treatment for ovarian cancer. On average, it costs $4939 for 6 cycles of chemotherapy in the standard treatment, but $16,107 in the drug shortage scenario. So, the cost difference is $11,1168 per patient over 6 cycles of treatment (Havrilesky et al. 2012). There have been different strategies developed for large-scale production of paclitaxel, including extraction, chemical synthesis and fermentation of genetically modified microbials (Ajikumar et al. 2010; Wani et al. 1971; Yamin Wei 2012).

From 1967 to 1994, the production of paclitaxel was mainly from extracts of barks of *Taxus* spp. (Renneberg 2007). However, the drawbacks of this approach are: firstly, it is not an environment-friendly and renewable process, because the removal of barks from yew trees leads to the death of the tree and it takes hundreds of years for yew trees to grow and accumulate a large amount of paclitaxel; secondly, the amount of extracted drug is only about 0.01% of the dry weight of the bark. For example, 1200 kg of *Taxus* bark only yielded 10g of pure paclitaxel, which is about 8.3 mg per kg of bark (McCoy and O'Connor 2008). Purification steps further reduce the drug yield. As a consequence, the synthetic approach may be promising. Due to paclitaxel's complexity in chemical structure, the steps required in the route of its total chemical synthesis are 35 to 51, with a highest yield of 0.4%. Although paclitaxel can be chemically synthesised from its precursor, 10-DAB, which can be extracted from the renewable part of yew trees (needles or leaves), this semi-synthetic approach is still not economically efficient for drug production, because of limitations on productivity and scalability (Denis 1988; Suffness 1995; Nicolaou et al. 1994; Holton 1994). Therefore, the synthetic approach is not the best alternative for paclitaxel production. Furthermore, the production of antibiotics from genetically modified microbials on an industrial scale has been successful, for example, the production of artemisinic acid (the antimalarial drug precursor) in engineered yeast (Ro et al. 2006). This strategy could shed light on the production of paclitaxel by engineered microbials. Endophytic fungi isolated from the bark of *Taxus sp*. can naturally produce paclitaxel, such as EFY21 isolated from *Taxus*

chinensis var. mairei. After overexpression of *TASY,* the engineered EFY21 produce 5 times as much as paclitaxel than the wildtype (Yamin Wei 2012). Another example of microbial engineering for paclitaxel precursor overproduction is engineered *E. coli*, from which, the titers of the first committed paclitaxel intermediate (taxadiene) is increased by about 15,000 fold (up to 1 gram per liter) (Ajikumar et al. 2010). However, the downsides of fermentation technologies, such as large, expensive infrastructures and low and variable yields, make it economically less favorable (Malik 2011). Overall, the drawbacks each of these strategies have greatly limited the large-scale paclitaxel production.

1.3.3 Production of natural products by plant cell culture

The unreliable harvesting of high-value and low-abundance natural products from plants has great negative impacts not only on patient treatments and clinical evaluation requirements, but also on the environment. This has been a key driving force behind efforts towards the development of renewable and environment-friendly production strategies.

Plant cell culture, pioneered by a German botanist in 1898 (Thorpe 2012), provides a sustainable and environmentally friendly production platform for high-value natural products. It has several advantages over natural harvesting: firstly, the conditions for plant cell culture can be well controlled to meet the requirements for drug production; secondly, the culture conditions can be optimised to increase the yield of a given target drug. In addition, production and engineering of novel drugs can be achieved through synthetic biology approaches in plant cell culture (Kolewe et al. 2008). Finally, it provides renewable source of natural products, as cell harvesting for drug production does not damage the environment. Therefore, production of natural products by plant cell culture is a promising approach to meet the demand for producing pharmaceutically active natural compounds. Table 2 lists natural products produced by plant culture with their pharmaceutical applications (Caldentey 2002). Most of the drugs that are produced through plant cell culture have anticancer properties.

Table 2. [Adapted from(Caldentey 2002)] List of natural products with pharmaceutically applications produced through industrial cell culture.

Cell culture	Product	Application	Manufacturer
Taxus spp.	paclitaxel	antitumor	Phyton Biotech (USA) Samyang Genex (Korea)
Coleus blumei	rosmarinic acid	anti-inflammatory	Nattermann (Germany)
Panag ginseng	ginseng	dietary supplement	Nitto Denko (Japan)
Lithospermum erythrohizon	shikonin	anti-HIV, antitumor, anti-inflammatory	Mitsui Petrochemical Industries (Japan)
Geramineae spp.	geraniol	antitumor	Mitsui Petrochemical Industries (Japan)
Podophyllum	podophyllotoxin	antitumor	Nippon Oil (Japan)
Coptis japonica Thalictrum minus	berberines	antitumor	Mitsui Petrochemical Industries (Japan)

1.3.3.1 Production of paclitaxel by plant cell culture

Because of the limitations of traditional methods for paclitaxel production, researchers have successfully conducted production of paclitaxel by plant cell culture. Table 3 summarises plant cell cultures of *Taxus* spp. used for production of paclitaxel. Although plant cell culture for industrial production is still somewhat limited, industrial production of paclitaxel and its precursor by this approach has been scaled up; for example, bioreactors up to 75,000 L have been used by chemical companies, such as ESC genetics (USA), Phyton (USA) and Samyang Genex (Korea), for paclitaxel production (Caldentey 2002). Furthermore, there have been several strategies developed to improve paclitaxel yields in *Taxus* spp. plant cell cultures, including elicitation, selection of high paclitaxel-accumulating *Taxus* spp. cell lines, culture media and environmental optimisation. These non-molecular approaches are collectively known as traditional strategies. On the basis of molecular and biochemical technologies, newer strategies have been developed and employed to gain a complete understanding of the paclitaxel synthesis, such as differentially expressed genes (DEGs) (Hao da et al. 2011).

Table 3. [Adapted from (McCoy and O'Connor 2008)] Paclitaxel production by MeJA-treated plant cell culture of *Taxus* spp. The concentration of the extracellular paclitaxel is measured in mg per litre of culture media.

Cell type	Paclitaxel (mg/L)	Reference
Taxus canadensis	117	(Parsons and Shuler 2002)
Taxus media	21	(Matsuda et al. 1991)
Taxus media	110	(Kayser 2007)
Taxus cuspidata	22	(Doran 2002)
Taxus chinensis	138	(Jennewein et al. 2004b)

1.3.3.2 Increased paclitaxel yield in cell culture by elicitation

An elicitor is defined as any biotic or abiotic factor that leads to an increased yield of secondary metabolites. Fig 1.7 summarises elicitors used for increased production of paclitaxel and precursors in *Taxus* spp. cell suspension culture.

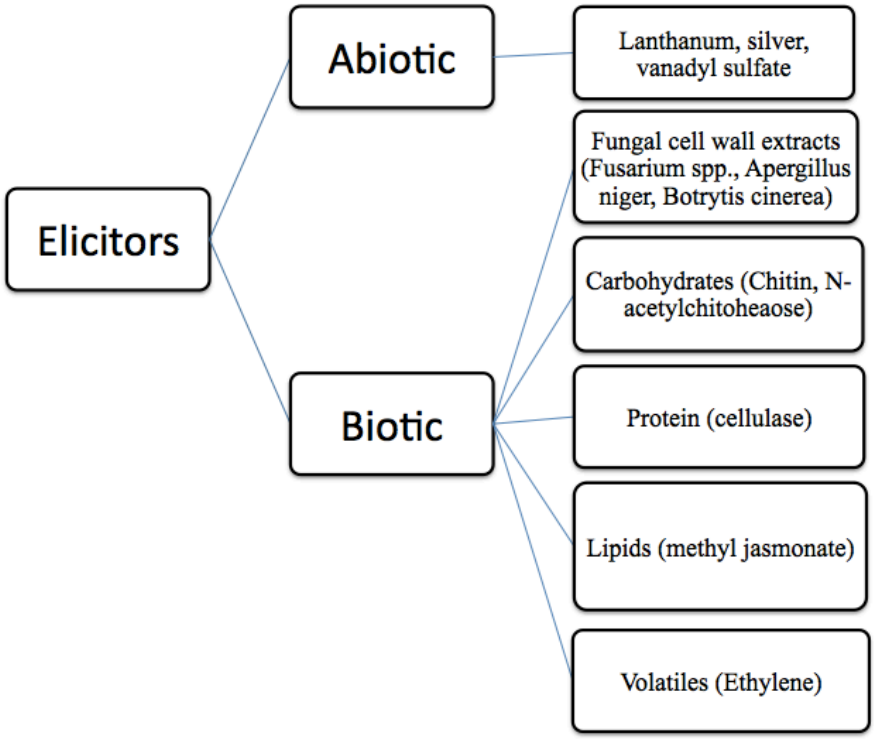


Fig 1.7 [adapted from (Vongpaseuth and Roberts 2007)] A summary of abiotic and biotic elicitors used to increase paclitaxel and its precursors production in *Taxus* spp. cell suspension culture.

Among these elicitors, MeJA, which has been shown to induce production of a range of secondary metabolites from alkaloids in *Catharanthus*, anthocyanins in soybean seedlings to rosmarinic acid in *Lithospermum* cell suspension cultures (Aerts 1994; Franceschi and Grimes 1991; Mizukami 1993), acts as key signal transducer and plays an important role in the production of secondary metabolites (Gundlach et al. 1992). Furthermore, it has been revealed that MeJA can induce overproduction of paclitaxel and Baccatin III in *Taxus* cell suspension cultures; in addition, two regions of MeJA (C_1 and C_3 position) are known to be important for this induction, because

the promoting activity of MeJA is greatly reduced by either the reduction of the keto group at the C_3 position or the deletion of the carboxyl group at the C_1 position (Yukimune et al. 1996). Interestingly, paclitaxel, the end product of paclitaxel biosynthetic pathway, was shown to stimulate cell-associated paclitaxel content by up to 32.7 fold in the presence of 200 mg/L exogenous paclitaxel in *Taxus baccata* suspension cultures (Exposito et al. 2009).

Furthermore, paclitaxel yield can also be improved by treating cell suspension culture with more than one elicitor, which is known as combinatorial elicitation. For example, in studies eliciting *Taxus chinensis* cell culture with a combination of 50mg chitosan l^{-1}, 60μM methyl jasmonate and 30μM Ag^{+}, the paclitaxel yield was found to be significantly higher than that of the control (almost 40 times) and the culture with Ag^{+} (10 times), chitosan (6 times) or methyl jasmonate (twice), respectively (Zhang 2000).

1.3.3.3 Increased paclitaxel yield in cell culture by molecular-based methods

The goal of molecular investigations to globally elucidate paclitaxel metabolism and accumulation is to bioengineer a superior paclitaxel-producing *Taxus* spp. cell line by transforming *Taxus* cells with rate-limiting structural genes involved in and outside of paclitaxel biosynthetic pathway, such as genes involved in the paclitaxel-accumulating process (transport and degradation). There are two different molecular approaches applied toward this goal: one is a targeted approach and it focuses on elucidating the structural genes and corresponding enzymes involved in paclitaxel biosynthesis in *Taxus* spp. A homology-based PCR cloning strategy has been used to identify the majority of structural genes and gene families and corresponding enzymes responsible for the enzymatic steps in the pathway, such as terpene cyclases, cytochrome P450 hydroxylases and acyl transferases; the other is a non-targeted approach. This molecular strategy focuses on not only elucidation of structural genes in the pathway, but also genes responsible for the unknown mechanisms outside the pathway. The main idea behind this approach is to construct two cDNA libraries: a cDNA library of paclitaxel-accumulating and non-paclitaxel-

accumulating state in *Taxus* spp. cell cultures i.e. a cDNA library of differentially expressed transcripts between these two libraries. In conjunction with genetic information obtained from traditional methods, both structural and regulatory genes with putative functions in and outside the paclitaxel pathway can also be identified (Vongpaseuth and Roberts 2007). This new strategy has been successfully employed in expression profiling of DEGs with *Nicotiana tobaccum* and *Arabidopsis thaliana* (Costa et al. 2000; Imanishi et al. 1998).

1.3.3.4 Contemporary limitations on plant cell culture for industrial utility

Traditionally, the establishment of plant cell suspension cultures begins with plant calli. Plant calli are undifferentiated somatic cells and they are initiated from differentiated material (explants) and maintained in a dedifferentiated state by exogenous phytohormones (auxins and cytokins) on solid medium. Plant cell suspension cultures are generated when calli are suspended in liquid media, in which the callus breaks into small cell aggregates approximately 40-200 μm in diameter. Compared to callus culture, cell suspension culture proliferates more rapidly, and therefore is suitable for large-scale industrial culture. In contrast to undifferentiated cells present in callus culture, hairy root cultures are established after plants infected with *Agrobacterium rhizogenes* and comprise differentiated cells. Hairy root cultures are especially utilised for production of certain natural products that can only be robustly synthesised in differentiated tissues (Shanks and Morgan 1999).

There are several challenges presented by plant cell cultures. The first obstacle present in the routine industrial utility of plant cell cultures for natural product production is the variability of natural product yield. Compared to the utility of microbial cultures for target compound production, where the yield is generally stable, large variations in product yield in plant cell culture system often limits their potential utility in industry. For example, paclitaxel accumulation between different cell lines of the same species is markedly different (Table 3) (Ketchum 1996). The second challenge that limits the potential commercial utility of plant cell culture is the existence of large cellular aggregates in suspension cultures when cultured in

large bioreactors. Plant cells naturally tend to aggregate, because the pectin-rich middle lamella bonds cell together after cell division. As a consequence, a spectrum of diverse heterogeneous microenvironments is established within these cellular aggregates that differ in size. Thus, there are metabolic and morphologic differences between cells at the centre and near the periphery of the aggregates (Naill and Roberts 2004). In addition, the formation of those cellular microenvironments may contribute to the striking variability often observed in natural product production from plant cell culture. Aggregate size affects paclitaxel accumulation in *Taxus* suspension cultures. For example, the *T. cuspidata* P991 culture with mean aggregate size of 400 μm produced 6 mg/L paclitaxel; in contrast, the same culture with mean aggregate size of 840 μm produced only 0.3 mg/L paclitaxel. So, there is a 20-fold increase in paclitaxel production in smaller aggregates (Kolewe et al. 2011). The third obstacle that limits the application of plant cell culture on an industrial scale is their sensitivity to mechanical stresses, such as sheer stress. Cell aggregation further intensifies this trait even in a sheer stress reducing bioreactor (stirred tank) (Joshi 1996).

1.3.3.5 Isolation and culture of plant cambial meristematic cells

As the establishment of plant cell suspension culture begins with initiation of calli from differentiated plant material through what was believed to be a dedifferentiation process, the type of cells proliferating in the resulting cell suspension culture is a dedifferentiated cell line (DDC) (Thorpe 2007). The differentiated plant cell is reversed into a "stem cell"-like state in this process when exposed to plant hormones auxin and cytokinin. The "stem cell"-like (dedifferentiated) state makes it easy for the dedifferentiated cells to re-enter into cell cycle and subsequent cell proliferation. Ultimately, a multicellular mixture of proliferating cells is generated from dedifferentiation of specialised cell types (Grafi et al. 2007). Cell suspension cultures established from such cellular assortments often show some deficient traits, such as poor growth properties, low and variable yields of natural products (Roberts 2007), this might be because of deleterious genetic and epigenetic changes that happen during the dedifferentiation process (Fereol et al. 2005). So, the dedifferentiated cell

line has not been a good source for the establishment of plant cell cultures on an industrial scale (Yun et al. 2012).

Recently, an innately undifferentiated cell line derived from cambium cells has been isolated and known as cambial meristematic cells (CMCs). This actively proliferating cambium cells can be naturally separated from phloem, cortex and epidermis (Fig 1.8). The cambium cells, from which CMCs are derived, function as vascular stem cells, and differentiate to form secondary xylem and phloem on the inside and outside of the ring, respectively. CMCs were first isolated from *T. cuspidata*, and the antineoplastic agent paclitaxel is most abundant within the region containing the CMCs (Stierle et al. 1993). This novel procedure for isolation of CMCs is described in Fig. 8 (Yun et al. 2012). Apart from CMCs of *T. cuspidata*, CMCs from a variety of plant species were also isolated, including ginseng (*Panax ginseng*), ginkgo (*Ginkgo biloba*) and tomato (*Solanum lycopersicon*). This suggests that this CMCs isolation technology can be broadly utilised.

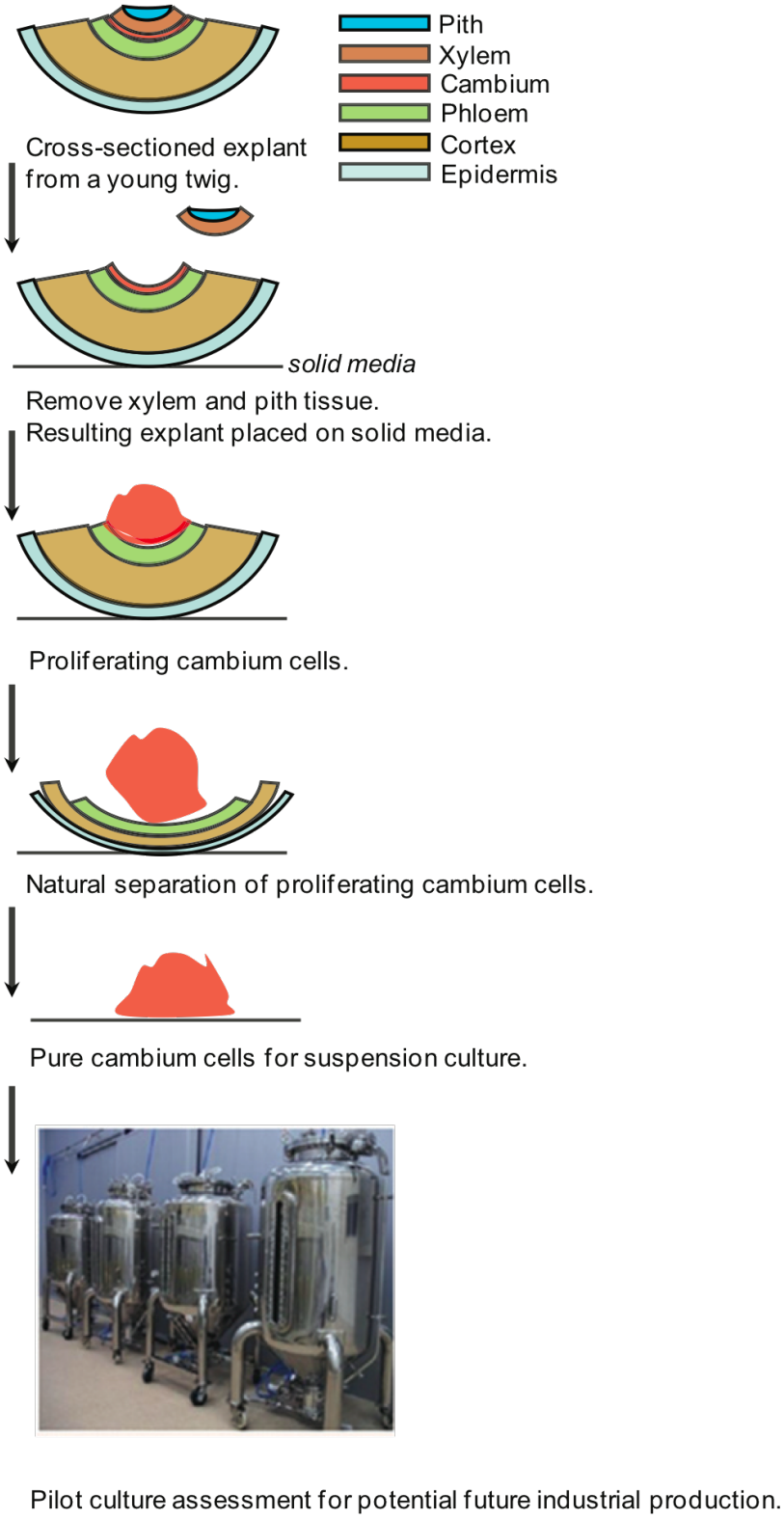


Fig 1.8 (Yun et al. 2012) The procedures for isolating and culturing CMCs.

1.3.3.6 CMCs versus DDCs in performance during small and large-scale culture

It has been shown that the growth of CMCs in suspension cultures is far superior to DDCs in several aspects. The dry cell weight (d.c.w) of CMCs derived from either needles or embryo of *T. cuspidata* was greater than that of standard *T. cuspidata* DDCs by a magnitude of 10^5 after subculturing both cell types every two weeks for 22 months. In addition, after 22 months of culturing, CMCs were still growing rapidly; in contrast, DDCs developed prominent necrotic patches and showed signs of a rapid decrease in their biomass. Secondly, DDC suspension culture often produces pronounced cell aggregates, which is a common feature of DDC containing suspension culture. This can lead to the formation of different microenvironments between cells within these aggregates. Ultimately, DDC cultures show reduced growth rates and variable and decreased natural product production (Lee et al. 2010). Typically, in CMC cultures, there is a large majority (93%) of cell aggregates having a diameter of <0.5 mm and many cells exist as singletons. Conversely, in DDC cultures derived either from needles or embryo of *T. cuspidata*, the proportion of cell aggregates with a diameter less than 0.5 mm was only 2% or 5%, respectively. Thirdly, on a large scale culturing, the growth of plant cells is limited by shear stress and cell aggregation further potentiates this problem. It has been shown that *T. cuspidata* CMCs in cultures grow significantly faster than DDCs in a 10 L stirred tank bioreactor. In response to shear stress, the survival rate of CMCs is much greater than that of DDCs. Further, the growth of CMCs in a 20 L air-lift bioreactor, which is often used for production of natural products on an industrial scale, is also found to be noticeably greater than that of DDCs, in contrast to CMCs, DDCs fail to grow. Due to the unique physiological and molecular characteristics of CMCs, including small and abundant vacuoles, reduced aggregation and thin cell walls, CMCs can tolerate greater shear stress than DDCs(Lee et al. 2010). Therefore, CMCs is a superior cell line to DDCs when utilised for natural product biosynthesis. Finally, on an industrial scale, the stability of cell growth in culture is often an important trait for the exploitation of plant suspension cells. Over the time of cell culturing, CMCs showed a relatively constant growth rate, whereas the growth rate of DDCs widely fluctuated (Lee et al. 2010; Yun et al. 2012).

1.3.3.7 Utility of CMCs for production of natural products

T. cuspidata CMC line has been proved to perform much better than its DDC line when cultured on a large scale. So, the utility of CMCs for natural product biosynthesis on an industrial scale could be established. A promising example is the use of CMCs for paclitaxel production.

When *T. cuspidata* cells are treated with fungal elicitor or MeJA, the plant immune activator, it induces the synthesis of paclitaxel. It has been reported that the addition of fungal elicitor can induce an increase in paclitaxel production by 14,000 fold in CMCs cultured in a 3 L air-lift bioreactor. In contrast, the fold induction of paclitaxel production for elicited cultures of DDCs derived from needle and embryo is only 220 and 433, respectively. Further, paclitaxel can be secreted into the culture media, and the amount of paclitaxel secreted from the cultured cells varies significantly between *Taxus* species and also depends on culture conditions. Therefore, culture conditions are often optimised to maximize the secretion of natural products synthesised in cultured plant cells into culture media. For example, perfusion culture (Lee et al. 2010), in general, promotes the secretion of natural products into culture media so that those compounds can be relatively easily purified. Because some natural compounds may be toxic to the cells when accumulated to a certain threshold inside the cell, the secretion into culture media can promote cell survival. After CMCs were grown in perfusion culture for 45 days, a total of 264 mg of paclitaxel per kg of cells was produced and 74% of this was secreted directly into the medium. In contrast, there was only 2-3 mg of paclitaxel secreted by DDCs derived from needle and embryo of *T. cuspidata*. Therefore, culturing CMCs with optimised conditions would aid the purification of paclitaxel, hence improve product yield (Yun et al. 2012). In addition to the diterpenoid, paclitaxel, *Taxus* species can also produce abietane tricyclic diterpenoid derivatives, such as taxamairin A and C. Both have been shown to have anti-cancer properties (Yang 1999). It was found that elicited CMCs produced 4983 and 521mg/kg/fresh cell weight (f.c.w) of taxamairin A and C in a 3 L air-lift bioreactor, respectively. Conversely, only trace of both was produced in DDCs. As might be expected, CMCs is a better source for production of these abietanes than DDCs.

More broadly, like *T. cuspidata* CMCs, CMCs can also be isolated from other medical plants. For example, *Panax ginseng* CMCs have been shown to synthesis a significant amount of ginsenosides, which are a class of triterpenoid saponins and produced exclusively by the plant genus *Panax*. This class of compounds has been shown to have medical values, such as neuroprotection and antioxidant effects (Leung and Wong 2010). When cultured in a 3 L air-lift bioreactor, CMCs derived from *Panax ginseng* performed much better than DDCs as expected in terms of production of ginsenosides F2 and gypenoside XVII. Strikingly, it was reported that CMCs produced 23.8- and 24.1-fold more ginsenosides F2 and gypenoside XVII than DDCs, respectively. As a consequence, CMCs potentially might be a good source for other medically important natural products (Yun et al. 2012).

In summary, the isolation of CMCs and their subsequent culture on an industrial scale could provide an improved source for not only biomedical drugs, but also plant natural products that are used as pigments, insecticides and antimicrobials (Schmidt et al. 2007). More importantly, this technique may also offer a sustainable, cost-effective and environmentally friendly way to produce important natural products for human utilities.

1.4 *Taxus* species

1.4.1 Botany of *Taxus*

Taxus is commonly known as yew and is a genus of gymonosperms, as its seeds do not develop in an ovary. There are 24 species and 55 varieties in the genus *Taxus*. They are distributed across the northern temperate and subtropical regions as far south as El Salvador in Central America and Sumatra in Southeast Asia. Based on the leaf epidermal and stomatal features, the species in the genus *Taxus* are divided into three groups: the *Wallichiana,* the *Baccata* Group and the *Sumatrana* Group. Table 4 summarises the information about species included in each group and their geographic distributions. Furthermore, the *Wallichiana* group is divided into two subgroups: *Wallichiana* and *Chinensis,* and *Baccata* group is divided into the *Baccata* and *Cuspidata* species alliances (Spjut 2010).

Table 4 (Spjut 2010) Number of species and their geographic distributions of three groups of the *Taxus* genus

Groups of *Taxus* genus	Number of species	Geographic distribution
Wallichiana	11	Central Himalayas, Indonesia (Sumatera, Sulawesi), the Philippines, North America in the Pacific Northwest, Mexico, central America and an isolated occurrence in Florida
Baccata	9	Temperate Eurasia, northern Africa and eastern North America
Sumatrana	4	Asia, but absent from North America

The name *Taxus* is derived from the Greek word for yew, toxos, which is thought to be originally from the combination of the Greek words toxicon and toxon with the meaning of poison and bow, respectively. *Taxus* belongs to the yew family Taxaceae which are coniferous trees, as their leaves are similar to that of conifers (pines). The

heights of mature trees range from 1 to 40 m, with trunks up to 4 m in diameter. The leaves of yew are dark-green and their barks are reddish. Yew trees are slow growing, and it takes about 200 years for yew to reach their nature maximum height (Suffness 1995).

1.4.2 *Taxus cuspidata*

Taxus cuspidata (Fig 1.9), also known as the Japanese yew, is within the genus *Taxus*, the family Taxaceae and the order Taxales. It is native to Asian countries, including Japan and China, in addition to North and South Korea and the Russian Federation (Kuril, Primorye, Sakhalin) (IUC 2010, http://www.iucnredlist.org/). It is a small coniferous tree with a pyramidal shape, and it can grow up to 16 m tall (Hartzell 1991). Its shoots are red-brown, leaves are dark green and spirally arranged and cones are aril red when ripen (Ohwi 1965). In Japan, it was formerly valued for piles and foundations, water tank and chopsticks and was also medically used as an antidiabetic. Due to overexploitation, its commercial use has declined in recent decades (http://www.conifers.org/ta/*Taxus*_cuspidata.php). In 1998, *Taxus cuspidata* was reported to be in a lower risk status in the IUCN Red List of threatened species (IUC 2010, http://www.iucnredlist.org/).

Fig 1.9 *Taxus cuspidata* from Flora Japonica, Sectio Prima (Table band) (1870). (URL: http://caliban.mpiz-koeln.mpg.de/siebold/flora3/high/CRW_6726_RT8.html)

1.5 MeJA-induced gene expression

MeJA and its derivatives (jasmonates) are plant hormones derived from fatty acids (Creelman and Mullet 1997; Wasternack 2007). They have been found to play important roles in many developmental processes, including root elongation, fertility and sex determination (Acosta et al. 2009; Wasternack 2007). In addition, they can also activate plant defence systems to defend against both biotic (pathogens, insects and herbivores) and abiotic stresses (UV and drought) (Balbi and Devoto 2008; Browse and Howe 2008; Farmer et al. 2003). Transcriptional profiling of MeJA-treated cultured *Arabidospsis* cells suggested that MeJA transcriptionally reprogrammes cells to activate defence systems and, at the same time, arrest the cell cycle and subsequent growth (Pauwels et al. 2008; Reymond et al. 2000).

1.5.1 Jasmonate biosynthesis, activation and deactivation

The molecule, α-linolenic acid (18: 3) (α-LeA), which is released by lipase activity on chloroplast membranes, is the precursor for jasmonates. In fact, it is the precursor for a group of oxygenated compounds, collectively named oxylipins, including jasmonates. The initial step in the generation of MeJA is the formation of (13*S*)-hydroperoxyoctadecadienoic acid (13-HPOT) from α-linolenic acid by 13-lipoxygenase (13-LOX). Then, the conversion of 13-HPOT carried by 13-allene oxide synthase (13-AOS) results in an unstable allene oxide, 12,13-epoxylinolenic acid (12, 13-EOT). The subsequent enzyme, the allene oxide cyclase (AOC), catalyses the final chloroplast-located reaction in the pathway and leads to the AOC product *cis* (+)-12-oxophytodienoic acid (OPDA). Then, the OPDA is released from the chloroplast and transported into peroxisomes by an unknown mechanism, where OPDA is specifically reduced into OPC-8:0 by OPDA reductase 3 (OPR3). Through three steps of β-oxidation by acyl-CoA oxidase (ACX), a multifunctional protein (MFP) and a L-3-ketoacyl CoA thiolase 2 (KAT2), the unstable JA, (+)-7-*iso*-JA is formed, which is rapidly epimerised to stable epimer (-)-JA (JA) in peroxisomes (Fig 1.10a) (Wasternack 2007).

However, the stable JA is not biologically active in terms of promoting interaction between COI1 and JAZ proteins (section 1.5.2). It is further converted into a variety of biologically active derivatives through conjugation with amino acids (Fig 1.10b). In particular, JA conjugation with amino acid isoleucine (Ile) by Jasmonate Resistant 1 (JAR1) gives rise to the molecularly active hormone, (+)-7-*iso*-JA-$_{L}$-Ile (Fonseca et al. 2009b; Guranowski et al. 2007; Staswick and Tiryaki 2004) (Fig 1.10b). Furthermore, JA conjugation with other amino acids by JAR1, including Leu, Val and Phe, has also been reported to exist in other plant species (Wasternack 2007). In *Arabidopsis*, however, those amino acid conjugates are not biologically active molecules, because there is no detectable interaction between COI1 and JAZs promoted by any of those conjugates, and further none of them is able to rescue the JA-insensitivity of *jar1* mutants (Staswick and Tiryaki 2004). *In vitro*, the interaction between tomato homologues COI1 and JAZs has been shown to be promoted by JA-

Leu, JA-Val and JA-Ala, although their functionality remains to be shown *in vivo* (Thines et al. 2007). Overall, these data suggest that there are differences in specificities of hormone perception in different plant species.

In addition, the active JA hormones can be inactivated through different hormone inactivation mechanisms. For example, the active form of JA, (+)-7-*iso*-JA-$_{L}$-Ile, is reversibly epimerised into the inactive (-)-JA-L-Ile through alkali treatment. Further, its capacity is found to be reversibly reduced by methyl-esterification of (+)-7-*iso*-JA-$_{L}$-Ile (Fonseca et al. 2009b) (Fig 1.10b). Recently, JA hydroxylation and subsequent sulfonation and glycosylation have been reported to partially switch off JA signalling (Miersch et al. 2008). So, both activation and inactivation of JA work together to fine-tune JA signalling.

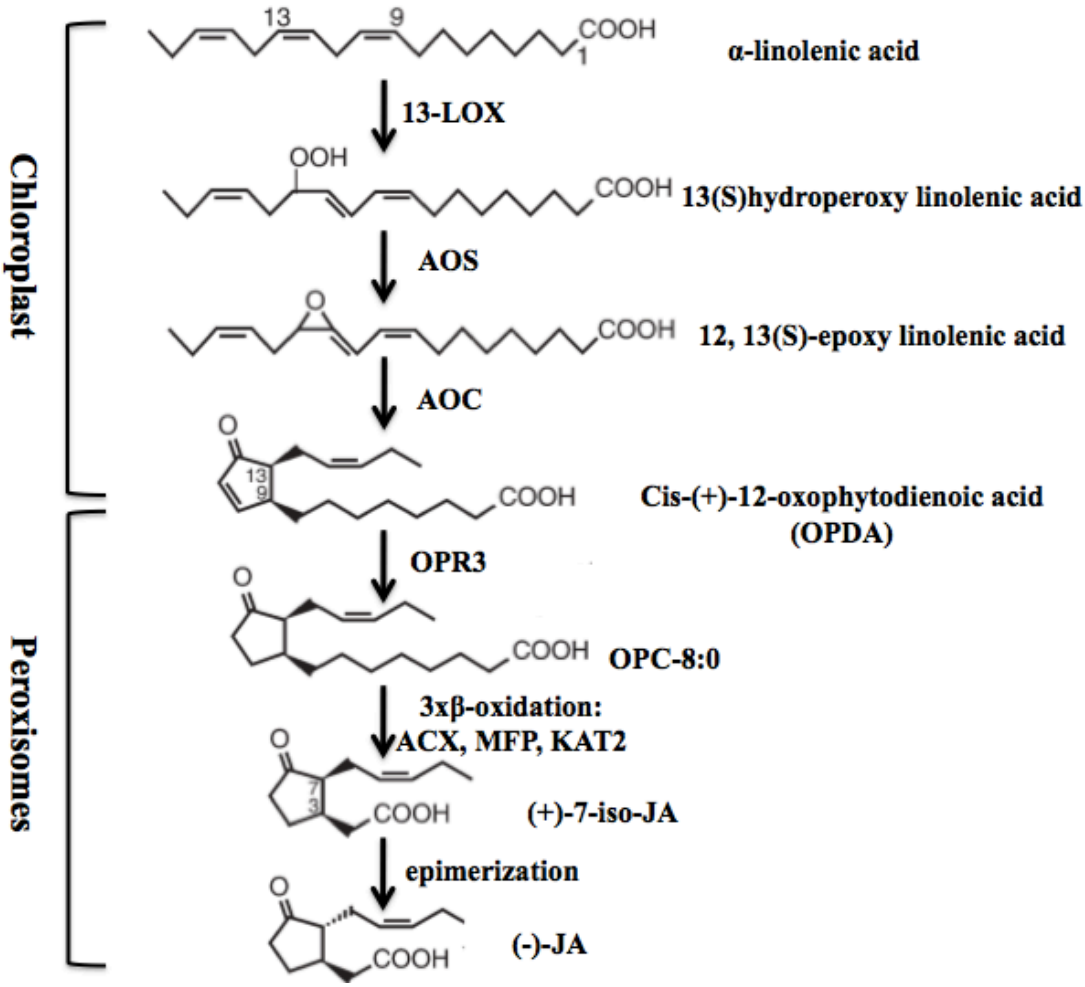


Fig 1.10a (Avanci et al. 2010; Wasternack 2007) JA biosynthetic pathway. The pathway first starts in chloroplast, in which α-linolenic acid is converted into *cis* (+)-12-oxophytodienoic acid (OPDA) through a series of enzymatic steps. In peroxisomes, *cis* (+)-12-oxophytodienoic acid is firstly reduced to OPC-8:0 by OPDA reductase 3 (OPR3). After three steps of β-oxidation by acyl-CoA oxidase (ACX), a multifunctional protein (MFP) and a L-3-ketoacyl CoA thiolase 2 (KAT2), the unstable JA, (+)-7-*iso*-JA is formed, which is rapidly epimerised to stable epimer (-)-JA (JA) in peroxisomes. 13-LOX: 13-lipoxygenase; AOS: allene oxide synthase; AOC: allene oxide cyclase.

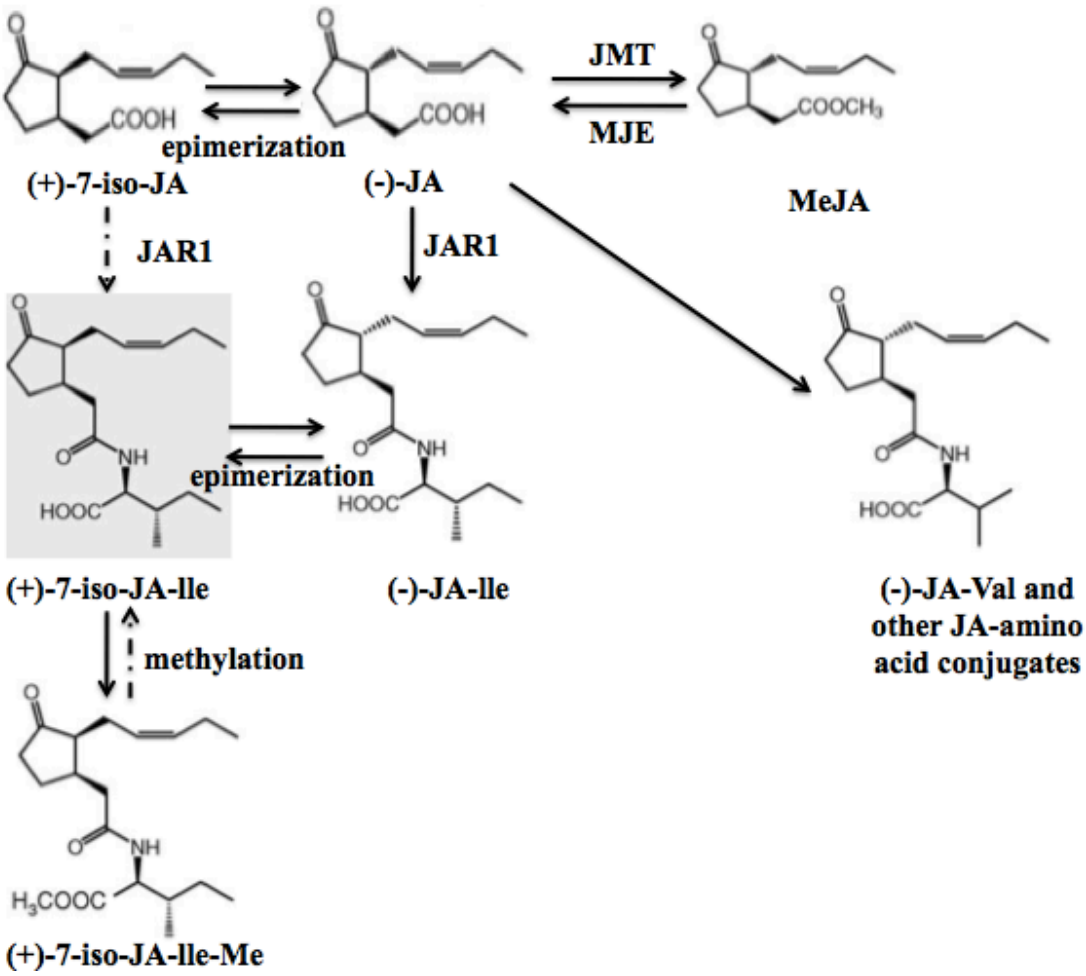


Fig 1.10b (Fonseca et al. 2009a) JA activation and deactivation. The stable JA can conjugate with amino acids to form a variety of derivatives. For example, JA conjugation with amino acid isoleucine (Ile) by Jasmonate Resistant 1 (JAR1) gives rise to the molecularly active hormone, (+)-7-*iso*-JA-$_{L}$-Ile (grey box). However, JA conjugation with other amino acids (leucine, valine and phenylalanine) is inactive. The active JA hormones can be inactivated through different hormone inactivation mechanisms. For example, the active form of JA, (+)-7-*iso*-JA-$_{L}$-Ile, is reversibly epimerised into the inactive (-)-JA-L-Ile through methylation or *in vitro* alkali treatment. JAR1: jasmonate resistant 1; JMT: Jasmonic acid carboxyl methylesterase; MJE: MeJA esterase. Dashed arrows indicate hypothetical modifications.

1.5.2 JA signalling pathway

JA signalling plays an important role during plant development and response to environmental stresses. The overview of the signalling pathway is illustrated in Fig 1.11A. The pathway comprises two important components: one is *COI 1*, which encodes an F-box protein as a component of a SKP1, Cullin, F-box protein E3 ubiquitin ligase ($SCF^{COI\ 1}$) (Xie et al. 1998). COI 1 has been known as a specificity factor that targets other proteins for degradation via the 26S proteasome (Devoto and Turner 2003); the other is the JAZ proteins, which are the targets of the $SCF^{COI\ 1}$ complex (Thines et al. 2007). In basal conditions, JAZs, which exist in the form of either homomeric or hetromeric complexes, interact with transcription factors (TFs) through the Jas domain and prevent TFs from activating JA-inducible genes. Under stress conditions, the active form of JA, (+)-7-*iso*-JA-$_L$-Ile, promotes the interaction between $SCF^{COI\ 1}$ and JAZs. As a consequence, JAZ proteins are ubiquitinated and degraded by the 26S proteasome, and TFs are released to activate transcription.

A well known example of JA signalling pathway involves the MYC2 basic Helix-Loop-Helix (bHLH) TF- based regulatory system in the JA-induced defence gene activation in *Arabidopsis* (Boter et al. 2004). In the absence of JA-Ile, MYC2, which binds to its target promoter through a G-box, is bound to JAZs through the Jas domain. The JAZ proteins inhibit the activation of MYC2 target genes by recruiting the corepressors TOPLESS (TPL) and TPL-related proteins through an interaction with an adaptor protein, known as Novel interactor of JAZ (NINJA) (Pauwels et al. 2010). Conversely, in the presence of JA-Ile and the cofactor, inositol pentakisphosphate (insP5, a molecule derived from inositol tetrakisphosphate by adding a phosphate group) (Sheard et al. 2010), JAZs are linked to COI 1 through the Jas domain, ubiquitinated and degraded by the 26S proteasome. This leads to the switch-on of JA-responsive gene expression (Fig 1.11B) (Pauwels and Goossens 2011). The re-establishment of MYC2 repression and restoration of the basal condition are completed by *de novo* synthesis of JAZ proteins. Interestingly, their corresponding *JAZ* genes are the early targets of MYC2 (Chini et al. 2007). MYC2

has been further reported to positively regulate a variety of genes involved in other processes, including oxidative stress response and flavonoid synthesis. Some other genes are negatively regulated by MYC2, such as genes for pathogen defence and tryptophan metabolism (Dombrecht et al. 2007).

The JA signalling pathway is regulated through both positive and negative feedback loops (Fig 1.12) (Farmer 2007). In the positive feedback loop (a), JA activates its own biosynthesis. This is supported by the fact that the genes encoding all the enzymes involved in JA biosynthesis are transcriptionally activated by JA (Wasternack 2007). Conversely, in the negative feedback loop (b), the *JAZ* genes are also transcriptionally activated through JA signalling. Therefore, JA action is down-regulated by *de novo* synthesised JAZs. For example, the transcription of *JAZ3*, which is a MYC2 transcriptional target, is repressed by its own product, JAZ3. The switch-on of *JAZ3* gene expression is mediated by JA-mediated destruction of JAZ3 repressors, the resulting JAZ3 proteins further down-regulate JA action (Chini et al. 2007). It seems that the JA pulsed response has to be precisely controlled to avoid runaway stress metabolism. Plants have developed many strategies to achieve this goal. The negative feedback loop could be one of the strategies. In addition, the inactivation of the bioactive (+)-7-*iso*-JA-$_{L}$-Ile by epimerisation or methylation after JA-induced protein degradation and the dominant repression by JAZΔJas repressors (JAZ proteins that lack Jas domains, which is caused by alternative splicing of *JAZ* mRNAs) could also contribute to the switching-off the pathway (Chung and Howe 2009; Fonseca et al. 2009b).

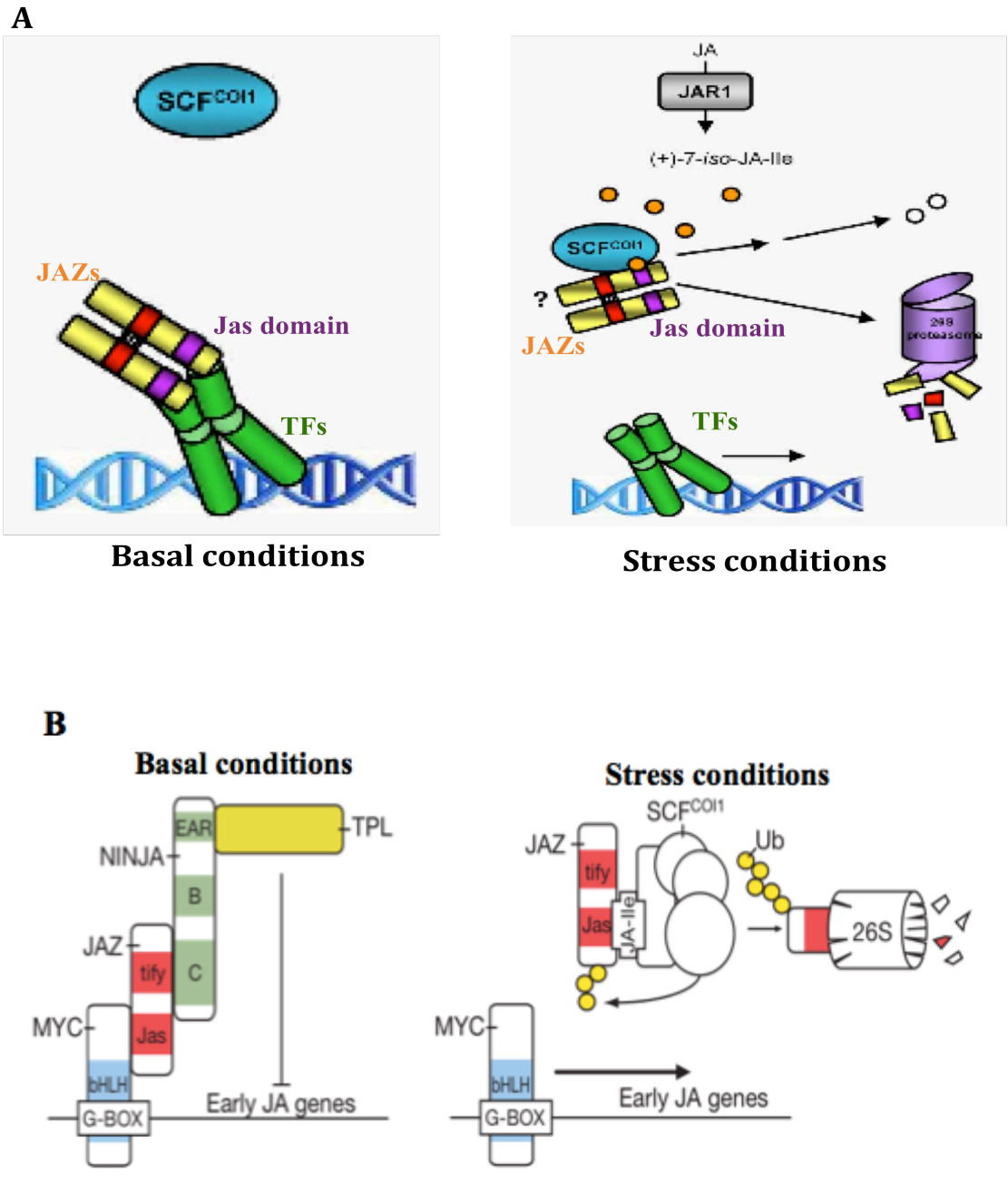


Fig 1.11 (Fonseca et al. 2009b; Pauwels et al. 2010) **(**A) a model for JA signalling: Under basal conditions, the transcriptional activity of TFs is repressed by JAZs. Under stress conditions, the active form of JA, (+)-7-*iso*-JA-$_{L}$-Ile, promotes the interaction between SCF$^{COI\ 1}$ and TFs. As a consequence, JAZ proteins are ubiquitinated and degraded by the 26S proteasome, and TFs are released to activate transcription (B) a representative example of JA signalling pathway: the JAZ

proteins inhibit the activation of MYC2 target genes by recruiting the corepressors TOPLESS and TPL-related proteins through an interaction with an adaptor protein, NINJA. Conversely, in the presence of JA-Ile, JAZs are linked to COI 1 through the Jas domain, ubiquitinated and degraded by the 26S proteasome. This leads to the switch-on of JA-responsive gene expression.

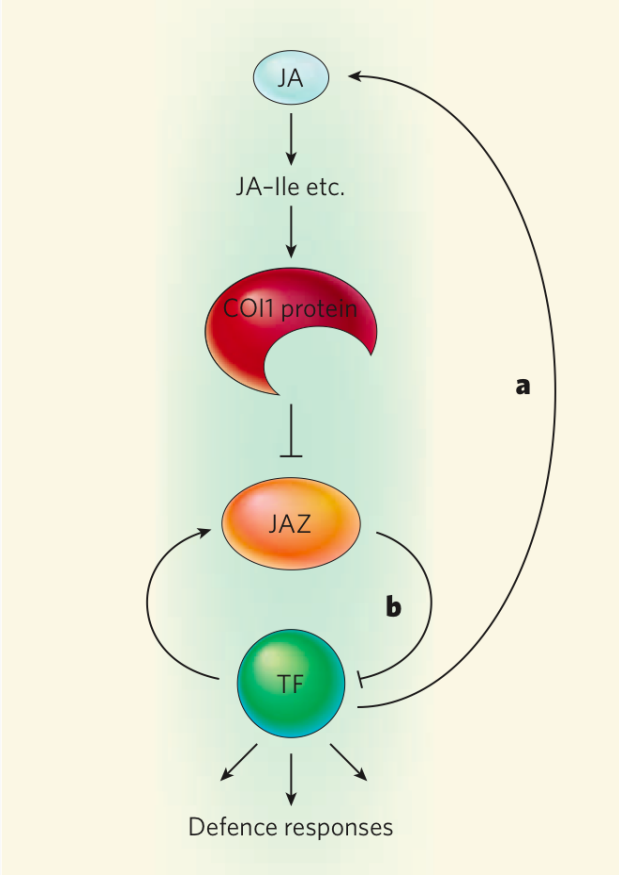


Fig 1.12 (Farmer 2007) The regulatory circuit of JA signalling that fine-tunes JA responses: (a) the positive feedback loop: under stress conditions, JA triggers the defence responses through JA signalling pathway. Meanwhile, the JA biosynthesis is activated and produces more JA, and therefore JA responses are enhanced. (b) the negative feedback loop: the *JAZ* genes are also transcriptionally activated through JA signalling. Therefore, JA action is down-regulated by *de novo* synthesised JAZs.

1.5.3 The JAZ family

JASMONATE-ZIM DOMAIN (JAZ) belongs to the TIFY family, which is previously known as Zinc-finger inflorescence meristem (ZIM). The renaming of the TIFY from ZIM is due to the inconsistency of its classification among different databases (Vanholme et al. 2007).

1.5.3.1 The structure and function of JAZ proteins

There are 12 putative JAZ proteins in the *Arabidopsis* JAZ protein family (JAZ1-12). The 28 amino acid ZIM and Jas domain are highly conserved among these 12 JAZs (Staswick 2008). Furthermore, the TIFY motif (TIFYAG) and the JAZ degron (LPIAR) are also highly conserved within the ZIM and Jas domain, respectively (Fig 1.13) (Vanholme et al. 2007).

It has been demonstrated both *in vitro* and *in vivo* that ZIM domains mediate JAZ-JAZ interactions to form either homomeric or heteromeric proteins. For example, JAZ10.4, a natural splice variant of JAZ10, which contains the ZIM domain but lacks the Jas motif, is still able to interact with other JAZs (Chini et al. 2009). In fact, all the JAZ-JAZ interactions in the *Arabidopsis* JAZ family have been confirmed by Yeast two-Hybrid (Y2H) (Chini et al. 2009; Chung and Howe 2009). In addition, JAZ has also been shown to interact with other regulatory proteins, such as NINJA and TPL. (Pauwels et al. 2010; Szemenyei et al. 2008). In Y2H database, JAZ1-6 and 9-12 are shown to interact with NINJA, whereas only JAZ 5 and 8 are shown to directly interact with TPL in a plant interactome project (Arabidopsis Interactome Mapping Consortium 2011).

The Jas domain is responsible for the interaction of JAZs with TFs, as well as COI 1 in the presence of JA-Ile and the cofactor, inositol pentakisphosphate. It contains a degron motif (LPIAR), which is required for its interaction with COI 1,and ultimately, the $SCF^{COI\ 1}$-meidated JAZ degradation (Sheard et al. 2010). Y2H has

confirmed the interaction of JAZs with different classes of TFs, including bHLH, R2R3 MYB and TFs involved in other hormonal signalling pathway (Pauwels and Goossens 2011). MYC2 was the first TF identified to be regulated by JAZ proteins (Chini et al. 2007). MYC2 interacts with most JAZs proteins and acts as both an activator and repressor in the JA signalling pathway (Dombrecht et al. 2007; Lorenzo et al. 2004). In addition, the closest homologs of MYC2, MYC3 and MYC4 are targets of JAZs and act as activators, but, additively with MYC2, regulate specifically different subsets of the JA responses (Cheng et al. 2011; Fernandez et al. 2011). Furthermore, two R2R3-MYB TFs, MYB21 and MYB 24, have been reported to be the target of JAZs and involved in JA-regulated stamen development in *Arabidopsis* (Song et al. 2011). Finally, the interaction of JAZs with WD-Repeat/bHLH/MYB complexes through bHLH and MYB, represses JA-regulated anthocyanin accumulation and trichome initiation. Upon JA-induced JAZ degradation, the transcription complex is released and functions to activate essential target genes involved in the regulatory pathway of anthocyanin accumulation and trichome initiation in *Arabidopsis* (Fig 1.14) (Qi et al. 2011).

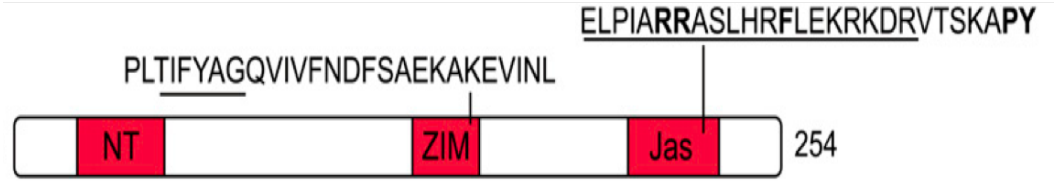


Fig 1.13 (Pauwels and Goossens 2011) The basic structure of JAZ proteins. It contains a ZIM and Jas domain. The TIFY motif (TIFYAG) and the JAZ degron (LPIAR) are also highly conserved within the ZIM and Jas domain, respectively.

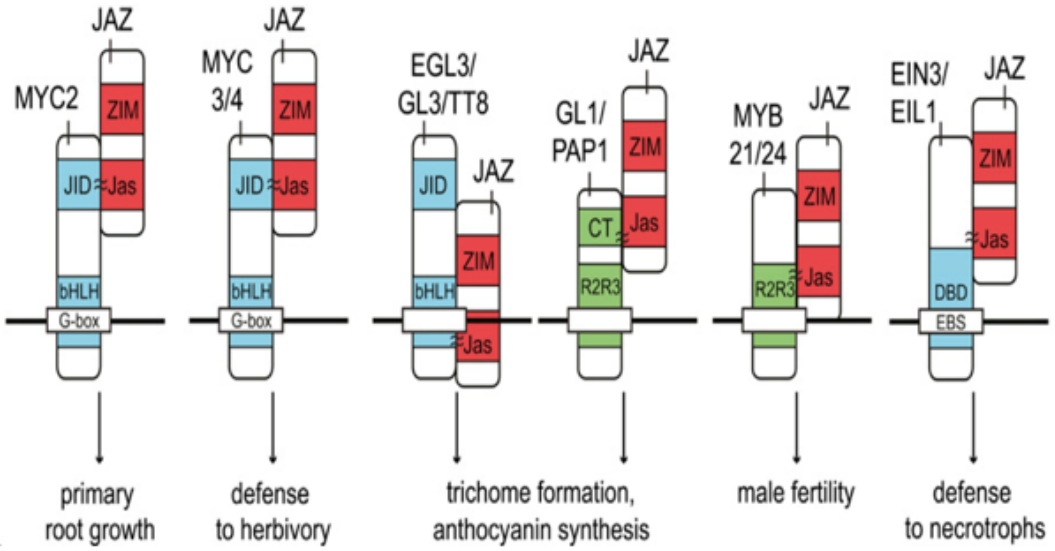


Fig 1.14 (Pauwels and Goossens 2011) Repression of different JA–mediated transcriptional responses through the interaction of JAZs with a wide array of TFs, including MYC2/3/4, EGL3/GL3/TT8, GL1/PAP1, MYB21/24 and EIN3/EIL1. The release of these TFs from JAZs-mediated transcriptional repression activates essential target genes involved in different biological processes.

The third functionally important motif is the Ethylene Responsive Factor-associated Amphiphilic Repression (EAR) motif present in the N-terminal domain of JAZ proteins. There are only four JAZ proteins containing the EAR motif, including JAZ5/6 and JAZ 7/8. The EAR motif is responsible for the recruitment of the corepressors, TPL, through the direct interaction between them. JAZ5 and JAZ8 have been confirmed to have direct interactions with TPL. In particular, JAZ8 has been shown to mediate transcriptional repression of JA responses through direct recruitment of TPL *in vivo* (Shyu et al. 2012). However, the interaction of JAZ6 or JAZ7 with TPL has not been described to date. In contrast, JAZs, which lack of the EAR motif, require an EAR-containing adaptor protein NINJA or heterodimerize with EAR-containing JAZs to recruit TPL for inhibition of JA responses (Pauwels et al. 2010). Fig 1.15 shows hypothetical models of inhibition of JA responses by JAZs (Pauwels and Goossens 2011).

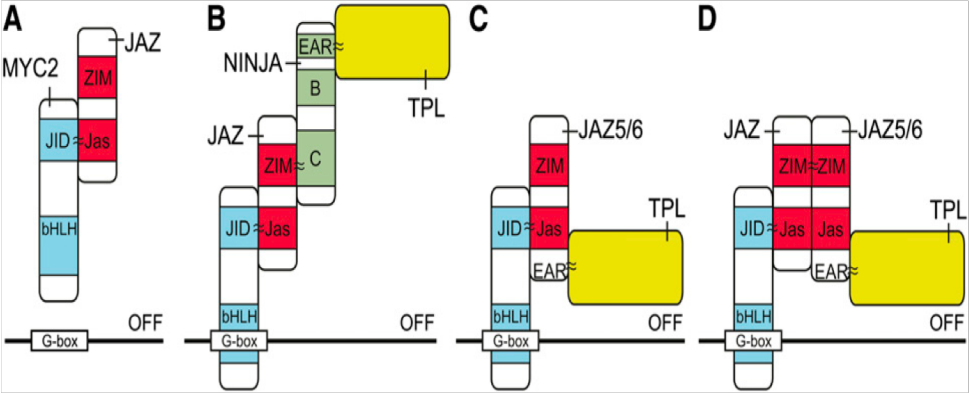


Fig 1.15 (Pauwels and Goossens 2011) Hypothetical models of inhibition of JA responses by JAZs. (A) Direct repression by JAZs only; (B) Repression by JAZ-recruited TPL through EAR-containing NINJA; (C) Repression by JAZ-recruited TPL without NINJA adapter; (D) Repression by JAZ dimer-recruited TPL without NINJA adapter.

