University of Northern Colorado
Scholarship & Creative Works @ Digital UNC

Dissertations | Student Research

8-1-2011

Transformation of Arabidopsis thaliana with heat stress-related genes from a Copaifera officinalis expressed sequence tag library

Samuel R. Zwenger

Follow this and additional works at: http://digscholarship.unco.edu/dissertations

Recommended Citation

Zwenger, Samuel R., "Transformation of Arabidopsis thaliana with heat stress-related genes from a Copaifera officinalis expressed sequence tag library" (2011). *Dissertations.* Paper 294.

This Text is brought to you for free and open access by the Student Research at Scholarship & Creative Works @ Digital UNC. It has been accepted for inclusion in Dissertations by an authorized administrator of Scholarship & Creative Works @ Digital UNC. For more information, please contact Jane.Monson@unco.edu.

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

TRANSFORMATION OF *ARABIDOPSIS THALIANA* WITH HEAT STRESS-RELATED GENES FROM A *COPAIFERA OFFICINALIS* EXPRESSED SEQUENCE TAG LIBRARY

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Samuel R. Zwenger

College of Natural and Health Sciences
School of Biological Sciences

August, 2011

THIS DISSERTATION WAS SPONSORED

BY

____

Chhandak Basu, Ph.D.

Samuel R. Zwenger

DISSERTATION COMMITTEE

Advisory Professor____

Susan Keenan, Ph.D.

Advisory Professor ____

Mitchell McGlaughlin, Ph.D.

Faculty Representative ____

Aichun Dong, Ph.D.

ASSISTANT VICE PRESIDENT FOR RESEARCH & EXTENDED STUDIES

DEAN OF THE GRADUATE SCHOOL

____

Robbyn R. Wacker, Ph.D.

Examination Date of Dissertation ____

ABSTRACT

Zwenger, Samuel R. *Transformation of* Arabidopsis thaliana *with Heat Stress-Related Genes from a C*opaifera officinalis *Expressed Sequence Tag Library.* Published Doctor of Philosophy dissertation, University of Northern Colorado, 2011.

This dissertation had three components: (a) identification of genes involved in heat stress-related processes derived from a tropical plant known as the diesel tree (*Copaifera officinalis*) by means of an expressed sequence tag (EST) library, (b) stable transformation of *Arabidopsis thaliana* with *C. officinalis* heat stress-related genes, and (c) assessing the degree to which *C. officinalis* heat-stress associated genes provided thermotolerance in transgenic *A. thaliana*. This dissertation had important plant biotechnological components as well as agriculturally important themes. Global climate change is predicted to result in elevated temperatures over the next century; thus, studying the influence of heat stress upon plants is becoming increasingly important. Subsequently, in this work, it was hypothesized that heat stress-related transgenes derived from *C. officinalis* would provide some level of thermotolerance to *A. thaliana*. Various stress assays, which included both *in vitro* and *in vivo* assays, were performed to help determine the degree of thermotolerance in each transgenic line. A novel and key component in this project was that the sequences from *C. officinalis* cDNA were ligated into a plant binary vector, which eliminated the need to perform individual ligation reactions for each gene of interest (GOI). Sequencing plates were obtained and stored at -80˚C so individual wells containing the *E. coli* harboring the binary vector could be

picked and grown overnight for subsequent plasmid miniprep. Plasmids were transferred into *Agrobacterium tumefaciens* for generating transgenic *A. thaliana* lines. This created a streamlined method of generating multiple transgenic plants, each with a unique GOI. Although others have overexpressed heat stress-related genes in *A. thaliana*, this project overexpressed genes from the tropical plant *C. officinalis*, which is novel. The results suggested that overexpression of *C. officinalis* heat stress-related genes in *A. thaliana* helped confer thermotolerance. Furthermore, the sequenced *E. coli* clones harboring *C. officinalis* genes can be readily obtained and investigated in future studies, which offers a rich resource for transgenic studies.

ACKNOWLEDGEMENTS

Many people have given me encouragement during my time at the University of Northern Colorado (UNC). Primary assistance in developing the research project was given by my committee. I am grateful to Dr. Chhandak Basu for accepting me as a graduate student so I could study plant biotechnology. Dr. Mit McGlaughlin has been crucial in providing sound input to give the dissertation scientific rigor. Dr. Susan Keenan stimulated my interest in bioinformatics, which gave me the courage to explore software for data analysis. I admire Dr. Aichun Dong for his academic status and his depth of knowledge on biological macromolecules. With this said, it is important that I thank all the faculty and staff in the School of Biological Sciences at UNC. Each is dedicated to stimulating minds and maintaining productive research. Over the last four years, I have had great conversations with almost all of them. Multiple graduate and undergraduate students at UNC have contributed to my persistence in the doctoral program; however, two very important people helped me in the research lab: Eric Reinsvold and Erich Hoffmann.

Additional thanks go to Dr. Ana Conesa at the Bioinformatics Department of the Centro de Investigatión Príncipe Felipe, Spain. Her assistance guided the bioinformatics analysis of the EST library; therefore, a major part of this work would not have been possible without her help. I thank the Colorado Office of Economic Development and the U.S. Department of Education who provided funding via a Colorado Bioscience Grant and the Graduate Assistance in Areas of National Need (GAANN) Fellowship,

respectively. I am also extremely grateful for funding provided by the School of Biological Sciences at UNC.

Several people helped stimulate a curiosity in my formative years but two individuals stand out among all others. Richard R. Schmidt taught me the importance of farming and showed me that hard work contributes towards developing patience and understanding. Karen S. Schmidt instilled a scientific curiosity early in my childhood, which continues to this day. I remember shuffling through her collection of strange ocean invertebrate fossils and wondering how such organisms came to exist on the prairie. Taken together, both of these people helped guide my development as a scientist and independent thinker.

Others have helped me stay focused on research and school. The individuals that have contributed most include Eric Gillock and Greg Farley. More recently, Stephen and Jan Kitzis have been extremely influential in their guidance while I have been in graduate school. They contributed toward several areas of the dissertation and it will be impossible to repay them. Finally, I thank Matthew J. Pfannenstiel for his endless demonstration of how to live life to its fullest.

No person has been as crucial to my happiness and satisfaction as Lianne. Whether she realizes it or not, she lives by the scientific process and is guided by logic. Much of my happiness in life can be attributed to her patience, creative cooking, and positive energy.

TABLE OF CONTENTS

CHAPTER I. INTRODUCTION AND OVERVIEW 1

Heat Stress in Plants... 2
Heat Stress Response 4
Geographical Influences on Plant Stress... 6
Macromolecular Responses to Heat... 7
Heat Shock Proteins 9
Detecting Heat Stress 14
Importance of Studying Heat Stress in Plants 15
Copaifera Species 17
Generating Sequence Data 17
Summary of Dissertation... 20

CHAPTER II. GENERATION AND ANALYSIS OF A *COPAIFERA* EXPRESSED SEQUENCE TAG LIBRARY 22

Introduction 22
Copaifera officinalis 22
Expressed Sequence Tag Libraries and Sequence Analysis.............. 23
Materials and Methods... 26
Results 28
Discussion.. 40

CHAPTER III. SELECTION OF GENES AND PLANT TRANSFORMATION 49

Introduction 49
Obtaining and Cloning Transgenes... 49
Plant Transgenic Research and *Arabidopsis thaliana*.................................. 51
Agrobacterium tumefaciens and Gene Transfer 51
pCHF3 Binary Vector Description 53
The Floral Dip 53
Materials and Methods... 54
Choice of Constructs 54
Electroporation of *A. tumefaciens* 55
Transformation of *A. thaliana*... 56
cDNA Synthesis and Detection of Transgene Expression 56
Results 57
Discussion.. 64

Factors Affecting Transgene Expression 65
Benefits of the Binary Library Approach 67

CHAPTER IV. ASSESSING THERMOTOLERANCE IN TRANSGENIC LINES... 69

Introduction 69
Association of Heat Stress with Transgenes 70
Assessing for Thermotolerance 72
Materials and Methods 74
In vitro Thermotolerance Assays 74
Hypocotyl elongation assay 75
In vivo Thermotolerance Assays 75
Results 76
Discussion 83

CHAPTER V. CONCLUSIONS AND FUTURE PERSPECTIVES 87

Summary of the *C. officinalis* Expressed Sequence Tag Library 87
C. officinalis EST library compared to other libraries 88
Summary of the Transformation Experiments 90
Summary of Heat Stress Assays 90
Future Studies 92
Measuring Heat Stress 93
Transcript Expression and Expression Levels 95
Transcriptional Influences of Transgene Expression 96
Detecting Translation 98
Constructs from the *C. officinalis* Library in Other Plants 98

REFERENCES 101

APPENDIX A. pCHF3 VECTOR MAP 124

APPENDIX B. pCHF3 VECTOR SEQUENCE 126

APPENDIX C. GO TERMS OVERREPRESENTED IN THE EXPRESSED SEQUENCE TAG LIBRARY 132

APPENDIX D. GENES AND PREDICTED PROTEIN SEQUENCES USED IN TRANSFORMATION STUDIES 136

LIST OF TABLES

1. Forward and Reverse Detection Primers for Each Gene of Interest.................. 59

2. Each Insert Identification (ID) Correlates to Plate and Well Position in Which Construct Was Obtained 60

3. ANOVA Results for Hypocotyl Elongation Assay 83

LIST OF FIGURES

1. Comparison of the number of copaiba ESTs and their associated lengths 30

2. Distribution of E-values according to the number of hits from the BLASTx database 31

3. Distribution of BLASTx hits based on other species. 32

4. Species distribution for all BLASTx hits 33

5. Distribution of sequences according to each step in library analysis 34

6. GO-level distribution for annotations of copaiba unigenes 35

7. More than half (488) of the biological processes category consisted of annotations for cellular and metabolic processes 36

8. The molecular function category consisted of relatively few annotations for enzyme regulator activity (4) and molecular transducer activity (5). However, this category had nearly half (246) of GO terms associated with binding 37

9. A large portion of unigenes categorized as cellular component were either cell (308), cell part (306), or organelle (238)... 38

10. PCR results using gene-specific detection primers for each vector-insert construct 61

11. RT-PCR using gene-specific detection primers for each transgenic line. 62

12. RT-PCR of 4B07. 63

13. RT-PCR using *nptII* detection primers in each transgenic line 64

14. Gene specific detection primers using *A. thaliana* cDNA 65

15. Results of the *in vitro* assay two days after heat shock (Nt, nontransgenic)...... 78

16. Results of the *in vitro* assay five days after heat shock (Nt, nontransgenic)...... 79

17. Results of the *in vitro* assay two days after heat shock (Nt, nontransgenic)..... 80

18. Results for the *in vivo* thermotolerance assay. 81

19. Comparison of hypocotyl elongation between transgenic and nontransgenic (Nt) plants that were either unstressed (dark gray) or given a heat treatment (light gray) 2.5 days after removal from 4°C......................... 82

CHAPTER I

INTRODUCTION AND OVERVIEW

Plant species exhibit variation in how they respond to environmental stress. Stress response mechanisms can be thought of as physiological responses to changing environmental conditions. Climate change is expected to have a major impact on crop plants by imparting multiple abiotic stresses, which reduces crop yield (Ainsworth & Ort, 2010). In addition, rising CO_2 levels are expected to have an indirect negative impact on grain quality by shifting the carbon: nitrogen ratios in food plants (Porter & Semenov, 2005). Heat is expected to have the greatest negative impact by reducing crop production (Halford, 2009). Since plants are unable to relocate or reposition themselves under heat stress conditions, internal mechanisms have evolved to help stabilize intracellular machinery. However, some plants, such as crop species, are deficient in adaptive mechanisms to cope with the predicted increases of temperatures in agricultural areas (Ahuja, de Vos, Bones, & Hall, 2010; Long & Ort, 2010). Characterizing genes that help alleviate heat stress in plants and transferring and expressing them in crop plants is one of the main goals of agrobiotechnology.

The three aims of this dissertation were to (a) characterize heat stress-related genes derived from a tropical plant known as the diesel tree (*Copaifera officinalis*) by means of expressed sequence tag (EST) library construction and analysis, (b) genetically transform *Arabidopsis thaliana* with novel heat stress-related genes from *C. officinalis*,

and (c) determine the level of heat tolerance that *C. officinalis* heat-stress associated genes provide to transgenic *A. thaliana*. Subsequently, it was hypothesized that the transgenes derived from *C. officinalis* would provide thermotolerance to *A. thaliana*. Various heat stress assays, which include both *in vitro* and *in vivo* assays, can be used to help determine the degree of thermotolerance in transgenic lines. If the *C. officinalis* transgenes confer thermotolerance in *A. thaliana*, then the transgenes might be good candidates for expression in crop plants.

Heat Stress in Plants

The heat stress response is a phenomenon that has been studied in yeast (Geiler-Samerotte et al., 2011), mosses (Shi & Theg, 2010), crop plants (Grigorova, Vaseva, Demirevska, & Feller, 2011), fungi (Montero-Barrientos et al., 2010), and mammals (Lindquist, 1986). Several plants including gymnosperms and angiosperms have been investigated. Many of the genetic and mutational investigations of plant heat stress have used *A. thaliana* due to its small genome and short life cycle. Binelli and Mascarhenas (1990) established that temperatures of 50˚C or more can be lethal to *A. thaliana*.

Many proteins have been shown to help confer heat tolerance in *A. thaliana* and other plants and these have been reviewed by Wang, Vinocur, Shoseyov, and Altman (2004). These include proteins important in stabilizing cellular structures or assisting in protein folding such as the heat shock proteins (HSPs). Other protein processes function in metabolic pathways that can yield metabolic products such as osmoprotectants. Both of these phenomena--an increase in HSPs and changes in metabolic pathways--are observed when a plant is exposed to elevated temperatures (Hua, 2009).

Entire transcriptomes have been studied in an attempt to elucidate heat tolerance mechanisms in plants. In one approach using microarrays, partial gene sequences are covalently bound to a glass support and fluorescently labeled mRNA is used as a probe. While thousands of genes can be bound to the array, hybridization of the probes only occurs if a particular gene is present in the mRNA sample. Comparisons can then be drawn between control or treatment samples based on their fluorescent patterns (Schena, Shalon, Davis, & Brown, 1995). In a comprehensive study, Larkindale and Vierling (2008) provided microarray data comparing *A. thaliana* plants that were gradually heat stressed (22˚C-45˚C) over a period of six hours and another group that was heat stressed without acclimation at 45˚C for 90 minutes to *A. thaliana* plants that were given a heat shock treatment of 38˚C for 90 minutes, 22˚C for 120 minutes, and 45˚C for 90 minutes. The three different treatments were expected to reveal the different cellular mechanisms associated with different heat acclimation periods. Plants that were given six hours to adjust to temperature changes showed greater survival compared to the plants given step-wise treatment. The data generated from these experiments aided researchers by identifying groups of genes that were involved in various levels of heat stress. Notably, plants that were gradually acclimated to elevated temperatures tended to express fewer heat stress-associated genes compared to plants given a shorter time to acclimate. These data suggested gradual changes such as progressive stabilization of the cytoskeleton and manageable repair processes were brought about within the cell as the temperature gradually increased which imparted preparedness to the plant.

Six different heat shock factors (HSFs) in Larkindale and Vierling's (2008) study were identified andHSFs helped regulate expression of heat stress-associated genes

(Chinnusamy & Zhu, 2009); the family of heat stress-associated transcription factors themselves was quite diverse (von Koskull-Döring et al., 2007). There are three classes of plant HSFs (A, B, and C) based on their oligomerization domains (Baniwal et al., 2004). These small proteins bind to promoter sites on DNA upstream of the target gene to induce expression of heat stress-associated genes. In addition, many other small non-regulatory proteins (i.e., heat shock proteins) that have been studied have been found to play an important role in the heat stress response (Sun, Van Montagu, & Verbruggen, 2002).

Heat Stress Response

In plants, the heat stress response has been intensely investigated; yet much remains to be discovered. Many physiological changes including cellular and metabolic reactions occur when a plant is exposed to elevated temperatures (Hua, 2009). For example, plant physiological responses might include stomatal closure while cellular responses might involve changes in gene expression. In particular, the influence of heat on gene regulation and expression has been studied extensively in plants (Frank et al., 2009; Kotak et al., 2007).

