Transcriptome Analysis of *Artemisia annua* Glandular Trichomes and Functional Study of AaWD40 in Arabidopsis

WANG, Wei

A Thesis Submitted in Partial Fulfillment of

the Requirements for the Degree of Doctor of Philosophy

in

Biology

The Chinese University of Hong Kong

September 2010

Thesis Committee:

Professor GUO Dianjing Diane (Supervisor)

Professor JIANG Liwen (Co-Supervisor)

Professor LAM Hon Ming (Chairman)

Professor WONG Yum Shing (Internal Committee)

Professor LI Ning (External Committee)

Statement:

All the works reported in this thesis were performed by the author, unless stated the otherwise.

_____

WANG, Wei

Abstract of thesis entitled

Transcriptome Analysis of *Artemisia annua* Glandular Trichomes and Functional Study of AaWD40

Submitted by Wang Wei

for the degree of Doctor of Philosophy

at the Chinese University of Hong Kong in July 2009

Artemisia annua L. is a common type of wormwood that grows throughout the world. Artemisinin, a terpene compound in *A. annua*, has recently been recognized as the most promising antimalaria drug. Artemisinin and other types of terpenoids are synthesized and accumulated in glandualr trichomes that appear on the surface of leaf, stem and flower bud. To identify new genes involved in artemisinin biosynthesis and trichome function in *A. annua*, a normalized glandular trichome cDNA collection was sequenced by Roche GS FLX pyrosequencing system. Two sequencing runs generated totally 85M nucleotides which were further assembled into 190,377 unigenes (42,678 contigs and 147,699 sigletons). Putative functions were assigned to the unigenes based on Blast search against GeneBank database. Many terpene biosynthesis pathway genes were identified from the pyrosequencing ESTs. Together with other identified *A. annua* terpene pathway genes, a global view of terpene biosynthsis in glandular trichomes of *A. annua* were re-established. Meanwhile, a WD repeat protein, AaWD40, which show high amino acid sequence similarity with its Arabidopsis ortholog, AtTTG1 (AT5G24520) was identified. To investigate the functional relevance of AaWD40 to its Arabidopsis counterpart, genetic complementation test using Arabidopsis mutants was conducted. When AaWD40 was transformed into Arabidopsis *transparent testa glabrous1* (*ttg1-1*) mutant, the anthocyanins and proanthocyanidin (PAs) production in seeds were restored, and the

trichomeless phenotype of *ttg1-1* mutant was rescued. In addition, over-expression of AaWD40 and AtTTG1 modulated the expression of WUS and CLVs genes which are required to maintain the stem-cell niche of Arabidopsis shoot apex. Transcriptomic profiling of transgenic Arabidopsis over-expressing AaWD40, TTG1, or *ttg1-1* mutant revealed lists of genes modulated by these two WD40 genes homologue and gene ontology (GO) analysis suggested that the top-ranked categories are defense, stress response and developmental programme. We hypothesize that WD40 repeat protein act as a crucial regulatory factor in a wide variety of cellular functions in *A. thaliana*.

摘要

青蒿是一種世界範圍內廣泛分布的蒿屬植物。從青蒿中提取的萜類化合物青蒿素爲一種高效安全的抗瘧藥物。青蒿素和其他多種萜類化合物的合成部位位於廣泛分布於青蒿植物葉、莖以及花蕾表面的分泌型腺毛。爲了鑒定青蒿素合成以及腺毛發育相關的新基因，我們使用羅氏 GS FLX pyrosequencing 系統對均一化處理後的青蒿分泌型腺毛 cDNA 進行轉錄組測序。兩輪測序得到合計 85M 核酸序列信息，經過處理拼接得到 190377 條高質量基因序列，其中包含 42678 條 contig 和 147699 條 singleton。通過使用 GeneBank 對預測的基因進行註釋，大量的萜類合成相關基因被發現。綜合其它已被功能鑒定的青蒿萜類合成基因，我們展示了青蒿分泌型腺毛中萜合成的全景。與此同時，我們從青蒿分泌型腺毛中克隆了 AaWD40 (WD40 家族)， AaWD40 與其擬南芥中同源基因 AtTTG1 在氨基酸序列水平上有著很高的同源性。爲了鑒定 AaWD40 的功能，我們將其轉入擬南芥 *ttg1-1* 突變型進行互補試驗，我們發現 AaWD40 可以恢復擬南芥 *ttg1-1* 突變型中的 Proanthocyanidin (PAs) 合成以及回復無纖毛表型。同時我們還發現，在擬南芥中超表達 AaWD40 和 AtTTG1 可造成 WUS 以及 CLVs 等莖尖分生組織相關基因表達異常。通過對 AaWD40 和 AtTTG1 過表達植株以及 ttg1-1 突變型轉錄組分析，我們發現大量與植物防禦、抗逆以及發育有關的基因位於 AaWD40 和 AtTTG1 調控的下遊。 我們認爲擬南芥中 WD40 家族蛋白作爲壹個關鍵調控因子作用於多種下遊調控通路。

Acknowledgements

First of all, my deepest gratitude goes to my supervisors, Prof. Guo Dianjing and Prof. Jiang Liwen, who have given me this precious opportunity to learn and study in the field of plant molecular and cell biology. Their patient instruction, professional advice and constant concern have given me great encouragement during my Ph.D study. Secondly, I am also grateful to Prof. Lam Hon Ming and Wong Yum Shing for serving as my thesis committee members. Not only they were sincerely concerned about my research progress, they also gave many valuable suggestions and comments on my seminar, experiments and thesis. In addition, I'm also grateful to my external examiner, Prof. Ning Li for his precious comments for my thesis.

Special thanks must be given to Qing Zhang and Yejun Wang, who performed microarray and pyrosequencing raw data processing. I also would like to give thanks to all my labmates in G94 and LG101, including Yan Qi, Man Yuan, Yang Liu, Yu Ding, Caiji, Gao, Priscilia Yu and Yiuman Chan for their kind help and technical support.

I also thank Mr. Patrick Lau from the core facility in the Faculty of Science at CUHK for performing the 454 pyrosequencing. The Affymetrix microarray service was provided by the Nottingham Arabidopsis Stock Centre (NASC). The work was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project no. CUHK 4603/06M).

Lastly, my thanks would go to my beloved family for their love and moral support throughout my Ph.D study.

Table of contents

List of Figures

List of Tables

List of Abbreviations

6-BA	6-benzylaminopurine
AaWD40	*Artemisia annua* WD40 repeats protein
A. annua	*Artemisia annua* L.
A. rhizogenes	*Agrobacterium rhizogenes*
A. tumefaciens	*Agrobacterium tumefaciens*
ABC transporters	ATP-binding cassette transporters
ACTs	artemisinin-based combination therapies
ADS	amorpha-4,11-diene synthase
AIDS	acquired immune deficiency syndrome
APS	adenosine 5′ phosphosulfate
ATP	adenosine triphosphate
AtTTG1	*Arabidopsis thaliana TRANSPARENT TESTA GLABRA 1*
bp	base pair
cDNA	complementary DNA
CYP	CYP71AV1
DDT	dichloro-diphenyl-trichloroethane
DMAPP	dimethylallyl diphosphate
DNA	Deoxyribonucleic acid
DOXP	1-deoxy-D-xylulose 5-phosphate
DW	dry weight
E. coli	*Escherichia coli*
EST	Expressed Sequence Tags
FPP	farnesyl diphosphate
FPS	farnesyl diphosphate synthase

GGPP	geranylgeranyl diphosphate
GO	Gene Ontology
GPP	geranyl diphosphate
GST	glandular secreting trichomes
IPP	isopentenyl diphosphate
Kan	Kanamycin
Kb	Kilo base pair
kDa	Kilodalton
Mb	megabase
MEP	methylerythritol phosphate
MVA	mevalonic acid
NAA	naphthaleneacetic acid
NCBI	National Center for Biotechnology Information, USA
PPi	pyrophosphate
SEM	scanning electron microscope
TFs	transcription factors
ttg1-1	*transparent testa glabrous1*

Chapter 1 Introduction

1. The Plant of *Artemisia annua* L.

Artemisia annua (*A.annua*, Asteraceae), also known as Sweet Wormwood, Qing Hao, Sweet Annie, Sweet Sagewort or Annual Wormwood, is an annual herb native to China. It is also widely distributed in western North America, temperate Asia and Eastern Europe. The developed plants often reach more than 2 m in height, usually single-stemmed with alternate branches and leaves. Leaves often range 2.5–5 cm in length. Leaves, stems and flowers contain both 10-celled biseriate trichomes and 5 cell filamentous trichomes. The nodding flowers (capitula), are greenish-yellow enclosed by numerous, imbricated bracts. Capitula are displayed in loose panicles containing numerous central bisexual florets and marginal pistillate florets, the latter extruding stigmas prior to the central flowers. Both flowers have synpetalous tubular corolla with the top split into five lobes in the hermaphroditic florets and 2–3 lobes in the pistillate florets. The receptacle is glabrous, not chaffy, and triangular in shape. Both florets and receptacle bear abundant 10-celled biseriate trichomes; T-trichomes (filamentous) occur at the pedicel and bracts (www.hort.purdue.edu/newcrop/cropfactsheets/artemisia.pdf). *A. annua* is a determinate short-day plant very responsive to photoperiodic stimulus and flower about two weeks after induction. The critical photoperiod is about 13.5 hours. *A. annua* is a diploid plant with chromosome number of 2n=18(Fig. 1.1).

Figure 1. 1 *A. annua* in greenhouse.

2. Malaria and Artemisinin

Malaria is one of the world's most severe parasitic diseases which causes more than one million deaths and 500 million infections each year. Despite the tremendous efforts on disease control, the global morbidity and mortality have not been significantly changed in the last 50 years (Riley, 1995). Because of the emerging resistance of *P falciparum* to conventional antimalarial drugs (quinine and chloroquine), alternative drugs are urgently needed to treat this deadly disease. Artemisinin, a sesquiterpene lactone (Fig.1.2) extracted from *A. annua*, together with Artemisinin-based combination therapy (ACT) has become and the most promising therapy against malaria (Luo and Shen, 1987). Artemisinin and its derivatives are effective against multi-drug-resistant *Plasmodium falciparum* strains mainly in

Southeast Asia and more recently in Africa, without any reported cases of resistance (Krishna et al., 2004). In the treatment of severe malaria, intravenous artesunate acts more rapidly than intravenous quinine in terms of parasite clearance with a reduced mortality (Faiz et al., 2005).

Artemisinin acts rapidly to the asexual stages of *P. falciparum*, the most malignant form of malaria. The biological mode of action of artemisinin has long been controversial. In general, most researchers believe the pecific reaction of artemisinin with TCTP (translationally controlled tumor protein), inhibition of the SERCA (sarco/endoplasmic reticulum Ca2+-ATPase) orthologue (PfATP6) of *P. falciparum*, and inhibition of *P. falciparum cysteine* (Kannan et al., 2005) are responsible for the drug activity. PfATP6 is thought to be the real molecular target of artemisinin in spite of some disagreements (Eckstein-Ludwig et al., 2003). A single amino acid (Leu263) in transmembrane segment 3 of SERCAs is proposed to determine its susceptibility to artemisinin (Jung et al., 2005; Uhlemann et al., 2005). In a recent study, artemisinin was also reported to selectively kill cancer cells. The efficacy of a combined treatment of dihydroartemisinin (DHA artemisinin analog) and butyric acid at low doses in killing cancer cells offered a less toxic, inexpensive, and effective cancer chemotherapy (Singh and Lai, 2005). Most recent data from malaria and yeast study showed that artemisinin directly acts on mitochondria. Specifically, artemisinin and its homologues exhibit correlated activities against malaria and yeast, with the peroxide bridge playing a key role for their inhibitory action in both organisms (Wang et al., 2010).

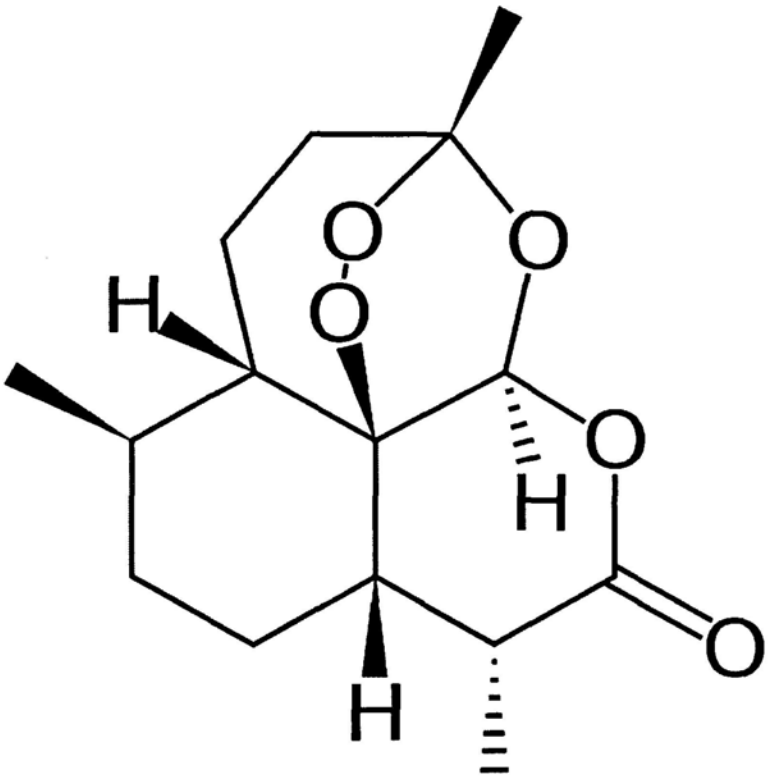


Figure 1. 2 Chemical Structure of Arteminsinin

3. *Biosynthesis of artemisinin*

Because of the structure complexity, the chemical synthesis of artemisinin is not feasible. Understanding the biosynthetic pathway of artemisinin as well as the regulatory mechanism holds the key to successful metabolic engineering for improved artemisinin production.

It is clear that the first committed step in the biosynthesis of arteminsinin is the cyclisation of farnesyl diphosphate (FDP) to amorpha-4,11-diene, the first specific precursor in atremisinin biosynthesis. This sesquiterpene is then oxidised to artemisinic alcohol and artemisinic aldehyde. Artemisinic aldehyde is either reduced at the C-11/C-12 double bond to yield dihydroartemisinic aldehyde, and subsequently

oxidized to dihydroartemisinic acid, or it is further oxidized to yield artemisinic acid. No enzymatic evidence about how artemisinic acid is further converted to dihydroartemisinic acid has been reported. Dihydroartemisinic acid can be converted into artemisinin non-enzymatically (Fig. 1.3) (Covello et al., 2007).

Artemisinin levels reach the highest during pre-flowering or flowering stages for both greenhouse and field conditions (Baraldi et al., 2008; Woerdenbag et al., 1991). In vitro production of artemisinin has been attempted using various tissue culture methods (Hairy roots and callus) (Liu et al., 1999; Weathers et al., 2004), but only trace level of artemisinin was obtained (De Jesus-Gonzalez and Weathers, 2003; Putalun et al., 2007). Qian etc. reported that artemisinin content was significantly enhanced (up to 2 - 3% dry weight) compared to control (1% dry weight) when plants were treated with 4 - 6 g/l NaCl (Qian et al., 2007). Synthetic biology strategy has been proved successful in engineering *Saccharomyces cerevisiae* to produce high titers (up to 100 mg/L) of artemisinic acid using an engineered mevalonate pathway, amorphadiene synthase, and a novel cytochrome P450 monooxygenase (CYP71AV1) from *A. annua* that performs a three-step oxidation of amorpha-4,11-diene to artemisinic acid (Ro et al., 2006).

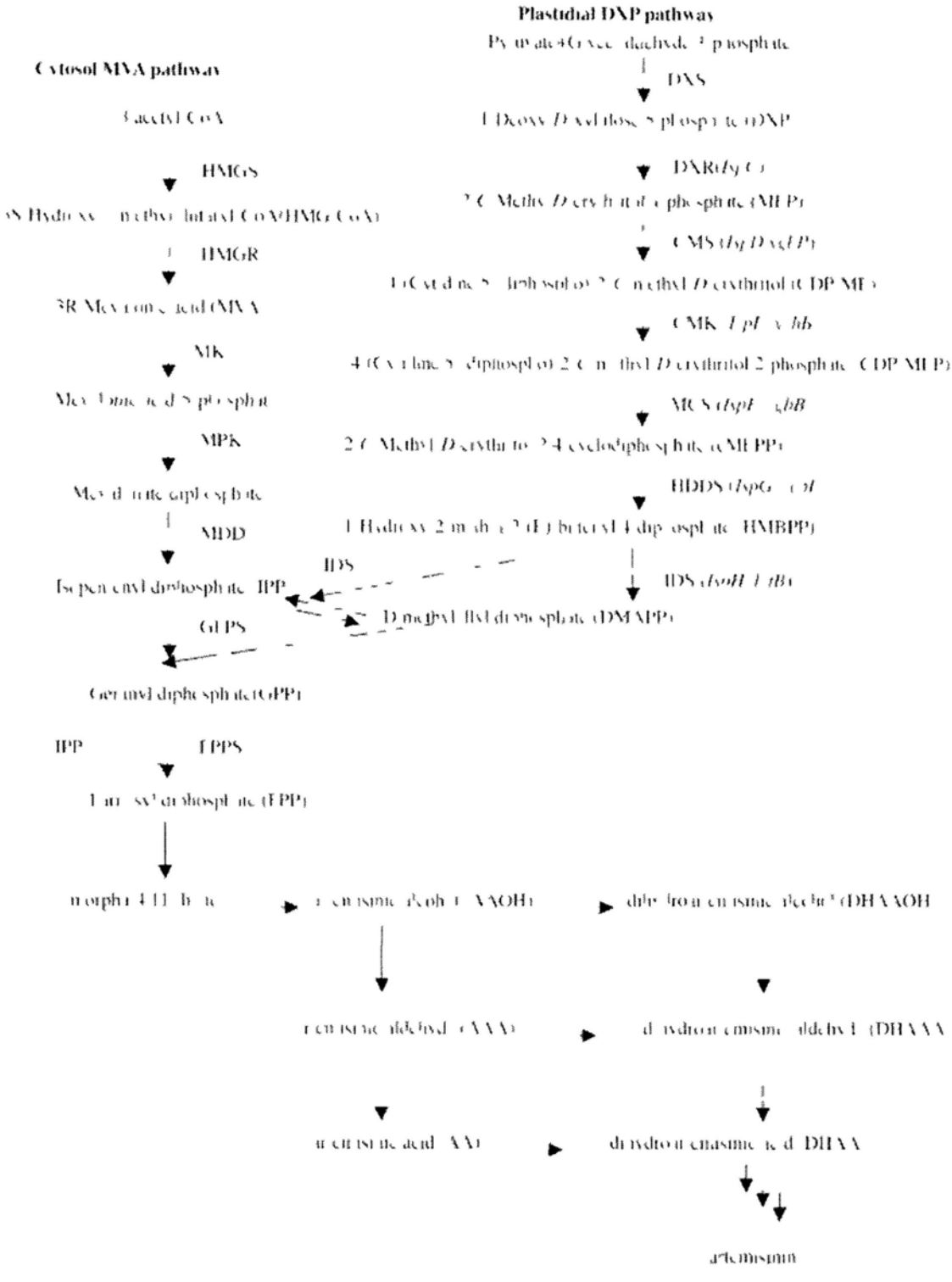


Figure 1. 3 Proposed pathway of artemisinin biosynthesis. Adapted from Bertea et al. (Bertea et al., 2006b).

4. Trichomes in plants

Trichomes (from the Greek τρίχωμα - trikhoma meaning "growth of hair") are specialized epidermal cells present in most aerial plants. Various structures of trichomes have been identified, e.g. unicellular or multicellular, branched or unbranched. In general, two types of trichome exist in plants: glandular trichomes, which contain a stalk terminating in a glandular head, and non-glandular trichome consisting of elongated tapering structures. A common type of trichome is the scale or peltate hair: a plate or shield-shaped cluster of cells attached directly to the surface or borne on a stalk of some kind. Glandular trichomes contain volatile oils and other secondary metabolites produced by the plants. Plants with leaf and stem trichomes have several basic functions or advantages. It is likely that in many cases, hairs interfere with the feeding of some herbivores depending upon the stiffness and irritability to the "palate". In windy locations, trichomes break up the flow of air across the plant surface to reduce evaporation. Dense coatings of trichomes reflect solar radiation, protecting the more delicate tissues underneath in hot, dry, open habitats. In locations where much of the available moisture comes from cloud drip, trichomes appear to enhance this process.

Different plant specials have diverse types of trichomes to perform various functions. For example, in the *Lamiaceae* (Mint Family), there are two types of trichomes commonly found on leaves and stems: peltate and capitate. Epiphytic bromeliads (air plants such as Spanish moss, *Tillandsia usneoides*; Bromeliaceae) absorb water and minerals via foliar trichomes. The glandular trichomes produce and secrete substances such as oils, mucilages, resins, and, in the case of carnivorous plants, digestive juices. Plants growing in soils with high salt content produce salt-secreting trichomes (e.g.,

saltbush, *Atriplex vesicaria*; Amaranthaceae) that prevent a toxic internal accumulation of salt. In other cases, trichomes help prevent predation by insects, and many plants produce secretory (glandular) or stinging hairs (e.g., stinging nettle, *Urtica dioica*; Urticaceae) for chemical defense against herbivores. In insectivorous plants, trichomes participate in trapping and digesting the insects. Prickles, such as those found in roses, are an outgrowth of the epidermis and are an effective deterrent against herbivores.

5. Glandular Trichomes are the specific sites of arteminsinin production and storage

Plants strongly rely on specialized secretary cells to produce and accumulate certain classes of secondary metabolites. Glandular trichomes occur on 20–30% of vascular plants and are believed to be responsible for synthesizing and secreting various types of terpenoids, such as, monoterpenes, sesquiterpenes, diterpene resin acids, etc. Leaves of *A. annua* possess not only glandular trichomes, but also nonglandular, multicellular filamentous trichomes. The glandular trichomes of *A. annua* are more prominent in the corolla and receptacles florets than in leaves, stems, or bracts (Fig. 1.4). The production of artemisinin occurs in specialized 10-celled biseriate glandular trichomes present on the leaves, stems and inflorescences of *A. annua* plants (Duke et al., 1994; Van Nieuwerburgh et al., 2006). Artemisinin content (% DW) is 4 to 11 times higher in the inflorescences as compared to leaves and the presence and development of glandular trichomes in the inflorescences is associated with artemisinin production based on extraction studies. Another evidence is neither artemisinin nor artemisitene can be detected from a glandless biotype. All of the above mentioned biosynthetic enzymes are highly expressed in glandular trichomes

(Bertea et al., 2005; Teoh et al., 2009; Teoh et al., 2006; Zhang et al., 2008), and most probably in the two outer apical cells exclusively (Olsson et al., 2009).