1.6 The JA-responsive TF families involved in transcriptional regulation of plant secondary metabolism

JA signalling has been involved in many processes in plant development and defence. In the signalling pathway, the bioactive JA-Ile links the JAZ repressors with $SCF^{COI\,1}$, which leads to that the degradation of JAZs by 26S proteasome and therefore the release of TFs from JAZ-mediated repression and subsequent transcriptional responses. TFs are proteins that regulate gene activity by binding to sequence-specific motifs present in gene promoters and modulate the rate of transcriptional initiation, which is carried out by the basal transcription machinery, including RNA polymerase II (Ranish and Hahn 1996). In particular, TFs and their corresponding promoter binding elements, which are employed for transcriptional gene activation in the JA signalling pathway, are said to be JA-responsive. Furthermore, there have been different TF families described to be involved in transcriptional regulation of plant secondary metabolism in JA responses, through which the secondary metabolites are produced in response to environmental stimuli, for example, pathogen attack (Schwinn 2003).

1.6.1 Apetala 2/ethylene-responsive element binding factor (AP2/ERF) family

The AP2/ERF family consists of a large group of plant-specific TFs, which is divided into four major subfamilies: the AP2, RAV, ERF and dehydration-responsive element-binding protein (DREB), and among which, the DREB subfamily is the largest subfamily (Mizoi et al. 2012). The *Arabidopsis* and rice contain 145 and 167 loci encoding family members, respectively (Sakuma et al. 2002). The AP2/ERF TFs commonly contain an AP2/ERF-type DNA binding domain, which consists of approximately 60 amino acids (Weigel 1995). The name of this family was derived from the *Arabidopsis* homeotic gene *APETALA 2* and tobacco (*Nicotiana tabacum*) ethylene-responsive element binding proteins (EREBPs), where this domain was first found (Jofuku et al. 1994; Ohme-Takagi and Shinshi 1995). The domain structure, which was revealed in the structure of the *Arabidopsis* ERF1 protein in complex with a target DNA by NMR, contains an N-terminal, a C-terminal α-helix and three-

strand β-sheet that is responsible for the recognition of the cognate target DNA sequence (Fig 1.16) (Allen et al. 1998).

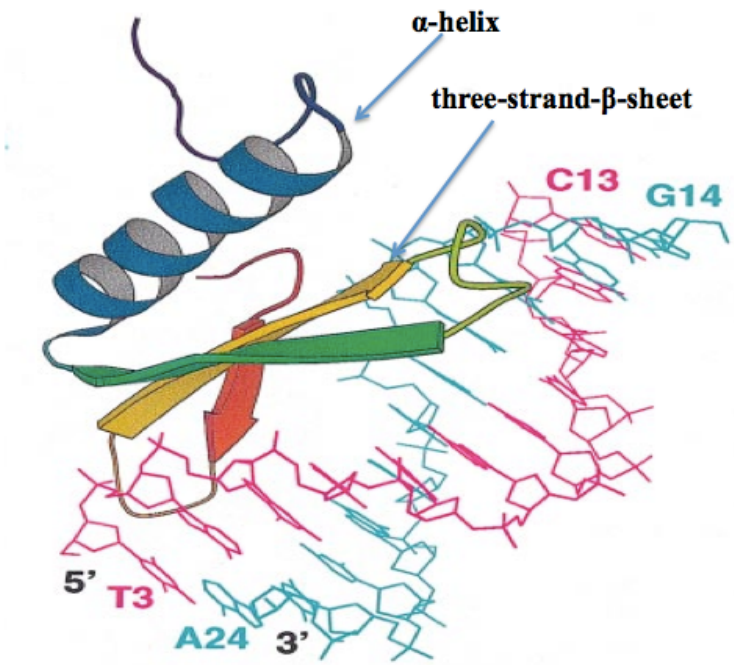


Fig 1.16 (Allen et al. 1998) The structure of AP2/ERF domain in complex with its target DNA sequence. It consists of a N-terminal α-helix and a three-stranded anti-parallel β-sheet.

Generally, the DNA binding element is specific to AP2/ERF subfamily proteins. DREB1A and DREB2A, the DREB subfamily proteins, have the highest affinity for the sequence A/GCCGAC, which is the core sequence of *cis*-acting dehydration-responsive element (DRE) (Sakuma et al. 2002). In contrast, five ERF subfamily TFs have the highest affinity for the ethylene-responsive element (ERE) core sequence, the GCC-box (AGCCGCC) (Fujimoto et al. 2000). However, there are exceptional cases for the core sequences. For example, TINY, a DRE-binding protein-like TF from the subgroup A-4 of the DREB subfamily has been shown to bind to both the DRE and GCC-box sequences (Sun et al. 2008).

A well-characterised example of JA-responsive AP2/ERF TFs is the Octadecanoid Responsive *C. roseus* AP2 (ORCA) TF family, including ORCA1, 2 and 3. ORCA2

and 3, but not ORCA1, are involved in terpenoid indole alkaloid (TIA) biosynthesis in *Catharanthus roseus* (Menke et al. 1999; van der Fits and Memelink 2000). Both ORCA2 and 3 bind to a jasmonate and elicitor responsive element (JERE) present in the promoter region of the *Strictosidine synthase* (*Str*) gene, which is involved in the initial step of TIA biosynthesis (Memelink 2007; Endt et al. 2007). Furthermore, ORCA3 has been suggested to be the central regulator of TIA biosynthesis, because it transcriptionally activates not only several structural genes in the TIA pathway, but also genes involved in the biosynthesis of TIA precursors (Memelink 2007). A close homologue of ORCA3, NtORC1 from tobacco, together with NtJAP1, are also shown to regulate the activity of structural genes involved in JA-responsive tobacco alkaloid biosynthesis in transient assays in tobacco protoplasts, such as the putrescine *N*-methyltransferase encoding gene, *PMT* (Sutter et al. 2005; Katoh 2005). The gene expression of both TFs is shown to be induced by MeJA (Goossens et al. 2003).

1.6.2 Basic Helix-loop-Helix TF family

The basic/helix-loop-helix (bHLH) proteins comprise a superfamily of TFs. It has been estimated that the *Arabidopsis* bHLH family contains at least 147 members, which is grouped into 21 subfamilies through phylogenetic analysis (Toledo et al. 2003). The bHLH family is characterised by the bHLH signature domain. This domain contains ~60 residues, which are functionally divided into two distinct regions. One is the basic region located at the N-terminal end of the domain. This region consists of ~15 amino acids, most of which are basic residues. It functions to bind target DNA sequences. The other is the HLH region at the C-terminal end. The amino acids within this region are mainly hydrophobic residues that form two amphipathic α-helices separated by a loop region. The amino acid sequence and length varies in the loop region (Nair and Burley 2000). The HLH region is responsible for protein dimerization (Ferre et al. 1994). The crystal structure of bHLH has been revealed in the PHO4-DNA complex (Fig 1.17) (Shimizu et al. 1997).

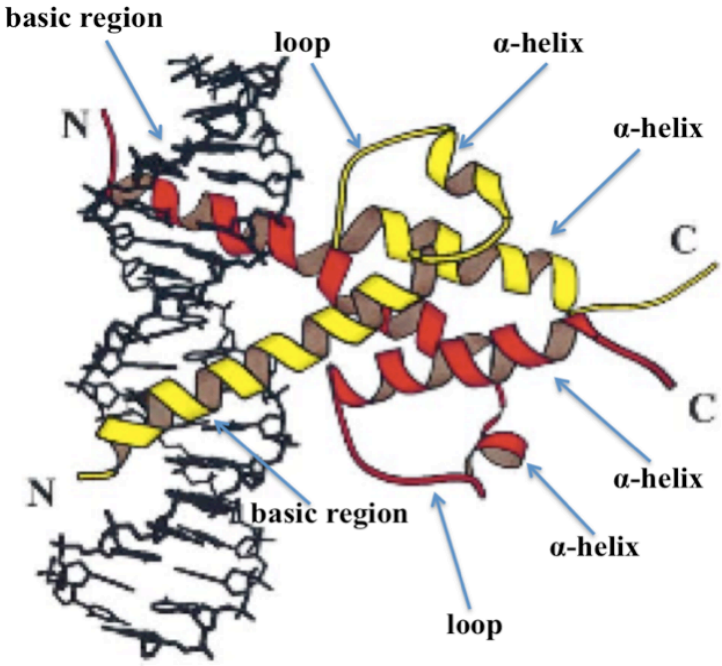


Fig 1.17 (Shimizu et al. 1997) The crystal structure of bHLH PHO4-DNA complex. It is a dimer. Each monomer contains a basic region located at the N-terminal end of the domain and a HLH region at the C-terminal end. The HLH region is responsible for protein dimerization.

A consensus hexanucleotide sequence, known as the E-box (5'-CANNTG-3'), is the core DNA sequence bound by bHLH TFs. The two central bases define the type of E-boxes, and the palindromic G-box (5'-CACGTG-3') is one of the most common types of E-box. The core sequence is recognized by the conserved residues within the basic region of the protein, whereas the specific recognition of a given type of E-box is provided by other residues in the domain (Robinson et al. 2000). In addition, the binding specificity is also shown to be influenced by the flanking nucleotides outside the core sequence (Massari and Murre 2000).

A well-studied example of bHLH TF is MYC2 that has been involved in differentially modulating diverse JA-dependent functions in *Arabidopsis* (Dombrecht

et al. 2007). As far as the secondary metabolites are concerned, it has been demonstrated that MYC2 positively regulates the JA-mediated flavonoid (anthocyanin or proanthocyanidin) biosynthesis by possibly modulating the expression of an intermediate spectrum of TFs, including MYBs (AtMYB75, 90 and TT2), bHLHs (TT8 and EGL3) and WD40 (TTG1) (Borevitz et al. 2000; Dombrecht et al. 2007; Nesi et al. 2000; Nesi et al. 2001; Walker et al. 1999). The Xie group (Qi et al. 2011) showed that WD-repeat, bHLH and MYB proteins form a complex to regulate JA-mediated anthocyanin accumulation in *Arabidopsis*. In addition, *Nt*MYC2a and *Nt*MYC2b have been reported to regulate the expression of JA-inducible multiple genes involved in nicotine biosynthesis by forming nuclear complexes with *Nt*JAZ1(Zhang et al. 2012). These two MYC proteins could have identical DNA-binding specificities for the same binding element, as they are from the same TF family and involved in the same JA signalling pathway. Instead of competing with each other for occupancy at the same location on promoters, one MYC protein can facilitate the binding of another (Stratmann and Schibler 2011).

1.6.3 MYB TF family

The MYB family consists of a large and functionally diverse number of TF proteins in all eukaryotes. The first identified plant TF gene is the *COLORED 1* (*C1*) that encodes a MYB domain protein involved in the anthocyanins biosynthesis in the aleurone of maize kernels (Paz et al. 1987). MYB proteins are characterised by the highly conserved MYB domain that is the DNA-binding motif. This domain generally comprises four imperfect amino acid repeats (R), with each having ~52 residues and forming three α-helices. In each repeat, a helix-turn-helix (HTH) structure is built by the second and third helices with three regularly spaced tryptophan residues, and the overall structure is a hydrophobic core (Ogata et al. 1996). In addition, it is the third helix in each repeat that recognizes DNA in the major groove (Jia et al. 2004).

Based on the number of adjacent repeats included in MYB proteins, this protein family can be divided into four different classes in *Arabidopsis*: MYB-related (one or

partial repeat), R2R3-MYB (two repeats), R1R2R3-MYB (3 repeats) and 4R-MYB (4 R1/R2-like repeats) (Fig 1.18). In particular, R2R3-MYB subfamily, which contains more than 100 members (Yanhui et al. 2006), is the largest subfamily in MYB proteins, and 4R-MYB proteins that form the smallest subfamily with the least number of protein members. Furthermore, in the modular structure of R2R3-MYB TF, the DNA-binding domain (MYB domain) is located at the N-terminus, whereas the activation or repression domain is located at the C-terminus. In contrast to these two highly conserved domains, the other regions of R2R3-MYB are highly variable. The R2R3-MYB subfamily is further divided into 23 subgroups (S1-23) (Dubos et al. 2010).

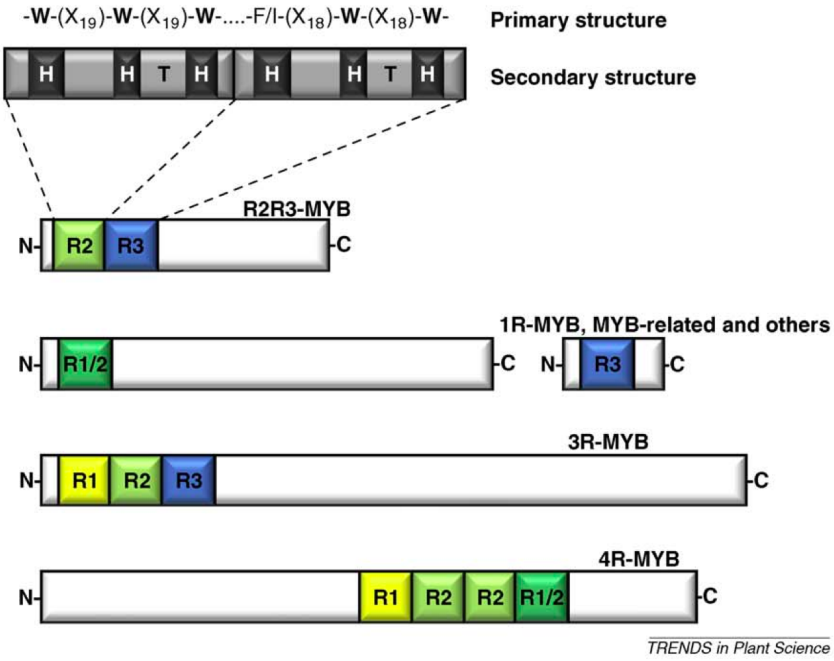


Fig 1.18 (Dubos et al. 2010) Classification of plant MYB TFs in *Arabidopsis*. There are four different classes: MYB-related (one or partial repeat), R2R3-MYB (two repeats), R1R2R3-MYB (3 repeats) and 4R-MYB (4 R1/R2-like repeats).

The large number of members within R2R3-MYB subfamily in plants fits well with the observation that many proteins from this subfamily are involved in plant–specific processes, such as primary and secondary metabolism, development and responses to

biotic and abiotic responses (Martin and Paz 1997). In terms of secondary metabolism, the regulation of flavonoid biosynthesis has been shown to involve several R2R3-MYB TFs in a tissue-specific manner in *Arabidopsis*. For example, *AtMYB75*, *AtMYB90*, *AtMYB113* and *114* from subgroup 6 regulate anthocyanin biosynthesis in vegetative tissues (Gonzalez et al. 2008), whereas *AtMYB123* from subgroup 2 controls proanthocyanidins (PAs) biosynthesis in the seed coat (Routaboul et al. 2006).

1.6.4 WRKY TF family

WRKY transcription factor proteins form one of the largest families of transcriptional regulators in plants. The members in this family are characterised by the DNA binding domain, which is called the WRKY domain. The WRKY domain consists of about 60 residues. Apart from the WRKY signature at the N-terminus, at the C-terminus it also has an atypical zinc-finger structure (Eulgem et al. 2000). In addition, the WRKY domain amino acid sequence is almost invariant except for a few WRKY proteins, in which the WRKY sequence is replaced by WRRY, WSKY, WKRY, WVKY or WKKY (Xie et al. 2005).

According to the number of WRKY domains and the structure of the zinc fingers ($Cx_{4-5}Cx_{22-23}HxH$/C2H2 or $Cx_7Cx_{23}HxC$/C2HC), the WRKY TFs can be divided into three groups: Group I (two WRKY domains with C2H2), Group II (one WRKY domain with C2H2) and Group III (one WRKY with C2HC). Group II can be further divided into Group IIa, b, c, d and e based on the primary amino acid sequence (Eulgem et al. 2000).

The structure of WRKY domain revealed that there is a four-stranded-β-sheet and a zinc-binding pocket formed by Cys/His residues (Fig 1.19). In addition, the conserved WRKYGQK contained at the most N-terminal β-strand makes contacts with a 6-bp region in the major DNA groove (Yamasaki et al. 2005). The remarkably conserved core sequence bound by almost all WRKY proteins is known as the W boxes (TTGACC/T). It has been confirmed that the W box is the minimal

consensus sequence required for DNA binding through several assays, such as gel shift experiment, yeast one-hybrid screens and co-transfection assays (Ciolkowski et al. 2008). The binding specificity of WRKY proteins for their target promoters is further determined by the adjacent binding sites. For example, the *At*WRKY6 preferentially binds to the core sequences that have a G base at the 5' direction of the W box element, whereas AtWRKY26 has a preference to W boxes that have T, C or A at the 5' direction of the core (Ciolkowski et al. 2008). However, there are a few exceptions that WRKY proteins bind to non-W box sequences. As well as binding to W box, *Os*WRKY13 also binds to a DNA sequence of TGCGCTT, known as the PRE4 element (Liu et al. 2008). A WRKY TF from barley, *Hv*WRKY46 binds to both W boxes and a sugar-responsive element (SURE) (Sun et al. 2003).

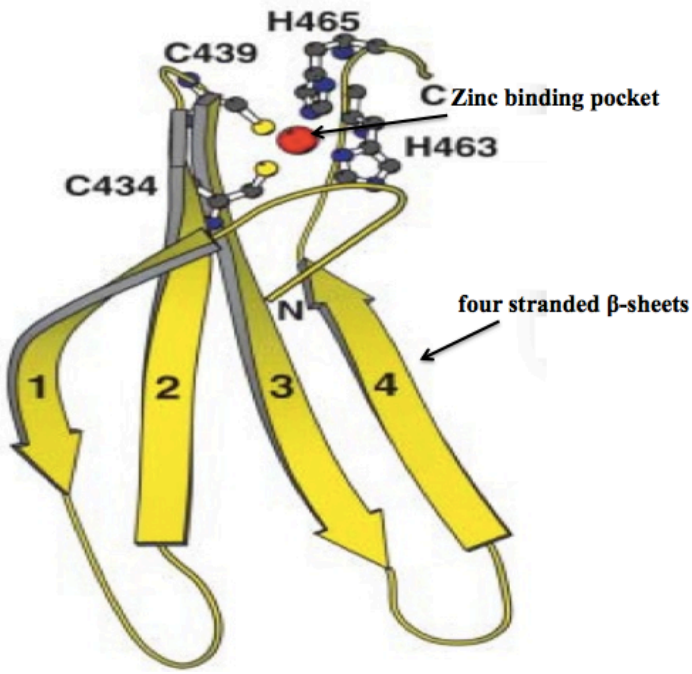


Fig 1.19 (Yamasaki et al. 2005) The structure of WRKY DNA binding domain. It consists of four stranded β-sheets and a zinc-binding pocket formed by Cys/His residues.

It has been evident that WRKY TFs play regulatory roles in a wide range of plant processes, such as abiotic and biotic stresses, plant development and senescence and

seed dormancy and germination (Rushton et al. 2010). Like other TFs described in this thesis, WRKY proteins have also been reported to be involved in biosynthesis of the secondary metabolites. The JA-inducible *Cr*WRKY1 is shown to positively regulate the TIA (serpentine) biosynthesis in roots of *C. roseus* plants. Together with other TFs, *Cr*WRKY1 stimulates the root-specific serpentine accumulation by activating several key TIA pathway genes, especially *Tryptophan Decarboxylase* (*TDC*) and zinc-finger *C. roseus* transcription repressors (*ZCT1*, *ZCT2* and *ZCT3*), as well as repressing key transcriptional activators, including *ORCA2*, *ORCA3* and *CrMYC2*, which directs the flux toward the vindoline and catharanthine pathways (Memelink 2007; Suttipanta et al. 2011; Wang et al. 2010).

1.6.5 NAC TF family

NAC (NAM, ATAF1, 2, CUC2) TFs constitute the NAC TF family, which is one of the largest of plant-specific TF families. There are more than 100 NACTF encoding genes found in *Arabidopsis* (Riechmann et al. 2000). The NAC domain at the N-terminal region of NAC proteins is the family-defining domain(Aida et al. 1997). By contrast, there are no known domains found at the C-terminal region, in which the residues are highly diverse (Ooka et al. 2003). Through forward genetics, *NAM* (no apical *meristem*) was the first NAC TF encoding gene to be identified, which was followed by the characterization of *AtCUC2* (*CUP-SHAPED COTYLEDON 2*) (Aida et al. 1997). Furthermore, NAC domains have been demonstrated to be responsible for both protein-DNA and protein-protein interactions (Ernst et al. 2004), and the core sequence they recognize in the promoter fragment is CACG (Tran et al. 2004). Whereas the transcription activation domains have been shown to locate at the C-terminal regions in yeast and plant cells (Hegedus et al. 2003; Tran et al. 2004), the structure of NAC domain (dimer) has been elucidated by X-ray crystallography. It shows that the NAC domain consists of a twisted antiparallel β-sheet surrounded by two α-helices located on each side (Fig 1.20) (Ernst et al. 2004).

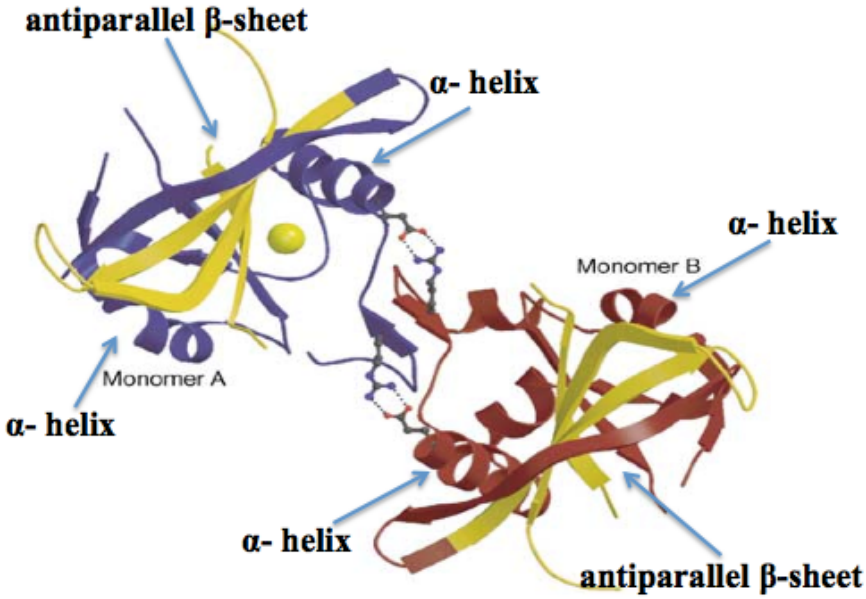


Fig 1.20 (Ernst et al. 2004) The dicyanoaurate derivative structure of NAC dimer. Each consists of a twisted antiparallel β-sheet surrounded by two α-helices located on each side.

NAC proteins have been implicated in various plant processes, including development, defence and abiotic stress responses (Olsen et al. 2005). In addition, they have been involved in regulating secondary metabolite synthesis. Two NAC TFs, NAC SECONDARY WALL THICKENING PROMOTING FACTOR1 (NST1) and NST3, have been suggested to be the key regulators of the formation of secondary walls in woody tissues of Arabidopsis (Mitsuda et al. 2007).

1.7 Regulation of TFs controlling plant secondary metabolism

A single TF alone is, often, not capable of regulating the whole biosynthetic pathway of secondary metabolite. They generally require cofactors, some post-translational modifications or other interacting proteins in order to become functionally active. These changes may be required to affect DNA binding activity, nuclear localization and protein stability. In addition, they typically work together with other TFs from either the same or different TF family to temporally and spatially regulate the biosynthesis of specific secondary metabolites in a TF regulatory network.

A key post-translational modification that plays important roles in plant processes is protein phosphorylation. Especially, it has been reported that phosphorylation of WRKY TFs by a phosphorylation cascade is involved in mediating many processes in plants, including plant immune reponses to pathogen attacks (WRKY33), plant defence against herbivore attack (WRKY3 and WRKY6) and dehydration (WRKY1), and senescence (WRKY53) (Rushton et al. 2010). Furthermore, the phosphorylation of *Nicotiana benthamiana* WRKY8 mediated by Mitogen-activated protein kinase (MAPK) also plays roles in plant defence (Ishihama et al. 2011). Recently, it has been revealed that, *in vivo*, the phosphorylation of WRKY33 by two MAPKs, MPK3 and MPK6, which function together in the same MAPK cascade, reprogrammes the expression of genes encoding enzymes required for camalexin biosynthesis, and therefore direct the metabolic flow toward camalexin metabolism in *Arabidopsis* when challenged with *Botrytis cinerea* (Mao et al. 2011).

Alternatively, TF activity can be regulated by protein-protein interactions. The stimulation of the binding activity of WRKY33 requires two nuclear-encoded activators, sigma factor binding protein 1(SIB1) and SIB, which recognize the C-terminal WRKY domain in plant defence against necrotrophic pathogens. A short FXXXVQXXTG (X, any amino acid) motif in these two proteins is important for interaction with WRKY33 (Lai et al. 2011). Furthermore, in the biosynthesis of flavonoids, MYB and bHLH TFs may functionally and physically interact with each other to direct the metabolic flow in a branch-specific manner, which leads to the production of different flavonoids. The physical interaction between certain MYB

and bHLH proteins has been confirmed by Koes (Mol 1998). In maize, C1 (MYB) and R (bHLH) together drive the biosynthetic pathway for the production of anthocyanins by activating the corresponding structural genes (Bruce et al. 2000). In contrast, PAP1 (MYB), a possible C1 ortholog, is able to activate anthocyanin production without co-expression of a bHLH TF in *Arabidopsis* (Borevitz et al. 2000); however, both TT2 (MYB) and TT8 (bHLH) are required to direct the metabolic flux toward proanthocyanidins biosynthesis in *Arabidopsis* (Nesi et al. 2000; Nesi et al. 2001). Furthermore, there is some evidence showing that JA-regulated secondary metabolite biosynthesis can be controlled by either two TFs that function cooperatively, or by a transcription factor cascade that consists of a TF regulating a down-stream TF gene expression, which in turn regulates the expression of a subset of structural genes involved in metabolite biosynthesis. For example, a bHLH TF is required for the maximum transactivation of ORC1 in the control of JA-elicited tobacco nicotine biosynthesis. In fact, the ORC1 (AP2) and bHLH binding element, GCC-motif and G-box, respectively, are both present in the promoters of the genes encoding enzymes involved in nicotine biosynthesis (De Boer et al. 2011). JA-responsive ORCA TF family (ORCA2 and ORCA3) has been shown to regulate alkaloid biosynthesis in *C. roseus* (van der Fits and Memelink 2000). Recently, the transcriptional regulator of these two TF-encoding genes has been identified to be a bHLH TF, CrMYC2. This bHLH TF up-regulates the JA-responsive *ORCA* gene expression, which in turn leads to the up-regulation of a subset of alkaloid biosynthesis genes and subsequent increase in alkaloid production (Figueroa and Browse 2012).

Realistically, the complexity of the transcriptional regulation of the biosynthesis of secondary metabolites by TFs in planta is far beyond the scope of single- or double-TF regulatory concept. In fact, it is more likely that the biosynthesis of secondary metabolites is regulated by a group of TFs that form a transcriptional network. Grotewold (2008) predicted a model of gene regulatory network in planta based on the findings in *Saccharomyces cerevisiae* and *Escherichia coli* (Yu and Gerstein 2006). Within this network, one TF often, together with others, positively or negatively regulates another in a regulatory motif. Then, when assembled into

regulatory modules, a variety of regulatory motifs build up a gene regulatory network (Babu et al. 2004). Moreover, TFs are situated at various levels in the pyramidal hierarchical arrangement. TFs situated near the top of the pyramid are master TFs, and they directly control these TFs present in the middle of the pyramid. In contrast, TFs located at the bottom of the pyramid are responsible for the control of the structural genes that are required for most of cellular processes. In other words, the higher the TF is in the pyramid, the more genes are under its control, which means an alteration in the activity of this TF will affect the expression of a large number of downstream genes. By contrast, the TFs at the bottom of the pyramid only regulate a small subset of genes; therefore, the change in the TF activity will only affect the expression of a small number of genes (Fig 1.21). Zhao and Dixon (2010) proposed a simple model of the transcriptional network of secondary cell wall biosynthesis in *Arabidopsis*. In their model, the master TFs, including NST1/2/3, can activate secondary cell wall synthesis via the centrally located MYB46/MYB83. Then, MYB58/63/85, which are under the control of MYB46/83, in turn activate the expression of lignin biosynthesis genes through AC elements, therefore leading to lignin production. Meanwhile, MYB4, which is controlled by MYB83, acts as a repressor to inhibit lignin biosynthesis (Zhao and Dixon 2011). Consistent with this concept, *Del* and *Ros1* in anthocyanin biosynthesis are likely located near the top of the hierarchical network so that over-expression of these master TFs leads to up-regulation of most of structural genes present in the pathway, thereby a significant increase in the production of anthocyanin in the fruit of tomato (Butelli et al. 2008).

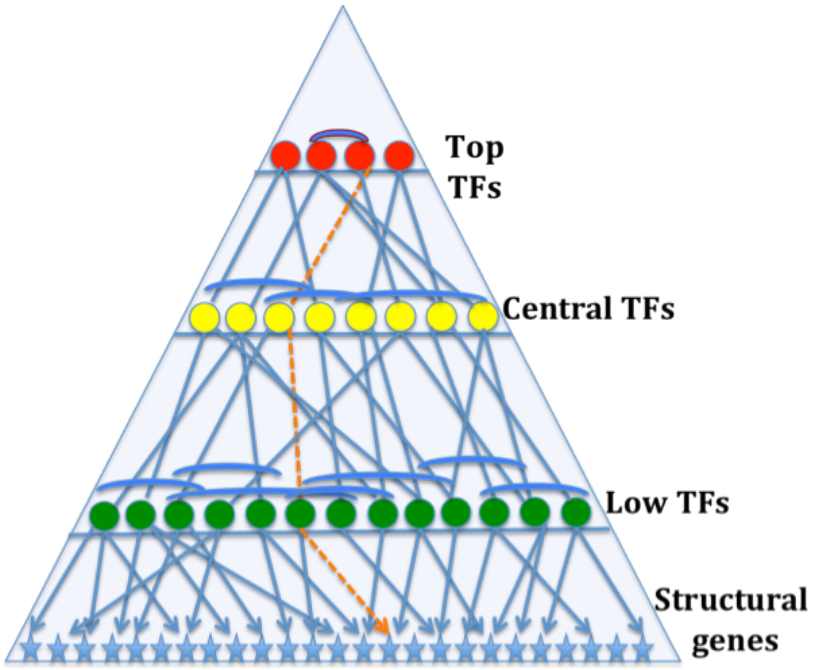


Fig 1.2 [Adopted from(Grotewold 2008)] A simplified diagram of a model gene transcriptional regulatory network involving TFs (circles) and structural genes (stars) arranged in a pyramidal hierarchical pattern. TFs are shown at various levels in the pyramidal hierarchical arrangement: top TFs (red circles), central TFs (yellow circle) and low TFs (green circles). Structural genes (stars) are located at the bottom of the pyramid. Blue lines indicate the direct interaction of one TF with another from different levels; horizontal blue lines at each level indicate the TF interactions at the same level; blue arrowed lines indicate the regulation of gene expression by TFs; orange dotted lines represent a regulatory motif.

1.8 Bioengineering of transcription factors

In the field of plant metabolic engineering to increase the production of medically valuable compounds, TFs have been the main targets for the manipulation of complex metabolic pathways in plant, because modification of TF activity has a dramatic effect on the transcriptome, which in turn leads to metabolic changes. Unlike structural genes that control individual pathway steps, TFs tend to regulate multiple pathway steps in metabolic pathways. Therefore, TF bioengineering would avoid the time-consuming steps of characterizing all the enzymes involved in a biosynthetic pathway. In addition, it provides powerful tools to overcome flux bottlenecks involving multiple enzymatic steps, or to deploy genes involved in metabolic pathways in specific organs, cell types and even plant where they normally do not express (Broun 2004; Grotewold 2008).

There have been some well-known examples of TF engineering with the aim of increasing the production of valuable compounds, which mainly include pharmaceutical and health-promoting products. Terpenoid indole alkaloids (TIAs) have attracted many attentions from pharmaceutical companies, because several TIA compounds are medically valuable. For example, the monomeric alkaloids serpentine and ajmalicine are used as a tranquilizer and to reduce hypertension, respectively. Moreover, the dimeric alkaloids vincristine and vinblastine are potent anticancer drugs (Memelink 2007). In *C. roseus*, ORCA3, a JA-responsive AP2 TF, has been shown to be a central regulator of TIA biosynthesis, and overexpression of this TF results in elevated expression levels of *STR* (strictosidine synthase) and other TIA biosynthesis genes and consequently, increased production of TIAs (van der Fits and Memelink 2000). Moreover, anthocyanins have been known to have health-promoting effects due to their high antioxidant activities (Chun et al. 2003). The upregulation of the *Pr* gene, which encodes a R2R3 MYB TF, specifically activates a bHLH TF and a subset of structural genes involved in anthocyanin biosynthesis and, consequently, leads to accumulation of pigments in the purple cauliflower (Chiu et al. 2010). Most strikingly, Butelli et al. (Butelli et al. 2008) reported that the co-expression of two heterologous TFs (Del and Ros1) isolated from the snapdragon *Antirrhinum majus* resulted in a substantial increase in anthocyanin production in the

fruit of transgenic tomatoes. In fact, the ectopic expression of those two TFs up-regulated a broad range of genes involved in the entire anthocyanin biosynthesis pathway, including structural genes required for anthocyanin biosynthesis, genes encoding enzymes for side-chain modification and two genes involved in transportation of the end-products into the vacuole. Sometimes, however, a target TF does not regulate the expression of all the structural genes involved in the metabolic pathway, and it needs to be co-expressed with key enzymes that are not regulated by this TF in order to achieve the desired flux level through a pathway (Grotewold 2008). Proanthocyanidins (PAs) are plant antioxidants that confer multiple human health benefits, as they give greater and more stable protection against free-radical-induced oxidative tissue damage than vitamins C or E (Shi et al. 2003). Moreover, they can also prevent pasture bloat in ruminant animals (McMahon et al. 2000). However, there are no PAs accumulating in vegetative tissue of tobacco. The Xie group developed a strategy for engineering PAs in tobacco leaves involving co-expression of anthocyanidin reductase (*ANR*) and a *PAP1* that encodes MYB TF. As a result, the anthocyanin pathway induced by PAP1 in tobacco leaves is diverted into the PAs-producing pathway upon ANR co-expression (Xie et al. 2006).

1.9 Strategies for identification and functional analysis of plant TFs

1.9.1 Strategies to identify plant TFs

Traditionally, there are two main routes, through which TFs that have defined functions in regulation of secondary metabolism can be identified. One is cloning of genes associated with mutations; the other is screening for proteins interacting with defined promoter *cis*-elements. The latter is often used for identification of TFs in non-model plants. For example, *MYC2/JIN1/JAI1*, which encodes a bHLH TF and acts as either an activator and a repressor of JA-responsive gene expression in *Arabidopsis*, was first identified by screening mutants with reduced sensitivity of its roots to exogenous JA (Berger et al. 1996). In addition, a homolog of *At*MYC2, *JA*MYC2 was identified using yeast one-hybrid assays, in which JA-responsive T/G-box motifs in tomato *LAP* and *pin2* promoters were used as baits to isolate the bHLH DNA-binding proteins (Boter et al. 2004). Furthermore, recently, a systematic approach to identify plant TFs has been developed. In this approach, the sequences of a group of putative TFs with unknown functions are first identified and then the TFs are systematically characterised through subsequent analysis, such as promoter binding activity, functional studies and phenotypes caused by their expression in transgenic plants (Schwinn 2003).

1.9.2 Strategies to functionally analyze plant TFs

After a high-throughout identification of sequences encoding putative TFs with unknown functions, several strategies have been developed to functionally analyze these TFs, including molecular functional analysis, expression analysis, phenotypic analysis, bioinformatic analysis and network analysis. These strategies are briefly discussed.

1.9.2.1 Molecular analysis

Molecular analysis of TFs is to characterize the activation or repression activities of TFs. It employs both reporter and effector genes. The effector gene normally contains a TF-coding sequence fused to a heterogeneous sequence that encodes a DNA binding domain (DBD), such as GAL4DB from yeast, and this chimeric construct is driven by a strong promoter, for example, cauliflower mosaic virus (CaMV) *35S*. In contrast, the reporter gene is usually the *E. coli* β-glucuronidase (*GUS*), or *HIS3* (Imidazoleglycerol-phosphate dehydratase) in *Saccharomyces cerevisiae* in the case of Y1H assay. This reporter gene is driven by a heterogeneous sequence, including a minimal promoter and the upstream cis-elements, which consists of binding sequences of GAL4DBD (Fig 4.1). In addition, a reference construct, which contains a constitutively expressed reporter gene, is used as a control. Then, all three constructs are transformed and transiently expressed in plant tissues, such as leaves, leaf protoplasts or cultured cells. Therefore, the TF activity (activator or repressor) can be indirectly examined through activity of reporter gene (Mitsuda and Ohme 2009).

1.9.2.2 Expression analysis

Expression analysis involves the identification of cells and tissues that express TF-encoding genes and further analysis of gene expression levels and patterns under certain conditions and chemical treatments. Several techniques, with each having their advantages and drawbacks, have been developed to establish the gene

expression profile of a given TF gene, such as quantitative reverse transcription-PCR (qRT-PCR), promoter-reporter experiments, in situ RNA hybridization and immunohistochemistry (Mitsuda and Ohme 2009).

1.9.2.3 Phenotypic analysis

Phenotypic analysis aims to identify a phenotype that is associated with a TF. Therefore, the phenotypes observed in mutants and transgenic plants need to be informative so that the biological functions of a given TF can be analyzed from them. Usually, the phenotypic changes are induced by either 'gain-of-function' or 'loss-of-function' strategies. In the 'gain-of-function' method, ectopic expression of a targeted TF is used to induce an informative phenotype; conversely, in the 'loss-of-function' method, T-DNA tagging is commonly used to inactivate genes. More specifically, RNA interference (RNAi) strategies have been adopted for the inactivation of genes of interest (Mitsuda and Ohme 2009).

1.9.2.4. Bioinformatic analysis

Bioinformatic analysis aims to identify and characterise TFs by using bioinformatic techniques on a large scale, in which, based on certain properties the proteins have in common, the putative TFs are aligned by sequences, grouped and compared with TFs with known functions in plant TF databases. These common properties include evolutionally conserved domains, subcellular localization, protein-protein interactions, miRNAs that regulate putative TFs and spatial and temporal TF gene expression profile. In addition to gene expression profiling, co-expression analysis is a particularly important strategy used to identify co-expressed genes, and therefore functionally related and/or physically interacting proteins (Mitsuda and Ohme 2009).

1.9.2.5 Network analysis

The elucidation of a TF regulatory network that regulates biosynthesis of secondary metabolites can lead to a more effectively targeted TF bioengineering, and ultimately increased yield of important medical molecules. The most recent and rapidly emerging technology that has great potential to unravel transcriptional regulatory networks is a combination of chromatin immunoprecipitation (ChIP) with either whole genome tiling arrays or next-generation sequencing technology (ChIP-chip or ChIP-seq). The application of ChIP confirms that TFs bind directly to the promoter region of putative target genes, and sequencing technologies further identify comprehensively the sequences of the TF-bound regions (Mitsuda and Ohme 2009).

Overall, with the completion of genome of model plants (*Arabidopsis* and rice) and advancements in Next-generation sequencing technology, plant TF isolation can be conducted more effectively and globally; as a result, many TFs that regulate secondary metabolism could be identified in both model and non-model plants (Schwinn 2003).

1.10 Summary

Despite extensive strategies for increasing paclitaxel yield to end the shortage of paclitaxel supply, *T. cuspidata* CMC has been proved to be a better-performing line, and subsequent cell culturing could provide an efficient, environment-friendly and sustainable platform to produce paclitaxel and its related precursors. The paclitaxel yield could be further improved using molecular approaches. Although a few paclitaxel biosynthesis pathway genes have been identified through the homology-based PCR approaches, there are still more metabolic pathway genes needed to be uncovered. The attempts to increase the paclitaxel yield by bioengineering a known rate-limiting pathway gene in heterologous microorganisms (*taxadiene synthase* or a P450-mediated *5α-oxidation of taxadiene*) cannot produce a significant increase in paclitaxel yield (Ajikumar et al. 2010; Huang et al. 2001; Yamin Wei 2012). Hence, more key pathway genes need to be engineered to achieve an outstanding increase in paclitaxel production. In addition, the upregulation of taxane biosynthesis pathway gene expression is not the only contributing factor that leads to the increased paclitaxel accumulation, because the differences in pathway genes are minor between high and low paclitaxel-accumulating MeJA-elicitated cultures. This indicates that there are additional factors that contribute to paclitaxel accumulation (Patil et al. 2012). The transcriptional regulators could be the potential targets, because it has been shown that overexpression of one or two master regulators can up-regulate most of the pathway genes and ultimately lead to a significant increase in the end products production (Butelli et al. 2008; Grotewold 2006). So far, there has been no paclitaxel pathway-specific TFs reported to up-regulate the entire pathway. Therefore, the prior step of my project is to identify and characterize TFs involved in controlling paclitaxel biosynthesis pathway, followed by elucidation of the TF regulatory network and targeted TF bioengineering to redirect the pathway flux towards taxane production.

1.11 Aims of my PhD project

A cambial meristematic cell (CMC) line of *Taxus cuspidata* has been isolated, cultured and demonstrated to be a cost-effective and environmentally friendly platform for the sustainable production of the key anticancer drug, paclitaxel (Lee et al. 2010). In my project, we aim to identify the transcription factors (TFs) that regulate the expression of paclitaxel biosynthetic genes using gene expression profiling followed by yeast one-hybrid analysis, gel shift assays and plant transient expression assays. The elucidation of a TF regulatory network that controls paclitaxel biosynthesis will guide the rational engineering of CMCs to ultimately increase yields of this important pharmaceutical.

Chapter 2 MATERIALS and METHODS

2.1 Cell suspension cultures of cambial meristematic cells and somatic cells

The medium (pH 5.5) used for culturing both cambial meristematic cells (CMCs) and somatic cells (SCs) consists of two parts: one part, which needs to be autoclaved, contains1011.1 mg/L KNO_3, 121.56 mg/L $MgSO_4.7H_2O$, 10 mg/L $MnSO_4.4H_2O$, 2 mg/L $ZnSO_4.7H_2O$, 0.025 mg/L $CuSO_4.5H_2O$, 113.23 mg/L $CaCl_2.2H_2O$, 0.75 mg/L KI, 0.025 mg/L $CoCl_2.6H_2O$, 130.44 mg/L $NaH_2PO4.H_2O$, 3.0 mg/L H_3BO_3, 0.25 mg/L $Na_2MoO_4.2H_2O$, 36.7 mg/L FeNaEDTA, 100 mg/L myo-inositol, 10 mg/L thiamine-HCL, 1 mg/L nicotinic acid, 1 mg/L pyridoxine-HCL, 133 mg/L aspartic acid, 175 mg/L arginine, 75 mg/L glycine, 115 mg/L proline and 1 mg/L picloram; the other part contains 0.5 mg/L gibberellic acid, 100 mg/L ascorbic acid, 150 mg/L citric acid, and 2% sucrose, which is filtered into the first part after autoclaving it. All chemicals used for culturing were ordered from Duchefa. Both CMCs and SCs were originally cultured in Unhwa Company in South Korea, delivered and cultured in our lab. In the initial inoculation, 4 ml of cells was cultured in 40 ml of culturing medium (cell: medium volume is 1: 10) in 100 ml conical flasks. When changing medium for subculturing after 12 days, one quarter of the old medium (including cell sediment) was inoculated with three quarters of freshly prepared medium in 100 ml conical flasks. After subculturing, the cell suspensions were shaken at 100 rpm in 24 hour dark at 21°C (Lee et al. 2010). The cultured CMCs and SCs were induced by MeJA and used to prepare cDNA libraries for Next-generation sequencing.

2.2 Gateway cloning system

Gateway cloning strategy (Life technologies) is an endonuclease digestion and ligation-free cloning technique, and it involves site-specific recombination and is a powerful method for cloning of PCR products and analysis of gene function. It shows a rapid and efficient cloning of PCR products with >99% of positive colonies. This cloning system consists of two site-specific recombination reactions: BP and LR reactions. In BP recombination reaction, PCR products or genes of interest flanked by attB1 and attB2 recombination sites recombine with pDONR vectors to generate Entry clones (pENTR), in which, PCR products of genes of interest are flanked by attL1 and attL2 recombination sites. The subsequent LR recombination reaction involves the attL sites recombine with attR sites in Destination vector (pDES) to generate expression clones (Fig 2.1) (Gateway^{TM} Cloning Technology Instruction Manual).

In BP reactions, 3 μl of attB site-flanked PCR products of TFs was mixed with 1 μl of pDONR, for example, pDONR^{TM} 221 (Fig 2.2b), in the presence of Gateway® BP clonase II Enzyme Mix (Life technologies). The mixture was incubated at 25°C for 1 hour, followed by introduction into *E. coli* strain, XL1-blue competent cells. In LR reactions, 3 μl of pENTR generated from BP reaction was mixed with 1 μl of pDES vector (pGAD424gatelapau) (Fig 2.2c) in the presence of Gateway® LR clonase II Enzyme Mix (Life technologies). The mixture was incubated at 25°C for 1 hour, followed by introduction into *E. coli* strain, XL1-blue competent cells. Positive colonies from BP and LR reaction were selected on kanamycin and ampicillin plates, respectively (Gateway^{TM} Cloning Technology Instruction Manual). This cloning system was used to facilitate the cloning of both TF encoding genes and promoters in the Y1H and plant transient expression assays. AttB and attP in BP reaction and attL and attR in LR reaction are two pairs of recombination sites. Site-specific recombination can occur between these two sites in each pair in the presence of the recombinase (Coates et al. 2005).

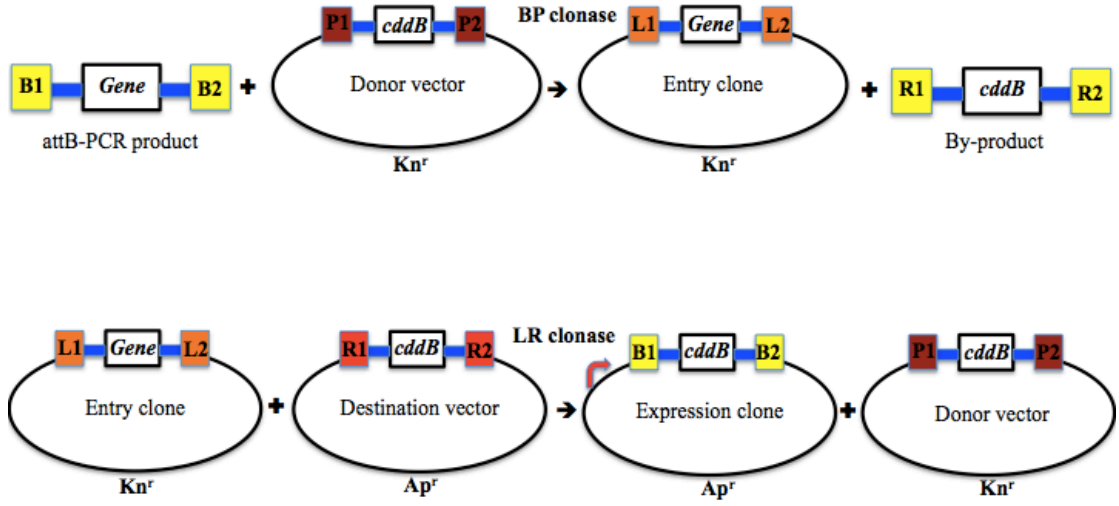


Fig 2.1 Gateway cloning system (Gateway^{TM} Cloning Technology Instruction Manual). BP reaction takes place between attB-sited gene products and pDONR vectors in the presence of BP clonase, and pENTR vectors are the products (upper panel). LR reaction takes place between pENTR and pDES in the presence of LR clonase, and the products are expression vectors (lower panel). Kn^r: Kanamycin resistance; Ap^r: ampicillin resistance.

2.3 Gateway-compatible Yeast one-hybrid assay

Yeast one-hybrid (Y1H) assay was used to screen for interactions between 5 paclitaxel biosynthetic gene promoters and 19 putative transcription factors in yeast strain, YM4271. Two vectors were made for this assay. One is the promoter-reporter construct, pHISi. Each of the five promoter fragments was inserted upstream of the *His3* reporter gene at the multiple cloning site (MCS) through conventional cloning strategy (Fig 2.2a). All five paclitaxel biosynthesis gene promoters were amplified from the genomic DNA of CMCs using primers (SI table 4), which were designed according to the published promoter sequences on NCBI (http://www.ncbi.nlm.nih.gov/nuccore/?term=taxus+cuspidata). The positive colony was then selected on ampicillin (Amp) plates. The promoter-*HIS3* vector was further linearized at *Xho* I site and integrated into yeast genome to construct *HIS3* reporter strains. The *P*minHIS3 enables *HIS3* to be constitutively expressed at a very low level in the absence of target elements; however, the level of *HIS3* expression is low enough to allow yeast transformants integrated with *HIS3* reporter vectors to grow on URA3 synthetic drop-out (SD/URA3$^-$) plates when constructing reporter strains. Therefore, the *HIS3* background can be eliminated by 3-AT. The other is the TF expression construct, pGADgate424 lapau. This vector was constructed by cloning a gateway cassette B into the pGAD424 backbone at *Sma* I site via blunt-end ligation (Fig 2.2c). This allows the cloning of TF open reading frames (ORF) downstream of *GAL4* activation domain (AD) driven by *P*ADH1 through the gateway cloning strategy (Deplancke et al. 2004; Hens et al. 2012). The resulting recombinant protein (GAL4-TFs) expression construct was introduced into yeast reporter strains by yeast transformation, and the transformants were selected on SD/Leu$^-$ plates (CLONTECH MATCHMAKER One-Hybrid System User Manual PT1031-1). All 19 putative TFs were isolated from the cDNA library of CMCs (Thomas Waibel). These primers are gateway-compatible and designed according to the TF sequecnes obtained from 454 pyrosequencing database (SI table 4).

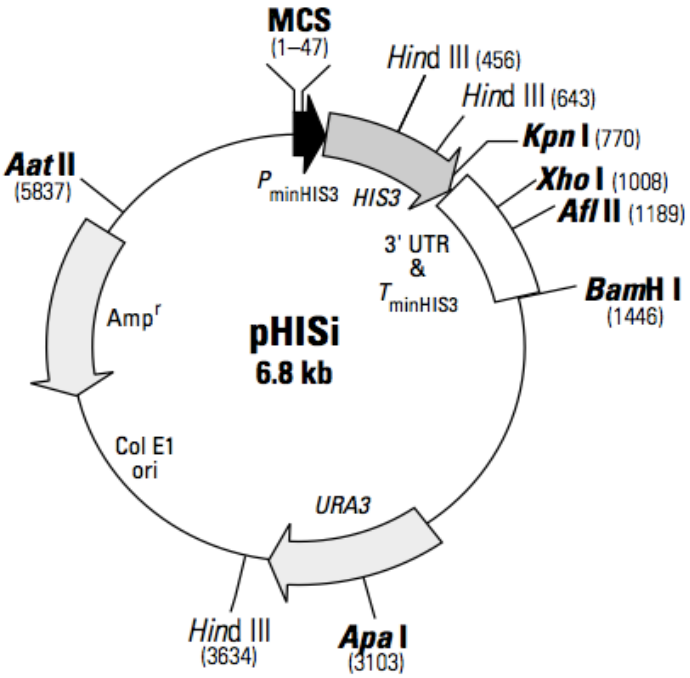


Fig 2.2a Map of pHISi (CLONTECH MATCHMAKER One-Hybrid System User Manual PT1031-1). The Multiple Cloning Site (MSC) is upstream of the *HIS3* reporter, into which the promoter fragments are inserted. *Xho* I is the site for linearization, followed by integration into YM4271 genome. *P*minHIS3 drives minimal constitutive *HIS3* expression. Amp^R allows for *E. coli* selection on ampicillin plates when constructing *HIS3* reporter vectors. URA3 is the selection marker for integration when constructing yeast reporter strains.

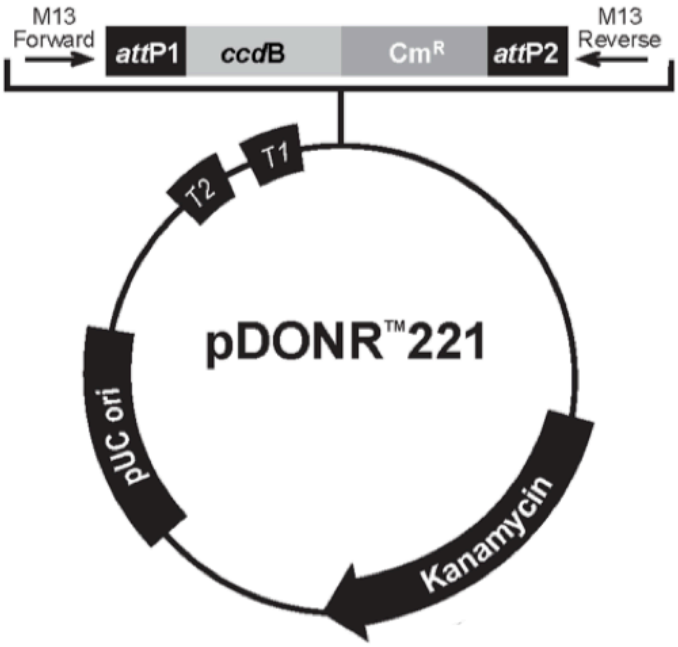


Fig 2.2b Map of pDONR^{TM} 221 (Life technology). *cddB* is flanked by attP1 and attP2 and a lethal gene to *E. coli*. Cm^{R}: chloramphenicol resistant gene; kanamycin: kanamycin resistant gene; T1 and T2 are rrnB T1 and T2 transcription termination sequence, respectively. M13 forward and reverse are priming sites.

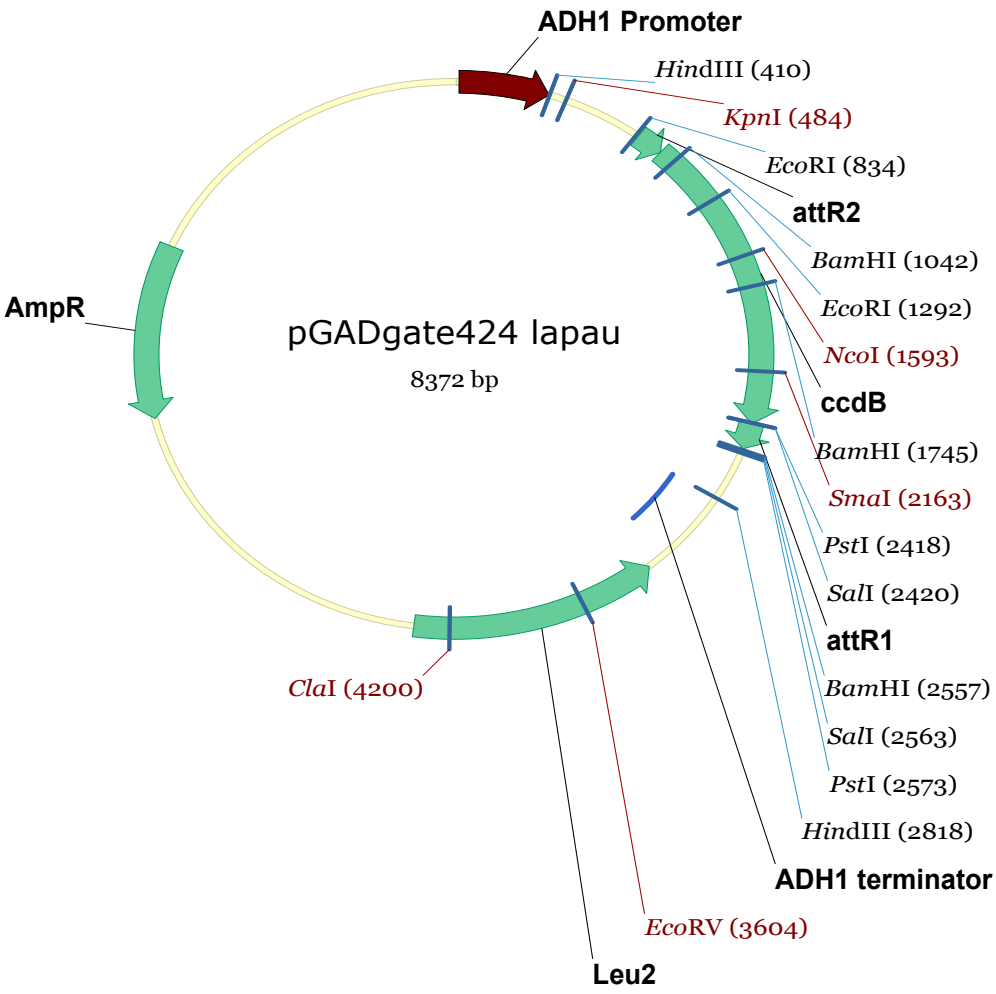


Fig 2.2c Map of pGADgate424 lapau (CLONTECH MATCHMAKER One-Hybrid System User Manual PT1031-1). *P*ADH1 is a constitutive promoter and drives the expression of GAL4 AD-TFs fusion protein. Amp^r allows for *E.coli* selection on ampicillin plates when constructing the Gal4 AD-TF expression vector. Leu2 is the selection marker for yeast transformation when transforming the recombinant protein expression vector into yeast reporter strains. Amp^R: ampicillin resistance; ADH1: alcohol dehydrogenases 1

2.3.1 Yeast transformation

Yeast strain used for transformation was YM4271 (His-, Leu-). YM4271 was cultured in 5 ml YPDA medium overnight. Then, ~3-4 ml of the overnight culture was diluted in 40 ml of YPDA and further incubated for 3 to 4 hours in a 30°C shaker. After incubation, it was centrifuged at 1000g for 5 minutes. The supernatant was discarded and the pellet was washed with 40 ml H_2O. Finally, the yeast pellet was resuspended in 600 µl of 0.5xTE/1xLiAc solution (200 µl of 3M Lithium acetate, 100 µl of 10x TE PH 7.5 and 1.7 ml of H_2O) to make yeast competent cells. 15 µl of plasmid DNA was mixed with 20 µl of denatured salmon sperm, 125 µl of yeast competent cells and 600 µl of 1xPEG/LiAC (8 ml of 50% PEG, 1ml of 10xTE PH7.5 and 1ml of 10xLiAc PH 7.5). The transformation mixture was then incubated for 60 minutes in a 30°C shaker. After incubation, 75 µl of DMSO was added into the mixture, followed by heat-shock for 40 minutes at 42°C. The heat-shocked mixture was incubated on ice for 2 minutes, followed by centrifugation at 14,000 rpm for 5 seconds. Finally, the pellet was resuspended in 200 µl of autoclaved H_2O and the whole suspended solution was plated on synthetic drop-out medium, followed by incubation at 30°C for 2-3 days (CLONTECH MATCHMAKER One-Hybrid System User Manual PT1031-1).

2.3.2 Yeast spotting assay

When screening the library of 19 putative TFs for their interactions with 5 promoter fragments, the background expression of *HIS3*, which is caused by the constitutive *P*minHIS3, is controlled by adding 3-Amino-1,2,4-triazole (3-AT). 3-AT is a competitive inhibitor of the product of the *HIS3*. Because it is heat-labile, it must be added into the autoclaved and cooled SD medium (~55°C). In this assay, the concentration of 3-AT used to test background *HIS3* expression of *HIS3* reporter strains is 45mM. According to the protocol, if yeast reporter strain only grows on SD/His- and <45 mM 3-AT, it indicates that the background *HIS3* expression is low, and therefore, this strain can be used for the second transformation with TF expression vectors. A single colony of yeast reporter strain was picked up and

suspended in 1 ml of TE buffer, then, a series of dilutions (0, 10^{-1}, 10^{-2} and 10^{-3}) were made. Furthermore, 5µl of each dilution was dropped onto SD/His- and 45 mM 3-AT plates, which were then completely dried and incubated at 30°C for 4-5 days (CLONTECH MATCHMAKER One-Hybrid System User Manual PT1031-1).