Epigenetic modifications--changes in gene expression without altering the sequence of bases--include processes such as histone modifications and DNA methylation and have also been suggested to occur after heat exposure (von Koskull-Döring, Scharf, & Nover, 2007). Although the genetic sequence itself is not changed, epigenetic modifications can influence whether transcription factors (and thus polymerase) can access a promoter site. Since lysine is positively charged and can bind negatively charged DNA, acetylation of histones renders the positive charge neutral and

thus loosens the packaging of DNA (Tian & Chen, 2001). Other histone tail modifications include phosphorylation, methylation, ubiquitination, and SUMOylation (Chinnusamy & Zhu, 2009). Addition of methyl groups to DNA bases is generally associated with decreased rates of transcription (von Kosgull-Doring et al., 2007). It has been shown that methylation patterns change in plants subjected to many types of stress (e.g., radiation, heat, osmotic changes); these epigenetic modifications can be passed onto subsequent generations (Boyko & Kovalchuk, 2011). In cold-stressed plants, decreased methylation of cytosine has been suggested to positively influence the movement of transposable elements (Hashida et al., 2006). All of these types of epigenetic modification can be induced upon exposure to heat stress and each mechanism serves to alter gene expression.

Interestingly, after exposing plants to abiotic stresses such as radiation, there is an increase in homologous recombination in successive generations (Molinari et al., 2007). This might suggest that other abiotic stresses (e.g., heat stress) might also boost such recombination events to help generate genomic diversity, which might be advantageous to subsequent generations. However, many other abiotic stresses besides heat stress influence the transcriptional machinery and enzymes within plant cells including water deficit, salinity, flooding, soil compaction, high light intensity, and oxidative stress.

One important consideration to plant stress research is how to determine what classifies as stress. Importantly, any stress condition is relative to the location in which that species has evolved. To some degree, all environments can be stressful. It has been suggested that many enzymes in plants have an optimal thermal range of 10ºC above or below their normal growing conditions (Mahan, McMichael, & Wanjura, 1995); enzymes

show reduced catalytic activity and stability outside of their optimal thermal range (Langridge, 1968). These insights might prove helpful when considering how to define stress. In general, stress conditions might best be defined as suboptimal growing conditions that have a negative impact on plant growth, homeostasis, or development. Pahlich (1993) and several others continue to cite Walter Larcher (1987) who provided a rather descriptive definition of plant stress when he stated the following:

> A state in which increasing demands made upon a plant lead to an initial destabilization of functions, followed by normalization and improved resistance. If the limits of tolerance are exceeded and the adaptive capacity is overtaxed, permanent damage or even death may result. (p. 160)

Based on this definition, it follows that geographical location imparts selective pressure on the evolution of adaptive mechanisms for environmental factors.

Geographic Influences on Plant Stress

Tropical areas are defined as latitudes within 23.5˚N and 23.5˚S of the equator. Plant species that inhabit tropical ecosystems are interesting since they must carry out important physiological functions in the face of high humidity and low air circulation, both of which restrict evapotranspiration and thus inhibit cooling (Taiz & Zeiger, 2010). Limited data have been published regarding the optimal temperature ranges of *C. officinalis*. In one study, *Copaifera aromatica* was grown at an average temperature of 33˚C+/-0.3˚C with a range of 26˚C -39˚C and a mean relative humidity of 63+/-0.3% with a range of 51-83% (Lovelock, Virgo, Popp, & Winter, 1999). Although the authors were not interested in the heat stress response, the environmental conditions gave an estimate of ideal growth temperatures for *Copaifera*. One might hypothesize that since *Copaifera species* are found in the new and old world tropical areas (Arrhenius & Langenheim, 1983), they should harbor adaptations (e.g., increased protein or membrane

stability) that allow them to carry out biological functions under increased temperatures and humid climates. Further research is needed to determine how such adaptations function at the cellular and subcellular level in tropical plants, particularly *C. officinalis*.

Macromolecular Responses to Heat

In most plants, macromolecules have been suggested to be the primary sensors of elevated temperatures (Ruelland & Zachowski, 2010). For example, plant macromolecules such as DNA undergo three-dimensional conformational changes during temperature fluctuations to allow access for transcription factors (Ruelland & Zachowski, 2010), which then help initiate binding of DNA polymerase for gene expression. Membranes are also subject to increases in fluidity and the associated proteins often lose their conformational shape upon exposure to heat stress.

Membranes become destabilized in response to increasing temperatures due to the physical properties of the lipid bilayer. In heat stressed plants, the lipid bilayer is composed of more saturated fatty acids (FA) compared to non-heat stressed plants (Murata & Los, 1997). The increase in saturation decreases movement of the FA tails. Temperature impacts molecular movement; therefore, a higher temperature will increase movement of the fatty acids. To compensate, plants can manufacture and incorporate saturated fatty acids in the lipid bilayer, which decreases movement and ion leakage. Without this compensation, high temperatures can lead to increased membrane fluidity and perhaps ion leakage (Wang, Vinocur, & Altman, 2003). This in turn disrupts normal cellular processes, resulting in a stressful state for the plant.

In a study presented by Saidi et al. (2009), it was suggested that membrane proteins, perhaps those involved in calcium signaling, are the first responders to elevated

temperatures. The calcium signaling proteins in the lipid bilayer are responsible for an array of biochemical functions. Increasing temperatures change the three-dimensional structure of membrane proteins, thereby allowing influx of calcium ions. The resulting flow of ions transduces a signal by activating kinases involved in the heat stress signal. Large amounts of phospholipid (in the plasma membrane, endoplasmic reticulum, Golgi, and nuclear membrane) are found in developing seedlings. For this reason, young seedlings are especially prone to heat stress. Tolerance to elevated temperature is often assayed by germinating seedlings in the dark, heat stressing them, and then measuring hypocotyl elongation after a period of a few days. Interestingly, Wang et al. (2010) overexpressed a fatty acid desaturase (*FAD2*) in *A. thaliana* and presented evidence that, after a heat treatment, hypocotyl elongation was not inhibited. These data suggest that fatty acid desaturases, which aid in the formation of double bonds between carbons, play a role in plant heat tolerance.

Heat stress also influences the permeability of the lipid bilayer of the chloroplast. Integral membrane proteins and transport proteins embedded in the bilayer of the chloroplast are freer to move with increased temperature. However, there are many important proteins that play a role in photosynthesis, which quickly lose their ability to function when placed under high temperature conditions. One of the major protein complexes involved in photosynthesis is ribulose (ribulose-1,5-bisphosphate carboxylase oxygenase)--the main enzyme involved in carboxylation (or oxidation) of ribulose-1,5-bisphosphate. However, rubisco activase, which functions to remove phosphorylated sugars from rubisco's active site, is more sensitive to high temperatures compared to rubisco and other enzymes involved in photosynthesis. In a study by Salvucci and

Crafts-Brandner (2004), it was shown that the upper temperature limit for rubisco activase was 35˚C. For this reason, expression levels of rubisco activase and its correlation to grain productivity have been investigated. Ristic et al. (2009) found a decrease in rubisco activase when wheat was grown at elevated temperatures. Additional enzymes and molecular aspects of the photosynthetic machinery and their role in heat tolerance have also been studied in detail. For example, while photosystem II is inhibited by elevated temperatures, it is the D1 protein portion of photosystem II that is damaged and continually repaired (Allakhverdiev et al., 2008).

In the daily metabolism of a plant, there is a balance of carbon fixation and respiration. When high temperatures inhibit photosynthetic machinery, the plant uses more stored carbohydrates to compensate for loss of energy production. When equal amounts of CO_2 are being fixed and given off during respiration, it is known as the temperature compensation point (Taiz & Zeiger, 2010). Temperatures above the temperature compensation point lead to respiration outpacing photosynthetic CO_2 fixation. This metabolic process can be difficult to interpret in some plants. For example, a plant might have some leaves in full sun while other leaves might receive less direct sunlight, which can impart temperature stress to a localized region of the plant. Therefore, some parts of a plant can experience heat stress while other parts do not, leading to photosynthetic repair mechanisms in localized regions of a plant.

Heat Shock Proteins

Perhaps the most well studied molecular aspect of the plant heat response is heat shock proteins (HSPs). HSPs were first detected in *Drosophila melanogaster* (Ritossa, 1962, 1996) and have since been found in a diversity of organisms (Lindquist & Craig,

1988). Plant HSPs have been separated into six main groups based on their molecular weight (in kilo Daltons): HSP100, HSP90, HSP70, HSP60, HSP40, and small HSPs (sHSPs), which are 9-30 kDa. Some suggest that ubiquitin (8.5 kDa) should be classified as a separate HSP class (Vierling, 1991). Each class of HSPs functions in different ways within the cell to decrease the damaging effects of abiotic stress. Therefore, the type of HSP produced depends on the type and degree of stress. Although HSPs are produced without abiotic stimulus and are produced in normal physiological conditions, they are often associated with a response to abiotic stress.

Based on protein size the HSP100s are the largest of the HSPs. They are responsible for disaggregating or unfolding proteins and can be assisted by two cochaperones: HSP70 and HSP40 (Shaupp, Marcinowski, Grimminger, Bösl, & Walter, 2007). Even without these two cochaperones, HSP104 has been suggested to have properties that allow it to unfold misfolded proteins (Shaupp et al., 2007). There are two subclasses of HSP100s: subclass I and subclass II. Subclass I has been detected in the cytoplasm while subclass II members are localized within the chloroplast. Hong and Vierling (2000) investigated mutants of *A. thaliana* defective in HSP101 and showed that the effects of heat stress (45˚C) were lethal. They also generated *A. thaliana* over expressing HSP101, which displayed increased levels of thermotolerance. In another set of experiments by the same authors, it was shown that HSP101 displayed reduced expression in seeds and was not required for germination of *A. thaliana* seeds or development (Hong & Vierling, 2001).

The next class, HSP90, consists of proteins that have been detected in the chloroplast, cytosol, mitochondria, and endoplasmic reticulum (Krishna & Gloor, 2001).

HSP90s are associated with assisting nascent proteins to fold into their functional state (Pearl & Prodromou, 2006). The C-terminal domain functions in forming the dimer quaternary structure, which is required for proper refolding of client proteins. HSP90s are similar in their ATP-dependent cycles of client folding in that they bind ATP in the N-terminal domain while their central domain interacts with client proteins (Hessling, Richter, & Buchner, 2009). However, interactions between HSP90 and misfolded proteins are distinct for each client.

Members of HSP70 are diverse in function; however, nearly all members have an ATPase domain and a peptide-binding domain (Sung, Kaplan, & Guy, 2001). HSP70s have been found to play a role in prevention of protein aggregation, transducing signals, activating transcription, and facilitating protein transport. Their diversity of function is reflected by their diversity in location; they are found in the chloroplast, mitochondria, cytosol, and the endoplasmic reticulum. Diversity and location is important since developmental pathways of the chloroplast (Latijnhouwers, Xu, & Moller, 2010) and the plant itself (Su & Li, 2010) rely on HSP70s.

HSPs belonging to the HSP60 class have subclass I members that are restricted to the chloroplast and mitochondria while the subclass II members have been detected primarily in the cytosol. They play a major role in folding of denatured proteins and preventing disaggregation of important enzymes (Martin, Horwich, & Hartl, 1992). In a study by Salvucci (2008), modified forms of rubisco activase and affinity chromatography were used to determine that a chloroplast HSP60 binds to rubisco activase after heat stress. The author hypothesized that HSP60 might somehow protect rubisco activase from temperature stress by preventing its denaturation.

Perhaps the most diverse of the HSPs are the sHSPs. Members of this group of HSPs can form molecular complexes that can play a variety of roles within the cell. There are six subfamilies of sHSPs, each being found in distinct locations as follows: CsHSP I, cytosol; sHSP II, cytosol; sHSP III, chloroplast; sHSP IV, endoplasmic reticulum; sHSP V, mitochondrial; and sHSP VI, membranes. The sHSPs are thought to form complexes of 12 (dodecamers) or higher oligomers, which disassociate under stressful conditions (van Montfort, Basha, Friedrich, Slingsby, & Vierling, 2001). The monomeric forms of this group of HSPs share a common motif called the α-crystallin domain. The α-crystallin domain is important in keeping proteins from aggregating but has also been intensely studied because of its role in lowering the refractive index in the eye lens and its importance elsewhere in the mammalian body (Kantorow & Piatigorsky, 1998). Phylogenetic analyses of this group of HSPs suggest a diverse set of evolutionary patterns with shared conserved domains and differ in their subcellular locations (Siddique, Gernhard, von Koskull-Döring, Vierling, & Scharf, 2008; Waters, 1995). The importance of plant sHSPs to withstanding cellular stress is evident since the evolution of cytosolic sHSPs has been suggested to predate higher plants (Waters & Vierling, 1999). Ubiquitin and its associated proteins are sometimes classified with the sHSPs (Schrader, Harstad, & Matouschek, 2009).

HSPs are not limited to plants; some genes encoding HSPs in other organisms have been transformed into plants to determine their physiological roles. For example, *Trichoderma harzianum* is a fungus that produces several different HSPs. To explore the effects of heterologous expression, Montero-Barrientos et al. (2010) generated transgenic *A. thaliana* plants over-expressing *T. harzianum* HSP70. Homozygous transgenic seeds

and seedlings were able to tolerate suboptimal abiotic conditions when grown on MS media impregnated with 150 mM NaCl, 350 mM mannitol, or 3 mM H_2O_2 because HSP70s were hypothesized to protect against such abiotic stresses. Thermotolerance assays showed a normal germination and hypocotyl elongation in transgenic seeds and seedlings, respectively. In contrast, non-transgenic lines showed a shortened hypocotyl length and poor germination rates. Heat-treated adult plants showed greater thermotolerance and recovery after treatment than non-transgenic plants. Furthermore, quantitative real time PCR (qRT-PCR) of transgenic plants showed a down regulation of other stress-related *hsp* genes (*hsp101*, *hsp90*, and *hsp18*) and a heat shock transcription factor. qRT-PCR of ascorbate peroxidase (*APX1*) and superoxide dismutase (*SOS1*) showed a lower expression under non-heat stress conditions but was upregulated after heat treatment (Montero-Barrientos et al., 2010). These data suggest crosstalk between salt stress, oxidative stress, and heat stress response in plants. Overall, the results suggest that *HSP70* could be used to help confer heat and other abiotic stress tolerance in plants.

Interestingly, proteins associated with heat stress have also been shown to be necessary in plant development. A class of chaperones and membrane stabilizing proteins known as the late embryogenesis abundant (LEA) proteins were detected in developing leaves of the resurrection plant (*Boea hygrometrica*). The resurrection plant is able to tolerate extreme dehydration and other abiotic stresses. It has been determined that genes known as late embryogenesis abundant (*LEA*) genes help confer this stress tolerance by producing heat-shock-like proteins known as LEA proteins. These proteins help to stabilize the cell membrane and also prevent cellular proteins from aggregating once they have undergone denaturation.

Liu et al. (2009) provided greater insight into the LEA proteins and their role in heat stress tolerance by generating transgenic tobacco plants that overexpressed *LEA* genes. When compared to non-transgenic lines, the transgenic plants were better able to tolerate heat stress. In addition, multiple abiotic stress assays were performed on transgenic lines of tobacco. Chlorophyll fluorescence assays showed that photosystem II, which is usually the most sensitive molecular machinery to abiotic stress, showed a higher activity in transgenic plants than in non-transgenic lines. Similarly, cell membrane stabilization was measured using an electrolyte leakage assay that showed higher membrane integrity in transgenic plants. Overall, the transgenic plants were better able to cope with drought stress and recovery than non-transgenic plants. A key insight into this work was that although under constitutive expression, the levels of the LEA proteins increased as drought progressed.

Detecting Heat Stress

Plants often respond to heat stress and other environmental cues that cause fluctuation of intracellular calcium ion (Ca^{2+}) concentrations, typically 50-100 nM. Subsequently, calcium that enters the cell can do so by gated channels, which are often gated by other signaling molecules (e.g., calmodulin or particular nucleotides; Bowler & Fluhr, 2000). Upon exposure to heat stress, tobacco plants showed an increase in intracellular Ca^{2+} levels, perhaps from intracellular and extracellular regions (Gong, van der Luit, Knight, & Trewavas, 1998). Calcium binds calmodulin or activates kinases, which can further transduce the heat stress signal. In this way, the increase of cytosolic Ca^{2+} can induce a signaling cascade often with cross-talk between stress response pathways (i.e., response to cold can be similar to drought response; Knight & Knight,

2001). In an extensive review, Hey, Byrne, and Halford (2010) emphasized that stresses plants encounter have detection pathways that overlap with basic metabolic processes.

Although the mechanism by which Ca^{2+} influences the heat stress response has until recently been poorly understood, calmodulin-binding protein kinases (CBKs) are thought to play an important role. It has recently been suggested that a possible mechanism for expression of heat stress-associated genes (e.g., HSPs) relies on Ca^{2+} binding to calmodulin. This in turn can activate CBK that phosphorylate HSFs--the inducers of *HSP* gene expression. To facilitate a response to heat, cis-elements along the chromosome allow binding for phosphorylated heat shock factors (HSFs), which help regulate *HSP* expression. Schöffl, Prändl, and Reindl (1998) describe the HSFs present in the *A. thaliana* cytosol as being in a monomeric state. Upon phosphorylation, the monomer forms a trimer and localizes to the nucleus. Once in the nucleus, HSFs can bind heat shock elements, the cis-element recognition sequence, which is located on the promoter of many *HSPs* (Sakurai & Enoki, 2010). In *A. thaliana*, HSFs contain a helix-turn-helix functional domain that functions as the DNA binding domain when in the trimeric state (van Montfort et al., 2001). Such a response to heat stress has been shown to occur rapidly; in some cases, expression of HSP genes has been detected within minutes after exposure to heat (Åkerfelt, Morimoto, & Sistonen, 2010).