6. The morphogenesis of non-glandular trichome

The protein complex composed of MYB GLABRA1 (GL1) and bHLH GLABRA3 (GL3) transcription factors associated with a WD40 repeat protein TRANSPARENT TESTA GLABRA1 (AtTTG1) regulate multiple cellular differentiation pathways in various of plants (Ramsay and Glover, 2005). Years of studies have proved that this regulation network act in concert to activate leaf trichome initiation and patterning (Bouyer et al., 2008; Morohashi and Grotewold, 2009; Morohashi et al., 2007; Payne et al., 2000; Walker et al., 1999; Zhao et al., 2008), specify root epidermal cell fate and cell patterning (Galway et al., 1994; Schiefelbein, 2003; Zhang et al., 2003), and regulate phenylpropanoid pigment biosynthesis (Carey et al., 2004; Feyissa et al., 2009; Gonzalez et al., 2008; Lloyd et al., 1992; Walker et al., 1999). In vivo binding experiment reveals that, among the three components of the complex, bHLH protein directly binds to the promoter regions of downstream genes. GLABRA3 (GL3) transcription factors binds to the promoters of GLABRA2 (GL2), TRANSPARENT TESTA GLABRA2 (TTG2), CAPRICE (CPC) and ENHANCER OF TRIPTYCHON AND CAPRICE1 (ETC1). However, the binding ability of GL3 to the promoters of these genes requires the presence of the R2R3-MYB factor GL1. Meanwhile, GL3 is recruited to its own promoter in a GL1-independent manner, and this result in decreased GL3 expression, suggesting the presence of a GL3 negative regulatory loop (Morohashi et al., 2007). Morohashi etc. have taken a comprehensive systems approach combining ChIP-chip, candidate gene approaches and genome wide expression analysis to identify genes regulated by WD40-bHLH-Myb Complex.

Novel regulatory functions for GL3 and GL1 were highlighted by the identification of a set of previously unidentified GL3/GL1 immediate direct targets. Some of these target genes express before any of the morphological changes associated with epidermal cell differentiation, suggesting their very early functions in trichome developmental program. Others peak after the first changes are evident, suggesting a need for the progression from trichome initials into mature trichomes. The integration of this information provides a first blueprint for the regulatory network involved in trichome formation (Fig. 1.5) (Morohashi and Grotewold, 2009). The transcription of AtTTG1 is detected in all major organs of Arabidopsis and at all stages of leaf and trichome development (Zhao et al., 2008). Unlike bHLH and MYB proteins which present in a redundant manner, the WD40 (AtTTG1) appear to be irreplaceable in the trimeric complex, suggesting its critical role in downstream regulation.

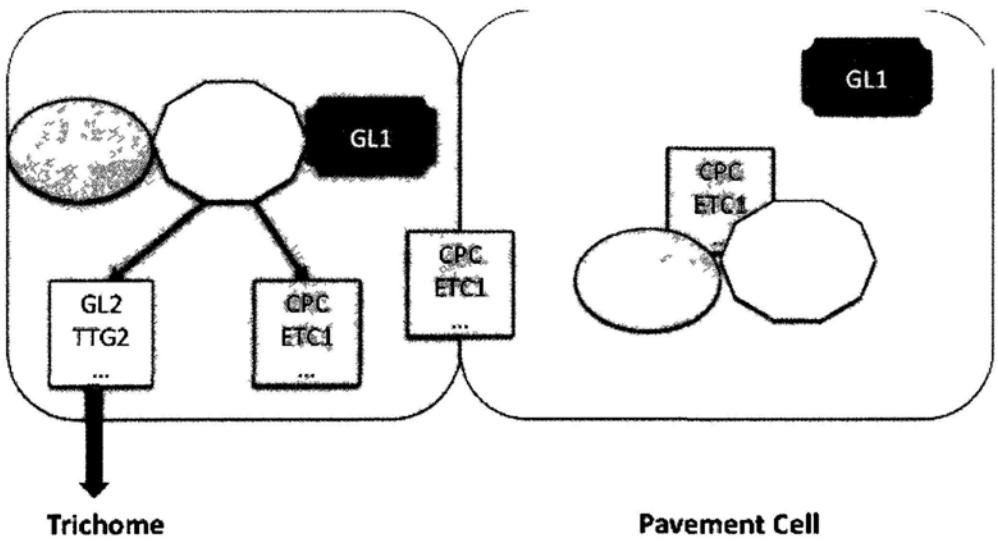


Figure 1. 4 Model for Arabidopsis trichome/non-trichome cell fate specification.

In trichome cells, the inhibitors (CPC, ETC1…) are directly activated by the activating complex and move into neighboring cells, to block the activity of the activating complex together with endogenous inhibitors; thereby the expression of activator (GL2/TTG2) is decreased to below a required initiating threshold level. Thus, the trichome cell fate is not triggered.

7. Objectives and significance

The specific objectives of this study include:

a. Global characterization of glandular trichome transcriptome in *A. annua*,

b. Identify genes potentially related to trichome formation based on the knowledge obtained from Arabidopsis.

c. Investigate the possible function of *A. annua* WD40 repeat gene (AaWD40) using model system Arabidopsis.

d. Clarify the regulatory functions of AtTTG1 like proteins (AtTTG1 and AaWD40).

Although the mechanism of how plants generate the non-glandular trichomes has been well-studied in model plant Arabidopsis, very little is known about the situation in glandular trichomes. It has been reported that phytohormones, including jasmonates, cytokinins, gibberellins and brassinosteroids can modulate the initiation of both glandular trichomes and non-glandular trichomes in Arabidopsis and tomato, indicating the initiation of glandular trichome may share certain degree of common regulatory mechanism with that of non-glandular trichomes. Through the functional study of AaWD40 and its putative Arabidopsis ortholog, we hope to clarify the role of the WD40 repeat protein in regulating trichome initiation. The ultimate goal is to elucidate the regulatory mechanism underlying glandular trichome morphogenesis, so that we can manipulate the density of glandular trichome for increased artemisinin production.

Chapter 2 Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing

1. Background

Secreting glandular trichomes (GTs) are a major site for biosynthesis and accumulation of a wide range of plant natural products. These plant natural products often function to protect the plants against insect predation (Ranger and Hower, 2001; Wagner et al., 2004), and contribute to the flavour and aroma of plants. Many of the natural products also have pharmacological effects, such as the analgesic drug morphine, the anticancer compound taxol, and the antimalarial drug artemisinin.

Artemisinin, a sesquiterpene lactone, is currently recognized as one of the most prominent anti-malarial treatment (Duke and Paul, 1993). A complete understanding of the artemisinin biosynthetic pathway and its regulatory mechanism holds the key to efficient metabolic engineering for increased artemisinin yield. In the past decades, research efforts have been dedicated to identification of enzymes and intermediate compounds leading to artemisinin production. Many genes encoding enzymes participate in the pathway have been cloned and functionally characterized (Bertea et al., 2005; Bouwmeester et al., 1999; Chang et al., 2000; Mercke et al., 2000; Ro et al., 2006; Teoh et al., 2006; Zhang et al., 2008). However, little is known about the regulatory aspects of sesquiterpene metabolism. This is partly due to the fact that *A.annua* is a non-model plant with limited genomic information available, and

sequencing of limited number of randomly selected cDNA clones often have insufficient coverage of less abundant transcripts, including important regulatory transcription factors (TFs). In addition, genes uniquely or preferentially expressed in trichomes may be under-represented in non-tissue-targeted EST sequencing projects. A comprehensive survey of genes expressed in glandular trichome will facilitate new gene discovery and contribute significantly to elucidating the terpenoid pathway regulation and trichome function in *A.annua*. Whole genome or transcriptome sequencing enables functional genomic studies based on global gene expression. The newly developed high throughput pyrosequencing technology allows rapid production of sequence data with dramatically reduced time, labor, and cost (Huse et al., 2007; Margulies et al., 2005; Moore et al., 2006; Weber et al., 2007; Wicker et al., 2006). So far, most applications of pyrosequencing have involved analysis of genomic DNA (Poinar et al., 2006). Published reports on 454 pyrosequencing of transcriptomes have been mostly restricted to model species with genomic or comprehensive Sanger EST data available (Bainbridge et al., 2006; Cheung et al., 2006; Emrich et al., 2007; Weber et al., 2007). Previous studies (Cheung et al., 2006; Weber et al., 2007) using genome or Sanger EST sequences for mapping and annotation of 454 ESTs were not able to accomplish de novo assembly of their 454 ESTs. We here present the global transcriptome characterization of *A.annua* glandular trichome, the so called biofactory for the production of artemisinin and other plant secondary metabolites. We assigned putative function to 28,573 ESTs, including previously undescribed enzymes likely involved in sesquiterpene biosynthesis. We verified the expression of 17 selected ESTs in glandular trichomes using semi-quantitative RT-PCR. These 454 ESTs were linked to metabolic process specific in glandular trichomes and form the basis for further investigation.

2. Materials and Methods

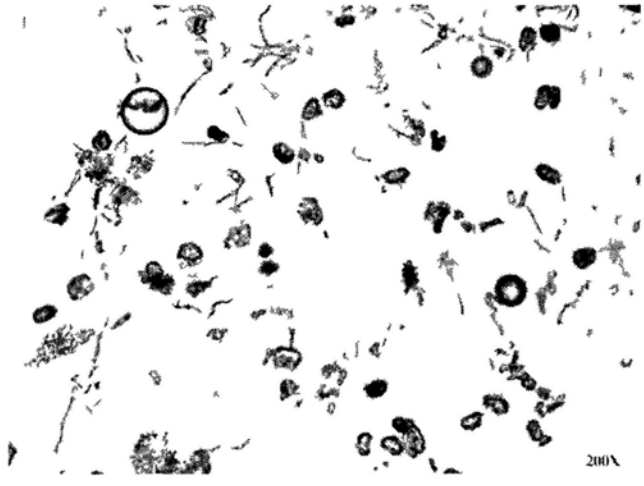
2.1 Plant Materials

A annua seeds were purchased from Youyang, Sichuang province of China. Seeds were sown into commercial potting mixture for germination. The germinated plantlets were grown under natural light conditions in the greenhouse located at The Chinese University of Hong Kong. Flower buds were collected for trichome isolation before flowering.

2.2 Isolation of glandular trichomes

Trichome cells were gently abraded from the surface of flower buds using glass beads and a commercial cell disrupter (BioSpec Products). The isolated secretary cells were separated from other cells and tissue fragments in the mixture by sequentially passing through a 40μm and a 30μm nylon sieves. Glandular cells were finally collected in 30μm meshes with minimum contamination of non-glandular trichomes (Fig 2.5).

A



B

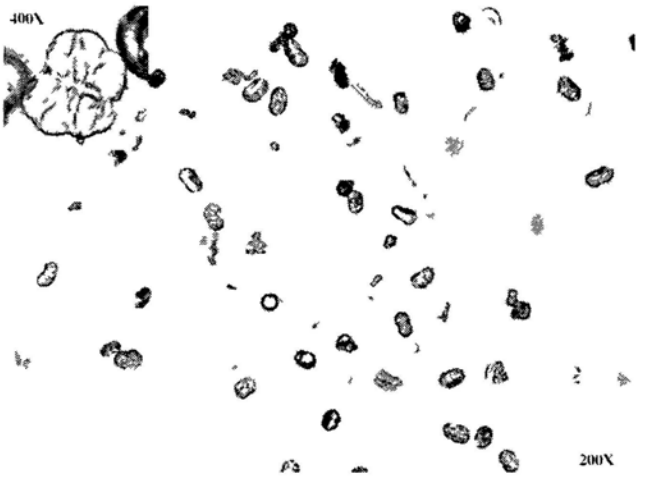


Figure 2. 1 Preparation of glandular trichomes

A: Crude extracts containing non-glandular trichomes. B: Extracted glandular trichomes.

2.3 RNA extraction, cDNA synthesis and normalization

Total RNA was extracted from glandular trichomes isolated from 30 g flower buds following the standard protocol of RNeasy Plant Mini Kit (Qiagen). cDNA was synthesized using the BD SMARTM PCR cDNA Synthesis Kit (Clontech). First-strand cDNA synthesis was performed with oligo(dT) primer as described in the provided protocol using 500ng total RNA. Double-strand cDNA was prepared from 2 μl of the first-strand reaction by PCR with provided primers in a 100μL reaction. cDNA was purified using Qiagen QIAquick PCR purification spin columns. Normalization was performed using TRIMMER cDNA normalization kit (EVRΩGEN) to decrease the prevalence of abundant transcripts before sequencing. Approximately 1 μg of normalized double stranded cDNA was used for 454 pyrosequencing on Roche GS FLX system.

2.4 454 pyrosequencing, data pre-process and assembly

Approximately 1 μg of the adaptor-ligated cDNA population was sheared by nebulization and DNA sequencing was performed following protocols for the

Genome Sequencer GS FLX System (Roche Diagnostic). Reads generated by the FLX sequencer were trimmed of low quality, low complexity [poly(A)] and adaptor sequences using the SeqClean software (http://compbio.dfci.harvard.edu/tgi/). The cleaned sequences were subject to CAP3 program (Huang and Madan, 1999) for clustering and assembly using default parameters.

2.5 Gene annotation using GO terms

After assembly, the resulting contigs and singlets were aligned with NCBI non-redundant protein database using blast2go software with cut-off e-value of 1e-10. The GI accessions of best hits were retrieved, and the GO accessions were mapped to GO terms according to molecular function, biological process and cellular component ontologies (ftp.geneontology.org/).

2.6 Semi-quantitative RT-PCR analysis

To verify the presence of pyrosequencing ESTs in glandular trichomes, we selected 17 contigs for RT-PCR analysis. Total RNA were extracted from glandular trichomes, non-glandular hairy trichomes, leaves and hairy roots respectively. The first-strand cDNA was synthesized from 10μL (about 1μg) total RNA using SuperScript™ II Reverse Transcriptase (Invitrogen) with Oligo(dT)12–18 Primer. PCR was performed using 0.5 to 2 μL of the cDNA in a total of 50 μL reaction volume. The PCR conditions were 2 min at 95°C, 30s at 95°C for 30 cycles, 30s at 47-56°C, 1 min at 72°C, followed by 5 min at 72°C. These conditions were chosen because none of the samples analyzed reached a plateau at the end of the amplification (i.e. they were at the exponential phase of the amplification). Actin was used as a loading control, and loading was estimated by staining the gel with ethidium bromide. Expression analysis

of each gene was confirmed in at least 2 independent RT-reactions using forward and reverse primers.

3. Results

3.1 Sequencing and assembly of 454 pyrosequencing ESTs

Totally 406,044 ESTs (minimal size > 50bp) averaging 210 bp were generated from two consecutive pyrosequencing runs. Cleaning (removal of primer, polyA tail, etc.) of the raw sequences resulted in a total of 386,881 high quality reads with an average length of 205 nucleotides totalling 85 Mb. After clustering and assembly using TGICL CAP3 clustering tools (Huang and Madan, 1999; Pertea et al., 2003), these reads were assembled into 42,678 contigs and 147,699 singletons. The average length for contigs and singletons are 334 bp and 191 bp respectively. The contigs and singletons are collectively referred to as unigenes. The length distribution of unigenes and their component reads are summarized in Table 2.1 and Table 2.2

Table 2. 1 Length distribution of assembled contigs and singletons

Nucleotides length (bp)	Contigs	Singletons
50-99	276	22,730
100-199	2,534	41,936
200-299	19,220	82,169
300-399	11,568	863
400-499	4,940	1
500-599	1,991	0
600-699	980	0
700-799	529	0

800-899	296	0
900-999	142	0
1,000-1,499	173	0
1,500-1,999	22	0
>2,000	7	0
Total	42,678	147,699
Maximum length	2,366 bp	411 bp
Average length	334 bp	191 bp

Table 2. 2 Summary of component reads per assembly

Number of reads	Number of contigs
2 to 10	39,135
11 to 20	3,799
21-30	15,798
31-40	11,817
41-50	5,632
51-100	6,667
101-150	1,321
151-200	430
>200	756

3.2 Pyrosequencing provides deep coverage of the *A.annua* trichome transcriptome

The contigs were searched against the NCBI non-redundant (NR) protein database using the blastx algorithm. Among the 190,377 contigs and singletons, 29,577(15.5%) had at least one significant alignment to existing gene model in blastx searches.

Majority (84.5%) of the pyrosequencing assemblies did not match sequences in the existing gene database and thus likely represent novel transcript sequences identified in this study.

Performing a second sequencing run increased the number of genes identified by approximately 30% (Tab 2.3), suggesting that two pyrosequencing runs detect a substantial fraction of genes expressed in glandular trichomes and provide deep coverage of the *A.annua* trichome transcriptome.

Table 2. 3 Summary of Blast hits from two pyrosequencing runs

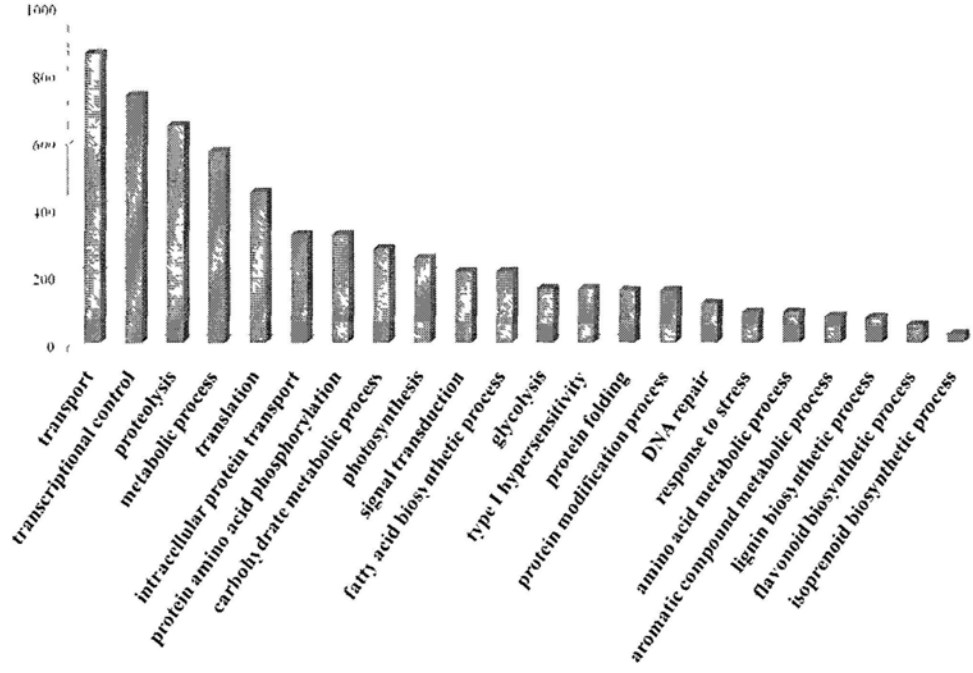
Pyrosequencing run	NCBI database hits
1st run	266,976
2nd run	289,467
Total	357, 843

3.3 Characterization and GO annotation of Novel Transcripts

The 357,843 sequences that had matches in the NCBI protein database could be condensed into 29,577 clusters based on their top protein hits. Each 454 contig was assigned a putative gene description and a GO classification based on the 'best hit' blastx search (bitscore > 45, e-value < 1–10), using the 'inferred from sequence similarity' (ISS) level of evidence (Ashburner et al., 2000) The unigenes were classified into three major functional categories: biological process, molecular function, and cellular component, according to the standard Gene Ontology terms (GO; http://www.geneontology.org). The assigned functionality of genes covers a broad range of GO categories. The top 20 most highly represented GO categories are illustrated in Figure 2.1 The abundance of the gene transcripts suggests their critical roles in glandular trichome metabolism. Under the category of biological process,

transport, transcriptional control, and metabolic process were among the most highly represented categories, indicating the important metabolic activities in *A.annua* glandular trichomes. Other categories include photosynthesis, secondary metabolism (lignin, flavonoid, and isoprenoid biosynthesis process) and primary metabolism (fatty acid, glycolysis, carbohydrate process etc.).

Figure 2. 2 Top-ranked functional categories of pyrosequencing ESTs



3.4 Comparison of 454 sequence contigs to trichome ESTs from other plant species

TrichOME is a publicly available database of genes and metabolites expressed in plant trichomes. It currently contains 37,017 conventional ESTs derived from 8 plant species, including *Medicago sativa*, *Humulus lupulus*, *Mentha x piperita*, *Nicotiana benthamiana*, *Ocimum basilicum, Solanum habrochaites, Solanum lycopersicum* and

Solanum pennellii. A tblastx search against Trichome DB showed that 17,372 (9%) of our 454 sequences had best blast hits (e-value < 1e-10) to 8,095 EST clusters with unique descriptions. Thus 454 sequencing has revealed many transcripts not previously detected in *A.annua*. This suggests that deep pyrosequencing using 454 pyrosequencing technology can identify a larger number of expressed sequences than conventional EST sequencing. ESTs homologous to photosynthesis-related proteins (chlorophyll a/b binding protein, ribulose bisphosphate carboxylase small subunit) are among the top 10 most highly expressed transcripts. The top ranked common molecular function of ESTs identified from all 9 plant species are listed in Table 2.4. Regulation of metabolic process, metabolic process, oxidation reduction, and transport categories have the highest number of contigs. Trichomes are known to be active in photosynthesis, as well as for their roles in storage and secretion of toxic compounds e.g. heavy metals (Choi et al., 2001; Kupper et al., 2000), which requires the function of transporters. In our pyrosequencing EST collections, we identified a large number of contigs homologous to ABC transporter, which is one of the most important families of membrane transport proteins that may play critical roles in the transmembrane transport of secondary metabolites in plants.

Table 2. 4 Shared GO terms (biological process) in all trichome EST databases

GO ID	GO term	No. of *A.annua* contigs
GO:0006464	positive regulation of protein metabolic process	147
GO:0006730	metabolic process	36
GO:0008152	positive regulation of metabolic process	28
GO:0055114	oxidation reduction	21
GO:0006006	glucose metabolic process	21
GO:0006334	nucleosome assembly	16
GO:0006412	positive regulation of biosynthetic process	14

GO:0006096	positive regulation of glycolysis	5
GO:0006810	transport	5
GO:0006869	positive regulation of lipid transport	2

3.5 Representation of genes related to secondary metabolism

Numerous sesquiterpene and monoterpene compounds have been identified in *A. annua* leaves, stems (Bouwmeester et al., 1999; Ma et al., 2007; Ma et al., 2008) and isolated glandular trichomes (Bertea et al., 2006a) The genes corresponding to enzymes involved in the biosynthesis of major sesquiterpenes have been cloned and characterized (Bertea et al., 2006a; Bouwmeester et al., 1999; Cai et al., 2002; Hua and Matsuda, 1999; Jia et al., 1999; Matsushita et al., 1996; Mercke et al., 1999; Picaud et al., 2005). To investigate the trichome function in secondary metabolism, the annotated unigenes were searched for enzymes participate in terpenoids biosynthesis. As shown in Table 2.5, pyrosequencing ESTs corresponding to all the known enzymes in the terpenoids MEP and MVA pathway were identified.