2.4 Cre/loxP recombination and gateway-based plant transient expression assays

Cre/loxP recombination engineering is used to recombine a binary promoter-reporter plasmid (pMDC162) (Fig 2.3a) with a TF expression vector (pJET1.2GWTF) (SF 4) through the loxP site (Nagy 2000; Shigaki et al. 2005). The resulting construct contains two independent gateway-compatible cassettes: one is the promoter-reporter; the other is the protein expression cassette. First of all, a protein expression cassette was amplified from pMDC32 (Fig 2.3b) and cloned into pJET1.2 backbone through blunt-end ligation. Then, a 34-base pair loxP site was inserted into each of the two vectors through digestion and ligation method. A loxP site was cloned into pMDC162 at *Pac* I site (Fig 2.3a), whereas it was cloned into pJET1.2GWTF at *Cla* I site (Fig 2.3c). The loxP fragment was directly synthesized using the overlapping primers (SI table 4). After confirming the presence of the loxP site in each vector by sequencing, in order to make a positive control for the plant transient expression assay, the *cddB* gene in pMDC162 and pJET1.2GWTF was replaced by the tandem repeat of 3xGCCGCC and open reading frame of ORA47 through gateway cloning, respectively. In addition to the use of the tandem repeat of ORA47 binding motif, a *35S* minimal promoter (*35Sm*/-49) is required to initiate reporter gene transcription. Therefore, a 51-base pair DNA fragment (*35Sm*) was directionally inserted upstream of 3xGCCGCC repeat at *Spe* I and *Hind* III site (Fig 2.3a). The *35Sm* fragment was directly synthesized using the overlapping primers (SI table 4). Finally, a mixture of 6 μl of pJET1.2GWORA47-loxP and 2 μl of pMDC162-3xGCCGCC-35Sm-loxP was incubated at 37°C for 1 hour in the presence of 2 μl of Cre recombinase (New England Biolabs). The mixture was then introduced into *E. coli* XL1-blue by transformation, and positive colonies were selected on ampicillin and kanamycin double selection plates. The recombinant plasmids (Fig 2.3d) were extracted using GeneJET Plasmid Miniprep kit (Thermo Scientific) and introduced into *Agrobacterium* (GV3101) through *Agrobacterium* transformation for *Agrobacterium*-mediated plant transient expression assay in tobacco (*N. benthamiana*) leaves.

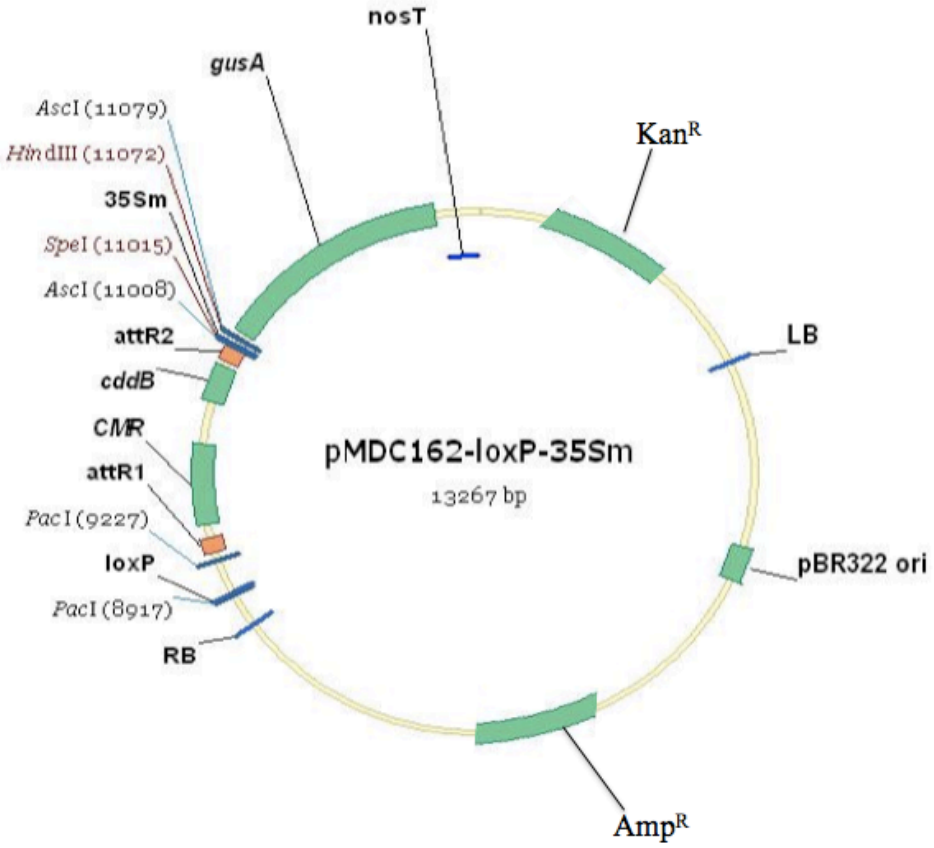


Fig 2.3a Map of pMDC162-loxP-35Sm. It contains a promoter-reporter cassette (attR1-*cddB*-attR2-gusA-nosT). loxP is inserted at *Pac* I site. 35Sm is cloned at *Spe* I and *Hind* III. nosT: *nos* terminator *GUS*: beta-glucuronidase. 35Sm: cauliflower mosaic virus (CaMV) 35S minimal promoter. Amp^R: ampicillin resistance. Kan^R: kanamycin resistant. LB: left border. RB: right border.

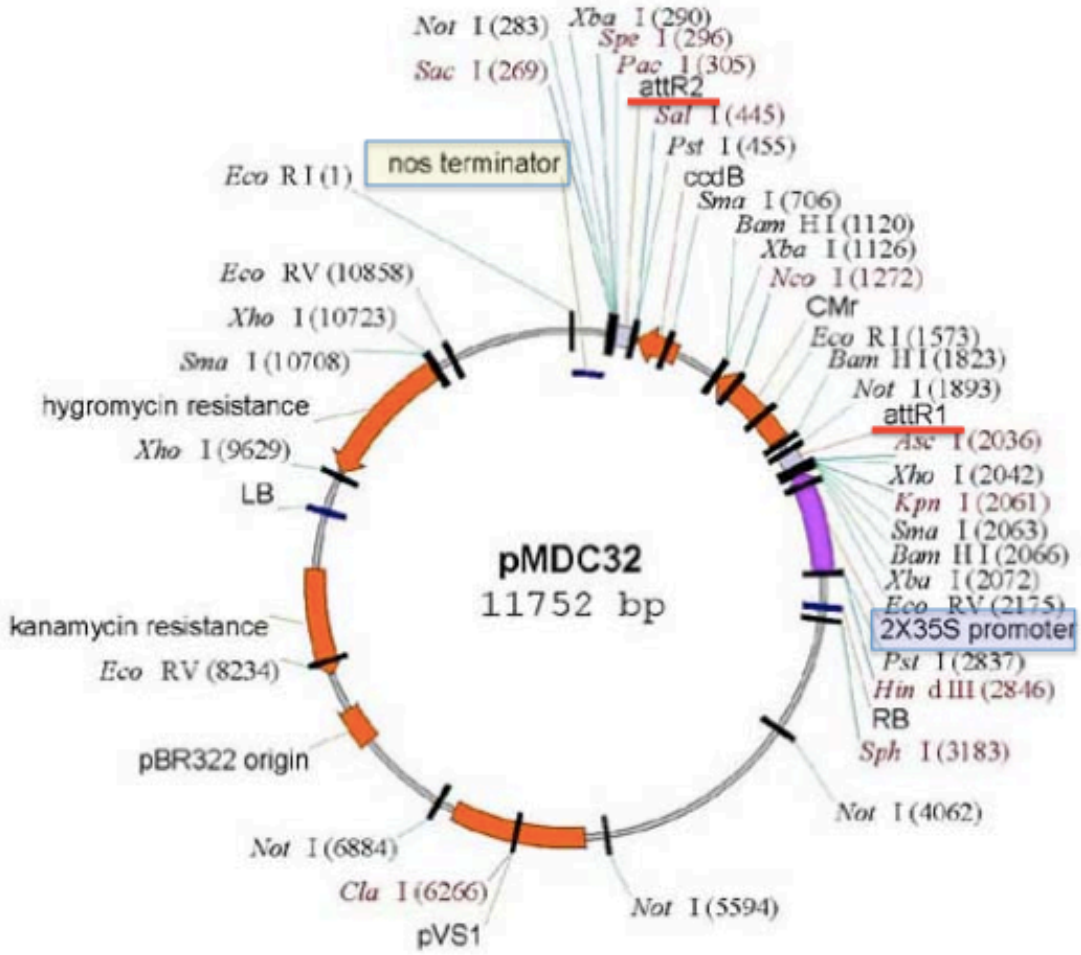


Fig 2.3b Map of pMDC 32 (Tair: The Arabidopsis Information Resourse). A protein expression cassette, 2x*35S*-attR1-*cddB*-attR2-nosT, was amplified and cloned into pJET1.2 backbone to give rise to pJET1.2GWTF (Fig 2.3c). nosT (highlighted in yellow box): *nos* terminator; 35S (highlighted in blue box): cauliflower mosaic virus (CaMV) 35S promoter. Kan^R: kanamycin resistant. Hyg^R: hygromycin resistance. LB: left border. RB: right border. AttR1 and R2 site are underlined in red.

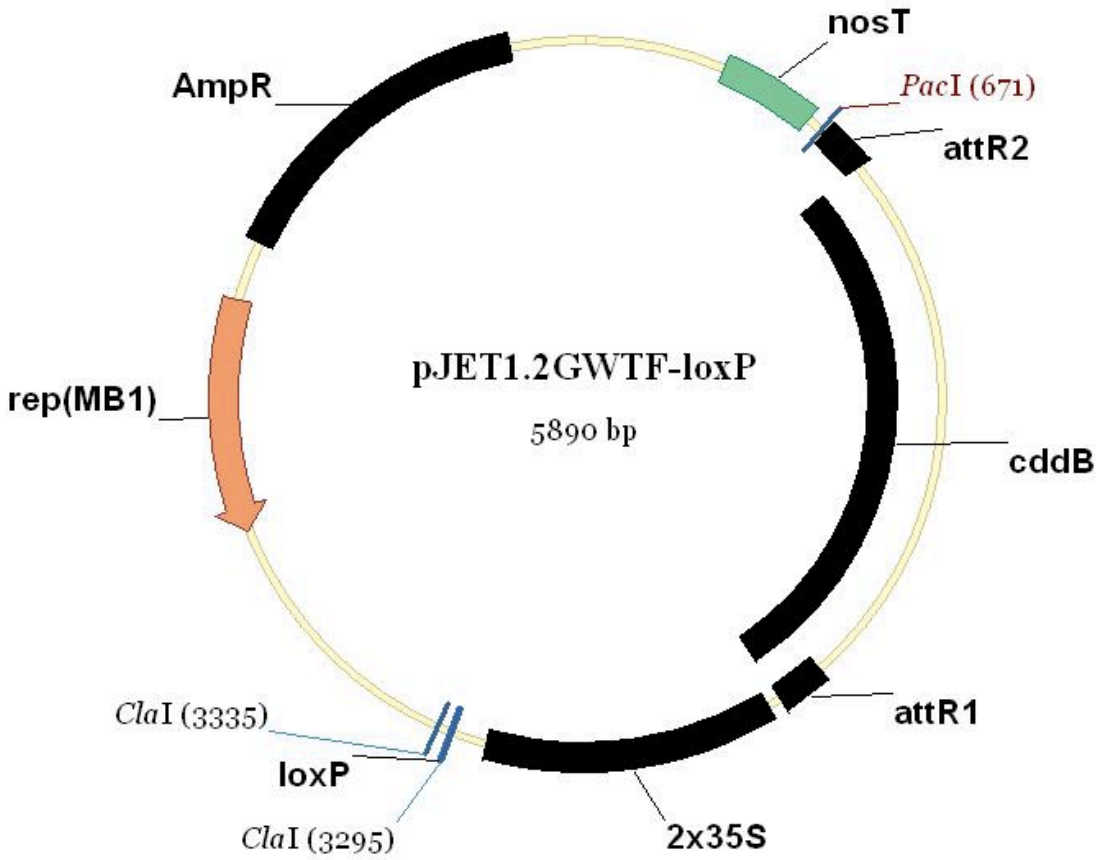


Fig 2.3c Map of gateway-compatible protein expression plasmid. The expression cassette (2x*35S*-attR1-*cddB*-attR2-nosT) is cloned into the backbone of pJET1.2. LoxP site is inserted at *Cla* I. 2x35S: cauliflower mosaic virus (CaMV) *35S* promoter. nosT: *nos* terminator. Amp^R: ampicillin resistance.

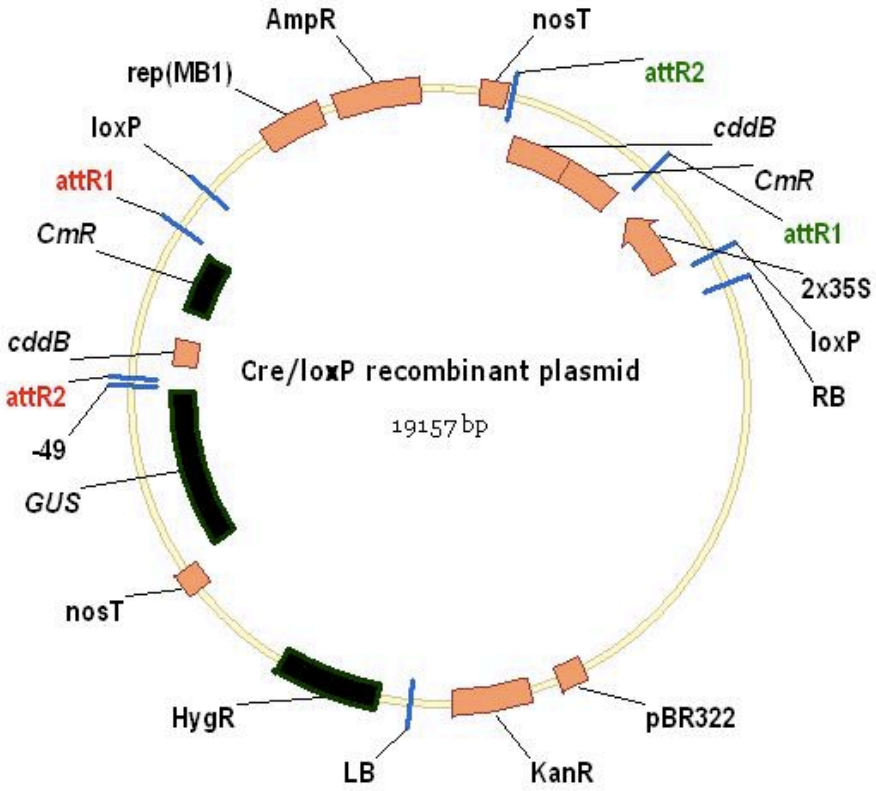


Fig 2.3d Map of a recombinant plasmid made by fusing pMDC162-loxP-35Sm (Fig 2.3a) with pJET1.2GWTF-loxP (Fig 2.3b) through loxP sites. It includes a TF expression cassette (2x35S-attR1-*cddB*-attR2-nosT) and promoter-reporter cassette (attR1-*cddB*-attR2-GUS-nosT). p35S: cauliflower mosaic virus (CaMV) *35S* promoter. T35S: CaMV *35S* terminator. nosT: *nos* terminator *GUS*: beta-glucuronidase. -49: *35S* minimal promoter. Sp^R: spectinomycin resistant. Hyg^R: hygromycin resistance. Amp^R: ampicillin resistance. Kan^R: kanamycin resistance. LB: left border. RB: right border.

2.5 Multi-gateway based plant transient expression assay

Because the Cre/loxP recombination based construction of a recombinant plasmid did not work (chapter 5), a multi-gateway cloning vector was constructed by cloning an expression cassette into the backbone of a binary gateway vector, which already contains a reporter cassette. Therefore, the resulting construct contains two gateway cassettes, and these two cassettes are flanked by two different sets of attR sites. The TF expression cassette is flanked by attR4 and attR3 sites, and the reporter cassette is flanked by attR1 and attR2 sites (Fig 2.4c). This allows a simultaneous cloning of TF encoding genes and promoter fragments into their corresponding cassette in the same vector. A p35S-R4-*cddB*-R3-t35S TF expression cassette was amplified from pK7m34GW2-8m21GW3,0 (Fig 2.4 a) (VIB) using both F_Sac I and R_Sac I primers (SI table 4) and subsequently cloned into a binary *GUS* reporter vector, pHGWFS7,0 (Fig 2.4 b) at *Sac* II site. As a result, the recombinant vector contains both p35S-R4-*cddB*-R3-t35S TF expression cassette and promoter fragments-R1-cddB-R2-GUS reporter cassette (Fig 2.4 c). Therefore, each of 5 promoter fragments and 19 putative TFs can be simultaneously cloned into the two-gateway cloning pDES vector through pDONR_P4-P3 and pDONR_P2-P1 (Fig 2.2b & 2.4d) (Life technology). In total, there were 95 constructs made to screen for TF-DNA interactions in plant transient expression assays. Before screening, a positive control was established to show the interaction between ORA47 and the tandem repeat of 3xGCCGCC.

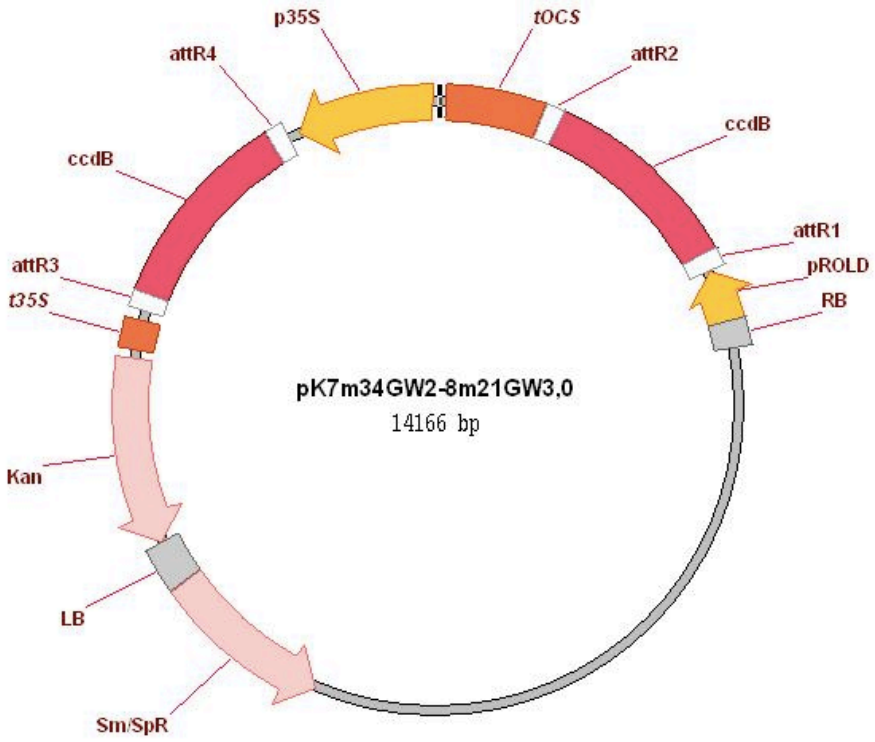


Fig 2.4a (VIB, a life sciences research institute, based in Flanders, Belgium) Two-gateway TF expression binary vector, pK7m34GW2-8m21GW3,0. It includes two TF expression cassettes: pROLD-attR1-*cddB*-attR2-tOCS and p35S-attR4-*cddB*-attR3-t35S. p35S-R4-*cddB*-R3-t35S TF expression cassette is amplified and cloned into the backbone of pHGWFS7,0 (Fig 2.4b) to make a two-gateway construct (Fig 2.4c). p35S: cauliflower mosaic virus (CaMV) *35S* promoter. T35S: CaMV *35S* terminator. *GUS*: beta-glucuronidase. Sp^R: spectinomycin resistance. Hyg^R: hygromycin resistance. Kan^R: kanamycin resistance. LB: left border. RB: right border.

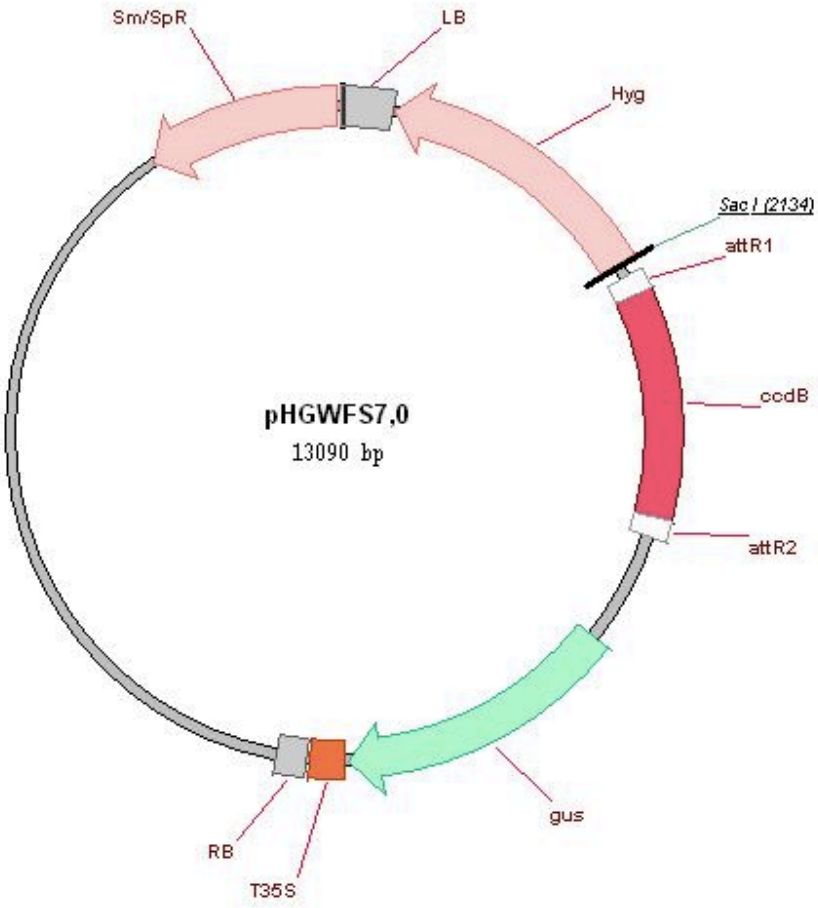


Fig 2.4b (VIB, a life sciences research institute, based in Flanders, Belgium) Map of gateway-compatible promoter-reporter binary vector. A gateway-compatible TF expression cassette is cloned at *Sac* I site to make a two-gateway recombinant construct (Fig 2.4c). T35S: cauliflower mosaic virus (CaMV) *35S* terminator. *GUS*: beta-glucuronidase. Sp^R: spectinomycin resistance. Hyg^R: hygromycin resistance. LB: left border. RB: right border.

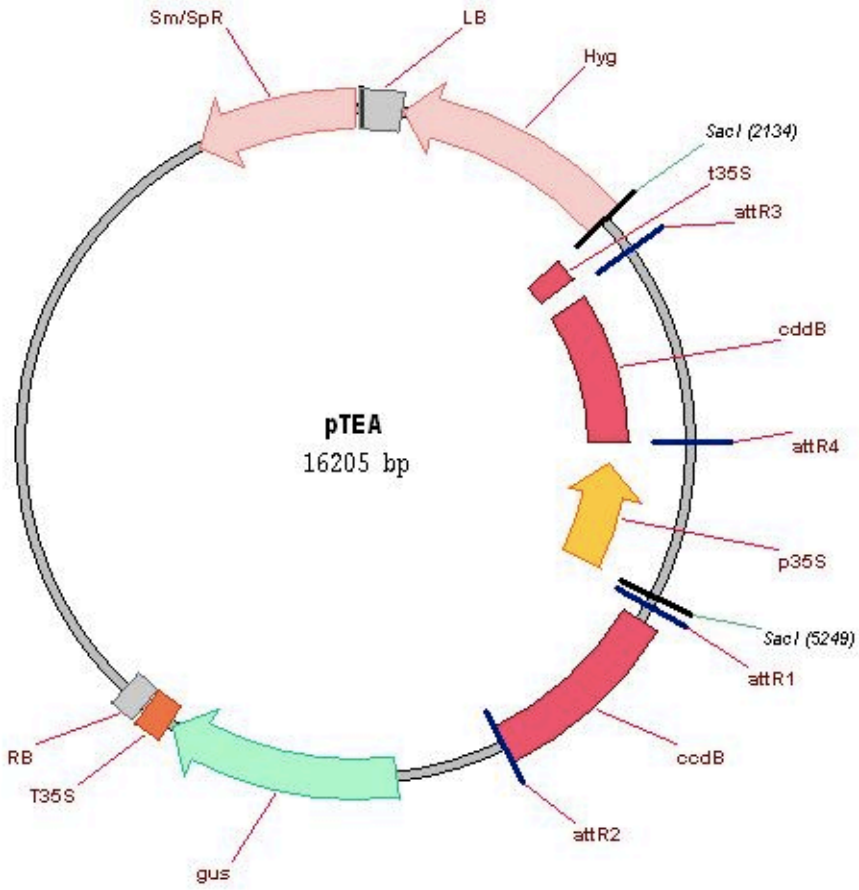


Fig 2.4c Map of two-gateway plant binary vector. It includes a TF expression cassette (p35S-attR4-*cddB*-attR3-t35S) and a promoter-reporter cassette (attR1-*cddB*-attR2-*gus*-T35S). The ORA47 TF and tandem repeats of AP2-binding motif (GCCGCC) replace cddB in TF expression cassette and cddB in promoter-reporter cassette through gateway cloning system, respectively. p35S: cauliflower mosaic virus (CaMV) *35S* promoter. T35S: CaMV *35S* terminator. *GUS*: beta-glucuronidase. Sp^R: spectinomycin resistance. Hyg^R: hygromycin resistance. LB: left border. RB: right border.

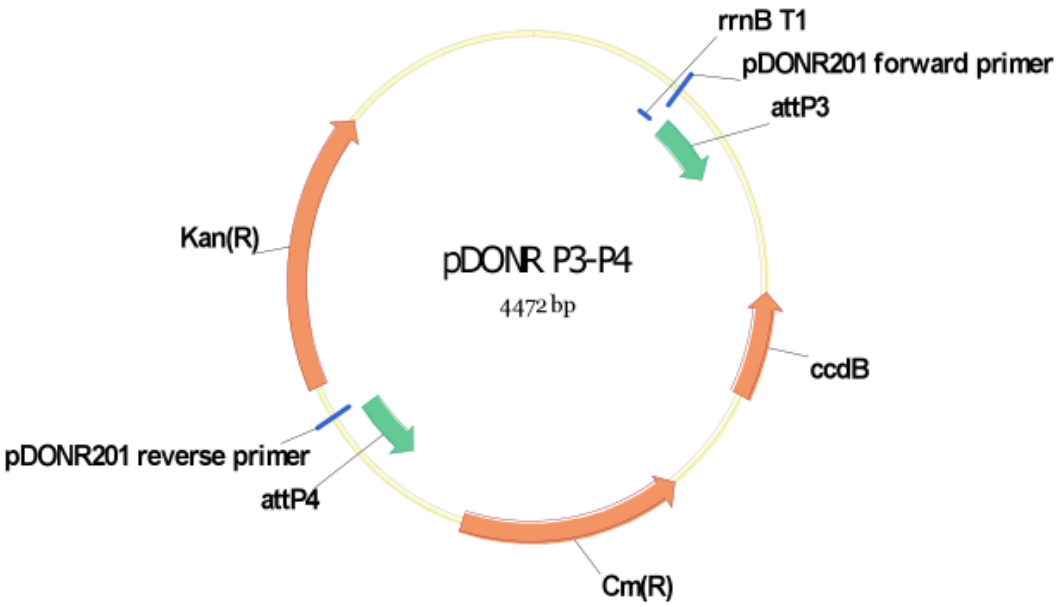


Fig 2.4d Map of pDONR_P3-P4 (Life technology). *cddB* is flanked by attP3 and attP4 and a lethal gene to *E. coli*; Cm^R: chloramphenicol resistant gene; Kan(R): kanamycin resistant gene; T1 and T2 are rrnB T1 and T2 transcription termination sequence, respectively. M13 forward and reverse are priming sites.

2.6 PCR-based synthesis of tandem repeats of TF-associated DNA binding motifs

In order to enhance the binding effect of between ORA47 (AP2 type TF) and plox3 (promoter of lipoxygenase gene), the tandem repeats of AP2-associated motif (GCCGCC) (<100 base pairs) were directly synthesized from primers by PCR amplification (SF 1). When designing primers, it is important to have overlapping base pairs between the forward and reverse primer (SI Table 4). The 50 μl PCR mixture contained 10 μl of Phusion buffer (5x), 2 μl of dNTPs (10 μM), 2 μl of each forward and reverse primer (10 μM), 0.3 μl of Phusion® High-Fidelity DNA polymerase (New England Biolabs®) and dH_2O (33.7 μl). In addition, the PCR condition shown in Table 5.

. The PCR products were then purified using GeneJET PCR purification kit (Thermo Scientific) and run on 2% agarose gel to check the quality. The PCR products were also cloned into a sequencing vector pJET1.2 using CloneJET PCR cloning kit (Thermo Scientific) and introduced into *E. coli* XL1-blue by transformation. The plasmids were extracted using GeneJET Plasmid Miniprep kit (Thermo Scientific) and sequenced using pJET1.2 Sequencing Primers (Thermo Scientific) in the Gene pool sequencing centre in Ashworth at King's Buildings. Using this method, the 34-bp loxP, 51-bp 35S minimal promoter and attB1-3xGCCGCC-35Sm were also synthesized and confirmed by sequencing (SF 1, 2 & 3).

Table 5 PCR conditions used in PCR-based synthesis of short DNA fragments

PCR conditions for synthesis of short DNA fragments			
Steps	**Temperatures**	**Time**	
1	98°C	1 min	
2	98°C	10 sec	Step 2 to 4 is cycled for 5 times
3	42°C	10 sec	
4	72°C	10 sec	
5	98°C	10 sec	Step 5 to 7 is cycled for 30 cycles
6	55°C	10 sec	
7	72°C	10 sec	
8	72°C	10 min	
9	**4**°C		

2.7 *Agrobacterium*-mediated plant transient expression assay

2.7.1 *Agrobacterium* transformation

The freeze-thaw method was used for transformation of *Agrobacterium* GV3101. Two constructs (pTEA-ORA47-3xGCCGCC-35Sm-GUS and pTEA-3xGCCGCC-35Sm-GUS) for plant transient expression assays were introduced into GV3101 by this method. Firstly, the competent *Agrobacterium* was thawed on ice, and 10 μl of plasmid DNA was added to 250 μl of the thawed *Agrobacterium*. Secondly, the mixture was incubated on ice for 5 minutes, then transferred into liquid nitrogen for 5 minutes and finally incubated in a 37°C water bath for further 5 minutes. Finally, 1 ml of Luria broth (LB) was added into the mixture, and the whole mixture was incubated at 28°C fro 2-4 hours with gentle shaking. *Agrobacterium* cells were collected by spinning briefly in a microcentrifuge and spread on LB agar plates that contain the appropriate antibiotic. The plates were incubated at 28°C for two days.

2.7.2 Preparing *Agrobacterium* infiltration solution

The transformed *Agrobacterium* cells GV3101 were grown in 4-5 ml of YEP medium for 1-2 days until the culture became cloudy. Then, the culture was centrifuged at 2500 rpm for 10-15 minutes at 4°C, and the pellet was resuspended in 1 ml of infiltration medium. The infiltration medium contains 1M $MgCl_2$, 0.5M MES, and 10mM acetosyringone. The absorbance at 600 nm of a 1 in 100 dilution of cells was measured and the OD of the original culture was calculated by multiplying the OD value by 100. The working solution with OD at 600 nm equalling 0.1 was therefore made from the original culture by calculating the dilution factors. After preparing the infiltration solution with OD_{600}=0.1, the samples were kept in dark for 2 hours at room temperature before infiltration (protocol provided by the Oparka lab).

2.7.3 Infiltrating *Agrobacterium* into epidermal layer of tobacco leaves

Tobacco leaves were chosen from healthy plants grown for 2-3 weeks in the growth chamber. Before infiltrating, the leaf sector was labelled for ease of identification. Then, a small region on abaxial side of the leaf to be infiltrated was gently rubbed to remove the wax cuticle with a small needle for ease of infiltration. The suspended *Agrobacterium* solution was taken up by 1ml syringe (no needle). The tip of the syringe was placed against the leaf over the rubbed area. The plunger was then gently pressed down while the adaxial side of the leaf being supported with fingers. After infiltrating, the infiltrated areas were outlined. Finally, the infiltrated plants were incubated in a growth cabinet for 48 hours to allow *GUS* gene expression (Sparkes et al. 2006).

2.8 GUS staining

The *Agrobacterium*-infiltrated tobacco leaves were first incubated with solution 1 (0.05 M NaHPO4, 0.3% formaldehyde, 0.5 mM $K_3Fe(CN)_6$, 0.5 mM $K_4Fe(CN)_6$, 1 mM EDTA and1% Triton X-100) at room temperature, and the solution 1 was removed after 1 hour. After rinsing the leaves with H_2O, the leaves were then incubated with rinsing solution (0.05 M NaHPO4, 0.3% formaldehyde, 0.5 mM $K_3Fe(CN)_6$, 0.5 mM $K_4Fe(CN)_6$ and 1 mM EDTA) for 20 minutes at room temperature. Finally, after removing the rinsing solution, the leaves were incubated in staining solution (0.05 M NaHPO4, 0.3% formaldehyde, 0.5 mM $K_3Fe(CN)_6$, 0.5 mM $K_4Fe(CN)_6$, 1 mM EDTA, 1% Triton X-100 and 2 mM X-Gluc) at 37°C overnight.

2.9 *In vitro* synthesis of transcription factors by wheat germ extract

TF proteins can be used in gel shift assays to detect TF-promoter interactions. Therefore, TF proteins firstly need to be synthesized. However, TF protein synthesis in *E. coli* did not produce a significant amount of protein required for further experiments (see chapter 7). So, *in vitro* protein synthesis by wheat germ extract was employed. Compared to protein synthesis in microbial hosts, it only takes one day to make a sufficient amount of desired proteins. The first step was to prepare the double-stranded DNA template by PCR. This amplification step consists of two round PCRs. The first round PCR added a linker fragment on each end of the open reading frame of TF encoding genes. The 50 μl PCR reaction mixture included 5x Phusion reaction buffer (10 μl), 4 μl of dNTP mixture (2.5 mM), 2.5 μl of each first round forward and reverse primer, template of TF encoding genes (2 μl), 0.5 μl of Phusion High Fidelity DNA polymerase (Thermo Scientific) and ddH_2O (28.5 μl). The first round PCR condition is listed in Table 6. The second round PCR further added T7 promoter and flag tag upstream of the open reading frame of TF encoding genes. The 50 μl reaction mixtures consisted of 5xreaction buffer (10 μl), 4 μl of dNTP mixture (2.5 mM), 2.5 μl of each SpT7u forward (10 μM) and Sp7dFLAG forward (100 nM) primer, 2.5 μl of each Reverse_U (10 μM) and deReverse (100 nM), 0.5 μ of template (PCR product from first round PCR), 0.5 μl of Phusion High Fidelity DNA polymerase (Thermo Scientific) and ddH_2O (23.5 μl). The second round PCR condition is shown in Table 6. The second step was to transcribe double-stranded DNA into sing-stranded mRNA *in vitro*. The 25 μl reaction mixture for in vitro transcription included 10x transcription buffer (2.5 μl), 2.5 μl of 25mM NTPs, RNase inhibitor (0.6 μl), 1.25 μl of 0.1M DTT, 1 μl of 1x T7 RNA polymerase (New England Biolabs), 2.5 μl of template (second round PCR product) and 14.65 μl of RNase free water (DEPC). The transcription reaction mixture was incubated for 3 hours at 37°C. After it, 1 μl of the sample was run on agarose gel for checking mRNA quantity and quality. Then, 10 μl of 4M ammonium acetate and 100 μl of ethanol were added to the sample, mixed, centrifuged for a few seconds and incubated at -20°C for 1 hour. After incubation, the sample was centrifuged at 15,000 rpm for 20 minutes at 4°C, followed by the removal of the supernatant. Finally, the

pellet was dried and dissolved in 35 μl of DEPC water. The final step was the translation of mRNA into proteins. The 55 μl of translation mixture consisted of 10 μl of wheat germ extract provided by the Tata lab, 10 μl of translation mixture (including 4x amino acid mixture), 35 μl of mRNA. The mixture was incubated at 16°C overnight. After it, the overnight sample was centrifuged at 15,000 rpm for 10 minutes at 4°C. The supernatant was collected and stored at -80°C (Fig 2.5) (Tada et al., unpublished).

Table 6 PCR conditions for synthesis of double-stranded *Taxus* TF encoding genes for *in-vitro* protein synthesis by wheat germ (Tada et al., unpublished).

First round PCR conditions			
1	98°C	30 sec	Step 2 to 4 is cycled for 40 cycles
2	98°C	5 sec	
3	65°C	30 sec	
4	72°C	30 sec	
5	72°C	10 min	
6	4°C		
Second round PCR conditions			
1	98°C	30 sec	Step 1 to 3 is cycled for 10 cycles
2	98°C	5 sec	
3	65°C	30 sec	
4	72°C	30 sec	Step 4 to 6 is cycled for 40 cycles
5	98°C	5 sec	
6	65°C	30 sec	
7	72°C	30 sec	
8	72°C	10 min	
9	4°C		

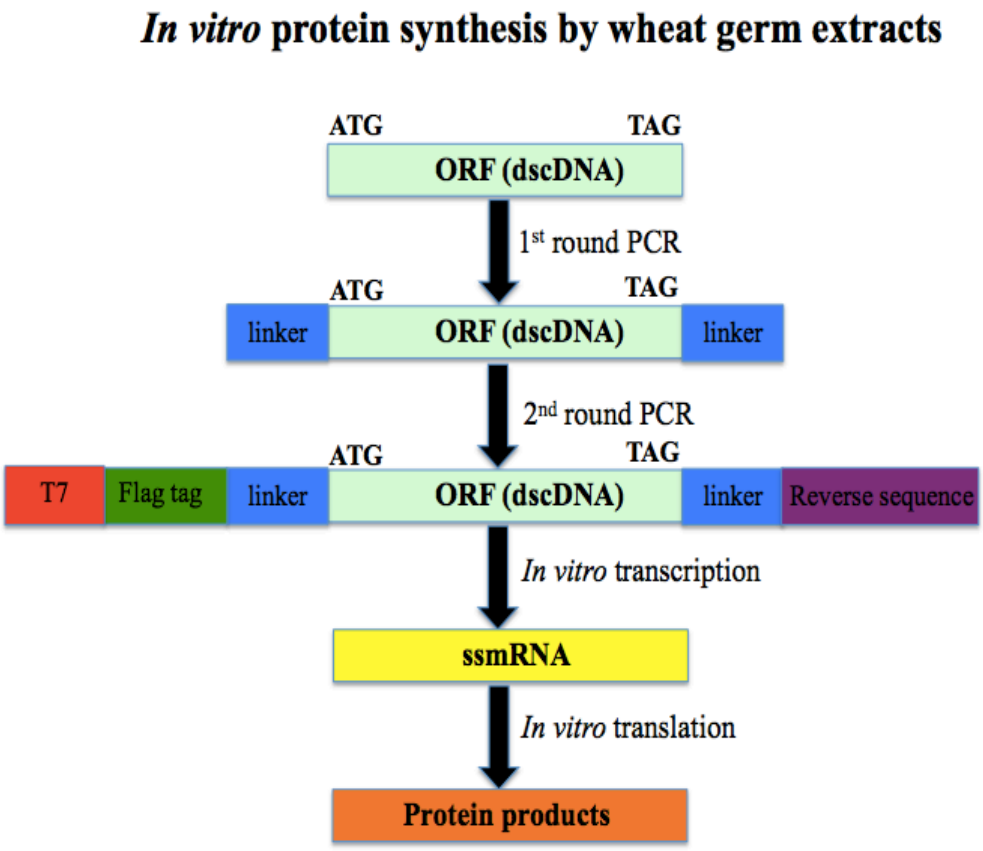


Fig 2.5 (Tada et al., unpublished) A schematic representation of in vitro protein synthesis by wheat germ extracts.

2.9.1 SDS-PAGE Gel and Western blots

The wheat germ-synthesized proteins (20 µl for each sample contains 5µl of protein, 4 µl of SDS, 1 µl of dithiothreitol (DTT) and 10 µl of H_2O) were run on 10% SDS-PAGE (Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis) gel in 1x running buffer, which is diluted from 10x stock solution (20 mM Tris from Sigma, 200 mM glycine from Sigma and 0.1% SDS from Sigma). The method for making both stacking and running gel is listed in Table 7. After completion of electrophoresis, protein samples on the gel were transferred onto the nitrocellulose membrane in 1x transferring buffer (TB) (14.4g of glycine from Sigma, 3.03g of Tris

base from Sigma and 200ml methanol from Sigma in 1 L of distilled water). After transferring, the membrane was incubated in 25 ml of blocking buffer (BB) for 1 hour at room temperature (5% w/v milk, 0.1% tween 20 from Sigma and 1x phosphate buffered saline (PBS) to adjust the final volume). After washing the membrane with BB for three times, 10 minutes each time, the membrane was incubated with the primary antibody (anti-FLAG from Sigma) in 1: 5000 dilutions in 5 ml of BB for 1 hour at room temperature, followed by washing with BB for three time, 20 minutes each time. Then, the secondary antibody (anti-mouse IgG from Sigma) was applied to incubate with the membrane in 1: 1000 dilution in 10 ml BB for 1 hour at room temperature, followed by washing with BB for three time, 20 minutes each time. Finally, the membrane was washed with PBS twice, 10 minutes each time. The membrane was then exposed for 10 seconds before development using Konica SRX-101A.

Table 7 The protocol for making stacking/running gel

Ingredients	Running gel	Stacking gel
30% acrylamide mix	3.3 ml	0.33 ml
1.5 M Tris pH8.8	2.5 ml	0.25 ml
10% SDS	0.1 ml	0.02 ml
10% ammonium persulphate (APS)	0.1 ml	0.02 ml
TEMED	0.004 ml	0.002 ml
H_2O	4 ml	1.4 ml
Total	10 ml	2 ml

Chapter 3 Analysis of MeJA-elicitated transcriptome of CMCs

3.1 Experimental design of transcriptomic profiling of MeJA induced CMCs

The cell suspension culture (csc) of CMCs from *T. cuspidata* was the starting material for transcriptome sequencing. Prior to MeJA elicitation, the csc of *T. cuspidata* CMCs (5 days after subculturing) was adjusted with fresh medium to 100 ml from 40 ml of CMC suspension culture and grown for 2 h in the dark to acclimatise. The cultures were then treated with 100 μM MeJA. The samples were taken in biological triplicates at three post elicitation (p.e.) time points, including 0.5 h, 2 h and 12 h. In addition, untreated samples taken at 0 h were used as the control. The total RNA of each sample (12 in total) was extracted and treated with DNAse. These RNA samples were used for Illumina/Solexa sequencing. Furthermore, for Roche/454 sequencing, the equal amounts of total RNA were used for subsequent preparation steps, including cDNA synthesis and sample normalization (Fig 3.1) (Thomas Waibel). Finally, the data generated by these two sequencing technologies were statistically analysed and annotated.

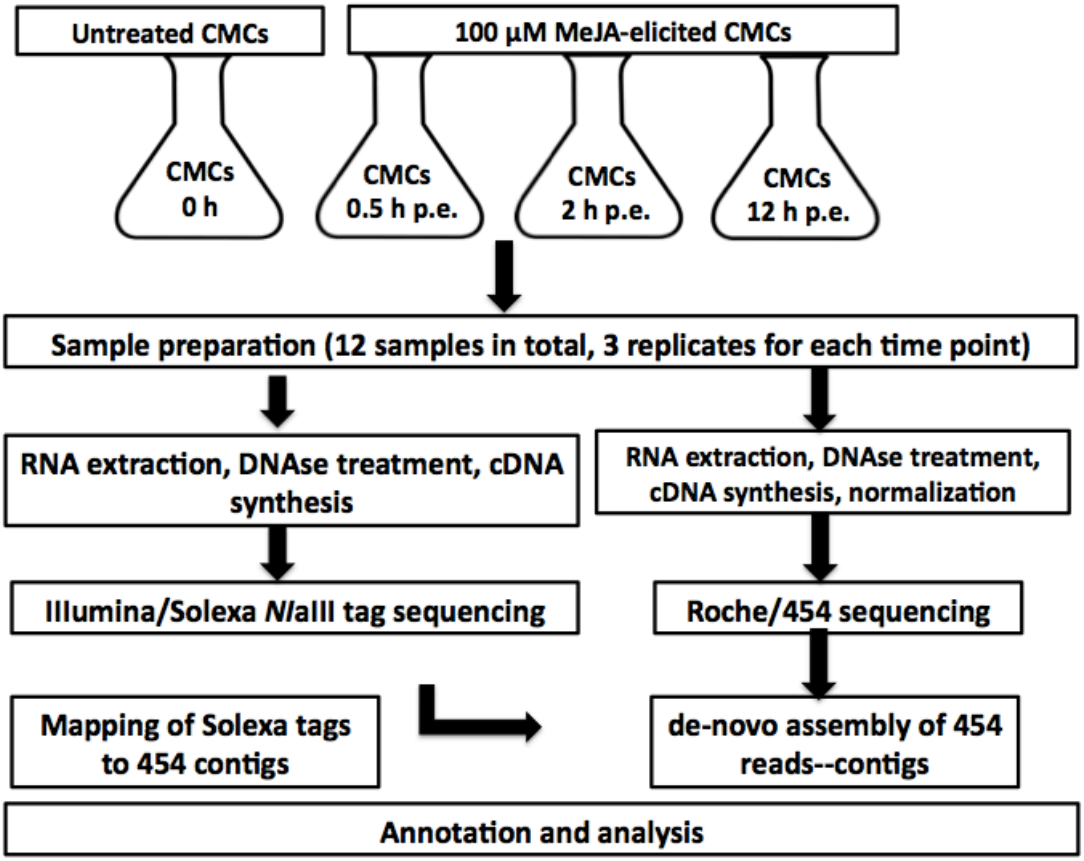


Fig.3.1 Schematic illustration of transcriptome sequencing of MeJA-elicited CMCs.

3.2 Transcriptome of MeJA elicited CMCs

Provided that contigs show differential gene expression in at least one of the three employed post-elicitation time points (0.5h, 2h and 12h), statistical analysis was carried out to compare the gene expression of MeJA-treated CMCs at each time point with that at 0h. This resulted in a list of 1646 differentially expressed contigs with a significant false discovery rate (FDR)<=0.05 (5%) (Simon Tomlison). FDR is a statistical method used to correct for multiple comparisons in multiple hypothesis testing. It represents the expected proportion of false positive among all statistically significant discoveries. In our Solexa data, the transcript counts for each significantly induced contig at one-time point are normalised and transformed to transcripts per million (TPM). Therefore, FDR (<=0.05) means that less than 50000 counts would be expected to be false positive among 1 million of transcript counts, which is statistically significant (Benjamini 1995). The data set of 1646 differentially expressed genes (DEGs) was used to perform hierarchical clustering analysis, which is visualised as a dendrogram and heatmap (Fig 3.2) (Florian Halbritter). Overall, this map shows the sample integrity of the biological triplicates, in which all 12 samples are mapped as close to each other as possible based on the gene expression pattern of MeJA-regulated 1646 contigs. In cluster A, all five samples, which include CMC #1, #2 and #3 at 0.5h and CMC #1 and #3 at 2h, exhibit relatively similar gene expression pattern. In addition, in cluster B, CMC #3 at 12h exhibits the most similar pattern to samples in cluster A. In contrast, CMC #2 shares less similarity with samples in clade A. All three CMC samples at 0h in cluster C exhibit relatively similar gene expression pattern between each other. However, the gene expression pattern of CMC #2 at 2h and CMC #1 at 12h is exceptionally different from that of the rest.

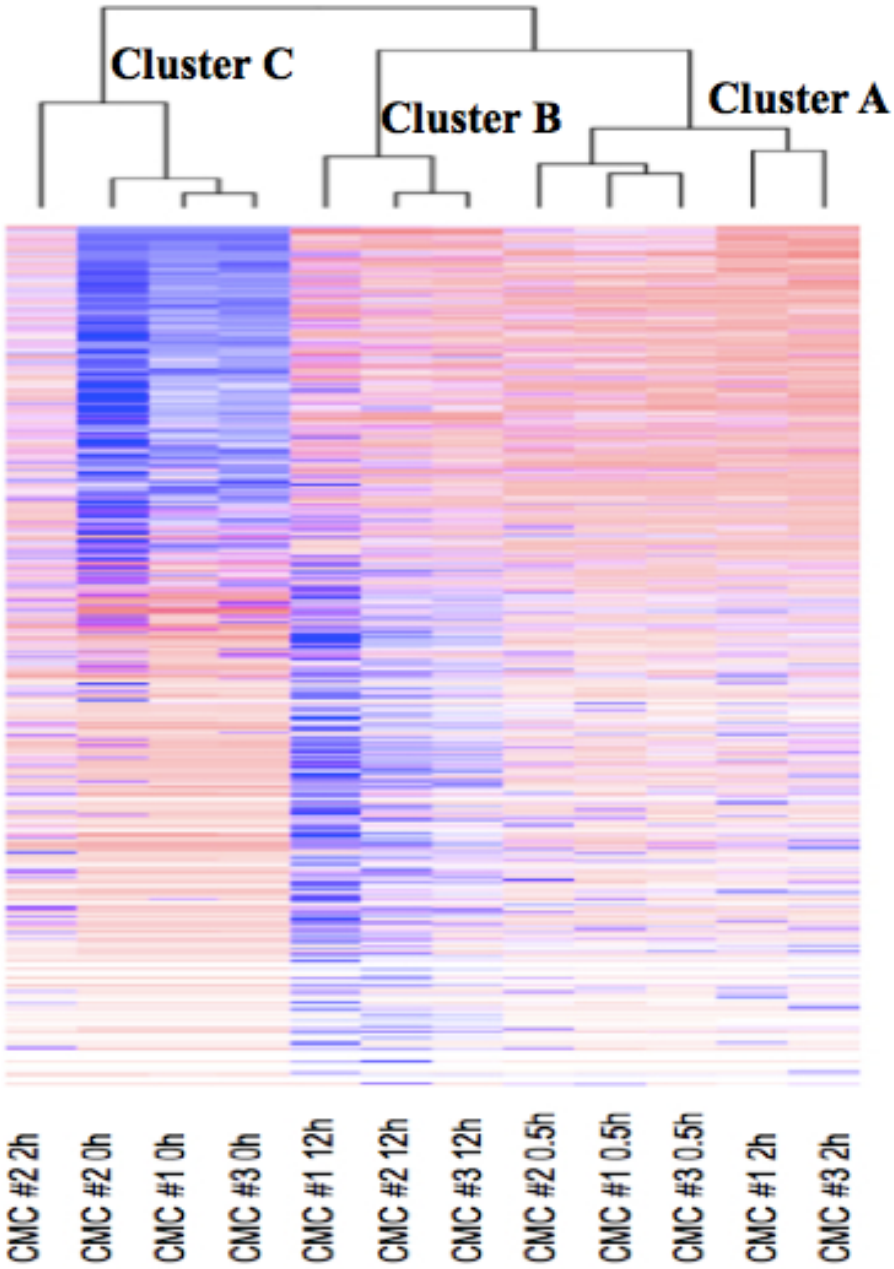


Fig. 3.2 Heatmap and dendrogram of 1646 DEGs identified in response to MeJA treatment in biological triplicates at three post-elicitation time points (Florian Halbritter). 12 MeJA-induced samples are clustered in 3 clusters based on the gene expression pattern of MeJA-regulated1646 contigs. Red indicates up-regulated genes; blue denotes down-regulated genes in MeJA-treated *T. cuspidata* CMC samples.

3.3 MeJA-responsive putative TF encoding genes

In the list of 1646 DEGs, 78 contigs were annotated as putative TF encoding genes. Based on the conserved domains found in their amino acid sequences by using Blast in NCBI, these 78 putative TFs were categorised into 19 different TF families (Fig 3.3a). Among them, the AP2/ERF group has the largest number of members, followed by MYB, bHLH, C3H, NAC and LOB TF groups. TF groups that contain less than 5 members include WRKY, GRAS, bZIP, E2F, CCAAT, IAA13, GATA, Trihelix, PHD, C2H2, ARF, FHA and SBP. In addition, Fig 3.3b shows the number of up- or down-regulated TF members in each family. In total, there are 50 up- and 28 down-regulated TFs, respectively. Among the up-regulated TFs, AP2/ERF (6), MYB (7), NAC (5), bHLH (5), LOB (6), C3H (6), WRKY (3), E2F (3) and GRAS (3) groups contain at least three members, whereas the rest of groups have only one or none.

Apart from Trihelix and SBP TF family, at least one member in each family has been assigned with a Gene Ontology (GO) term (Fig 3.3c). In total, there are 51 TFs, each of which has more than one GO term assigned. On the basis of GO terms, these TFs are further grouped into three different but inter-connected categories: cellular location, molecular function and biological process (Fig 3.3d). 36 TFs belonging to 12 different TF families are all found in these three categories (Fig 3.3e). The GO terms assigned to these 36 TFs show that they locate in nucleus (cellular location), have DNA binding and transcription factor activities (molecular function) and are involved in regulation of transcription (biological process). Together, these indicate that these proteins have TF-like properties. In 12 TF families, the AP2/ERF family contains 10 members, followed by NAC with 6 members, LOB with 5 members, bHLH and bZIP with 3 members each, MYB and WRKY with 2 members each, and C2H2, CCAAT, E2F, FHA and GATA with only one member each. A typical TF protein contains both a DNA binding and a transcriptional activation or repression domain. However, according to the assigned GO terms, one member in AP/ERF, 2 members in GRAS and one member in IAA13 family show that they are present in the nucleus, participate in transcriptional regulation, have transcription factor activity, but have no DNA binding activity. Thus, it seems that these TFs contain

transcription domain only, and therefore, they might need to interact with other proteins that contain DNA binding domains in order to regulate gene expression. Meanwhile, these could be assembly errors of sequencing data. Fig 3.3f shows the number of up- and down-regulated TF encoding genes in each of 12 families.

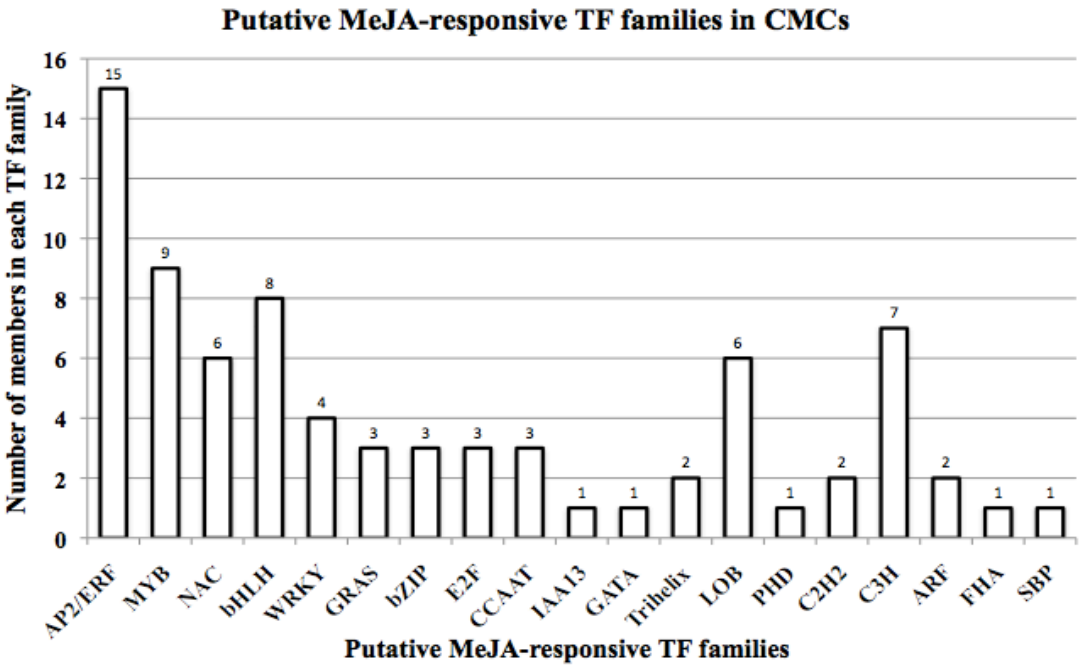


Fig 3.3a Number of TF members in each of 19 differentially regulated TF families in MeJA-treated CMCs.

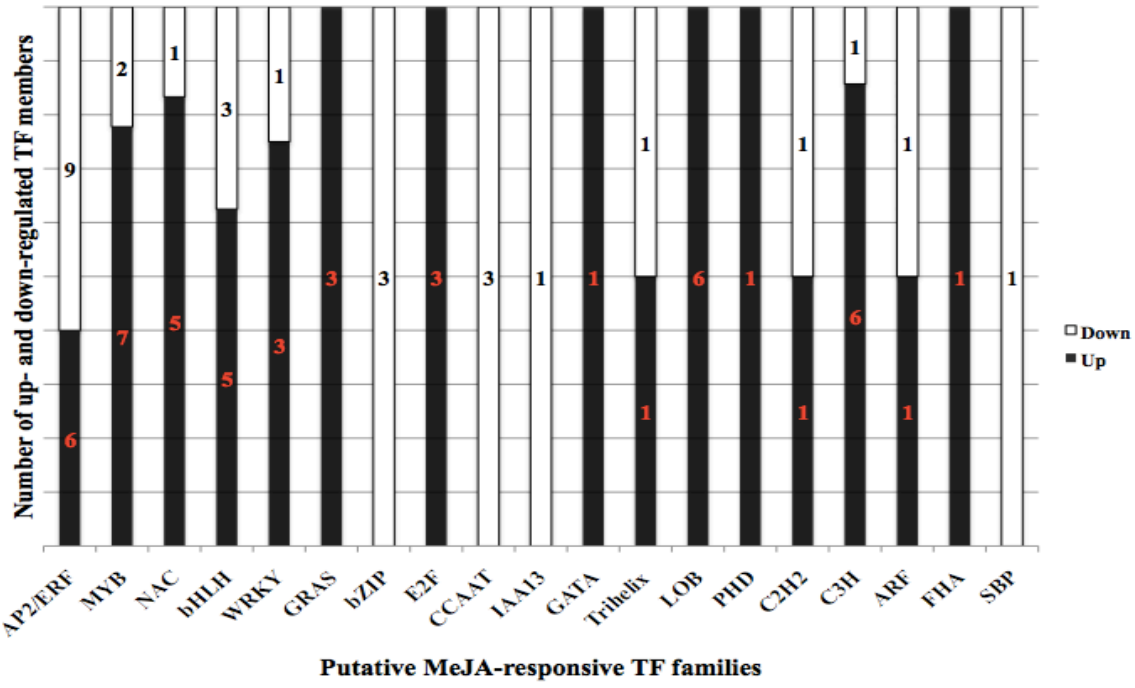


Fig 3.3b Number of up- or down-regulated TF members in each TF family (In total, 78 putative TF encoding genes) in response to MeJA in CMCs.