Importance of Studying Heat Stress in Plants

Atmospheric researchers agree that climate change is expected to result in an increase in global temperatures over the next century; therefore, it is important to characterize heat stress-related genes. One aspect of the change in global climate includes increased temperatures in many agricultural areas of the world. Therefore,

increased temperatures will require increased heat tolerance by agricultural crops. A majority of plant heat stress responses involve up-regulation of heat stress-related genes. To help crop plants cope with climate change and increased temperatures, heat shock proteins could be over-expressed. Under a constitutive promoter, HSPs might facilitate protein disaggregation, allow stabilization of membranes, and facilitate normal plant development and growth, even in the face of heat stress.

Tropical plants could be a rich source to understand heat stress-related processes. Limited data on heat-induced gene expression in tropical plants currently exist. Much of the genetic research on tropical plant gene expression has focused on tropical food crop species (Paterson, Felker, Hubbell, & Ming, 2008). A knowledge gap exists between tropical plants and their molecular mechanisms of mitigating elevated temperatures. In addition to their potentially unique tropical physiology, the importance of tropical plants in society is difficult to overstate. Many have been used for agricultural purposes such as banana (*Musa acuminata*), papaya (*Carica papaya*), chocolate (*Theobroma cacao*), and sugar cane (*Saccharum* spp.). In addition, important compounds have been discovered in tropical plant species such as periwinkle (*Catharanthus roseus*), coffee (*Coffea arabica*), and cocaine (*Erythroxylum coca*). Currently, there are also many examples of commercial and industrial products that were either originally discovered in or are presently derived from tropical plants. This is in part due to the great diversity of plant species located in the diverse ecological niches in tropical environments (Mittermeier, Myers, Thomsen, Da Fonseca, & Olivieri, 1998). The diversity provides genetic prospectors with a rich source of novel genes that might be involved in interesting and potentially beneficial biochemical pathways such as the heat stress response.

Copaifera Species

Tropical plants, whether cultivated for food or other purposes, must carry out essential physiological functions in the face of high humidity and low air circulation, both of which restrict evapotranspiration and thus limit cooling of organs such as leaves. One tropical plant that has stimulated interest among researchers is the diesel tree (*C. officinalis*, Fabaceae), which is known primarily for its characteristic oleoresin (Plowden, 2003). Plant secondary metabolites (e.g., terpenes) are thought to be a major component of the oleoresin (Chen et al., 2009). Although originally described for its medicinal benefits, the oleoresin has stimulated interest in the biofuel sector. However, demand for oleoresin would outpace production; thus, it would not be sustainable to continually harvest the oleoresin directly (Plowden, 2003). In this respect, a two-fold purpose for construction and analysis of a *C. officinalis* expressed sequence tag (EST) library becomes apparent. First, EST sequences might shed light on the biodiesel-like pathways involved in oleoresin production. Second, sequence data might also reveal biochemical pathways and/or cellular mechanisms involved in the heat stress response.

Generating Sequence Data

A central goal of this project was to identify genes expressed under heat stress conditions. Since limited nucleotide sequence data currently exist for heat stress-associated genes for *C. officinalis*, generating sequence data was an important first step in attempting to identify genes of interest. This dissertation also sought to insert and express these genes in the model plant species, *A. thaliana*. For these reasons, identification and subsequent transformation of *A. thaliana* with heat stress-related genes

was partially based on analysis of a small EST library and relied on similarity matches to characterized sequences in the NCBI database using BLASTx.

While complimentary DNA, or cDNA (in the context of a cDNA library), refers to a complete gene sequence, ESTs represent a partial sequence of a gene (Hatey, Tosser-Klopp, Clouscard-Martinato, Mulsant, & Gasser, 1998). Although ESTs might not be the complete coding sequence for a gene, they can be of sufficient length for comparison to sequences of proteins whose function has been determined. Generating ESTs from an organism requires obtaining the mRNA, reverse transcribing the mRNA into complimentary DNA (cDNA), and ligating the cDNA sequences into a vector. Primer sites with the vector that flank the cDNA can then be used to sequence cDNA inserts. An EST represents a portion of this cDNA. Similarity to other known sequences is determined through analysis using the basic local alignment search tool (BLAST).

In the case of constructing an EST library, total RNA is usually first obtained. Important to this step is that *C. officinalis* seeds can be germinated and grown in a greenhouse. Once grown, total RNA can be extracted and cDNA can be produced as described above. The cDNA ligated into the vector can then be transformed into *E. coli*. Transformed *E. coli* cells that harbor the vector-insert constructs are grown as colonies on selective media and the cDNA can be amplified and sequenced.

After sequencing, translation of nucleotide sequence data in all six reading frames allows comparisons to known protein sequences in a public database (Altschul, Gish, Miller, Myers, & Lipman, 1990). The resulting information provides insight into the transcriptional state by identifying proteins that might contribute to heat stress physiology.

For this project, cDNA was ligated into a plant expression vector. Sequencing relied on selecting *E. coli* from selective media and growing them in 96-well plates. Partial sequencing of each individual cDNA insert was performed directly from each well of the 96-well plate using an ABI3730xl sequencing machine. This was important since the 96-well plate harbors each vector-insert construct and can be used for downstream applications such as plant transformation studies. Using this approach provided advantages such as developing numerous transgenic lines of *A. thaliana*, each expressing a unique *C. officinalis* gene. However, this method required screening of each candidate construct to verify the complete coding sequence was present within the vector. If the insert contained the entire coding region, the vector-insert could be electroporated into *A. tumefaciens*. Finally, with the help of *A. tumefaciens*, the *C. officinalis* gene could be stably integrated into *A. thaliana.*

A central goal of the plant transformation portion of the study was to show that integration of the transgene occurred and determine if the gene was indeed expressed. Importantly, to determine if the gene was advantageous to conferring heat tolerance, a series of heat stress assays was performed. Expression of 10 independent transgenes and the ability of each to confer tolerance in *A. thaliana* were investigated.

This study had both agricultural and important biotechnological themes. Although HSPs are required for normal development, they have been shown to accumulate in response to stress. This is important since some have argued that heat stress is a major hindrance for food production, especially in cereal crops (Maestri et al., 2002). Therefore, this dissertation attempted to contribute toward understanding whether or not thermotolerance was confered by overexpressing *C. officinalis* genes in *A.*

thaliana. Subsequently, overexpression of these genes in crop plants might be useful in helping them cope with heat stress. With the findings from this research, it was hoped that the increase in global temperatures, much of which has occurred over the last 30 years (Walther et al., 2002), and its effects on crop plants might be at least partially mitigated using heat stress-associated genes derived from tropical plants.

Summary of Dissertation

The first goal of this dissertation, which is described in Chapter II, was to generate an EST library of *C. officinalis*. The library was constructed using a plant expression vector; sequence data generated from the library were analyzed using the Blast2GO software suite. The resulting BLASTx hits were searched for high sequence similarity to heat stress-associated genes. Since the heat stress-associated genes of interest had identifiers based on location of their well positions, the sequenced 96-well plates were obtained.

The second goal, which is described in Chapter III, was to transform *A. thaliana* with *C. officinalis* heat stress-related genes identified in the EST library. In this portion, the *E. coli* harboring the vector-insert construct was picked from a specific well of the 96-well plate and grown overnight. A plasmid miniprep was performed; the insert from the plasmid was amplified using the polymerase chain reaction (PCR) and using primers flanking the cDNA insert. Inserts that showed similarity in length to known coding sequences were used in transformation studies. The construct from the plasmid miniprep was then electroporated into *A. tumefaciens* EHA105. Flowering *A. thaliana* plants were transformed using *A. tumefaciens*, generating a total of 10 transgenic *A. thaliana* plant

lines each incorporating an independent gene. After the plants matured, seeds were collected and screened on selective media.

The third and final goal, which is described in Chapter IV, entailed using homozygous lines for three different heat stress assays. These included an *in vitro* assay, a hypocotyl elongation assay, and an *in vivo* assay. In each assay, transgenic plants were expected to show less heat damage compared to control (nontransgenic plants).

The final chapter summarizes the overall findings resulting from the research, describes the major conclusions from the research, and provides suggestions for future research. Other potential genes generated from the EST library that might be good candidates for additional transformation studies are also discussed. The final chapter also summarizes why identifying heat stress-related genes derived from tropical plants and expressing these in crop plants should be given more attention by plant biotechnologists.

CHAPTER II

GENERATION AND ANALYSIS OF A *COPAIFERA* EXPRESSED SEQUENCE TAG LIBRARY

Introduction

Unique insights into physiological processes can be gained by sequencing portions of mRNA transcripts, which are known as expressed sequence tags (ESTs) (Boguski, Lowe, & Tolstoshev, 1993). Since little sequencing work has been done for *Copaifera officinalis*, potential discoveries might lead to novel insights into its metabolic and biosynthetic pathways. The present chapter seeks to illustrate the importance of generating and analyzing ESTs from *C. officinalis*, interpret their putative role and significance, and discuss potential future research for interesting sequences.

Copaifera officinalis

Copaifera officinalis is a tropical plant (also known as the “diesel tree”) previously noted for production of diesel-like oleoresin. The volatile component of the oleoresin largely consists of sesquiterpenes and the resinous component consists of diterpenes (Plowden, 2003). Multiple differences exist in terpene profiles between seedlings and young trees. In a study by Chen et al. (2009), terpenes were qualitatively and quantitatively compared in *C. officinalis* stems, leaves, and roots. Interestingly, the roots of two-year old trees lacked many of the terpenes found in the stems and leaves. Importantly, the biochemical constituents of the two-year old trees closely matched

oleoresin composition studied in mature trees. This suggests a young specimen can be investigated for metabolic pathways, which might contribute to the oleoresin production and biodiesel-like characteristics.

In addition to its oleoresin production, the diesel tree lives in tropical environments that are often subjected to elevated temperatures coupled with high humidity. Currently, sequencing data for tropical plants are largely restricted to important tropical food crop species such as papaya (Paterson et al., 2008). Generating sequence data from expressed sequences derived from *C. officinalis* might contribute toward an understanding of molecular processes that mitigate heat stress in tropical plants. Sequence data for *Copaifera* species is limited to conserved genes (e.g. *trnL*) that have been used in phylogenetic studies (Bruneau, Fores, Herendeen, & Klitgaard, 2001). Since whole genome sequencing can be costly and computationally intensive, development of a small EST library provides a novel opportunity to initially investigate *C. officinalis* genes.

Expressed Sequence Tag Libraries and Sequence Analysis

Advancements of molecular tools such as EST library construction provide an excellent opportunity for insight into molecular plant physiology. Other researchers have relied on small sets of sequences to gain insight into molecular processes related to plant stress. Machuka et al. (1999) sequenced 169 ESTs to help identify genes associated with responses to abscisic acid in the moss *Physcomitrella patens*. Identified sequences were similar to multiple stress-related proteins and included matches to heat shock proteins, cold and drought responsive pathways, and protein products known to be responsible for coping with oxidative stress. Wood, Duff, and Oliver (1999) constructed a small set of

ESTs for *Tortula ruralis*, also called star moss. From the 152 ESTs analyzed, 44 were similar to previously described genes and 11 (25%) of those played a role in responding to environmental stress.

The use of expression vectors for library construction (expression libraries) has been performed by many researchers over the last two decades. Classically, expression libraries often use a phage or bacterial expression system and are screened for a protein of interest using hybridization assays (Schäfer & Brown, 2009). However, reports using plant binary vectors in library construction and later obtaining the sequencing plates for picking clones are difficult to find. Such an approach might speed the process of generating a transgenic plant.

EST libraries can yield a copious amount of information. Therefore, analysis of sequence data often relies on the most current bioinformatics tools that facilitate sequence comparisons to public databases. Perhaps the most common public database is the NCBI database--GenBank. The Basic Local Alignment Search Tool (BLAST) is often used to search this database (Altschul et al., 1990). There are several variations on the BLAST algorithm for querying a sequence against GenBank: nBLAST requires a nucleotide sequence to search against other nucleotide sequences; BLASTp requires input of a protein sequence to search against other proteins; BLASTx requires input of a nucleotide sequence that is automatically translated in all six reading frames to search protein sequences. Since a translated nucleotide query gives more accurate results, the BLASTx algorithm is commonly used to find similar protein sequences to a translated nucleotide (e.g., EST) sequence. The NCBI database has been improved since its inception (Johnson et al., 2008; McGinnis & Madden, 2004; Ye, McGinnis, & Madden, 2006). In

addition, a low complexity filter can be used to remove regions in a query that are less likely to have distinct matches. Statistical significance of alignment to the BLAST database is important; an expect value (E-value) is provided with each sequence query to indicate the level of significance. A smaller E-value indicates a more likely match. A useful equation to determine the E-value is

$$\text{Evalue} = m \times n \times P$$

This equation means that an E-value is equal to the number of residues in the database (m) times the number of residues in the query sequence (n) times the P value derived from the likelihood of finding a match with a random sequence. This later value is determined from an extreme distribution curve generated by comparing a sequence against another sequence repetitively (>1000x) and reshuffling one of the sequences each time they are compared. Since a database changes over time (sequences are added), a determined E-value can increase, which might render a match unfound in subsequent searches.

Sequences that have been compared to the NCBI database using BLASTx can be mapped, which involves retrieving associated gene ontology (GO) terms according to the gene identifiers and accessions (Götz et al., 2008). All relevant GO terms are mapped to the sequence regardless of the method used originally (experimentally or computationally) to assign those terms. This results in a large pool of GO terms that might vary in their strength in association. For example, GO terms that were assigned to the accession of a particular protein in the NCBI database using experimental evidence might be more reliable then GO terms assigned based on a computer algorithm (Götz et al., 2008).

After mapping, sequences can be annotated by selecting only relevant GO terms. This is based on the evidence for each GO assignment in the mapping step. Each term is assigned an annotation score according to the annotation rule (Ashburner et al., 2000). The terms themselves are dictated by the Gene Ontology Consortium (Harris et al., 2004), which is a group working to assign a standard set of terms to sequence data. Therefore, gene ontology is a collaborative project and a set of controlled vocabulary. Each sequence can have multiple terms associated with it depending on its biological process, molecular function, or cellular location. Assignment can be automated; however, members of the Gene Ontology Consortium can also perform manual annotations. Such a consortium is needed to facilitate common understanding between disciplines (e.g., plant biologists and mammalogists) when discussing the biological function, molecular process, or cellular location of homologous proteins (Ashburner et al., 2000).

Presented here is an analysis of *C. officinalis* ESTs that might help in providing insight into the molecular physiology of this tropical plant. To better understand the proteins expressed, a software suite that integrated BLASTx and GO term annotation was used. Based on the BLASTx and GO terms, the *C. officinalis* EST library provides insight into developmental and heat stress-related processes.

Materials and Methods

Seeds from *C. officinalis* were obtained from the University of Puerto Rico, San Juan. These were germinated in approximately 0.5 cubic meter of Miracle Grow and 5-7-5 potting soil (Scotts, USA) in controlled conditions (30˚C, ~630 lux) at the University of Northern Colorado, Greeley, Colorado, USA. Leaf material was harvested from 18-

month-old trees, snap frozen, and ground in liquid nitrogen. Total RNA was extracted using Trizol Reagent (Invitrogen, USA) according to manufacturer's protocol and shipped on dry ice to Advanced Biotekservices (San Diego, USA) for cDNA library construction. PolyA mRNA was purified from the total RNA with magnetic oligo-dT beads. The mRNA was reverse transcribed to make cDNA using oligo-dT primers and dNTPs SuperScriptIII RT (Invitrogen, USA). The reverse transcriptase reaction proceeded for 1 hour at 50˚C. Second strand cDNA was synthesized using first strand cDNA, polymerase I, DNA ligase, RNase H, and dNTP mix. This reaction proceeded for 2 hours at 16˚C. The double-stranded cDNA was size selected (>500 bp) and directionally cloned into pCHF3, which is a reliable binary vector, using proprietary technology of Advanced BiotekServices. The vector-insert plasmids were electroporated into DH10B electrocompetent *E. coli*. Glycerol was added to give a final concentration of 20% and the cDNA library was stored at -80˚C.