Table 2. 5 ESTs encoding putative enzymes of terpenoids metabolism

	Enzymes	ESTs	No.of hit
MEP pathway	1-deoxy-D-xylulose 5-phosphate synthase (DXS)	Contig26863, Contig5078, Contig9122, Contig9122, Contig18267, Contig27338, Contig29817, Contig15938, Contig13588, Contig18365, Contig33752, Contig688	129
	1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXP)	Contig3726, Contig30566, Contig40163, Contig5373, Contig22259, Contig5238, Contig9143	92
	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCT)	Contig23991, Contig27774	67
	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK)	Contig4107, Contig13278, Contig22091, Contig26285, Contig33971	56
	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS)	Contig5224, Contig23617, Contig23991, Contig27774, Contig28159	66
	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (HDS)	Contig16829, Contig1645	46
	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (HDR)	Contig10323	13
	isopentenyl diphosphate/ dimethylallyl diphosphate isomerase (IDI, also involved in MVA pathway)	Contig2745, Contig28949	70
	isopentenyl diphosphate/ dimethylallyl diphosphate synthase (IDS)	Contig2173, Contig38106, Contig22643, Contig3323, Contig3077, Contig3261	76
MVA pathway	acetoacetyl-coenzyme A thiolase (AACT)	Contig1372, Contig6398, Contig7766, Contig8730, Contig13733, Contig16680, Contig28687, Contig35350	44
	3-hydroxy-methylglutaryl coenzyme A synthase (HMGS)	Contig22601, Contig22850, Contig35313	20
	3-hydroxy-methylglutaryl coenzyme A reductase (HMGR)	Contig16434, Contig8757, Contig17100, Contig34709, Contig25845	43
	Mevalonate kinase (MK)	Contig1666, Contig7660, Contig8060, Contig24269	18
	phosphomevelonate kinase (PMK)	Contig40618	3
	mevalonate disphosphate decarboxylase (MDC)	Contig8855, Contig11288, Contig15468, Contig32395, Contig42566	36
Sesquiterpene biosynthetic pathway	famesyl diphosphate synthase (FDPS)	Contig8404, Contig16347, Contig684, Contig32292, Contig39117, Contig1599, Contig9587, Contig6625, Contig15762, Contig17434, Contig29080, Contig31705	240
	Undefined sesquiterpene synthase	Contig5701, Contig8435, Contig15269, Contig15156, Contig5703, Contig33181, Contig40578, Contig4268, Contig6795, Contig11547, Contig12831, Contig14375,Contig15398, Contig27432, Contig41183	519
	amorpha-4,11-diene synthase	Contig5974, Contig9665, Contig1130, Contig13788, Contig14375	329
	valencene synthase	Contig11535	3
	delta-cadinene synthase	Contig5569, Contig7798	24
	germacrene A synthase	Contig11184, Contig38698, Contig40427, Contig17763	44
	germacrene D synthase	Contig4645	21
	beta-caryophyllene synthase	Contig4071	18
	beta-farnesene synthase	Contig6571, Contig8230	25
	5-epi-anstolochene synthase	Contig17572	19
	8-epicedrol synthase	Contig424, Contig27643, Contig42482	60
Monoterpene biosynthetic pathway	geranyl pyrophosphate synthase	Contig17179	5
	geranylgeranyl-diphosphate synthase	Contig9896, Contig11304, Contig6374, Contig41173, Contig17179	62
	gamma-terpinene synthase	Contig10635	30
	delta-limonene synthase	Contig7080, Contig16521	13
	nerolidol/linalool synthase	Contig33252	6
	(3R)-linalool synthase	Contig1594, Contig3219, Contig5497, Contig7532, Contig15497, Contig15888, Contig23791, Contig23997, Contig34392	163
	undefined monoterpene synthase	Contig10107	16
	alpha-terpineol synthase	Contig6442, Contig6748, Contig13975	21
	beta-pinene synthase	Contig5154, Contig5299, Contig9113	55
Other	dammarenediol-II synthase	Contig8149	2
	undefined terpene synthase	Contig11946, Contig26547,Contig35096	36

Unigenes encoding putative enzymes in terpenoids metabolism. The data represent all the unigenes encoding putative enzymes in terpenoids metabolism.

In higher plants, terpenoids precursor isopentenyl diphosphate (IPP) can be produced from both MVA and MEP routes, which is then converted to its isomer DMAPP (Fig 2.2) (Lichtenthaler et al., 1997). The cytosolic MVA terpenoids pathway, which starts from acetyl-CoA and proceeds through the intermediate mevalonate (MVA), provides the precursors for sterols and ubiquinone (Disch et al., 1998). The plastidial MEP pathway, which involves a condensation of pyruvate and glyceraldehyde-3-phosphate, is used for the synthesis of isoprene, carotenoids, abscisic acid, and the side chains of chlorophylls and plastoquinone (Arigoni et al., 1997; Hirai et al., 2000; Schwender et al., 1997). Although the subcellular compartmentation allows both pathways to operate independently, there is ample evidence that cross-talk exist between these two pathways (Laule et al., 2003).

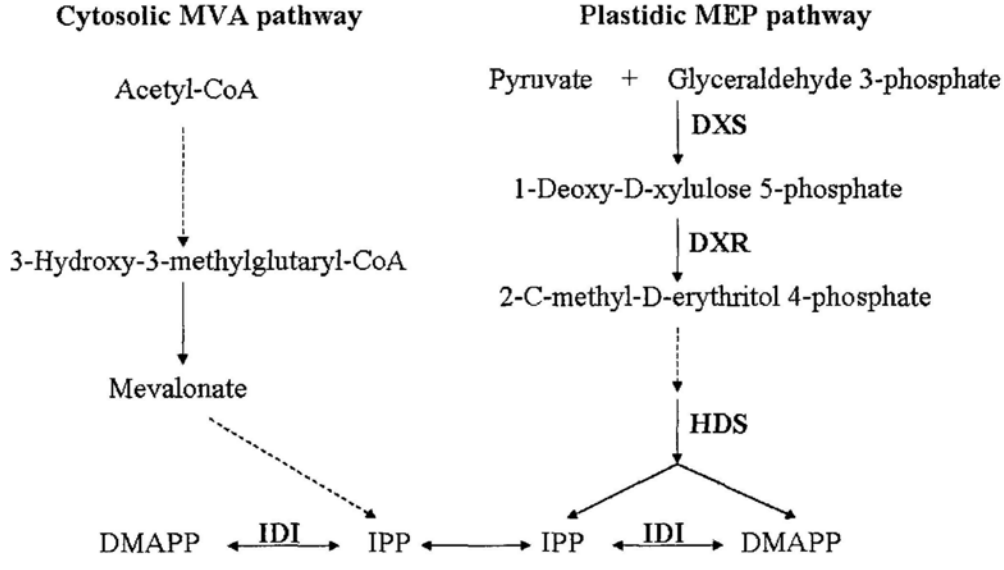


Figure 2. 3 Simplified graphical representation of terpenoid biosynthetic pathway DXR: deoxy-D-xylulose 5-phosphate synthase; DXP: 1-deoxy-Dxylulose-5-phosphate reductoisomerase; HDS: 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase; IDI: isopentenyl diphosphate/dimethylallyl diphosphate isomerase. DMAPP: Dimethylallyl Diphosphate. IPP: isopentenyl diphosphate.

Unigenes encoding the MEP and MVA pathway enzymes and all the sesquiterpene

artemisinin pathway enzymes were present in our pyrosequencing collection. It is noteworthy that although the sequences were derived from normalized cDNA collections, ESTs corresponding to MEP pathway enzymes were two fold more abundant as compared with MVA pathway transcripts. This likely suggests that the MEP pathway may serve as a major route for DMAPP/IPP production in the *A.annua* trichomes. The MEP pathway has previously been shown to provide precursors for both mono-and sesqui-terpene biosynthesis in snapdragon flowers (Dudareva et al., 2005). In a recent report on hops, the ESTs encoding MEP pathway enzymes are also found more abundant than those of MVA pathways (Wang et al., 2008).

Except for those well characterized terpenoid pathway genes, other ESTs annotated as sesquiterpene synthase and monoterpene synthase were identified. Three ESTs (Contig02039, Contig16267, Contig14765) annotated as sesquiterpine synthases were selected for RACE PCR to retrieve the full length cDNAs (data not show). Sequence analysis indicated that the conserved sesquiterpene synthase functional domain exists in all three genes (Fig 2.3).

Furthermore, large amount of ESTs annotated as phenylpropanoids and flavanoids pathway enzymes were present in the pyrosequencing EST collection (Tab. 2.4), indicating the metabolic function of glandular trichomes in *A.annua* secondary metabolism.

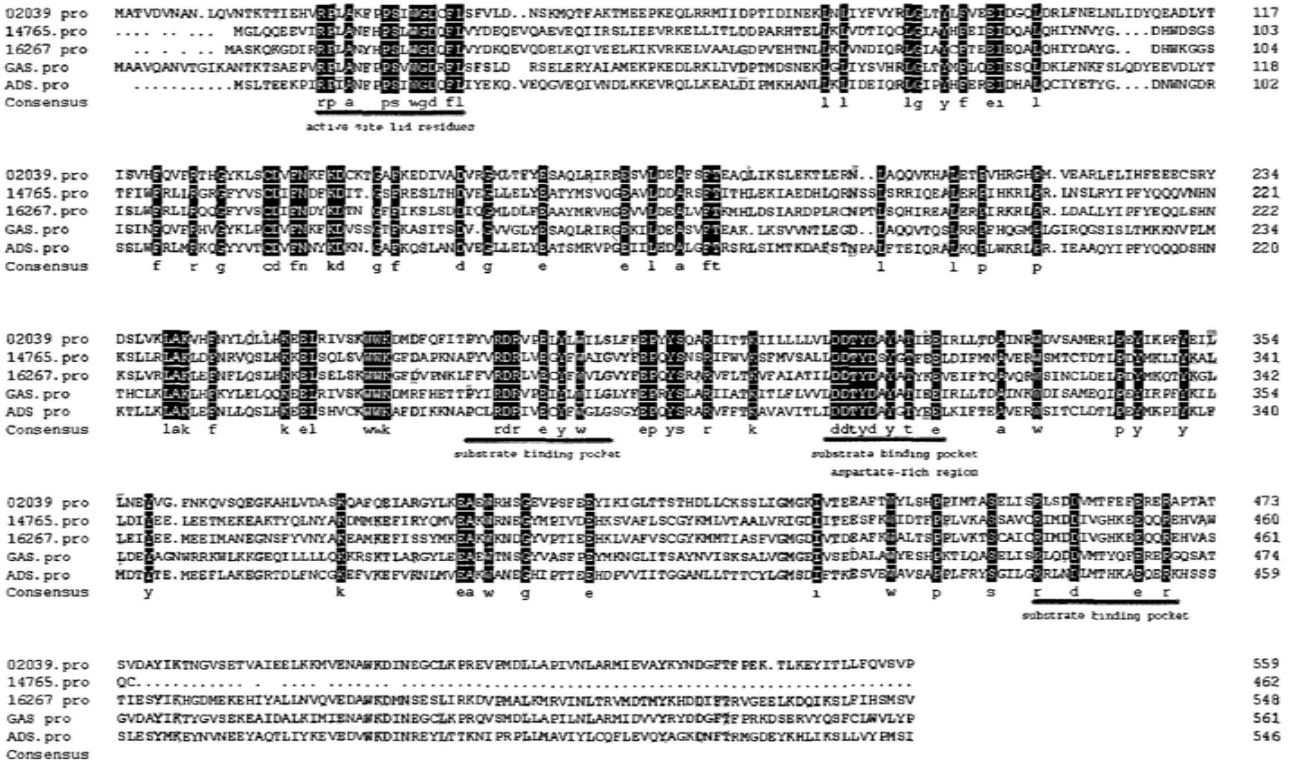


Figure 2. 4 Alignment of putative sesquiterpene synthases with homologs from other plant species (accession no. DQ447636 andAY006482). Identical amino acids are highlighted. The functional motifs are underlined.

Table 2. 6 Unigenes annotated as phenylpropanoids and flavanoids pathway enzymes presented in assembled pyrosequencing EST collection. The data represent all the unigenes annotated as phenylpropanoids and flavanoids pathway enzymes.

GenBank ID	Unigene	Length	BLAST hit	E value
CAL91169	Contig7123	282	phenylalanine ammonia-lyase 3 [Cynara scolymus]	0.0012+000
AAR31107	Contig7654	330	phenylalanine ammonia-lyase [Quercus suber]	0.0012+000
CAJ43711	Contig14494	503	phenylalanine ammonia lyase [Plantago major]	0.0012+000
BAF36972	Contig24666	394	phenylalanine ammonia-lyase [Lotus japonicus]	0.0012+000
CAL91169	Contig31045	684	phenylalanine ammonia-lyase 3 [Cynara scolymus]	0.0012+000
S60043	Contig35786	275	phenylalanine ammonia-lyase[Populus sieboldii x Populus grandidentata]	0.0012+000
O23865	Contig41492	326	Phenylalanine ammonia-lyase 1[Daucus carota]	0.0012+000
ACE95171	Contig3097	322	cinnamate-4-hydroxylase [Populus tomentosa]	0.0012+000
BAB71717	Contig4587	270	cinnamic acid 4-hydroxylase [Lithospermum erythrorhizon]	0.0012+000
Q43240	Contig5605	262	Cinnamic acid 4-hydroxylase[Zinnia violacea]	0.0012+000
P37115	Contig22082	357	Cinnamic acid 4-hydroxylase[Vigna radiata var. radiata]	0.0012+000
AAV36374	Contig23192	419	cinnamate 4-hydroxylase [Pinus taeda]	0.0012+000
ACB78016	Contig28778	254	cinnamic acid 4-hydroxylase [Trifolium pratense]	0.0012+000
AAP03018	Contig2934	665	4-coumarate-CoA ligase-like protein [Arabidopsis thaliana]	0.0012+000
ACC63869	Contig3493	318	4-coumarate:CoA ligase [Populus trichocarpa]	0.0012+000
AAF91309	Contig18394	181	4-coumarate:coA ligase 2 [Rubus idaeus]	0.0012+000
BAB89961	Contig19196	252	putative 4-coumarate:CoA ligase [Oryza sativa Japonica Group]	0.0012+000
NP_192425	Contig26707	372	4-coumarate:CoA ligase[Arabidopsis thaliana]	0.0012+000
ABV44809	Contig31054	245	4-coumarate coenzyme A ligase [Eriobotrya japonica]	0.0012+000
CAJ41420	Contig31194	421	4-coumarate-CoA ligase-like protein [Coffea arabica]	0.0012+000
AAP03021	Contig33156	435	4-coumarate-CoA ligase-like protein [Arabidopsis thaliana]	0.0012+000
AAP03018	Contig37043	331	4-coumarate-CoA ligase-like protein [Arabidopsis thaliana]	0.0012+000
AAM65672	Contig39422	358	4-coumarate-CoA ligase-like protein [Arabidopsis thaliana]	0.0012+000
ABC71308	Contig1057	502	chalcone synthase [Saussurea medusa]	0.0012+000
ABC71308	Contig2946	347	chalcone synthase [Saussurea medusa]	0.0012+000
AAO43482	Contig3955	387	chalcone synthase [Camellia yunnanensis]	0.0012+000
ABC71308	Contig4235	241	chalcone synthase [Saussurea medusa]	0.0012+000
BAB40787	Contig6156	705	chalcone synthase [Lilium hybrid division I]	0.0012+000

CAK19318	Contig23652	766	naringenin-chalcone synthase [Humulus lupulus]	0.0012+000
Q9ZU06	Contig28621	409	Naregenin-chalcone synthase[Persea americana]	0.0012+000
ABO42635	Contig29505	341	putative chalcone synthase [Euryops virgineus]	0.0012+000
ABC71308	Contig39546	534	chalcone synthase [Saussurea medusa]	0.0012+000
BAB40786	Contig33429	499	chalcone synthase [Lilium hybrid division I]	0.0012+000
AAY96729	Contig35933	409	chalcone synthase family protein [Arabidopsis thaliana]	0.0012+000
CAA11226	Contig11726	534	chalcone reductase [Sesbania rostrata]	0.0012+000
AAT94362	Contig29637	605	putative chalcone isomerase 4 [Glycine max]	0.0012+000
Q08704	Contig24736	643	Chalcone-flavonone isomerase[Zea mays]	0.0012+000
A1E260	Contig14484	264	Chalcone-flavonone isomerase 1 [Chrysanthemum x morifolium]	0.0012+000
A1E261	Contig4615	443	Chalcone-flavonone isomerase 2 [Chrysanthemum x morifolium]	0.0012+000
CAA48775	Contig3607	516	chalcone isomerase [Malus sp.]	0.0012+000
Q8LKP9	Contig41267	446	Chalcone-flavonone isomerase[Saussurea medusa]	0.0012+000
AAC15414	Contig3240	470	flavanone 3-hydroxylase [Nicotiana tabacum]	0.0012+000
BAE16364	Contig6420	331	flavanone 3-hydroxylase [Triticum aestivum]	0.0012+000
AAM00948	Contig2599	326	Putative flavonoid 3'-hydroxylase [Oryza sativa]	0.0012+000
ABN79672	Contig7107	838	flavonol synthase [Rudbeckia hirta]	0.0012+000
Q9ZWQ9	Contig9338	325	Flavonol synthase/flavanone 3-hydroxylase (CitFLS) (FLS)	0.0012+000
BAC10995	Contig13897	454	flavonol synthase [Nierembergia sp. NB17]	0.0012+000
Q9M547	Contig26275	381	Flavonol synthase/flavanone 3-hydroxylase (FLS)	0.0012+000
AAC15414	Contig32102	508	flavanone 3-hydroxylase [Nicotiana tabacum]	0.0012+000
ABF21085	Contig41956	308	flavanone 3-hydroxylase [Chrysanthemum x morifolium]	0.0012+000
NP195268	Contig6274	1241	dihydroflavonol 4-reductase	0.0012+000
AAM63604	Contig5721	773	putative anthocyanidin synthase [Arabidopsis thaliana]	0.0012+000
AAM63604	Contig12629	563	putative anthocyanidin synthase [Arabidopsis thaliana]	0.0012+000
AAM63604	Contig28588	346	putative anthocyanidin synthase [Arabidopsis thaliana]	0.0012+000

3.6 RT-PCR validation

A set of 17 contigs was selected for semi-quantitative RT-PCR analysis to confirm the expression of novel transcripts detected among the 454-ESTs (Fig 2.4). These selected contigs encode enzymes involved in artemisinin biosynthesis, and putative transcription factors. PCR experiments were conducted on four pools of cDNAs derived from (1) glandular trichomes, (2) non-glandular trichomes (3) leaves, and (4) roots. The results demonstrate that all of the novel transcripts detected among the 454-ESTs are indeed expressed in glandular trichomes, including those with low expression levels. This suggests that deep pyrosequencing is effective in revealing the expression of many rare transcripts, e.g. transcription factors, which provides further evidence for the value of tissue specific 454 pyrosequencing for gene discovery. Most of the tested genes were also expressed in leaf and non-glandular trichome cDNA pools, except for one contig40477, which was only expressed in glandular trichomes and roots. Interestingly, three tested ESTs likely encode enzymes needed for sesquiterpene biosynthesis were also strongly expressed in non-glandular type of trichomes. This raises the question as to whether glandular trichome is the sole site for the biosynthesis of artemisinin and other sesquiterpenes in *A.annua*.

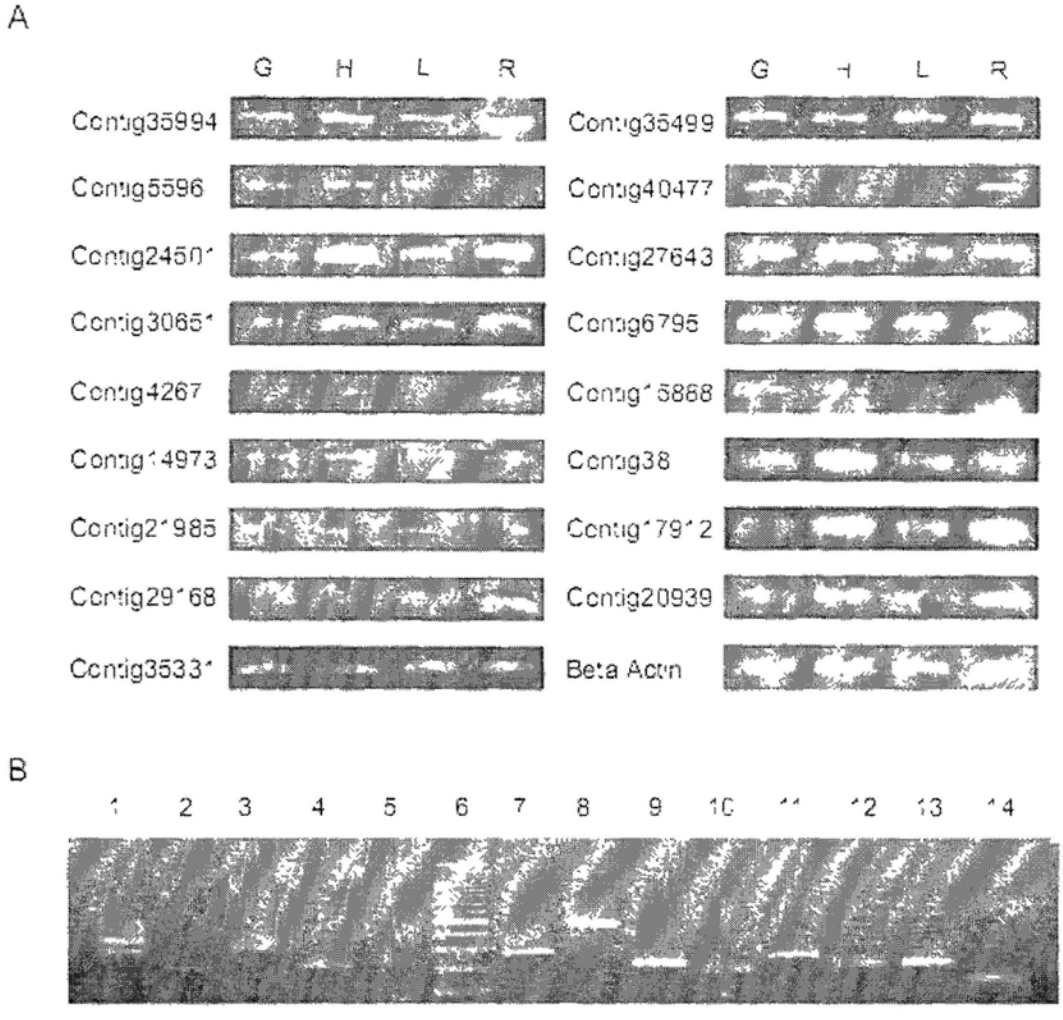


Figure 2. 5 Semi-quantitative RT-PCR analysis of selected ESTs

A. Expression of selected contigs in different tissue types. G: Glandular trichome, N: Non-glandular tirhcome, L: Leaf, R: Root. bHLH family proteins: Contig35994 and Contig5596; WD family proteins: Contig24501 and Contig30651; Myb family proteins: Contig14973, Contig21985, Contig29168, Contig35331, Contig35499, and Contig40477; Terpene synthases: Contig27643; amorpha-4,11-diene synthase [*Artemisia annua*] ABM88787: Contig 6795; sesquiterpene cyclase [*Artemisia annua* | AAG24640: Contig15888 (3R)- linalool synthase [*Artemisia annua*] AAF13356; WRKY Proteins: Contig38, Contig17912, Contig20939. B. Expression of novel transcripts and singletons in GT. Lane 1-5: singletons S122859, S078690, S091943, S154166, and S174533; Lane 6: DNA marker; Lane 7-14: novel transcripts C3719, C13021, C15708, C1441, C20920, C29103, C445, and C14916

4. Discussion

As the sole plant source for artemisinin production, the *A.annua* has been studied extensively for the past decades. Like most other non-model plant species, it has lacked genetic and genomic resources necessary for mechanistic study. Although a precise estimate of transcriptome coverage is unattainable without full genomic sequence, we appear to have recovered a significantly portion of the *A.annua* glandular trichome transcriptome. Novel transcripts detected highlights the hypothesis-expanding aspects of 454 deep pyrosequencing approach, which potentially facilitate the understanding of glandular trichome metabolic function. The assembled sequence data also provided a rich source of information for further investigation.