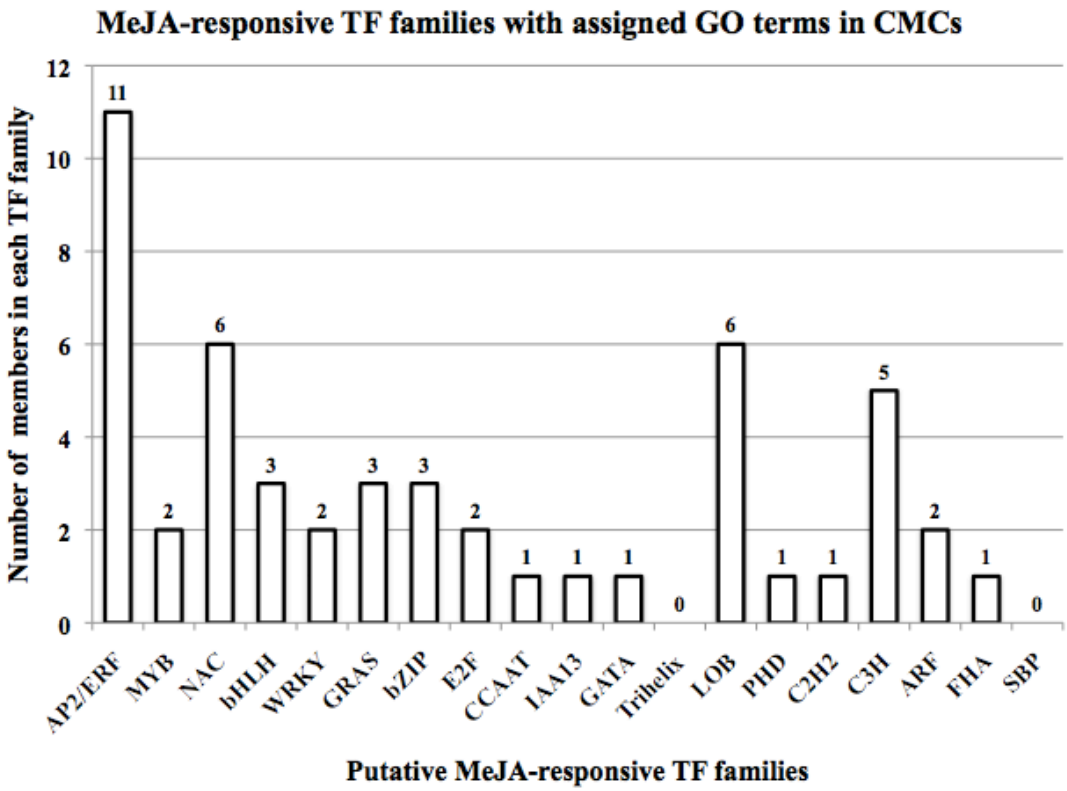


Fig 3.3c Number of MeJA-responsive putative TFs assigned with GO terms in each family. GO: gene ontology.

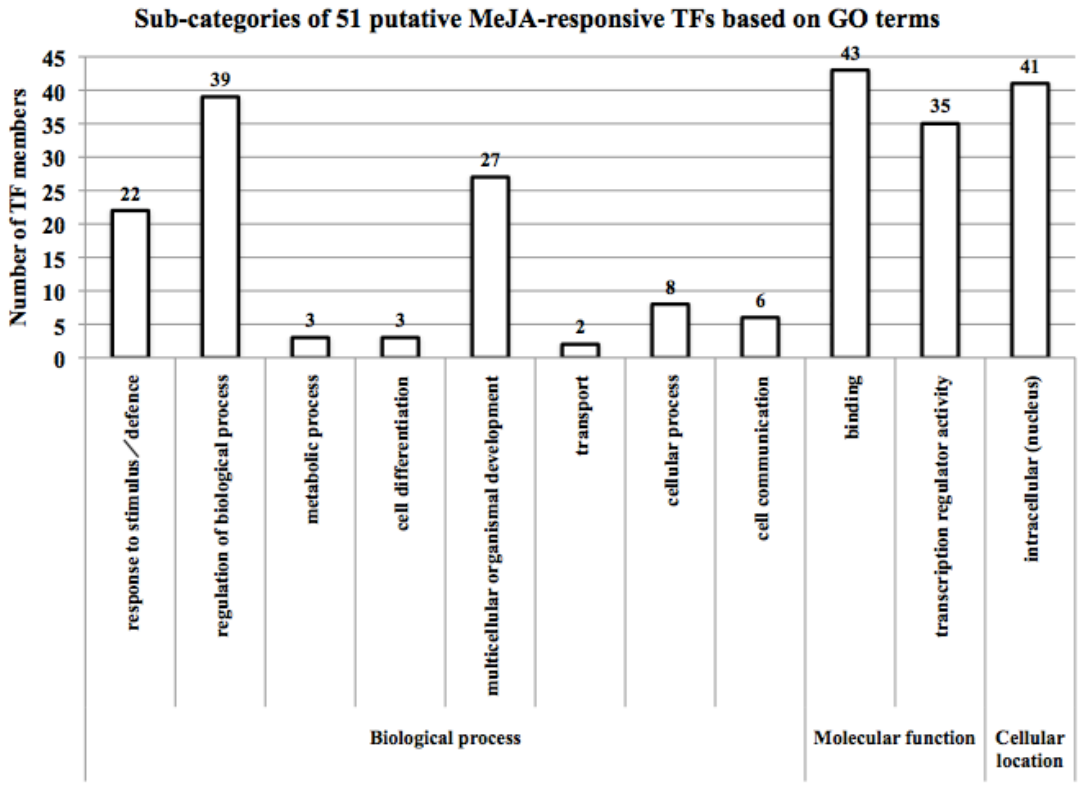


Fig 3.3d 51 putative MeJA-induced TFs assigned with GO terms are further classified into three different categories: (d) biological process, (e) molecular function and (f) cellular component.

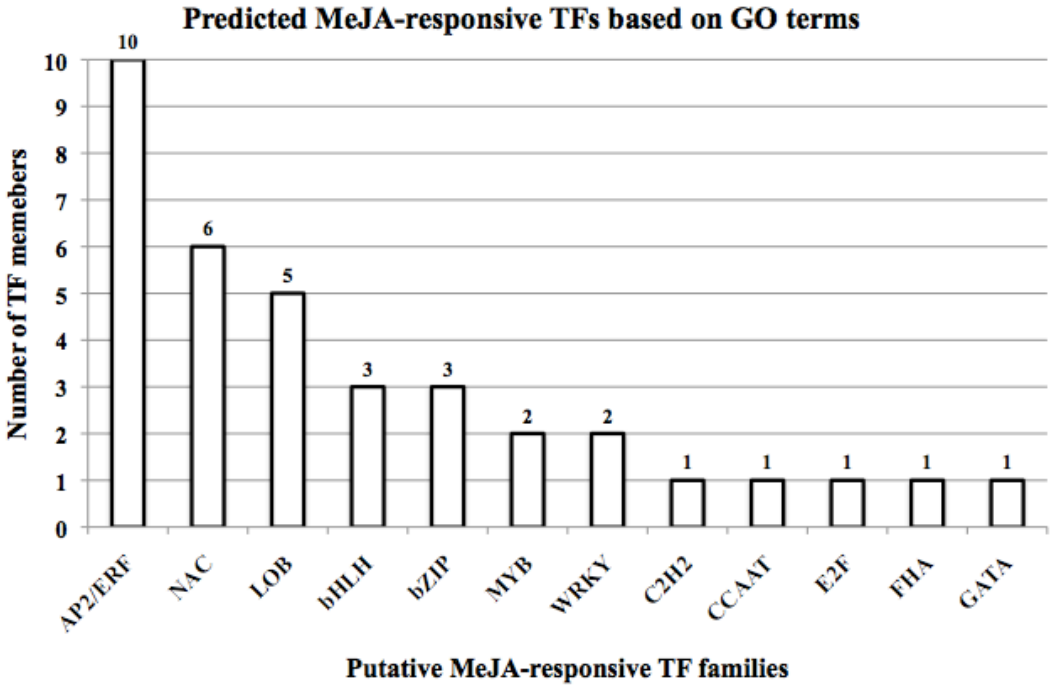


Fig 3.3e 36 TFs belonging to 12 TF families are predicted to be involved in MeJA-induced transcriptional regulation of gene expression.

In addition, there are 6 ZIM motif-containing proteins annotated as putative JAZ proteins. All 6 proteins are up-regulated upon MeJA treatment, and 4 of which are assigned with GO terms. As expected, these putative JAZ proteins are found in nucleus (Thines et al. 2007) and are able to interact with other proteins involved in the regulation of transcription mediated by JA signalling, such as MYC proteins (Cheng et al. 2011; Fernandez et al. 2011; Niu et al. 2011; Zhang et al. 2012)

3.3.1 Bioinformatic analysis of MeJA up-regulated putative TFs

Among 36 differentially regulated TF encoding genes by MeJA at all three time points (0.5, 2 and 12 h), 22 are up-regulated and 14 are down-regulated (Fig 3.3h). 15 out of the 22 belong to 5 commonly studied TF families, which have been reported to be involved in the regulation of secondary metabolite biosynthesis (Martin and Paz-Ares 1997; Mitsuda et al. 2007; Suttipanta et al. 2011; van der Fits and Memelink 2000; Zhang et al. 2012). In addition, there are 4 more putative TF encoding genes (two AP2/ERFs and two MYBs), which are not assigned with GO terms but are also significantly induced by MeJA. Furthermore, these genes blasted using Blastp in NCBI all show high E-values ($<e^{-04}$) and contain conserved DNA binding domains (DBD) in their amino acid sequences. Therefore, in total, there are 19 MeJA up-regulated genes that encode putative TFs (Table 8). The Solexa data show the up-regulation of these genes at the early post-elicitation time points (0.5 h) relative to 0 h. Contig07245 and contig09658 show down-regulated gene expression at 12 h, whereas the rest of all contigs show up-regulated gene expression at all three time points (Fig 3.4A). This was further confirmed by RT-PCR results (Fig 3.4B).

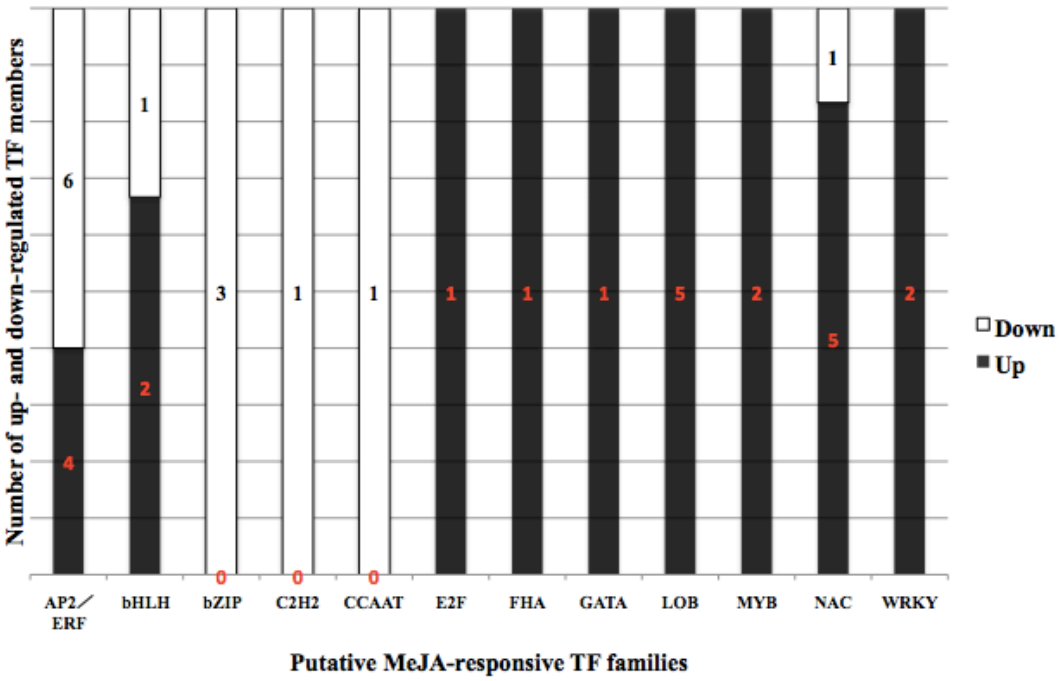


Fig 3.3f Number of up- and down-regulated TF members in each family (In total, 36 putative TF encoding genes are found in all three sub-categories based on their GO terms) upon MeJA treatment.

Table 8 19 MeJA up-regulated putative *Taxus* TF encoding genes were selected for biological assays based on E-values. They are obtained using BLASTp to blast their amino acid sequences against GeneBank non-redundant protein sequences in National Centre for Biotechnology Information (NCBI). (http://blast.ncbi.nlm.nih.gov/). This program can find similar protein sequences to the target sequence. The S score is a measure of the similarity of the query to the sequences shown in the blasted list. The E-value is a measure of the reliability of the S scores, and it is defined as the probability due to chance, that there is another alignment having a similarity greater than the given S score. The threshold for E-value in my analysis is e-04=(10^{-4}) or lower. Therefore, an $E<e^{-04}$ of an alignment means that this alignment is highly unique. The E-values for 19 putative *Taxus* TFs are equal to or less than the threshold, which indicates that the corresponding alignments for each TF are highly unique.

Contigs	Gene homology	Score (S)	E-value	Conserved DBD	TF family
03304*	ethylene-responsive transcription factor 10 [*Arabidopsis thaliana*]	45.8	2e-04	AP2	AP2/ERF
04485	ethylene-responsive transcription factor ERF018-like [*Cucumis sativus*]	125	6e-33	AP2	AP2/ERF
07245*	AP2/ERF domain-containing transcription factor [*Populus trichocarpa*]	92.4	4e-20	AP2	AP2/ERF
01431	Ethylene-responsive transcription factor ERF113 [*Aegilops tauschii*]	124	4e-32	AP2	AP2/ERF
00499	Ethylene-responsive transcription factor [Medicago truncatula]	112	2e-27	AP2	AP2/ERF
22386	Ethylene-responsive transcription factor [*Medicago truncatula*]	126	7e-31	AP2	AP2/ERF
12379	Tannin-related R2R3 MYB transcription factor, partial [*Trifolium repens*]	174	9e-49	MYB	MYB
10385	R2R3-MYB transcription factor MYB12 [*Picea glauca*]	247	7e-76	MYB	MYB
10855*	R2R3 MYB [*Solanum lycopersicum*]	238	8e-70	MYB	MYB
15401*	MYB transcription factor MYB99 [*Glycine max*]	162	2e-45	MYB	MYB
06771	NAC protein 1 [*Elaeis guineensis*]	276	1e-87	NAM	NAC
09658	NAC domain class transcription factor [*Malus x domestica*]	112	3e-27	NAM	NAC
05638	NACa protein [*Ginkgo biloba*]	356	4e-117	NAM	NAC
00172	NAC transcription factor [*Arachis hypogaea*]	288	1e-90	NAM	NAC
08447	NAC protein 1 [*Elaeis guineensis*]	308	2e-99	NAM	NAC
11748	transcription factor bHLH27 [*Arabidopsis thaliana*]	95.1	1e-20	bHLH	bHLH
08058	JAMYC [*Taxus cuspidata*]	231	4e-68	bHLH	bHLH
09595	transcription factor WRKY [*Taxus wallichiana var. chinensis*]	297	2e-99	WRKY	WRKY
19284	transcription factor WRKY [*Taxus wallichiana var. chinensis*]	169	1e-46	WRKY	WRKY

* Indicates genes that are not assigned with GO terms.

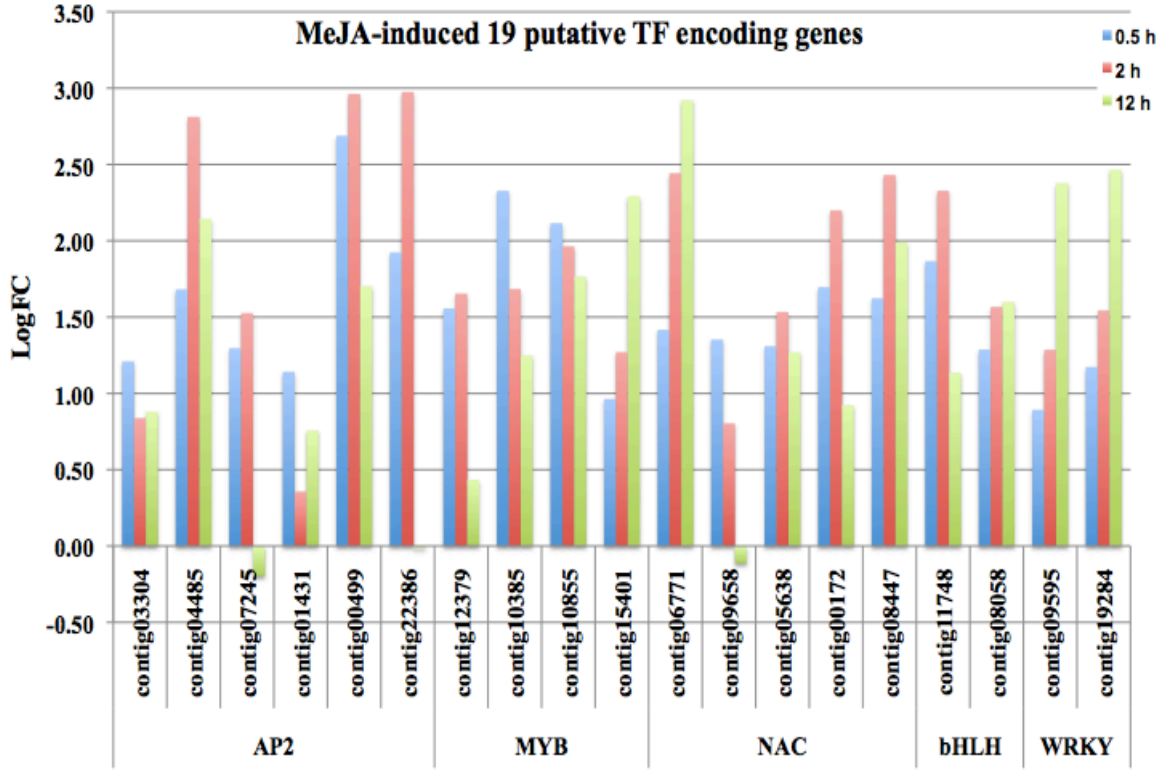


Fig 3.4A Fold change in the expression of 19 putative TF encoding genes belonging to 5 different TF families in at least one of the three post elicitation time points (0.5, 2 and 12 hour) relative to 0 hour. Log FC is the fold change on the log2-scale between the normalised and stabilised transcript counts.

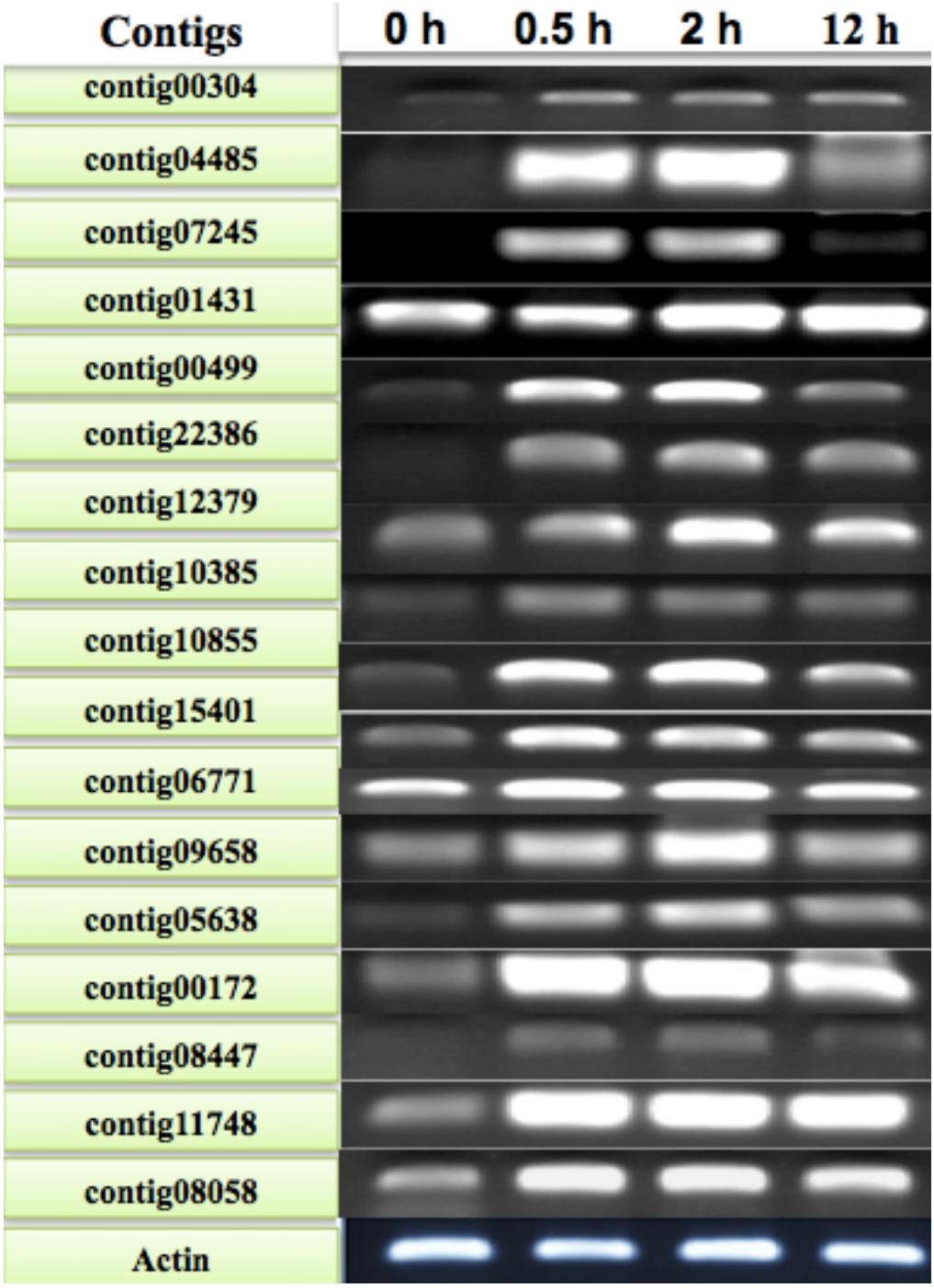
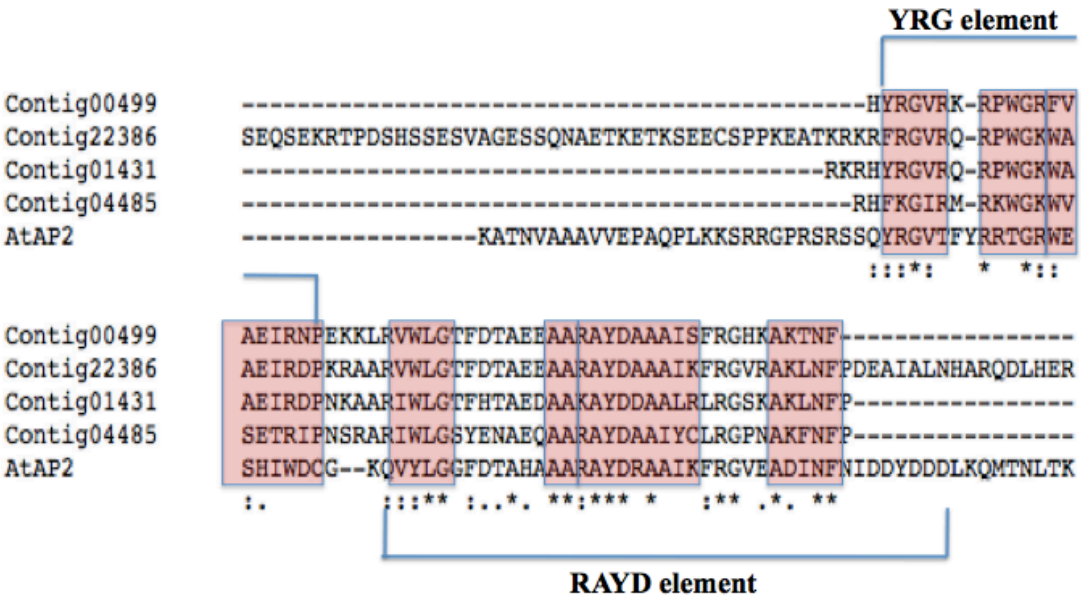


Fig 3.4B (Rabia et al. unpublished) Confirmation of MeJA-induced expression of 19 putative *Taxus* TF encoding genes by RT-PCR. Left panel: contig number of each gene; right panel: PCR confirms the induction of 19 putative TF genes by MeJA.

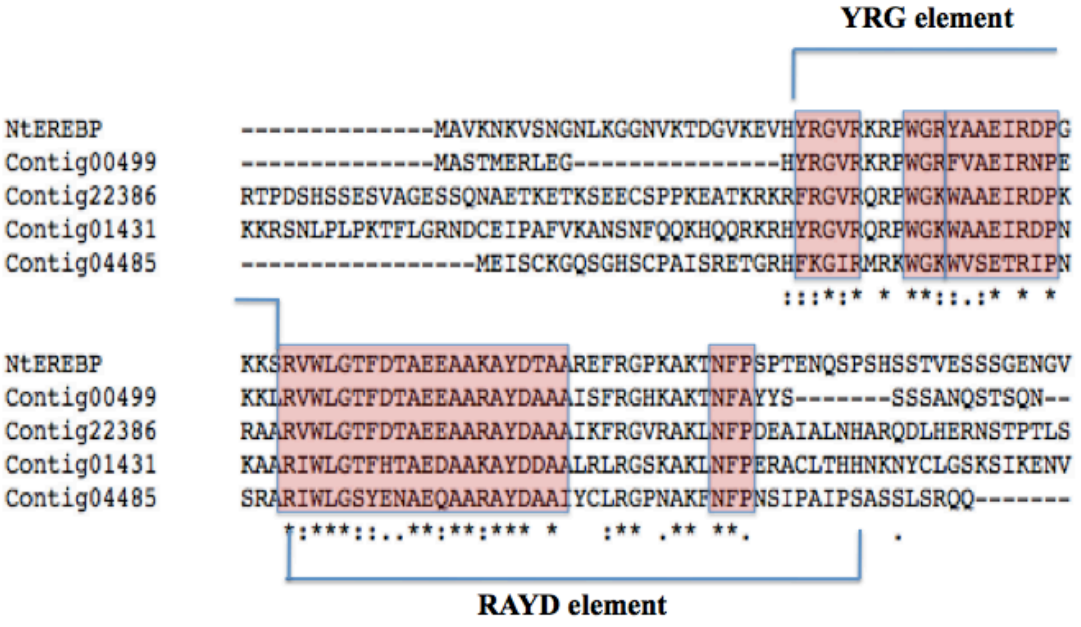
In order to further characterise all 19 TFs, the amino acid sequences of those TFs from 5 families are aligned with that of well-characterised TFs from corresponding TF families using the sequence aligner from Just-bio (http://www.justbio.com/index.php?page=aligner) (Fig 3.4C-G).

AP2 TFs are characterised by the AP2 domain present in their polypeptides, and both YGR and RAYD elements are located in this domain (Okamuro et al. 1997). The amino acid sequences of 6 MeJA-induced putative AP2 TFs are aligned against that of the AP2 domain of *At*AP2 and *Nt*EREBP (NCBI), respectively (Fig 3.4C-a to f). Among them, 4 putative AP2 TFs, which include Contig00499, 22386, 01431 and 04485, show high sequence similarity with both *At*AP2 and *Nt*EREBP in the conserved YGR and RAYD elements (Fig 3.4C-a to b). This could indicate that all 4 TFs are AP2-like and EREBP-like proteins. In contrast, Contig03304 only shows sequence similarity in the RAYD element when aligned with *At*AP2, but the YGR region is missing (Fig 3.4C-c). In addition, when aligned with *Nt*EREBP, the YGR element in Contig03304 is highly conserved, but the RAYD element is partially conserved (Fig 3.4C-d). This could indicate that Contig03304 is more closely related to *Nt*EREBP, and it is a more EREBP-like TF. Interestingly, when aligned with *At*AP2, Contig07245 shows partially conserved YRG element but highly conserved RAYD element (Fig 3.4C-e). In contrast, it shows the opposite when aligned with *Nt*EREBP (Fig 3.4C-f). Therefore, Contig07245 could be either an AP2-like or an EREBP-like protein.

a ***At*AP2 Vs Contig00499, 22386, 01431 and 04485**

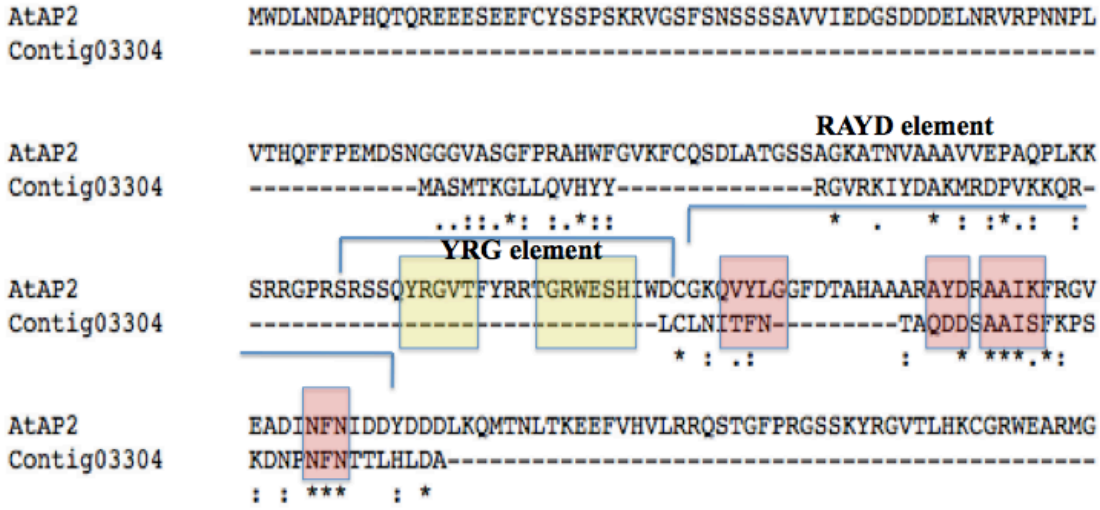


b ***Nt*EREBP Vs Contig00499, 22386, 01431 and 04485**



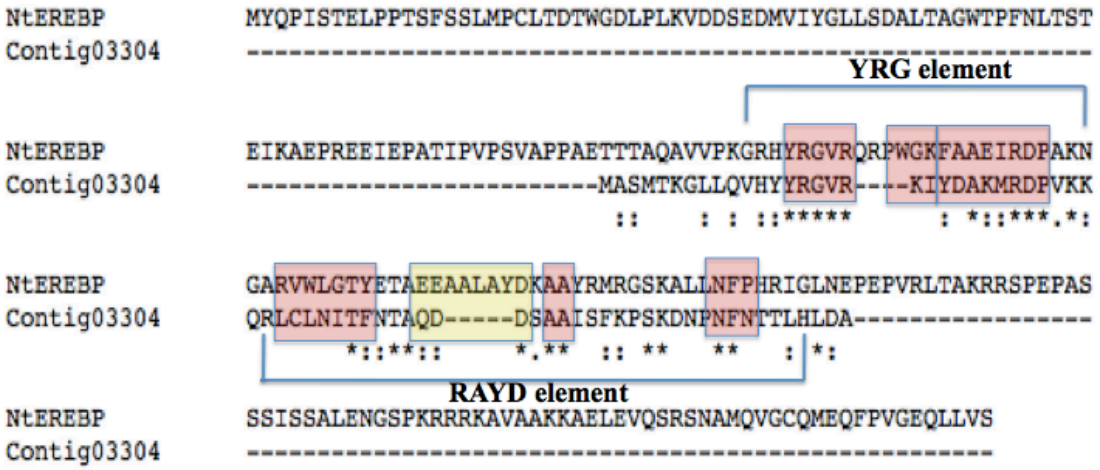
c

***At*AP2 Vs Contig03304**



d

***Nt*EREBP Vs Contig03304**



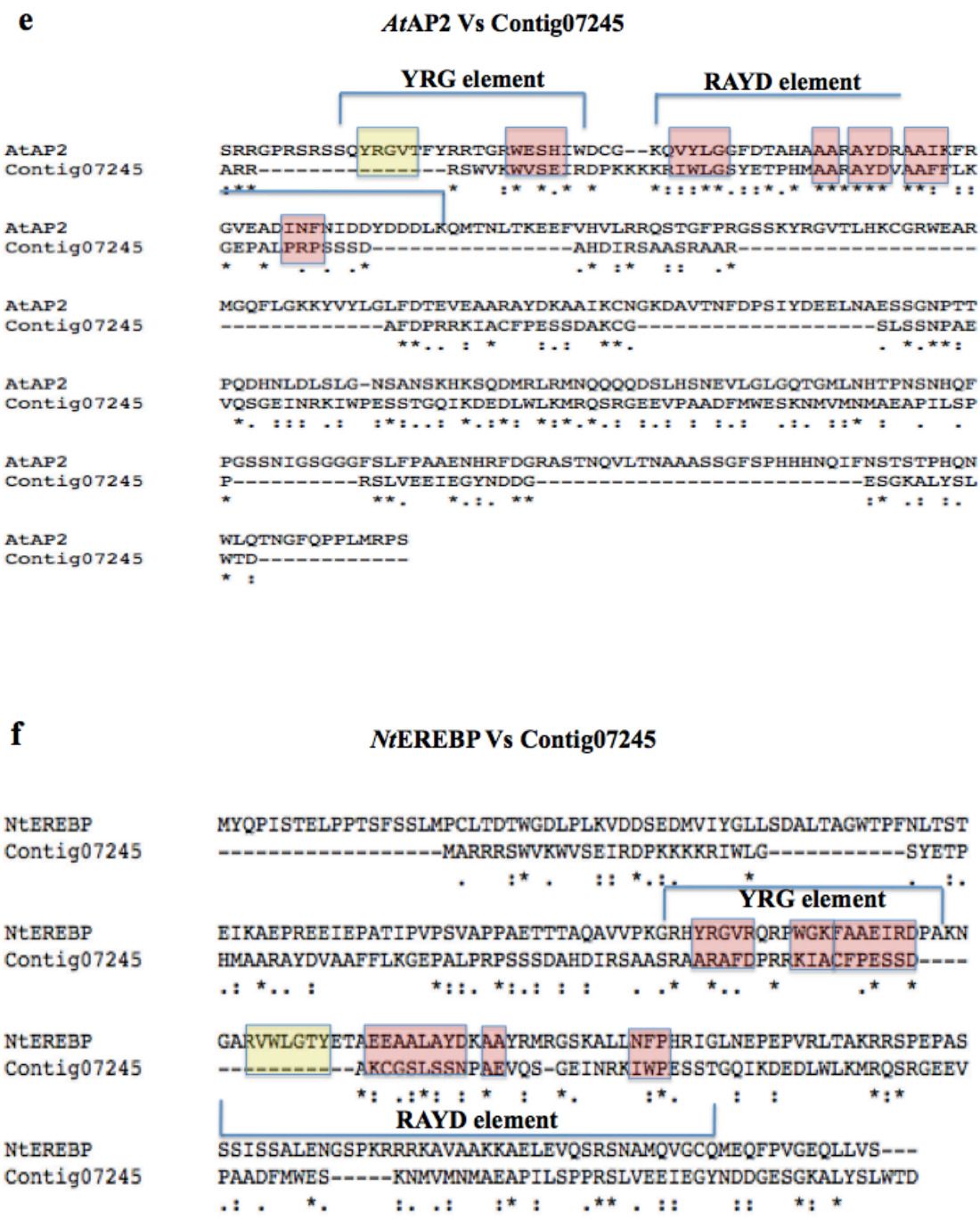


Fig 3.4C (a to f) Amino acid sequence alignment of the AP2 domain between *At*AP2/*Nt*EREBP and 6 putative MeJA-induced AP2 TFs. Brackets indicate the location of the two conserved elements in the AP2 domain: YRG and RAYD. Brackets indicate location of the conserved YRG and RAYD elements. Red shaded boxes highlight regions of sequence similarity. Yellow shaded boxes highlight the conserved regions that do not appear in putative AP2 TFs (Contig03304 and 07245). Asterisk indicates the conserved single amino acids.

NAC TFs are defined by the NAC DNA-binding domains located at the N-terminal of the peptides. When the amino acid sequences of *Os*NAC6, *At*ANAC019, *At*NAM and *At*NAC1 (NCBI) are aligned against 4 MeJA-induced putative NAC TFs, including Contig00172, 08447, 10855 and 05638, all these 4 TFs contain highly conserved regions found at the N-termini of NAC proteins (Fig 3.4D-a). Furthermore, these conserved regions were also found in the alignment between ANAC019, 055 and 072 (Tran et al. 2004) and in the NAC domain of NAC proteins from a wide range of plant species (Olsen et al. 2005). In contrast, the C-terminal regions are highly variable in sequence. They are characterised by the frequent appearance of simple amino acid repeats and contain regions that are rich in serine (S) and threonine (T), proline (P) and glutamine (Q), or acidic residues (SF 10 a & b). This is a common feature of the C-terminal region of NAC proteins (Duval et al. 2002; Hegedus et al. 2003; Kikuchi et al. 2000). All these features indicate that these four putative TFs are NAC-like proteins. By contrast, the N-terminal region of Contig09658 is partially conserved (Fig 3.4D-b). This might indicate that this protein has diverged away from the NAC TF FAMILY through evolution. Finally, all 5 putative TFs possess a putative nuclear localisation signal (NLS) sequences (Tran et al. 2004), which is a signal for nuclear import.

a

ANAC019, OsNAC6, AtNAM and AtNAC1 Vs Contig00172, 08447, 10855 and 05638

b

ANAC019, OsNAC6, AtNAM and AtNAC1 Vs Contig09658

Fig 3.4D (a to b) Amino acid sequence alignment of NAC domain between *Os*NAC6, *At*ANAC019, *At*NAM and *At*NAC1and 5 putative MeJA-induced NAC TFs. Red shaded boxes highlight regions of sequence similarity. Blue double-headed arrow below the sequence indicates putative nuclear localization signal (NLS). Asterisk indicates the conserved single amino acids.

Most plant *MYB* genes encode the R2R2-MYB type proteins. They typically contain a R2R3 domain, which has an amino acid sequence of -W-(X19)-W-(X19)-W-….F/I-(X19)-W-(X18)-W- (Dubos et al. 2010). The amino acid alignment of 4 putative MYB TFs with that of R2R3 domain of *Es*R2R3-MYB2 and *At*R2R3-MYB proteins (NCBI) shows that they all contain the conserved sequence of the R2R3 domain, which defines them as *Taxus cuspidata* R2R3-MYB TFs.

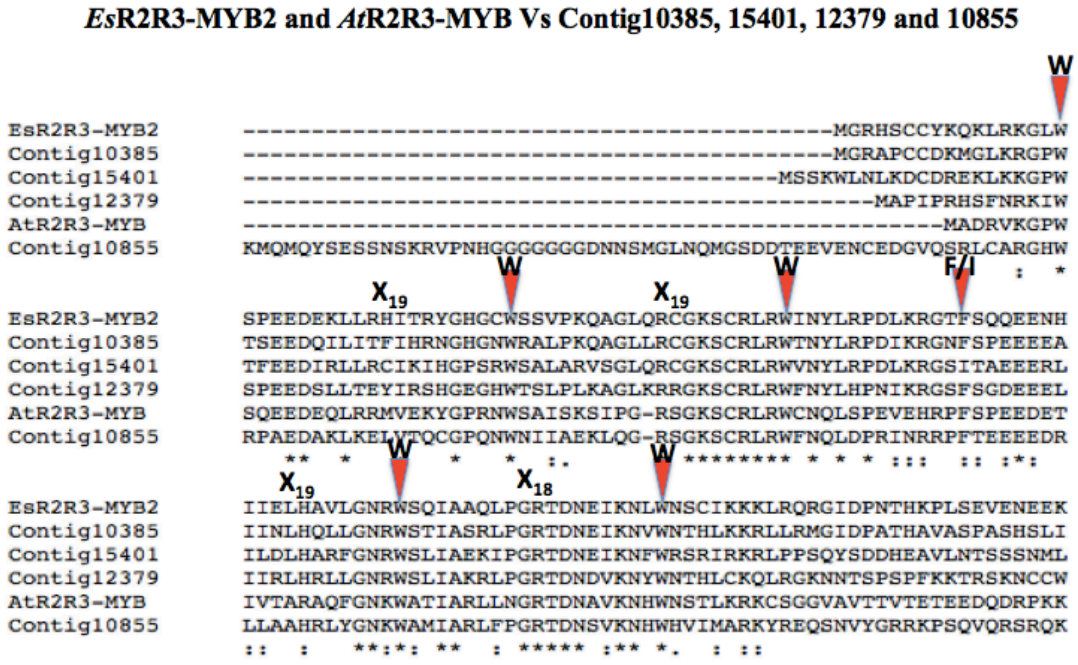


Fig 3.4E Amino acid sequence alignment of R2R3 domain of MYB TFs between *Es*R2R3-MYB2 and *At*R2R3-MYB and 4 putative MeJA-induced MYB TFs. Red arrows indicate the conserved amino acid sequence (-W-(X19)-W-(X19)-W-….F/I-(X19)-W-(X18)-W-) found in R2R3 domain of R2R3-MYB TFs. Asterisk indicates the conserved single amino acids.

WRKY TFs typically consist of WRKY domains (WRKYGQK) and zinc fingers (C2H2 or C2HC). In the amino acid alignment between *At*WRKY6, 18 and 60 (NCBI) and Contig09595 and 19284, both a WRKY domain and a zinc finger are found in the two MeJA-induced putative WRKY TFs (Fig 3.4F). This feature defines these two proteins as group II WRKY TFs (Eulgem et al. 2000). In addition, there is a putative NLS (KKRK) found in the upstream of the WRKY domain in Contig19284, but not in Contig09595. This may indicate that it is only Contig19284 that can be imported into the nucleus.

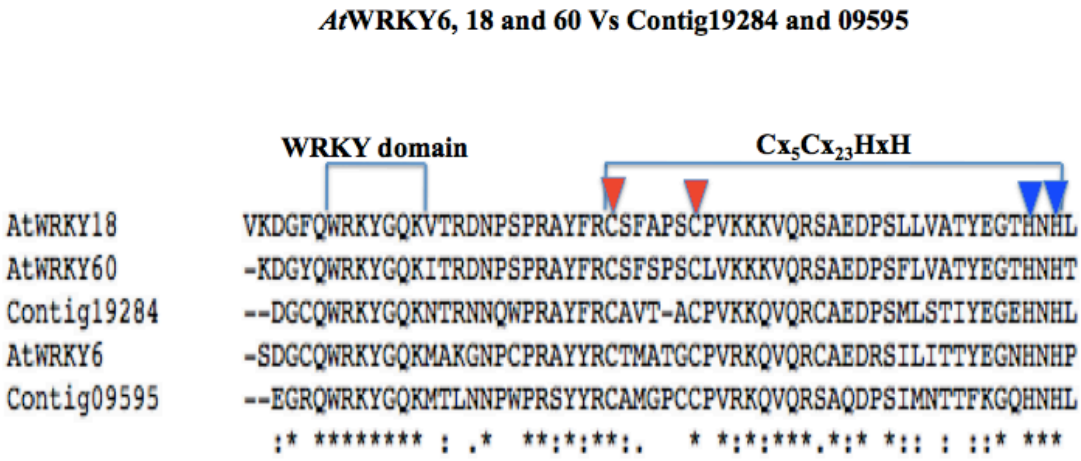


Fig 3.4F Amino acid sequence alignment of WRKY TFs between *At*WRKY18, 6 and 60 and 4 putative MeJA-induced WRKY TFs. Brackets indicate both WRKY and C2H2 zinc finger domains. Red and blue arrows indicate C and H residue in the C2H2 zinc finger, respectively. Asterisk indicates the conserved single amino acids.

bLHLH TFs are featured by a bHLH signature domain that locates at the C-terminal region of the peptide. It is about 60 residues long and divided into two distinct regions, including a basic at the N-termini and a HLH region at the C-termini. The amino acid sequence of the bHLH domain from two putative bHLH TFs is 58-amino acids long. After alignment with the bHLH domains of *At*MYC2 and 4 (NCBI), they show high sequence similarity in the two distinct regions in bHLH domains (Toledo et al. 2003) (Fig 3.4G). Furthermore, the highly conserved amino acids found in these two TFs are also found in the sequences of a large number of *At*bHLH TF members (Toledo et al. 2003).

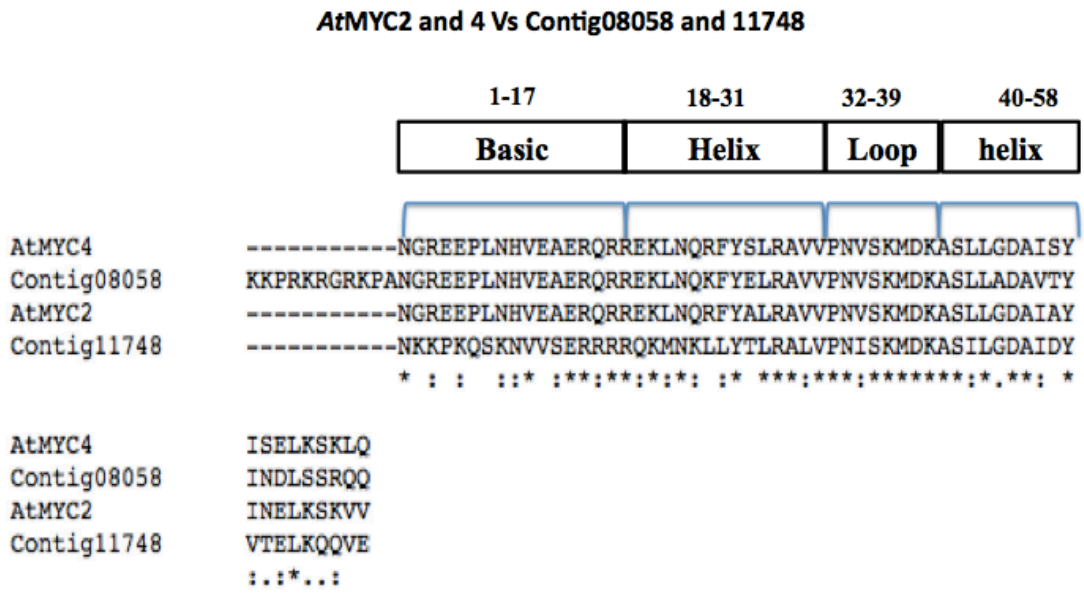


Fig 3.4G Amino acid sequence alignment of bHLH domain between *At*MYC2 and 4 and 2 putative MeJA- induced bHLH TFs. Brackets indicate the basic and HLH regions. Asterisk indicates the conserved single amino acids.

In conclusion, after the sequence alignment with well-known TFs of 5 different families, each of 19 MeJA-induced putative TFs has been assigned to a specific TF family, which is supported by the fact that the conserved signature domains are present in these 19 putative TFs.

3.4 MeJA up-regulated paclitaxel biosynthetic pathway genes

The increased abundances for transcripts of paclitaxel biosynthesis pathway genes were first revealed through random sequencing of a MeJA-induced *Taxus cuspidata* cDNA library (Jennewein et al. 2004b). Consistently, the paclitaxel biosynthesis pathway genes are also significantly up-regulated after MeJA treatment in at least at one of the three post-elicitation time points in our Solexa database. The seven pathway genes, which include three early (*GGPP*, *TASY* and *T5αH*) and four late (*DBAT*, *PAM*, *BAPT* and *DBTNBT*) genes, are all significantly induced at least at one of the three time points tested (Fig 3.5). Both RNA gel blot analysis and RT-PCR results showed that these 7 pathway genes were all up-regulated by MeJA elicitation within 6 h in the *Taxus cuspidata* cell line P991, although the transcript level of *BAPT* and *DBTNBT* were much lower than that of early pathway genes (Nims et al. 2006; Patil et al. 2012). Furthermore, the middle pathway genes involved in hydroxylation and acyl/benzoyltransfer are also significantly induced by MeJA, including *T10βH*, *T13αH*, *T2αH*, *T7βH*, *T14bH* and *TBT* (Fig 3.6). In addition, similar results were obtained in the transcriptional profile of *Taxus chinensis* cells in response to MeJA by RNA sequencing and the randomly sequenced cDNA library derived from MeJA induced *Taxus cuspidata* cells (Jennewein et al. 2004b; Li et al. 2012). Finally, plant multidrug and toxic compound extrusion (MATE) transporters encoding genes have been reported to be involved in the transport of secondary metabolites. For example, a MATE transporter, which is localized in the vacuolar membrane, has been reported to mediate nicotine transportation in *Nicotiana tabacum* (Morita et al. 2009). Another MATE transporter, TT12 from *Arabidopsis*, and its ortholog MATE1 from *Medicago truncatula* have also been found to facilitate the uptake of epicatechin 3'-O-glucoside into the vacuole, which is the precursor for proanthocyanidin biosynthesis (Zhao and Dixon 2009; 2010). However, there are no MATE encoding genes found in our 1646 annotated contigs. This may indicate that MATE is not involved in paclitaxel transportation into vacuoles. Instead, two putative ATP binding cassettes (ABC) transporter encoding genes are found to be induced by MeJA at all three post-elicitation time points (Fig 3.7). This ABC transporter family has been shown to transport a wide range of chemical compounds,

including metal ions, auxins and defensive secondary metabolites (Rea 2007). For example, a multidrug resistance-associated protein (MRP)-type ABC transporter, ZmMrp3, has been reported to be involved in anthocyanin transport in *Zea mays*. In addition, the *ZMMrp3* is co-expressed with other anthocyanin structural genes (Goodman et al. 2004). Therefore, these putative ABC transporters could be involved in mediating the transport of paclitaxel and its precursors, as their encoding genes are induced together with the paclitaxel biosynthesis structural genes. The Walker group has shown enrichment of not only several known paclitaxel biosynthetic pathway genes but also novel transcripts that may be involved in taxane modification (acylation and hydroxylation), transport and degradation in the MeJA up-regulated EST libraries (Lenka et al. 2012).

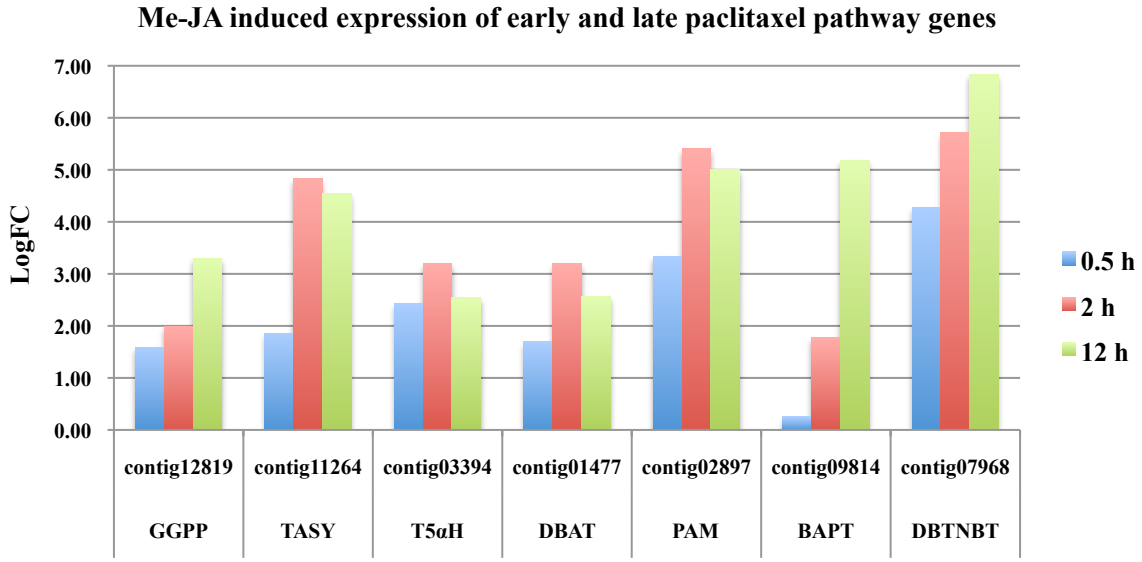


Fig 3.5 Fold change in the expression of the early and late paclitaxel biosynthesis pathway genes after MeJA treatment at three different post-elicitation time point (0.5, 2 and 12 hour) relative to 0 hour (Fig 1.6). Log FC is the fold change on the log2-scale between the normalised and stabilised transcript counts. *GGPPS*: geranylgeranyl diphosphate synthase; *TASY*: taxadiene synthase; *T5αH*: taxadiene-5α-hydroxylase; *DBAT*: 10-deacetylbaccatin 10β-*O*-acetyltransferase; *PAM*: phenylalanine aminomutase; *BAPT*: taxoid C13-O-phenylpropanoytransferase; *DBTNBT*: 3'-*N*-debenzoyl-2'-deoxytaxol *N*-benzoyltransferase.

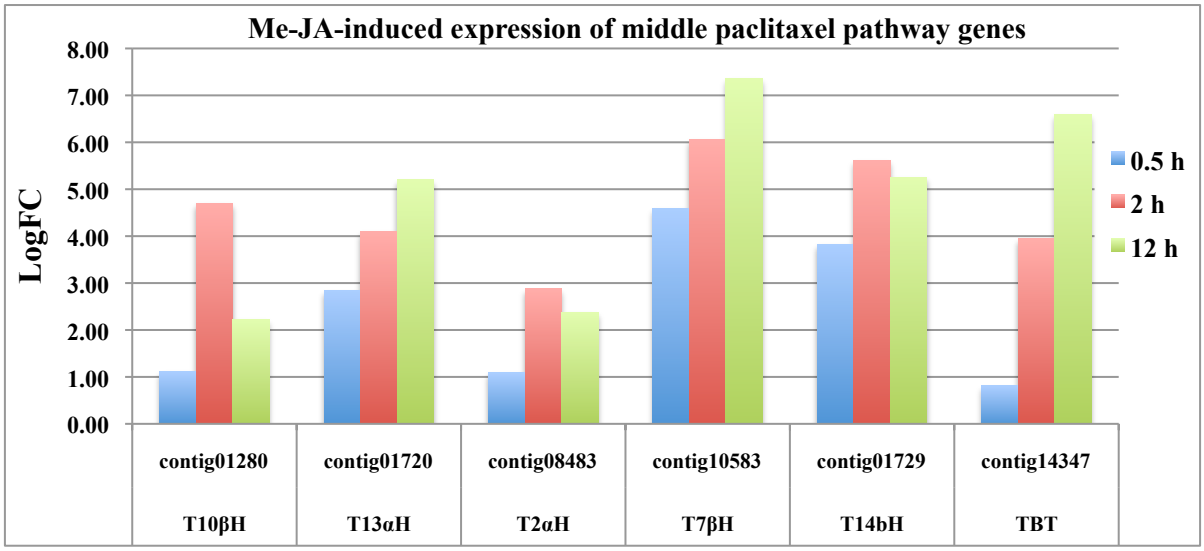


Fig 3.6 Fold change in the expression of the middle paclitaxel pathway genes after MeJA treatment at three different post-elicitation time point (0.5, 2 and 12 hour) relative to 0 hour. Log FC is the fold change on the log2-scale between the normalised and stabilised transcript counts. *T10βH*: taxane 10β-hydroxylase; *T13αH*: taxane 13α-hydroxylase; *T2αH*: taxoid 2α-hydroxylase; *T7βH*: taxoid 7β-hydroxylase; *T14bH*: taxane 14b-hydroxylase; *TBT*: 2α-hydroxytaxane 2-O-benzoyltransferase.

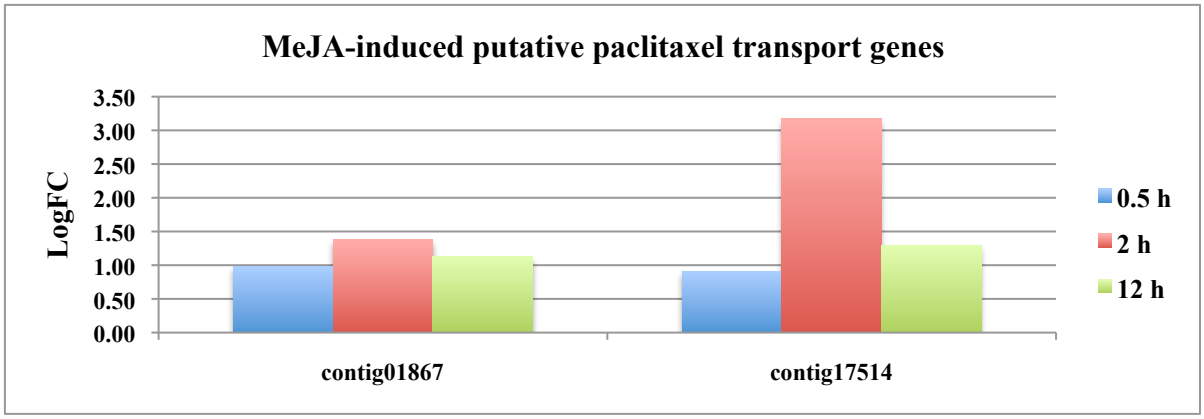


Fig 3.7 Fold change in the expression of two putative ABC transporter encoding genes after MeJA treatment at three different post-elicitation time point (0.5, 2 and 12 hour) relative to 0 hour. Log FC is the fold change on the log2-scale between the normalised and stabilised transcript counts.

3.5 19 putative TF-associated DNA-binding elements in paclitaxel biosynthesis gene promoters

In order to regulate gene transcription, TFs have to interact with DNA by binding to the promoter of a given gene via specific binding motifs, which typically contain a core sequence of 4-6 nucleotides and are present within the promoter. Each TF family has family-specific binding motifs. For example, AP2 and WRKY TFs specifically bind to GCCGCC and TTGACC, respectively (Ciolkowski et al. 2008; Sakuma et al. 2002). The putative TF binding motifs, which are present within the promoters of five paclitaxel biosynthesis genes, are summarised in Table 9. The *PAM* promoter only contains 3 E-boxes. W-box, E-box and MYB binding sites (MYBBS) appear frequently in the promoter of *TASY*, *T5αH*, *DBAT* and *DBTNBT*. In contrast, the T/G-box appears only once in the four promoters except *PAM* and GCCGCC is only present in *TASY, DBAT* and *T5αH*. The core sequence of the NAC binding site (NACBS) is part of the E-box motif if the two middle nucleotides are CA (Appendix I). There are reasons that these five genes (*TASY, T5αH*, *DBAT*, *PAM* and *DBTNBT)* were chosen in this project. They are involved in the early and late steps of the paclitaxel biosynthesis pathway (Fig 1.6) (Guerra et al. 2012). In particular, *TASY*-encoding enzyme catalyses the committed step of paclitaxel biosynthesis (Wildung and Croteau 1996). In addition, their expression has been experimentally shown to be up-regulated by MeJA (Nims et al. 2006; Patil et al. 2012). Finally, since *Taxus* spp. is a non-model plant, and therefore its genome is not completely sequenced, the promoters of these five genes are the only ones available while doing my experiments.

Table 9 Summary of JA-responsive putative TF associated DNA binding motifs present in five paclitaxel biosynthetic gene promoters by searching promoter sequences in the database of Plant *Cis*-acting Regulatory DNA Elements (PLACE) (http://www.dna.affrc.go.jp/PLACE/).