Importantly, the cDNA was ligated into the plant expression vector pCHF3 (Borevitz, Xia, Blount, Dixon, & Lamb, 2000), which has the CaMV35S promoter and spectinomycin bacterial selectable marker (see Appendix A). Library complexity was estimated to be 2.1 x 10^6. Average insert size was estimated to be 1.6 kb. Single pass sequencing of the cDNA clones was performed with 96-well plates on an ABI 3730xl Genetic Analyzer by Lucigen Corporation (Middleton, Wisconsin, USA).

A total of 1008 sequences were analyzed with Sequencher software (Gene Codes Corporation, USA). The pCHF3 vector ends were removed from the sequence of interest by comparing partial flanking vector sequences with the trace files. To improve quality of sequences, end trimming was performed with removal of ambiguous bases at 3´ and 5´

ends. ESTs that are >200 bases in length can fail to find matches using BLASTx; therefore, these were removed from the data set. The remaining ESTs were assembled through 635,722 comparisons into 545 fragments and 68 contigs that yielded 613 unigenes.

The resulting sequences were translated in all six reading frames and compared against the BLASTx database. To facilitate batch handling of sequence data, mapping, and annotation, the Blast2GO (B2G, v.2.4.9) software suite (http://www.blast2go.org/) was implemented (Götz et al., 2008).

A cutoff E-value of 1.0E-5 was used to select for significant matches. Gene ontology information was obtained by comparing BLASTx accessions to the GO database. Annotation was carried out using default settings (E-value hit filter of 1.0E-6 and an annotation cutoff of 55 terms).

Results

A total of 716 ESTs was used to form 68 contigs and 545 singletons (single EST). The 68 contigs were each composed of two or more ESTs that included a total of 171 singletons. Most sequences had a length that ranged from 700-900 bases (see Figure 1). The E-value distribution for all sequences ranged from 1.5E-6 to 3.2E-165 (see Figure 2). Eighty-four percent of contigs showed positive hits with the NCBI database. Top hits (see Figure 3) for sequences closely matched *Vitis vinifera* (143), *Glycine max* (138), *Ricinus communis* (83), *Populus trichocarpa* (56), and *Medicago truncculata* (31). Based on all the BLASTx hits (at least 20 per sequence), the species that *C. officinalis* sequences matched included *Vitis vinifera* (1612), *Arabidopsis thaliana* (1164), *Populus trichocarpa* (1133), and *Glycine max* (633; see Figure 4). In general, shorter sequences

failed to match other sequences in the database. Overall redundancy of the library was estimated to be 24% and was calculated as number of clustered EST's/total ESTs x 100.

Although all sequences were compared against the BLASTx database, 89 sequences failed to present a positive match (see Figure 5). In addition, 31 sequences could not be mapped, 61 were not annotated, and 432 sequences were annotated (see Figure 5).

GO level distribution, which can be thought of as the depth of description of each GO term, varied across sequences (see Figure 6). Higher values of a GO level equate to a more descriptive function of each gene product. A total of 2046 annotations were made with a mean GO level of 5.6. The GO level 2 had 15 sequences with biological processes and 52 sequences with terms associated with molecular function. The GO level 7 had the most sequences (308) associated with cellular location while GO level 9 only had seven sequences associated with cellular location. GO level 6 had the most sequences (178) associated with biological processes. No sequences were annotated with more than 11 terms.

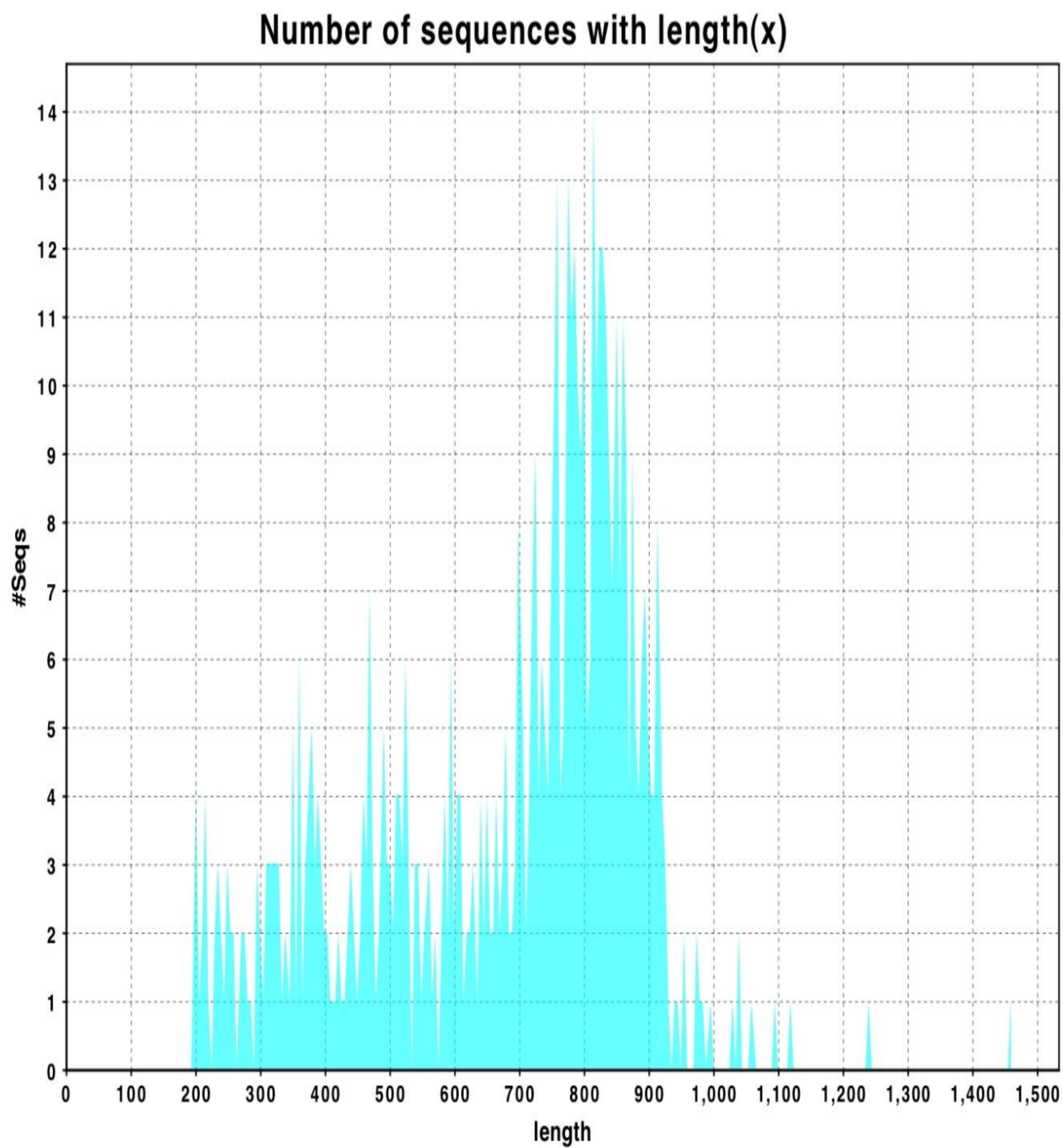


Figure 1. Comparison of the number of copaiba ESTs and their associated lengths. Overall sequence lengths varied but most ranged from 700-900bp.

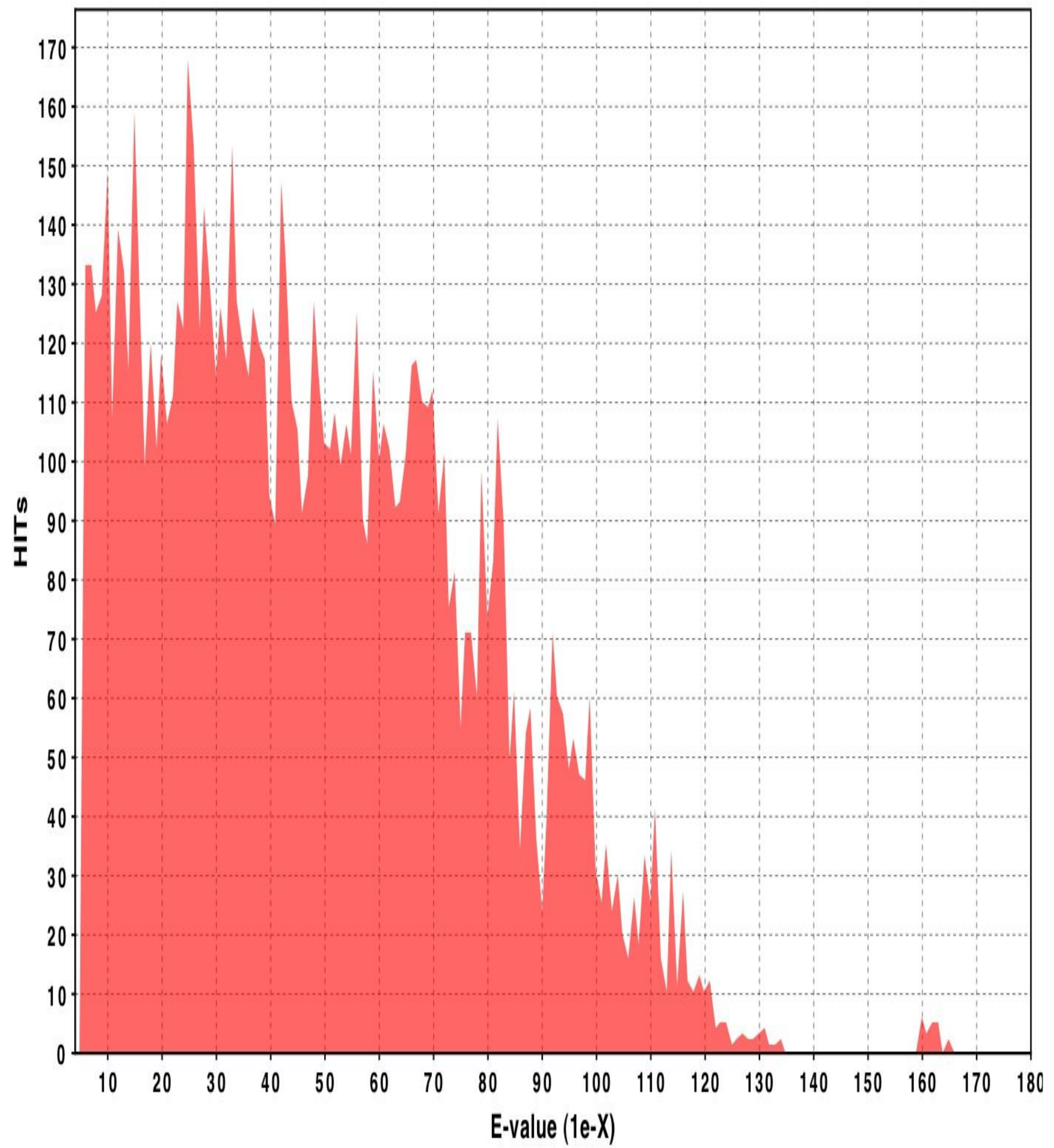


Figure 2. Distribution of E-values according to the number of hits from the BLASTx database. All but 89 sequences matched other sequences within the database.

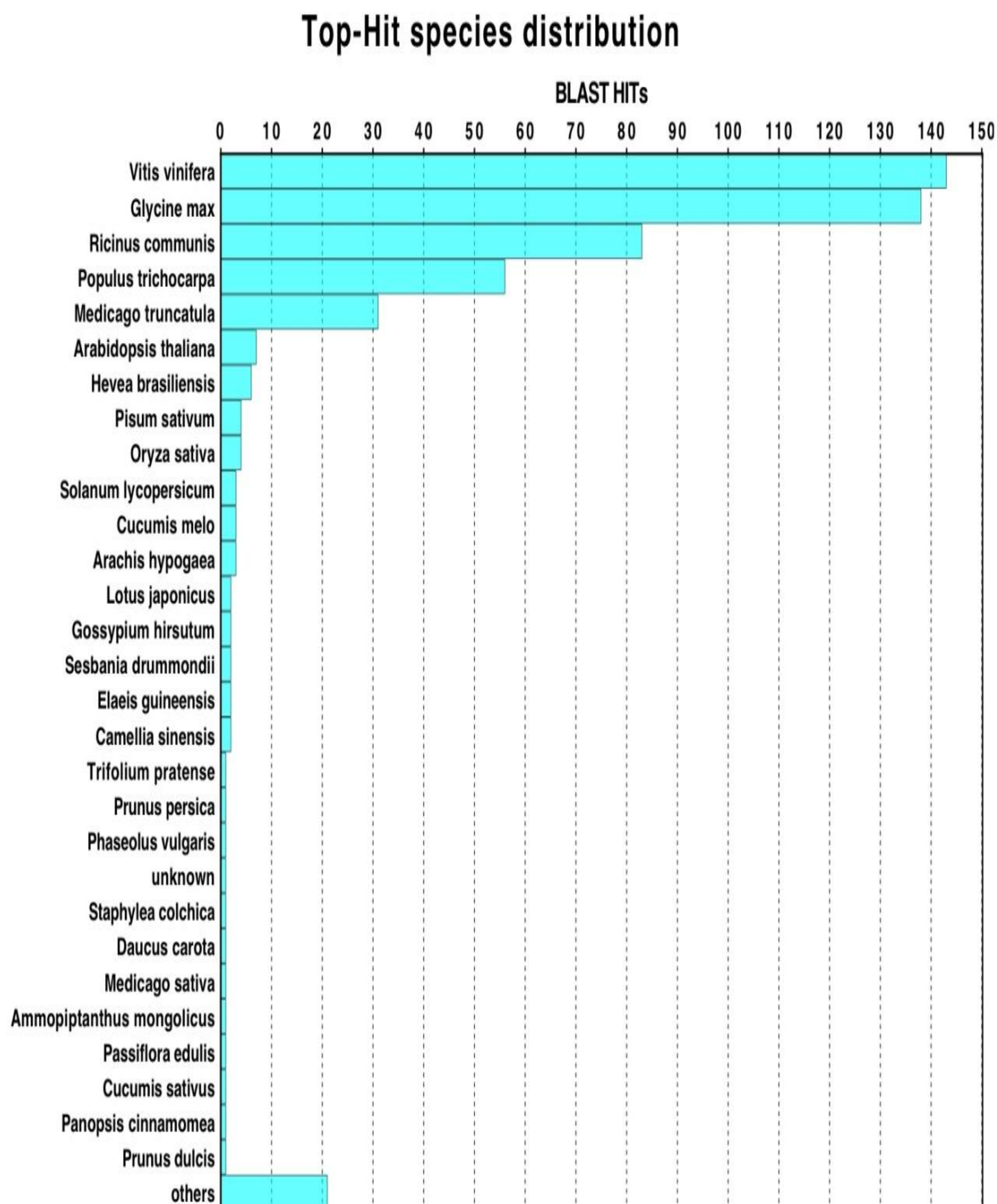


Figure 3. Distribution of BLASTx hits based on other species. *Arabidopsis thaliana*, a model organism, had 7 hits while *Hevea brasiliensis* (rubber tree) had 6. Most hits were associated with species that have all or most of their genome sequenced.