Two consecutive pyrosequencing runs identified a large number of genes expressed in glandular trichomes. In data analysis, approximately 85% of the pyrosequencing ESTs did not align to any ESTs available in GenBank. This high proportion could reflect the specialized cell type that was sampled or perhaps the greater complexity of the *A.annua* genome. Because our priority goal in this study is gene discovery, we therefore chose normalized cDNA population to reduce oversampling of abundant transcripts and to maximize coverage of less abundant transcripts present in the

sample. The average contig length was fairly short (~334 bp), and only 62% of the sequence reads assembled into contigs, leaving 147,699 singletons. The currently available software tools have problems with assembly of a large numbers of short sequences generated by pyrosequencing, largely due to the short overlaps. Improvement in software will greatly facilitate assembly of full-length cDNA sequences.

Genes involved in plant secondary metabolism have frequently been identified by EST approach (Bao et al., 2002). The lower cost and greater sequence coverage offered by pyrosequencing makes it possible to identify more candidate genes involved in plant natural product biosynthetic pathways, esp. those with low abundance and often missed by conventional EST projects. For non-model species with little or no genomic data available, such as *A.annua*, pyrosequencing offers more great advantage. The ability to rapidly characterization of a large portion of the transcriptome can provide a comprehensive tool for gene discovery. However, one limitation of pyrosequencing is that one must rely on RACE PCR in order to obtain full-length sequence data for a given gene of interest.

Comparison between our glandular trichome 454 ESTs with conventional ESTs generated from trichomes of other plant species revealed likely common function in non-glandular and glandular trichomes. In addition, some ESTs corresponding to enzymes in sesquiterpene biosynthesis were found to be highly expressed in both trichome types in our RT-PCR analysis. Although it has been suggested that glandular trichomes are the site for synthesis and accumulation of plant secondary metabolites, it will be interesting to further investigate whether non-glandular trichomes also contribute to artemisinin biosynthesis in *A.annua*.

5. Summary

We conducted a global analysis of glandular trichome in *A.annua* using massively parallel pyrosequencing. Mining the pyrosequencing ESTs resulted in the identification of many unigenes likely involved in terpenoid biosynthesis and trichome function. Functional characterizations of selected genes are being carried out. These pyrosequence data form the basis for further characterization of the molecular basis of glandular trichome metabolism in *A.annua*. The results also highlight the value of using high throughput pyrosequencing technnology for gene discovery in non-model plants. Access to all EST data obtained in this study is facilitated through

an Excel workbook available in the supplemental data.

http://www.biomedcentral.com/1471-2164/10/465/additional/

Chapter 3 An *Artemisia annua* WD40-repeat gene Regulates Multiple Cellular Functions in Arabidopsis

1. Background

WD repeat (WDR) protein is characterized by 4-16 tandem WD (also called Trp-Asp or WD-40) motifs, which comprise 44–60 residue sequence that typically delineated by the Gly-His (GH) dipeptide 11–24 residues from its N-terminus and ending in Trp-Asp (WD) dipeptide. In between GH and WD locates a core sequence that exhibits only limited amino acid sequence conservation (Smith et al., 1999). The WD-repeat proteins are found in all eukaryotes but not in prokaryotes. They are implicated with a wide range of diverse functions, such as, cell division, cell-fate determination, signal transduction, cytoskeleton dynamics, protein trafficking, nuclear export, RNA processing and chromatin modification (van Nocker and Ludwig, 2003). Currently, it is believed that WDR proteins provide binding sites and foster transient interactions for other proteins. In some cases, WDR proteins act as integral component of protein complexes or modular interaction domain of larger proteins by bringing the protein and associated ancillary domain(s) into proximity of its target(s) (van Nocker and Ludwig, 2003). One of the most well studied WDR proteins is TRANSPARENT

TESTA GLABRA1 (AtTTG1), which regulates several developmental and biochemical pathways, including the formation of root hair, differentiation of trichome and the production of seed mucilage and anthocyanin pigments in Arabidopsis (Galway et al., 1994; Gonzalez et al., 2008; Walker et al., 1999). A glabrous phenotype on leaf surfaces or stem base is observed in Arabidopsis *ttg1* mutants. Other phenotypes include: absence of purple anthocyanin pigments in all tissues, absence of ruthenium red-staining seed mucilage, yellow seeds due to the absence of brown pigment in seed coat, and abnormal cell appearance of dry seed coat under scanning electron microscope (Galway et al., 1994; Zhao et al., 2008). It is believed that Arabidopsis TTG1 and the R2R3-MYB transcription factor GLABRA1 (GL1) (Oppenheimer et al., 1991) bind simultaneously to the basic helix-loop-helix (bHLH) proteins GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) (Payne et al., 2000; Zhang et al., 2003) to form a trimeric regulatory complex, thereby activating the transcription of downstream genes controlling epidermal differentiation and anthocyanin biosynthesis. More recent study found that *ttg1-1* mutants tend to have more inflorescences compared to the wild type (Buer and Djordjevic, 2009), indicating TTG1 may also regulate genes involoved in shoot development. Although one possible explanation is that the perturbation of flavonoids may directly affects auxin movement and consequently cause the developmental abnormality (Peer and

Murphy, 2007), it is still unclear whether flavonoids act directly as regulatory agents or indirectly through auxin accumulation or movement, or alternatively, TTG1 may regulate the shoot development directly.

Several WD40 repeat proteins that control epidermal cell fate identity and anthocyanin biosynthesis have been identified from Arabidopsis, Petunia, Perilla, Zea mays, cotton, Japanese morning glory, *Medicago truncatula*, *Malus x domestica.* Phylogenetic analysis indicated that these proteins belong to the same clade. Ectopic over-expression of these non-Aradidopsis WD40 proteins in Arabidopsis *ttg1-1* mutant background could complement several mutant phenotypes (Brueggemann et al., 2010; Carey et al., 2004; de Vetten et al., 1997; Humphries et al., 2005; Morita et al., 2006; Pang et al., 2009; Sompornpailin et al., 2002; Walker et al., 1999), suggesting that they can interact with similar bHLH and MYB classes of protein partners and have similar function as Arabidopsis TTG1.

In this part, I focus on the functional study of a *A.annua* WD40 repeat gene, AaWD40. Using cell biology, molecular biology, and functional genomics approach, I found that over-expression of AaWD40 can fully rescue the trichome and PA phenotype in Arabidopsis *ttg1-1* mutant. Subcellular localization study suggested that the

translocation of AaWD40 into nuclear requires the assistance of bHLH family protein. Ectopic over-expression of AaWD40 in Arabidopsis altered the expression of CLV1, CLV2, CLV3 and WUS transcripts, which are required to maintain the stem-cell niche of Arabidopsis shoot apex. Finally, microarray analysis revealed the possible new function of AaWD40 and Arabidopsis TTG1 in regulating plant development and response to biotic and abiotic stress.

2. Materials and methods

2.1 Plant material and growth conditions

Arabidopsis seeds are purchased from the Arabidopsis Biotechnology Resource Centre (Ohio State University). All seeds used in this study are newly harvested (within 3 month) and stored in airtight tubes after drying at 4°C. Seeds were germinated on MS Basal Medium (Sigma M5519) with 1% sucrose and 0.8% Agar. The plants were placed at 22°C under long day condition (16h light / 8 dark). Seedlings were harvested for real-time RT-PCR, microarray analysis and microscopic analysis at 7 day following germination unless otherwise stated.

2.2 *A. thaliana* cell suspension cultures

Arabidopsis cell suspension cultures (ecotype Landsberg erecta, Ler) PSB-L were grown in 250 ml flasks under a light/ dark cycle of 16h/8h (22°C). The cultures were incubated on a shaker with an agitation speed of 130 rpm. Cells were sub-cultured in MS medium supplemented with 4.3 g /liter Murashige and Skoog Basal Salt Mixture, 100 mg / liter myo-inositol, 0.4 mg / liter thiamine hydrochloride, 50 mg / liter kinetin, 800 mg / liter 1-naphthaleneacetic acid and 30 g / (adjust pH to 5.7 with KOH). Sub-culture was conducted by transferring 5 ml of old cells into 45 ml of fresh Arabidopsis MS medium every 5 days. The cultures obtained on the 3rd day after sub-culture were used for protoplast preparation.

2.3 Molecular cloning and sequence analysis of AaWD40

Total RNA was isolated from 100mg *A. annua* leaves following the standard protocol using RNeasy Plant Mini Kit (Qiagen). The 5′ and 3′-end sequences of AaWD40 were recovered by rapid amplification of cDNA ends (RACE)-PCR based on the Contig24501 sequence using SMART™ RACE cDNA Amplification Kit (Clontech) to obtain the open reading frame of AaWD40. First-strand cDNA was synthesized with oligo(dT) primer using SuperScript® II Reverse Transcriptase (Invitrogen) from 500 ng total RNA, followed by 30 cylcles two-step (denaturation at 95°C followed by annealing and extensionat 68°C) PCR using Advantage® 2 Polymerase Mixes

(Clontech) and gene-specific oligonucleotide primers (5'-ATGGACACCAACTCAACCTTAG-3' and 5'- AACTTTCAGAAGCTGCAGCTTATT- 3'). The genomic sequence of AaWD40 was obtained with the same primers using genomic DNA as template. The PCR product was cloned into the vector Pmd18-T Simple (Takara) for sequencing.

2.4 Gene structure and phylogenetic analysis

The presence of nuclear localization signals was analyzed using PSORT (http://psort.ims.u-tokyo.ac.jp). Phylogenetic analysis was performed to investigate the relationship between AaWD40 and other WD40 repeat proteins from other plant species. The multiple alignments of the deduced amino acid sequences of AaWD40 with other WD40 repeat domain proteins were constructed using ClustalW2 (Larkin et al., 2007). The consensus tree was drawn by MacVector (MacVector, Inc) using neighbor-joining method. The scale bar indicates the number of amino acid substitutions per site.

2.5 Constructs

The coding region of AtTTG1 and GL3 was isolated from Arabidopsis Ler and that of AaWD40 was isolated from *A.annua* by PCR respectively. Using the following primers (forward primer, 5 AAAGCTTAACCGAGAATGTCTCCCGACTTCTA

T-3; reverse primer, 5-AGTCGACTCAAACTCTAAGGAGCTGCATTTG-3), XbaI and XhoI restriction sites were added to generate 35S::AtTTG1::GFP, 35S::AtTTG1::CFP, 35S::AaWD40::GFP, 35S::AaWD40::CFP 35S::GL3::RFP and 35S::GL3::YFP construct respectively. Gel-purified PCR product was digested with XbaI and XhoI, and the resulting fragment was cloned into the modified transient vector PBI221 and binary vector PBI121.

2.6 Transient expression in Arabidopsis protoplast

Transient expression of fluorescent fusion proteins in protoplasts of suspension cell cultures were performed as described by Miao etc (Miao and Jiang, 2007). For each sample, 40 μ g PBI221 plasmid were mixed with protoplasts prepared from *Arabidopsis thaliana* (ecotype *Landsberg erecta*) PSB-L cell suspension cultures and used for electroporation. After electroporation, the transfected protoplasts were incubated at 27°C before observation under Olympus FluoViewTM FV1000 confocal laser scanning biological microscope. Protoplasts were observed for fluorescent signals 8-12 h after electroporation.

2.7 Generation of transgenic plants

A. thaliana plant transformation were performed using floral dipping method and *A. tumefaciens* strain GV3101 pMP90-RK (Zhang et al., 2006) was used for

transformation. Selection of transgenic plants was conducted by germination on MS medium containing 50 mg/L kanamycin and 3% sucrose. Kanamycin resistant plants were transferred to soil and grown in the greenhouse as described before. The presence of transformed gene was confirmed by PCR in T1 plants of independent lines surviving the selection, using primers 35 5- TCCTTCGCAAGACCCTTCCTC -3 and AaWD40 RT R 5- CGGTGTATCCGTTGGACT -3 for AaWD40 gene; and 35S 5- TCCTTCGCAAGACCCTTCCTC -3 and AtTTG1RT R 5- GAGACTGATGCGAAAACCCTA-3 for AtTTG1 gene.

2.8 Genomic DNA isolation and Southern-Blot analysis (Font size not consistent)

Arabidopsis genomic DNAs were extracted by CTAB buffer. For genomic blots, DNA was subject to restriction digestion with EcoRI, agarose gel electrophoresis, and blotting according to standard procedures (Sambrook and Russell, 2006). DIG-labeled probes were synthesized with a DIG-High Prime DNA Labeling and Detection Starter Kit II from Roche according to the manufacturer's instructions.

For Southern-blot analysis of AaWD40 and AtTTG1 over-expression lines, the 400bp 35S promoter specific probe was generated by random-priming. Blots were probed, processed, and visualized according to DIG-labeling kit instructions from Roche

Biochemicals. Genomic Southern blots were hybridized at 42°C for over-night and washed at high stringency (68°C in 0.5× SSC).

2.9 Confocal microscopy

Imaging of YFP fusions was performed on Olympus FluoViewTM FV1000 confocal laser scanning biological microscope. Laser scanning microscope with excitation at 488 or 561nm and emissions at 500-540 nm were used for GFP and 580-630nm for RFP. Collected images were processed for maximum intensity projection.

2.10 Microarray and data analysis

Probe labeling, hybridization, and scanning for microarray analysis were conducted at Nottingham Arabidopsis Stock Centre (http://arabidopsis.info/) according to the manufacturer's instructions (Affymetrix). For each sample, raw probe intensities were normalized using RMA (Irizarry et al., 2003). Data analysis was conducted using Limma package (Gentleman, 2005; Smyth, 2004) in Bioconductor. Using a cut-off value of $p<0.05$, genes with significant expression changes were selected. All the microarray data have been deposited in NASCArrays (http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl) with respective experiment name as follows: "Arabidopsis *Landsberg erecta* ecotype "; "Arabidopsis

ttg1-1 mutant", "Arabidopsis AtTTG1 over-expression" "Arabidopsis AaWD40 over-expression".

2.11 Quantitative Real-time RT-PCR

100mg 7-day-old seedlings were frozen in liquid nitrogen for RNA preparation. Total RNA was isolated and purified using RNeasy plant mini kit (Qiagen). After DNase treatment using RQ1 RNase-free DNase (Promega), the first-strand cDNA was synthesized from ~1 μg total RNA using SuperScript™ II Reverse Transcriptase (Invitrogen) with Oligo(dT)12-18 Primer. Quantitative real-time PCR was carried out with the IQ5 Multicolor Real-time PCR Detection System (Bio-Rad, USA) using the 0.5X iQTM SYBR Green Supermix (Bio-Rad). Gene-specific primers were designed using Primer3 (http://frodo.wi.mit.edu/primer3). The cycling conditions were: 95°C for 10 min followed by 40 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 30s. Experiments were performed using three biological replicates (three transgenic lines) and three technical replicates. Expression levels of target genes were calculated relative to AtUBC21/At5g25760. Primers used for QRT-PCR were listed in Table 3.1

Table 3. 1 Primers used for QRT-PCR

Primer Name	Primer Sequence
WUSRT F	5- GCAACAACAACAACAACAAGTCCG -3
WUSRT R	5- GCCCTCAATCTTTCCGAACTGTCTC -3
BANRT F	5- GCACGGGAAACTTAGCCTCTATTC -3
BANRT R	5- TCTTCATCAGTCAAATCTGCCTTG -3
ETC1RT F	5- ATAGCAATGGCTCAGGAAGAAGAGG -3
ETC1RT R	5- AAATCCCACCTTTCACCGACAAG -3
TTG2RT F	5- ACCTTCTTCAAGTTCTGGCTTCAGG -3
TTG2RT R	5- AAGTCGTTTGCGGAGATACGGTAAC -3
CPC RT F	5- GGAGACAGAGCAAAGCCAAGG -3
CPC RT R	5- CAACGAGTTTATACATCCGAGAA -3
CLV1RT F	5- ATCTCTCGCCTGGAAACTAACCGC -3
CLV1RT R	5- GCTCCGCCTTTACCGATTATGTTC -3
CLV2RT F	5- CCTTCTTGCTTCGGTTCGTTGC -3
CLV2RT R	5- GACTTACAAATGTGGCGGGTATCG -3
CLV3RT F	5- GGGTTGGAGCAAATGGAGAAGC -3
CLV3RT R	5- TTGGCTGTCTTGGTGGGTTCAC -3

3. Results

3.1 **Molecular characterization of AaWD40**

Gene tag (Contig24501) encoding a putative WD40 protein was identified from an in-house-generated *A. annua* glandular trichome EST library. Tissue specific expression analysis showed that this contig was preferentially expressed in filamentous trichomes and roots. The contig was also expressed in glandular trichomes and leaves (Fig 3.1), although with lower abundance (Wang et al., 2009). A full length cDNA of 999bp was obtained by RACE PCR, and hereafter was referred to as AaWD40. The deduced amino acid sequence has 60%-73% sequence similarity to other known WD40 repeat proteins from different plant species (Fig 3.2). Phylogenetic analysis showed that

AaWD40 is most closely related to GhAtTTG1 from *Gossypium hirsutum* (Fig 3.3).

The 2,974bp genomic sequence of AaWD40 spans 3 exons and 2 introns.

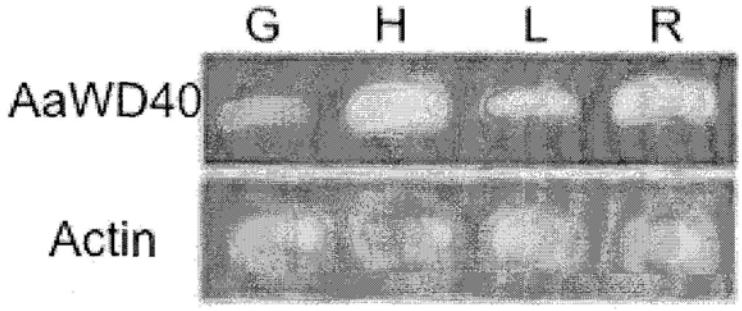


Figure 3. 1 Expression of AaWD40 in different tissue types.

G: Glandular trichome, N: Non-glandular tirhcome, L: Leaf, R: Root.

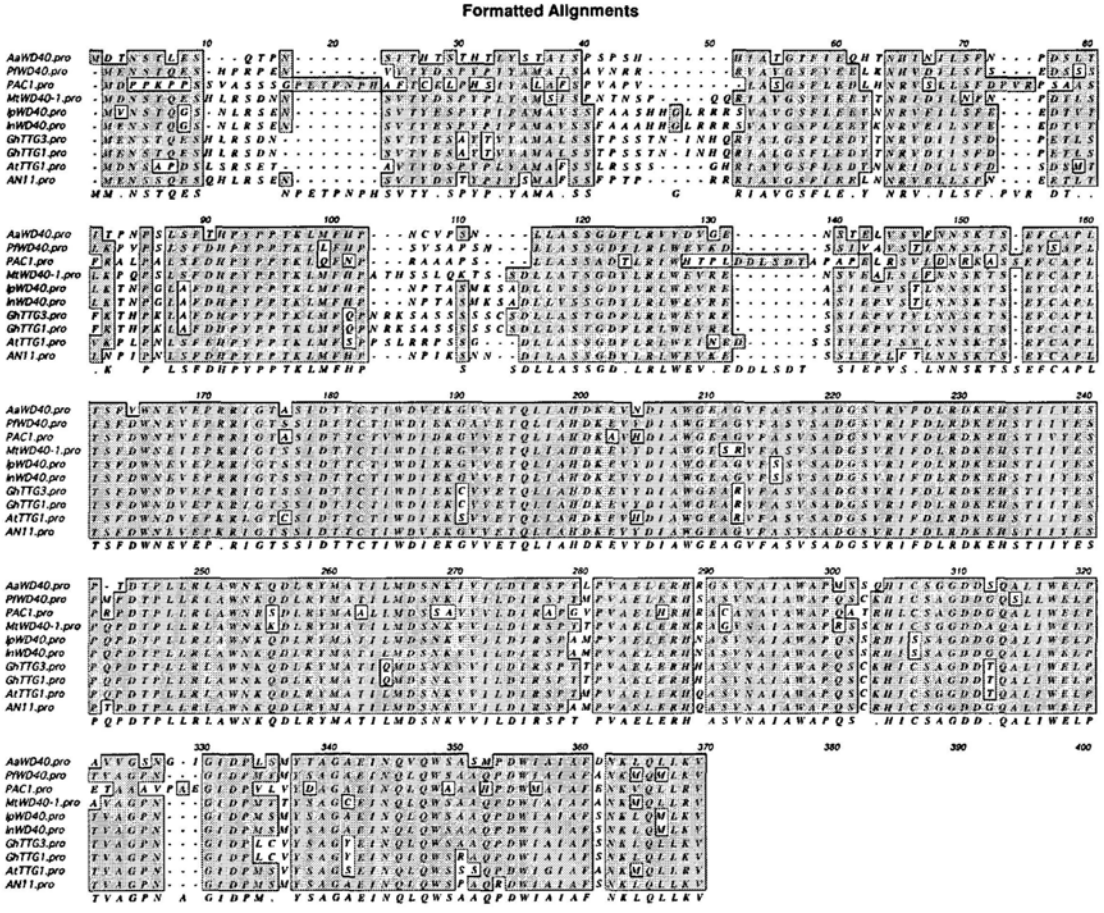


Figure 3. 2 Multiple sequence alignment of deduced amino acid sequences of plant WD40 repeat proteins.

Identical residues are highlighted on a black background. The region underlined indicates the conserved WD40 repeat domain. Alignments were performed using the using ClustalW. The GenBank accession numbers are as follows: AaWD40, from *A. auuna*; NP_851070, AtAtTTG1 from Arabidopsis; AAC18914, AN11 from *Petunia hybrida*; AAM76742, PAC1 from *Zea mays*; BAB58883, PfWD40 from *Perilla frutescens*; ABW08112, MtWD40-1 from *Medicago truncatula*; BAE94396, IpWD40 from *Ipomoea purpurea*; BAE94398, InWDR40 from *Ipomoea nil*; AAM95645, GhTTG3 from *Gossypium hirsutum*; AAK19614, GhAtTTG1 from *Gossypium hirsutum.*

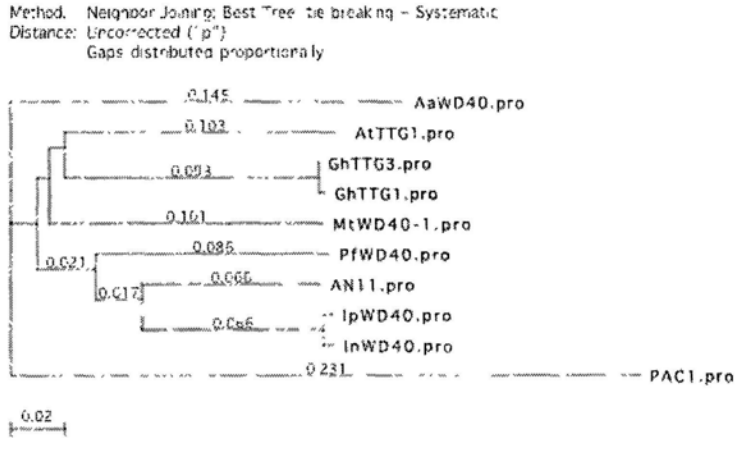


Figure 3. 3 Phylogenetic tree of AaWD40 and other WD40 repeat proteins.