Motif Name		T/G-box	W-box	MYBBS	NACBS	GCCGCC	E-box
Core sequence		AAACGTG	TGAC	Variable	CACG	GCCGCC	CANN TG
Putative binding TFs		bZIP	WRKY	MYB	NAC	AP2/ERF	bHLH/ MYC
Promoters	**TASY (1408 bp)**	1	12	6	3	1	7
	T5αH (1611 bp)	1	10	13	4	1	9
	PAM (257 bp)	0	0	0	1	0	3
	DBAT (1433 bp)	1	10	4	1	1	8
	DBTNBT (1898 bp)	1	3	9	3	0	10

The core sequences determine the DNA-binding selectivity of TFs from different families. In other words, the TF family to which the corresponding DNA-binding proteins for a given motif belong can be predicted. Furthermore, the surrounding sequences outside the core motif appear to be important, because they partially determine the binding site preferences of the type of TFs from the same family. For example, *At*WRKY6 and *At*WRKY11 have a binding preference to W-boxes that contain a G residue directly 5' adjacent to W-box (5'AGTTGACCAA). In contrast, *At*WRKY26, 38 and 43 prefer to bind to the same sequence that has a T, C or A in the place of this G residue (5'A [T/C/A] TTGACCAA) (Ciolkowski et al. 2008). Furthermore, the surrounding sequences also have been reported to be functionally important for gene transcription regulation. It has been shown that WRKY53 can bind to different promoters that contain W-boxes. However, it can act as either a transcriptional activator or repressor depending on the surrounding sequences of W-boxes (Miao et al. 2004). Since the surrounding sequences required for efficient WRKY binding are shown to include 4 bp proximal to and 3 bp distal of W-boxes (Ciolkowski et al. 2008), there is assumption that this rule is also applied to other TF binding motifs in our database. After aligning all TF binding motif sequences listed

in Appendix I, the core sequences of W-boxes (WRKY), E-boxes (bHLH/MYC), NACBS (NAC) and GGCGCC (AP2/ERF) are highly conserved throughout all five promoters. However, the sequences that surround each core are different both within a promoter and between promoters. This could indicate that the DNA-binding selectivity of TFs from the same family depends on the surrounding sequences of their conserved core binding sequence. In fact, there is more than one MeJA-responsive TF found in each TF family. For example, the AP2/ERF, NAC, bHLH, and WRKY family have 15, 6, 8 and 4 members, respectively (Fig 3.3b). This might indicate that each member of a TF family has their specific binding motif that differs in surrounding sequences with each other. Furthermore, the MYBBS are found to be variable in both the core and surrounding sequences, which are represented by TAAC [T/G] G, GGATA [T/A/G/C], AAACCA, [T/C] [A/G/T/C] GT [T/A] [A/T/GG/TC/GT] or C [A/T] A [C/T] C [G/C] A (Appendix I). MYB TF family has been known to contain a large and functionally diverse number of TF proteins in all eukaryotes. There are 9 MeJA-responsive MYB proteins (Fig 3.3b). They might have their own specific DNA-binding motifs that differ with each other in surrounding sequence only or both core and surrounding sequence.

3.6 Discussion

In conclusion, after bioinformatic analysis of the transcriptome of MeJA-elicited CMCs, the expression of 1646 genes is significantly induced by MeJA in at least one of the three post-elicitation time points. Based on the conserved TF signature domains, 78 genes are annotated as putative TFs. Then, 19 putative TF-encoding genes were chosen for further assays, because they are all induced at the early time point (0.5h) (Fig 3.4 A) and show high E-values ($<e^{-04}$) when blasted against the database of non-redundant protein sequences of vascular plants (Table 3.1). RT-PCR further confirmed their induction by MeJA at all three time points (Fig 3.4 B). Finally, the amino acid sequence alignment results show all 19 TFs contain the signature domains when aligned with their well-characterised TF counterparts (Fig 3.4 C-G). Furthermore, a number of known paclitaxel biosynthetic pathway genes, including the early, middle and late pathway genes, are also induced by MeJA at the early time points (Fig 3.5 & 6). The induction of the early and late pathway genes by MeJA was confirmed by RT-PCR in the literature, including *GGPP*, *TASY*, *T5αH, DBAT*, *PAM*, *BAPT* and *DBTNBT* (Nims et al. 2006; Patil et al. 2012). Finally, the TF associated DNA-binding motifs are found in the promoters of two early and three late pathway genes (Table 8 & Appendix I). Collectively, this leads to the prediction that these TFs could regulate the expression of these five genes as well as other known and unknown paclitaxel biosynthesis pathway genes. Biological assays are required for the identification of the positive regulators.

Chapter 4 Gateway compatible Yeast One-Hybrid Assay

4.1 Experimental design

The Yeast One-Hybrid (Y1H) assay has been widely used as a straightforward technique for detecting interactions between DNA fragments and proteins. It has been shown to be a rapid and efficient method to screen for transcription factor-DNA interactions (Ouwerkerk and Meijer 2011). Therefore, this method was employed to detect interactions between 19 putative TFs and five paclitaxel biosynthesis gene promoters. Fig 4.1 illustrates the experimental design of Y1H used in my project. There were two constructs introduced into the yeast nucleus by transformation. One is reporter vector ($P_{promoter}$: *His*), which is integrated into the yeast genome (Fig 2.2a); the other is effector vector (PADH1: *GAL4AD-TF*), which is used to express the fusion proteins of *GAL4* activation domain and putative TFs (Fig 2.2c). The interactions between DNA fragments (promoters) and putative TFs are detected as a result of *His* expression, which enables the yeast to grow on SD/His- medium. In addition to the effector construct, the cloning of TFs into this construct is facilitated by gateway cloning technology. The TFs are cloned downstream of and in-frame with the *GAL4 AD*.

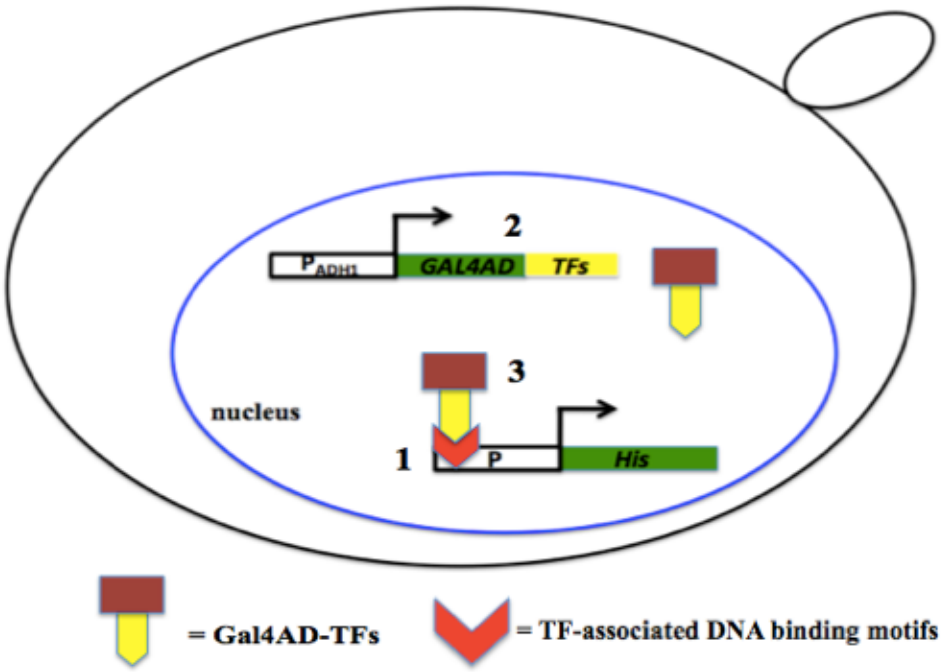


Fig 4.1 Experimental design of Y1H. A linearized $P_{promoter}$: *His* reporter construct is integrated into the yeast genome to make a *His* reporter strain (1). An effector vector (PADH1: *GAL4AD-TF*) is introduced into the reporter strain (2). If TF proteins bind to the bait (promoter), *His* gene expression is activated by GAL4 activation domain.

4.2 Y1H positive control

In order to test if Y1H is a suitable assay to detect protein-DNA interactions, a positive control was established for Y1H assay before screening interactions between 19 putative TFs and paclitaxel biosynthesis promoters. A known interaction between *At*STZ/ZAT10 (a C2H2 zinc finger TF of *Arabidopsis*) and plox3 (the promoter of *Lipoxygenase 3* in JA biosynthesis) was used as the positive control in Y1H assay (Pauwels et al. 2008).

4.2.1 Construction of plox3: *His* reporter and pADH1: *GAL4-STZ/ZAT10* effector

The construction of Plox3: *His* reporter vector began with PCR amplification of *lox3* promoter, followed by cloning into pHISi at MCS (Fig 4.2). The size of PCR-amplified fragment shown on gel agreed with the 1997 bp-long plox3. The enzymatic digestions further confirmed the insertion of plox3 into pHISi. As expected, both *Xba* I and *Xho* I gave rise to linearized plox3: *His*. By contrast, *Hind* III digestion resulted in three fragments: 5.6 kb, 3.0 kb and 187 bp. As the smallest *Hind* III fragment is much smaller than the lowest marker (500 bp), which run off the gel. Furthermore, the amplification of attB1- and attB2-sited STZ/ZAT10 was carried out by a nested PCR (Fig 4.3). Two sets of primers were used to successfully amplify the attB1-*STZ/ZAT10*-attB2 fragment (~0.75 kb). Finally, the *STZ/ZAT10* was cloned into pGADgate424 through gateway cloning. The insertion of it into both pDONR 221 and pGADgate424 (pDES) was detected by PCR (Fig 4.4). As expected, the amplified PCR fragments (~0.75 kb) were in between 0.5 and 1.0 kb.

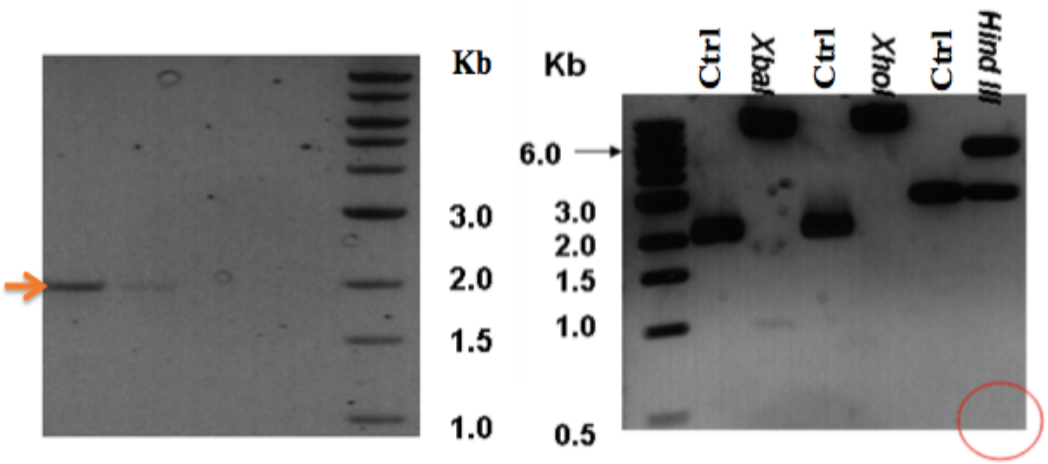


Fig 4.2 PCR amplification and cloning of Plox3 into pHISi. Left panel, PCR amplification of ~2.0 kb *lox3* promoter (indicated by red arrow); right panel, confirmation of cloning of plox3 into pHISi by enzymatic digestion. The red circle indicates the 187 bp *Hind* III fragment that run off the gel.

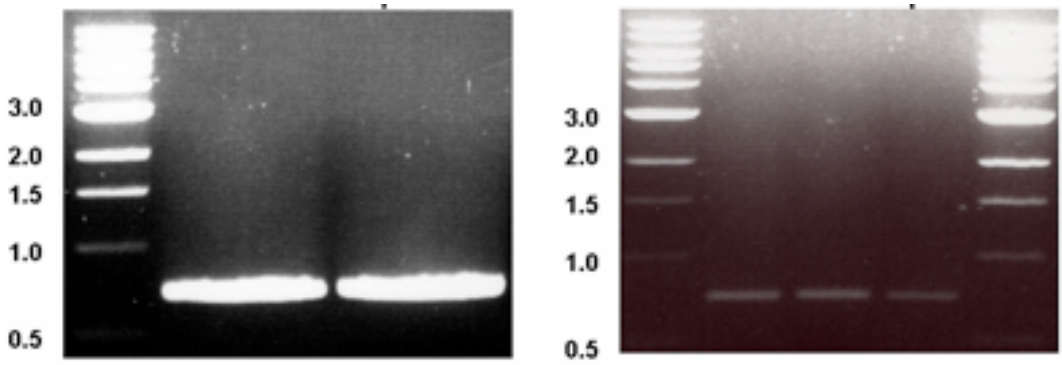


Fig 4.3 Nested PCR amplification of attB1-STZ/ZAT10-attB2 fragment. Left panel, first round PCR fragments; right panel, second round PCR fragments.

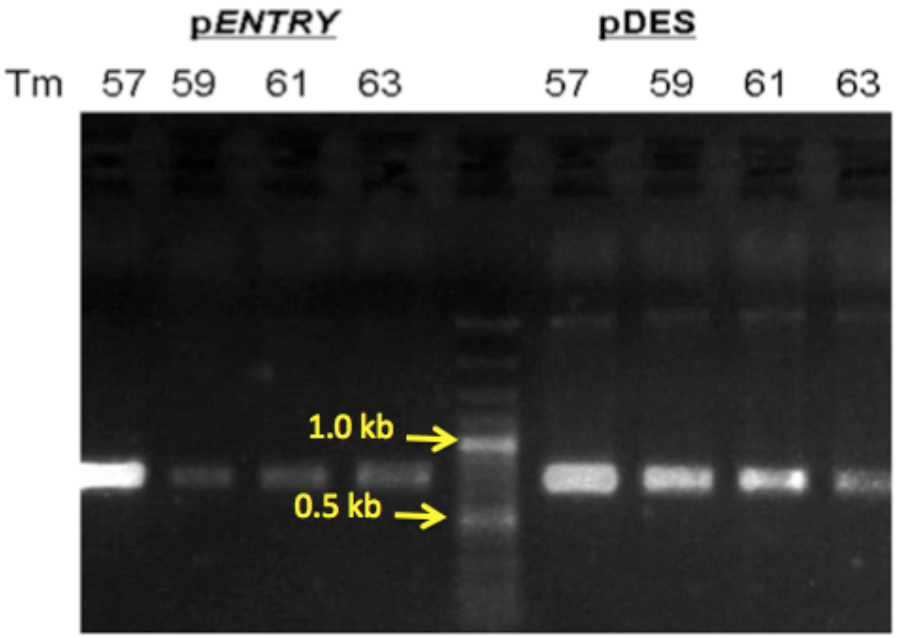


Fig 4.4 Gateway cloning of *STZ/ZAT10* into pGADgate424 (pDES). Temperature gradient PCR confirmed the insertion of it into both pDONR 221 (pENTY) and pGADgate424 (pDES).

4.2.2 Yeast genomic integration of plox3: *His* reporter

The YM 4271 yeast strain cannot grow on SD/URA-, LEU- or HIS- plates, as the genotype of this strain is *Ura*-52, *His*3-200, *and Leu2*-3. Before transformation of plox3: *His* reporter, the strain can only grow on YPDA plates (Fig 4.5a), but not on SD/URA- plates (Fig 4.5b). After transformation, the strain transformed with plox3: *His* reporter can grow on SD/URA- plates (Fig 4.5c). However, it still cannot grow on SD/HIS-plates (Fig 4.5d). This indicated that plox3: *His* reporter was integrated into yeast genome, which was further confirmed by PCR amplification of plox3 from the genomic DNA extracted from these transformants. Plox3 insertion was detected in them (Fig 4.6).

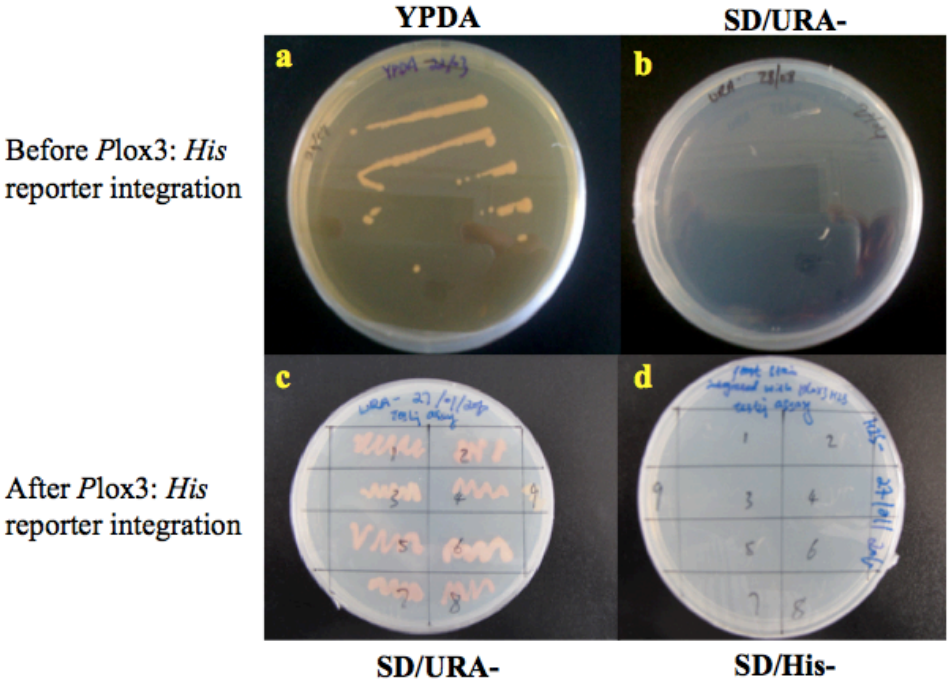


Fig 4.5 (a-d) Genomic integration of *P*lox3: *His* reporter into YM4271. Before integration, yeast strain can grow on YPDA (a), but not on SD/URA- (b); after integration, yeast strain can grow on SD/URA- (c), but not on SD/His- (d).

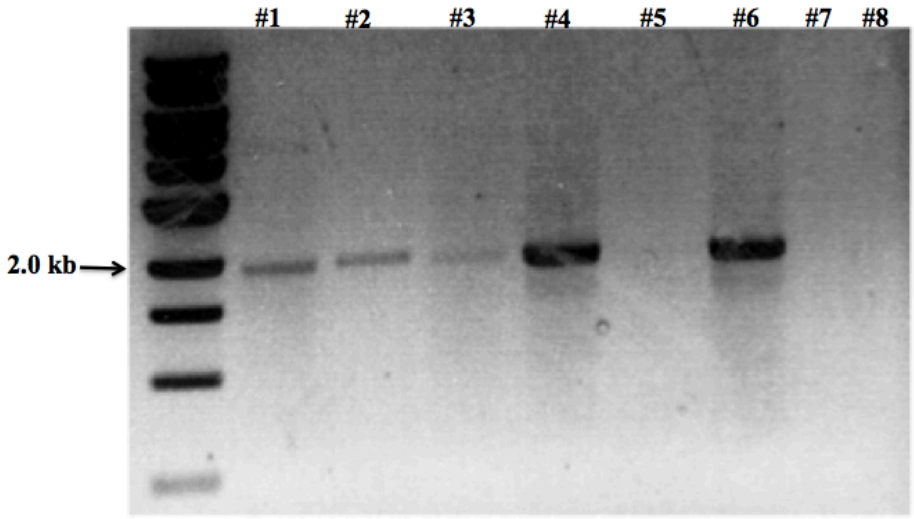


Fig 4.6 PCR confirmation of genomic integration of *P*lox3: *His* reporter into YM4271 using extracted genomic DNA from 8 transformants that were selected from the SD/URA- plate in Fig 4.5c.

4.2.3 Yeast spotting assays

Because of the minimal promoter of *HIS3* (*P*minHIS3) present downstream of plox3, which is used for transcriptional initiation, there is a basal *His* expression after genomic integration of plox3. 3-Amino-1,2,4-triazole (3-AT), a competitive inhibitor of *HIS3* gene products, was used to eliminate the background *His* expression on SD/His- plates before the transformation of the effector vectors. Five plox3: *His* yeast colonies (colony 1, 2, 4, 6 and 8), which were selected from the SD/URA- plate in Fig 3.14c, were subject to spotting assays. As a result, colony 1, 4 and 8 showed reduced growth as [3-AT] increased from 0 to 30mM. No growth was found when it reached both 45 and 60mM (Fig 4.7). So, all these three colonies showed low background of *His* expression upon inhibition by 3-AT, and therefore, they were used for transformation with effector vectors.

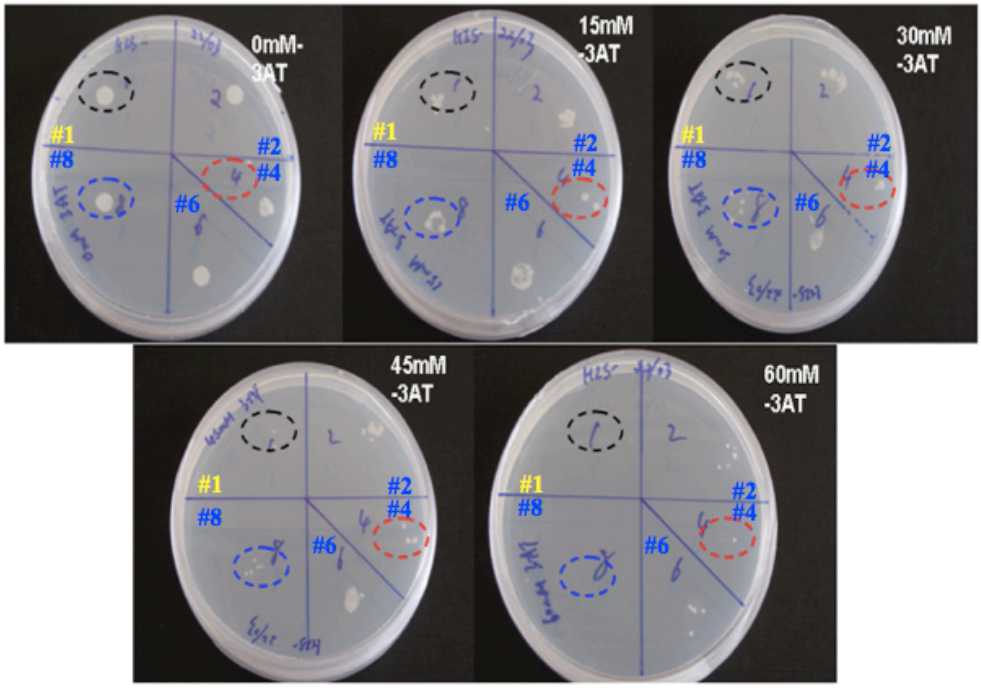


Fig 4.7 Yeast spotting assay of the positive control on SD/His- plates. Colony 1, 4 and 8 showed low background of *His* expression as [3-AT] increased from 0 to 60mM. 3-AT: 3-amino-1,2,4-triazole.

After the selection of *P*lox3: *His* reporter transformants that had low background *His* expression, the reporter transformant 1 was transformed with the effector construct (pADH1: *GAL4AD-STZ/ZAT10*) and selected on SD/Leu- plates. 4 colonies from SD/Leu- plates were further spotted on SD/His-/60mM [3-AT] plates. As expected, the negative controls, which include *His*-, YM4271 and plox3: *His* strains, did not show any growth. By contrast, like the positive control (*His*$^{+}$ strain), all 4 transformants showed strong growth at three series of dilutions. This indicates that there is an interaction between plox3 and STZ/ZAT10 TF (Fig 4.8).

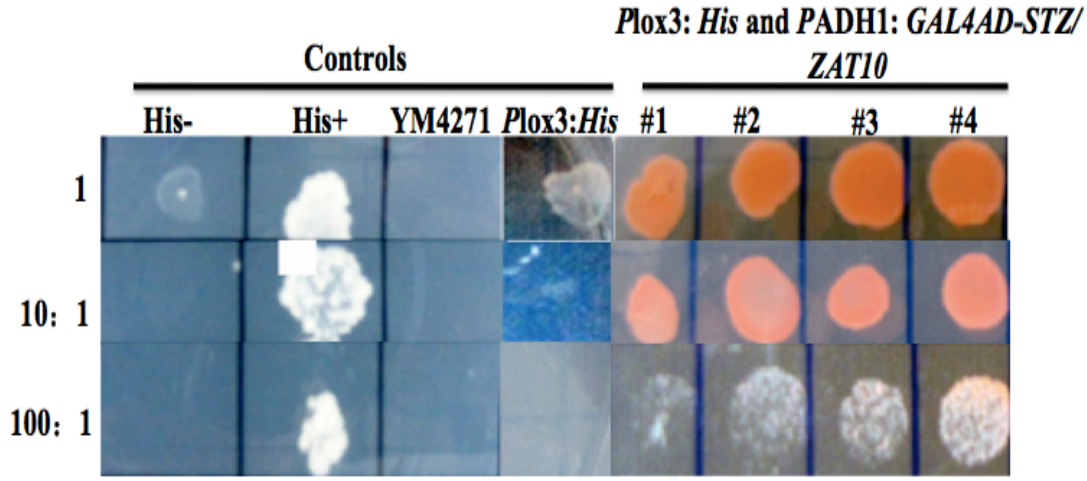


Fig 4.8 Spotting assay of yeast strains containing both plox3: *His* and pADH1: *GAL4AD-STZ/ZAT10* on SD/His-/60mM [3-AT] plates. As expected, like His+, yeast transformants with both constructs grew on SD/His-/60mM [3-AT] plates. In contrast, the controls, including His-, YM4271 and YM4271 transformants containing only plox3: *His,* did not grow. 10: 1 is 1 in 10 dilution of yeast strain suspension; 100: 1 is 1 in 100 dilution of yeast strain suspension.

4.3 Screening for interactions between 19 putative TFs and 5 paclitaxel biosynthesis gene promoters

Following the establishment of the positive control of Y1H, all 5 promoters were amplified from the genomic DNA of CMCs and cloned into pHISi reporter vectors to make *P*promoter: *His* reporter constructs (Fig 4.9). There were two DNA fragments amplified for the *DBAT* promoter, because the forward primer sequence was also found in the middle of the promoter sequence (Fig 4.9 E). Fig 4.10 shows the confirmation of promoter insertions into pHISi through enzymatic digestion of either *Hind* III or *Xho* I, or both.

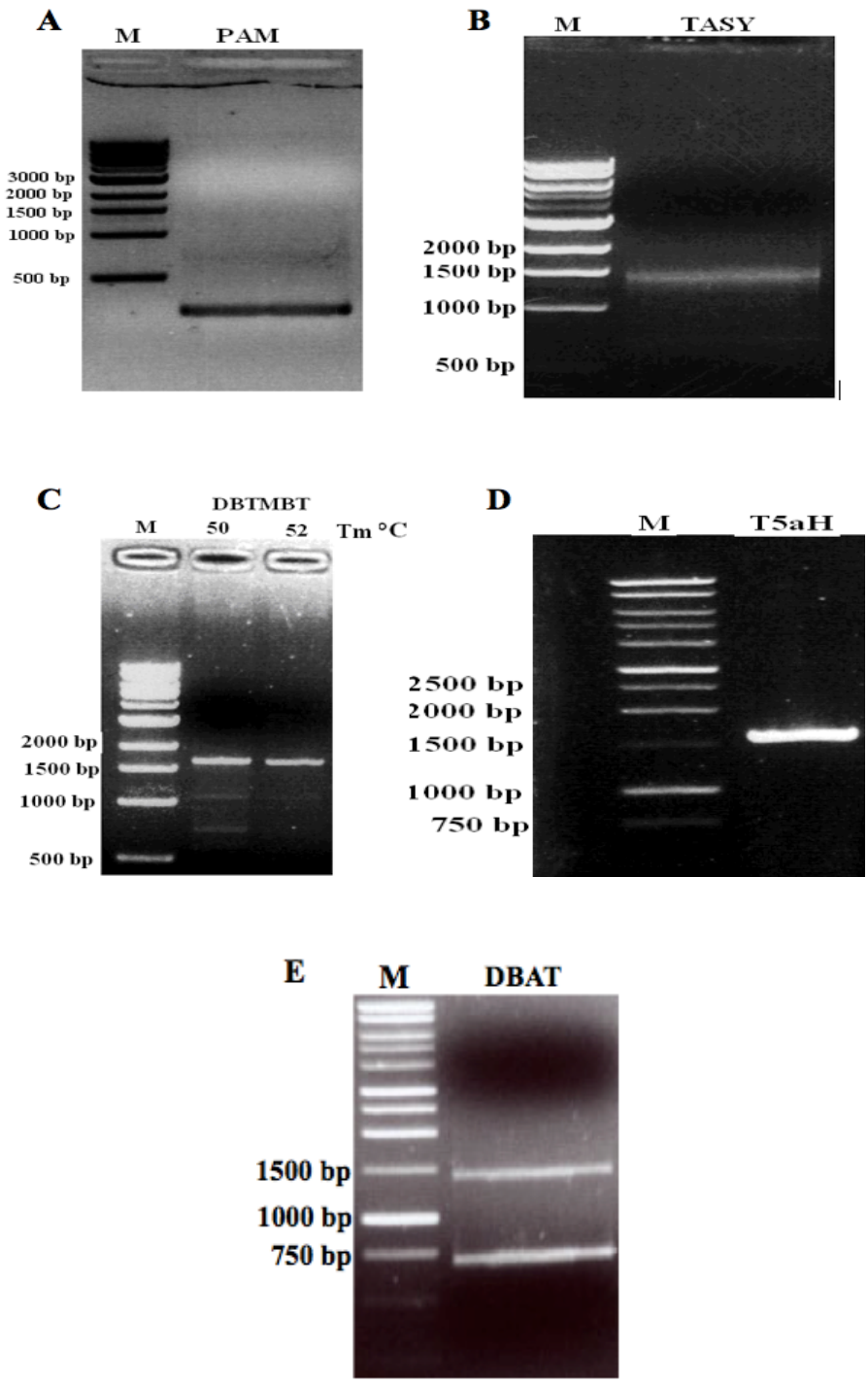


Fig 4.9 (A-E) PCR amplification of 5 paclitaxel biosynthetic gene promoters. (A): *PAM* (257 bp), (B): *TASY* (1433 bp), (C): *DBTNBT* (1665 bp), (D): *T5αH* (1611 bp), (E): *DBAT* (1408 bp). M: 1kb ladder.

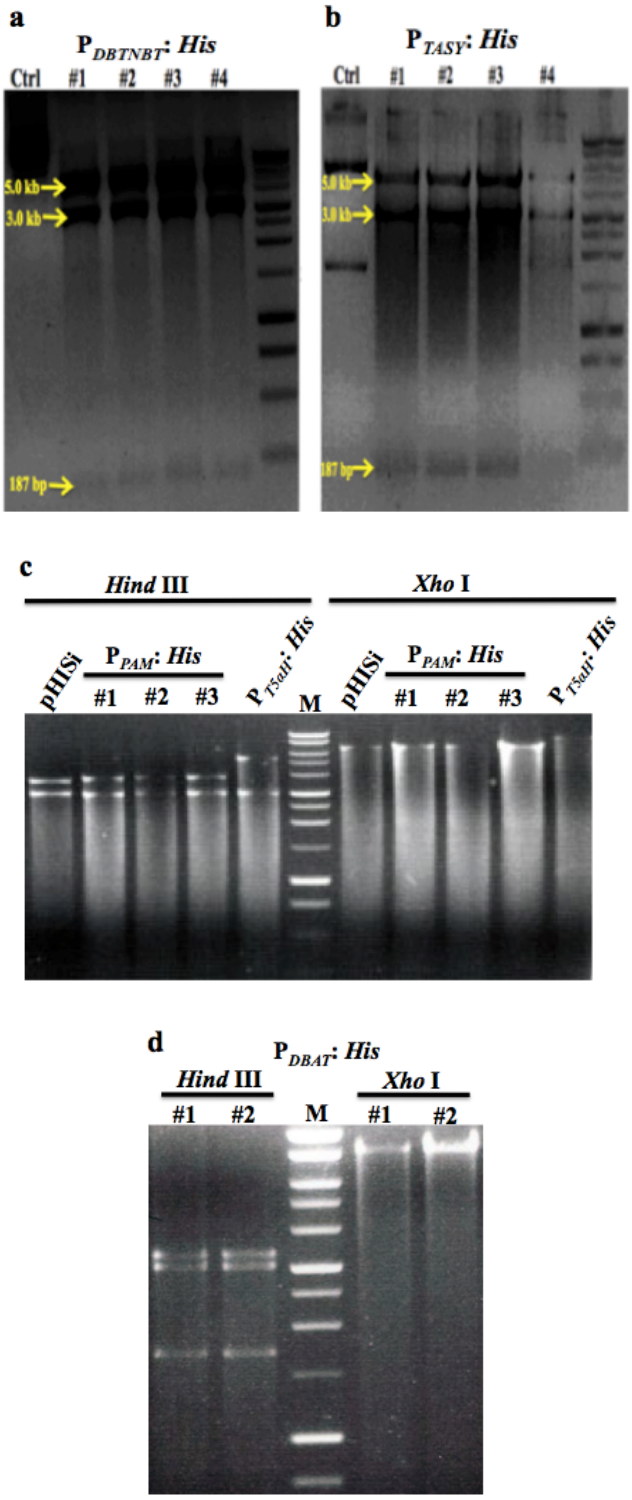
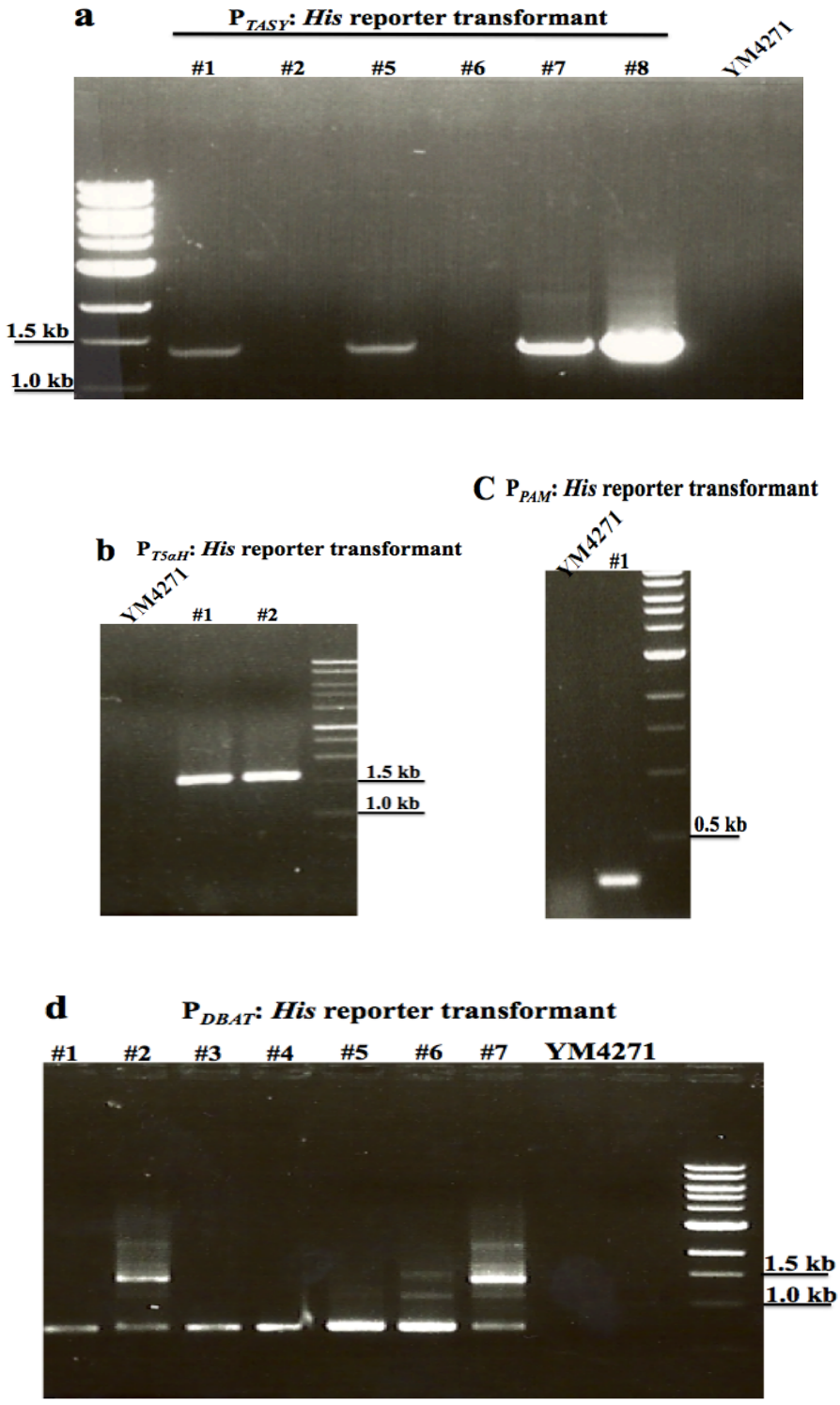


Fig 4.10 Confirmation of the insertion of paclitaxel biosynthesis gene promoters into pHISi. (a): *Hind* III-P_{DBTNBT}: *His* reporter, (b): *Hind* III-P_{TASY}: *His* reporter, (c): *Hind* III & *Xho* I-P_{PAM}: *His* reporter and *Hind* III & *Xho* I-$P_{\mathrm{T5\alpha H}}$: *His* reporter, (d): *Hind* III & *Xho* I-P_{DBAT}: *His* reporter.

Five different *P*promoter: *His* reporter yeast transformants were obtained and selected on SD/URA- plates after integrating each of five reporter vectors into YM4271 genome. The promoter integration was confirmed by PCR using genomic DNA extracted from the five reporter transformants (Fig 4.11). The PCR-confirmed positive transformants of each *P*promoter: *His* reporter were spotted on SD/His-/various [3-AT] plates to select for transformants with lowest *His* expression. As a result, the growth of the five reporter transformants was suppressed at 0.5mM [3-TA]. Furthermore, those five promoter-reporter transformants were each transformed with *P*ADH1: *GAL4-TFs* effectors and selected on SD/Leu- plates. PCR further confirmed the existence of TFs in each reporter strain (Fig 4.12). Then, the positive strains were spotted on SD/His-/0.5, 1.0, 3.0, 5.0, 15, 30 to 45mM [3-AT] plates. However, there was no convincing yeast growth on any of the plates with an increasing 3-AT concentration.


a
P$_{TASY}$: *His* reporter transformant
#1
#2
#5
#6
#7
#8
YM4271
1.5 kb
1.0 kb
b
P$_{T5αH}$: *His* reporter transformant
YM4271
#1
#2
1.5 kb
1.0 kb
C
P$_{PAM}$: *His* reporter transformant
YM4271
#1
0.5 kb
d
P$_{DBAT}$: *His* reporter transformant
#1
#2
#3
#4
#5
#6
#7
YM4271
1.5 kb
1.0 kb

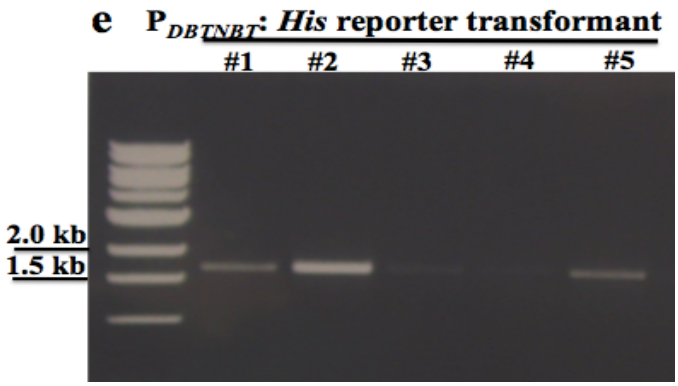
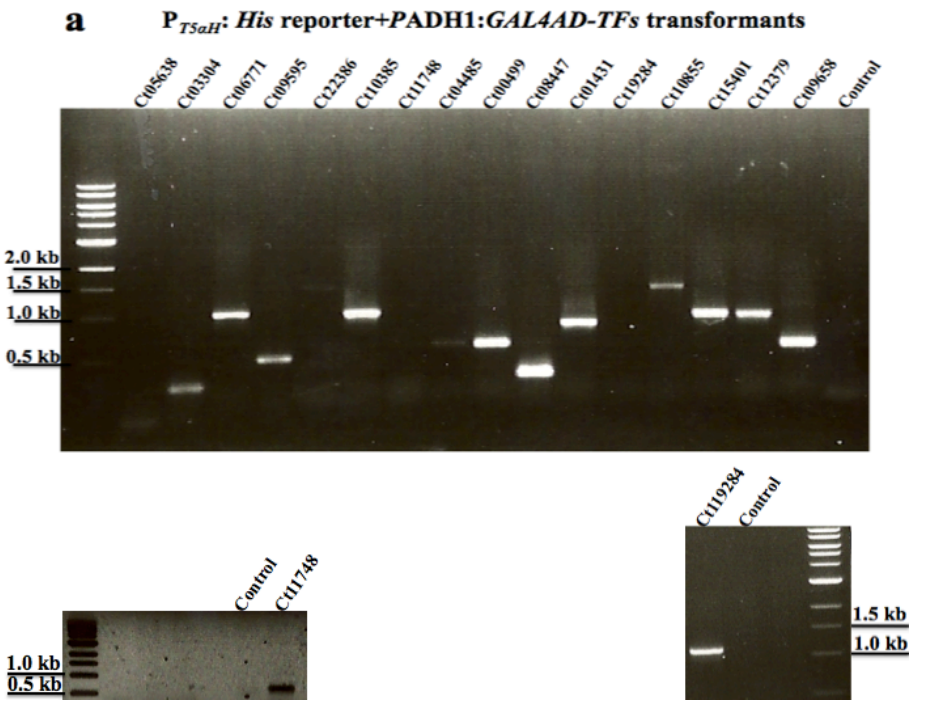
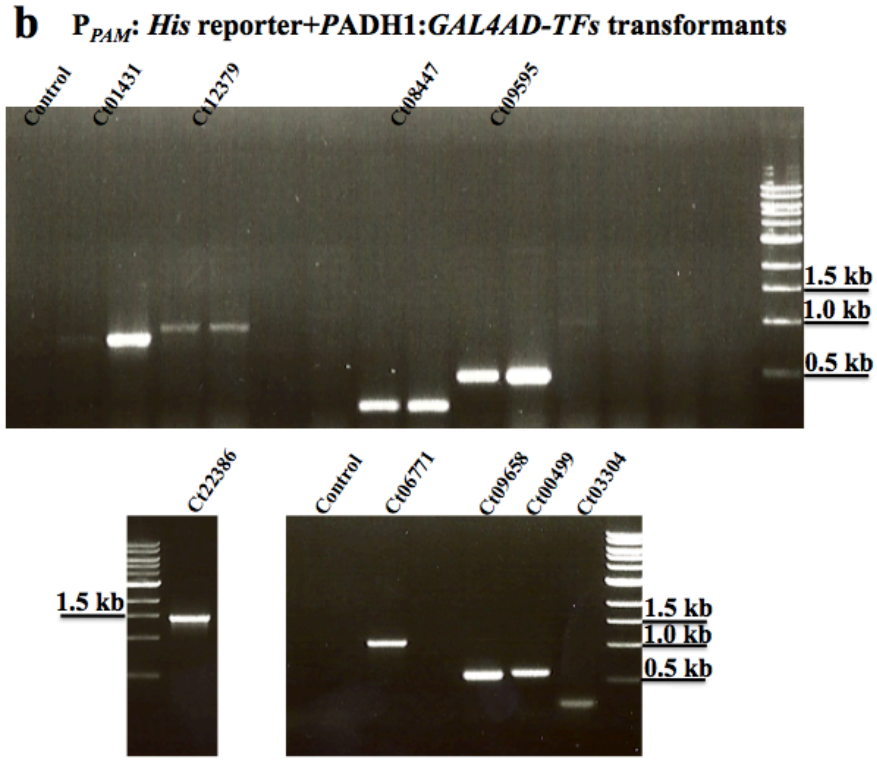


Fig 4.11 (a-e) Confirmation of integration of five paclitaxel biosynthetic gene promoters into the yeast genomic DNA of YM4271 strain by PCR. Five paclitaxel biosynthetic gene promoters, including TASY, T5αH, PAM, DBAT and DBTNBT, were cloned into pHISi (Fig 4.10). The $P_{promoter}$: *His* was then linearized by *Xho* I and integrated into YM4271 genome by transformation. Promoter-specific primiers (SI Table 4) were used in PCR to confirm the genomic intergration of P_{TASY}: *His* (a), $P_{T5\alpha H}$: *His* (b), P_{PAM}: *His* (c), P_{DBAT}: *His* (d) and P_{DBTNBT}: *His* (e).


a
P T5aH: His reporter+PADH1:GAL4AD-TFs transformants
Ct05638
Ct03304
Ct06771
Ct09595
Ct22386
Ct10385
Ct11748
Ct04485
Ct00499
Ct08447
Ct01431
Ct19284
Ct10855
Ct15401
Ct12379
Ct09658
Control
2.0 kb
1.5 kb
1.0 kb
0.5 kb
Control
Ct11748
1.0 kb
0.5 kb
Ct119284
Control
1.5 kb
1.0 kb


b
P PAM: His reporter+PADH1:GAL4AD-TFs transformants
Control
Ct01431
Ct12379
Ct08447
Ct09595
1.5 kb
1.0 kb
0.5 kb
Ct22386
1.5 kb
Control
Ct06771
Ct09658
Ct00499
Ct03304
1.5 kb
1.0 kb
0.5 kb

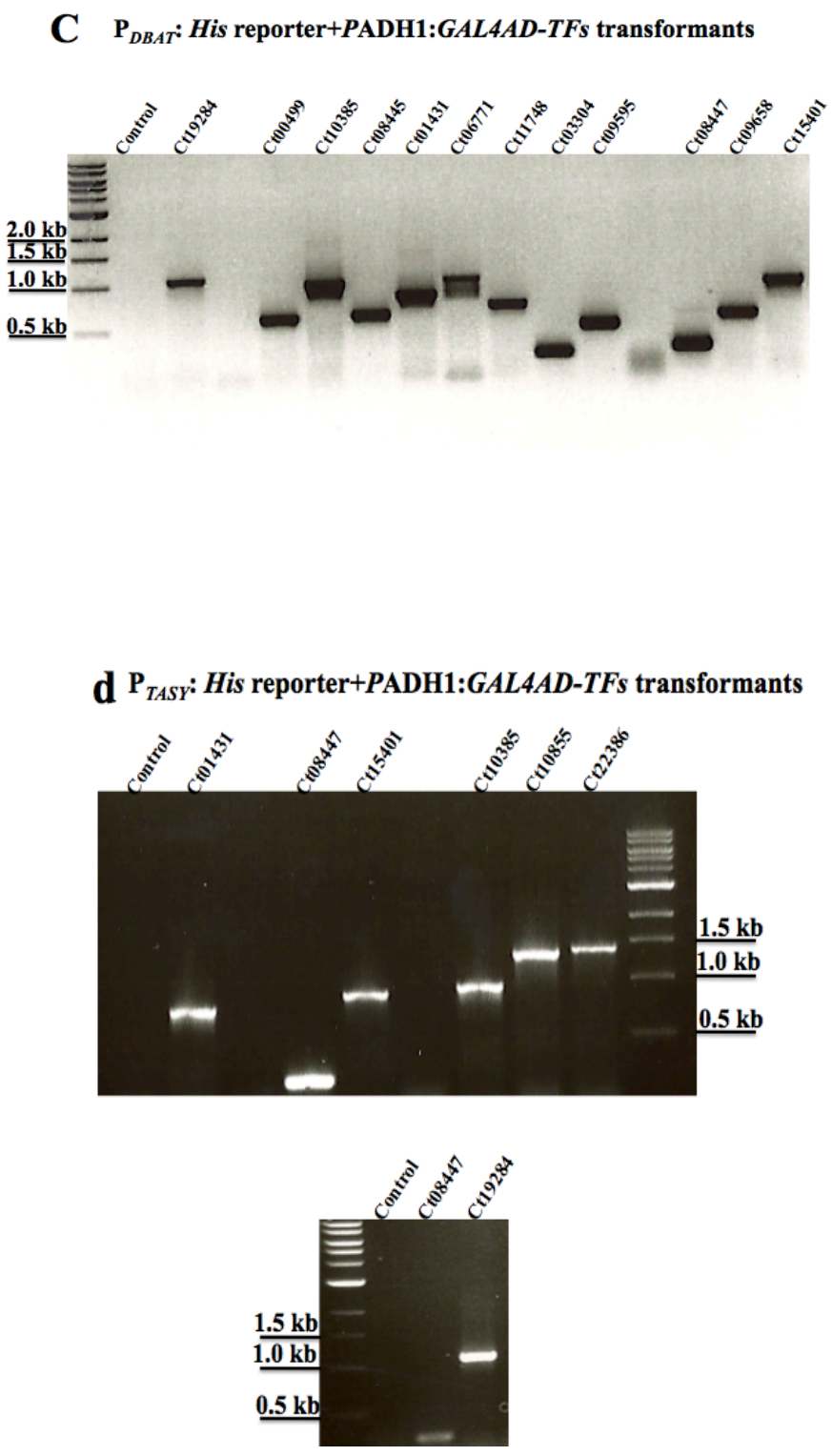


Fig 4.12 (a-d) Confirmation of introduction of PADH1: *GAL4-TFs* into each of four Ppromoter: *His* reporter strains by PCR. The effector constructs, PADH1: *GAL4-TFs,* were introduced into each of the four different YM4271 promoter reporter

strains (Fig 4.11) by transformation. TF-specific primers (SI Table 4) were used in PCR to confirm that 13 different PADH1: *GAL4-TF* effector constructs were successfully introduced into $P_{T5_{-35}H}$: *His* (a), 9 different PADH1: *GAL4-TF* effector constructs were successfully introduced into P_{PAM}: *His* (b), 12 different PADH1: *GAL4-TF* effector constructs were successfully introduced into P_{DBAT}: *His* (c) and 8 different PADH1: *GAL4-TF* effector constructs were successfully introduced into P_{TASY}: *His* (d).

4.4 Discussion

In conclusion, the positive control of this assay was established; however, this system could not be successfully deployed to screen TF-promoter interactions (Fig 4.10). There was no convincing yeast growth shown for any of the 95 putative TF protein-DNA interactions (5 promoters against 19 TFs). One reason could be that plant-specific post-translational modifications, such as phosphorylation, are required for TF protein binding activity (Ishihama et al. 2011). Some of *Taxus* TF proteins might not be post-translationally modified (PTM) in the yeast system due to the lack of plant PTM machineries. Therefore, their DNA binding activity is not activated. Another possible reason is that some TFs might require a TF dimer (either homodimer or heterodimer) or cofactor to form a higher order protein complex for protein-DNA interactions (Bruce et al. 2000; Lai et al. 2011; Xu et al. 2006). As a result, false negatives, which cannot be detected in the yeast system, can be produced. Because there was only one TF expressed in each promoter-reporter strain, a protein-DNA interaction might not be detected in the absence of another one or a cofactor. Hence, it might be necessary to co-express two TFs in one type of promoter-report strain. A candidate TF dimer, which could be co-expressed, is a MYB/bHLH heterodimer or WRKY/WRKY homodimer (Bruce et al. 2000; Xu et al. 2006). However, a *Taxus chinensis* WRKY (*Tc*WRKY) has been shown to interact with *cis*-elements present in *DBAT* promoter by Y1H assay and regulate its expression in *T. chinensis* suspension cells (Li et al. 2013). There could be a technical issue that led to the results that no TF-promoter interactions were detected in Y1H assay. One possible problem could be that the protein expression vector (pGADgate424 lapau), which is designed to express the recombinant proteins, was not able to express *Taxus* TF-GAL4AD recombinant proteins. The evidence for this was that only GAL4AD, but not GAL4AD-TF recombinant proteins, was detected by western blots (Eunjung et al., unpublished data). Prior to detection of TF proteins, RT-PCR could be used to detect expression of TF encoding genes.

Chapter 5 Gateway-compatible and *Agrobacterium*-mediated Plant Transient Expression Assays

Plant transient expression assays (TEAs) have been used to study gene function, such as protein-DNA interactions, in planta (Yang et al. 2000). It is a rapid screening system for assigning gene function. In addition, this method is frequently used when there is a difficulty in producing stably transformed transgenic plants for analysis of gene function *in planta* (Shang et al. 2007). An *Agrobacterium*-mediated transient gene expression assay has been developed for the rapid study of interactions between plant transcription factors and promoters *in vivo* (Yang et al. 2000). Therefore, this method might be suitable for screening interactions between 5 paclitaxel biosynthetic gene promoters and 19 putative TFs in tobacco leaves.

5.1 Experimental design

Two gateway-compatible vectors were initially constructed. One was a gateway-compatible plant promoter-reporter (*GUS*) binary vector using either the whole promoter or tandem repeats of isolated TF binding motifs as the bait, such as repeats of the AP2/ERF TF binding sequence (GCCGCC). In the case of tandem repeats, a minimal promoter is needed for transcriptional initiation of the *GUS* reporter gene (Fig 5.1b). Another construct is the effector vector for expression of TFs driven by a *35S* promoter. Cre/lox recombination is a site-specific recombination technology. The Cre recombinase recombines a pair of target sequences, which is known as the *Lox* sequences (loxP site). The Cre recombinase and loxP sites are originally derived from a bacteriophage P1 (Sauer and Henderson 1988). Both constructs in this plant TEA contain a loxP site. They were therefore recombined into a single recombinant construct through Cre/loxP recombination. The recombinant plasmid was introduced into *Agrobacterium* and then infiltrated into tobacco leaves by *Agrobacterium*-infiltration (Fig 5.1a) (Yang et al. 2000). *GUS,* which encodes the beta-glucuronidase from *E. coli*, is often used in a reporter gene system in transgenic plants (Jefferson 1989). This enzyme catalyses the hydrolysis of a wide range of glucuronides. In the case of 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc), the product of the reaction gives rise to a clear blue colour. GUS is very stable and still active after prolonged storage. Plants that express *GUS* are healthy and fertile (Jefferson et al. 1987). The *GUS* reporter gene system can be used to study *in vivo* protein-DNA interactions in tobacco leaves (Fan and Dong 2002; Yang et al. 2000). Upon *GUS* expression, which can be either visualised or quantitatively analysed (Jyoti Kapila 1997; Yang et al. 2000), it indicates not only the interaction between plant promoters and TFs, but also that the candidate TF is a positive regulator in the regulation of gene expression.

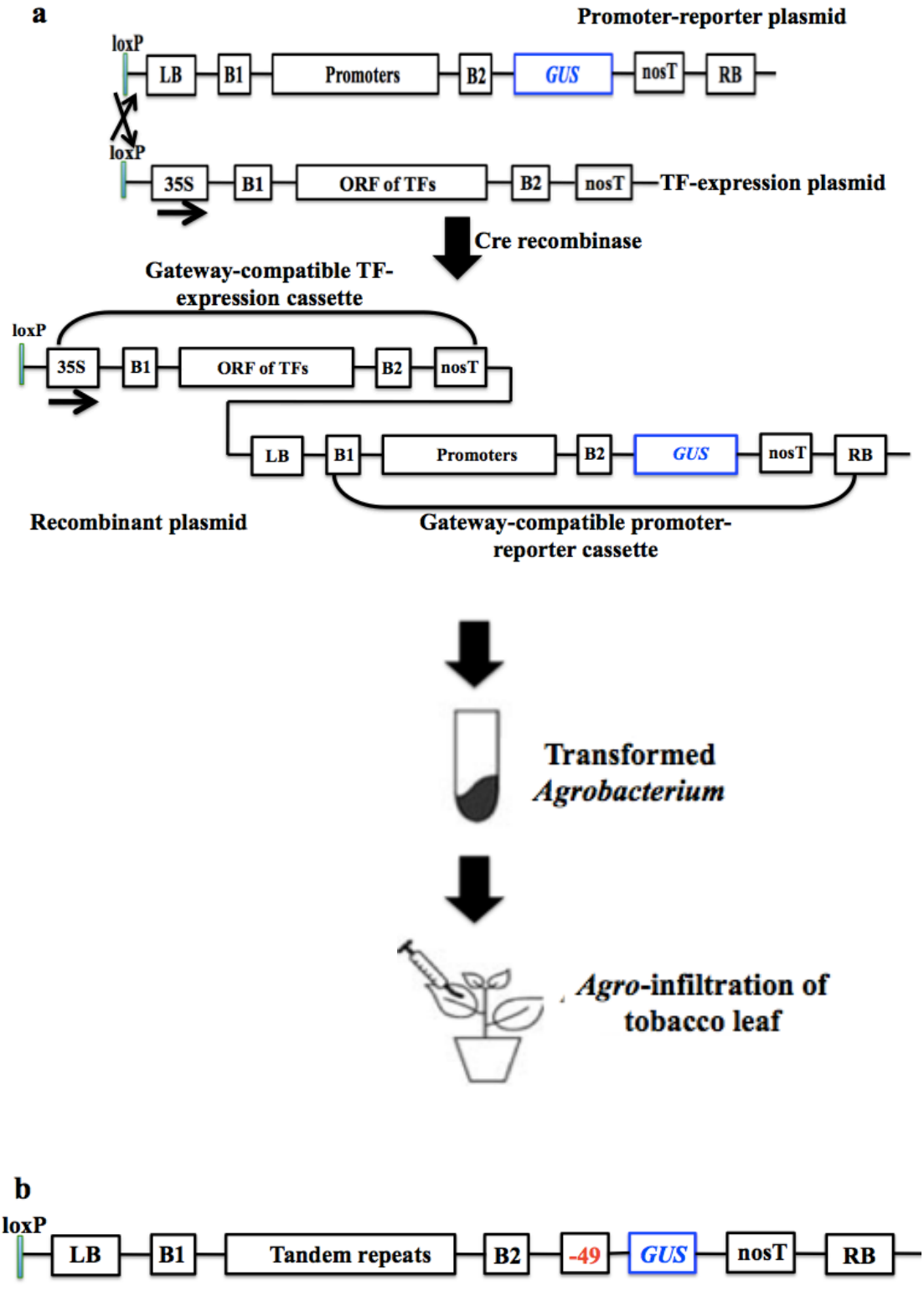


Fig 5.1 a: Construction of a plant gateway-compatible recombinant plasmid and *Agrobacterium*-mediated infiltration: a gateway-compatible promoter-reporter

(GUS) plasmid is fused with a gateway-compatible TF expression plasmid via the loxP site through Cre/loxP recombination. The resulting two-gateway recombinant plasmid is introduced into tobacco leaves through *Agrobacterium*-mediated infiltration. b: the construction of the promoter-reporter plasmid in the case of using tandem repeats of TF binding motif as the bait. *GUS*: beta- glucuronidase gene; *35S*: cauliflower mosaic virus (CaMV) *35S* promoter; nosT: *nos* terminator; -49: cauliflower mosaic virus (CaMV) *35S* minimal promoter; B1: attB1site (Fig 2.1); B2: attB2 site (Fig 2.1); LB: left border; RB: right border.

5.2 Positive control for plant transient expression assay

The positive control for plant TEA is the already known interaction between ORA47 (AP2 TF) and plox3 in *Arabidopsis* (Pauwels et al. 2008). Instead of using the whole *lox3* promoter, the tandem repeat of AP2 DNA binding motif (GCCGCC) was used as the bait to screen for putative AP2 TFs (Fig 5.2).