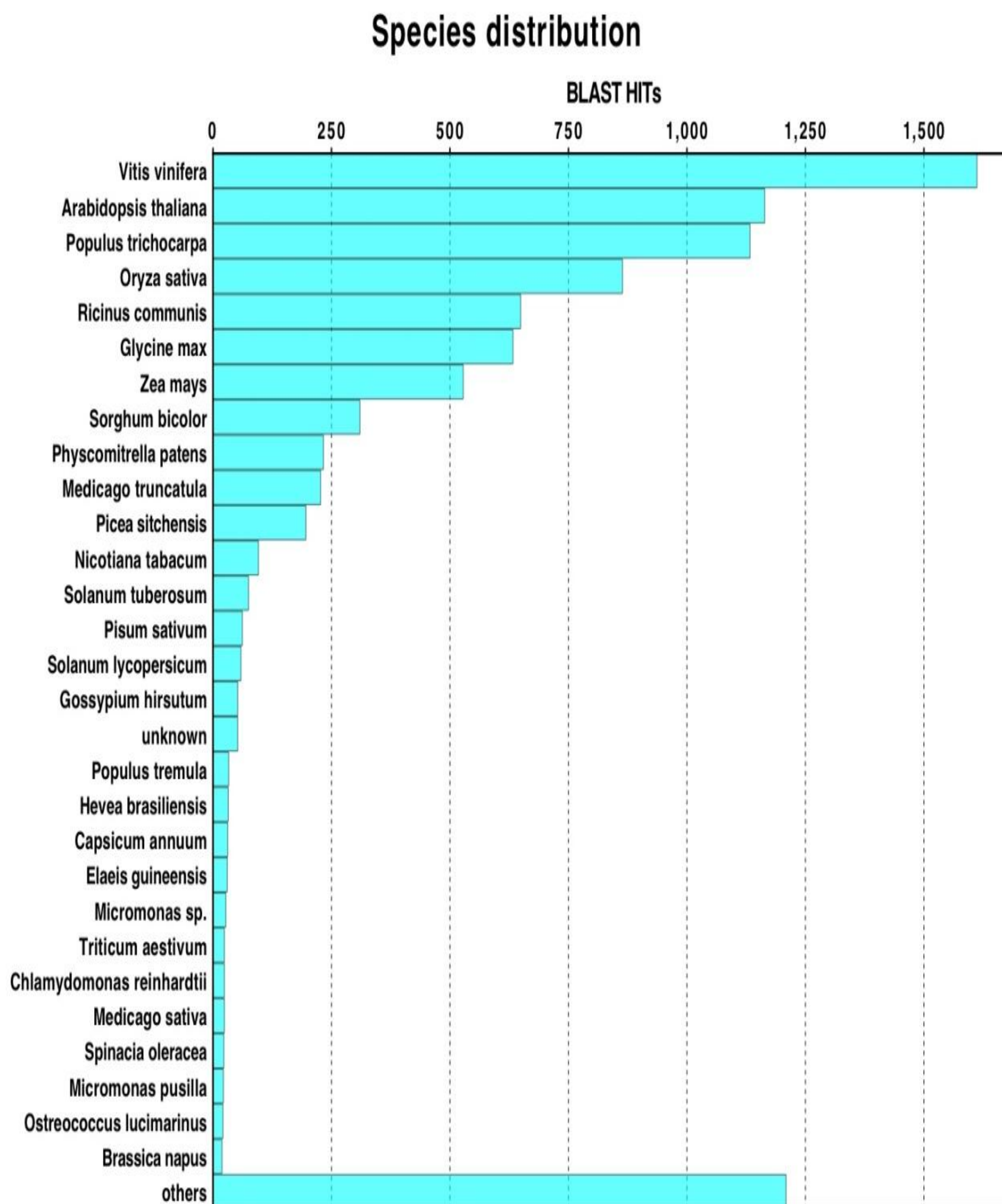


Figure 4. Species distribution for all BLASTx hits. A maximum of 20 hits were provided with each queried sequence. Species with a lot of sequence information were found to be similar such as *Vitis vinifera* (1612), *Arabidopsis thaliana* (1164), *Populus trichocarpa* (1133), and *Glycine max* (633).

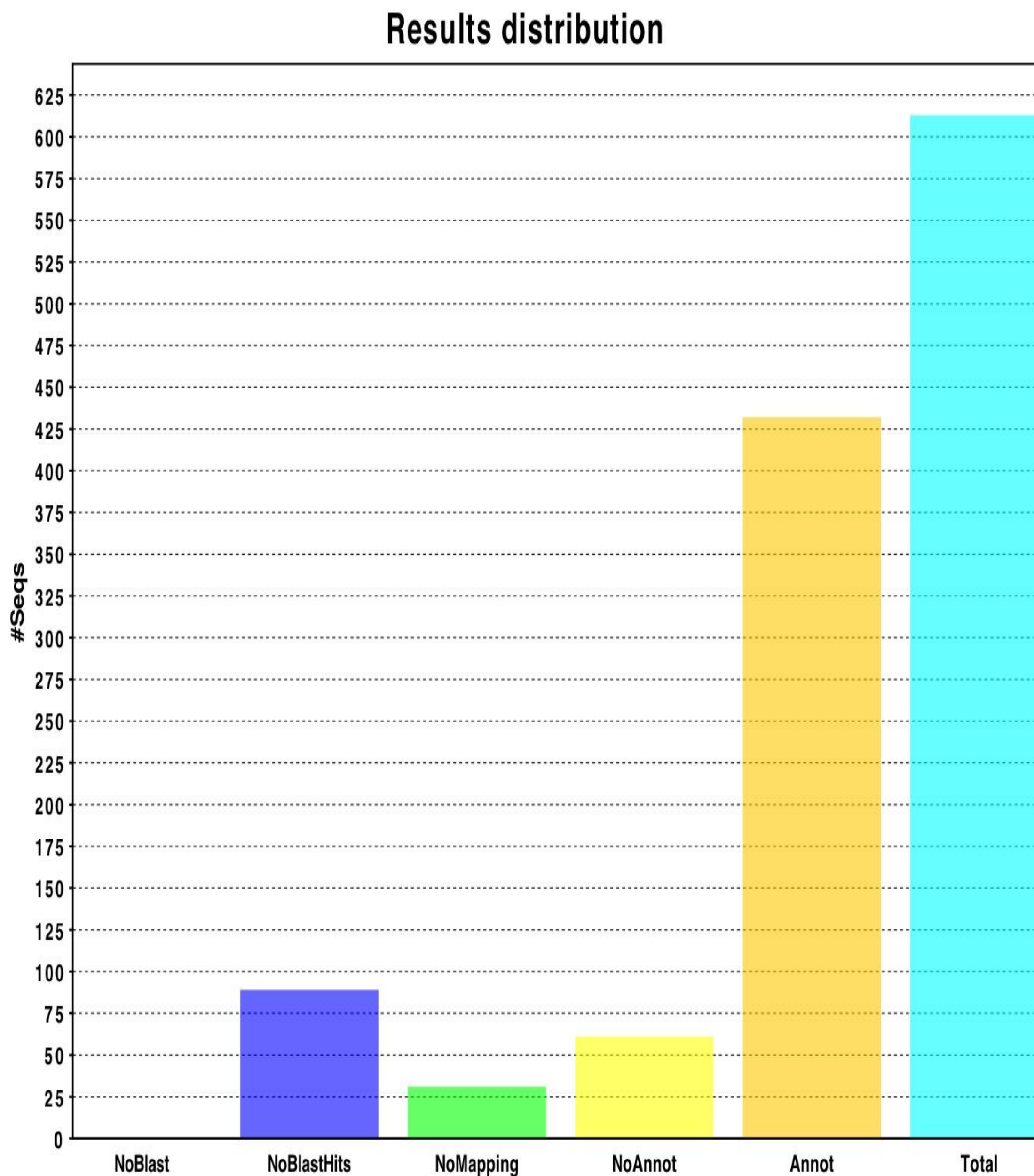


Figure 5. Distribution of sequences according to each step in library analysis. A total of 89 sequences could not be matched with BLASTx (NoBlastHits); 31 no mapping; 61 no annotation (NoAnnot); 432, annotations (Annot); 613 total).

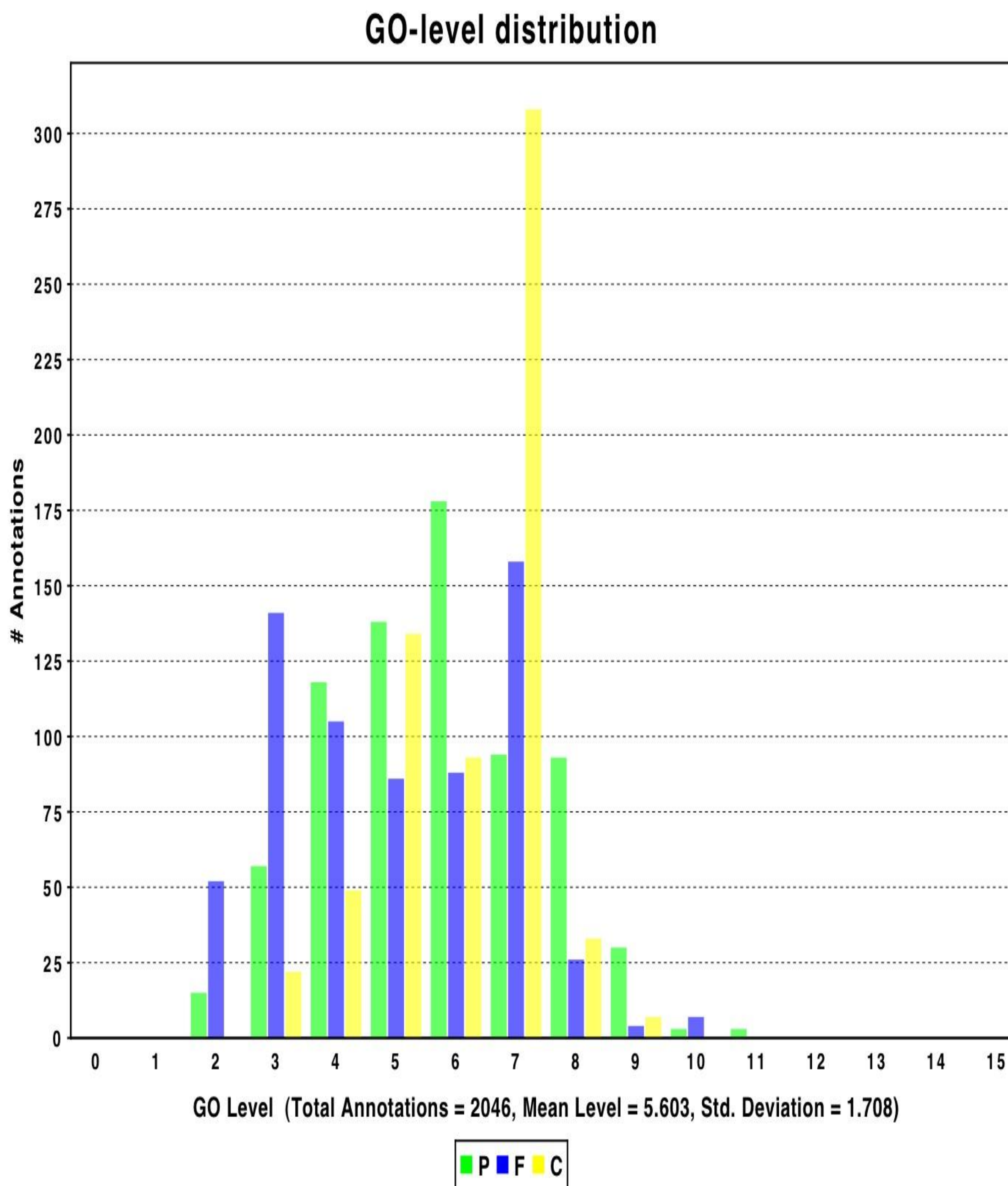


Figure 6. GO-level distribution for annotations of copaiba unigenes. A total of 2,046 annotations were given across all levels (P, biological process; F, molecular function; C, cellular component).

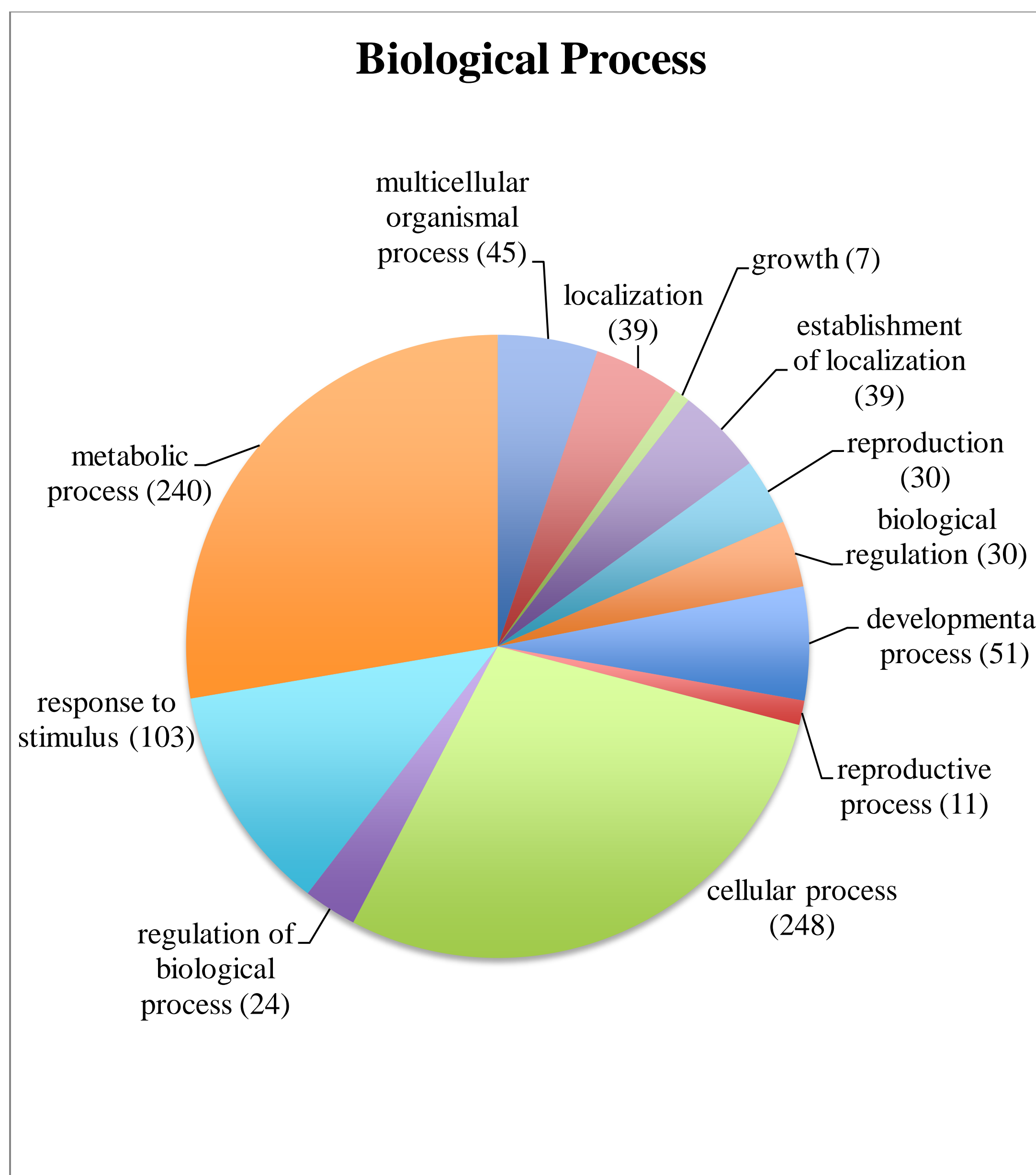


Figure 7. More than half (488) of the biological processes category consisted of annotations for cellular and metabolic processes. Growth (7) and reproductive process (11) were the two lowest occurring GO terms.

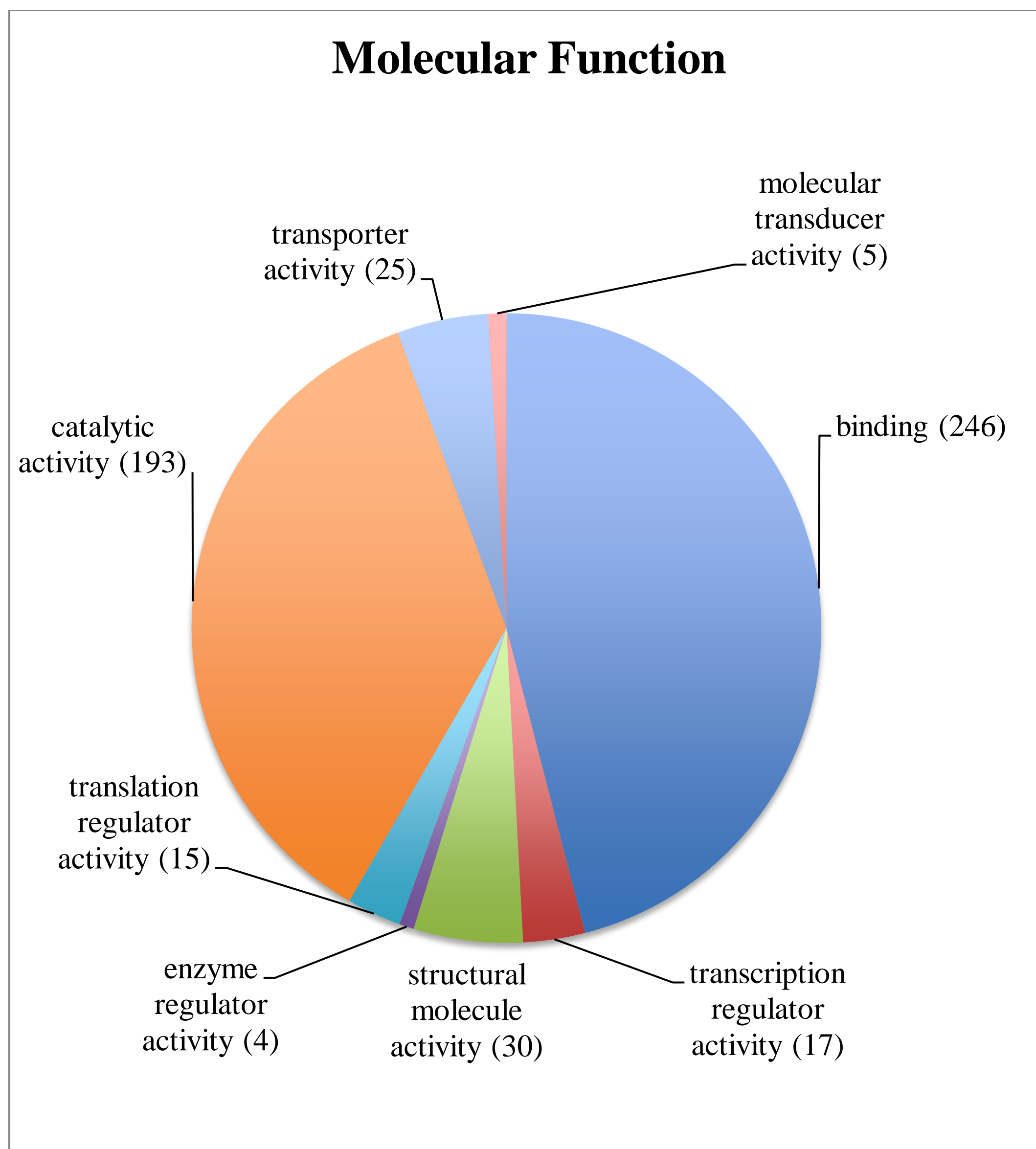


Figure 8. The molecular function category consisted of relatively few annotations for enzyme regulator activity (4) and molecular transducer activity (5). However, this category had nearly half (246) of GO terms associated with binding.