Full-length amino acid sequences were aligned using ClustalW2. The presented consensus tree was drawn with MacVector using neighbour-joining analysis. The scale bar indicates the number of amino acid substitutions per site.

3.2 Complementation of *A. thaliana ttg1-1* mutant by AaWD40

Because AaWD40 shares high sequence similarity with AtTTG1, we first examined if they are functional homologous by complementation test. The mutation of AtTTG1 results in severe defects in anthocyanin pigmentation, seed coat pigmentation, seed coat mucilage, root hair positioning, and trichome differentiation. The PBI121 binary vectors, designated as 35S::AtTTG1::GFP and 35S::AaWD40::GFP, were transformed into Arabidopsis *ttg1-1* mutant. AtTTG1 (35S::AtTTG1::GFP fusion) and AaWD40 (35S::AaWD40::GFP fusion) were able to fully restore PAs , production in seeds and to rescue the trichomeless phenotypes in leaves and stems in homozygous transgenic lines. However, in heterozygous lines, density of trichomes was found less than that in the wild type (Fig 3.4). Moreover, in some of the heterozygous plants, a portion of seeds with changed seed coat color were observed.

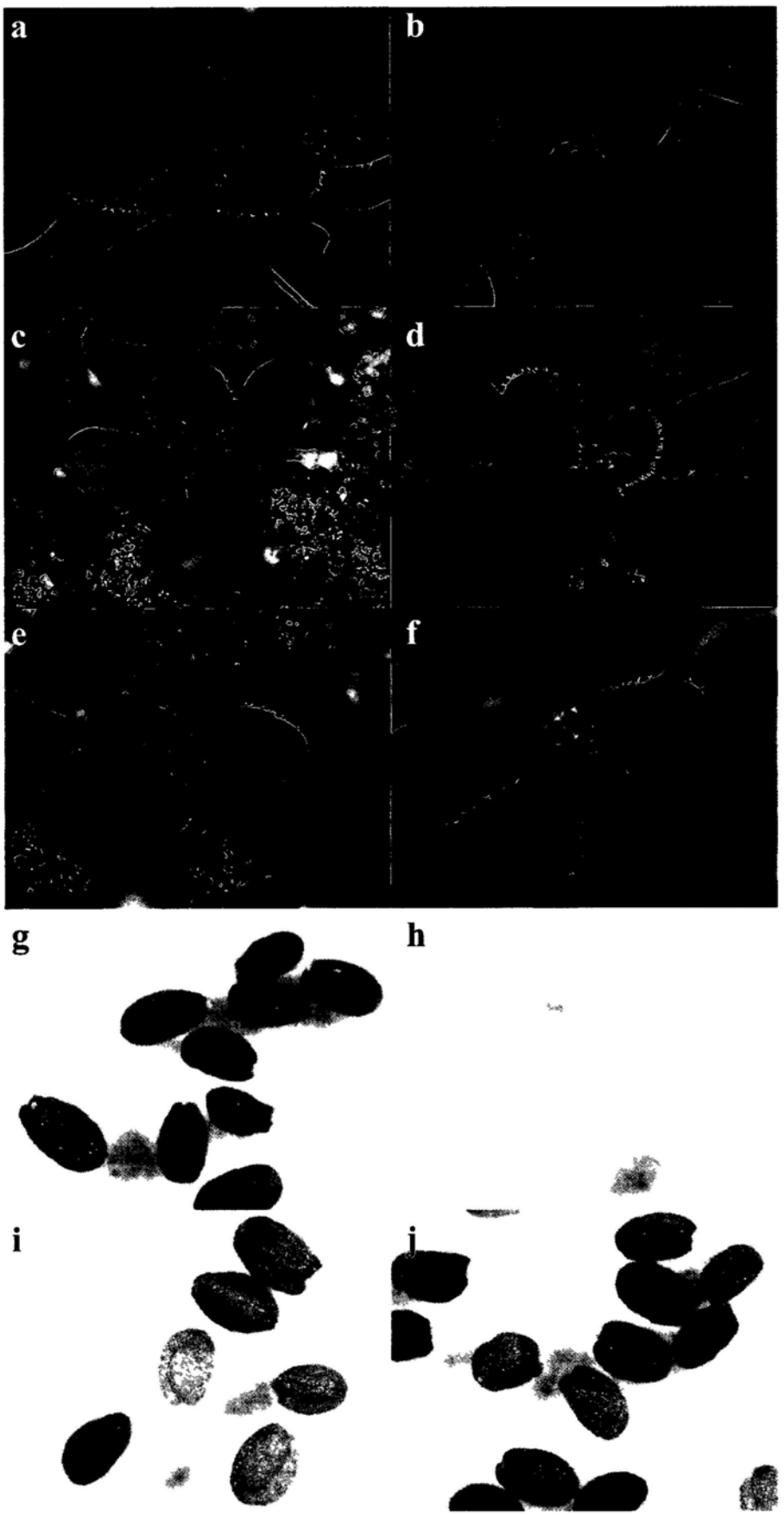


Figure 3. 4 Genetic complementation of the Arabidopsis *ttg1-1* mutant.

a, Leaves of the wild type (Ler). b, Leaves of the *ttg1-1* mutant line. c, Leaves of the *ttg1-1* mutant expressing 35S::AaWD40::GFP (heterozygous). d, Leaves of the *ttg1-1* mutant expressing 35S::AtTTG1::GFP (heterozygous). e, Leaves of the *ttg1-1* mutant expressing 35S::AaWD40::GFP (homozygous). f, Leaves of the *ttg1-1* mutant expressing 35S::AtTTG1::GFP (homozygous). g-j, Seed coat pigmentation of the wild type, *ttg1-1*, *ttg1-1* mutant expressing 35S::AaWD40::GFP (homozygous) and *ttg1-1* mutant expressing 35S::AtTTG1::GFP (homozygous). Leaves of *ttg1-1* show the glabrous phenotype. Leaves of the *ttg1-1* mutant expressing 35S::AaWD40::GFP (heterozygous) and 35S::AtTTG1::GFP (heterozygous) showing partial restoration of the trichome phenotype (only a few trichomes arise on the edges), while in homozygous transgenic lines the phenotype was fully rescued. *ttg1-1* have yellow seeds due to absence of brown pigment in seed coat and abnormal appearance, and over expression of 35S::AaWD40::GFP or 35S::AtTTG1::GFP can rescue this phenotype.

3.3 Subcelluar localization of AaWD40

Unlike most nuclear-bound WD40 proteins, no typical nuclear localization signal (NLS) was detected in the N-terminal region of AaWD40. In root tip cells of transgenic Arabidopsis plants over-expressing 35S::AaWD40::GFP, strong fluorescent signals were detected in both nuclear and cytoplasmic membrane (Fig 3.5 A). AaWD40 was also expressed transiently as a C-terminal fusion to GFP in Arabidopsis protoplast, and a cytoplasmic pattern was observed. A similar expression pattern was also observed for AtTTG1 (Fig 3.5 B). This result is somewhat contradictory to previous report that AtTTG1 was mostly located in the nuclear region (Bouyer et al., 2008). Other WD40 repeat proteins from petunia (AN11) and Perilla (PFWD) have been suggested to have a cytoplasmic pattern (de Vetten et al., 1997;

Sompornpailin et al., 2002). To further investigate the subcellular translocation pattern of WD40 protein, AtTTG1::GFP and AaWD40::GFP fusion proteins were co-expressed with AabHLH::RFP fusion protein in Arabidopsis protoplast respectively. AabHLH is a putative AtGL3 homolog with clusters of full term nuclear localization signal (NLS) (http://psort.ims.u-tokyo.ac.jp) cloned from *A.annua* (data not shown). As shown in Figure 3.6, co-expression of these two proteins resulted in an obvious nuclear pattern of AtTTG1 and AaWD40. These results indicate that the protein-protein interactions between WD40 repeat protein and bHLH protein may facilitate the nuclear translocation of WDR protein

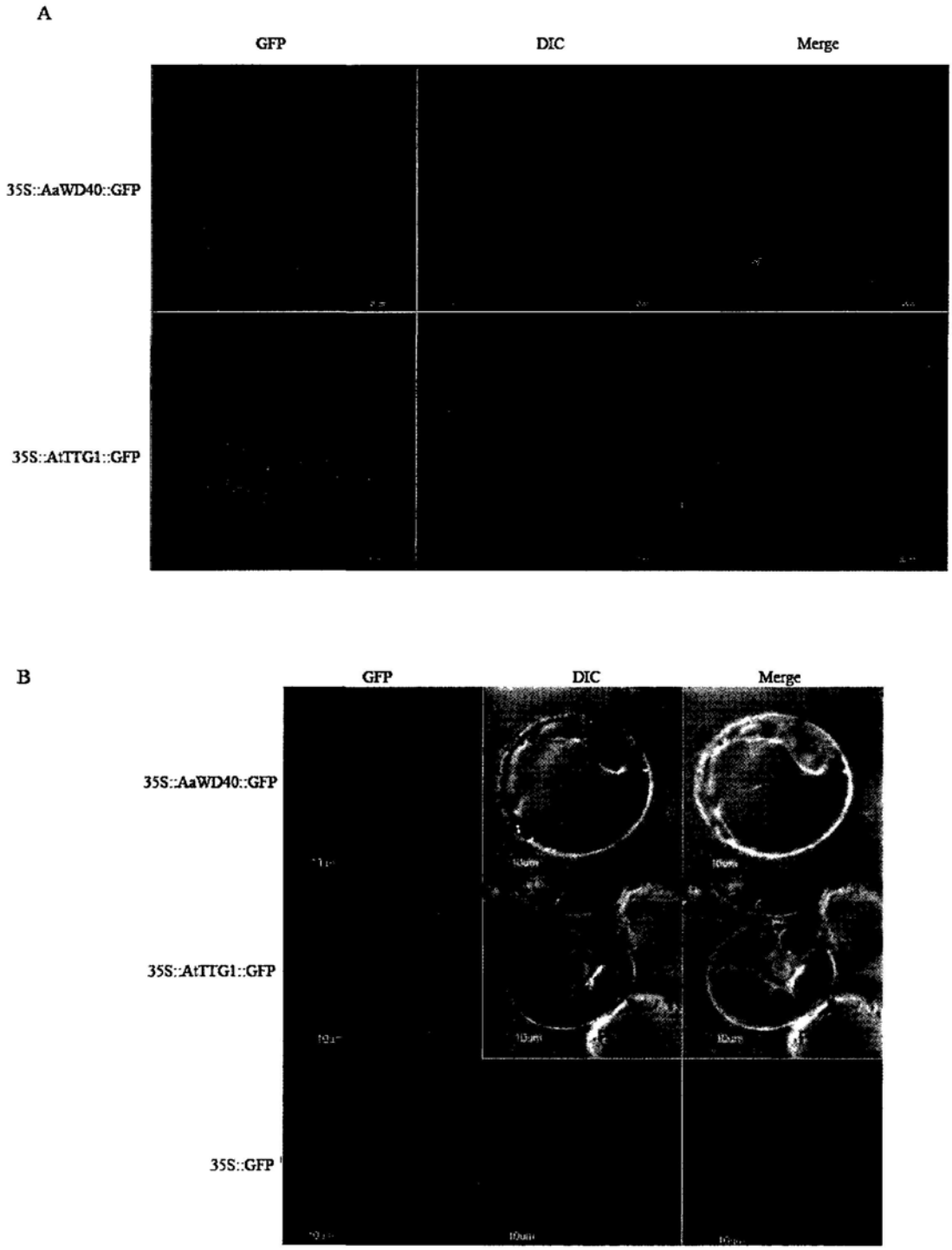


Figure 3. 5 Subcellular Localization Analysis of AaWD40 and AtTTG1 in Arabidopsis root tips and protoplast.

A. In the root tip cells of Arabidopsis, AaWD40 and AtTTG1 represent both cytosolic and nuclear pattern. Upper panel: 35S::AaWD40::GFP; Lower panel:

35S::AtTTG1::GFP. B. In Arabidopsis protoplasts, AaWD40 and AtTTG1 represent both cytosolic and nuclear pattern. 35S::GFP was used as control.

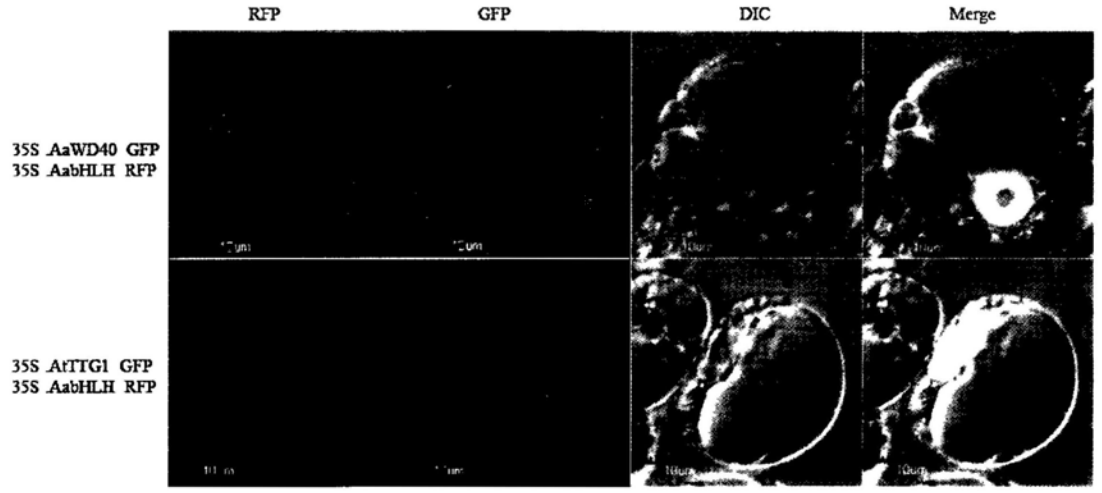


Figure 3. 6 Co-expression of AaWD40 and AtTTG1 with AabHLH in protoplast

When Artemisia bHLH protein (AabHLH) was co-expressed, the nuclear signal of AaWD40 or TTG1 was observed. Upper panel: Co-expression of 35S::AaWD40::GFP and 35S::AabHLH::RFP; Lower panel: Co-expression of 35S::AtTTG1::GFP and 35S::AabHLH::RFP

3.4 Southern-blotting analysis of AaWD40 and AtTTG1 over-expression lines

Southern-blotting analysis was conducted to verify the transgenic lines are derived from independent transgenic event. As illustrated in Figure 3.7, all the 8 tested samples were derived from independent transgenic events. b, d, e, f, g and h were used for further real-time RT-PCR and microarray analysis.

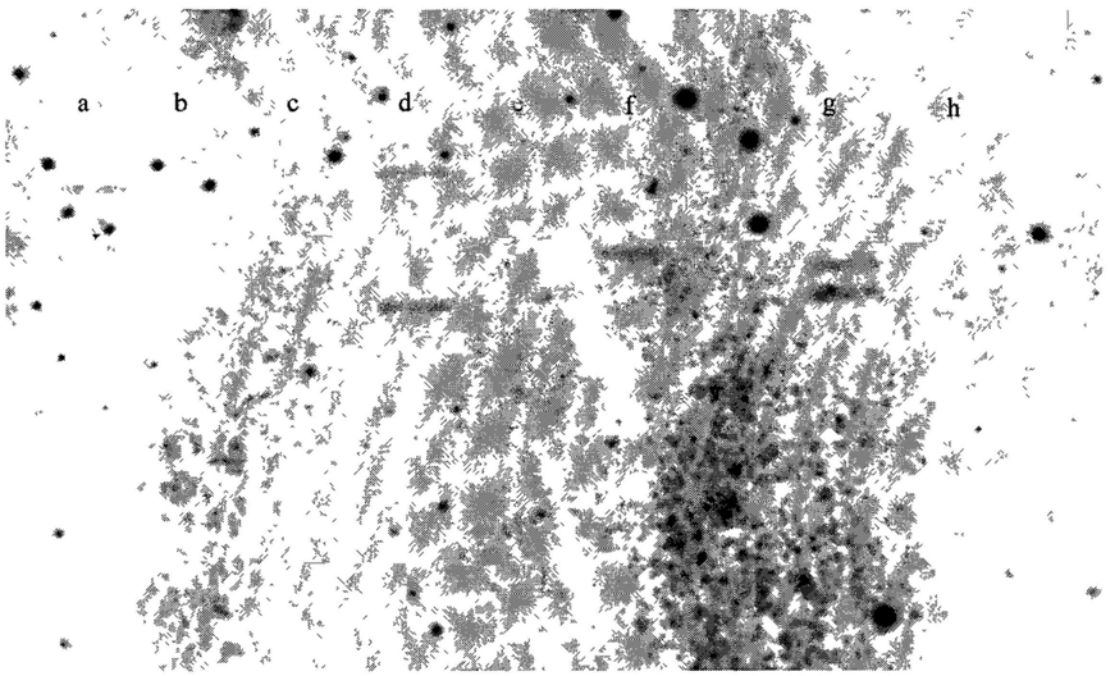


Figure 3.7 Southern-blot analysis of genomic DNA from AaWD40 and AtTTG1 over-expression lines.

Genomic DNA isolated from AaWD40 over-expression lines (a-e) and AtTTG1 over-expression lines (f-h) was digested with EcoRI, resolved by agarose gel electrophoresis, Southern blotted, and probed with the 35S specific probe.

3.5 Effects of AaWD40 over-expression on trichome morphogenesis and PA biosynthesis

In Arabidopsis, TTG1, GL1 and GL3 forms a trimeric regulatory complex to activate the expression of GLABRA2 (GL2), a homeo-domain Zip (HD-Zip) transcription factor, and TRANSPARENT TESTA GLABRA2 (TTG2), a WRKY factor (Zhao et al., 2008). MYB/bHLH/TTG1 regulatory complex also activates the expression of a subset of partially redundant single-repeat R3 MYB proteins, including CAPRICE (CPC) (Wada et al., 1997), ENHANCEROF TRY and CPC1 (ETC1) (Kirik et al., 2004a; Kirik et al., 2004b). CPC and ETC1 play central roles in lateral inhibition, by inactivating specific components of the trimeric regulatory complex (Schellmann et al., 2002). To characterize the effects of AaWD40 on trichome morphogenesis, the mRNA transcripts of 4 target genes of MYB/bHLH/TTG1 complex, including GL2, CPC, ETC, and TTG2, were evaluated by qRT-PCR (Fig 3.8). Arabidopsis *ttg1-1* mutant and over-expression line were used as reference for comparison. The mRNA level of GL2, which encodes a direct positive regulator of trichome morphogenesis located downstream of the trimeric complex, was significantly increased in TTG1 and AaWD40 over-expression plants. The mRNAs of CPC, ETC, TTG2 were strongly down-regulated in Arabidopsis *ttg1-1* mutant. In TTG1 and AaWD40 over-expression lines, the expression of TTG2 was significantly up-regulated, which was different

from that of CPC and ETC1. BANYULS (BAN), which encodes a core enzyme in PA biosynthesis, is another direct target of MYB/bHLH/AtTTG1 complex (Baudry et al., 2004). The mRNA transcripts of BAN were found to be decreased in the *ttg1-1* mutants and increased in AtTTG1 and AaWD40 over-expression lines (Fig 3.8).

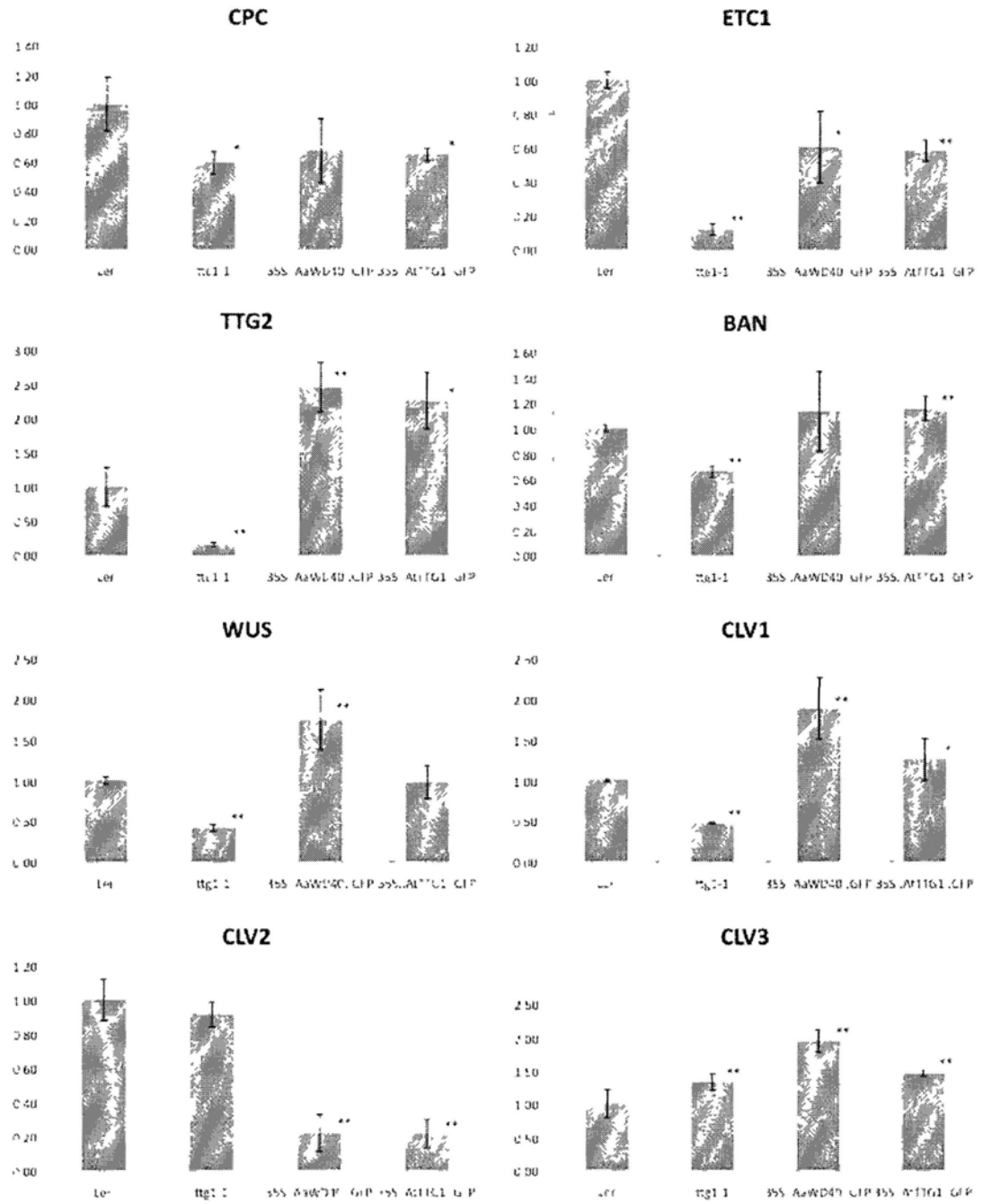


Figure 3. 8 Expression of genes involved in trichome formation, anthocyanidin biosynthesis and shoot meristem development were verified by qRT-PCR analysis.