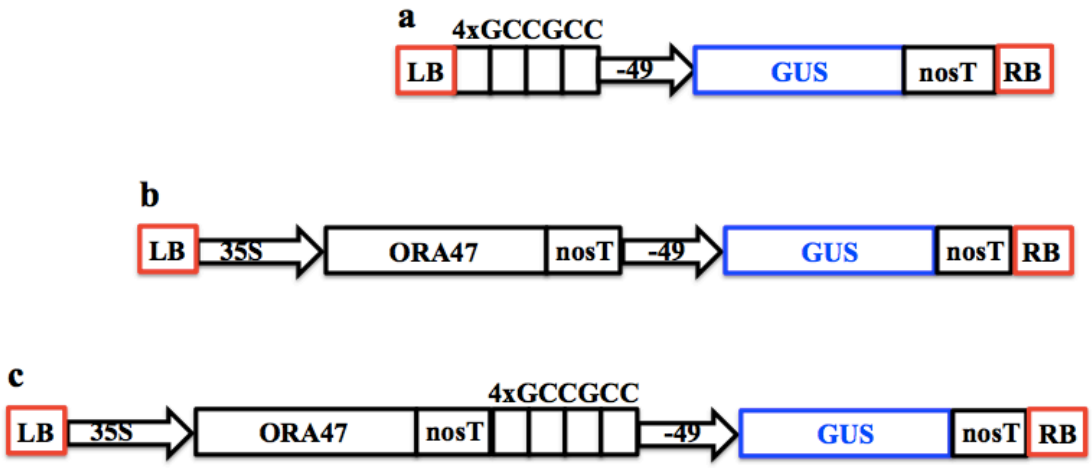


Fig. 5.2 Construction of the binary vectors for the positive control of plant TEA. The two internal control constructs (a & b) and the target construct (c). Two internal controls (a & b) are used to eliminate the possibility that the endogenous AP2 TFs bind to GCCGCC motif, which results in false positive results. -49: the 35S minimal promoter. ORA47: AP2/ERF domain TF protein; GCCGCC: AP2/ERF associated DNA binding motif; *GUS*: beta- glucuronidase gene; *35S*: cauliflower mosaic virus (CaMV) *35S* promoter; nosT: *nos* terminator; -49: cauliflower mosaic virus (CaMV) *35S* minimal promoter; B1: attB1; B2: attB2; LB: left border; RB: right border.

5.2.1 Cre/loxP mediated plasmid recombination

Before cloning the ORF of putative TFs and GCC tandem repeats into pJET1.2-GW-loxP (ampicillin selection) (SF 5) and pMDC162-GW-35Sm-loxP (kanamycin selection) (SF 6) through the gateway system, respectively, these two backbones were first mixed together to see whether the recombination would occur in the presence of Cre recombinase. After Cre recombination, there were a small number of colonies growing on both ampicillin and kanamycin double selection plates (Fig 5.3 c & f), compared to the number of colonies growing on single selection plates (either ampicillin or kanamycin) before Cre recombination (Fig 5.3 a & b). This was probably because of the low recombinant efficiency (20%-30%). Furthermore, as expected, there was no colony growth on double selection plates when only one of the plasmids was introduced into the relevant *E. coli* strain by transformation (Fig 5.3 d & e).

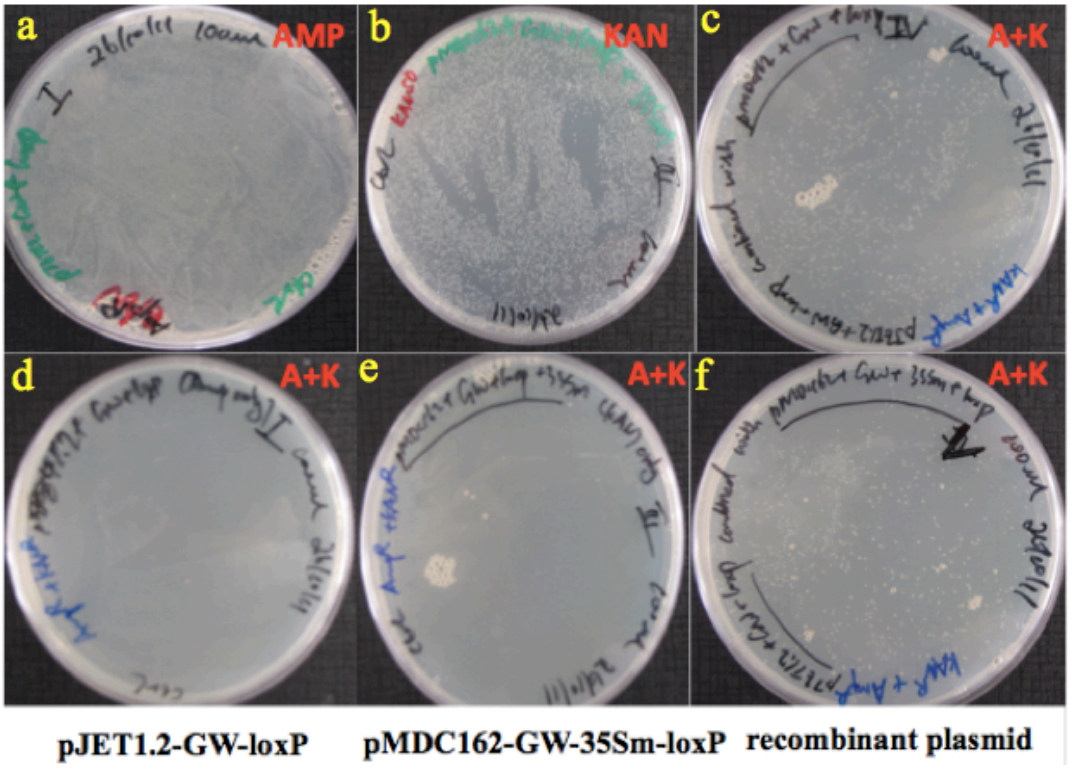


Fig 5.3 Introduction of pJET1.2-GW-loxP, pMDC162-GW-35Sm-loxP and putative recombinant plasmid into DB3.1 *E.coli* strain, and selection on ampicillin (A),

kanamycin (K) and ampicillin and kanamycin (A+K). As expected, transformants with pJET1.2-GW-loxP grew on ampicillin plates, but not on ampicillin and kanamycin plates (a & d), and transformants with pMDC162-GW-35Sm-loxP grew on kanamycin plates, but not on ampicillin and kanamycin plates (b & e). Transformants with putative recombinant plasmid of both pJET1.2-GW-loxP and pMDC162-GW-35Sm-loxP grew on ampicillin and kanamycin plates (c & f).

5.2.2 Evidence for co-existence of two incompatible plasmids within the same host

There are three possibilities that would enable the transformants to grow on double selection plates after Cre recombination. The first one is that the two plasmids are fused into one recombinant plasmid and the recombined plasmid replicates within the host. The second is that the two plasmids are not fused together, instead that they co-exist and independently replicate within the same host. The third reason is that either of the two plasmids or both of them co-exists with the recombinant plasmid within the same host (Fig 5.4).

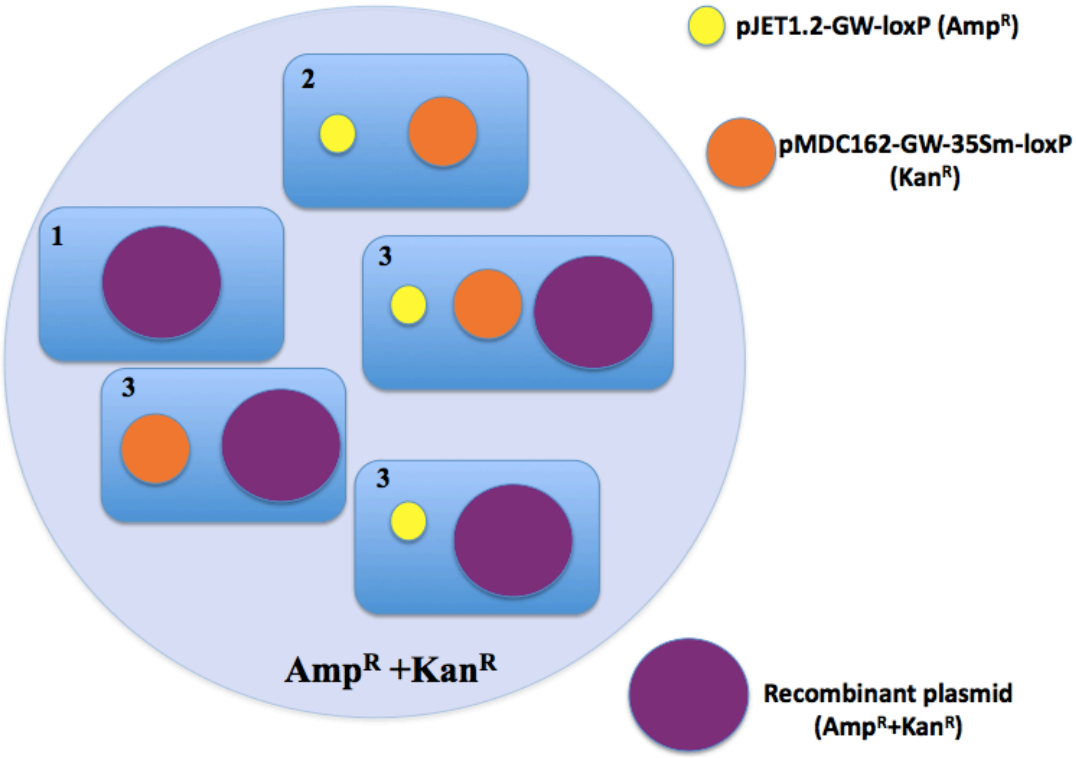


Fig 5.4 The possible plasmid combinations that exist within the same host after Cre recombination, which allows the transformants to grow on kanamycin and ampicillin double selection plates. 1=the first possibility; 2=the second possibility; 3=the third possibility.

In order to show which of the possible plasmid combinations exists inside the host, the putative recombinant plasmids were extracted and digested with *Pac* I enzyme. In theory, *Pac* I digestion of putative recombinant plasmids (Fig 5.5c), pJET1.2-GW-loxP (Fig 2.3b) and pMDC162-GW-35Sm-loxP (Fig 2.3a) is supposed to give rise to three *Pac* I fragments (~13.0, ~3.1 and ~2.9 kb) (Fig 5.5c), a single large fragment of linearized pJET1.2-GW-loxP (~6.0 kb) and pMDC162-GW-35Sm-loxP (~13.0 kb), respectively. However, there were two *Pac* I fragments shown after *Pac* I digestion in 12 samples: the upper band had the same size as the linearized pMDC162-GW-35Sm-loxP (control 2b Fig 5.5a); the lower band was at the same level as the linearized pJET1.2-GW-loxP (control 1b Fig 5.5a). Furthermore, *Asc* I was also used

to digest the putative recombinant plasmids. The result was the same as that of *Pac* I digestion (Fig 5.5c & 5.6). Therefore, it is most likely that the majority of transformants had two plasmids co-existing and independently replicating within the same host. In addition, both plasmids contain the same replication origins, pBR322. Therefore, they are incompatible to each other when replicating in the same host. However, it has been shown that two incompatible plasmids are not rapidly lost, but are able to persistently replicate for multiple overnight growth cycles (Velappan et al. 2007). Furthermore, the small plasmids showed much stronger bands than the large ones on a gel; this could be due to the fact that the replication of the small plasmid gradually outcompetes the large one (Fig 5.5 a & b). In other words, the large plasmid will be gradually lost, and the small one will remain replicating in the host.

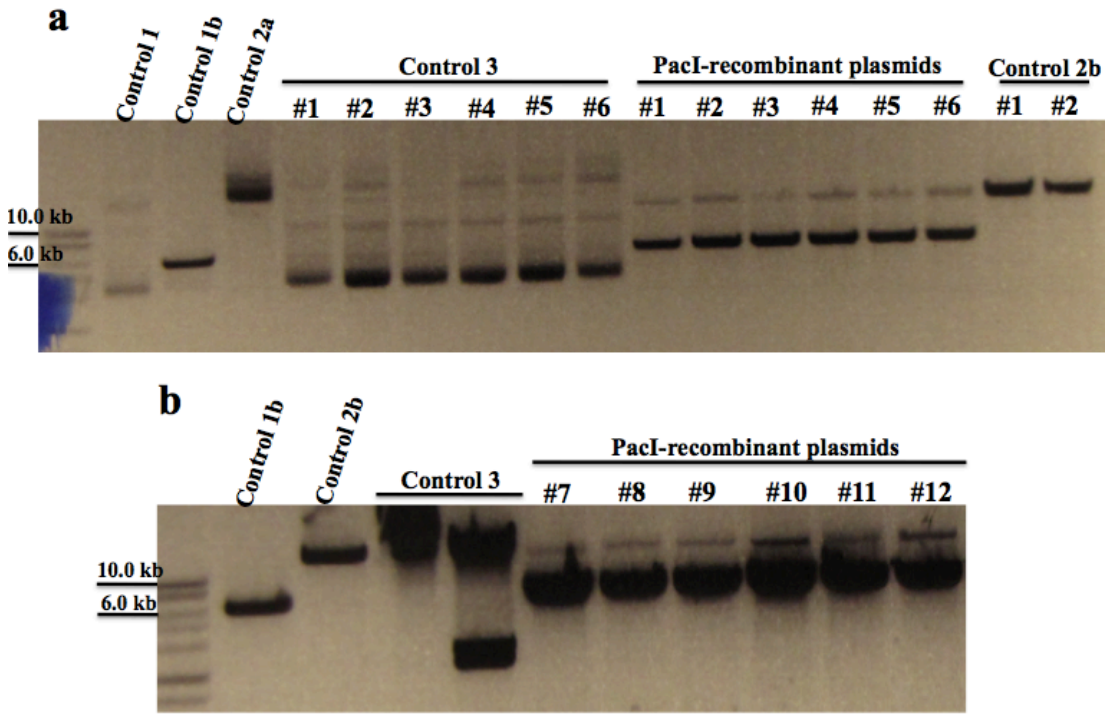


Fig 5.5 a & b: *Pac* I-digestion of putative recombinant plasmids, pJET1.2-GW-loxP and pMDC162-GW-35Sm-loxP. Control 1a: Uncut pJET1.2-GW-loxP; Control 1b: PacI- pJET1.2-GW-loxP; Control 2a: Uncut pMDC162-GW-35Sm-loxP; Control 3b: PacI- pMDC162-GW-35Sm-loxP; Control 3: Uncut putative recombinant plasmid.

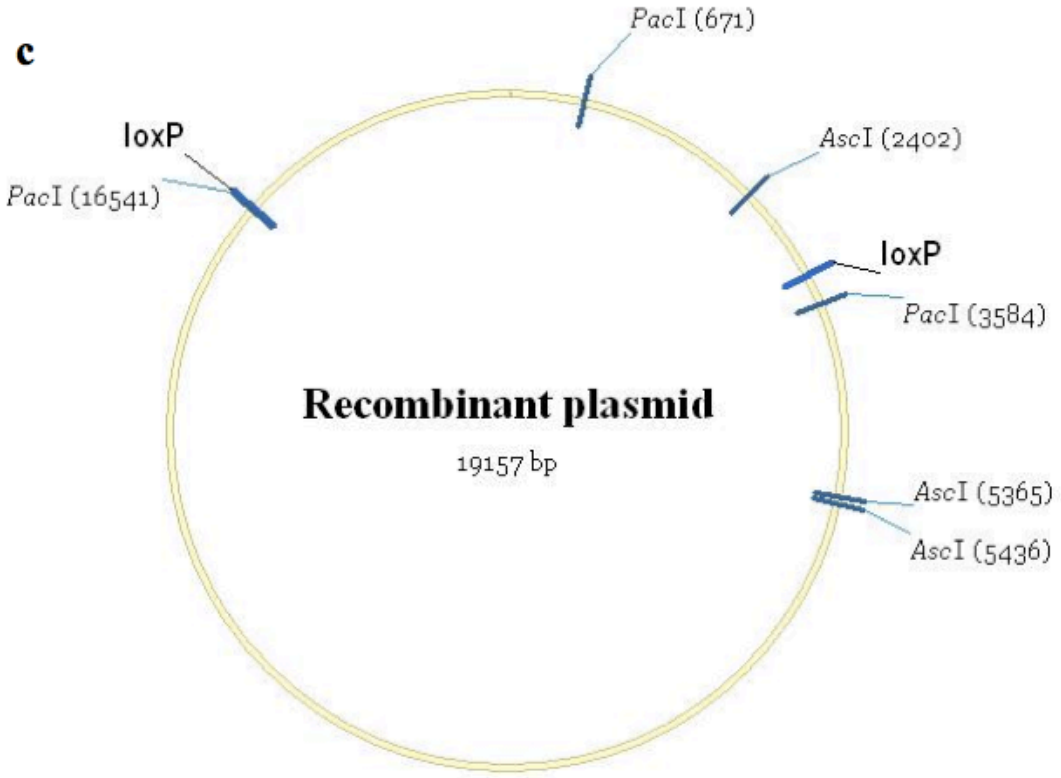


Fig 5.5 c: Map of the recombinant plasmid. Three *Pac* I sites are highlighted in red dotted box; three *Asc* I sites are highlighted in blue dotted box. The size of three expected *Pac* I fragments: ~3.1 Kb, 2.9 Kb and 13.0 Kb.

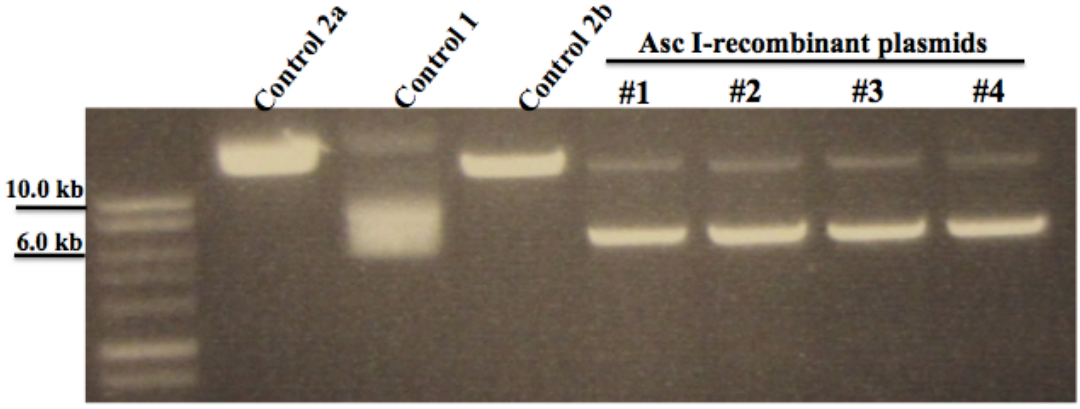


Fig 5.6 *Asc* I-digestion of putative recombinant plasmids and pMDC162-GW-35Sm-loxP. Control 1: uncut putative recombinant plasmid; Control 2a: uncut pMDC162-GW-35Sm-loxP; Control 2b: AscI- pMDC162-GW-35Sm-loxP.

The co-existence of two independent plasmids within the same host could be firstly explained by the fact that the recombination reaction did not take place before transformation so that two plasmids were independently co-introduced into the same *E. coli* host. This might be due to that the Cre/loxP recombination system did not work at all. Regarding this possible explanation, the insertions of loxP sites into both binary vectors were confirmed by analytic digestion as well as sequencing results (SF 4 & 5). Furthermore, in order to test whether the inserted loxP site functioned to recombine another loxP-containing plasmid in the presence of Cre recombinase, the pJET1.2-GW-loxP (Amp^R) was chosen to react with pUNI 51 (Kan^R) through Cre/loxP recombination. This was followed by introduction into DB3.1 *E. coli* strains through transformation. pUNI 51 vectors also contain loxP sites and are commonly used as shuttle vectors to construct a recombinant DNA without restriction enzymes (Liu et al. 1998). As expected, there was no colony growth for the introduction of either pUNI 51 or pJET1.2-GW-loxP into DB3.1 on double selection plates before Cre/loxP recombination (Fig 5.7 a, b, c & d). pUNI 51 can only be propagated in BW cells (Pir^+), but not in DB3.1 (Pir^-). Because *E. coli* needs to express the P protein (the *pir* gene product) for replication of plasmids containing the R6kr origin of replication (Yang et al. 2000) and pJET1.2-GW-loxP can only be selected on ampicillin selection plates but not on kanamycin selection plates. After recombination, the DB3.1 transformants were able to grow on double selection plates (Fig 5.7 e). This indicates that the plasmid recombination occurred between the inserted loxP site in pJET1.2-GW-loxP and that in pUNI 51 in the presence of Cre recombinase. As a result, the fused pUNI 51 was replicated under the replication origin contained in pJET1.2-GW-loxP in DB 3.1 cells. Meanwhile, the kanamycin resistant gene on pUNI 51 was activated, which allowed the transformants to be selected on kanamycin and ampicillin double selection plates. In addition, the sequencing results confirmed the appearance of loxP site in both vectors. So, the recombination between pJET1.2-GW-loxP and pUNI 51 further confirmed that the Cre/loxP recombination system works when a loxP site is inserted into a plasmid. Therefore, these observations prove that the inserted loxP allows the recombination reaction to take place in the presence of Cre recombinase before transformation.

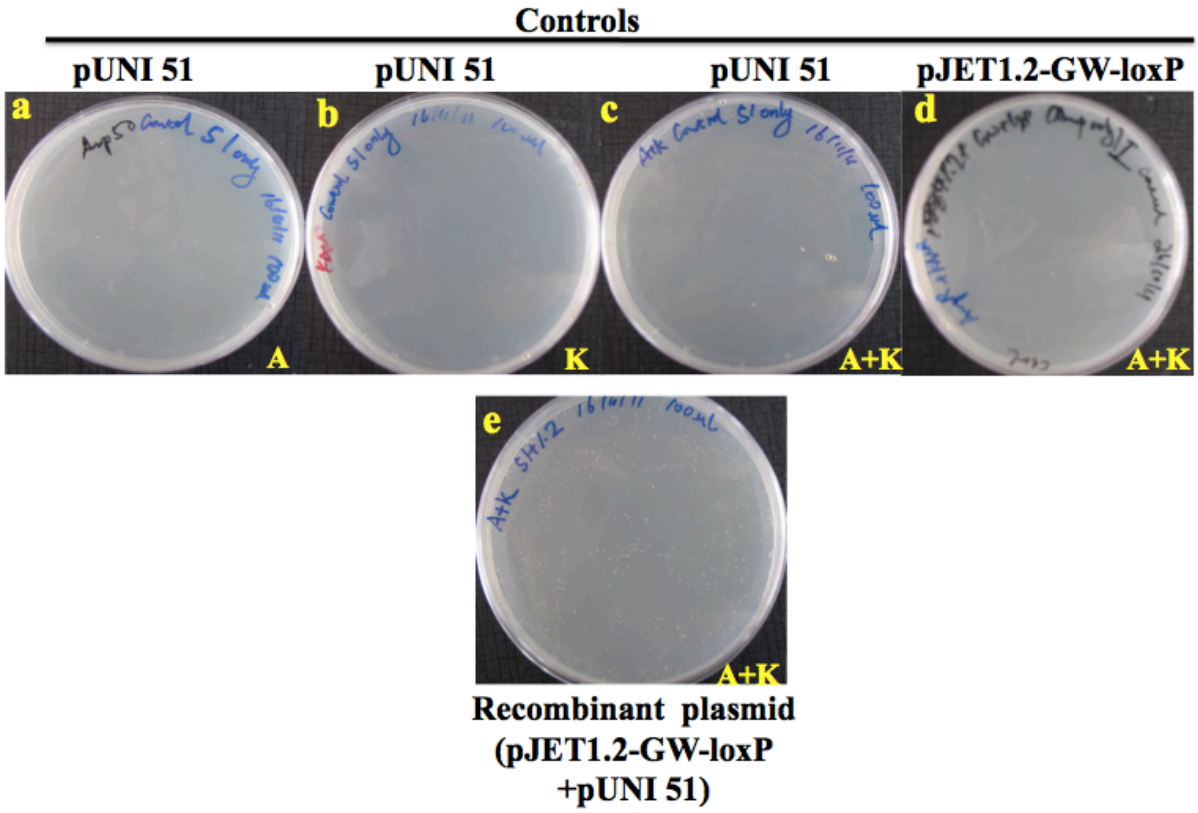


Fig 5.7 Cre/loxP recombination of pUNI 51 with pJET1.2-GW-loxP. The controls: a, b, c & d. e, the growth of DB 3.1 transformed with recombinant plasmid on ampicillin (A) and kanamycin (K) double selection plates. As expected, compared to the controls, the transformants with recombinant plasmids grew on ampicillin and kanamycin plates.

The co-existence of two independent plasmids within the same host could also be explained by the fact that the recombination reaction occurred before transformation, the recombinant plasmid, however, was not able to replicate itself inside the host. This could be caused by the presence of two identical replication origins (pBR322) in one recombinant plasmid. However, it has been suggested that in ColE1 plasmid, replication initiates at a single pBR322 replication origin even when there are several identical potential origins present in the same plasmid. In other words, only one

pBR322 replication origin is active per plasmid (Martin et al. 1992). Therefore, the recombinant plasmid (pTEA) might be able to replicate inside the host.

Finally, because the recombinant plasmid contains two identical *cddB* genes (Fig 5.5 c), the third explanation could be that homologous recombination most likely happened between these two identical sequences; as a result, the recombinant DNA was split into two separate plasmids once introduced into DB3.1 competent cells and these two dependent plasmids co-existed in the same host. However, the genotype of DB3.1 shows this strain is Δ (sr1-recA). Therefore, it has reduced occurrence of unwanted homologous recombination.

5.3 Discussion

Through the Cre/loxP recombination, a recombinant binary vector was constructed by recombining a TF expression vector with a promoter-*GUS* reporter vector *in vitro*. However, instead of the existence of a recombinant vector inside the *E. coli* after transformation, these two plasmids independently co-existed (Fig 5.5 a & b). After two possible reasons were eliminated, one possible explanation for this is that there could still be a recombination event occurring between two *cddB* genes present in each gateway cassette after introducing the recombined plasmid into DB 3.1 *E. coli* strain, although the genotype of DB 3.1 shows that this strain has reduced occurrence of DNA recombination. As a result, two independent plasmids co-existed and independently replicated inside the host. Another possible explanation is that there were only a tiny minority of cells transformed with recombinant plasmids because of a low recombinant efficiency (20%-30%). The transformation efficiency became even lower when transforming a large sized vector (the recombinant vector is ~19 kb) (SI Table 1). Altogether, this resulted in a low possibility of successfully screening for colonies that contain the desired recombinant plasmid. Since there were 16 colonies picked up for extraction of recombinant plasmids, and none of them was found to contain the desired recombinant plasmid (Fig 5.5 a & b). Therefore, a positive colony could possibly be selected if a large number of colonies were screened.

In short, this plasmid construction has two drawbacks: one is the occurrence of recombination events; the other is that the size of the recombinant is too large to be effectively introduced into *Agrobacterium* by transformation. Therefore, a more efficient plasmid construction is needed.

Chapter 6 Technically improved plant transient expression assays

The Cre/loxP-mediated plasmid recombination between a gateway-compatible TF expression vector (pJET.12GWTF) and promoter-reporter vector (pMDC162-GW-35Sm) did not work. This might be caused by technical issues, for example, low recombination efficiency and homologous recombination. In addition, provided that the recombinant could be constructed, the size of which (~ 19 kb) would be too large to be efficiently transformed. Therefore, in order to overcome these problems, a relatively small sized construct needs to be made, which itself contains both gateway-compatible TF expression and promoter-reporter cassettes. In addition, both TF-encoding genes and target promoters can be cloned simultaneously, because of the use of four different attB sites (B1, B2 B3 and B4) (Fig 6.1A).

6.1 Experimental design

In this technically improved plant TEA, a two-gateway cloning system is formed by cloning the entire TF expression gateway cassette, which is amplified from pK7m34GW2-8m21GW3,0 (Fig 2.4a), into the backbone of a promoter-reporter binary vector, pHGWFS7,0. Therefore, the resulting plasmid contains two gateway cassettes. The TF expression cassette, which is driven by a *35S* promoter, contains attB4 at the left end and attB3 at the right end of a TF ORF sequence. The bait-reporter cassette contains attB1 at one end and attB2 at the other end of the bait sequence (Fig 2.4c and 6.1A). As a result, the cloning of TFs and baits can be carried out simultaneously. In the case of using tandem repeats of TF binding motifs as the bait, a *35S* minimal promoter downstream of the bait is needed to initiate the expression of *GUS* genes. In addition, there is one control construct, which contains the bait-reporter cassette (Fig 6.1B). Finally, these two constructs are introduced into *Agrobacteria* by transformation, and the transformed *Agrobacteria* are infiltrated into tobacco leaves through *Agro*-infiltration (Chapter 2).

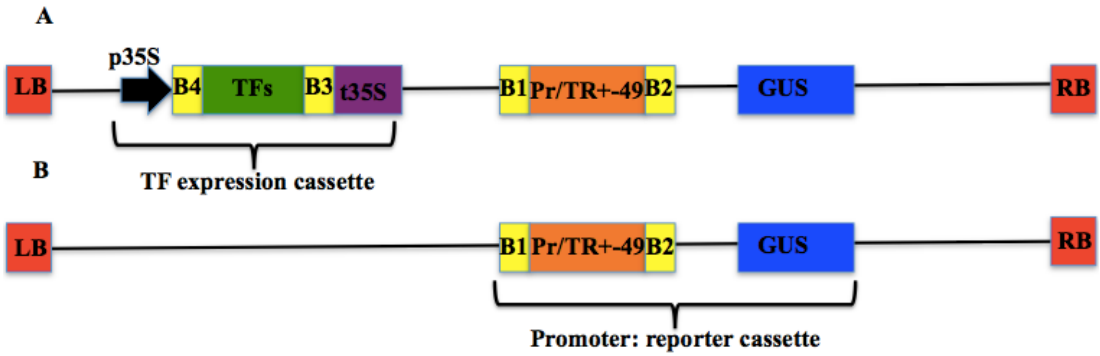


Fig 6.1 Constructs of *Agrobacterium*-mediated plant TEA. A: A two-gateway target construct. This construct was made by inserting a gateway-compatible TF expression cassette (p35S-B4-*cddB*-B3-t35S) into the backbone of pHGWFS7,0. TF encoding genes are cloned through B4 and B3-mediated gateway cloning system, while either promoters or tandem repeats is cloned through B1 and B2-mediated gateway cloning system. These two reactions can occur simultaneously. B: the internal control, pHGWFS7,0. Both constructs are infiltrated into tobacco leaves through *Agrobacterium*-mediated infiltration. -49: *35S* minimal promoter; *GUS*: beta-glucuronidase gene. Pr: promoters; TR: tandem repeats; B4: attB4; B3: attB3; B2: attB2; B1: attB1; LB: left border; RB: right border.

6.2 Positive controls for improved plant TEA

The interaction between *P*lox3 and ORA47 was used as a positive control of plant TEA. A tandem repeat of a TF binding motif was used as bait, because tandem repeats of a given *cis* motif result in an increased chance of TF occupancy relative to a single motif and hence result in increased reporter gene expression (Yang et al. 2000). Therefore, three tandem repeats of GCCGCC, the AP2 TF-associated binding motif, were used in conjunction with ORA47 as a positive control in the plant TEA. There were two constructs in the positive control, including the target construct and one internal control (Fig 6.2). The *Agrobacteria* strain GV3101 transformants of these two plasmids were selected on spectinomycin plates and further confirmed by PCR (Fig 6.3).

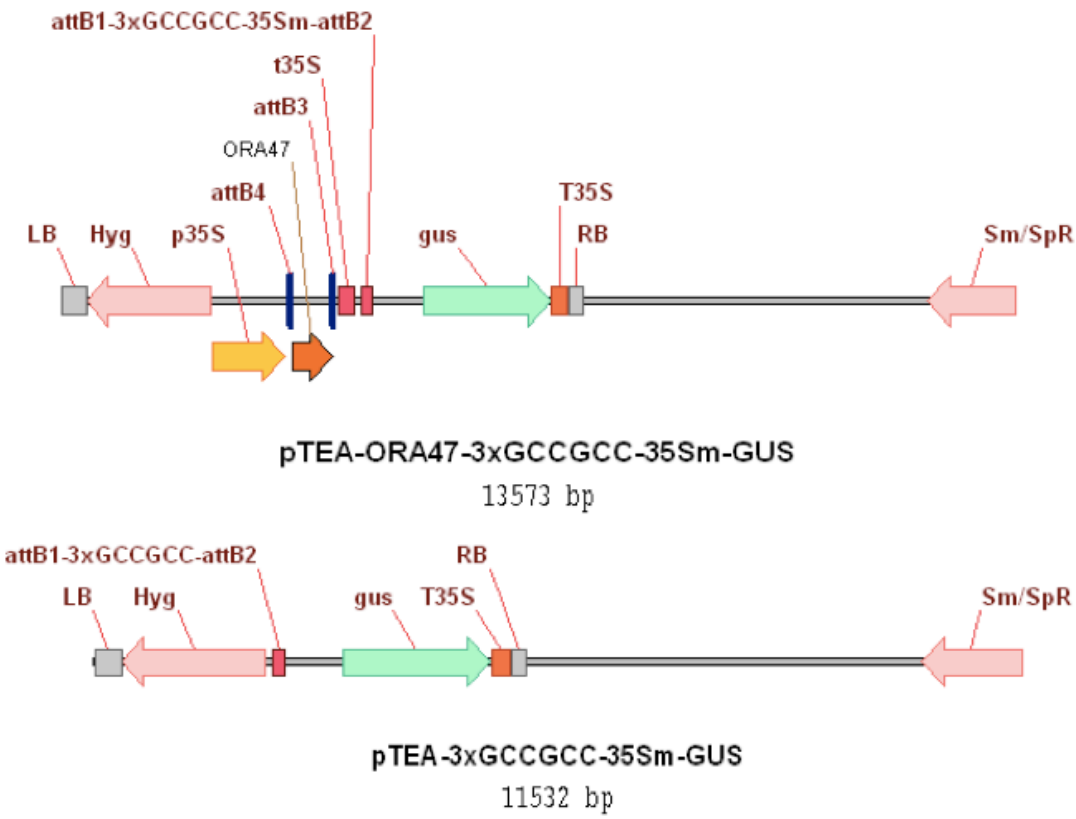


Fig 6.2 The constructs for positive control of plant transient expression assay. Upper panel: the target construct. This construct contains a TF expression cassette (p35S-attB4-ORA47-attB3-t35S), from which ORA47 is expressed, and a GCCGCC tandem repeat-GUS reporter cassette (attB1-3xGCCGCC-attB2). Lower panel: the internal control. This construct only contains a GCCGCC tandem repeat-GUS reporter cassette. ORA47: AP2/ERF domain TF protein; GCCGCC: AP2/ERF associated DNA binding motif; *GUS*: beta- glucuronidase gene; *35S*: cauliflower mosaic virus (CaMV) *35S* promoter; nosT: *nos* terminator; *35Sm*: cauliflower mosaic virus (CaMV) *35S* minimal promoter; B4: attB4; B3: attB3; B1: attB1; B2: attB2; LB: left border; RB: right border. Hyg: hygromycin resistance; Sm/Sp^R: spectinomycin resistance.

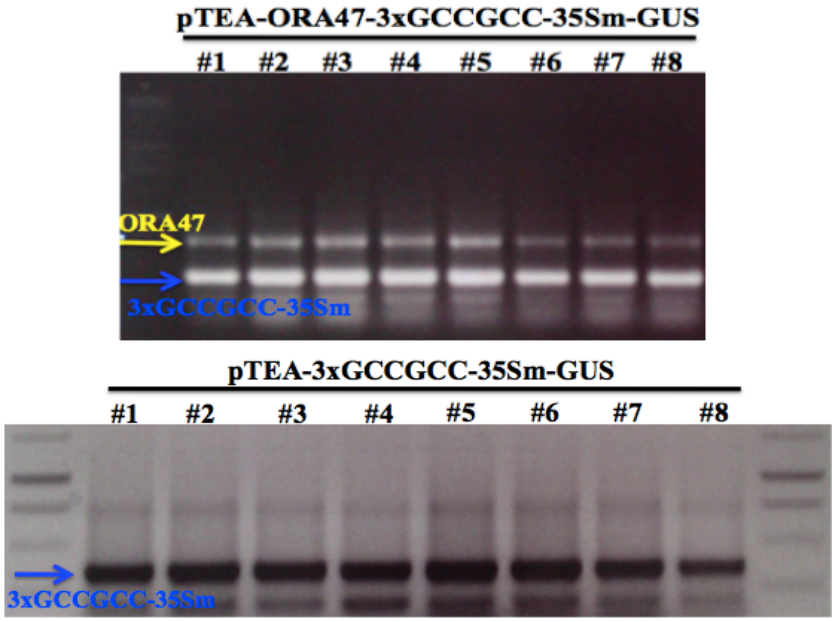


Fig 6.3 Confirmation of *Agrobacteria* GV3101 transformants of pTEA-*ORA47*-3xGCCGCC-*35Sm*-GUS and pTEA-3xGCCGCC-*35Sm*-GUS by amplifying *ORA47* (yellow arrow) and 3xGCCGCC-*35Sm* (blue arrow) by PCR using gene-specific primers (SI Table 4). Upper panel, both ORA47 and 3xGCCGCC-*35Sm* were confirmed to introduced into *Agrobacterium* transformants; lower panel, 3xGCCGCC-*35Sm* was confirmed to introduced into *Agrobacterium* transformants.

The PCR-confirmed positive *Agrobacterium* transformants for each construct were infiltrated into tobacco leaves. As expected, the control, which was infiltrated with infiltration buffer only, did not show blue colour (Fig 6.4 C). Furthermore, the internal control, pTEA-3xGCCGCC-*35Sm*-GUS showed little background (Fig 6.4 B). Finally, the transformants with pTEA-ORA47-3xGCCGCC-*35Sm*-GUS showed strong blue colour (Fig 6.4 A). This experiment was repeated three times.

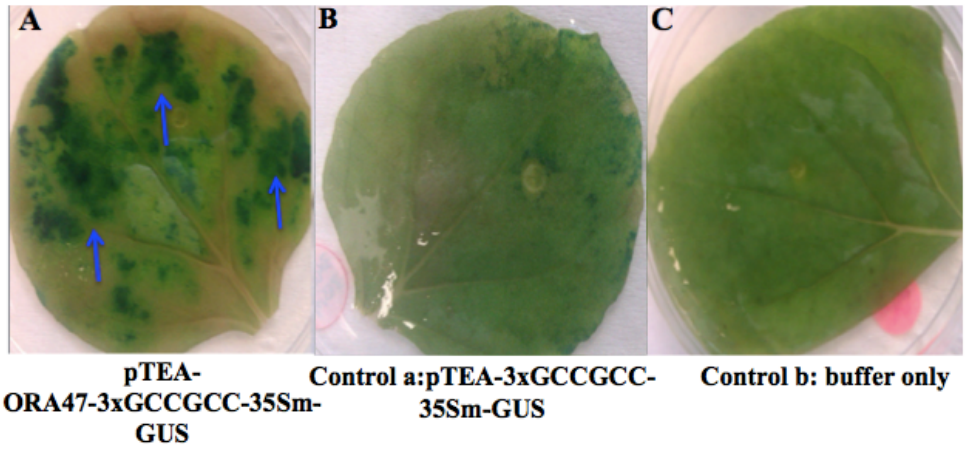


Fig 6.4 *Agro*-infiltration of GV3101 transformants in tobacco leaves at OD_{600}=0.1. A: pTEA-ORA47-3xGCCGCC-35Sm-GUS; B: pTEA-3xGCCGCC-35Sm-GUS; C: infiltration buffer only. As expected, the target construct A showed *GUS* expression, whereas B and C showed no *GUS* expression.

6.3 Discussion

In this improved plant TEA, a TF expression cassette was cloned into a promoter-*GUS* reporter backbone to form a pTEA destination binary vector. The size of this plasmid is ~13 kb, which facilitates its introduction into *Agrobacterium* by transformation. After confirming the presence of pTEA-ORA47-3xGCCGCC-35Sm-GUS in *Agrobacterium* (GV3101) by PCR (Fig 6.3), the results from three replicates of *Agro*-mediated infiltration showed *GUS* expression (Fig 6.4 A) for this positive control but little or no *GUS* expression for the two negative controls (Fig 6.4 B & C). This indicates that there is an interaction between ORA47 and its associated DNA-binding motifs present in plox3, and ORA47 is an activator. This assay was therefore validated as a suitable system to screen plant TFs for binding to their cognate promoter sequences. This system will be used to identify putative *Taxus* TFs (19) that bind to the promoters (5) of paclitaxel biosynthesis pathway genes and activate the expression of the reporter gene (*GUS*).

Chapter 7 Synthesis of TF proteins by *in vitro* transcription and translation using wheat germ

Regarding the rapid progress in the area of proteomics, the production of sufficient amount of protein becomes crucial. Therefore, the development of a protein synthesis system is an important endeavour in biotechnology. There are three systems currently used to synthesize desired proteins: chemical synthesis, *in vivo* expression and cell-free protein synthesis. The first two strategies each have their drawbacks: chemical synthesis is not suitable for synthesis of long peptides because of low yield (Blaschke et al. 2000), and *in vivo* expression can only synthesize proteins that would not interfere with the physiology of the host cells (Chrunyk et al. 1993; Goff and Goldberg 1987). In contrast, cell-free protein synthesis can produce large peptides as accurately as ones synthesized by *in vivo* translation (Kurland 1982; Pavlov and Ehrenberg 1996). In addition, this system can express proteins that have a significant effect on the physiology of the host cells. One of the most convenient eukaryotic cell-free protein synthesis systems is based on the wheat germ extracts. This cell-free protein synthesis system, which is prepared from wheat embryos, is highly efficient and robust (Madin et al. 2000). The wheat germ-based cell-free system has several advantages such as low cost, capacity to synthesize large proteins and easy availability in large amounts. Furthermore, it is more suitable for synthesizing eukaryotic proteins, as it is a eukaryotic system. However, a study has shown that this system is not generally stable and thus insufficient to synthesize proteins because of the presence of endogenous ribosome-inactivating proteins (RIPs). RIPs are N-glycosidases and can inhibit protein synthesis by removing a single adenine residue present in a stem-loop structure of 28S ribosomal RNA through the cleavage of the N-glycosidic bond (Stirpe et al. 1992). Most commonly RIPs are single-chain proteins (Type I RIPs) such as tritin, colocin 1/2, mapalmin and asparin 1/2 (Bolognesi et al. 1990; Madin et al. 2000). Therefore, the elimination of these translation inhibitors by extensive washing gives rise to wheat germ extracts that contain stable and highly active translational apparatuses. With such an extract, translation proceeded for more than 60 hours when performed in a dialysis bag, which enables the continuous supply of substrates as well as the continuous removal of small byproducts. It has been claimed that 1-4 milligram (mg) of enzymatically

active proteins, for example, 1.2 mg of green fluoresce protein (45 kDa), and 0.6 mg of a 126 kilodalton (kDa) tobacco mosaic virus protein per millilitre of reaction volume were produced (Madin et al. 2000).

Electrophoretic mobility-shift assay (EMSA) is a commonly used *in vitro* assay to detect protein-DNA interactions. When the target protein is bound to ^{32}P radioactively labelled probes (TF-associated DNA binding motifs or DNA fragments that contain TF-associated DNA binding motifs), a protein-DNA complex forms, thus it migrates more slowly than the free probes (unbound ^{32}P radioactively labelled TF-associated DNA binding motifs or DNA fragments that contain TF-associated DNA binding motifs) (Carey et al. 2012; Garner and Revzin 1981). This technique was therefore chosen to detect interactions between the promoters of five paclitaxel biosynthesis pathway genes and 19 *Taxus* TF candidate proteins. Prior to EMSA, 19 *Taxus* TF candidate proteins need to be synthesized. *In vivo* protein expression in *E. coli* was initially chosen. However, the synthesized proteins could be not detected. This could be due to the fact that expression of these *Taxus* proteins interferes with the physiology of the *E. coli* host, and therefore insoluble inclusion bodies form (Chrunyk et al. 1993). In addition, it is also likely that because of significant levels of mRNA degradation by endogenous *E. coli* ribonucleases, large mRNAs are more likely to be degraded than small ones, and therefore this system is not suitable for producing proteins translated from large mRNAs (Madin et al. 2000). The molecular weights (MW) of these 19 *Taxus* TFs (SI Table 2) range from ~8 to 51 kDa. 10 of them have MW greater than 30 kDa, and the rest has MW below 20 kDa. So, these proteins (MW>30 kDa) might not be synthesized in the *E. coli* host because of their corresponding mRNA degradation. As a consequence, wheat germ cell-free protein synthesis was deployed to synthesize these 19 *Taxus* TF candidate proteins because of its numerous advantages.

The *in vitro* synthesis of *Taxus* TF candidate proteins started with first round PCR, in which linker sequences were first added to both 5' and 3' end of the open reading frame (ORF) of these TFs. In total, 18 out of 19 TF encoding sequences each were

amplified with linker sequences added except contig 07425. This exception could be caused by technical problems such as PCR reaction conditions (Fig 7.1 and 7.2). The DNA sequences of the T7 promoter and FLAG-tag were subsequently added to the 5' end of the first round PCR products through the linker sequences in the second round PCR (Fig 7.2). FLAG-tag is a polypeptide protein tag and has an amino acid sequence of DYKDDDDK. It can be added to a target protein using recombinant DNA technology. This allows one to detect the FLAG-tagged protein with an anti-FLAG antibody when there is no antibody against the target protein. Then, the second round PCR products were used to generate mRNA by *in vitro* transcription, from which Flag tagged TF proteins were produced by *in vitro* translation in wheat germ extract (Dr. Yasuomi Tada at Kagawa University, Japan). In total, 15 proteins were produced, including 14 *Taxus* TF candidates and the ORA47 (Fig 7.3). The synthesis of the other 4 candidates (contig 07425, 05638, 15401 and 22386) was successfully repeated (Rabia Amir and Gary Loake, unpublished data).

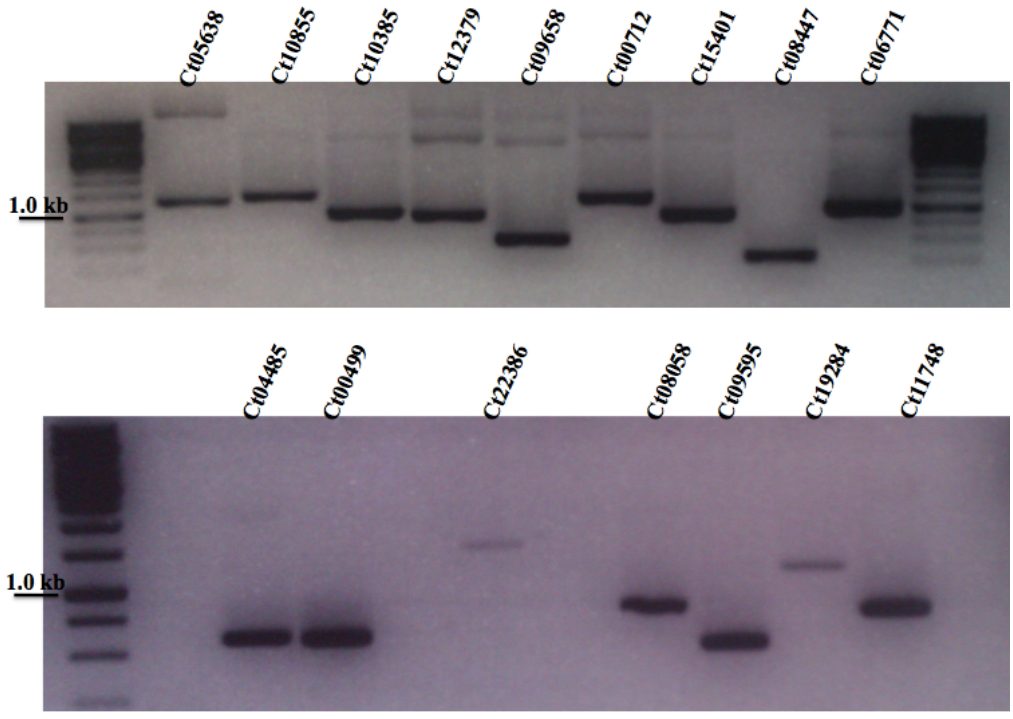


Fig 7.1 First round PCR in the process of synthesizing TFs by wheat germ cell-free system. In total, 19 (16 shown in Fig 7.1 and 3 shown in Fig 7.2) out of 20 TF encoding sequences were amplified with linker sequences added on both 5' and 3' end of the ORF of each TF. These 19 genes include 18 *Taxus* TF candidates and the *ORA47* that encodes an AP2/ERF TF. The exception was contig 07425 (Fig 7.2).

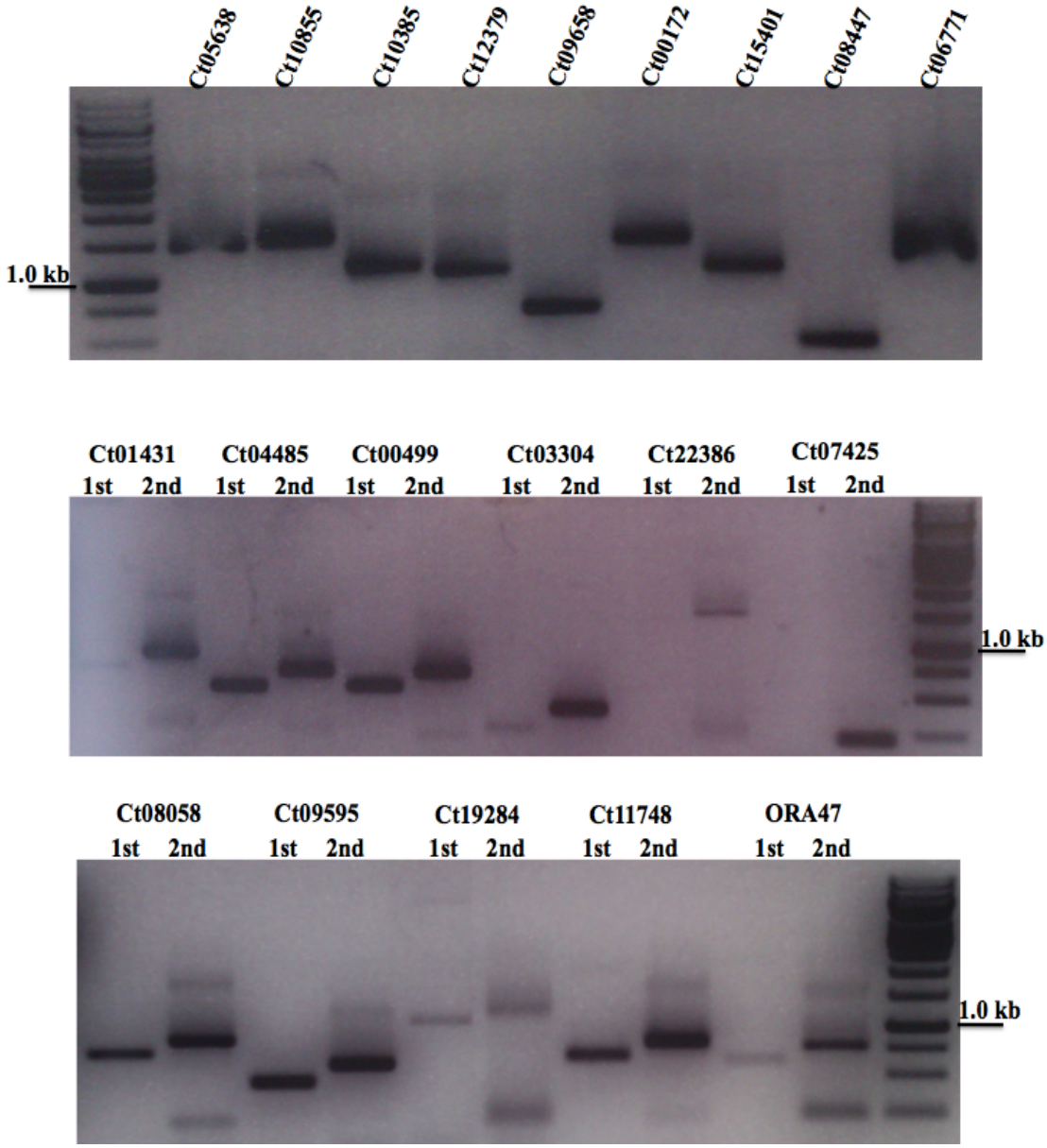


Fig 7.2 Second round PCR in the process of synthesizing TFs by wheat germ cell-free system. Based on the first round PCR products, the second round PCR products were subsequently amplified with the T7 promoter and Flag tag added to the 5' end the through the linker sequences. In total, all 19 TF encoding sequences from the first round PCR were amplified, including 18 *Taxus* TF candidates and the *ORA47* that encodes an AP2/ERF TF.

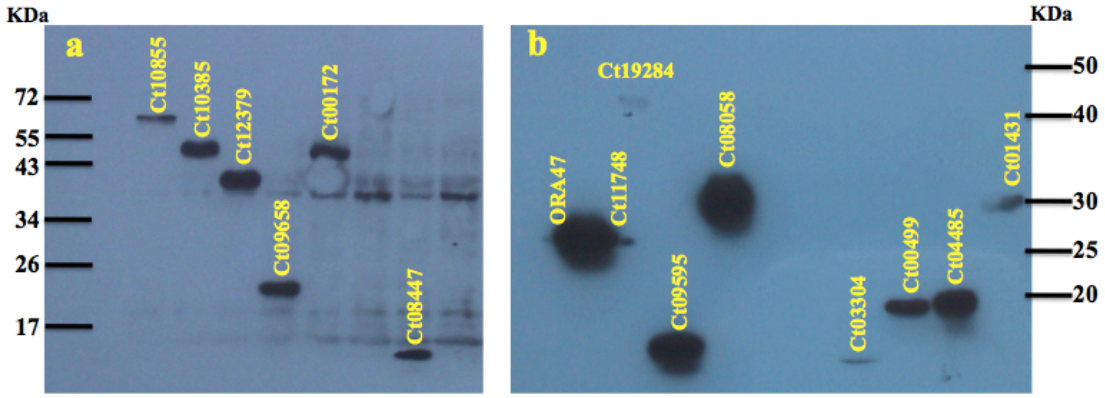


Fig 7.3 Western blots of synthesized FLAG-tagged *Taxus* TF candidate proteins and ORA47 using wheat germ cell-free system. In total, 14 FLAG-tagged *Taxus* TF candidates and ORA47 were synthesized. The primary antibody against 15 FLAG-tagged proteins was anti-FLAG in 1 in 5000 dilution. The secondary antibody agaisnt the primary antibody was anti-mouse (IgG) in 1 in 1000 dilution. The exposure time was 10 seconds.

In conclusion, because these 19 *Taxus* TF candidate proteins might interfere with *E. coli* physiology, it was difficult to express these proteins and the amount of recombinant proteins could not even be visualized (Rabia et al. unpublished). In contrast, the *in vitro* protein synthesis by wheat germ is a simple, quick and efficent way to synthesize these *Taxus* TF proteins. As shown in Fig 7.4, 14 putative candidate *Taxus* TF proteins and one AP2/ERF TF (ORA47) were produced using this system and visualized. Because EMSA only requires ~2 to 20 µl of crude protein solutions (wheat germ extracts) for detecting protein-DNA binding, in which target

proteins are synthsized, the amount of synthesized proteins in the wheat extract does not need to be determined (Carey et al. 2012). However, it has been experimentally proved that the specific *Taxus* TF candidate protein-DNA interactions were clearly detected using the wheat germ extracts that contain the desired *Taxus* protiens (Rabia Amir and Gary Loake, unpublished data).

Chapter 8 General discussion

8.1 Summary of my project

In this project, Illumina Solexa was first performed to digitally profile differently expressed genes at the early time points from the cDNA library of MeJA-elicitated CMCs. Consequently, this led to the discovery of 1646 significantly differently expressed genes, and further annotation showed that there were 78 MeJA-induced putative TF-encoding genes. Based on the E-values, 19 TF-encoding genes all have high E-values ($<e^{-04}$) when blasted against the non-redundant protein sequences in the vascular database. In addition, they were all significantly induced at 0.5 h time point after MeJA treatment (Fig 3.4 A & B). These 19 genes were therefore selected for further characterisation. Finally, the bioinformatic analysis showed that all 19 TFs contain the conserved signature domains (CSDs) when aligned with their well-known counterparts. Based on the CSDs (Fig 3.4 C to G), these 19 TFs were therefore grouped into 5 different TF families, including AP2/ERF, NAC, MYB, WRKY and bHLH. The promoter analysis by PLACE (http://www.dna.affrc.go.jp/PLACE/) shows that these five TF families-associated DNA-binding motifs are present in the promoters of five paclitaxel biosynthesis pathway genes, including *TASY*, *T5αH*, *DBAT*, *PAM* and *DBTNBT* (Appendix I). All together, it indicates that these 19 TFs could regulate the expression of these five pathway genes as well as other known and unknown pathway genes.

The Y1H assay was firstly deployed to screen interactions between 19 TFs and 5 promoters (95 combinations). The positive control for Y1H showed that ZAT10 interacted with plox3; however, no convincing interactions were detected among 95 TF-promoter combinations. This was most likely caused by some technical issues such as protein expression vectors used to express the recombinant GAL4AD-TF proteins, as only the GAL4AD, but not the recombinant proteins, was detected (Eunjung et al., unpublished data) (chapter 4). Because of the numerous advantages the plant transient expression assay has, such as expression of plant proteins in *planta* system, thus, it was chosen to screen *Taxu*s protein-promoter interactions in the leaves of *Nicotiana benthamiana*. However, the positive control for the assay was

not successfully established. Because the recombinant plasmid, which contains both TF expression cassette and promoter-reporter cassette, was not successfully constructed through Cre/loxP-mediated recombination of two dependent vectors. Evidence showed that the recombination event could happen between two separate loxP-containing vectors (pJET1.2-GW-loxP and pMDC162-GW-35Sm-loxP). However, the probability of selecting a positive colony that contains the recombinant plasmid is very low. This is most likely caused by low recombination efficiency, difficulty in transforming *E. coli* with large vectors (~ 19 kb) and occurrence of reverse recombination event between two *cddB* genes present in each gateway cassette. On the other hand, the interaction between ORA47 (AP2 TF) and its associated GCCGCC binding motif was confirmed through the expression of the *GUS* reporter in the technically improved plant assay (Fig 6.4 A). This indicates that this re-designed plant transient expression assay is therefore validated as a suitable system to detect interactions between plant proteins and promoters.

It would be more informative to validate *Taxus* TF-promoter interactions using an *in planta* system, if a list of interactions between 19 TFs and their family associated binding motifs found in each of five pathway gene promoters is quickly produced and used as a guide. An *in vitro* gel shift assay, which can quickly and efficiently detect interactions between proteins and DNA fragments, was therefore undertaken by a lab colleague, Rabia Amir. As a positive control, ORA47 was first shown to bind to the AP2/ERF associated GCCGCC motif present in plox3. Furthermore, two bHLH TFs (Ct11748 and 08058) were shown to bind to the promoter of *TAS*Y, *PAM*, *DBAT* and *T5αH*. A WRKY TF Ct09595 was shown to bind to all 4 promoters except PAM, because PAM does not contain any WRKY-binding sites. In contrast, another WRKY TF, Ct19284, interacted with DBAT, DBTNBT and T5αH. Finally, three AP2 TFs (Ct03304, 01431 and 04485) were found to bind with the GCCGCC-containing promoters, including TASY and T5αH (Rabia Amir and Gary Loake, unpublished data).

The gel shift assay is not able to distinguish between positive and negative regulators. Therefore, the interactions obtained in a gel shift assay are typically further validated by another *in-vivo* transient expression, for example utilising

protoplasts of *Arabidospsis*. Instead of using TF-associated DNA-binding motifs, whole promoters or multimers of *cis*-elements are used to identify positive transcriptional regulators that activate lucifrase gene expression when binding to their associated DNA-binding motifs present in the promoters (Pauwels et al. 2008). As a positive control, ORA47 was shown to bind the plox3 promoter and activate lucifrase gene expression. This interaction was also confirmed by Goossens group (Pauwels et al. 2008). We have identified a number of positive regulators through this assay, including MYB (Ct10855), AP2 (Ct22386) and bHLH (Ct08058) interacting with the TASY promoter and WRKY (Ct19284) and MYB (Ct15401) interacting with the DBAT promoter (Howat and Gary Loake, unpublished data).

In conclusion, the plant transient expression assay in tobacco leaves was found to be a suitable system to validate plant TF-promoter interactions. In parallel, gel shifting and plant transient expression assays have been shown to work well and are identifying positive transcriptional regulators controlling the expression of five paclitaxel biosynthesis pathway genes.