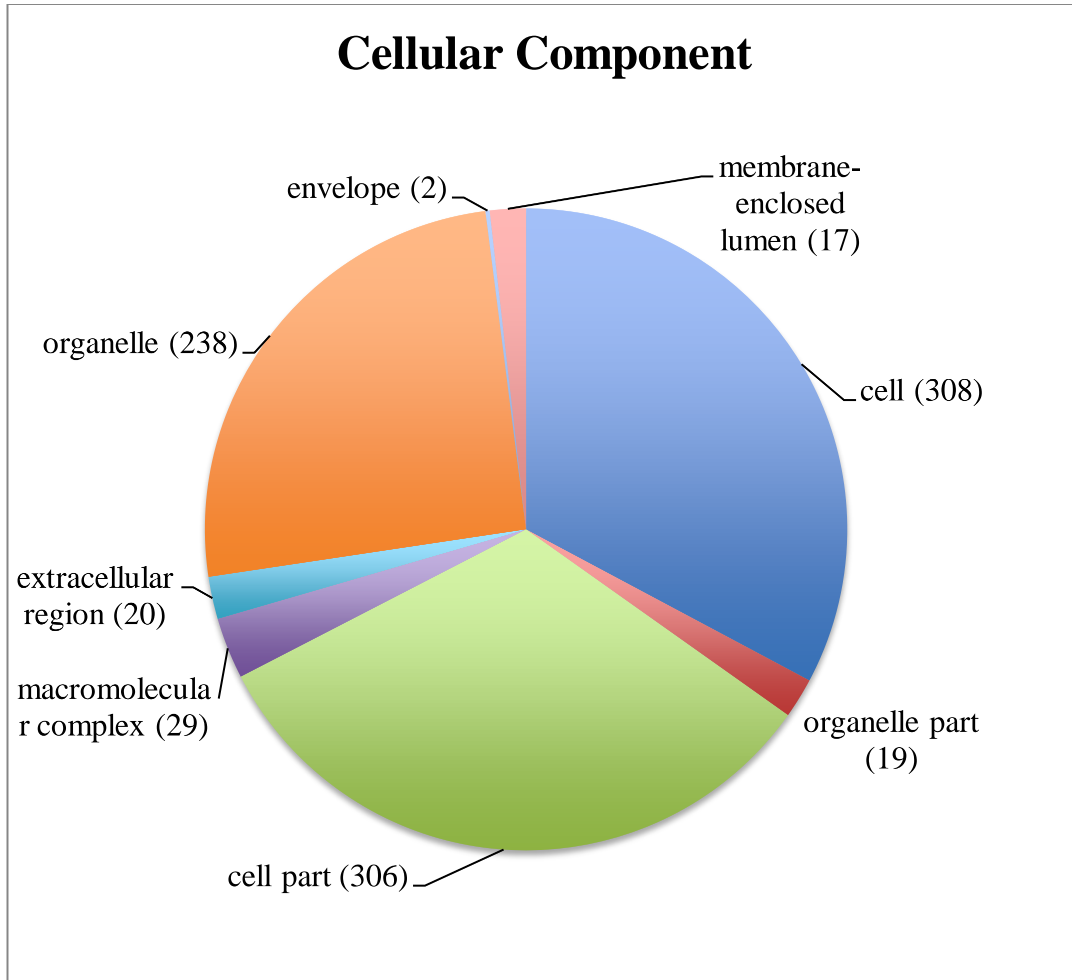


Figure 9. A large portion of unigenes categorized as cellular component were either cell (308), cell part (306). or organelle (238).

Contigs consisted of two ESTs (50), three ESTs (12), four ESTs (2), or five ESTs (2). The two largest contigs were composed of nine and eight ESTs, respectively. Sequences with annotations were assigned gene ontologies according to biological process (see Figure 7), molecular function (see Figure 8), and cellular component (see Figure 9). The GO level 2 was used in annotating our data. A total of 839 annotations were observed as a biological process. More than half (58%) of these were annotations for either a metabolic or cellular process. Response to stimulus represented 12% (103) of annotations while 6% (51) and 3% (24) of annotations were associated with developmental processes and regulating biological processes, respectively. Less than 1% (7) of annotations was for growth.

A total of 939 annotations were observed for cellular component. More than 90% (852) of these were associated with one of three components: organelle (238), cell (308), and cell part (306). About 2% (20) were located in an extracellular region and 3% (29) were associated with a macromolecular complex.

Sequences annotated as having molecular function consisted of eight categories with a total of 535 annotations. Nearly half (46%) of these were observed as having some type of binding (e.g., nucleotide binding) activity. A large portion (36%) of sequences was annotated as having catalytic activity. The remaining six annotation categories were for structural molecule activity (30), translation regulator activity (15), transporter activity (25), molecular transducer activity (5), transcription regulator activity (17), and enzyme regulator activity (4).

Based on the results for the Fisher's exact test, GO terms associated with heat stress were overrepresented in *C. officinalis* when compared to an *A. thaliana* reference

set (see Appendix C). The Fisher's exact test relied on comparing two sets of GO terms using a contingency table, similar to a Chi-square goodness of fit test (Blüthgen et al., 2008). However, the Fisher's test accounted for differences in the total number of sequences used in the two data sets and a probability value (p-value) was attributed to each term. A low *p*-value (< 0.05) indicated that the GO term was not occurring due to chance; therefore, genes with those GO terms were being upregulated. The GO term, response to stimulus, had a *p*-value of 6.89E-05 with 103 occurrences in *C. officinalis* and 3,734 in the *A. thaliana* reference set. Response to stress (*p*-value = 2.67E-09) occurred 80 times in *C. officinalis* and 2,086 times in the reference set; response to abiotic stimulus (*p*-value = 1.01E-09) occurred 57 times in *C. officinalis* and 1,217 times in *A. thaliana*.

Discussion

The results from this research suggested that under conditions of high humidity and high temperature, *C. officinalis* expressed a vast array of genes to mitigate detrimental heat stress-associated affects. Much of this evidence came from the GO terminology and BLASTx results used to help characterize each sequence. Each major category of the GO (biological process, cellular component, and molecular function) had associated parent-child terms that helped describe a translated EST. Using GO terms, strict vocabulary was used to associate protein names with more specificity (termed 'child') below a less specific category (termed 'parent'). This not only allowed for rigorous control of language but aided in relating each original nucleotide sequence to a GO match of a specific category. The three main annotation categories of cellular

component, biological process, and molecular function showed repeated annotations of the same sequence.

The EST library of *C. officinalis* provided limited yet interesting insight into its transcriptional and molecular (protein) composition. Contig0007 consisted of eight ESTs and closely matched (E-value = 3E-65) an early light inducible protein. GO terms associated with Contig0007 included its cellular location (plastidial membrane), molecular function (binding), and its biological process (response to stress and response to abiotic stimulus). These terms agreed with previous research that has shown early light inducible proteins to be associated with light stress (Adamska, Kloppstech, & Ohad, 1992).

Proteins that detect light are a basic requirement for plants. Hits associated with other fundamental metabolic activities were observed. For example, P7G08 closely related (E-value = 2E-16) with an enoyl-acyl carrier protein reductase, an enzyme necessary for *de novo* fatty acid synthesis (Slabas & Fawcett, 1992). Its associated GO terms were metabolic process and catalytic activity.

EST 6E01 was associated with GO terms relating to molecular function (catalytic activity and binding), biological process (response to stress, response to abiotic stimulus), and cellular location (peroxisome). Contig0052 (composed of two sequences) and contig 0077 (composed of two sequences) showed similarity to polyubiquitin. Ubiquitin is a small protein that is covalently attached to damaged or irreparable proteins to mark them for destruction (Dreher & Callis, 2007). Once a protein is denatured beyond repair, it might be more energy efficient to degrade the protein rather than repair the damaged protein (Dreher & Callis, 2007). Polyubiquitination targets the damaged protein to enter

a proteasome complex. Other ESTs were found closely related to a proteolytic pathway. For example, EST 2E01 closely matched a senescence-associated protein and 9B01 closely matched a ubiquitin conjugating enzyme.

A major finding of this work was that when grown at 30°C, ~630 lux and high humidity, *C. officinalis* expressed a multitude of heat stress-related genes. This was based on BLASTx hits and the fact that 103 of the biological process GO annotations were associated with response to stimuli (e.g., abiotic factors). This was surprising since most tropical plants are accustomed to living at these (or higher) temperatures.

A contributing factor of expression of heat stress-related genes might have been the humidity of the growth chamber. High humidity tends to decrease water transpiration (and, thus cooling) from stomata (Taiz & Zeiger, 2010). To compound matters, within the growth chamber, airflow was almost completely restricted, thus decreasing the plant's ability to thermoregulate via evaporative cooling. Multiple genes were associated with different abiotic stresses: heat, osmotic, salt, and oxidative stress. Of major interest in plant biotechnology is production of glutathione (May & Leaver, 1994; Yoshimura et al., 2004). One of the *C. officinalis* transcripts (6E01) closely matched a glutathione-dependent formaldehyde dehydrogenase, a protein shown to be important in using reduced glutathione to inactivate toxic species such as formaldehyde. Contig0020 was composed of two ESTs and closely matched (E-value = 1.1E-97) a glutathione s-transferase and had only one GO term (transferase activity). Glutathione has been shown to not only reduce oxidative damage but is also an important signaling molecule (May, Vernoux, Leaver, Montagu, & Inzé, 1998).

Sequence 2C03 closely matched (E-value = 6.3E-11) a VTC2-like protein and had associated GO molecular process of response to heat and response to jasmonic acid stimulus. VTC2 proteins are involved in ascorbate biosynthesis, which is important in deactivating reactive oxygen species (Müller-Moulé, 2008). Jasmonic acid, a relatively recently described plant hormone, is involved in abiotic and biotic stress signaling cascades (Schaller, Biesgen, Mussig, Altmann, & Weiler, 2002).

EST 8D11 closely matched (E-value = 7.5E-69) a plant SNARE (soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor). SNARE proteins are a large group of proteins involved in many functions including directing the transport of vesicles, signaling, and solute transport (Grefen & Blatt, 2008). SNAREs have been suggested to play a role in plant stress response. For example, guard cells must rapidly change in surface area to accommodate stomatal closing; SNAREs are thought to function by transporting cellular contents for secretion (Leyman, Geelen, Quintero, & Blatt, 1999).

Some ESTs were suggested to be involved in synthesis of compatible solutes. For example, EST 7A05 was similar (E-value = 4.7E-41) to allene oxide cyclase. This enzyme is important in synthesizing glycine betaine, an important compatible solute that has been studied extensively in transgenic crop plants. Alia, Hayashi, Sakamoto, and Murata (1998) overexpressed choline oxidase, a key enzyme in glycine betaine producing, in *A. thaliana*. They found that transgenic seeds were able to germinate at higher temperatures compared to controls. Compatible solutes (e.g., glycine betaine and proline) are synthesized in response to drought stress (Ashraf & Foolad, 2007). Upon loss of water from the plant cell (an increasing water potential) causes an increase in ion

concentration. This causes proteins to become destabilized due to the loss of the hydration shell, which is crucial to minimize interactions with polar solutes. Compatible solutes help counter the loss of the hydration shell and often contain charged groups, making them very soluble in water; however, their direct interaction with proteins is minimal. In the case of glycine betaine, three methyl groups are associated with the quaternary ammonium group and a negatively charged carboxyl group allows interactions with water and small solutes while minimizing protein disruptions (Sakamoto & Murata, 2002).

Many ESTs and contigs generated closely matched heat shock proteins (HSPs). These can be small (>30 kDa) proteins that act to prevent protein aggregation upon heat stress or in other cases HSPs can be much larger (>100 kDa). EST 2G07 closely matched (E-value = 4.5E-45) a chaperonin 60. This HSP60 group of chaperonins is able to actively assist denatured proteins to fold back into their native conformation. In addition, both EST 3B03 and EST 1H10 closely matched (E-values = 2.2E-65 and 7.5E-47, respectively) chaperones. Contig0005 (composed of three ESTs) was found to closely match (E-value = 2.6E-59) a mitochondrial HSP. Contig0010, which had a high similarity (E-value = 3E-70) to a heat shock protein, consisted of nine ESTs. Contig0022 (composed of three ESTs), contig0031 (consisted of two ESTs), and contig0049 (consisted of two ESTs) were all found to closely match (E-values < 2.1E-37) heat shock proteins. Contig0067 (E-value = 3.1 E-96) and contigs0071 (E-value = 4.9 E-66) both consisted of two ESTs and were observed to have BLASTx hits matching heat shock proteins as well. A large amount of ESTs also had BLASTx hits relevant to heat stress or closely matched heat shock proteins.

Heat shock proteins are expressed by animals, plants, fungi, and algae. Although different families of HSPs are found in different organisms, they can be grouped based on sequence similarity and molecular weight. Interestingly, although they are found in a diversity of species, they have retained sequence similarity (Kriehuber et al., 2010). Misfolding of proteins can be prevented by sHSPs; however, larger molecular weight HSPs assist as molecular chaperones so-called because they carry out energy-dependent processes. HSP90 and HSP70 both have an ATP-binding domain, which functions in the chaperone cycle of binding, folding, and releasing of clients (misfolded proteins) (Daugaard, Rohde, & J\ä\ättel\ä, 2007).

Response to oxidative stress was observed with multiple BLASTx hits that matched cytochrome c oxidase and peroxidase. Contig0043 showed high similarity (1E-109) to an ascorbate peroxidase previously found in *Citrus maxima*. Ascorbate peroxidase (AsPX), an enzyme unique to plant and algae, functions as a vital protectorate from highly reactive hydrogen peroxide and hydroxyl radicals, particularly within the chloroplast (Asada, 2006). Higher levels of AsPX in fruit have been associated with longer postharvest shelf life and increased antioxidant activity (Lester, 2008). Heat stress can quickly cause build up of reactive oxygen species (ROS); therefore, finding enzymes that mitigate the effects of ROS further support the data that *C. officinalis* was growing in stress conditions.

Although stress-related sequences were abundant, finding sequences similar to terpene or oleoresin pathways was difficult. Gramosa and Silveira (2005) performed a GC/MS analysis on *C. officinalis* specimens from Brazil. They found more than 40 different compounds with an abundance of γ-muurolene and β-caryophyllene in the

leaves. They captured a different terpene profile from separate parts of the plant (e.g., leaves, seed, bark, etc.) that suggested multiple proteins were required to generate such diverse metabolites. Since our library was taken from developing leaves and petioles, we expected to find correlating terpene synthase expression. However, this was not the case and might have been partly due to lack of sequence data and age or growth conditions of the organism.

In some cases, it is important to closely examine BLASTx results for less similar hits. For example, although no matches to genes encoding for terpene synthases were observed, the presence of a sequence similar to rubber elongation factor (REF) previously described in the rubber tree (*Hevea brasiliensis*; Priya, Venkatachalam, & Thulaseedharan, 2006) was found by examining BLASTx hits with higher E-values. This protein facilitates the interaction between a growing cis-polyisoprene unit (i.e., rubber), from isopentenyl pyrophosphate (IPP), and a prenyltransferase (Dennis & Light, 1989). REF is believed to affect the stereochemistry of IPP, allowing for a cis addition to the growing cis-polyisoprene units (Priya et al., 2006). Its absence prevents further addition of IPP onto cis-polyisoprene by prenyltransferase (Dennis & Light, 1989). The presence of a potential homolog to REF in *C. officinalis* provides key insight into possible molecular pathways involved in the production of large terpenes seen in extracted oleoresin.