Numbers in the ordinate give the fold induction relative to the expression level in the control sample (Ler). Error bars are SE (n=3). The relative standard method was used to measure the relative expression level of the target gene and was calculated relative to normalization of AtUBC21/At5g25760. Means were generated from three independent measurements; bars indicate standard errors. Expressions of the samples were compared using two-tail T-Test. The significance is expressed at 5%

significance level ($p < 0.05$, marked with *) and 1% significance level ($P < 0.01$, marked with **).

3.6 Effects of AaWD40 over-expression on WUS and CLV pathway

The shoot apical meristem of higher plants gives rise to the entire aboveground part of a plant. In Arabidopsis, shoot meristem has a properties of self-regulatory system in which is mediated by WUS/CLV interactions establish a feedback loop between the stem cells and the underlying organizing center (OC) (Schoof et al., 2000). WUS-mediated signaling from the OC specifies stem cell identity in the outermost cell layers, which signal back via the CLV pathway to limit the size of the WUS-expressing OC (Carles and Fletcher, 2003). Because the Arabidopsis *ttg 1-1* mutant clearly have more inflorescences than wild type (Buer and Djordjevic, 2009), we analyzed the mRNA transcripts of, CLAVATA1, CLAVATA2 and CLAVATA3 (Fig. 6) using real time RT-PCR in *ttg 1-1* mutant and AaWD40 over-expression line. As expected, WUS transcripts were down-regulated in *ttg1-1* mutant and up-regulated in AaWD40 over-expression line. As a consequence, CLV3 transcripts were significantly increased in AaWD40 line due to the positive feedback regulation of WUS expression. Interestingly, no up-regulation of WUS was detected in Arabidopsis TTG1 over-expression lines. CLV1 transcript was increased in both AaWD40 and TTG1 over-expression lines, but was decreased in *ttg1-1* mutant. Moreover, over-expression of AaWD40 and TTG1 resulted in dramatic decrease of CLV2 transcripts.

The CLV cell signaling model, which is supported by joint genetic and biochemical studies, suggests that CLV2 and CLV1 probably associates as a heterodimer to conduct CLV3 signaling (Brand et al., 2000; Jeong et al., 1999; Trotochaud et al., 1999). Although the transcription of CLV1 was activated by over-expression of TTG1 and AaWD40, the decreased CLV2 level likely affect the stability of CLV complex and further weaken the function of CLV3 negative regulation loop, and consequently resulted in increased WUS expression. Further investigation is still needed in order to understand the role of WD40 repeat proteins in mediating WUS/CLV pathway.

3.7 Microarray Analysis of Arabidopsis *ttg1-1* mutant, AaWD40 and AtTTG1 over-expression Lines

We performed genome-wide gene expression analysis using Affymetrix ATH1 arrays. RNAs were extracted from 7 days-old Ler, *ttg1-1*, 35S::AtTTG1::GFP and 35S::AaWD40::GFP transgenic seedlings. Genes with more than 2 folds change in mRNA expression relative to wild type were selected for further investigation.

The result revealed that 164, 96, 1914 genes were differentially expressed in 35S::AaWD40::GFP, 35S::AtTTG1::GFP and *ttg1-1* plants respectively. Out of the

1914 genes, 40 genes (Appendix Table) were commonly altered in all three types of plants, and 220 were altered in both of the over-expression lines (Fig 3.9).

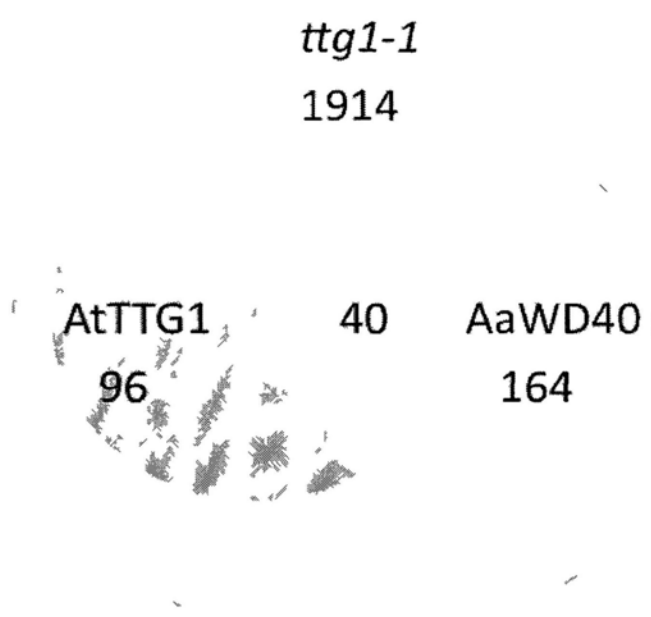


Figure 3. 9 Differentially expressed genes in *ttg1-1* mutant, AaWD40 and AtTTG1 over-expression lines

The assigned functionality of genes covers a broad range of GO categories. The top 20 most highly represented GO categories were illustrated in Figure 3.13-3.16. The genes differentially expressed in AaWD40 and AtTTG1 over-expression lines have similar GO categories, with "regulation of transcription" and "defense/stress response" listed as the top-ranked GO categories. Those defense and stress related categories includes: defense response to fungus, response to jasmonic acid stimulus,

response to chitin, defense response to bacterium, response to heat and response to wounding, etc.

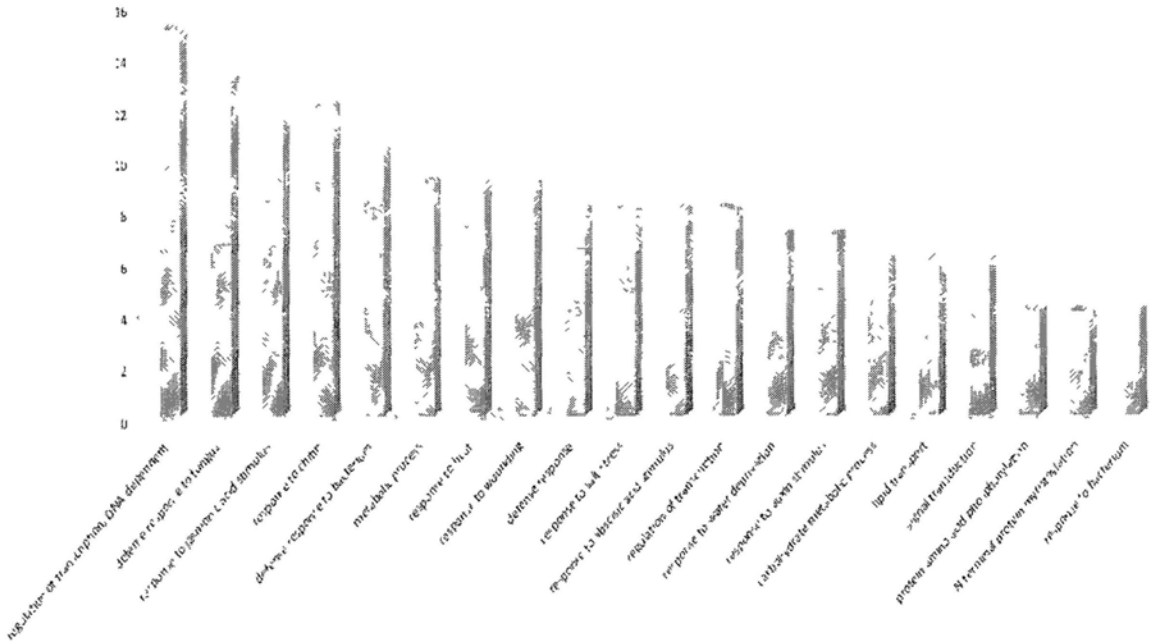


Figure 3. 10 Top-ranked GO categories (biological process) of genes differentially expressed in AaWD40 over-expression plants

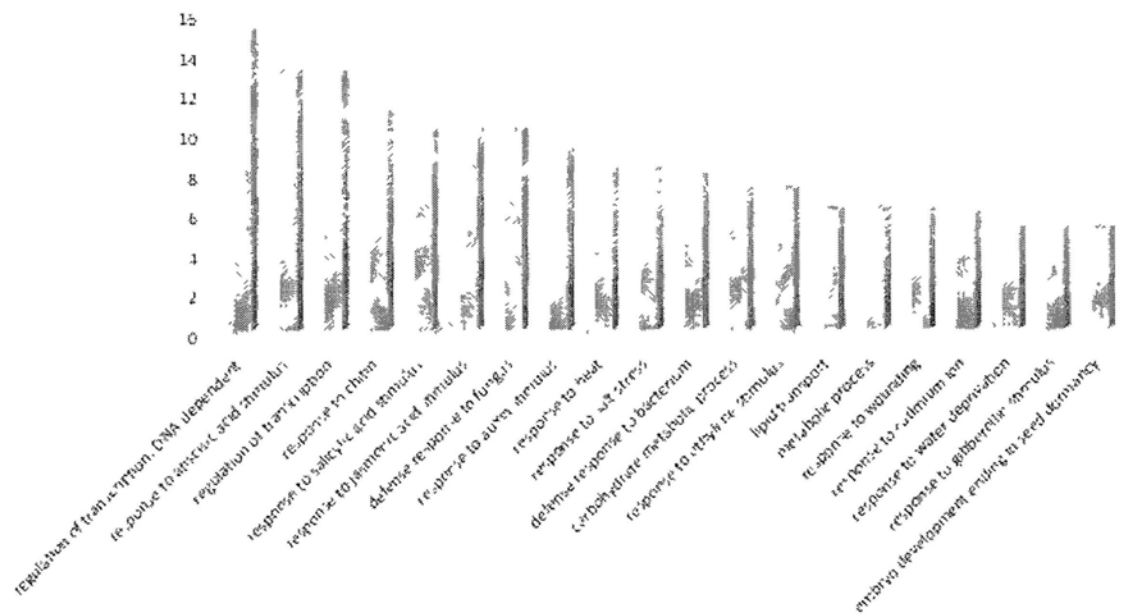


Figure 3. 11 Top-ranked GO categories (biological process) of genes differentially expressed in AtTTG1 over-expression plants

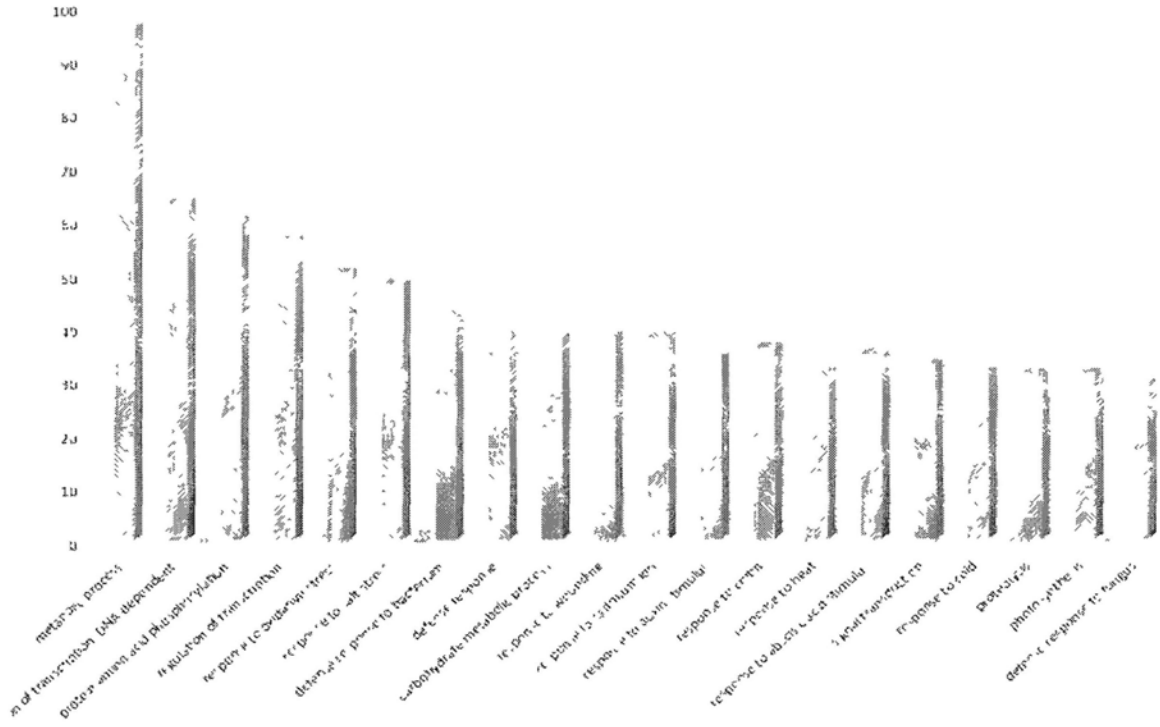


Figure 3. 12 Top-ranked GO categories (biological process) of genes differentially expressed in *ttg1-1* mutant

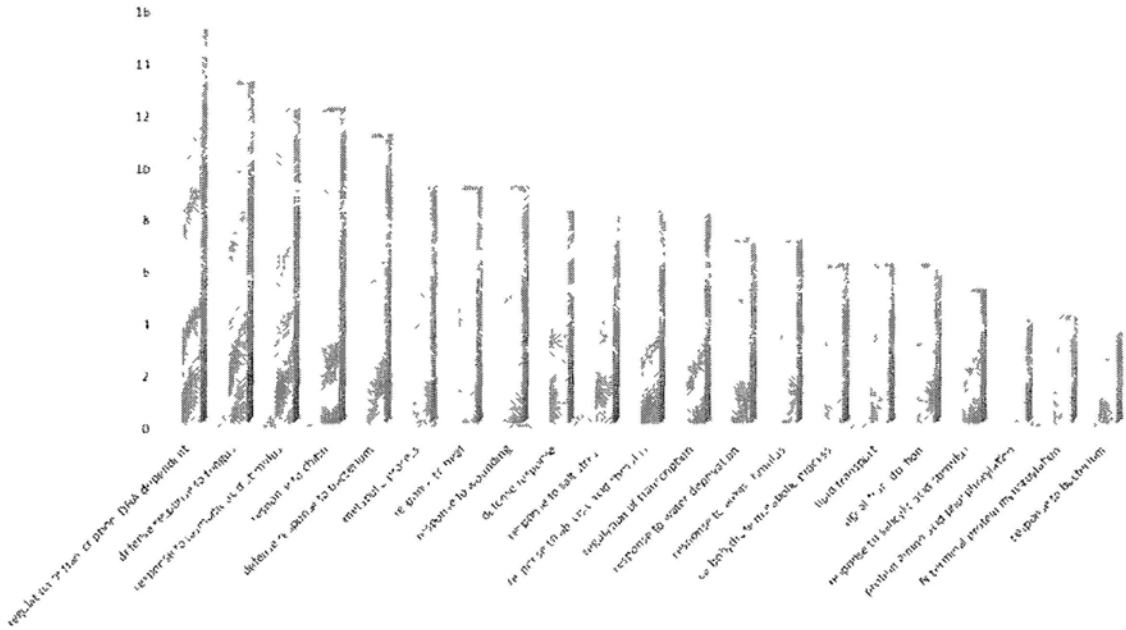


Figure 3. 13 Top-ranked GO categories (biological process) of genes differentially expressed in differentially expressed in *ttg1-1* mutant and also changed in AtTTG1::GFP or AaWD40 over-expression plants

Among the 1914 genes differentially expressed in *ttg1-1* plants, 123 are annotated as transcriptional factors (TFs). The AP2 domain family, basic helix-loop-helix (bHLH) family, MYB family, WRKY family, NAC domain proteins and bZIP family proteins were top-listed (Fig 3.14). It is well known that these TFs regulate a broad range of biological functions, including developmental programming, cell fate determination, and defense/stress response. Mutation of AtTTG1 resulted in altered expression of genes involved in development and defense/stress response highlight the possible

cross-talks among these pathways.

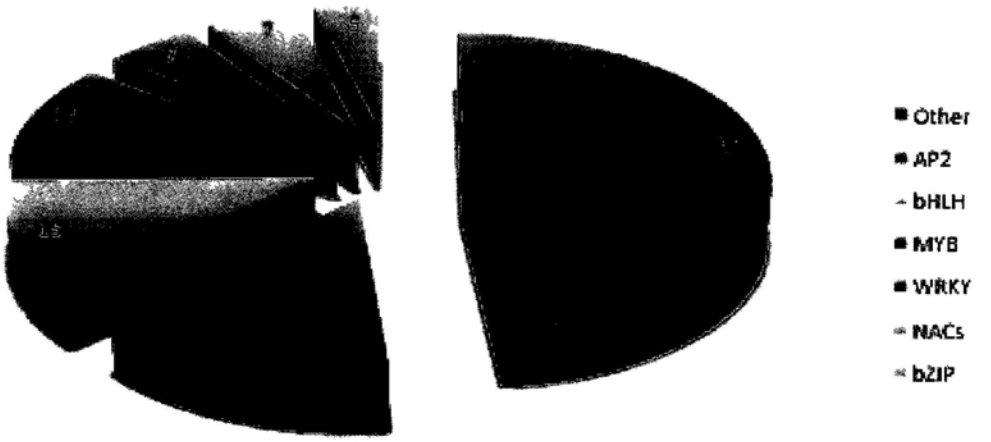


Figure 3. 14 Transcriptional factors differentially expressed in *ttg1-1* mutant

In total, 14 MYB factors, including TRY and MYBL2, showed altered expression in *ttg1-1* mutant. It is known that TRY and MYBL2 interact with AtTTG1 and GL3 to modulate the epidermal cell differentiation and trichome formation (Tab. 3.2). Most of the MYB factors were down-regulated except for AT3G23250 and AT1G18570. Seven NACs, including the well characterized ANAC019 and ANAC055 (Bu et al., 2008) were up-regulated (Tab. 3.3). The largest TF family is AP2/ERFs (19), followed by bHLH family (15), WRKYs family (9). Interestingly, most of AP2/ERFs

and WRKYs were up-regulated with only three exceptions (AP2/ERF family AT2G22200, AT2G39250 and AT5G67180). A detailed description of the differentially expressed TFs can be found in Table 3.4, 3.5 and 3.6).

Table 3. 2 MYB family proteins differentially expressed in *ttg1-1* mutant

Locus	Gene Name	Fold Change	Subfamily
AT1G71030	ATMYBL2	0.068896246	
AT5G53200	TRIPTYCHON (TRY)	0.273693768	
AT1G01520	F22L4.18	0.368582632	Single Repeat
AT3G09600	F11F8.19	0.382470463	
AT2G46410	CAPRICE (CPC)	0.573605899	
AT1G01380	ENHANCER OF TRY AND CPC 1 (ETC1)	0.962588898	
AT3G23250	ATMYB15	16.97698154	
AT1G18570	ATMYB51	4.645660422	
AT5G15310	ATMYB16	0.289889329	
AT1G22640	ATMYB3	0.298893646	
AT4G21440	ATMYB102	0.340762006	
AT1G49010	F27J15.20	0.375257274	R2R3
AT1G74650	ATMYB31	2.515164444	
AT1G18710	ATMYB47	0.481168351	
AT5G07690	ATMYB29	0.489313456	
AT1G06180	ATMYB13	2.004185705	
AT5G14750	WEREWOLF (WER)	0.920788037	
AT5G40330	ATMYB23	0.522703318	
AT3G27920	GLABRA 1 (GL1)	0.667790483	

Table 3. 3 NAC domain proteins differentially expressed in *ttg1-1* mutant

Locus	Gene Name	Fold Change
AT1G02220	ANAC003	7.085830651
AT1G52890	ANAC019	5.858436003
AT1G34180	ANAC016	5.11031205
AT3G15500	ANAC055	3.354145323
AT3G04070	ANAC047	3.281119742
AT5G18270	ANAC087	2.778977121
AT2G17040	ANAC036	2.502001425

Table 3. 4 AP2/ERF family proteins differentially expressed in *ttg1-1* mutant

Locus	Gene Name	Subfamily	Fold Change
AT5G05410	DREB2A	A-2	3.13521483
AT1G21910	T26F17.14	A-5	6.218971536
AT3G50260	DEAR1	A-5	3.238248713
AT1G77640	T5M16.23	A-5	3.076629459
AT2G23340	DEAR3	A-5	2.075916431
AT2G22200	T26C19.14	A-6	0.4826182
AT2G47520	HRE2	B-2	2.036089683
AT4G17500	ATERF-1	B-3	5.589583807
AT5G47220	ATERF-2	B-3	6.949270014
AT4G34410	RRTF1	B-3	3.823831467
AT4G17490	ATERF6	B-3	2.448832227
AT2G44840	ATERF13	B-3	2.071638642
AT5G61600	ERF104	B-3	2.314888505
AT5G51190	MWD22.13	B-3	2.234727683
AT1G43160	F1I21.18	B-4	4.58019868
AT5G13330	RAP2.6L	B-4	4.242060039
AT2G39250	SCHNARCHZAPFEN		0.231042773
AT5G67180	K21H1.22		0.342663784
AT3G25730	EDF3		2.075265952

Table 3. 5 WRKY family proteins differentially expressed in *ttg1-1* mutant

Locus	Gene Name	Fold Change
AT5G13080	ATWRKY75	8.850787881
AT1G80840	ATWRKY40	6.71074283
AT4G31800	ATWRKY18	6.186890846
AT3G56400	ATWRKY70	5.649160151
AT4G23810	ATWRKY53	4.522091695
AT5G49520	ATWRKY48	4.071067477
AT2G38470	ATWRKY33	3.663289447
AT2G46400	ATWRKY46	3.027562389
AT3G01970	ATWRKY45	2.053565819
AT2G25000	ATWRKY60	1.853176124

Table 3. 6 bHLH family proteins differentially expressed in *ttg1-1* mutant

Locus	Gene Name	Fold Change
AT1G10585		12.04689072
AT5G50915		4.078493351
AT5G46690	BHLH071	0.270142143
AT5G39860	BHLH136	0.198136789
AT2G43140	F14B2.8	3.350685666
AT1G02340	FBI1	0.312295643
AT4G36540	BEE2	0.323013119
AT2G18300	T30D6.19	0.338687853
AT3G56980	BHLH039	2.732724016
AT3G25710	BHLH32	0.37163965
AT4G01460	F11O4.13	0.37597731
AT4G00050	F6N15.11	0.394199748
AT1G12860	F13K23.12	0.421165531
AT1G73830	BEE3	0.467796875
AT3G05800	AIF1	0.474403592

4. Discussion

Plants have developed effective trichome-based anti-herbivory traits for self-defense after millions of years of co-evolution with insects (Rausher, 2001). Thus, the regulatory networks underlying defense response and trichome development are suggested to have cross-talks. In fact, the plant hormone participates in the regulation of a variety of developmental processes and serves as a key mediator of plant responses to biotic and abiotic stress factors also promotes different trichome types of various plant species. Molecular analysis indicated that JA, BAP, and GA3 stimulate trichomes initiation through transcriptional regulation of the components of TTG1 activator/inhibitor complex (Maes and Goossens, 2010; Maes et al., 2008). Besides, several other hormone classes also affect trichomes density indirectly or directly. Ethylene, gibberellin, and auxin mutants indirectly showed alteration in trichome density in tomato, through effects on epidermal cell area. However, brassinosteroids (BRs) directly affect trichome density and allelochemical content in an opposite fashion (Campos et al., 2009). In our study, microarray analysis revealed that genes involved in a variety of developmental processes and plant stress response were modulated by AaWD40 or AtTTG1. That makes them very hopeful candidates for genetic engineering to enhance the glandular trichome-based defence in *A. annua*.