8.2 CMCs: culturing and potential in plant natural product production

In terms of plant cell culturing at an industrial scale, CMCs have several advantages over DDCs, such as greatly reduced cell aggregate size (<0.5 mm in diameter), relatively fast and constant growth rate and significantly increased production of secondary metabolites. Therefore, this superior cell line could be potentially used for production of medically and economically important natural products. The increased production of paclitaxel from CMCs of *Taxus cuspidata* has shown that the use of CMCs has great potential in natural drug production (Lee et al. 2010). Apart from anticancer drugs (paclitaxel, taxamairin A and taxamairin C) produced from *Taxus cuspidata* CMCs, ginsenosides (F2 and XVII), medically valuable natural products, have also been produced through CMC culturing technology (Yun et al. 2012). Furthermore, CMCs have been isolated from other plant species, including ginseng (*Panax ginseng*), ginkgo (*Ginkgo biloba*) and tomato (*Solanum lycopersicon*). This

further indicates that this plant cell isolating technique could be potentially used to isolate CMCs from a wide range of selected plant species for production of natural products that are used as antimicrobials, perfumes, pigments and insecticides. Overall, culturing of CMCs may therefore provide a cost-effective and sustainable platform for the production of both medically and economically important plant natural products.

8.3 Employment of advanced techniques in my project

In terms of efficiency in time, the use of advanced biological techniques throughout my project has greatly facilitated the profiling of gene expression of MeJA-elicited CMCs, Y1H, gel shift and plant transient expression assay.

8.3.1 Next Generation sequencing

Next Generation sequencing technology was employed to help determine MeJA-induced TFs in *T. cuspidata*. In this context, 454 pyrosequencing sequencing was employed to determine the *T. cuspidata* transcriptome producing a total of 860,800 reads of 351 bp average length (Lee et al. 2010). These reads were assembled into longer reads (isotigs), which were blasted against available sequence databases. In addition, Illumina Solexa digital gene expression (DEG) profiling was used to identify MeJA-elicitated transcripts in CMCs. Thus, 1646 differentially expressed genes were uncovered as being significantly up-regulated (Fig 3.2). After annotation of these 1646 DEGs, 78 contigs were annotated as putative TF encoding genes (Fig 3.3a). Furthermore, according to the E-values ($<e^{-04}$) (Table 8), 19 putative TFs, which are induced at 0.5 h time point by MeJA, were selected and assayed for their interactions with the promoters of five paclitaxel biosynthesis pathway genes. These promoters were analyzed in PLACE and found to be rich in the cognate binding motifs of the 19 putative *Taxus* TFs (Appendix I).

8.3.2 Gateway cloning technology

When screening interactions between 5 *Taxus* promoters and 19 putative TFs, gateway cloning technology was used to clone all 19 TFs into pGADgate424 lapau destination vector (Fig 2.2c) in Y1H assays. Furthermore, in the two-component system cloning, both promoters and TFs were spontaneously cloned into pTEA destination vector through two different gateway cassettes, attR1-cddB-attR2 and attR3-cddB-attR4 (Fig 2.4c). Compared to the conventional cloning technology (ligation-directed cloning), gateway cloning has several advantages such as rapid, directional and efficient cloning and expression of PCR products with a wide rang of size and high percentage of positive colonies. Therefore, the deployment of this technology in Y1H and plant transient expression assays has greatly increased the efficiency of transferring a large number of genes between different vectors. In addition, it is a versatile cloning system; any standard vector can be converted into a gateway-compatible vector through the insertion of gateway cassettes. For example, after inserting a gateway cassette, pGAD424 is converted into pGADgate424 lapau (Fig 2.2c), into which all 19 TF-encoding genes are spontaneously cloned through the gateway cloning system.

8.3.3 *In-vitro* transcription and translation using wheat germ extracts

In order to screen TF protein-DNA interactions by gel shift assay, *Taxus* TF proteins first need to be synthesized. It was found difficult to express these TF proteins in *E. coli*, which is probably due to the fact that these proteins interfere with *E. coli* physiology. In contrast, they were synthesized successfully by a wheat germ-based cell free protein synthesis system (Fig 7.4) (Tada et al. unpublished). This *in-vitro* system is an efficient and cost-effective way to produce enough protein for EMSA. 14 *Taxus* TF candidate proteins were synthesized through this method and they were used to detect TF associated DNA-binding motif interactions in gel shift assays (Amir et al., unpublished).

8.4 Further work

8.4.1 Transcriptional regulation network for paclitaxel synthesis

A gene regulation network (GRN) is the understanding of transcriptional regulatory networks at a genome-wide scale, and it can be constructed by working out interactions between regulators (TFs) and with their targets through different approaches. Y1H assays have been used as an approach to identify upstream TFs that bind to promoters or DNA fragments of interest (target-centered approach) and systematically map GRNs. In this assay, a collection of TF proteins or a cDNA library of TFs (prey proteins) needs to be developed, and it will be assayed against target promoters or DNA fragments of interest (baits). In addition, the gateway-compatible Y1H allows a high-throughput screening of protein-DNA interactions (Deplancke et al. 2006; Hens et al. 2012). This technique could be further enhanced when coupled with a robot, the BioRad VersArray, which is known as enhanced Y1H (eY1H) (Gaudinier et al. 2011). By contrast, Chromatin immunoprecipitation (ChIP) coupled with microarrays (ChIP-chip) or Next-Generation sequencing (ChIP-seq), on the other hand, are TF protein-centered approaches to identify downstream targets of given TFs and subsequently map GRNs. In this approach, together with their bound target DNA (promoters), the given TFs are immunoprecipitated, which is followed by subsequent identification of the bound target DNA. Therefore, ChIP-chip or ChIP-seq can construct a GRN that describes the *in vivo* interactions between given TFs and their targets. However, there are some drawbacks in these two approaches. The in-*planta* TF-DNA binding can be easily missed in yeast system, because of the heterologous nature of Y1H assay. As a result, there will be TF-DNA interactions that cannot be detected in yeast, but can be detected if assayed in *plant*a. In contrast, although the ChIP approach can show *in vivo* planta TF-DNA binding, high-quality antibodies against the TFs are required to perform specific and sufficient immunoprecipitation. In common, both do not provide information about the regulatory effect of TFs after binding with their targets. To circumvent this problem, an *in vivo planta* system is needed to show the effect of TFs upon binding. The transient expression assay performed in either protoplasts or leaves can provide

information about the TF's effects on the expression of target genes (Lau et al. 2011; Yang et al. 2000). Furthermore, comprehensive data on transcript abundance and on TF-DNA interactions could be generated by RNA sequencing (RNA-seq) and ChIP-seq (Ouyang et al. 2009).

So, based on the data from screening interactions between 19 *Taxus* TF candidates and 5 promoters using both gel shift and protoplast transient expression approaches, a GRN for paclitaxel biosynthesis can be mapped. For example, it has been found that AP2 (Contig03304) simultaneously interacts with DBAT, TASY and T5αH promoters in TEA (Howat et al. unpublished). It indicates that this AP2 could act as a master TF that regulates the expression of three paclitaxel biosynthesis pathway genes. However, this mini GNR only shows the regulatory relationship between a limited number of TFs and 5 paclitaxel biosynthesis structural genes, but no information about TF-TF interactions and microRNA-mediated regulation of gene expression. Therefore, a more comprehensive GRN could be constructed, if including more TFs and their regulated paclitaxel biosynthesis pathway genes and combining other approaches, such as Y2H and computationally and experimentally derived microRNA-mRNA interactions (Brady et al. 2011).

8.4.2 Establishment of a superior paclitaxel-producing CMC line

The identification of TFs controlling the expression of paclitaxel biosynthesis genes and further unravelling of a TF regulatory network for paclitaxel biosynthesis would enable bioengineering of CMC to establish a superior paclitaxel-producing line. It has been established that a non-embryonic *Taxus cuspidata* cell line is stably transformed with two *GUS* reporter binary vectors (pCAMBIA1301 and pCAMBIA1305.2) through the *Agrobacterium*-mediated transformation and these two resulting transgenic *Taxus cuspidata* cell lines showed GUS activity. In addition, both transgenic cell lines have been maintained in suspension cultures for more than 20 months without the loss of *GUS* gene expression and antibiotic resistance (Ketchum et al. 2007). The transformation of CMCs with TF expression binary

vectors is currently ongoing. The over-expression of key TFs in CMCs, which control the expression of rate-limiting paclitaxel biosynthesis pathway genes could lead to a significant increase in paclitaxel production. In addition, the silencing of TF-encoding genes, which regulate the expression of branched pathway genes, could further enhance paclitaxel production.

Together with the advantages CMCs have over DDCs in large-scale culturing, the cell culturing of this superior paclitaxel-producing CMC line could significantly enhance paclitaxel production.

8.4.3 Identification of unknown genes involved in paclitaxel metabolism

Based on the assumption that genes sharing a common function or being involved in a given biological process are most likely transcriptionally co-regulated (the 'guilt-by-association' principle), the co-expression approaches aim to construct coexpression networks, from which biologically relevant information can be extracted (Usadel et al. 2009). For example, identification of unknown structural genes involved in various biological processes (Bassel et al. 2011). Ectopic overexpression of a TF in plants is one of the common means to identify the remaining structural genes that constitute metabolic networks. For example, in combination with microarray analysis, a suite of co-regulated genes, which are involved in synthesis and further metabolism of phenylalanine, were identified in transgenic tomato overexpressing a MYB TF, *Petunia hybrida* ODORANT1. Among these co-expressed genes, a gene encoding prephenate aminotransferase was discovered, which converts prephenate to arogenate (Dal et al. 2011). Therefore, after identifying *Taxus* TFs, one or more TFs could be ectopically expressed in CMCs, and subsequently, the unknown structural genes involved in paclitaxel metabolism, including unknown pathway genes and genes involved in paclitaxel transport and storage, could be discovered through the analysis of coregulated genes and further validated by biochemical approaches.

8.4.4 Gene-to-metabolite network

Furthermore, by extending the concept of the 'guilt-by-association' principle, a co-expressed relationship between the pathway genes involved in synthesis of secondary metabolites and secondary metabolite accumulation patterns could also be established, from which, the function of genes involved in secondary metabolic pathways can be decoded, and therefore, a gene-to-metabolite network can be elucidated. For example, through the BL-SOM (Batch-learning self-organizing map) analysis of the combined time-series data of the transcriptome and of the metabolome obtained after sulfur deprivation, which causes changes in gene expression and metabolite accumulation patterns in *Arabidopsis*, two gene clusters for synthesis of glucosinolate and anthocyanin were found, respectively. In addition, a set of genes involved in glucosinolate biosynthesis was found to encode sulfotransferases and R2R3-MYB transcription factors (Hirai et al. 2005; Hirai et al. 2007). Another example of assigning functions to genes through the correlation of transcripts and metabolites is the identification of genes involved in the late step of anthocyanin biosynthesis in the background of MYB-overexpressing transgenic *Arabidopsis,* including genes that encode two flavonoid glysosyltransferases and two acyltransferases (Tohge et al. 2005). In this project, an in-house time-series (0.5, 2 and 12 hour samples) of the transcriptome was obtained from the MeJA-elicitated CMC line, in which, a set of pathway genes involved in the paclitaxel biosynthesis are co-upregulated at all three time points (Fig 3.5, 6 &7). Collaboraions have been established to generate metabolomic data from a MeJA-elicited *T. cuspidata* CMC line using mass spectrometry-based techniques (Kueger et al. 2012) and this will be combined with the in-house transcriptomic data. The function of pathway genes involved in paclitaxel biosynthesis and gene-metabolite correlation can then be predicted by combining both transcriptome and metabolome data of the MeJA-elicited CMC line. The function of identified genes could then be validated using reverse genetics (e.g. RNAi knockout) and biochemical approaches (e.g. analysis of recombinant protein activity). Finally, molecular networks of gene-to-metabolite for

paclitaxel biosynthesis can be constructed through a comprehensive identification of function of paclitaxel pathway genes.

8.4.5 Gene clusters for paclitaxel synthesis

Operon-like gene clusters for secondary metabolic pathways are commonly found in bacteria and filamentous fungi, but not in planta, in which functionally related genes are co-ordinated (Osbourn 2010b). However, recent studies have discovered a small number of plant secondary metabolic gene clusters in different plant species, which are mainly implicated for the biosynthesis of defence compounds, for example, the avenacin clusters in *Avena* spp. (oat) (Qi et al. 2004), the momilactone clusters for the synthesis of diterpenes in *Oryza sativa* (rice) (Sakamoto et al. 2004), the cluster for the synthesis of 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) in maize (Jonczyk et al. 2008) and the cluster for the synthesis of the anticancer alkaloid noscapine (Winzer et al. 2012). In addition, TFs, in turn, are found to be key regulators of plant secondary metabolic gene clusters. For example, OsTGAP1, a basic leucine-zipper TF, is found to positively regulate both momilactone and phytocassane synthesis in rice (Okada et al. 2009). Interestingly, there are more than one TF found to regulate a plant secondary metabolic gene cluster, and the regulatory genes that encode these TFs also tend to cluster together. For example, at least seven ERF (Ethylene Response Factor) encoding genes were found to cluster at the *NIC2* locus (locus *B*), which positively regulates nicotine biosynthesis. Furthermore, these seven genes were experimentally proven to show unequal effectiveness in promoting nicotine biosynthesis (Legg 1971; Shoji et al. 2010). This finding might provide new insights into how regulatory genes for biosynthesis of secondary metabolites are organized, and further shed light on the studies on how a downstream transcriptional network for synthesis of secondary metabolites is effectively activated through the jasmonate signalling pathway (Shoji et al. 2010). In our MeJA-upregulated TF database, more than one member of each TF family shows up-regulation after MeJA elicitation. These families include AP2/ERF (6), MYB (7), NAC (5), bHLH (5), WRKY (3), GRAS (3), E2F (3), LOB (6) and C3H (6) (Fig 3.3b). As these regulatory genes are all activated upon MeJA elicitation, it is likely that these TFs from the same TF family are organized in a clustered manner, which could be tested

using fluorescence in situ hybridisation (FISH) (Trask 1991), and that each of these TFs shows differences in effectiveness in promoting paclitaxel biosynthesis.

Furthermore, paclitaxel is a defence compound against plant pathogenic fungi (Flores 1994). In our MeJA-responsive paclitaxel biosynthetic gene database, genes involved in the early, middle and late paclitaxel biosynthetic pathway and transportation of paclitaxel into vacuoles are all co-upregulated at three different time points (Fig 3.5, 6 & 7). So, there could be a gene cluster for the synthesis of paclitaxel in *Taxus cuspidata.* A striking example to support this hypothesis is the discovery of a 10-gene cluster for synthesis of noscapine, an anticancer alkaloid from *Papaver somniferum* (opium poppy). Unlike paclitaxel, which functions to promote tubulin polymerization, this anticancer drug binds to tubulin and inhibits its polymerization, and thereby arrests cell division and induces apoptosis (Winzer et al. 2012). Furthermore, an upstream regulatory gene cluster could regulate this downstream structural gene cluster for synthesis of paclitaxel, and in turn, the jasmonate signalling pathway could regulate the whole transcriptional cascade for paclitaxel biosynthesis, because the production of paclitaxel is significantly increased in CMC suspension cultures after MeJA treatment (Lee et al. 2010).

Gene clusters of plant secondary metabolites are thought to form through gene duplication events, followed by independent evolutionary events, which leads to subsequent establishment of two dependent clusters for synthesis of two different end products. For example, the Field group proposed that both thalianol and mameral cluster formed after the α whole-genome duplication event within the Brassicales (Chu et al. 2011; Field et al. 2011). There are two primary advantages of gene clustering: one is co-inheritance; the other is co-regulation (Osbourn 2010a; b). Firstly, co-inheritance prevents the disruption of the most effective and beneficial gene sets, such as plant secondary metabolic gene clusters for synthesis of defence compounds. As a result, the pathway genes are maintained in a clustered manner, and subsequently the selection for clustering might be further enhanced. The disruption of secondary metabolic gene clusters can result in not only the loss of the beneficial end products, but also the accumulation of harmful intermediates, such as toxic or

bioactive compounds that have negative effects on plant growth and development (Mylona et al. 2008). Secondly, co-regulation allows the pathway genes to be expressed in a similar pattern because they share similar promoters. In addition, it may facilitate the processing of transcripts, including chromatin remodelling as well as export from nucleus to cytoplasm for protein synthesis (Chu et al. 2011). For example, the histone modification, H3 lysine 27 (H3K27), has been marked in the thalianol cluster for synthesis of triterpenes (Field and Osbourn 2008).

8.5 Obstacles and solutions in my research

The first obstacle in my research was the unsequenced genome of *Taxus* spp. This makes it difficult to carry out the molecular research on unravelling the paclitaxel metabolism with the ultimate goal to improve paclitaxel production. The solution to this was to establish a transcriptomic database of MeJA-elicitated CMCs. We performed Solexa sequencing to digitally profile DEGs from the cDNA library of the MeJA-elicitated CMCs of *Taxus cuspidata*. Meanwhile, 454 pyrosequencing further sequenced the DNA sequences of these DEGs so that we were able to isolate and characterise candidate genes annotated as putative TFs.

The second obstacle is the use of Cre/loxP recombination to create a recombinant binary vector by combining two separate ones for *Agrobacterium*-mediated infiltration in transient expression assays in tobacco leaves. However, it was found that the recombinant plasmid could not be replicated; instead, these two separate vectors independently replicate inside *E. coli.* Therefore, a two-component gateway binary vector was constructed by cloning a TF-expression cassette into another vector that already contains a promoter-reporter cassette in its backbone. Through this method, the pTEA-ORA47-3xGCCGCC-35Sm-GUS was made and tested for *GUS* activity in tobacco leaves.

The third obstacle is that there was no colony growth when introducing a two-component gateway binary vector, which contains two *cddB* genes in each of two gateway cassettes it contains, into either DB3.1 or 2T1 *cddB*-resistant *E. coli.*

However, when replacing one of the *cddB* with a different DNA sequence, such as *ORA47*, the resulting construct was successfully introduced into both strains. This finding is not expected, because a commercial two-component gateway vector (pK7m34GW2-8m21GW3,0) (Fig 2.4), which contains two *cddB* genes, was extracted from 2T1 strain. However, when it was re-introduced into the same strain, no colony grew. As a result, both *cddB* genes had to be replaced by the DNA sequence of TF and promoters, respectively, prior to cloning the *cddB*-containing TF expression cassette into the backbone of *cddB*-containing reporter vector (SF 8). However, this cloning strategy is not what is expected from using two-component gateway cloning. Because both TF-encoding gene and TF-associated promoter sequences are supposed to simultaneously replace each *cddB* gene present in TF expression and promoter-reporter cassette, respectively, in the strategy of two-component gateway cloning. Therefore, the attempted two-component gateway system did not work as planned. In order to avoid the labour-intensive cloning steps, TF-promoter interactions were first screened by gel shift assays (Amir et al. unpublished); then, the TF-promoter pairs that show interactions will be validated by the plant assay.

Finally, there was difficulty in cloning small DNA fragments into large-sized vectors, such as the cloning of loxP (34 bp) into pMDC162 (~13 kb). The solution to this was to amplify a loxP-containing fragment (300 bp) from pUNI 51, and clone this fragment into pMDC163 by conventional cloning (SF 6). Another difficulty was the directional cloning of *35Sm* upstream of the *GUS* reporter in pMDC162, because there is only one site (*Asc* I) available for the insertion of this 51 bp-long fragment. Therefore, two sites (*Spe* I and *Hind* III) were added in the place of *Asc* I, and the amplified *SpeI*-35Sm-*Hind* III was directionally inserted upstream of *GUS* gene (SF 7). In addition, it was also difficult to amply short DNA fragments such as loxP site (34 bp), *35Sm* (51 bp), B1-GCCGCC tandem repeats-B2, B1-*35Sm*-B2 (120 bp) and B1-3xGCCGCC-*35Sm* (~125 bp) by PCR, because of difficulties in designing primers for such short DNA sequences (loxP) and the absence of the DNA templates for PCR amplification of 35Sm, B1-GCCGCC tandem repeats-B2, B1-*35Sm*-B2 and

B1-3xGCCGCC-*35Sm*. They were synthesized by directly annealing both forward and reverse primers (Fig SF 1, 2 & 3).

8.6 Overview of paclitaxel project

My project can be summarized in Fig 8.1 Our current aim is to identify TFs controlling the expression of paclitaxel biosynthesis pathway genes among the 19 putative *Taxus* TF candidates and produce high paclitaxel-producing transgenic *Taxus cuspidata* CMC lines. Further work could be the elucidation of a gene-to-metabolite network of paclitaxel and transcription regulatory network of paclitaxel biosynthesis and the discovery of unknown paclitaxel pathway genes and a gene cluster for paclitaxel synthesis.

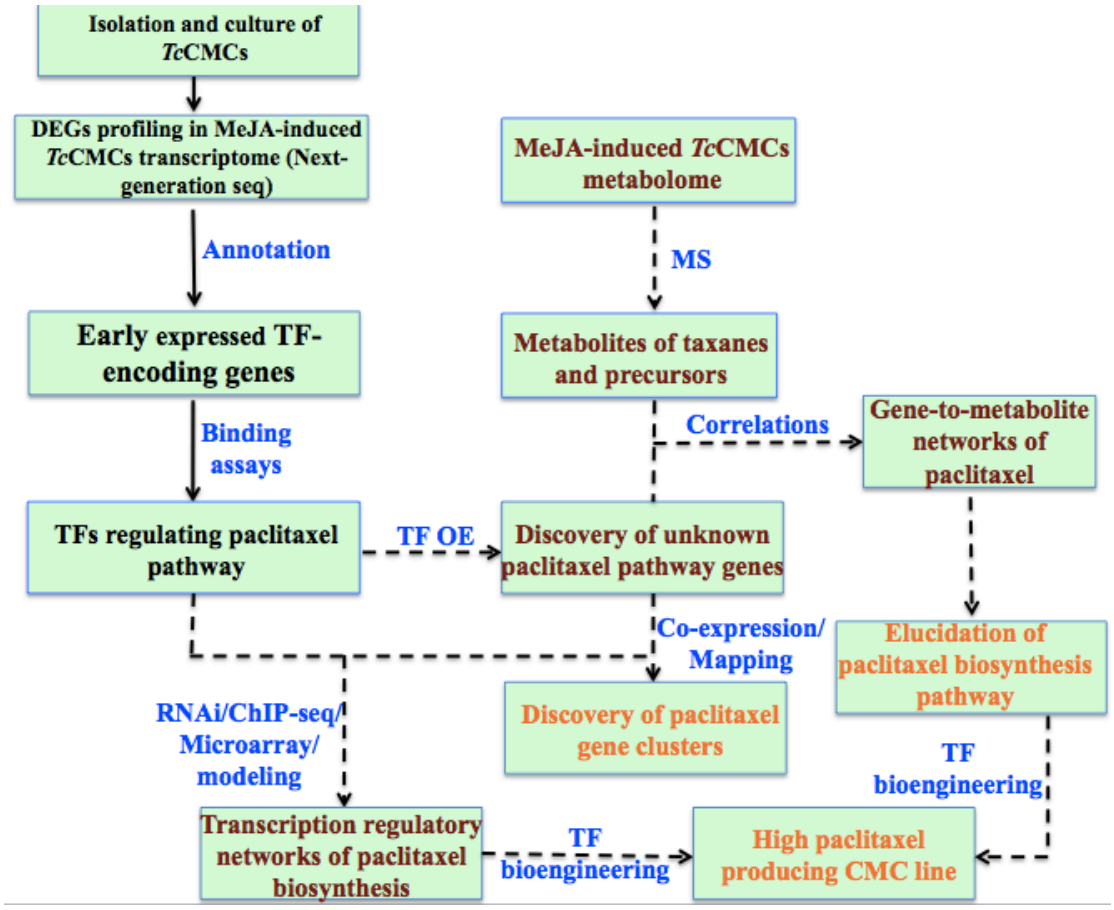
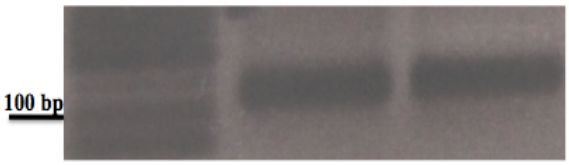


Fig 8.1 An overview of paclitaxel project. The isolation and culture of *Taxus cuspidata* CMCs (*Tc*CMCs) provide a platform for industrial paclitaxel production. After profiling differentially expressed genes (DEGs) in MeJA-induced *Tc*CMCs using Next Generation sequencing technology, early induced putative TF encoding genes are annotated. Then, different binding assays, such as gel shift assay, Y1H and plant transient expression assays, are used to verify interactions between these *Taxus* TFs and promoters of paclitaxel biosynthesis genes (solid black arrows). Provided TFs that regulate the paclitaxel biosynthetic pathway have been identified, overexpression (OE) of key regulators would lead to the identification of previously unknown pathway genes. Meanwhile, if the metabolome of MeJA induced *Tc*CMCs could have been profiled using mass-spectrometry (MS), taxanes and their precursors could be identified. Therefore, this could lead to a construction of gene-to-metabolite networks by the concept of the 'guilt-by-association' principle, and ultimately the elucidation of paclitaxel biosynthesis pathway. On the other hand, gene clusters for paclitaxel synthesis could be discovered through mapping co-expressed pathway genes on the genome of *Tc*CMCs. Finally, based on the discovery of TFs and pathway genes, a transcription regulatory network of paclitaxel biosynthesis could be constructed using a combination of techniques, including RNAi, microarray, chromatin immunoprecipitation sequencing (ChIP-seq) and computational modelling. This gene transcriptional network not only shows both TF-TF and TF-structural gene interactions, but also indicates the hierarchical relationships between regulatory genes (TF-TFs) and between TFs and structural genes. Therefore, *Tc*CMCs could be engineered to over-express the most centred TFs that regulate the entire paclitaxel biosynthesis pathway. In addition, paclitaxel production could be further enhanced by silencing TFs that control branched pathways. In other words, all the branched pathways could be re-directed toward paclitaxel synthesis. Solid black arrows indicate work that has been done or in the process. Dotted black arrows indicate further work.

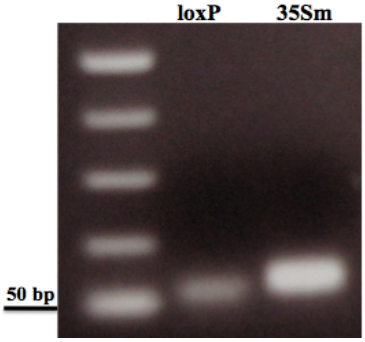
Chapter 9 Supplementary Information

SI. 1 PCR-based synthesis of short DNA fragments

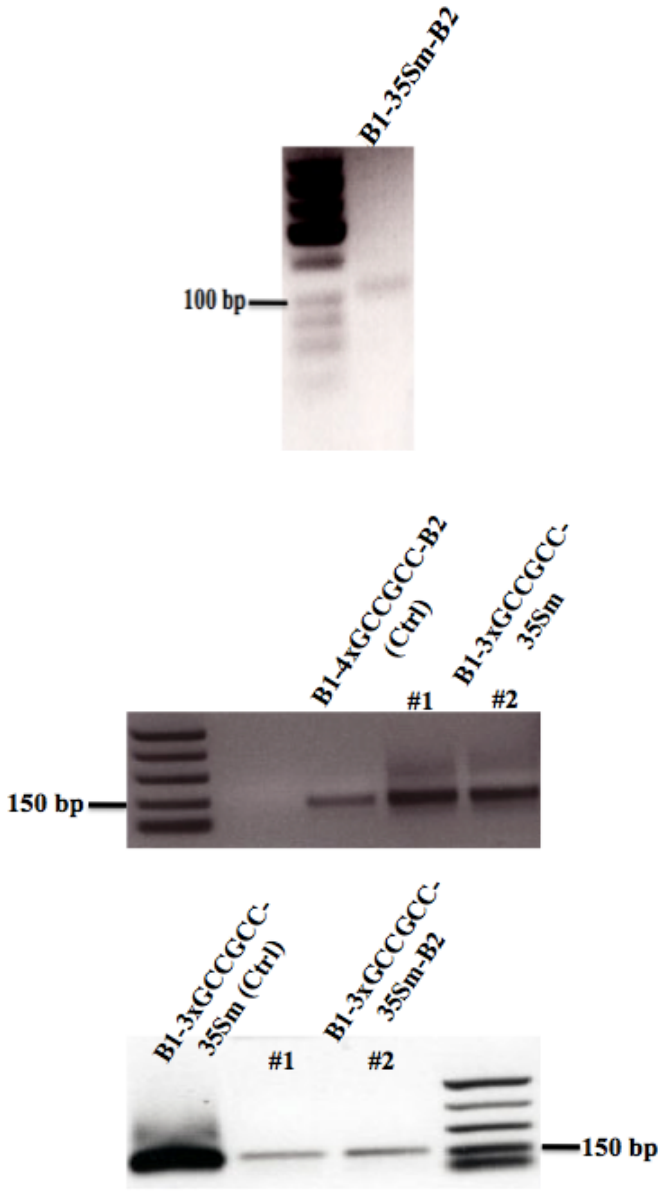
It was difficult to amplify short DNA fragments, including loxP (34 bp), cauliflower mosaic virus (CaMV) *35S* minimal promoter (*35Sm*) (51 bp), attB1-*35Sm*-attB2 (~100 bp), attB sited -3xGCCGCC tandem repeats (~100 bp) and attB1-3xGCCGCC-*35Sm* (~120 bp). Because the DNA templates were not available, and it was difficult to design amplifying primers for such short DNA sequences. Therefore, these fragments were directly synthesized by annealing both forward and reverse primers that have overlapping sequences (SI table 3) (SF 1, 2 & 3).



SF 1 PCR-based synthesis of GCCGCC tandem repeats flanked by attB sites. The PCR products were purified from agarose gel and cloned into sequencing vector (pJET1.2) for sequencing. Each GCCGCC repeat is 17-bp long, including 6 and 5 bases on the left-hand side and on the right-hand side of GCCGCC core, respectively. Both attB1 and attB2 site contain 25 base pairs.



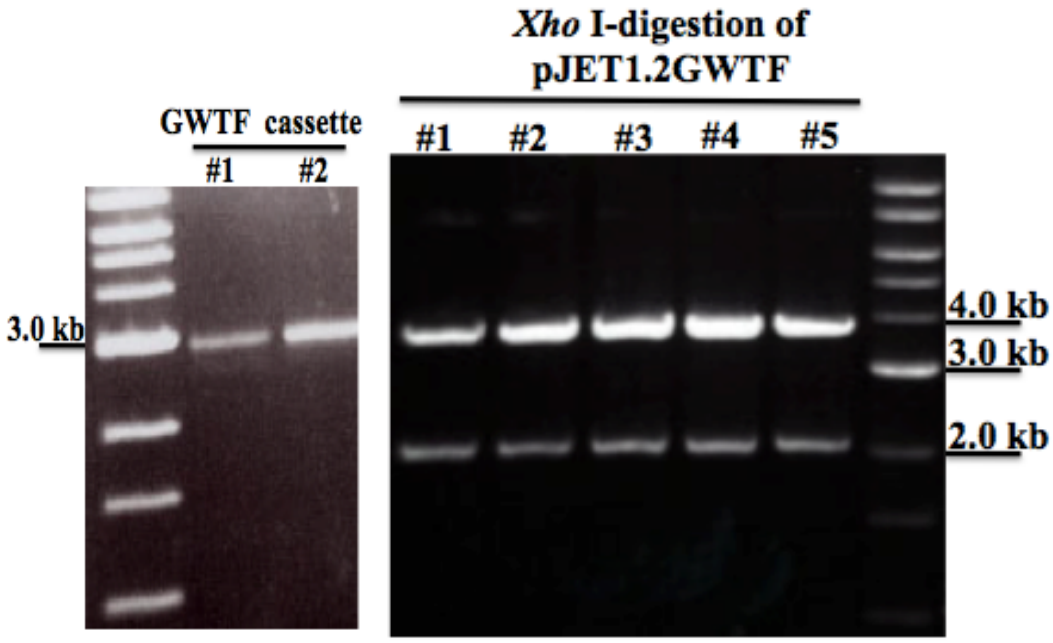
SF 2 PCR-based synthesis of loxP 35S minimal (*35Sm*) promoter shown on 1.8% agarose gel. The loxP site and 35Sm is 34-bp and 51-bp long, respectively. *35Sm*: cauliflower mosaic virus (CaMV) *35S* minimal promoter.



SF 3 PCR-based synthesis of attB1-35Sm-attB2 and attB1-3xGCCGCC-*35Sm*-attB2 shown on 1.8% agarose gel. As expected, the synthesized B1-*35Sm*-B2 is about 120 bp long (Upper panel). In the synthesis of B1-3xGCCGCC-35Sm-B2, the fragment B1-3xGCCGCC-*35Sm* was firstly synthesized by primer annealing. Then, B1-3xGCCGCC-*35Sm*-B2 was synthesized using the B1-3xGCCGCC-*35Sm* as templates (Middle and lower panel). *35Sm*: cauliflower mosaic virus (CaMV) *35S* minimal promoter.

SI. 2 Construction of pJET1.2GWTF expression vector

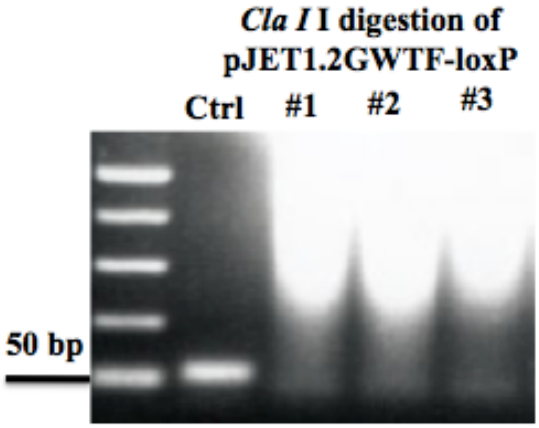
A gateway-adapted TF expression cassette was amplified from pMDC32 and clone into pJET1.2 by blunt-end ligation to form pJET1.2GWTF. *Xho* I digestion further confirmed its insertion into pJET1.2 backbone (SF 4).



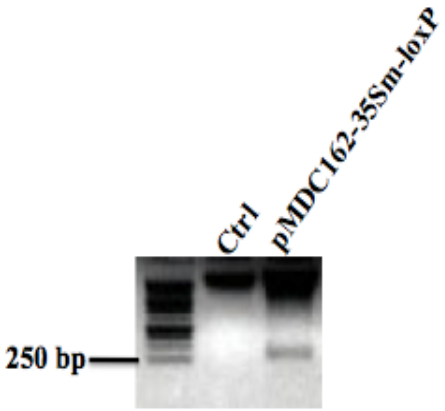
SF 4 Construction of pJET1.2GWTF expression vector. Left panel: PCR amplification of gateway-adapted TF expression cassette (2x35S-R1-cddB-R2-nosT); right panel: *Xho* I digestion of pJET1.2GWTF. As expected, it is cut into two *Xho* I fragments: one is ~3.8 kb; the other is ~2.1 kb.

SI. 3 Cloning of loxP site and *35S* minimal promoter

In order to combine pJET1.2GWTF and pMDC162 through Cre/loxP recombination, a loxP site was cloned into each of the backbones. The loxP was inserted into pJET1.2GW at *Cla* I. However, it was difficult to clone loxP into pMDC162. This might be due to the fact that it is not easy to clone a very small fragment (34 bp) into a large backbone (~13 kb), and the transformation efficiency of large vectors is relatively low. As a result, a ~300 bp loxP-containing fragment was amplified from pUNI 51 and subsequently cloned into pMDC162 backbone at *Pac* I site.

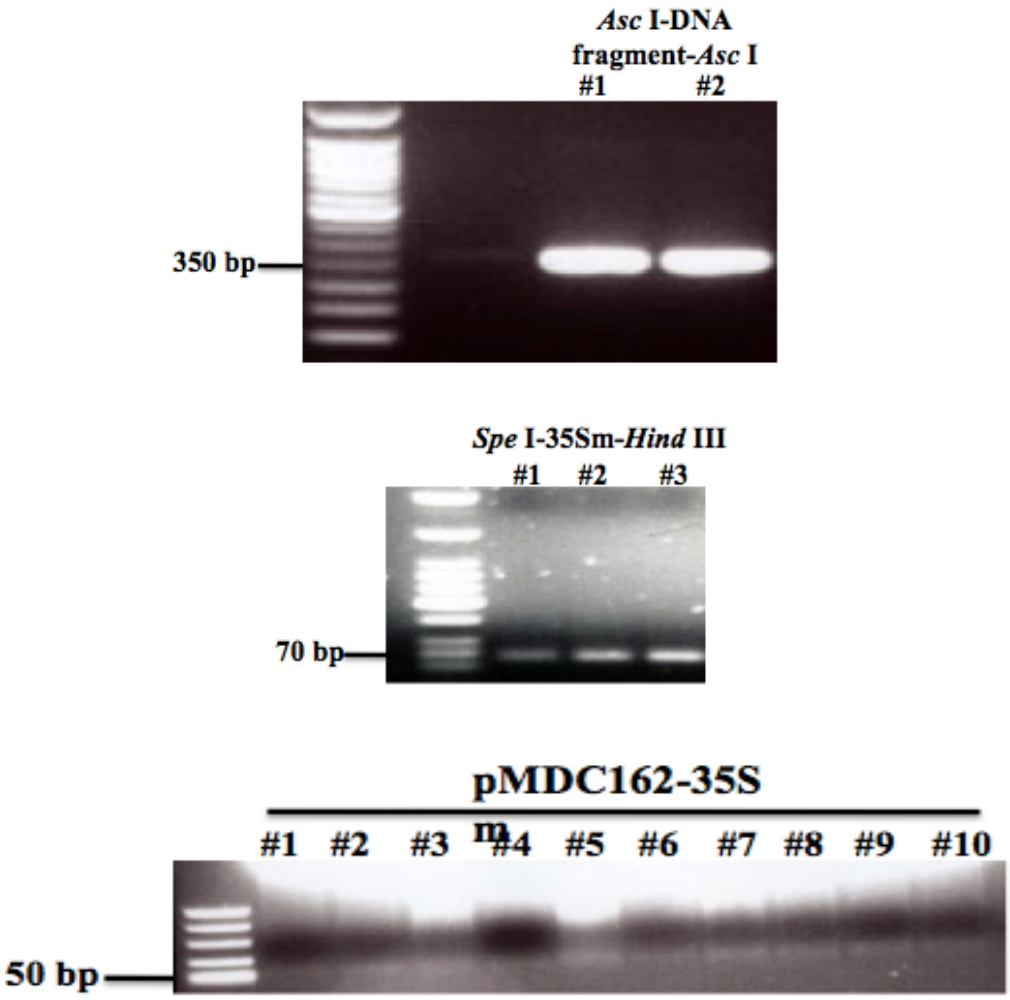


SF 5 Cloning of loxP site into pJET1.2GWTF at *Cla* I site. *Cla* I digestion of Miniprep products of pJET1.2GWTF-loxP showed the bands at the expected size (~34 bp) shown on 1.8% agarose gel. Ctrl: PCR product of synthesized loxP was used as the control.



SF 6 Cloning of loxP site-containing fragment into pMDC162-*35Sm* at *Pac* I site. The loxP-containing fragment (~300 bp) was amplified from pUNI 51 (Upper panel); *Pac* I-digestion of Miniprep products of pMDC162-*35Sm*-loxP showed the bands at the expected size (lower panel). Ctrl: un-cut plasmids. *35Sm*: cauliflower mosaic virus (CaMV) *35S* minimal promoter.

In order to initiate *GUS* expression in the case of using AP2-associated TF DNA binding motifs (GCCGCC), the *35Sm* must be directionally cloned into pMDC162 backbone. However, there was only one site (*Asc* I) available for the insertion of *35Sm* upstream of *GUS*. As a result, two unique sites (*Spe* I and *Hind* III) were introduced by cloning ~300 bp fragment into pMDC162 at *Asc* I site, which contains both sites. Then, *35Sm* was directionally inserted upstream of *GUS* at these two sites in pMDC162 (SF 7).



SF 7 Directional Cloning of *35Sm* into pMDC162. A DNA fragment, which contains two unique sites (*Spe* I and *Hind* III), was amplified and subsequently cloned into pMDC162-loxP at *Asc* I site (Upper panel); *Spe* I-*35Sm*-*Hind* III was synthesized (Middle panel) and cloned into *Spe* I and *Hind* III-linearized pMDC162-loxP, which is further proved by *Spe* I and *Hind* III double-digestion of Miniprep products of pMDC162-*35Sm* shown on 1.8% agarose gel (lower panel). *35Sm*: cauliflower mosaic virus (CaMV) *35S* minimal promoter.

SI. 4 The correlation between plasmid size and *E. coli* transformation efficiency

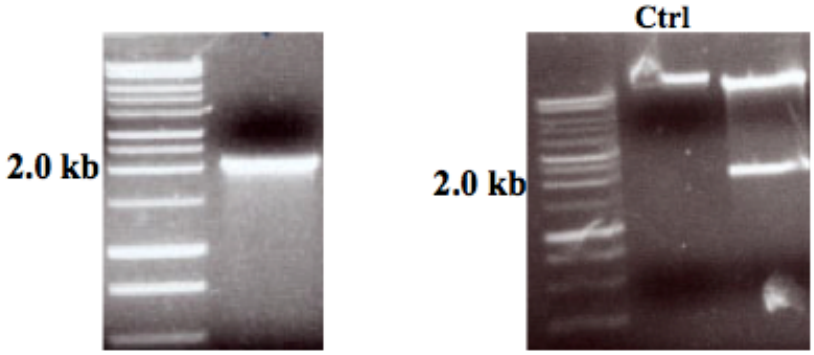
In order to test the hypothesis that the plasmid size affects transformation efficiency of *E. coli*, a large and small sized plasmid were introduced into E. coli (XL1-blue) by transformation. The transformation of pMDC162 (~13 kb), which is more than 4 times larger than pJET1.2 (~3 kb), showed much less number of colony, when compared to the number of colony in the transformation of pJET1.2 (SI table 1). So, the larger the plasmid is, the less efficient its transformation with *E. coli* becomes.

SI table 1 *E. coli* transformation efficiency. The transformation for each plasmid was repeated three times.

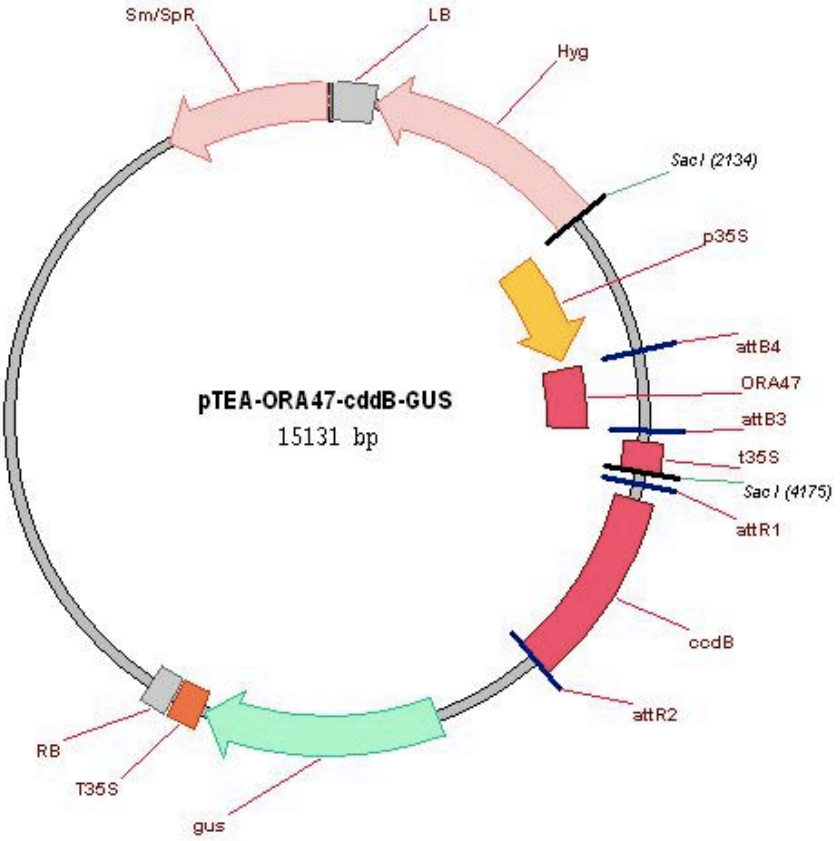
Plasmids	Size (kb)	Selection markers	Number of colony
pJET1.2	~3.0	ampicillin	Full of plates
pMDC162	~13.0	kanamycin	<100

SI. 5 Construction of pTEA-ORA47-3xGCCGCC-*35Sm-GUS* for the positive control of transient expression assay in tobacco leaves

The ORA47 expression cassette (p35S-*ORA47*-t35S) was firstly amplified and ligated into the backbone of pHGWFS7,0 at *Sac* I site to give rise to pTEA-ORA47-*cddB*-*GUS* (SF 8 & 9). Then, the attB-sited 3xGCCGCC-35Sm fragment was cloned upstream of *GUS* through the gateway cloning.



SF 8 Construction of pTEA-ORA47-3xGCCGCC-35Sm-*GUS*. Left panel: PCR amplification of p35S-ORA47-t35S; right panel: *Sac* I digestion of pTEA-ORA47-3xGCCGCC-*35Sm-GUS*. *35Sm*: cauliflower mosaic virus (CaMV) *35S* minimal promoter.



SF 9 The pTEA-ORA47-*cddB-GUS* construct. The p35S-ORA47-t35s cassette was inserted at *Sac* I site. *cddB* is replaced by 3xGCCGCC-35Sm through gateway cloning at the attR1 and ttR2 site. *ORA47* encodes a AP2/ERF TF. p35S: cauliflower mosaic virus (CaMV) *35S* promoter. T35S: CaMV *35S* terminator. *GUS*: beta-glucuronidase. Sp^R: spectinomycin resistance. Hyg^R: hygromycin resistance. LB: left border. RB: right border.

SI. 6 Five paclitaxel biosynthetic gene promoter sequences amplified from our *Taxus cuspidata* CMCs

Five paclitaxel biosynthesis gene promoters were amplified from the genomic DNA of our *Taxus cuspidata* CMCs, and sequenced in the GenePool centre at Kings' Buildings. There were three replicates sequenced for each promoter. After comparing the sequences of each promoter with the corresponding one published on NCBI (http://www.ncbi.nlm.nih.gov/nuccore/?term=taxus+cuspidata), some minor differences in the promoter length were found between our and on-line database except the PAM promoter.

SI table 2 Minor differences in promoter length between our database and NCBI

Promoters	**Promoter length (bp)**	
	Our database	**NCBI**
PAM	257	257
TASY	1408	1405
T5αH	1433	1436
DBAT	1611	1610
DBTNBT	1898	1891

SI.7 The common feature of the C-terminal region of NAC proteins

The C-terminal regions are highly variable in sequence. They are characterised by the frequent appearance of simple amino acid repeats and contain regions that are rich in serine (S) and threonine (T), proline (P) and glutamine (Q), or acidic residues (Duval et al. 2002; Hegedus et al. 2003; Kikuchi et al. 2000) (SF 10 a & b).

SF 10a Sequence alignment of the C-terminal regions of NACs between well-characterised (ANAC019, *Os*NAC6, NAM and *At*NAC1) and putative *Taxus* NACs (Contig 00172, 08447, 10855 and 05638). The C-terminal regions are highly variable in sequences. Regions that that are rich in serine (S) and threonine (T), proline (P) and glutamine (Q), or acidic residues such as aspartic acid (D), asparagine (N) and glutamic acid (E), are underlined in blue.

SF 10b Sequence alignment of the C-terminal regions of NACs between well-characterised (ANAC019, *Os*NAC6, NAM and *At*NAC1) and putative *Taxus* NACs (Contig 09658). The C-terminal regions are highly variable in sequences. Regions that that are rich in serine (S) and threonine (T), proline (P) and glutamine (Q), or acidic residues such as aspartic acid (D), asparagine (N) and glutamic acid (E), are underlined in blue.

SI table 3 Molecular characterisation of 19 MeJA induced putative *Taxus* transcription factors.

Contig number	TF class	ORF (bp)	Protein (aa)	Protein MW (kDa)
03304	AP2	210	69	7.89
04485	AP2	537	178	19.87
07245	AP2	375	124	14.69
01431	AP2	759	252	28.65
00499	AP2	528	175	19.75
22386	AP2	1392	463	50.85
12379	MYB	867	288	33.64
10385	MYB	936	311	35.21
10855	MYB	1332	443	49.36
15401	MYB	870	289	32.7
06771	NAC	963	320	36.5
09658	NAC	501	166	18.42
00172	NAC	1233	410	45.85
05638	NAC	1260	419	47.22
08447	NAC	1140	379	9.14
11748	bHLH	630	209	23.42
08058	bHLH	678	225	25.01
09595	WRKY	801	266	15.38
19284	WRKY	1029	341	38.47

The open reading frames (ORFs) were determined using Expert Protein Analysis System (ExPASy) (http://web.expasy.org/translate/). Then, according to the amino acis sequences, the molecular weight (MW) of each protein was calculated using ProtCalc in JustBio (http://www.justbio.com/index.php?page=protcalc). These proteins were classified using a specialized BALST tool that identifies the conserved domains (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

SI table 4 Primers used in PCR amplification of DNA fragments

Primers used to amplify paclitaxel biosynthesis gene promoters	
PAM_F	5’TATACAATTGCTACCTAAACAGACAAGACAACGA
PAM_R	5’TGTATCTAGATGCAGAGCAAGGAAAATAACTTTAAAA
TASY_F	5’ATTAGAATTCAACTCGCAATAGCATGGACATCTT
TASY_R	5’CGTATCTAGATTCTGCAGAGAGGCAGGGGAACTA
DBTNBT_F	5’AATT CAATTGCTCACTTGGCCTCAA
DBTNBT_R	5’TTAATCTAGACAGCTGGATTATTCACG
DBAT_F	5’TATAGAGCTCTCCCACAAAACTAGACAAGTCATT
DBAT_R	5’TGTAACTAGTCTCAGATCAAGCTCAGAAAATCTT
T5αH_F	5’TTAAGAATTCCTTGTCAGAAGCAGCTCCCTCTCAATAAG
T5αH_R	5’AATTTCTAGAAAGAAACCAAAAGAAAGAGAACAGAGAACAGAGA
attB sites used in gateway cloning	
attB1	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTTC
attB2	5’GGGGACCACTTTGTACAAGAAAGCTGGGTC
attB3	5’GGGGCAACTTTGTATAATAAAGTTG
attB4	5’GGGGCAACTTTGTATAGAAAAGTTGGAAGGAGATAGAACC
Primers used to amplify attB-sited transcription factors	
Ct01431_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGGATCACCAGCATATGAGAGTCA
Ct01431_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAAATGTGCTTCTCGTGAAGGATACTC
Ct019284_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGACCAGCACAAAATTGAGCTTTTG
Ct019284_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTTTGGGAATAACTCTTGTGTATTTCTGATAAAA
Ct08447_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGTGAAAGGAAATGCCGAAGC
Ct08447_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATGTTACAAATGATAAATTTTCATATAAACCAGA
Ct10855_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTGTCAAACACAAGAGGAAGGAGCC
Ct10855_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATGTAGTTCCCACTCCCAGAAAATCA

Ct08058_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAGAACTGAAGATACCCTTCATT
Ct08058_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATTTAGCAATCACTGTCTGAATC
Ct00172_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGGAGAGAAGGACCTC
Ct00172_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTGTGGTCTAGGTTGCAGA
Ct15041_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAGCTCCAAGTGGTTGAATTTGA
Ct15041_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTCTAGTATTGGCTGTTGCATTGGAAT
Ct07425_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCGCGTCGTCGCAGTTGGGTAA
Ct07425_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGTCAGTCCACAGCGAGTACAAT
Ct00499_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCGTCGACAATGGAAAGGTTAG
Ct00499_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATTGCCCTTCTTTTTCTTCTTCCTCC
Ct10385_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGGCGAGCTCCGTGCTGTGATA
Ct10385_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACAACGTAGGCGATGATGCGGCT
Ct04485_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAAATATCATGCAAAGGACAGAGCG
Ct04485_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGTATTCCTCAAATAGTTCCGACTCCCACA
Ct22386_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCCGGCCAGGAAGAGGATCCAAA
Ct22386_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAGAGAAATGATCAGAAGTAGGAGAAGGG
Ct11748_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGTTGGCAGGTGACCAGGGTT
Ct11748_R	5'GGGGACCACTTTGTCAAGAAAGCTGGGTCTTACTCACTCAAAATGGTTGCCCCTAAATTC
Ct06771_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGTAAAGATGGTCTAGAAGGGAAGG
Ct06771_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACTGGAAAATCATATCAGACGAGTGG
Ct05638_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGTCCCTGGCGGTGAATGTGA
Ct05638_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAATAGCCAGACCACAAACAGTCTGGC
Ct09595_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAGTGCACATAGTAGAACAAAGACAAC

Ct09595_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATGAATCTTGAGTCAGAAACCCTTG
Ct09658_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCCGGTGGCTCACAAACTGTGG
Ct09658_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAATTGAGGGTAATACTTGTTGTTGTTGTT
Ct03304_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTTCCATGACGAAGGGATTATT
Ct03304_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGCATCCAAGTGTAGGGTGGTG
Ct12379_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTCCAATCCCCAGGCATTCTT
Ct12379_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATAGACTTTCCCAGAAGTCACATAAATCTG
ORA47_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATGGTGAAGCAAGCGATGAAGG
ORA47_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAAAATCCCAAAGAATCAAAGAT
ZAT10_F	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTAACCATGGCGCTCAAGTGGTTGAATTTGA
ZAT10_R	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAAGTTGAAGTTTGACCGGAAAGTC
35Sm_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCTTCGCAAGACCCTTCCTCTATATAAGGAAG
35Sm_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTCTCCTCTCCAAATGAAATGAACTTCCTTATATAGAG
B1-3xGCC GCC-35Sm-B2	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTCACGGGCCGCCCGTCCCACGGGCCGCCCGTCCCACGGGCCGCCCGTCC
	5’TCCTCTCCAAATGAAATGAACTTCCTTATATAGAGGAAGGGTCTTGCGAAGGACGGGCGGCCCGTGGGACGGGCGGCCCGTG
	5’GGGGACCACTTTGTACAAGAAAGCTGGGTTCCTCTCCAAATGAAATGAACAA
Primers used to amplify attB-sited promoters	
TASY_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTAACTCGCAATAGCTAGGACATCTT
TASY_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTTTCTGCAGAGAGGCAGGGGAACTA
PAM_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTCTACCTAAACAGACAAGACAACGA
PAM_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTTGCAGAGCAAGGAAAATAACTTTAAAA
DBAT_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCCACAAAACTAGACAAGTCATT

DBAT_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGATCAAGCTCAGAAAATCTT
T5αH_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTCTTGTCAGAAGCAGCTCCCTCTCAATAAG
T5αH_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTAAGAAACCAAAAGAAAGAGAACAGAGAACAGAGA
DBTNBT_F	5’GGGGACAAGTTTGTACAAAAAAGCAGGCTAGGCTCCGAATTCGCCCTTCAAAGGAGATG
DBTNBT_R	5’GGGGACCACTTTGTACAAGAAAGCTGGGTGACTGGATCAAAGATGAAACGATGGGCTGAG
Primers used to synthesize short DNA fragments	
loxP_F	5’GGGGGATCGATATAACTTCGTATAGCATACATTATA
loxP_R	5’TAAAAATCGATATAACTTCGTATAATGTATGCTATA
35Sm_F	5’AAAAA-ACTAGTCTTCGCAAGACCCTTCCTCTATATAAGGAAG
35Sm_R	5’AAAAAAAGCTTTCCTCTCCAAATGAAATGAACTTCCTTATATAGAG
GCCGCC-TR_F	5’CCATTGGGCCGCCCATGTCCCCATTGGGCCGCCCATGTCCCCATTGGGCCGCCCATGTCCCCATTGGGCCGCCCATGTCCCCATTGGGCCGCCCATGTCC
GCCGCC-TR_R	5’GGTAACCCGGCGGGTACAGGGGTAACCCGGCGGGTACAGGGGTAACCCGGCGGGTACAGGGGTAACCCGGCGGGTACAGGGGTAACCCGGCGGGTACAGG
Primers used in cloning loxP sites	
loxP (51)_F	5’AAAGGGTTAATTAAACAATTAACCCTCACTAAAGGG
loxP (51)_R	**5’** TTTAAAGGTACCCCTTCCAGGGTCAAGGAAGG
Directional 35Sm_F	**5’** AAAAAGGCGCGCCACTAGT-CTGCAGGTCGACCATATGG
Directional 35Sm_R	**5’** AAAAAGGCGCGCCAAGCTT-AAGAGCGCCCAATACGCAAA
Primers used in cloning TF expression cassette into pHGWFS7,0 backbone	
F_Sac I (47)	5’ ATTATGAGCTCCCCGGGGATCCTCTAGAGG
R_Sac I (47)	5’ ATAATGAGCTCTCCCATATGGTCGAC
Primers used in sequencing	
loxP in pJET1.2GWT	5’ATCTTTCTAGAAGATCTCCTAC

F_F	
loxP in pJET1.2GWT F_R	5’TCATGAGAGTCGATTGCCAAG
loxP in pMDC 162_F	5‘CTATTACGCCAGCTGGCGAA
loxP in pMDC 162_R	5’GCCATAGTGACTGGATATGTTG
35Sm in pMDC 162-loxP_F	5’CATTTTACGTTTCTCGTTCAGC
35Sm in pMDC 162-loxP_R	5’CCACAGGCCGTCGAGTTTTT
pJET1.2 sequencing vector_F	5’ CGACTCACTATAGGGAGAGCGGC
pJET1.2 sequencing vector_R	5’ AAGAACATCGATTTTCCATGGCAG

Bibliography:

Acosta IF, Laparra H, Romero SP, Schmelz E, Hamberg M, Mottinger JP, Moreno MA, Dellaporta SL (2009) Tasselseed1 is a lipoxygenase affecting jasmonic acid signaling in sex determination of maize. Science 323: 262-265

Aerts RJ, Gisi D, Carolis ED, Luca VD, Baumann TW (1994) Methyl jasmonate vapor increases the developmentally controlled synthesis of alkloids in *Catharathus* and *Cinchona seedlings*. Plant J 5: 635-643

Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M (1997) Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. Plant Cell 9: 841-857

Ajikumar PK, Xiao WH, Tyo KE, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G (2010) Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. Science 330: 70-74

Alexandre J, Batteux F, Nicco C, Chereau C, Laurent A, Guillevin L, Weill B, Goldwasser F (2006) Accumulation of hydrogen peroxide is an early and crucial step for paclitaxel-induced cancer cell death both in vitro and in vivo. Int J Cancer 119: 41-48

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: 5484-5496

Amos LA, Lowe J (1999) How taxol stabilises microtubule structure. Chem Biol 6: 65-69

Avanci NC, Luche DD, Goldman GH, Goldman MH (2010) Jasmonates are phytohormones with multiple functions, including plant defense and reproduction. Genet Mol Res 9: 484-505

Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14: 283-291

Balbi V, Devoto A (2008) Jasmonate signalling network in *Arabidopsis thaliana*: crucial regulatory nodes and new physiological scenarios. New Phytol 177: 301-318

Bassel GW, Lan H, Glaab E, Gibbs DJ, Gerjets T, Krasnogor N, Bonner AJ, Holdsworth MJ, Provart NJ (2011) Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. Proc Natl Acad Sci USA 108: 9709-9714

Benjamini Y, Hochberg, Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Statis Soci 57: 289-300

Berger S, Bell E, Mullet JE (1996) Two methyl jasmonate-insensitive mutants show altered expression of AtVsp in response to methyl jasmonate and wounding. Plant Physiol 111: 525-531

Blaschke UK, Silberstein J, Muir TW (2000) Protein engineering by expressed protein ligation. Methods Enzymol 328: 478-496

Bolognesi A, Barbieri L, Abbondanza A, Falasca AI, Carnicelli D, Battelli MG, Stirpe F (1990) Purification and properties of new ribosome-inactivating proteins with RNA N-glycosidase activity. Biochim Biophys Acta 1087: 293-302

Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12: 2383-2394

Boter M, Ruiz-Rivero O, Abdeen A, Prat S (2004) Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. Genes Dev 18: 1577-1591

Brady SM, Zhang L, Megraw M, Martinez NJ, Jiang E, Yi CS, Liu W, Zeng A, Taylor-Teeples M, Kim D, Ahnert S, Ohler U, Ware D, Walhout AJ, Benfey PN (2011) A stele-enriched gene regulatory network in the Arabidopsis root. Mol Syst Biol 7: 459

Broun P (2004) Transcription factors as tools for metabolic engineering in plants. Current Opin Plant Biol 7: 202-209

Browse J, Howe GA (2008) New weapons and a rapid response against insect attack. Plant Physiol 146: 832-838

Bruce W, Folkerts O, Garnaat C, Crasta O, Roth B, Bowen B (2000) Expression profiling of the maize flavonoid pathway genes controlled by estradiol-inducible transcription factors CRC and P. Plant Cell 12: 65-80

Buchanan, Wilhelm G, Russell J (2000) *Biochemistry & Molecular Biology of Plants*. 2nd ed. The United State of American: the American Society of Plant Biologists.