Some researchers have taken on broader studies of metabolite production. Da Silva Medeiros and Vieira (2008) have provided work on a related species, *C. multijuga*, and found abiotic and biotic factors (e.g., termites, age and size of tree) contributed to oleoresin production. A separate study found little relationship between soil types

(variations in nitrogen and moisture) and leaf sesquiterpene content in leaves from *C. multijua* (Nascimento & Langenheim, 1986). Similar results were found by Feibert and Langenheim (1988); they attributed higher sesquiterpenes in leaves mostly to herbivory (response to stimuli) as did Langenheim, Convis, Macedo, and Stubblebine (1986).

Although our *Copaifera* cDNA library shed little insight into the biochemical pathways and genes involved in oleoresin synthesis, this library was found to be important in understanding *Copaifera* developmental and heat stress-related gene expression. Incorporation of a plant expression vector in library construction allows for transfer and expression of novel heat stress-related genes expressed in non-*Copaifera* species (e.g., *A. thaliana*). Based on the diversity of sequences obtained from this tropical tree, further research might be able to include studies on heat, drought, and other predicted climatic responses on other tropical species.

Construction of cDNA or EST libraries provides information on the transcriptional state of an organism. Hundreds of thousands of plant ESTs have been uploaded to public databases, providing researchers with an inordinate amount of genomic data without the burden and expense of full genome sequencing. Sequence information from these libraries has been important in understanding plants in relation to invasiveness (Wang, Yang, Liu, Jiang, & Wu, 2006), weediness, (Anderson et al., 2007; Broz et al., 2007), biotic stress responses (Zhang et al., 2007), and flowering times (Carlson et al., 2006). *C. officinalis* ESTs are lacking in databases and the research presented here contributed to filling this gap.

Finally, GO terms in this small EST collection were compared to an *A. thaliana* reference set using a Fisher's exact test. GO terms that were significantly

overrepresented included response to stress and response to abiotic stimulus. This, coupled with the fact that a majority of BLASTx matches were closely related to stress-related proteins, suggested the *C. officinalis* was growing under stressful conditions. Since the library construction incorporated a binary vector and sequencing plates were obtained, future work should focus on expressing *C. officinalis* stress-related genes in a model organism such as *A. thaliana*. This would help facilitate a rapid approach for plant transformation studies and also help determine what advantage *C. officinalis* stress-related genes might provide when expressed in *A. thaliana* growing at elevated temperatures.

CHAPTER III

SELECTION OF GENES AND PLANT TRANSFORMATION

Introduction

Many different plant species have been genetically modified and several different methods of plant transformation have been used (Wang, Vinocur, & Altman, 2003). The diversity of genes demonstrates not only the diverse potential of plant biotechnology but also suggests this technology might offer unique benefits to society. This chapter discusses each clone chosen from the *Copaifera officinalis* EST library. In addition, its subsequent insertion in *Arabidopsis thaliana* and confirmation of expression are also discussed. The potential for using the *C. officinalis* library for other investigations is also briefly considered along with a comparison to other research in *A. thaliana* transformation.

Obtaining and Cloning Transgenes

Many researchers rely on cloning methods to obtain their gene of interest (GOI). Often, genes that have been previously described can be obtained by using cloning primers to amplify the region of DNA containing the transgene. However, as in the case of eukaryotes, this includes non-coding intervening sequences (introns) that might not be part of a functional transcript. For this reason, RNA can be reverse transcribed to yield cDNA and primers can be used to PCR amplify the desired sequence. This is followed by gel purification and ligation into an acceptable expression vector.

In this project, genes of interest from *C. officinalis* were cloned into a binary plasmid for construction of a binary library and are described by their location on 96-well plates (see Chapter II). Six constructs (TPG03, 1A08, 3F04, 4B07, 4H05, and 7A05) were chosen based on several important considerations.

First, association of the product of each insert with heat stress was sought; this was based on the function from the BLASTx results from Chapter II. Gene Ontology terms also helped characterize the function of each gene product. Second, it was necessary for each insert be of such a length so that it would encode an entire protein from the ATG start site to a stop codon. This required sequencing the full length of each insert and followed by comparison to homologous genes in which the entire transcript length is known. If the insert contained an ATG translation initiation codon and was similar length to homologous genes, then it was a good candidate to express in *A. thaliana*.

Finally, and also of high importance, the predicted function of each insert's protein product was determined. Inserts with protein products involved in metabolic pathways were undesirable unless otherwise justified. For example, if overexpression of a particular enzyme was thought to confer thermotolerance, it was of interest to determine if that protein was also the rate-limiting step of the pathway. Since a rate-limiting step can limit the formation of an important metabolite, overexpressing enzymes not involved in a rate-limiting step might not confer an advantage to the transgenic plant when exposed to heat stress.

Plant Transgenic Research and
Arabidopsis thaliana

A commonly used plant in transgenic experiments is *A. thaliana* (thale cress). The advantages of using *A. thaliana* plants include efficiency of transformation (~1.0%), short generation time (~7 weeks), and large number (>5,000) of seeds produced (Meinke, Cherry, Michael, Rounsley, & Koornneef, 1998). It is also useful in genetic studies because of its small genome (~27,000 genes; 125Mb) and diploid level (2x=10). Various methods of introducing a foreign gene into a plant exist (Sanford, 1990); however, the floral dip method (Clough & Bent, 1998), which utilizes molecular machinery of *Agrobacterium tumefaciens*, is a common and reliable technique for transformation of *A. thaliana*.

***Agrobacterium tumefaciens* and**
Gene Transfer

Agrobacterium are Gram negative, rod shaped bacteria that reside in soil (Lippincott & Lippincott, 1975). Upon wounding or breach of a plant's epidermis, *A. tumefaciens* cells are able to invade plant cells (Ream, 1989). The success of infection depends on the tumor inducing (Ti) plasmid. This is a large plasmid (~200,000bp) that has Ti genes responsible for transferring bacterial genes into the plant cell (Bevan & Chilton, 1982). Residing on the Ti plasmid is a section of DNA called transfer DNA (T-DNA)--the genes transferred from the bacterium to the plant cell. Some of the Ti genes encode for a type IV secretion system (T4SS), which is similar to a molecular syringe and is responsible for shuttling the T-DNA from *A. tumefaciens* to the plant cell (Christie & Cascales, 2005).

The expression and assembly of the T4SS is dependent upon sensing sugars or phenolic compounds that bind to a histidine kinase receptor. This membrane protein phosphorylates a transcription initiation factor that binds to promoters of virulence operons. In a natural setting, the T-DNA that gets introduced into the plant cell encodes metabolic pathways for compounds (octopine, nopaline, agropine) that *A. tumefaciens* can use, but which the plant does not recognize and therefore cannot use. In conjunction with these major metabolites, hormones are also produced that act upon the plant cell machinery to produce a cellular compartment around the *A. tumefaciens* cells. This space serves as a refuge in which the bacterial cells will live and appears as a gall around the site of infection.

A. tumefaciens was first described to have utility in introducing foreign DNA into plants after it was determined that the flanking border regions (left border/LB and right border/RB) of the T-DNA can be separated on a plasmid, aptly named a binary plasmid. In addition, any DNA can be inserted between the LB and RB to be transferred to a plant genome (Hoekema, Hirsch, Hooykaas, & Schilperoort, 1983).

To transform *A. thaliana*, a compatible transformation system is necessary and is contingent upon a binary plasmid and an *A. tumefaciens* strain, each with different antibiotic selection markers. Binary plasmids are usually small (<10kb) and have a multiple cloning site downstream of a particular promoter, a plant selection marker, and a bacterial selection marker within the T-DNA border sequences. An origin of replication is also present that allows manipulation of the plasmid in either *E. coli* or *A. tumefaciens*. One part of the T-DNA that becomes incorporated into the plant genome is an antibiotic resistance gene that is placed adjacent to the GOI so that transfer of both genes occurs

simultaneously. If a plant cell is successfully transformed, it will grow on selective media; nontransformants do not survive.

pCHF3 Binary Vector Description

The binary vector used in this portion of the study was pCHF3 (Liu, Blount, Steele, & Dixon, 2002). The plant selectable marker (*nptII*) and the insert are each located directly downstream of a CaMV35S promoter. The bacterial selection markers encode resistance to spectinomycin and streptomycin. The vector map and nucleotide sequence for pCHF3 are provided in Appendixes A and B, respectively.

Upstream and downstream elements can greatly affect mRNA processing and protein translation. Coding sequences contain a promoter, a coding region, and a terminator; all have distinct primary sequences that influence gene expression. The 5' untranslated region (UTR) is important in mRNA transcripts because it helps stabilize the mRNA within the cytosol and helps the small ribosomal subunit associate. In plants, the 5' UTR can range from 30-130 nucleotides (Kochetov et al., 2002). Important to translation, the presence of an of an ATG start codon is needed to initiate transcription by a charged tRNA molecule carrying the amino acid methionine. A stop codon is also necessary to stop translation. In many sequences, the presence of a polyA tail was observed.

The Floral Dip

In the floral dip method, *A. tumefaciens* harboring a binary vector is grown overnight and resuspended to a specific optical density. The resulting *A. tumefaciens* solution is used to dip *A. thaliana* flowers so that some of the *A. tumefaciens* cells in the suspension infect the ovule. Since *A. thaliana* has two carpels that do not fuse together

until flowering, there is an opportunity for *A. tumefaciens* to infect the ovules--the primary site of infection by *A. tumefaciens* (Desfeux, Clough, & Bent, 2000). After the flowers mature and dry siliques develop, the seeds are harvested and sown on selective media.

Murashige and Skoog (MS) media containing an antibiotic (e.g., kanamycin) allows for germination of all seeds (Valvekens, Montagu, & Lijsebettens, 1988). However, plants with the selectable marker gene appear green in color and non-transgenic plants appear small and yellowish in color. The transgenic plants can be transferred from the selective media to soil and allowed to mature and develop seeds. Transgenic plants are observed in the F1 generation. Transgenic plants grown on ½ MS media with kanamycin are apparent in 7-10 days. These plants appear as green plants with development of true leaves and long roots that extend into the media. Since the transgene and selectable marker gene are inherited in a Mendelian ratio (3:1), T3 seeds can be sown on selective media to determine homozygosity.

Materials and Methods

Choice of Constructs

A total of six vector-insert constructs obtained from the *C. officinalis* binary library were used. The insert of each construct chosen encoded a protein that functions in heat stress. Inserts were compared to homologous cDNAs from other species; only those that had a similar length were used in this study. BLASTx and Gene Ontology results were used to help characterize heat stress-associated transcripts. Initially, insert length was determined by PCR amplification of each insert. The complete sequence of each insert was subsequently obtained (UW Madison, WI) and is provided in Appendix C.

Presence of a methionine as the first amino acid in the translated nucleotide sequence indicated that the initiation codon was present for translation; therefore, protein production would likely result. After choosing six heat stress-associated genes that contained full-length coding regions from the *C. officinalis* library, the corresponding locations on each sequenced 96-well plate were picked for *E. coli* harboring each vector-insert construct and grown overnight in 5 mL broth with 50 µg mL^{-1} spectinomycin. Plasmid extraction was done using Quiagen's (USA) plasmid miniprep kit. Plasmid extraction was performed on each clone to obtain the construct for transformation into *A. tumefaciens*.

Electroporation of *A. tumefaciens*

A. tumefaciens EHA105 was grown in LB broth overnight and 5 mL was used to inoculate 500 mL of LB broth and grown overnight at 30˚C at 300 rpm until the optical density at 550 nm (OD550) was ~1.0. The cells were then centrifuged at 3000 x g for 10 minutes at 4˚C. After decanting the media, the pelleted cells were placed on ice and resuspended in deionized water with 50 mL of 10% glycerol (ice-cold). The cells were again centrifuged at 3000 x g for 10 minutes at 4˚C. After decanting supernatant, the cells were resuspended and washed again with 10% glycerol followed by centrifugation as described above. The resulting pellet was resuspended in 0.5 mL ice-cold 10% glycerol and 80 µL aliquots of electrocompetent cells were stored in -80˚C.

Electrocompetent *A. tumefaciens* was thawed in an ice bath for 10 minutes and 40 µL of cells were combined with ~80 ng of vector-insert construct and gently stirred with the pipette tip. The mixture was transferred to an ice-cold 0.1 cm cuvette and placed into the ShockPod of a Gene Pulser Xcell Electroporation System (BIO-RAD). After pulsing,

cells were recovered from the cuvette and allowed to grow for 3 hours at 30˚C with shaking at ~250 rpm. Selection of transformed cells was performed by growing all cells on LB agar plates with 10 µg mL^{-1} rifampicin and 50 µg mL^{-1} spectinomycin.

Transformation of *A. thaliana*

A. thaliana plants were grown for ~6 weeks until initial bolts appeared and those were clipped to increase the number of secondary bolts. Upon flowering, plants were dipped into an *A. tumefaciens* solution (50 g sucrose L^{-1}, 50 µL silwet L^{-1}) with an optical density of 0.8 at 600 nm. Plants were then placed in a plastic bag and placed under lights until the following day (18-24 hours). A second dip was performed to increase the number of transformed plants. After siliques matured, the seeds were harvested, sterilized, and sown on ½ MS with 1.0% sucrose and 50 µg mL^{-1} kanamycin. Transgenic seedlings displayed an elongated root and developed true leaves. Upon acquiring T2 generation, homozygous plants were selected by growing several lines and germinating seeds from all lines on selective media. If all seeds showed resistance to the selective media, the line was considered homozygous.

cDNA Synthesis and Detection of Transgene Expression

For each transgenic plant line, total RNA was extracted using Trizol (Invitrogen, USA). Less than 1 µg of total RNA from each line was treated with DNase I and reverse transcribed using the SuperScriptIII cDNA kit (Invitrogen, USA). Detection of expression of each insert, the kanamycin resistance gene (*nptII)* and the actin control gene (*act2)*, was performed using the following PCR conditions: three minutes at 94˚C, 30 seconds at 94˚C, 30 seconds at 55˚C, and 30 seconds at 72˚C for 33 cycles followed by a five minute extension at 72˚C.

Results

A. thaliana plants were successfully transformed using each of the six constructs chosen. Each clone chosen contained an ATG start site and, after further sequencing, the polyA site was observed (see Appendix D). Starting from the +1 nucleotide and stopping at the polyA tail, the longest insert encoded glutathione peroxidase (4H05) and was 1,282 bases long. The shortest length was for a mitochondrial HSP (3F04) and was 830 bases long. The average length of all inserts was 1,221 bases.

Gene specific detection primers (see Table 1) were used for PCR in each vector-insert construct (see Figure 10). All primers amplified the expected size of fragments (see Table 2). After generating cDNA from total RNA as previously described, each line was also tested using the same gene specific detection primers (see Figure 11).

Initially, two lines (3F04 and 4B07) did not amplify (lanes 4 and 5; see Figure 11). However, one of these lines (4B07) amplified upon a second attempt after increasing the number of PCR cycles from 33 to 40 (see Figure 12). The other line (3F04) consistently failed to show a band at several different annealing temperatures (50-60˚C) or when it did amplify, the negative control also amplified.

RT-PCR showed *nptII* expression in all transgenic lines (see Figure 13). The gene specific detection primers failed to produce bands in nontransgenic *A. thaliana* cDNA (see Figure 14).

Table 1

Forward and Reverse Detection Primers for Each Gene of Interest

Sequence Name	Forward Primer (5'-3')	Reverse Primer (5'-3')	Expected Length (Bases)
TPG03	**TACCCCGCTTTTCATTTTTG**	**CTTGTTCTCCATGGGTTGCT**	**105**
1A08	**AGAACTATGCGCCAGATGCT**	**CGCCTCTGATGACAAGAACA**	**219**
3F04	**TCCAGCAGGATTGACCTACC**	**ATCCGACCTGTCCTCCTCTT**	**116**
4B07	**ACGTTGTTTATTCAAACTTC**	CAGTATACATCAAAGTCTGC	**159**
4H05	GCTTTGTCCAGGGTTGTGAT	AGCAGCAAGGGCAGTAATGT	188
7A05	TTCCCAAACTCCAAAGCAAC	CTCAAGTATGCAGGGCTTCC	**230**
nptII	GCATACGCTTGATCCGGCTACC	TGATATTCGGCAAGCAGGCAT	**231**
act2	ACCAGCTCTTCCATCGAGAA	CAGCGATGCCTGAGAACATA	**255**

Note. Detection primers for each GOI were generated using Invitrogen's online primer design software (http://tools.invitrogen.com/content.cfm?pageid=9716). Each primer set had a Tm of 60°C and the same elongation time (30 seconds) was used for each.