Arabidopsis *ttg1-1* mutant to investigate the downstream regulatory targets of AtTTG1 like proteins.

4.2 AtTTG1 like proteins may be related to glandular trichome based defence

Plants have developed effective trichome-based anti-herbivory traits for self-defense after millions of years of co-evolution with insects (Rausher, 2001). Thus, the regulatory networks underlying defense response and trichome development are suggested to have cross-talks. The plant hormone participates in the regulation of a variety of developmental processes and serves as a key mediator of plant responses to biotic and abiotic stress factors. Loss-of-function of CORONATINE-INSENSITIVE1 (COI1), the tomato homolog of Arabidopsis JA signal F-box protein, in tomato jai1 mutant resulted in several defense-related phenotypes, including the inability to express JA-responsive genes and abnormal development of glandular trichomes (Li et al., 2004). These findings suggest a role for JA in the promotion of glandular trichome and enhancement of plant defense (Li et al., 2004). Besides JA, several hormone classes, e.g. ethylene, gibberellin, auxin, and brassinosteroids also affect trichomes density directly or indirectly. (Campos et al., 2009). In our study, microarray analysis revealed that genes involved in a variety of developmental processes and plant stress response were modulated by AtTTG1 and AaWD40. This highlights the potential of

using WDR protein as genetic engineering target to enhance the glandular trichome-based defense in *A. annua*.

4.3 AtTTG1 modulates anthocyanin biosynthesis and epidermal differentiation by activating MYBs

The AtTTG1,GL3 and GL1 trimeric complex can directly activate most of the known single MYB repeat inhibitor proteins, including CPC, ETC1, MYBL2, CPL3 and TRY(Morohashi and Grotewold, 2009). From our result, we found that almost all single repeat MYB genes were down-regulated in *ttg1-1* mutant except AT3G23250 and AT1G18570. Among the differentially expressed 14 MYB family genes, 3 previously uncharacterized single repeat MYB AT1G71030, AT1G01520 and AT3G09600, which are conserved in sequence with CPC, ETC1, CPL3, MYBL2 and TRY, also showed similar down-regulated expression pattern in *ttg1-1*. This indicates that they may also act downstream of the AtTTG1-GL3-GL1 complex.

Arabidopsis GLABROUS1 (GL1), WEREWOLF (WER) and ATMYB23, which constitute a subgroup of R2R3-MYB proteins, share 95% amino acid sequence identity in the MYB motifs for DNA binding and bHLH partner protein interaction

and 65% similarity in a C-terminal domain responsible for transcriptional activation (Galway et al., 1994; Kwak et al., 2005). These R2R3 MYB protein competent with single MYB domain protein to form WD40-bHLH-R2R3 MYB complex and provides positive feedback loops by binding to their own promoters (Kang et al., 2009). Among the 10 R2R3 MYBs with altered transcripts in *ttg1-1* mutant, AT1G22640 (ATMYB3) has been reported to repress phenylpropanoid biosynthesis pathway genes, Thus, it may be another important target of AtTTG1 in phenylpropanoid metabolism.

4.4 AtTTG1 affects developmental programmes, defense and abiotic stress responses by regulating NACs

NAC proteins constitute one of the largest plant-specific TF families and are present in a wide range of land plants. Although more than a hundred NAC members have been identified in Arabidopsis (Riechmann et al., 2000), only a small portion of the NAC proteins have been studied. This protein family has been implicated to function in diverse processes, including developmental programmes, defense and abiotic stress responses. Three NACs, including NAM, CUC1 (CUP-SHAPED COTYLEDON 1), and CUC2 (CUP-SHAPED COTYLEDON 2) have been well-studied. Mutants of NAM, CUC1 and CUC2 exhibit fused cotyledons and absence of embryonic shoot apical meristem (SAM) (Aida et al., 1997). Another Arabidopsis NAC gene, NAC1,

WD40-bHLH-MYB regulatory complex is a well studied model, which can explain lots of AtTTGl based regulatory events perfectly. WD40 protein acts as an upstream regulator of multiple biological processes by recruiting various co-functional components. However, other model also been proved recently. AtTTG1 was found can physically interacted with GEM, a GL2-expression modulator that regulate cell division and represses the expression of GL2 and CPC in Arabidopsis roots (Caro et al., 2007). It is possible that AtTTG1 or it's functional homlogue may recruit various kinds of co-factors in order to regulate multiple biological functions in plant.

4.1 AaWD40 and AtTTG1 have similar function in Arabidopsis

Complementation test proved that AaWD40 is a functional homlogue of AtTTG1. Over-expression of AaWD40 in Arabidopsis *ttg 1-1* mutant can fully restore the trichome initiation and anthocyanin biosynthesis. Furthermore, over-expression of either AaWD40 or AtTTG1 can modulate the expression of WUS and CLVs, which are critical for the normal morphology of shoot meristem. Microarray data indicated that over-expression of AaWD40 and AtTTG1 resulted in differential expression of genes related to regulation, defense and stress response. All these results indicated AaWD40 and AtTTG1 has similar functions in Arabidopsis. Thus, we used

can promote lateral root development by mediating auxin signaling (Xie et al., 2000). In our study, NAC1 transcript was increased to 1.55 folds in *ttg1-1* mutant compared to WT. NARS1, also known as NAC2, or NAM, can affect embryogenesis by regulating the development and degeneration of ovule integuments (Kunieda et al., 2008). NACs are also involved in defense and abiotic stress responses. For example, transgenic plants over-expressing three different Arabidopsis NAC genes (ANAC019, ANAC055 and ANAC072) showed significantly increased drought tolerance (Tran et al., 2004). Further evidence suggest that Arabidopsis NAC family protein ANAC019 and ANAC055 might function as downstream transcription activators of AtMYC2 to regulate JA-induced defense gene expression (Bu et al., 2008). We found that all the 7 NAC domain containing proteins, including ANAC019 and ANAC055, were dramatically increased in *ttg1-1* mutant, indicating the AtTTG1-bHLH-MYB complex may act as negative regulator upstream of NACs.

4.5 AtTTG1 regulates a broad range of AP2/ERFs and WRKYs

Transcriptional reprogramming, which is involved in both basal disease resistance and stress response, is mediated by several classes of TFs, including AP2/ERF, WRKY, and basic region/leucine zipper (bZIP) family members (Chen and Chen, 2002; Chen et al., 2002; Kalde et al., 2003; Kim and Delaney, 2002). The AP2 transcription factor

family, found only in plants, includes several genes that encode proteins involved in the regulation of disease resistance and stress pathways. One of them is DREB2A, which can specifically interact with a cis-acting dehydration-responsive element (DRE) to modulate the expression of cold and dehydration response genes in Arabidopsis (Maruyama et al., 2009). ATWRKY75 plays a role in regulating Pi starvation responses in Arabidopsis. In addition, when ATWRKY75 expression was suppressed, lateral root length and number, as well as root hair number, were significantly increased (Devaiah et al., 2007). Pathogen-induced ATWRKY18, ATWRKY40, and ATWRKY60 factors can form both homo-complexes and heterocomplexes to bind to the W-box in vitro. They have partially redundant roles in response to the hemibiotrophic bacterial pathogen *Pseudomonas syringae* and the necrotrophic fungal pathogen *Botrytis cinerea* (Xu et al., 2006). Our microarray analysis revealed that all the AP2/ERF and WRKY proteins mentioned above were significantly up-regulated in *ttg1-1*, indicating that AtTTG1 may function in defense and stress response, as well as in developmental programme by modulating the transcription of AP2/ERF and WRKY transcriptional factors.

4.6 bHLHs also act downstream of AtTTG1

Arabidopsis bHLH family proteins participate in a broad range of growth and developmental processes at all stages of the plant life cycle, ranging from

phytochrome signaling (Fairchild et al., 2000; Huq et al., 2000; Ni et al., 1998), gynoecium development (Heisler et al., 2001; Rajani and Sundaresan, 2001), flavonoid biosynthesis (Nesi et al., 2000), trichrome differentiation (Payne et al., 2000), microspore development (Sorensen et al., 2003), abscisic acid–induced gene expression (Abe et al., 2003; Abe et al., 1997), tryptophan biosynthesis (Smolen et al., 2002), brassinosteroid signaling (Friedrichsen et al., 2002), and chilling and freezing tolerance responses (Chinnusamy et al., 2003) and MP-dependent root initiation (Schlereth et al., 2010). One of the bHLH gene (bHLH 32), which showed 2.73 fold increase in *ttg 1-1* mutant are known to be involved in phosphate starvation response, negative regulator of root hair development, anthocyanin formation and Pi content (Chen et al., 2007; Schlereth et al., 2010). BEE2 and BEE3, which are key components in brassinosteroid signaling pathway, were also decreased in *ttg 1-1.* These results further support our presumption that WD40 protein AtTTG1 acts as a key regulator in a wide range of cellular events.

5. Summary

In the present study, we conducted functional analysis of a WD40 repeat proteins (AaWD40) isolated from *A.annua* glandular trichome. Previous study showed that TTG1 is localized in the nuclear of Arabidopsis leaf epidermal cell and trichome cell

with the present of endogenous GL3, but diffused in *gl3* mutant background (Bouyer et al., 2008). The mechanism of how this WD40 protein enters the nucleus has never been investigated. We provided the direct evidence that AabHLH protein can facilitate the correct targeting of AaWD40 and AtTTG1 into nucleus. We also showed that AaWD40 is a functional homolog of Arabidopsis WD repeat protein TTG1, and it plays critical roles in epidermal cell differentiation and secondary metabolite biosynthesis. When AaWD40 and AtTTG1 gene was ectopically over-expressed in Arabidopsis *transparent testa glabrous1-1* (*ttg1-1*) mutants of *A. thaliana*, proanthocyanin (PA) production in seeds was restored, and the trichomeless phenotypes of mutant was rescued. Real-time PCR analysis results revealed that ETC1, CPC, TTG2 and BAN (the downstream targets of TTG1 depends regulatory complex), which regulate the epidermal differentiation and anthocyanins biosynthesis, were modulated by AaWD40 and AtTTG1. Furthermore, the expression of CLV1, CLV2, CLV3 and WUS, which are required to maintain the stem-cell niche of Arabidopsis shoot apex, were also affected. Transcriptomic profiling results showed that the transcription of AP2/ERF, bHLH, MYB, WRKY and NACs family proteins, which are mostly involved in defense, stress response and development regulation, were remarkably modulated by AaWD40 and AtTTG1.

Chapter 4 Conclusions and Perspective

In this study, we first conducted a global analysis of glandular trichome in *A.annua* using massively parallel pyrosequencing. Mining the pyrosequencing ESTs resulted in the identification of many unigenes likely involved in terpenoid biosynthesis and trichome function. Functional characterizations of selected genes are being carried out. These pyrosequence data form the basis for further characterization of the molecular basis of glandular trichome metabolism in *A.annua*. The results also highlight the value of using high throughput pyrosequencing technnology for gene discovery in non-model plants. Then we focused on the functional analysis of a WD40 repeat proteins (AaWD40) isolated in *A.annua* glandular trichome. We showed that AaWD40 is a functional homolog to AtTTG1, which play critical roles in epidermal cell differentiation and secondary metabolite biosynthesis, it also regulates the WUS and CLVs genes to affect shoot development. Subcellular localization study demonstrated that bHLH protein facilitates the correct targeting of both AaWD40 and AtTTG1 into nucleus.

Although the Arabidopsis TTG1 has been extensively studied previously, to our knowledge, a comprehensive analysis of differentially expressed genes regulated by TTG1 has never been reported. In this study, we conducted gene-chip analysis on

transgenic Arabidopsis over-expressing AaWD40 and TTG1 as well as *ttg1-1* mutant, and lists of differentially expressed genes were identified. Among them, genes participate in transcription regulation and stress/defense response are highly represented. This indicates that apart from the previously described molecular functions, TTG1 and AaWD40 may also act as important upstream regulator at the regulatory network integrating multiple biological pathways, such as, defense, biotic and abiotic stress responses, and developmental programme.

Our results highlight the critical role WD40 repeat protein may play in mediating the plant hormone signaling and defense/stress response. More investigations are needed in order to further understand the genetic and hormonal basis of induced trichome production. Such knowledge would be important for the development of trichome-based defense and stress tolerance, as well as for enhanced plant natural production through manipulation of trichome density.

Appendix Table Genes commonly altered in over-expression and mutant lines

Gene ID	Gene Name	Description	Fold Change in AaWD40	Fold change in AtTTG1	Fold change in *ttg 1-1*
AT1G08650	PPCK1	phosphoenolpyruvate carboxylase kinase	0.336292822	0.459184397	0.49662166
AT3G14200	MAG2.19	DNAJ heat shock N-terminal domain-containing protein	0.393960706	0.47157111	2.161714123
AT5G05140	OBSOLETE		2.069491562	2.226931483	2.255640596
ATCG00180	RPOC1	RNA polymerase beta' subunit-1	2.454036536	2.415213368	2.323774263
AT3G44450	F14L2.1	unknown protein	0.418570749	0.430141856	0.427882433
ATCG00170	RPOC2	RNA polymerase beta' subunit-2	2.35736191	2.17420153	2.526550776
AT5G62130	MTG10.16	Per1-like protein-related	0.435941088	0.431080156	0.373870085
AT1G56600	ATGOLS2	Arabidopsis thaliana galactinol synthase 2 (AtGolS2)	0.381408298	0.4198378	0.367116791
AT1G69890	T17F3.8, T17F3_8	molecular_function unknown	2.443227031	2.174824375	2.874231765
AT4G21440	MYB102	Member of the R2R3 factor gene family.	0.25577374	0.438306315	0.340762006
ATMG00660	ORF149	hypothetical protein	2.24092505	2.139095893	2.985591297
AT4G30280	XTH18	putative xyloglucan endotransglycosylase/hydrolase	2.11126814	2.600857765	3.157659593
ATCG01080	NDHG	NADH dehydrogenase ND6	0.488133176	0.493761225	0.315093108
AT1G49450	F13F21.11	WD-40 repeat family protein	0.414313966	0.437536454	0.31301478
AT3G15500	ANAC055	ATAF-like NAC-domain transcription factor	2.741069615	3.774577792	3.354145323
AT5G57560	TCH4	cell wall-modifying enzyme	4.066236575	2.408751081	3.515404045
AT5G06760	LEA4-5	Late Embryogenesis Abundant (LEA) protein	0.325579715	0.482365163	0.278712361
AT5G19970	F28I16.120	unknown protein	0.441004556	0.482347957	0.253378565
AT3G21890	MZN24.1	zinc finger (B-box type) family protein	0.289028052	0.396127357	0.245970897

AT2G34600	JAZ7	JASMONATE-ZIM-DOMAIN PROTEIN 7 (JAZ7)	2.084087478	3.235583289	4.272715586
AT3G15310	K7L4.11	transposable element gene	0.393247672	0.390359131	0.228510741
AT5G67080	K21H1.4	member of MEKK subfamily	2.230548728	2.357413599	4.416346426
AT3G50400	F11C1.240	GDSL-motif lipase/hydrolase family protein	0.4698502	0.425603141	0.219172117
AT4G17500	ATERF-1	ERF/AP2 transcription factor family (ATERF-1)	2.324806214	2.657752241	5.589583807
AT3G56400	ATWRKY70	member of WRKY Transcription Factor; Group III	2.373261507	2.215291344	5.649160151
AT3G50770	CML41	calmodulin-related protein	2.306481662	3.019409918	5.860379987
AT2G15020	T15J14.6	unknown protein	0.275939268	0.479698696	0.161538019
AT1G21910	T26F17.14	DREB subfamily A-5 of ERF/AP2 transcription factor	4.209843163	2.122553553	6.218971536
AT1G30135	JAZ8	JASMONATE-ZIM-DOMAIN PROTEIN 8 (JAZ8)	3.066870875	3.027600111	6.233893426
AT1G80840	ATWRKY40	Pathogen-induced transcription factor	2.816636138	4.914388945	6.71074283
AT1G35140	EXL7	At1g35140 (At1g35140/T32G9_32) mRNA	5.28278199	2.337524051	6.876368829
AT4G23210	CRK13	Cysteine-rich receptor-like kinase (CRK13)	2.452198822	2.319289919	7.085150467
AT1G72900	F3N23.10	disease resistance protein (TIR-NBS class)	2.127202337	2.260644474	7.507154118
AT2G26150	HSFA2	Heat Stress Transcription Factor (Hsf)	0.493388588	0.306077995	7.632528005
AT2G24850	TAT3	tyrosine aminotransferase	2.08231324	3.139669118	8.039911227
AT1G33960	AIG1	induced by avirulence gene avrRpt2 and RPS2	2.12316172	2.149802172	11.84366478
AT5G39580	MIJ24.50	peroxidase	5.996284159	2.547202809	16.4084679
AT3G23250	ATMYB15	Member of the R2R3 factor gene family	2.457021338	3.103335282	16.97698154
AT1G69930	T17F3.4	glutathione transferase belonging to the tau class of GSTs	5.625498213	3.523364831	23.83606958
AT2G02990	RNS1	member of the ribonuclease T2 family	2.913970007	2.344369894	31.46213488

References

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15, 63-78.

Abe, H., YamaguchiShinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., and Shinozaki, K. (1997). Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell 9, 1859-1868.

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

Arigoni, D., Sagner, S., Latzel, C., Eisenreich, W., Bacher, A., and Zenk, M.H. (1997). Terpenoid biosynthesis from 1-deoxy-D-xylulose in higher plants by intramolecular skeletal rearrangement. Proc Natl Acad Sci U S A 94, 10600-10605.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics 25, 25-29.

Bainbridge, M.N., Warren, R.L., Hirst, M., Romanuik, T., Zeng, T., Go, A., Delaney, A., Griffith, M., Hickenbotham, M., Magrini, V., et al. (2006). Analysis of the

prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. BMC Genomics 7, -.

Bao, X., Katz, S., Pollard, M., and Ohlrogge, J. (2002). Carbocyclic fatty acids in plants: biochemical and molecular genetic characterization of cyclopropane fatty acid synthesis of Sterculiafoetida. Proc Natl Acad Sci U S A 99, 7172-7177.

Baraldi, R., Isacchi, B., Predieri, S., Marconi, G., Vincieri, F.F., and Bilia, A.R. (2008). Distribution of artemisinin and bioactive flavonoids from Artemisia annua L. during plant growth. Biochem Syst Ecol 36, 340-348.

Baudry, A., Heim, M.A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. Plant Journal 39, 366-380.

Bertea, C.M., Freije, J.R., van der Woude, H., Verstappen, F.W.A., Perk, L., Marquez, V., De Kraker, J.W., Posthumus, M.A., Jansen, B.J.M., de Groot, A., et al. (2005). Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in Artemisia annua. Planta Med 71, 40-47.

Bertea, C.M., Voster, A., Verstappen, F.W., Maffei, M., Beekwilder, J., and Bouwmeester, H.J. (2006a). Isoprenoid biosynthesis in Artemisia annua: cloning and heterologous expression of a germacrene A synthase from a glandular trichome cDNA library. Arch Biochem Biophys 448, 3-12.

Bertea, C.M., Voster, A., Verstappen, F.W.A., Maffei, M., Beekwilder, J., and Bouwmeester, H.J. (2006b). Isoprenoid biosynthesis in Artemisia annua: Cloning and heterologous expression of a germacrene A synthase from a glandular trichome cDNA library. Arch Biochem Biophys 448, 3-12.

Bouwmeester, H.J., Wallaart, T.E., Janssen, M.H.A., van Loo, B., Jansen, B.J.M., Posthumus, M.A., Schmidt, C.O., De Kraker, J.W., Konig, W.A., and Franssen, M.C.R. (1999). Amorpha-4,11-diene synthase catalyses the first probable step in artemisinin biosynthesis. Phytochemistry 52, 843-854.

Bouyer, D., Geier, F., Kragler, F., Schnittger, A., Pesch, M., Wester, K., Balkunde, R., Timmer, J., Fleck, C., and Hülskamp, M. (2008). Two-dimensional patterning by a trapping/depletion mechanism: the role of TTG1 and GL3 in Arabidopsis trichome formation. PLoS Biol 6, e141.

Brand, U., Fletcher, J., Hobe, M., Meyerowitz, E., and Simon, R. (2000). Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity. Science 289, 617-619.

Brueggemann, J., Weisshaar, B., and Sagasser, M. (2010). A WD40-repeat gene from Malus x domestica is a functional homologue of Arabidopsis thaliana TRANSPARENT TESTA GLABRA1. Plant Cell Rep 29, 285-294.

Bu, Q.Y., Jiang, H.L., Li, C.B., Zhai, Q.Z., Zhang, J.Y., Wu, X.Q., Sun, J.Q., Xie, Q., and Li, C.Y. (2008). Role of the Arabidopsis thaliana NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses. Cell Res 18, 756-767.

Buer, C., and Djordjevic, M. (2009). Architectural phenotypes in the transparent testa mutants of Arabidopsis thaliana. J Exp Bot 60, 751-763.

Cai, Y., Jia, J.W., Crock, J., Lin, Z.X., Chen, X.Y., and Croteau, R. (2002). A cDNA clone for beta-caryophyllene synthase from Artemisia annua. Phytochemistry 61, 523-529.

Campos, M.L., de Almeida, M., Rossi, M.L., Martinelli, A.P., Litholdo Junior, C.G., Figueira, A., Rampelotti-Ferreira, F.T., Vendramim, J.D., Benedito, V.A., and Peres, L.E. (2009). Brassinosteroids interact negatively with jasmonates in the formation of anti-herbivory traits in tomato. J Exp Bot 60, 4347-4361.

Carey, C., Strahle, J., Selinger, D., and Chandler, V. (2004). Mutations in the pale aleurone color1 regulatory gene of the Zea mays anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in Arabidopsis thaliana. Plant Cell 16, 450-464.

Carles, C.C., and Fletcher, J.C. (2003). Shoot apical meristem maintenance: the art of a dynamic balance. Trends in Plant Science 8, 394-401.

Caro, E., Castellano, M.M., and Gutierrez, C. (2007). A chromatin link that couples cell division to root epidermis patterning in Arabidopsis. Nature 447, 213-217.

Chang, Y.J., Song, S.H., Park, S.H., and Kim, S.U. (2000). Amorpha-4,11-diene synthase of Artemisia annua: cDNA isolation and bacterial expression of a terpene synthase involved in artemisinin biosynthesis. Arch Biochem Biophys 383, 178-184.

Chen, C., and Chen, Z. (2002). Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen-induced Arabidopsis transcription factor. Plant Physiol 129, 706-716.