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EG, Hall RD, Bovy AG, Luo J, Martin C (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat Biotechnol 26: 1301-1308

CH Zhang, XG Mei, L Liu, LJ, Yu (2000) Enhanced paclitaxel production induced by the combination of elicitors in cell suspension cultures of Taxus chinensis. Biotechnol Lett 22: 1561-1564

Carey MF, Peterson CL, Smale ST (2012) Experimental strategies for the identification of DNA-binding proteins. Cold Spring Harb protoc 2012: 18-33

Chau M, Croteau R (2004) Molecular cloning and characterization of a cytochrome P450 taxoid 2alpha-hydroxylase involved in Taxol biosynthesis. Arch Biochem Biophys 427: 48-57

Chau M, Jennewein S, Walker K, Croteau R (2004) Taxol biosynthesis: Molecular cloning and characterization of a cytochrome P450 taxoid 7 beta-hydroxylase. Chem Biol 11: 663-672

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J 66: 212-229

Cheng Z, Sun L, Qi T, Zhang B, Peng W, Liu Y, Xie D (2011) The bHLH transcription factor MYC3 interacts with the Jasmonate ZIM-domain proteins to mediate jasmonate response in *Arabidopsis*. Mol Plant 4: 279-288

Chini A, Fonseca S, Chico JM, Fernandez-Calvo P, Solano R (2009) The ZIM domain mediates homo- and heteromeric interactions between *Arabidopsis* JAZ proteins. Plant J 59: 77-87

Chini A, Fonseca S, Fernandez G, Adie B, Chico JM, Lorenzo O, Garcia-Casado G, Lopez-Vidriero I, Lozano FM, Ponce MR, Micol JL, Solano R (2007) The JAZ family of repressors is the missing link in jasmonate signalling. Nature 448: 666-671

Chiu LW, Zhou X, Burke S, Wu X, Prior RL, Li L (2010) The purple cauliflower arises from activation of a MYB transcription factor. Plant Physiol 154: 1470-1480

Chrunyk BA, Evans J, Lillquist J, Young P, Wetzel R (1993) Inclusion body formation and protein stability in sequence variants of interleukin-1 beta. J Biol Chem 268: 18053-18061

Chu HY, Wegel E, Osbourn A (2011) From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. Plant J 66: 66-79

Chu Q, Vincent M, Logan D, Mackay JA, Evans WK (2005) Taxanes as first-line therapy for advanced non-small cell lung cancer: a systematic review and practice guideline. Lung Cancer 50: 355-374

Chun OK, Kim DO, Lee CY (2003) Superoxide radical scavenging activity of the major polyphenols in fresh plums. J Agri Food Chem 51: 8067-8072

Chung HS, Howe GA (2009) A critical role for the TIFY motif in repression of jasmonate signaling by a stabilized splice variant of the JASMONATE ZIM-domain protein JAZ10 in *Arabidopsis*. Plant Cell 21: 131-145

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: 81-92

Coates CJ, Kaminski JM, Summers JB, Segal DJ, Miller AD, Kolb AF (2005) Site-directed genome modification: derivatives of DNA-modifying enzymes as targeting tools. Trends Biotechnol 23: 407-419

Costa CL, Arruda P, Benedetti CE (2000) An *arabidopsis* gene induced by wounding functionally homologous to flavoprotein oxidoreductases. Plant Mol Biol 44: 61-71

Creelman RA, Mullet JE (1997) Biosynthesis and Action of Jasmonates in Plants. Annu Rev Plant Physiol Plant Mol Biol 48: 355-381

Creemers GJ, Bolis G, Gore M, Scarfone G, Lacave AJ, Guastalla JP, Despax R, Favalli G, Kreinberg R, Van Belle S, Hudson I, Verweij J, Ten Bokkel Huinink WW (1996) Topotecan, an active drug in the second-line treatment of epithelial ovarian cancer: results of a large European phase II study. J Clin Oncol 14: 3056-3061

Croteau R, Ketchum RE, Long RM, Kaspera R, Wildung MR (2006) Taxol biosynthesis and molecular genetics. Phytochem Rev 5: 75-97

Dal Cin V, Tieman DM, Tohge T, McQuinn R, de Vos RC, Osorio S, Schmelz EA, Taylor MG, Smits-Kroon MT, Schuurink RC, Haring MA, Giovannoni J, Fernie AR, Klee HJ (2011) Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit. Plant Cell 23: 2738-2753

De Boer K, Tilleman S, Pauwels L, Vanden Bossche R, De Sutter V, Vanderhaeghen R, Hilson P, Hamill JD, Goossens A (2011) APETALA2/ETHYLENE RESPONSE FACTOR and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. Plant J 66: 1053-1065

De Sutter V, Vanderhaeghen R, Tilleman S, Lammertyn F, Vanhoutte I, Karimi M, Inze D, Goossens A, Hilson P (2005) Exploration of jasmonate signalling via automated and standardized transient expression assays in tobacco cells. Plant J 44: 1065-1076

Denis AE, Greene, D (1988) A highly efficient, practical approach to natural taxol. J. Am. Chem. Soc 110: 5917

Deplancke B, Dupuy D, Vidal M, Walhout AJ (2004) A gateway-compatible yeast one-hybrid system. Genome Res 14: 2093-2101

Deplancke B, Vermeirssen V, Arda HE, Martinez NJ, Walhout AJ (2006) Gateway-compatible yeast one-hybrid screens. CSH Protoc 2006

Devoto A, Turner JG (2003) Regulation of jasmonate-mediated plant responses in arabidopsis. Ann Bot 92: 329-337

Dixon RA (2001) Natural products and plant disease resistance. Nature 411: 843-847

Dombrecht B, Xue GP, Sprague SJ, Kirkegaard JA, Ross JJ, Reid JB, Fitt GP, Sewelam N, Schenk PM, Manners JM, Kazan K (2007) MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. Plant Cell 19: 2225-2245

Doran P (2002) *Plant biotechnology and transgenic plants*. New York: Marcel Dekker, Inc.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in *Arabidopsis*. Trends Plant Sci 15: 573-581

Duval M, Hsieh TF, Kim SY, Thomas TL (2002) Molecular characterization of AtNAM: a member of the *Arabidopsis* NAC domain superfamily. Plant Mol Biol 50: 237-248

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: 297-303

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends Plant Sci 5: 199-206

Exposito O, Bonfill M, Onrubia M, Jane A, Moyano E, Cusido RM, Palazon J, Pinol MT (2009) Effect of taxol feeding on taxol and related taxane production in Taxus baccata suspension cultures. N Biotechnol 25: 252-259

Fan W, Dong X (2002) In vivo interaction between NPR1 and transcription factor TGA2 leads to salicylic acid-mediated gene activation in *Arabidopsis*. Plant Cell 14: 1377-1389

Farmer EE (2007) Plant biology: jasmonate perception machines. Nature 448: 659-660

Farmer EE, Almeras E, Krishnamurthy V (2003) Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory. Current Opin Plant Biol 6: 372-378

Fereol L, Chovelon V, Causse S, Triaire D, Arnault I, Auger J, Kahane R (2005) Establishment of embryogenic cell suspension cultures of garlic (Allium sativum L.), plant regeneration and biochemical analyses. Plant Cell Rep 24: 319-325

Fernandez-Calvo P, Chini A, Fernandez-Barbero G, Chico JM, Gimenez-Ibanez S, Geerinck J, Eeckhout D, Schweizer F, Godoy M, Franco-Zorrilla JM, Pauwels L, Witters E, Puga MI, Paz-Ares J, Goossens A, Reymond P, De Jaeger G, Solano R (2011) The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. Plant Cell 23: 701-715

Ferre-D'Amare AR, Pognonec P, Roeder RG, Burley SK (1994) Structure and function of the b/HLH/Z domain of USF. EMBO J13: 180-189

Field B, Fiston-Lavier AS, Kemen A, Geisler K, Quesneville H, Osbourn AE (2011) Formation of plant metabolic gene clusters within dynamic chromosomal regions. Proc Natl Acad Sci USA 108: 16116-16121

Field B, Osbourn AE (2008) Metabolic diversification--independent assembly of operon-like gene clusters in different plants. Science 320: 543-547

Figueroa P, Browse J (2012) The Arabidopsis JAZ2 promoter contains a G-Box and thymidine-rich module that are necessary and sufficient for jasmonate-dependent activation by MYC transcription factors and repression by JAZ proteins. Plant Cell Physiol 53: 330-343

Flores LJWaHE (1994) Effect of Taxol and Related Compounds on Growth of Plant Pathogenic Fungi. Biochem Cell Biol 84: 1173-1178

Fonseca S, Chico JM, Solano R (2009a) The jasmonate pathway: the ligand, the receptor and the core signalling module. Current Opin Plant Biol 12: 539-547

Fonseca S, Chini A, Hamberg M, Adie B, Porzel A, Kramell R, Miersch O, Wasternack C, Solano R (2009b) (+)-7-iso-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. Nat Chem Biol 5: 344-350

Franceschi VR, Grimes HD (1991) Induction of soybean vegetative storage proteins and anthocyanins by low-level atmospheric methyl jasmonate. Proc Natl Acad Sci USA 88: 6745-6749

Fuchs C, Mitchell EP, Hoff PM (2006) Irinotecan in the treatment of colorectal cancer. Cancer Treat Rev 32: 491-503

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: 393-404

Garner MM, Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res 9: 3047-3060

Gaudinier A, Zhang L, Reece-Hoyes JS, Taylor-Teeples M, Pu L, Liu Z, Breton G, Pruneda-Paz JL, Kim D, Kay SA, Walhout AJ, Ware D, Brady SM (2011) Enhanced Y1H assays for *Arabidopsis*. Nat Methods 8: 1053-1055

Gibson JD, Khanal BP, Zubarev ER (2007) Paclitaxel-functionalized gold nanoparticles. J Am Chem Soc 129: 11653-11661

Goff SA, Goldberg AL (1987) An increased content of protease La, the lon gene product, increases protein degradation and blocks growth in *Escherichia coli*. J Biol Chem 262: 4508-4515

Gonzalez A, Zhao M, Leavitt JM, Lloyd AM (2008) Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. Plant J 53: 814-827

Goodman CD, Casati P, Walbot V (2004) A multidrug resistance-associated protein involved in anthocyanin transport in Zea mays. Plant Cell 16: 1812-1826

Goossens A, Hakkinen ST, Laakso I, Seppanen-Laakso T, Biondi S, De Sutter V, Lammertyn F, Nuutila AM, Soderlund H, Zabeau M, Inze D, Oksman-Caldentey KM (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. Proc Natl Acad Sci USA 100: 8595-8600

Grafi G, Ben-Meir H, Avivi Y, Moshe M, Dahan Y, Zemach A (2007) Histone methylation controls telomerase-independent telomere lengthening in cells undergoing dedifferentiation. Dev Biol 306: 838-846

Grotewold E (2006) The genetics and biochemistry of floral pigments. Annu Rev Plant Biol 57: 761-780

Grotewold E (2008) Transcription factors for predictive plant metabolic engineering: are we there yet? Curr Opin Biotechnol 19: 138-144

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: 683-696

Gundlach H, Muller MJ, Kutchan TM, Zenk MH (1992) Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures. Proc Natl Acad Sci USA 89: 2389-2393

Guranowski A, Miersch O, Staswick PE, Suza W, Wasternack C (2007) Substrate specificity and products of side-reactions catalyzed by jasmonate:amino acid synthetase (JAR1). FEBS Lett 581: 815-820

Hammond LA, Hilsenbeck SG, Eckhardt SG, Marty J, Mangold G, MacDonald JR, Rowinsky EK, Von Hoff DD, Weitman S (2000) Enhanced antitumour activity of 6-hydroxymethylacylfulvene in combination with topotecan or paclitaxel in the MV522 lung carcinoma xenograft model. Eur J Cancer 36: 2430-2436

Hao da C, 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: e21220

Hartzell H (1991). *The yew tree: a thousand whispers.* Eugene, Oregon: Hulogosi.

Havrilesky LJ, Garfield CF, Barnett JC, Cohn DE (2012) Economic impact of paclitaxel shortage in patients with newly diagnosed ovarian cancer. Gynecol Oncol 125: 631-634

Hefner J, Ketchum RE, 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. Arc Biochem Biophys 360: 62-74

Hegedus D, Yu M, Baldwin D, Gruber M, Sharpe A, Parkin I, Whitwill S, Lydiate D (2003) Molecular characterization of brassica napus NAC domain transcriptional activators induced in response to biotic and abiotic stress. Plant Mol Biol 53: 383-397

Hens K, Feuz JD, Deplancke B (2012) A high-throughput gateway-compatible yeast one-hybrid screen to detect protein-DNA interactions. Methods Mol Biol 786: 335-355

Hezari M, Ketchum RE, Gibson DM, Croteau R (1997) Taxol production and taxadiene synthase activity in *Taxus* canadensis cell suspension cultures. Arc Biochem Biophys 337: 185-190

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. J Biol Chem 280: 25590-25595

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proc Natl Acad Sci USA104: 6478-6483

Hounsome N, Hounsome B, Tomos D, Edwards-Jones G (2008) Plant metabolites and nutritional quality of vegetables. J Food Sci 73: 48-65

Huang Q, Roessner CA, Croteau R, Scott AI (2001) Engineering *Escherichia coli* for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. Bioorg Med Chem 9: 2237-2242

Imanishi S, Hashizume K, Kojima H, Ichihara A, Nakamura K (1998) An mRNA of tobacco cell, which is rapidly inducible by methyl jasmonate in the presence of cycloheximide, codes for a putative glycosyltransferase. Plant Cell Physiol 39: 202-211

Ishihama N, Yamada R, Yoshioka M, Katou S, Yoshioka H (2011) Phosphorylation of the *Nicotiana benthamiana* WRKY8 transcription factor by MAPK functions in the defense response. Plant Cell 23: 1153-1170

IUC (2010) The IUCN Red List of Threatened Species (http://www.iucnredlist.org/).

Jefferson RA (1989) The GUS reporter gene system. Nature 342: 837-838

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6: 3901-3907

Jennewein S, Long RM, Williams RM, Croteau R (2004a) Cytochrome p450 taxadiene 5α-hydroxylase, a mechanistically unusual monooxygenase catalyzing the first oxygenation step of taxol biosynthesis. Chem Biol 11: 379-387

Jennewein S, Rithner CD, Williams RM, Croteau RB (2001) Taxol biosynthesis: taxane 13α-hydroxylase is a cytochrome P450-dependent monooxygenase. Proc Natl Acad Sci USA 98: 13595-13600

Jennewein S, Wildung MR, Chau M, Walker K, Croteau R (2004b) Random sequencing of an induced *Taxus* cell cDNA library for identification of clones involved in taxol biosynthesis. Proc Natl Acad Sci USA 101: 9149-9154

Jia L, Clegg MT, Jiang T (2004) Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies indica and japonica genomes. Plant Physiol 134: 575-585

Jofuku KD, den Boer BG, Van Montagu M, Okamuro JK (1994) Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. Plant Cell 6: 1211-1225

Jonczyk R, Schmidt H, Osterrieder A, Fiesselmann A, Schullehner K, Haslbeck M, Sicker D, Hofmann D, Yalpani N, Simmons C, Frey M, Gierl A (2008) Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of Bx6 and Bx7. Plant Physiol 146: 1053-1063

Joshi JB, Elias, C.B. and Patole, M.S. (1996) Role of hydrodynamic shear in the cultivation of animal, plant and microbial cells. Chem Eng J Biochem Eng J 62: 121-141

Jyoti Kapila RDR, Marc Van Montagu, Geert Angenon (1997) An *Agrobacterium*-mediated transient gene expression system for intact leaves. Plant Sci 122: 101-108

Kantarjian HM, O'Brien S, Anderlini P, Talpaz M (1996) Treatment of myelogenous leukemia: current status and investigational options. Blood 87: 3069-3081

Kaspera R, Croteau R (2006) Cytochrome P450 oxygenases of taxol biosynthesis. Phytochem Rev 5: 433-444

Katoh A, Ohki, H., Inai, K and Hashimoto, T. (2005) Molecular regulation of nicotine biosynthesis. Plant Biotechnol 22: 389-392

Kayser O, Quax W (2007) *Medicinal plant biotechnology*: *from basic research to industrial application*. Wiley-VCH, Weinheim

Ketchum RE, Wherland L, Croteau RB (2007) Stable transformation and long-term maintenance of transgenic *Taxus* cell suspension cultures. Plant Cell Rep 26: 1025-1033

Ketchum REBaG, D.M. (1996) Paclitaxel production in suspension cell cultures of *Taxus*. Plant cell tissue organ culture 46: 9-16

Kikuchi K, Ueguchi-Tanaka M, Yoshida KT, Nagato Y, Matsusoka M, Hirano HY (2000) Molecular analysis of the NAC gene family in rice. Mol Gen Genet 262: 1047-1051

Kingston DG (2009) Tubulin-interactive natural products as anticancer agents. J Nat Prod 72: 507-515

Koksal M, Jin Y, Coates RM, Croteau R, Christianson DW (2011) Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. Nature 469: 116-120

Kolewe ME, Gaurav V, Roberts SC (2008) Pharmaceutically active natural product synthesis and supply via plant cell culture technology. Mol Pharm 5: 243-256

Kolewe ME, Henson MA, Roberts SC (2011) Analysis of aggregate size as a process variable affecting paclitaxel accumulation in *Taxus* suspension cultures. Biotechnol Prog 27: 1365-1372

Kueger S, Steinhauser D, Willmitzer L, Giavalisco P (2012) High-resolution plant metabolomics: from mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions. Plant J 70: 39-50

Kurland CG (1982) Translational accuracy in vitro. Cell 28: 201-202

Lai Z, Li Y, Wang F, Cheng Y, Fan B, Yu JQ, Chen Z (2011) *Arabidopsis* sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. Plant Cell 23: 3824-3841

Lange BM, Ahkami A (2013) Metabolic engineering of plant monoterpenes, sesquiterpenes and diterpenes-current status and future opportunities. Plant Biotechnol J 11: 169-196

Lau S, De Smet I, Kolb M, Meinhardt H, Jurgens G (2011) Auxin triggers a genetic switch. Nat Cell Biol 13: 611-615

Lee EK, Jin YW, Park JH, Yoo YM, Hong SM, Amir R, Yan Z, Kwon E, Elfick A, Tomlinson S, Halbritter F, Waibel T, Yun BW, Loake GJ (2010) Cultured cambial meristematic cells as a source of plant natural products. Nat Biotechnol 28: 1213-1217

Lee-Parsons CW, Shuler ML (2002) The effect of ajmalicine spiking and resin addition timing on the production of indole alkaloids from *Catharanthus roseus* cell cultures. Biotechnol Bioeng 79: 408-415

Legg PD, Collins GB (1971) Inheritance of per cent total alkaloids in *Nicotiana tabacum* I. ii. genetic effects of two loci in burley 21 × la burley 21 population. Cana J Genet Cytol 13: 287-291

Lenka SK, Boutaoui N, Paulose B, Vongpaseuth K, Normanly J, Roberts SC, Walker EL (2012) Identification and expression analysis of methyl jasmonate responsive ESTs in paclitaxel producing *Taxus cuspidata* suspension culture cells. BMC Genomics 13: 148

Leung KW, Wong AS (2010) Pharmacology of ginsenosides: a literature review. Chin Med 5: 20

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 (Stuttg) 15: 19-26

Li ST, Zhang P, Zhang M, Fu CH, Zhao CF, Dong YS, Guo AY, Yu LJ (2012) Transcriptional profile of *Taxus* chinensis cells in response to methyl jasmonate. BMC Genomics 13: 295

Lin YS, Tungpradit R, Sinchaikul S, An FM, Liu DZ, Phutrakul S, Chen ST (2008) Targeting the delivery of glycan-based paclitaxel prodrugs to cancer cells via glucose transporters. J Med Chem 51: 7428-7441

Liu Q, Li MZ, Leibham D, Cortez D, Elledge SJ (1998) The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. Curr Biol 8: 1300-1309

Liu SJ, Zhang DQ, Sui XM, Zhang L, Cai ZW, Sun LQ, Liu YJ, Xue Y, Hu GF (2008) The inhibition of in vivo tumorigenesis of osteosarcoma (OS)-732 cells by antisense human osteopontin RNA. Cell Mol Biol Lett 13: 11-19

Lorenzo O, Chico JM, Sanchez-Serrano JJ, Solano R (2004) JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in *Arabidopsis*. Plant Cell 16: 1938-1950

M. Suffness MEW (1995) In taxol: Sciences and Applications. CRC, Boca Raton, FL

Madin K, Sawasaki T, Ogasawara T, Endo Y (2000) A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. Proc Natl Acad Sci USA 97: 559-564

Malik RMC, Mohammad Hossein Mirjalili, Elisabetn Moyano, Javier Palazon, Mercedes Bonfill (2011) Production of the anticancer drug taxol in *Taxus baccata* suspension cultures: A review. Process Biochem 46: 23-34

Mao G, Meng X, Liu Y, Zheng Z, Chen Z, Zhang S (2011) Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in *Arabidopsis*. Plant Cell 23: 1639-1653

Martin C, Paz-Ares J (1997) MYB transcription factors in plants. Trends Genet 13: 67-73

Martin-Parras L, Hernandez P, Martinez-Robles ML, Schvartzman JB (1992) Initiation of DNA replication in ColE1 plasmids containing multiple potential origins of replication. J Biol Chem 267: 22496-22505

Massari ME, Murre C (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol 20: 429-440

Mastropaolo D, Camerman A, Luo Y, Brayer GD, Camerman N (1995) Crystal and molecular structure of paclitaxel (taxol). Proc Natl Acad Sci 92: 6920-6924

Matsuda J, Okabe S, Hashimoto T, Yamada Y (1991) Molecular cloning of hyoscyamine 6 beta-hydroxylase, a 2-oxoglutarate-dependent dioxygenase, from cultured roots of Hyoscyamus niger. J Biol Chem 266: 9460-9464

McCoy E, O'Connor SE (2008) Natural products from plant cell cultures. Progress in drug research. Fortschritte der Arzneimittelforschung. Pro Rec Pharmacol 65: 329, 331-370

Memelink JG, P (2007) Transcription factors involved in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. Phytochem Rev 6: 353-362

Menke FL, Champion A, Kijne JW, Memelink J (1999) A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene Str interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. EMBO J 18: 4455-4463

Miao Y, Laun T, Zimmermann P, Zentgraf U (2004) Targets of the WRKY53 transcription factor and its role during leaf senescence in *Arabidopsis*. Plant Mol Biol 55: 853-867

Miersch O, Neumerkel J, Dippe M, Stenzel I, Wasternack C (2008) Hydroxylated jasmonates are commonly occurring metabolites of jasmonic acid and contribute to a partial switch-off in jasmonate signaling. New Phytol 177: 114-127

Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M (2007) NAC transcription factors, NST1 and NST3, are key regulators of the

formation of secondary walls in woody tissues of *Arabidopsis*. Plant Cell 19: 270-280

Mitsuda N, Ohme-Takagi M (2009) Functional analysis of transcription factors in *Arabidopsis*. Plant Cell Physiol 50: 1232-1248

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) AP2/ERF family transcription factors in plant abiotic stress responses. Biochim Biophys Acta 1819: 86-96

Mizukami H, Tabira, Y., and Ellis, B.E (1993) Methyl jasmonate-induced rosmarinic acid biosynthesis in *Lithospermum erythrorhizon* cell suspension culture. Plant Cell Rep 12: 706-709

Mol J, Grotewold, E. and Koes, R (1998) How genes paint flowers and seeds. Trends Plant Sci Rev 3: 212-217

Morita M, Shitan N, Sawada K, Van Montagu MC, Inze D, Rischer H, Goossens A, Oksman-Caldentey KM, Moriyama Y, Yazaki K (2009) Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. Proc Natl Acad Sci USA 106: 2447-2452

Mylona P, Owatworakit A, Papadopoulou K, Jenner H, Qin B, Findlay K, Hill L, Qi X, Bakht S, Melton R, Osbourn A (2008) Sad3 and sad4 are required for saponin biosynthesis and root development in oat. Plant Cell 20: 201-212

Nagy A (2000) Cre recombinase: the universal reagent for genome tailoring. Genesis 26: 99-109

Naill MC, Roberts SC (2004) Preparation of single cells from aggregated *Taxus* suspension cultures for population analysis. Biotechnol Bioeng 86: 817-826

Nair SK, Burley SK (2000) Recognizing DNA in the library. Nature 404: 715-718

Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L (2000) The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in *Arabidopsis* siliques. Plant Cell 12: 1863-1878

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

Newman DJ, Cragg GM, Snader KM (2000) The influence of natural products upon drug discovery. Nat Prod Rep 17: 215-234

Nicolaou KC W-MD, and Rodney Kiplin Guy (1994) Chemistry and Biology of Taxol. Angen. Chem. lnt. Ed. Engl 33: 15-44

Nicolaou KC, Yang Z, Liu JJ, Ueno H, Nantermet PG, Guy RK, Claiborne CF, Renaud J, Couladouros EA, Paulvannan K, et al. (1994) Total synthesis of taxol. Nature 367: 630-634

Nims E, Dubois CP, Roberts SC, Walker EL (2006) Expression profiling of genes involved in paclitaxel biosynthesis for targeted metabolic engineering. Metab Eng 8: 385-394

Niu Y, Figueroa P, Browse J (2011) Characterization of JAZ-interacting bHLH transcription factors that regulate jasmonate responses in *Arabidopsis.* J Exp Bot 62: 2143-2154

Noble RL (1990) The discovery of the vinca alkaloids--chemotherapeutic agents against cancer. Biochem Cell Biol 68: 1344-1351

Ogata K, Kanei-Ishii C, Sasaki M, Hatanaka H, Nagadoi A, Enari M, Nakamura H, Nishimura Y, Ishii S, Sarai A (1996) The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. Nat Struct Biol 3: 178-187

Ohme-Takagi M, Shinshi H (1995) Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell 7: 173-182

Ohwi (1965) *Flora of Japan*. Washington, DC: Smithsonian Institution.

Okada A, Okada K, Miyamoto K, Koga J, Shibuya N, Nojiri H, Yamane H (2009) OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. J Biol Chem 284: 26510-26518

Okamuro JK, Caster B, Villarroel R, Van Montagu M, Jofuku KD (1997) The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. Proc Natl Acad Sci USA 94: 7076-7081

Oksman-Caldentey WB (2002) Plant biotechnology and transgenic plants. Marcel Dekker, New York

Olsen AN, Ernst HA, Leggio LL, Skriver K (2005) NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci 10: 79-87

Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K, Matsubara K, Osato N, Kawai J, Carninci P, Hayashizaki Y, Suzuki K, Kojima K, Takahara Y, Yamamoto K, Kikuchi S (2003) Comprehensive analysis of NAC family genes in Oryza sativa and *Arabidopsis thaliana*. DNA Res 10: 239-247

Osbourn A (2010a) Gene clusters for secondary metabolic pathways: an emerging theme in plant biology. Plant Physiol 154: 531-535

Osbourn A (2010b) Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. Trends Genet 26: 449-457

Ouwerkerk PB, Meijer AH (2011) Yeast one-hybrid screens for detection of transcription factor DNA interactions. Methods Mol Biol 678: 211-227

Ouyang Z, Zhou Q, Wong WH (2009) ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. Proc Natl Acad Sci USA 106: 21521-21526

Patil RA, Kolewe ME, Normanly J, Walker EL, Roberts SC (2012) Contribution of taxane biosynthetic pathway gene expression to observed variability in paclitaxel accumulation in *Taxus* suspension cultures. Biotechnol J 7: 418-427

Pauwels L, Barbero GF, Geerinck J, Tilleman S, Grunewald W, Perez AC, Chico JM, Bossche RV, Sewell J, Gil E, Garcia-Casado G, Witters E, Inze D, Long JA, De Jaeger G, Solano R, Goossens A (2010) NINJA connects the co-repressor TOPLESS to jasmonate signalling. Nature 464: 788-791

Pauwels L, Goossens A (2011) The JAZ proteins: a crucial interface in the jasmonate signaling cascade. Plant Cell 23: 3089-3100

Pauwels L, Morreel K, De Witte E, Lammertyn F, Van Montagu M, Boerjan W, Inze D, Goossens A (2008) Mapping methyl jasmonate-mediated transcriptional

reprogramming of metabolism and cell cycle progression in cultured *Arabidopsis* cells. Proc Natl Acad Sci USA 105: 1380-1385

Pavlov MY, Ehrenberg M (1996) Rate of translation of natural mRNAs in an optimized in vitro system. Arch Biochem Biophys 328: 9-16

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: 3553-3558

Perez EA (1998) Paclitaxel in Breast Cancer. Oncologist 3: 373-389

Qi T, Song S, Ren Q, Wu D, Huang H, Chen Y, Fan M, Peng W, Ren C, Xie D (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: 1795-1814

Qi X, Bakht S, Leggett M, Maxwell C, Melton R, Osbourn A (2004) A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. Proc Natl Acad Sci USA 101: 8233-8238

Ranish JA, Hahn S (1996) Transcription: basal factors and activation. Curr Opin Genet Dev 6: 151-158

Rea PA (2007) Plant ATP-binding cassette transporters. Annu Rev Plant Biol 58: 347-375

Renneberg R (2007) Biotech History: Yew trees, paclitaxel synthesis and fungi. Biotechnol J 2: 1207-1209

Reymond P, Weber H, Damond M, Farmer EE (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. Plant Cell 12: 707-720

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science 290: 2105-2110

Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MC, Withers ST, Shiba Y, Sarpong R, Keasling JD (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440: 940-943

Robert A. Holton CS, Hyeong-Baik Kim, Feng Liang, Ronald J. Biediger, P. Douglas Boatman, Mitsuru Shindo, Chase C. Smith, Soekchan Kim, Hossain Nadizadeh, Yukio Suzuki, Chunlin Tao, Phong Vu, Suhan Tang, Pingsheng Zhang, Krishna K, Murthi, Lisa N. Gentile, and Jyanwei H. Liu (1994) First Total Synthesis of Taxol. 1. Functionalization of the B Ring. J Am Chem SO 116: 1597-1598

Roberts SC (2007) Production and engineering of terpenoids in plant cell culture. Nat Chem Biol 3: 387-395

Robinson KA, Koepke JI, Kharodawala M, Lopes JM (2000) A network of yeast basic helix-loop-helix interactions. Nucleic Acids Res 28: 4460-4466

Routaboul JM, Kerhoas L, Debeaujon I, Pourcel L, Caboche M, Einhorn J, Lepiniec L (2006) Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. Planta 224: 96-107

Rowinsky EK DR (1991) The clinical pharmacology and use of antimicrotubule agents in cancer chemotherapeutics. Pharmacol Ther. 52: 35-84

Rowinsky EK, Onetto N, Canetta RM, Arbuck SG (1992) Taxol: the first of the taxanes, an important new class of antitumor agents. Semin Oncol 19: 646-662

Rushton PJ, Somssich IE, Ringler P, Shen QJ (2010) WRKY transcription factors. Trends Plant Sci 15: 247-258

Sakamoto T, Miura K, Itoh H, Tatsumi T, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Agrawal GK, Takeda S, Abe K, Miyao A, Hirochika H, Kitano H, Ashikari M, Matsuoka M (2004) An overview of gibberellin metabolism enzyme genes and their related mutants in rice. Plant Physiol 134: 1642-1653

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002) DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun 290: 998-1009

Sauer B, Henderson N (1988) Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc Natl Acad Sci USA 85: 5166-5170

Schiff PB, Fant J, Horwitz SB (1979) Promotion of microtubule assembly in vitro by taxol. Nature 277: 665-667

Schmidt BM, Ribnicky DM, Lipsky PE, Raskin I (2007) Revisiting the ancient concept of botanical therapeutics. Nat Chem Biol 3: 360-366

Schoendorf A, Rithner CD, Williams RM, Croteau RB (2001) Molecular cloning of a cytochrome P450 taxane 10 beta-hydroxylase cDNA from *Taxus* and functional expression in yeast. Proc Natl Acad Sci USA 98: 1501-1506

Schwinn KMDaKE (2003) Transcriptional regulation of secondary metabolism. Funct Plant Biol 30: 913-925

Sengottuvel V, Leibinger M, Pfreimer M, Andreadaki A, Fischer D (2011) Taxol facilitates axon regeneration in the mature CNS. J Neurosci 31: 2688-2699

Shang Y, Schwinn KE, Bennett MJ, Hunter DA, Waugh TL, Pathirana NN, Brummell DA, Jameson PE, Davies KM (2007) Methods for transient assay of gene function in floral tissues. Plant Methods 3: 1

Shanks JV, Morgan J (1999) Plant 'hairy root' culture. Curr Opin Biotechnol 10: 151-155

Sheard LB, Tan X, Mao H, Withers J, Ben-Nissan G, Hinds TR, Kobayashi Y, Hsu FF, Sharon M, Browse J, He SY, Rizo J, Howe GA, Zheng N (2010) Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. Nature 468: 400-405

Shi J, Yu J, Pohorly JE, Kakuda Y (2003) Polyphenolics in grape seeds-biochemistry and functionality. J Med Food 6: 291-299

Shigaki T, Vyzasatya RR, Sivitz AB, Ward JM, Sze H, Hirschi KD (2005) The Cre-loxP recombination-based reporter system for plant transcriptional expression studies. Plant Mol Biol 58: 65-73

Shimizu T, Toumoto A, Ihara K, Shimizu M, Kyogoku Y, Ogawa N, Oshima Y, Hakoshima T (1997) Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. EMBO J 16: 4689-4697

Shoji T, Kajikawa M, Hashimoto T (2010) Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. Plant Cell 22: 3390-3409

Shyu C, Figueroa P, Depew CL, Cooke TF, Sheard LB, Moreno JE, Katsir L, Zheng N, Browse J, Howe GA (2012) JAZ8 lacks a canonical degron and has an EAR motif that mediates transcriptional repression of jasmonate responses in *Arabidopsis*. Plant Cell 24: 536-550

Song S, Qi T, Huang H, Ren Q, Wu D, Chang C, Peng W, Liu Y, Peng J, Xie D (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: 1000-1013

Sparkes IA, Runions J, Kearns A, Hawes C (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc 1: 2019-2025

Spjut RW (2010) Overview of the genus *Taxus* (Taxaceae): The Species, Their Clasiification, and Female Reproductive Morphology. *The World Botanical Associates, Inc.*

Staswick PE (2008) JAZing up jasmonate signaling. Trends Plant Sci 13: 66-71

Staswick PE, Tiryaki I (2004) The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in *Arabidopsis*. Plant Cell 16: 2117-2127

Stierle A, Strobel G, Stierle D (1993) Taxol and taxane production by *Taxomyces andreanae*, an endophytic fungus of Pacific yew. Science 260: 214-216

Stirpe F, Barbieri L, Battelli MG, Soria M, Lappi DA (1992) Ribosome-inactivating proteins from plants: present status and future prospects. Biotechnology (N Y) 10: 405-412

Stratmann M, Schibler U (2011) Transcription factor loading: please take my place! Cell 146: 497-499

Suffness M (1995) Taxol: Science and Application. CRC Press, Inc., *Boca Raton, FL.*

Sun C, Palmqvist S, Olsson H, Boren M, Ahlandsberg S, Jansson C (2003) A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. Plant Cell 15: 2076-2092

Sun S, Yu JP, Chen F, Zhao TJ, Fang XH, Li YQ, Sui SF (2008) TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE- and ethylene-responsive element-mediated signaling pathways in *Arabidopsis*. J Biol Chem 283: 6261-6271

Suttipanta N, Pattanaik S, Kulshrestha M, Patra B, Singh SK, Yuan L (2011) The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. Plant Physiol 157: 2081-2093

Szemenyei H, Hannon M, Long JA (2008) TOPLESS mediates auxin-dependent transcriptional repression during *Arabidopsis* embryogenesis. Science 319: 1384-1386

Thines B, Katsir L, Melotto M, Niu Y, Mandaokar A, Liu G, Nomura K, He SY, Howe GA, Browse J (2007) JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. Nature 448: 661-665

Thorpe T (2012) History of plant tissue culture. Methods Mol Biol 877: 9-27

Thorpe TA (2007) History of plant tissue culture. Mol Biotechnol 37: 169-180

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, Noji M, Yamazaki M, Saito K (2005) Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant J 42: 218-235

Toledo-Ortiz G, Huq E, Quail PH (2003) The *Arabidopsis* basic/helix-loop-helix transcription factor family. Plant Cell 15: 1749-1770

Tran LS, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (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: 2481-2498

Trask BJ (1991) Fluorescence in situ hybridization: applications in cytogenetics and gene mapping. Trends Genet 7: 149-154

Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ 32: 1633-1651

van der Fits L, Memelink J (2000) ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. Science 289: 295-297

Vanholme B, Grunewald W, Bateman A, Kohchi T, Gheysen G (2007) The tify family previously known as ZIM. Trends Plant Sci 12: 239-244

Velappan N, Sblattero D, Chasteen L, Pavlik P, Bradbury AR (2007) Plasmid incompatibility: more compatible than previously thought? Protein Eng Des Sel 20: 309-313

Vom Endt D, Soares e Silva M, Kijne JW, Pasquali G, Memelink J (2007) Identification of a bipartite jasmonate-responsive promoter element in the Catharanthus roseus ORCA3 transcription factor gene that interacts specifically with AT-Hook DNA-binding proteins. Plant Physiol 144: 1680-1689

Vongpaseuth K, Roberts SC (2007) Advancements in the understanding of paclitaxel metabolism in tissue culture. Curr Pharm Biotechnol 8: 219-236

Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC (1999) The Transparent testa glabra1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. Plant Cell 11: 1337-1350

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*. Proc Natl Acad Sci USA 97: 583-587

Walker K, Fujisaki S, Long R, Croteau R (2002a) Molecular cloning and heterologous expression of the C-13 phenylpropanoid side chain-CoA acyltransferase that functions in *Taxol* biosynthesis. Proc Natl Acad Sci USA 99: 12715-12720

Walker K, Long R, Croteau R (2002b) The final acylation step in taxol biosynthesis: cloning of the taxoid C13-side-chain N-benzoyltransferase from *Taxus*. Proc Natl Acad Sci USA 99: 9166-9171

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: 371-380

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: 53947-53954

Wang CT, Liu H, Gao XS, Zhang HX (2010) Overexpression of G10H and ORCA3 in the hairy roots of *Catharanthus roseus* improves catharanthine production. Plant Cell Rep 29: 887-894

Wang TH, Wang HS, Soong YK (2000) Paclitaxel-induced cell death: where the cell cycle and apoptosis come together. Cancer 88: 2619-2628

Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT (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: 2325-2327

Wasternack C (2007) Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. Ann Bot 100: 681-697

Weigel D (1995) The APETALA2 domain is related to a novel type of DNA binding domain. Plant Cell 7: 388-389

Wheeler AL, Long RM, Ketchum RE, Rithner CD, Williams RM, Croteau R (2001) Taxol biosynthesis: differential transformations of taxadien-5 alpha-ol and its acetate ester by cytochrome P450 hydroxylases from *Taxus* suspension cells. Arch Biochem Biophys 390: 265-278

Wildung MR, Croteau R (1996) A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. J Biol Chem 271: 9201-9204

Williams DC, Wildung MR, Jin AQ, Dalal D, Oliver JS, Coates RM, Croteau R (2000) Heterologous expression and characterization of a "Pseudomature" form of taxadiene synthase involved in paclitaxel (Taxol) biosynthesis and evaluation of a potential intermediate and inhibitors of the multistep diterpene cyclization reaction. Arch Biochem Biophys 379: 137-146

Winzer T, Gazda V, He Z, Kaminski F, Kern M, Larson TR, Li Y, Meade F, Teodor R, Vaistij FE, Walker C, Bowser TA, Graham IA (2012) A Papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science 336: 1704-1708

Xie DX, Feys BF, James S, Nieto-Rostro M, Turner JG (1998) COI1: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. Science 280: 1091-1094

Xie DY, Sharma SB, Wright E, Wang ZY, Dixon RA (2006) Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. Plant J 45: 895-907

Xie K, Zhang J, Xiang Y, Feng Q, Han B, Chu Z, Wang S, Zhang Q, Xiong L (2005) Isolation and annotation of 10828 putative full length cDNAs from indica rice. Sci China CLife Sci 48: 445-451

Xu X, Chen C, Fan B, Chen Z (2006) Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors. Plant Cell 18: 1310-1326

Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2005) Solution structure of an Arabidopsis WRKY DNA binding domain. Plant Cell 17: 944-956

Yamin Wei LL, Xuanwei Zhou, Juan Lin, Xiaofen, Sun and Kexuan Tang (2012) Engineering taxol biosynthtic pathway for improving taxol yield in taxol-producing endophytic gungus EFY-21 (*Ozonium Sp*.). Afri J Biotechnol 11: 9094-9101

Yang SJ, Fang, J.M. and Cheng, Y.S. (1999) Lignans, flavonoids and phenolic derivatives from *Taxus mairei*. J Chi Chem Soc 46: 811-818

Yang Y, Li R, Qi M (2000) In vivo analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. Plant J 22: 543-551

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing-Wang D, Zhangliang C, Hongya G, Li-Jia Q (2006) The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol 60: 107-124

Yu H, Gerstein M (2006) Genomic analysis of the hierarchical structure of regulatory networks. Proc Natl Acad Sci USA 103: 14724-14731

Yuan L, Wu L, Chen J, Wu Q, Hu S (2010) Paclitaxel acts as an adjuvant to promote both Th1 and Th2 immune responses induced by ovalbumin in mice. Vaccine 28: 4402-4410

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: 1129-1132

Yun UW, Yan Z, Amir R, Hong S, Jin YW, Lee EK, Loake GJ (2012) Plant natural products: history, limitations and the potential of cambial meristematic cells. Biotechnol Genet Eng Rev 28: 47-59

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: 73-84

Zhao J, Dixon RA (2009) MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. Plant Cell 21: 2323-2340

Zhao J, Dixon RA (2010) The 'ins' and 'outs' of flavonoid transport. Trends Plant Sci 15: 72-80

Zhao Q, Dixon RA (2011) Transcriptional networks for lignin biosynthesis: more complex than we thought? Trends Plant Sci 16: 227-233

Appendix I: TFs associated DNA binding motifs on paclitaxel biosynthesis pathway gene promoters

The TFs associated DNA binding *cis*-elements are analysed by searching promoter sequences in the database of Plant *Cis*-acting Regulatory DNA Elements (PLACE) [http://www.dna.affrc.go.jp/PLACE/]. The sequences of five paclitaxel biosynthesis pathway gene promoters were analysed. The TF family-associated DNA binding motifs are shown for the 19 MeJA induced putative *Taxus* TFs grouped into five different TF families. GCC-box (GCCGCC) is bound by AP2/ERF TFs, E-box (CANNTG) and T/G-box (AACGTG) are bound by bHLH TFs, W-box (TGAC) is bound by WRKY TFs, NAC binding site (NACBS) (CACG) is bound by NAC TFs and MYB binding sites (MYBBS) are variable in core sequences (Ciolkowski et al. 2008; Dubos et al. 2010; Fujimoto et al. 2000; Toledo-Ortiz et al. 2003; Tran et al. 2004).

PAM: 257 bp

CTACCTAAACAGACAAGACAACGAATACGTTCCCCGGGGGCCAGAACCATGTGTTAAAAAAATTAAATCA
E-box

NACBS
TTGAGTGTTTCAGCATGTGATAAATGCAACGTCTCGCCCTTCCCTAGACGAATTCCGGGTTCACGTGTCT
E-box E-box

ATATATATAAGTGCGACGCAACGATAAATTGCAGTGCAGCGTCCATTAGAGATTTTTCAGTTTTATCTCGCT

CAAGTTTCAATCTTTTAATTTTAAAGTTATTTTCCTTGCTCTGCA

TASY: 1433 bp

AACTCGCAATAGCTAGGACATCTTTCATAGTATCAATTTTAAGGTATATTTTTTGTGATTTGATTCAATTTTT

W-box W-box
GTAGAAGAATGGATGTCATGAAGATTAATAAATTTAACTGACAAGAATGCAGCCACACTTGACCTTTACC
W-box MYB2BS E-box

GTTACTATACAGAGTTTGATTAGTGTGACAAATATAGATAGGAAAAAAGGAAGAGAGACATCAAATTACA
MYB2BS W-box

CATATTAAGGGTATCTACCTTTCAAGTCCAATATGATTTGCATCAAGTTTTAAAGTGGAGTTACATAATTAA

ATTTT

GTGAAAATGTAAGATATTAAATCAAGACTGAAGAGGTGGACATAAAAAGGGTATAATAAATGTCATAG
W-box

GAAAATGAGTAGTTTGGATATGTGGTAAAATAAGGAGTATTTAATAATTGAAAATTTGTATTGGTCGTG
MYBSt1BS

GGTGTGACTTTTAAAAGATTGAGATATAGATTTGGAAATATGAGTATTATTTAATATCCTGTGAATGCAT
W-box MYBSt1BS

ATGATAGGCAATAATTTGATATAGCGAATCCCTCGTGACTTTCGAAGTTATCTTCAAAGTACAATGCCTC
E-box W-box

AAGAATATGTGCA

E-box
TCACCCCCTTTCTCTTTCAAAAGGTTCATAAAATTGACAAGTGTTCCGTATGTGACTAACTAGAAGAAAGAA
W-box W-box

GACGTTGAGCTACGATCCATCCAACAATTTCTCTACCATTGAAAATTTCATGCTTCAACCCCTTCCTTGTTACC

CCTACTCTTCTTGCAATGTTTCAACTTCACACCTTTAACTTTAACTTAACTTTAACTTCTTCAATGCCTCCCTAT

CAACATATCCATTACATTGCTTTGATCTAATTGCTAATATCTATAATACTCCTAAAAAATATAGTGGTG
MYBSt1 /
OsMYBS1-3BS

AACGACGAAGATCATGTTATGCATCAACAACACGCCACTACCAATCTTCATCGTCCACTATCTACTTTGAA
NACBS

AATACCTCTCAACACAAAACTAAGCAATAACAAAGCACGAGTGCCACTAAAAGAGCGCCTAGACCATAA
NACBS

AAGTATAAAAAAATAACGCATTAAAGAATTAAAAACGATTTGTTTGCTAGTAATTACCTAAAATAGAAAC

GAGTCAAGAGGGAATCTTTTTTGGATTTTTCGAGACAATAATACTATTGTTTTTATATAACTATTTAAATAT
W-box

TTTTCTCT

CTTCGAAAGTGCTTAAAAACGTGATGTTTTTAAGAGGGGGAAGATGGAACGTTGGGAAATGCTGCG
T/G-box

NACBS
GGCAAGTGTAATGGGCGTCGTCAGATCAAAACATGAATATCTGATTAAACACGTGATATGCGCCTGCC
E-box W-box E-box GCC-box

GCCTGACTTACTTCACTCGCTGCCACCTCTATTTAAAGTGGTTCTCCATACCATCGAAGACGTGTCCAA
W-box

CTCAGCATTTGATTTGGCATTTGCAATTTGGAACTCGTAATTTCAGTAGTTCCCCTGCCTCTCTGCAGA
E-box E-box

A

DBAT: 1433 bp

TCCCACAAAACTAGACAAGTCATTTTTTTGGTCTGATTTTTTTGGGTGTTGTTTATATCACCCTCTAGACCAAT
W-box

TTCCACAAACTACCACTAACAAAAGCTTCTAAAGCACATATAAAGACACTGATATCAACATATTTACGTCTCAT

CTTTCTAAACATTTAATTATTTTAAATATATTTCCACAATTTCCCACACTTTCTTTAATATTATATAATTAAGTATG

AACCAATTGTATTTTTTTTTCCAAACCCACTTCCCAATATCTTTGTCCTTACCAAAATTAGTTGTTAACAAGTGG
E-box E-box

AGAATCGAGATGATTAGTCACATTCTTGGGTGTGTGAGTACAATATTAAACAT
W-box

TTCCTTTTAGGTATAAACAAATATTGCTTTAGAATCTTATATCACCCACCCAAATTCCCAAAAAAAACTTCTAT

ATTGTTTTTTGTAATTTCCTAAGGGAAGTGTGCGAAATGCATGTTAAGTATATTTTATTAATTATTTTAAATATA

TTTCCACAATTTCCCACACTTTCTTTAATATTATATAATTAAGTATGAACCAATTGCATTTTTTTTTTCCAAACCC
E-box

ACTTCCCAATATCTTTGTCCTTACCAAAATTAGTTGTCAACAAGTGGAGAATCGAGATGATTAGTCACATTCT
W-box E-box W-box

TGGGTGTGTGAGTACAATATTAAACATTTTCTTTTCTCAACCCAGACAAAACCA
AtMYBBS

TTTTTCATAGATTTGAGGTCTCCACAAAACTAGACAAGTCAATTTTTTGGTTTGATTTTTTTGGGTGT
W-box AtMYBBS

TGTTTATATCACCCTCTAGACCAATTTCACGAACTACCACTAACTAAAGCACATATAAAGACACTGATAT
NACBS

CAACATATTTACATCTCATCTTTCTAAACATTTAATTATTTTAATTATATTTCCACAATTTCCCACACTTTCT

TTAATATTATATAATTAAGTATGAACCAATTGCATTTTTTTTCCAAACCCACTTCCCAATATCTTTGTCCTC
E-box

ACCAAAATTAGTTGTCAACAAGTGGAGAATCTAGATGATTAATCACATTCTTAGGTGTGTGACTACAA
W-box E-box W-box

TATTAAACATTTCCTTTTAGGTATAAACAAATATTGCTTTGTTTTTTGTAATTTCCTAAGGCAAGCGTGCGA

AATGCATGTTAAGTATATTTTATAAAATTTTGACCCACTTTGAAAACTAATTTAAAAAAAAACCAAATAATA
W-box
AtMYBBS

ATATGAATTGACTCGAAAAACAGAAAACCTTTAGACATGTGAGAAATCCACTCATATTGAATTCGACGAA
W-box
E-box

GACAAACCATATTTCATAGAAAACGTGAACAATCAAGCGAAATACAGTCAGCAATTTCCTTTTCCCCTACA
AtMYBBS
T/G-box
W-box

TAAACAGTGATCCTGTACGTAATTTCCATCCTATTCTACAGTTGAAGATTTTCTGAGCTTGATCTGAG
E-box

T5αH: 1611 bp

CTTGTCAGAAGCAGCTCCCTCTCAATAAGCTCAATTCTCGTTGGAGCTTGCATTCCACCCATTTTGGTTGGA
W-box
MYBBS

AACAATGCTTTCAGGTTGACTGGGCCATTCCTATAGAGCCCAAAGTTGCTCACTTCCTTTGGTGTATTCTCCA
W-box

CCAAGGGCTTCCTGTGGGCAAACGCCTCATTCTTATGGGAGCTCCTCCGCTGTGCTGCCCTTTTTGCAACC

ACATTGAAACTCTTCTTCATCTTTTTTGGTTCTGCCCCCACGCTCAGTCTGCTTGGCACTGGATTCATACCCTT
NACBS

NACBS
TTTCAACCTTTTCTGCCCACGTTCTTTACGTGGAAAATGGTTCTGCTGGGGGACCCAGCCTTTGTTCCTGCG
T/G-box

AAATTTTTCAACATTTGGCATTCTTTTCGGGCTACTATTTTATATCG
E-box

CATATGGAAAACTAGAAATTCAGCTATTTTTAACAAAGAGACAGTGGATAGAACTTTTATTATTTCTAACAA
E-box MYBSt1BS

ACTGGCTGTTATTTCCAATGCCTCGCTTCAGATCCAAGTTCAAGCGGAGAAAGTCCGCATTGAGGTGGAA
*At*MYBBS

CACCTGCAGACCCTGCTAGATCACTGGGGAAGCGCCCCTTGCGTTGTCAGTCGTGTTCCGCCTGACTCTC
E-box W-box W-box

CTCATCGCCGCTCCCGTTCCCGTGGGCAACACAATTTTGATCGAGGTCCGGTTCATGGGCAGCGGAGCG
MYBBS RAV1BS

CCTCCCCTCCTCGGCAGGGTCCTTCTCCGCCTCTTCTTGGGCTCCTCCCGCACCAGGGGTGTTTGGCCGG

CTGGGTAGTGGCGGTTCTGGGTGGCAGGTGCCTGATCAATCCAAGTTCCCTCGGC
E-box

TCTCCTTAGCCATCTCGCCCAACCAGCGGGGCACTGAGGAGGAGAAGCCGCAGACTGACGGCTGGACGC
MYBBS W-box

CGATTTCTGACACCCCCGCCAACCCTCTTGGTGGGTGGGGTGACTGTCCGAAGAGCCCCCCGCAATAGCCA
W-box MYBBS W-box

CCTACCTCCATGTTCGCCACTGCCGCCTAGCTCCATGCCTGGCTTGGTTTCTGCAAATTTTGCTCCAACTTTT
MYBBS GCC-box *At*MYBBS

TTTGATTGATAAGTTTATCTTAATATGATTATGTAATATGTACCACTCGTAGCCTTTTTATGGCTACTTTTGCTCT

GCATGTGGGGTGTGATCTTTTCCACCCGATTTTGTATGATTAGCATTGTACATTTCATCATTATTATATGGATAT
E-box MYBSt1BS

GCTTTTCAATCAAAAAAAAAAAAGTTCCTCGACATGTCTCTTTCAATGTATCCATT
*Os*MYBBS

NACBS
GAATTACCCACGTGCAATTGCCAGTTGAAAAATTGTTCAAATATAAAGCTACTTTAAATGATTTATCACATA
E-box E-box E-box

TAAGGAAGATAGCCTTATTAACTGATAATATGGAATCTAACGTTTGAAATGACCTTTGAAAACAGTCCAAG
AtMYB2BS W-box

NACBS
AAAACAATGACTTTTCACGTGGTACCTAGAAACCGTATAAATCCATTATCCTGTGATCCTATATTTTTCGCG
W-box E-box MYBSt1BS

bZIPBS
TGCTACGAACAAAGGCATATAAGTCCAACGACCCTACATTTTTGACGTGGTGCAAACATCTCGGAGTAGA
W-box

AAGTGTTATATTAACTCTTACAATCCATCCATCCATCCATCATATCATCACGGATTTTCTCTGTTCTCTGTTCT

CTTTCTTTTGGTTTCTT
AtMYBBS

DBTNBT: 1898 bp

AGGCTCCGAATTCGCCCTTCAAAGGAGATGGGGAACTGGAAGTGGATTGCACGGGGGAAGGTGCTCTGT
NACBS

TTGTAGAAGCCATGGGTTCATATTCTTAAGTTATATCGTCGAGGCTTGCGCACTCTTAATGACATCTTTTGCA
W-box

GGCAGTCCAAAATGCTTCTCACTTGGCCTCAAGCTAAAGCCAAATTTTTGGCCTGTTTCAACCGACAGAGA
E-box AtMYBBS

GTCTTATCAGGAAATTCTTGATTCCATTCCTGCTGAGATACTTCACAAAGTGGGTATTGTCTCTCTCAGGCCT

TGGTGGAAGGATTGGAAATGGTTTCCAAATACCCCGTTCTTAAATTACTCCCCAAGCAAAGGGTATGCTCTT
MYBBS

CTCACCCCCCGGCTGCCAATGGTGGCTTTGCTAAACTCTAGGTGGT

CCATCAACTCTGGATGGTCCTCTTGGGAAGCTAGATTTCACACAGTTTTGAAGGCCATCCTTCAGCCCAA

GCTAGCATGTTTTTCTTGGTGCCTGTTCTTTCAAGGCCTCCCTATTGGGCGCAAATTAGCGCATTACGGGA

TTGCTCCTCGGCTTTGCCCTTTTTGTGGGCATATTGAGACACTTAGGCATATGTTCTGGTTCTGCCCAAAA
E-box

GCTCAGGGTATTTGGGTCTGGATCCACGATTTCTTTCGGCCATTTCGCCCAGCTCCATTTCTTTGGCAACA
NACBS

TGTTTTGCTTGGGGATTCTATTTCCATTCCACATGCTTTTGCGCAAGTGTGGCATACTTTCAAAGTGGAGA
E-box E-box

TCCTTTTTTTATATGGAAAGGACGCAATTCTGTTGTTTTTGCTAATTCCGCAG
AtMYBBS

TTTCTTGCTTTCTTTCGCATTATGATAAGGCCATGCTCTTCTTGAATGTGTTCTTACAGATCCAGGATCAGGCC

CGCAAGGTTGAGCTGGAAGTGTGGCACCTCCAGACCCTCGTGGATCATTGGGGCCAACGCTCCTTACACT

GTGGCACAGGTGCCCCCCACCTCCCCTTCTAGGAGTGTTCGTTCCAGAGTGTTCGTTCCAGAGGTCATCGT
E-box W-box

GGCCAAGGCCAGTCGAATGAACGTACCCACCATCATCGCTCCCCCAGGTGGCAACCCGAGCATCGGCAGA
E-box

GCCTTGTTAATCAGGGTGGTTAAGGGTCCCCGCAGGGGCCCGTGCATCACAGTGGTTGTGGCTCCTGCCA
AtMYBBS

CCACTCCCCACCTCCTGCGCAATGGGATGGATGGGATTCTGGCAAAGTCCCCGGTTG
AtMYBBS

GCTCTTGCCATATCTCCTAATCTGCAAGACAACAAAGAAGAAGAAGAAGAAGAAGTTGGCACCGG

CTGGCTGCCCCTTCCGGATGGAAATGATCTTCCCAGTGGTAACCTGGAGGATAAAGGGCCTCCACCACC
MYBSt1BS

GCCCCCTGCTGCTTTGTTTGTTGCGGCAGCTTAGTTTGTTTTCCTTATGTTTTGTGGCATTGTTGCCTTTTG

ATACTTTTGTCTCCTAGCCTTCGGCTAATTTGAACGTGAGGTGTGCCTTTTACCAACCATATTTTTTATGGA
T/G-box MYBBS *Os*MYBBS

TAGCTGTAATATTTTTATTAAATAATATAGTCTATTTCATCAAAAAAAAAAATGATATCATTCTTTTTGCTTCTT

GTTGTGCATGCTCTAAATGCACTAATTAATATACCTATAAATGTACAAACTTCTCGCATTCCAA

CAAAAACCCCAAACTCAATCTTAATAACTTTTTAACAAACCAAAACTATAAAATACAATGGAATGAGGATTTT
MYBBS

ATACTCGATTCCATTAATTTCTATTGGAAGCGTGTGAAATCCACTTAGATATTAAAAAAAACAAATAAACATGT
E-box

W-box NACBS
GACTTGAATCAAAAAGCCCACGTGAATAATCCAGCTGAATCCCCTACAAAAATGGATTACCCTACAGCTGAA
E-box E-box E-box

TCCCCATTCTCAGCCCATCGTTTCATCTTTGATCCAGTC