Chen, W., Provart, N.J., Glazebrook, J., Katagiri, F., Chang, H.S., Eulgem, T., Mauch, F., Luan, S., Zou, G., Whitham, S.A., et al. (2002). Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell 14, 559-574.

Chen, Z.H., Nimmo, G.A., Jenkins, G.I., and Nimmo, H.G. (2007). BHLH32 modulates several biochemical and morphological processes that respond to Pi starvation in Arabidopsis. Biochem J 405, 191-198.

Cheung, F., Haas, B.J., Goldberg, S.M.D., May, G.D., Xiao, Y.L., and Town, C.D. (2006). Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology. BMC Genomics 7, -.

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B.H., Hong, X.H., Agarwal, M., and Zhu, J.K. (2003). ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Gene Dev 17, 1043-1054.

Choi, Y.E., Harada, E., Wada, M., Tsuboi, H., Morita, Y., Kusano, T., and Sano, H. (2001). Detoxification of cadmium in tobacco plants: formation and active excretion of crystals containing cadmium and calcium through trichomes. Planta 213, 45-50.

Covello, P.S., Teoh, K.H., Polichuk, D.R., Reed, D.W., and Nowak, G. (2007). Functional genomics and the biosynthesis of artemisinin. Phytochemistry 68, 1864-1871.

De Jesus-Gonzalez, L., and Weathers, P.J. (2003). Tetraploid Artemisia annua hairy roots produce more artemisinin than diploids. Plant Cell Reports 21, 809-813.

de Vetten, N., Quattrocchio, F., Mol, J., and Koes, R. (1997). The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. Genes Dev 11, 1422-1434.

Devaiah, B.N., Karthikeyan, A.S., and Raghothama, K.G. (2007). WRKY75 transcription factor is a modulator of phosphate acquisition and root development in Arabidopsis. Plant Physiol 143, 1789-1801.

Disch, A., Hemmerlin, A., Bach, T.J., and Rohmer, M. (1998). Mevalonate-derived isopentenyl diphosphate is the biosynthetic precursor of ubiquinone prenyl side chain in tobacco BY-2 cells. Biochem J 331 (Pt 2), 615-621.

Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., Boland, W., and Gershenzon, J. (2005). The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. Proc Natl Acad Sci U S A 102, 933-938.

Duke, M.V., Paul, R.N., Elsohly, H.N., Sturtz, G., and Duke, S.O. (1994). Localization of Artemisinin and Artemisitene in Foliar Tissues of Glanded and Glandless Biotypes of Artemisia-Annua L. Int J Plant Sci 155, 365-372.

Duke, S.O., and Paul, R.N. (1993). Development and Fine-Structure of the Glandular Trichomes of Artemisia-Annua L. Int J Plant Sci 154, 107-118.

Eckstein-Ludwig, U., Webb, R.J., van Goethem, I.D.A., East, J.M., Lee, A.G., Kimura, M., O'Neill, P.M., Bray, P.G., Ward, S.A., and Krishna, S. (2003). Artemisinins target the SERCA of Plasmodium falciparum. Nature 424, 957-961.

Emrich, S.J., Barbazuk, W.B., Li, L., and Schnable, P.S. (2007). Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res 17, 69-73.

Fairchild, C.D., Schumaker, M.A., and Quail, P.H. (2000). HFR1 encodes an atypical bHLH protein that acts in phytochrome A signal transduction. Gene Dev 14, 2377-2391.

Faiz, M.A., Bin Yunus, E., Rahman, M.R., Islam, F., Hoque, M.G., Hasan, M.U., Samad, R., Aung, S., Thein, S., Than, M., et al. (2005). Artesunate versus quinine for treatment of severe falciparum malaria: a randomised trial. Lancet 366, 717-725.

Feyissa, D., Løvdal, T., Olsen, K., Slimestad, R., and Lillo, C. (2009). The endogenous GL3, but not EGL3, gene is necessary for anthocyanin accumulation as induced by nitrogen depletion in Arabidopsis rosette stage leaves. Planta 230, 747-754.

Friedrichsen, D.M., Nemhauser, J., Muramitsu, T., Maloof, J.N., Alonso, J., Ecker, J.R., Furuya, M., and Chory, J. (2002). Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. Genetics 162, 1445-1456.

Galway, M., Masucci, J., Lloyd, A., Walbot, V., Davis, R., and Schiefelbein, J. (1994). The TTG gene is required to specify epidermal cell fate and cell patterning in the Arabidopsis root. Dev Biol 166, 740-754.

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

Heisler, M.G.B., Atkinson, A., Bylstra, Y.H., Walsh, R., and Smyth, D.R. (2001). SPATULA, a gene that controls development of carpel margin tissues in Arabidopsis, encodes a bHLH protein. Development 128, 1089-1098.

Hirai, N., Yoshida, R., Todoroki, Y., and Ohigashi, H. (2000). Biosynthesis of abscisic acid by the non-mevalonate pathway in plants, and by the mevalonate pathway in fungi. Biosci Biotechnol Biochem 64, 1448-1458.

Hua, L., and Matsuda, S.P. (1999). The molecular cloning of 8-epicedrol synthase from Artemisia annua. Arch Biochem Biophys 369, 208-212.

Huang, X.Q., and Madan, A. (1999). Cap3: A DNA Sequence Assembly Program. Genome Res 9, 868-877.

Humphries, J., Walker, A., Timmis, J., and Orford, S. (2005). Two WD-repeat genes from cotton are functional homologues of the Arabidopsis thaliana TRANSPARENT TESTA GLABRA1 (TTG1) gene. Plant Mol Biol 57, 67-81.

Huq, E., Tepperman, J.M., and Quail, P.H. (2000). GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis. P Natl Acad Sci USA 97, 9789-9794.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Mark Welch, D. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biology 8, -.

Jeong, S., Trotochaud, A.E., and Clark, S.E. (1999). The Arabidopsis CLAVATA2 gene encodes a receptor-like protein required for the stability of the CLAVATA1 receptor-like kinase. Plant Cell 11, 1925-1934.

Jia, J.W., Crock, J., Lu, S., Croteau, R., and Chen, X.Y. (1999). (3R)-Linalool synthase from Artemisia annua L.: cDNA isolation, characterization, and wound induction. Arch Biochem Biophys 372, 143-149.

Jung, M., Kim, H., Nam, K.Y., and No, K.T. (2005). Three-dimensional structure of Plasmodium falciparum Ca2+-ATPase(PfATP6) and docking of artemisinin derivatives to PfATP6. Bioorg Med Chem Lett 15, 2994-2997.

Kalde, M., Barth, M., Somssich, I.E., and Lippok, B. (2003). Members of the Arabidopsis WRKY group III transcription factors are part of different plant defense signaling pathways. Mol Plant Microbe Interact 16, 295-305.

Kang, Y.H., Kirik, V., Hulskamp, M., Nam, K.H., Hagely, K., Lee, M.M., and Schiefelbein, J. (2009). The MYB23 gene provides a positive feedback loop for cell fate specification in the Arabidopsis root epidermis. Plant Cell 21, 1080-1094.

Kannan, R., Kumar, K., Sahal, D., Kukreti, S., and Chauhan, V.S. (2005). Reaction of artemisinin with haemoglobin: implications for antimalarial activity. Biochemical Journal 385, 409-418.

Kim, H.S., and Delaney, T.P. (2002). Over-expression of TGA5, which encodes a bZIP transcription factor that interacts with NIM1/NPR1, confers SAR-independent resistance in Arabidopsis thaliana to Peronospora parasitica. Plant Journal 32, 151-163.

Kirik, V., Simon, M., Huelskamp, M., and Schiefelbein, J. (2004a). The ENHANCER OF TRY AND CPCl gene acts redundantly with TRIPTYCHON and CAPRICE in trichome and root hair cell patterning in Arabidopsis. Developmental Biology 268, 506-513.

Kirik, V., Simon, M., Wester, K., Schiefelbein, J., and Hulskamp, M. (2004b). ENHANCER of TRY and CPC 2 (ETC2) reveals redundancy in the region-specific control of trichome development of Arabidopsis. Plant Molecular Biology 55, 389-398.

Kunieda, T., Mitsuda, N., Ohme-Takagi, M., Takeda, S., Aida, M., Tasaka, M., Kondo, M., Nishimura, M., and Hara-Nishimura, I. (2008). NAC Family Proteins NARS1/NAC2 and NARS2/NAM in the Outer Integument Regulate Embryogenesis in Arabidopsis. Plant Cell 20, 2631-2642.

Kupper, H., Lombi, E., Zhao, F.J., and McGrath, S.P. (2000). Cellular compartmentation of cadmium and zinc in relation to other elements in the hyperaccumulator Arabidopsis halleri. Planta 212, 75-84.

Kwak, S.H., Shen, R., and Schiefelbein, J. (2005). Positional signaling mediated by a receptor-like kinase in Arabidopsis. Science 307, 1111-1113.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Laule, O., Furholz, A., Chang, H.S., Zhu, T., Wang, X., Heifetz, P.B., Gruissem, W., and Lange, M. (2003). Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in Arabidopsis thaliana. Proc Natl Acad Sci U S A 100, 6866-6871.

Li, L., Zhao, Y., McCaig, B.C., Wingerd, B.A., Wang, J., Whalon, M.E., Pichersky, E., and Howe, G.A. (2004). The tomato homolog of CORONATINE-INSENSITIVE1 is required for the maternal control of seed maturation, jasmonate-signaled defense responses, and glandular trichome development. Plant Cell 16, 126-143.

Lichtenthaler, H.K., Schwender, J., Disch, A., and Rohmer, M. (1997). Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. FEBS Lett 400, 271-274.

Liu, C.Z., Wang, Y.C., Ouyang, F., Ye, H.C., and Li, G.F. (1999). Production of artemisinin by Artemisia annua hairy root culture in a internal loop airlift bioreactor. Acta Bot Sin 41, 181-183.

Lloyd, A., Walbot, V., and Davis, R. (1992). Arabidopsis and Nicotiana anthocyanin production activated by maize regulators R and C1. Science 258, 1773-1775.

Luo, X.D., and Shen, C.C. (1987). The chemistry, pharmacology, and clinical applications of qinghaosu (artemisinin) and its derivatives. Med Res Rev 7, 29-52.

Ma, C., Wang, H., Lu, X., Li, H., Liu, B., and Xu, G. (2007). Analysis of Artemisia annua L. volatile oil by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. J Chromatogr A 1150, 50-53.

Ma, C., Wang, H., Lu, X., Xu, G., and Liu, B. (2008). Metabolic fingerprinting investigation of Artemisia annua L. in different stages of development by gas chromatography and gas chromatography-mass spectrometry. J Chromatogr A 1186, 412-419.

Maes, L., and Goossens, A. (2010). Hormone-mediated promotion of trichome initiation in plants is conserved but utilizes species- and trichome-specific regulatory mechanisms. Plant Signal Behav 5, 205-207.

Maes, L., Inze, D., and Goossens, A. (2008). Functional specialization of the TRANSPARENT TESTA GLABRA1 network allows differential hormonal control

of laminal and marginal trichome initiation in Arabidopsis rosette leaves. Plant Physiol 148, 1453-1464.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z.T., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376-380.

Maruyama, K., Takeda, M., Kidokoro, S., Yamada, K., Sakuma, Y., Urano, K., Fujita, M., Yoshiwara, K., Matsukura, S., Morishita, Y., et al. (2009). Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A. Plant Physiol 150, 1972-1980.

Matsushita, Y., Kang, W., and Charlwood, B.V. (1996). Cloning and analysis of a cDNA encoding farnesyl diphosphate synthase from Artemisia annua. Gene 172, 207-209.

Mercke, P., Bengtsson, M., Bouwmeester, H.J., Posthumus, M.A., and Brodelius, P.E. (2000). Molecular cloning, expression, and characterization of amorpha-4,11-diene synthase, a key enzyme of artemisinin biosynthesis in Artemisia annua L. Arch Biochem Biophys 381, 173-180.

Mercke, P., Crock, J., Croteau, R., and Brodelius, P.E. (1999). Cloning, expression, and characterization of epi-cedrol synthase, a sesquiterpene cyclase from Artemisia annua L. Arch Biochem Biophys 369, 213-222.

Miao, Y., and Jiang, L. (2007). Transient expression of fluorescent fusion proteins in protoplasts of suspension cultured cells. Nat Protoc 2, 2348-2353.

Moore, M.J., Dhingra, A., Soltis, P.S., Shaw, R., Farmerie, W.G., Folta, K.M., and Soltis, D.E. (2006). Rapid and accurate pyrosequencing of angiosperm plastid genomes. Bmc Plant Biology 6, -.

Morita, Y., Saitoh, M., Hoshino, A., Nitasaka, E., and Iida, S. (2006). Isolation of cDNAs for R2R3-MYB, bHLH and WDR transcriptional regulators and identification of c and ca mutations conferring white flowers in the Japanese morning glory. Plant Cell Physiol 47, 457-470.

Morohashi, K., and Grotewold, E. (2009). A systems approach reveals regulatory circuitry for Arabidopsis trichome initiation by the GL3 and GL1 selectors. PLoS Genet 5, e1000396.

Morohashi, K., Zhao, M., Yang, M., Read, B., Lloyd, A., Lamb, R., and Grotewold, E. (2007). Participation of the Arabidopsis bHLH factor GL3 in trichome initiation regulatory events. Plant Physiol 145, 736-746.

Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M., and 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.

Ni, M., Tepperman, J.M., and Quail, P.H. (1998). PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. Cell 95, 657-667.

Olsson, M.E., Olofsson, L.M., Lindahl, A.L., Lundgren, A., Brodelius, M., and Brodelius, P.E. (2009). Localization of enzymes of artemisinin biosynthesis to the apical cells of glandular secretory trichomes of Artemisia annua L. Phytochemistry 70, 1123-1128.

Oppenheimer, D.G., Herman, P.L., Sivakumaran, S., Esch, J., and Marks, M.D. (1991). A Myb Gene Required for Leaf Trichome Differentiation in Arabidopsis Is Expressed in Stipules. Cell 67, 483-493.

Pang, Y., Wenger, J., Saathoff, K., Peel, G., Wen, J., Huhman, D., Allen, S., Tang, Y., Cheng, X., Tadege, M., et al. (2009). A WD40 repeat protein from Medicago truncatula is necessary for tissue-specific anthocyanin and proanthocyanidin biosynthesis but not for trichome development. Plant Physiol 151, 1114-1129.

Payne, C., Zhang, F., and Lloyd, A. (2000). GL3 encodes a bHLH protein that regulates trichome development in arabidopsis through interaction with GL1 and TTG1. Genetics 156, 1349-1362.

Peer, W., and Murphy, A. (2007). Flavonoids and auxin transport: modulators or regulators? Trends Plant Sci 12, 556-563.

Pertea, G., Huang, X.Q., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., et al. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19, 651-652.

Picaud, S., Brodelius, M., and Brodelius, P.E. (2005). Expression, purification and characterization of recombinant (E)-beta-farnesene synthase from Artemisia annua. Phytochemistry 66, 961-967.

Poinar, H.N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R.D.E., Buigues, B., Tikhonov, A., Huson, D.H., Tomsho, L.P., Auch, A., et al. (2006). Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. Science 311, 392-394.

Putalun, W., Luealon, W., De-Eknamkul, W., Tanaka, H., and Shoyama, Y. (2007). Improvement of artemisinin production by chitosan in hairy root cultures of Artemisia annua L. Biotechnol Lett 29, 1143-1146.

Qian, Z.H., Gong, K., Zhang, L., Lv, J.B., Jing, F.Y., Wang, Y.Y., Guan, S.B., Wang, G.F., and Tang, K.X. (2007). A simple and efficient procedure to enhance artemisinin content in Artemisia annua L. by seeding to salinity stress. Afr J Biotechnol 6, 1410-1413.

Rajani, S., and Sundaresan, V. (2001). The Arabidopsis myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. Current Biology 11, 1914-1922.

Ramsay, N., and Glover, B. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci 10, 63-70.

Ranger, C.M., and Hower, A.A. (2001). Role of the glandular trichomes in resistance of perennial alfalfa to the potato leafhopper (Homoptera : Cicadellidae). J Econ Entomol 94, 950-957.

Rausher, M.D. (2001). Co-evolution and plant resistance to natural enemies. Nature 411, 857-864.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., et al. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290, 2105-2110.

Riley, E. (1995). Malaria Vaccine Trials - Spf66 and All That. Curr Opin Immunol 7, 612-616.

Ro, D.K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440, 940-943.

Sambrook, J., and Russell, D.W. (2006). The condensed protocols from Molecular cloning : a laboratory manual (Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press).

Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, A., Thumfahrt, J., Jurgens, G., and Hulskamp, M. (2002). TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in Arabidopsis. Embo Journal 21, 5036-5046.

Schiefelbein, J. (2003). Cell-fate specification in the epidermis: a common patterning mechanism in the root and shoot. Curr Opin Plant Biol 6, 74-78.

Schlereth, A., Moller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E.H., Schmid, M., Jurgens, G., and Weijers, D. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. Nature 464, 913-916.

Schoof, H., Lenhard, M., Haecker, A., Mayer, K.F., Jurgens, G., and Laux, T. (2000). The stem cell population of Arabidopsis shoot meristems in maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. Cell 100, 635-644.

Schwender, J., Zeidler, J., Groner, R., Muller, C., Focke, M., Braun, S., Lichtenthaler, F.W., and Lichtenthaler, H.K. (1997). Incorporation of 1-deoxy-D-xylulose into isoprene and phytol by higher plants and algae. FEBS Lett 414, 129-134.

Singh, N.P., and Lai, H.C. (2005). Synergistic cytotoxicity of artemisinin and sodium butyrate on human cancer cells. Anticancer Res 25, 4325-4331.

Smith, T., Gaitatzes, C., Saxena, K., and Neer, E. (1999). The WD repeat: a common architecture for diverse functions. Trends Biochem Sci 24, 181-185.

Smolen, G.A., Pawlowski, L., Wilensky, S.E., and Bender, J. (2002). Dominant Alleles of the basic helix-loop-helix transcription factor ATR2 activate stress-responsive genes in Arabidopsis. Genetics 161, 1235-1246.

Sompornpailin, K., Makita, Y., Yamazaki, M., and Saito, K. (2002). A WD-repeat-containing putative regulatory protein in anthocyanin biosynthesis in Perilla frutescens. Plant Mol Biol 50, 485-495.

Sorensen, A.M., Krober, S., Unte, U.S., Huijser, P., Dekker, K., and Saedler, H. (2003). The Arabidopsis ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. Plant Journal 33, 413-423.

Teoh, K.H., Polichuk, D.R., Reed, D.W., and Covello, P.S. (2009). Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in Artemisia annua. Botany 87, 635-642.

Teoh, K.H., Polichuk, D.R., Reed, D.W., Nowak, G., and Covello, P.S. (2006). Artemisia annua L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a

cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. Febs Letters 580, 1411-1416.

Tran, L.S., Nakashima, K., Sakuma, Y., Simpson, S.D., Fujita, Y., Maruyama, K., Fujita, M., Seki, M., Shinozaki, K., and 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.

Trotochaud, A.E., Hao, T., Wu, G., Yang, Z., and Clark, S.E. (1999). The CLAVATA1 receptor-like kinase requires CLAVATA3 for its assembly into a signaling complex that includes KAPP and a Rho-related protein. Plant Cell 11, 393-406.

Uhlemann, A.C., Cameron, A., Eckstein-Ludwig, U., Fischbarg, J., Iserovich, P., Zuniga, F.A., East, M., Lee, A., Brady, L., Haynes, R.K., et al. (2005). A single amino acid residue can determine the sensitivity of SERCAs to artemisinins. Nat Struct Mol Biol 12, 628-629.

Van Nieuwerburgh, F.C.W., Casteele, S.R.V., Maes, L., Goossens, A., Inze, D., Van Bocxlaer, J., and Deforce, D.L.D. (2006). Quantitation of artemisinin and its biosynthetic precursors in Artemisia annua L. by high performance liquid

chromatography - electrospray quadrupole time-of-flight tandem mass spectrometry. J Chromatogr A 1118, 180-187.

van Nocker, S., and Ludwig, P. (2003). The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. BMC Genomics 4, 50.

Wada, T., Tachibana, T., Shimura, Y., and Okada, K. (1997). Epidermal cell differentiation in Arabidopsis determined by a Myb homolog, CPC. Science 277, 1113-1116.

Wagner, G.J., Wang, E., and Shepherd, R.W. (2004). New approaches for studying and exploiting an old protuberance, the plant trichome. Ann Bot-London 93, 3-11.

Walker, A., Davison, P., Bolognesi-Winfield, A., James, C., Srinivasan, N., Blundell, T., Esch, J., Marks, M., and Gray, J. (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.

Wang, G., Tian, L., Aziz, N., Broun, P., Dai, X., He, J., King, A., Zhao, P.X., and Dixon, R.A. (2008). Terpene biosynthesis in glandular trichomes of hop. Plant Physiol 148, 1254-1266.

Wang, J., Huang, L.Y., Li, J., Fan, Q.W., Long, Y.C., Li, Y., and Zhou, B. (2010). Artemisinin Directly Targets Malarial Mitochondria through Its Specific Mitochondrial Activation. PLoS One 5, A158-A169.

Wang, W., Wang, Y., Zhang, Q., Qi, Y., and Guo, D. (2009). Global characterization of Artemisia annua glandular trichome transcriptome using 454 pyrosequencing. BMC Genomics 10, 465.

Weathers, P.J., DeJesus-Gonzalez, L., Kim, Y.J., Souret, F.F., and Towler, M.J. (2004). Alteration of biomass and artemisinin production in Artemisia annua hairy roots by media sterilization method and sugars. Plant Cell Reports 23, 414-418.

Weber, A.P.M., Weber, K.L., Carr, K., Wilkerson, C., and Ohlrogge, J.B. (2007). Sampling the arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiology 144, 32-42.

Wicker, T., Schlagenhauf, E., Graner, A., Close, T.J., Keller, B., and Stein, N. (2006). 454 sequencing put to the test using the complex genome of barley. BMC Genomics 7, -.

Woerdenbag, H.J., Pras, N., Bos, R., Visser, J.F., Hendriks, H., and Malingre, T.M. (1991). Analysis of Artemisinin and Related Sesquiterpenoids from Artemisia-Annua L by Combined Gas-Chromatography Mass-Spectrometry. Phytochem Analysis 2, 215-219.

Xie, Q., Frugis, G., Colgan, D., and Chua, N.H. (2000). Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev 14, 3024-3036.

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

Zhang, F., Gonzalez, A., Zhao, M., Payne, C., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development 130, 4859-4869.

Zhang, X., Henriques, R., Lin, S.S., Niu, Q.W., and Chua, N.H. (2006). Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat Protoc 1, 641-646.

Zhang, Y., Teoh, K.H., Reed, D.W., Maes, L., Goossens, A., Olson, D.J.H., Ross, A.R.S., and Covello, P.S. (2008). The molecular cloning of artemisinic aldehyde Delta 11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in Artemisia annua. Journal of Biological Chemistry 283, 21501-21508.

Zhao, M., Morohashi, K., Hatlestad, G., Grotewold, E., and Lloyd, A. (2008). The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. Development 135, 1991-1999.