Table 2

Each Insert Identification Correlates to Plate and Well Position in Which Construct Was Obtained

Insert name	Putative function	Length (bases)	Expected coding region (bases)	Source
TPG03	rubber elongation factor	1,136	743	Goyvaerts et al., 1991
1A08	sHSP	907	468	Simoes-Araujo et al., 2008
3F04	mitochondrial HSP	830	606	Lenne et al., 1995
4B07	auxin-induced protein/nodulin	1,208	597	Mohammad et al., 2004
4H05	glutathione peroxidase	1,282	766	Skipsey et al., 1997
7A05	allene oxide cyclase	924	732	Ziegler et al., 2000

Note. The putative function is based on the BLASTx result. The length is the number of bases from the +1 ATG site to the start of the polyA tail.

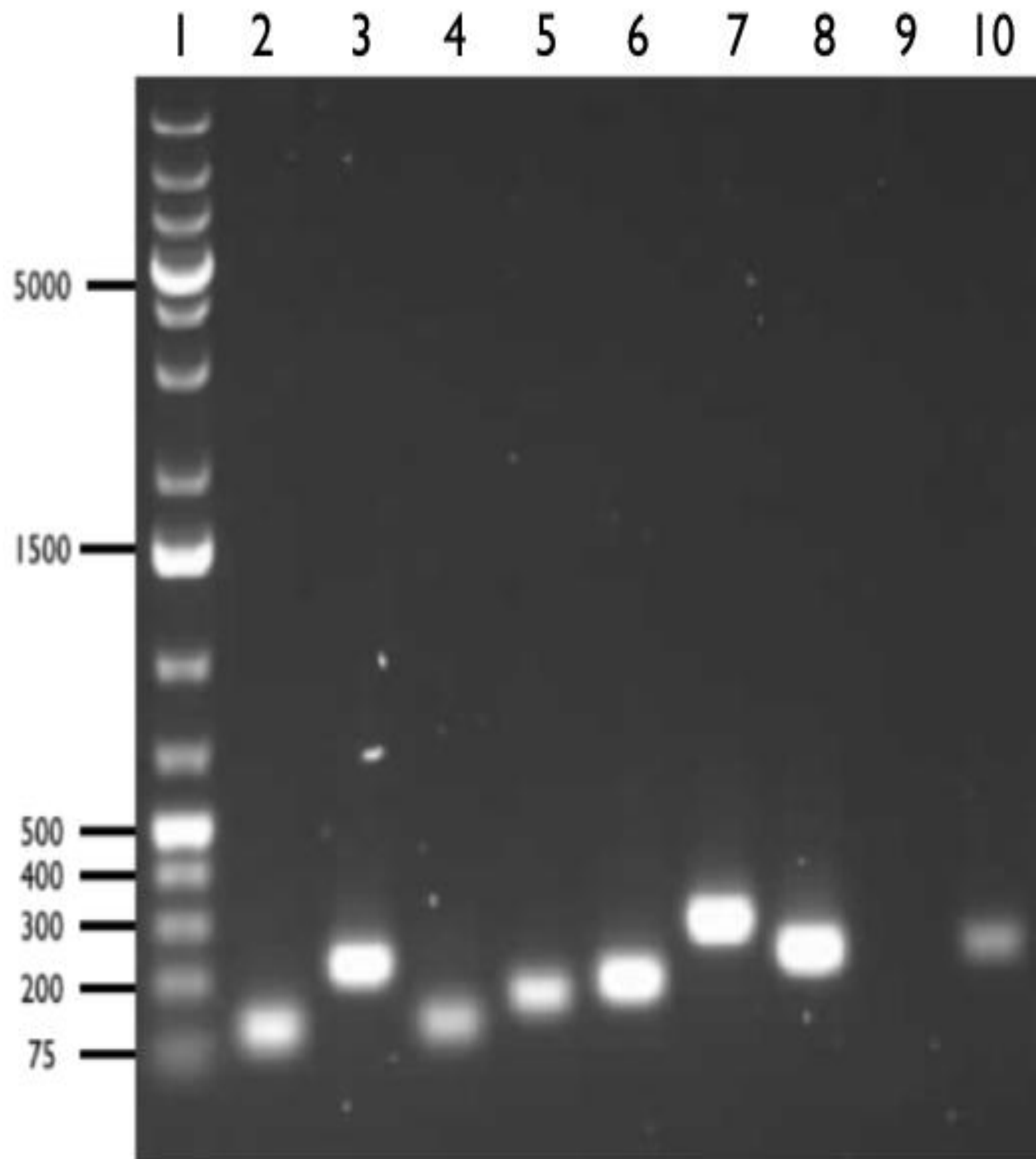


Figure 10. PCR results using gene-specific detection primers for each vector-insert construct. Lane 1-10; 1 kb ladder; TPG03, 1A08, 3F04, 4B07, 4H05, 6G10, 7A05, negative control, positive control (*act2*). Although shown in this gel image, 6G10 was later removed from the study.

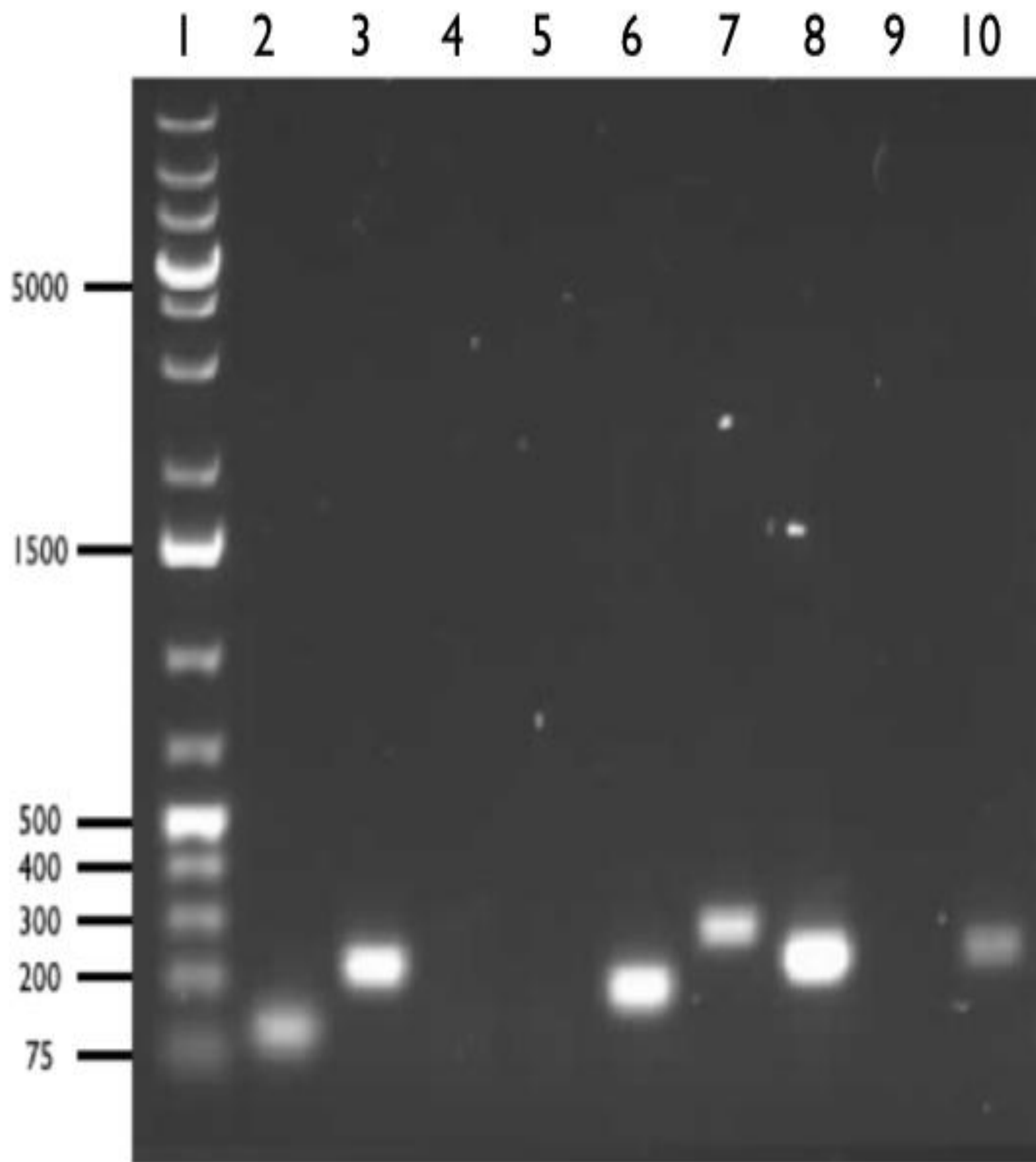


Figure 11. RT-PCR using gene-specific detection primers for each transgenic line. Lane 1-10; 1 kb ladder; TPG03, 1A08, 3F04, 4B07, 4H05, 6G10, 7A05, negative control, positive control (*act2*, 255bp). In this reaction, 3F04 and 4B07 did not amplify. Although shown in this gel image, 6G10 was later removed from the study.

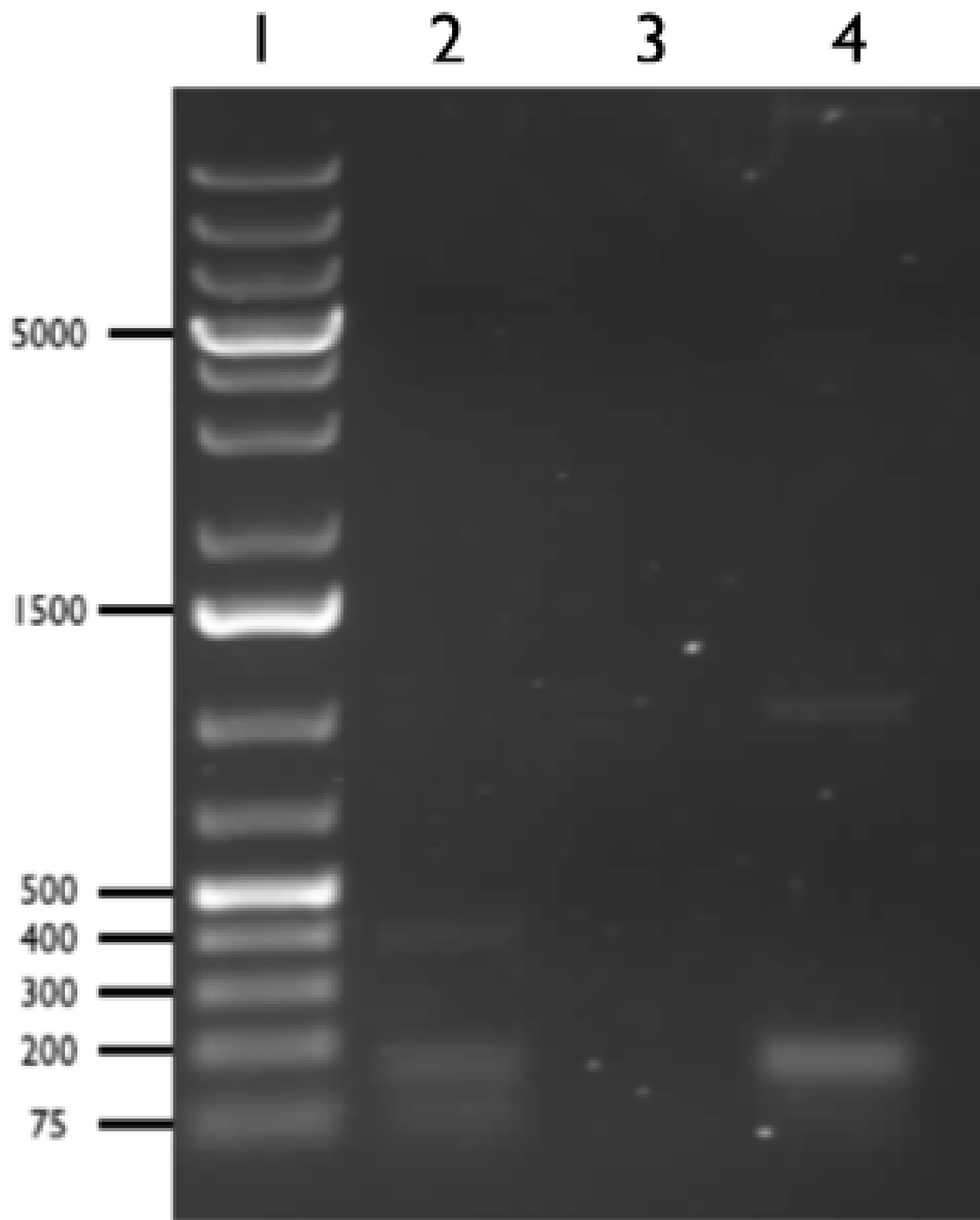


Figure 12. RT-PCR of 4B07. Lanes 1-4, 1 kb ladder, 4B07 cDNA, negative control, 4B07 insert-vector construct. After increasing the number of PCR cycles from 33 to 40, gene specific detection primers showed a product (159 bases) but also showed some bands due to non-specific priming.

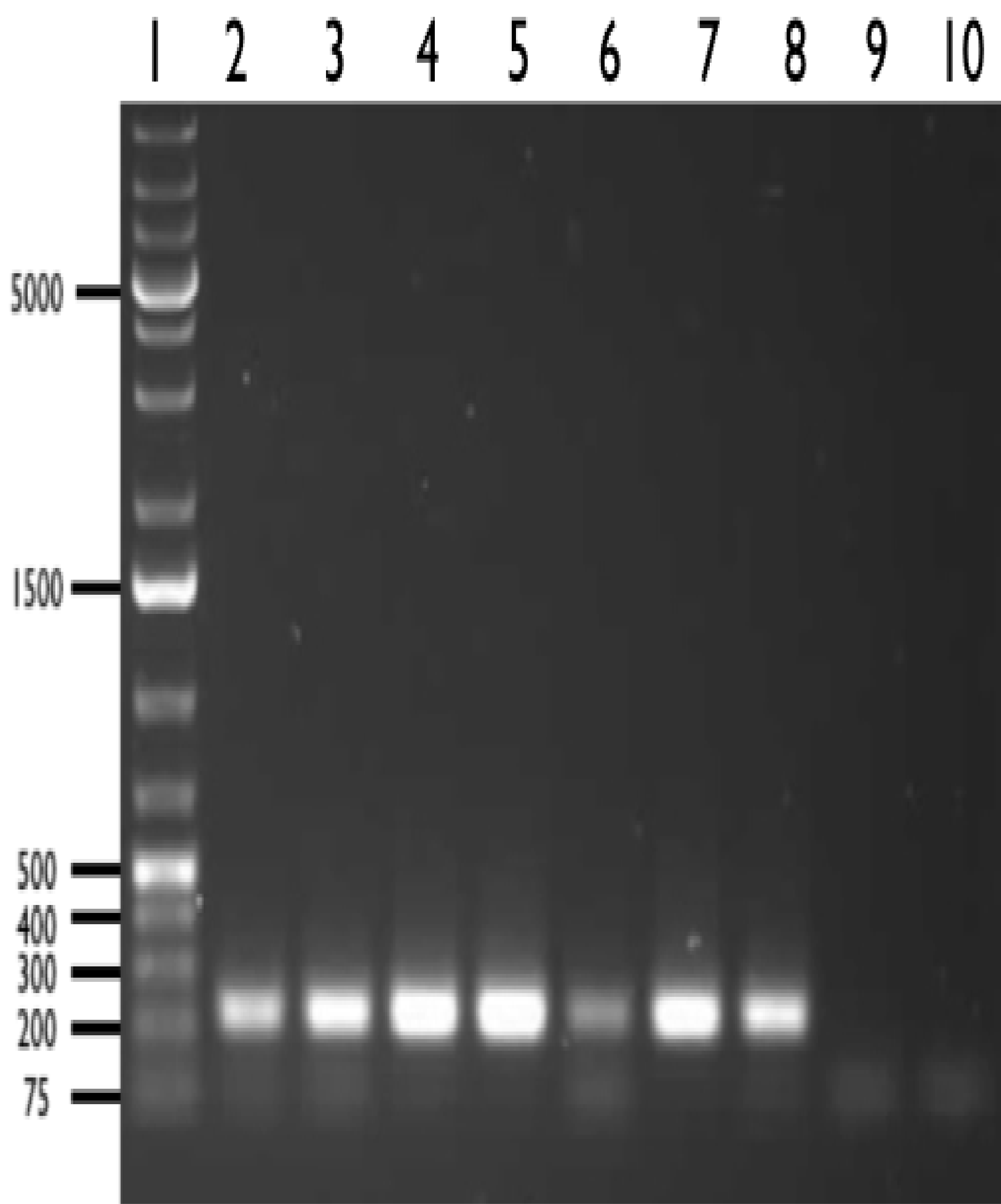


Figure 13. RT-PCR using *nptII* detection primers in each transgenic line. Lanes 1-10, 1 kb ladder, TPG03, 1A08, 3F04, 4B07, 4H05, 6G10, 7A05, negative control, *A. thaliana* cDNA with *nptII* primers. The expected band size was 231bp.

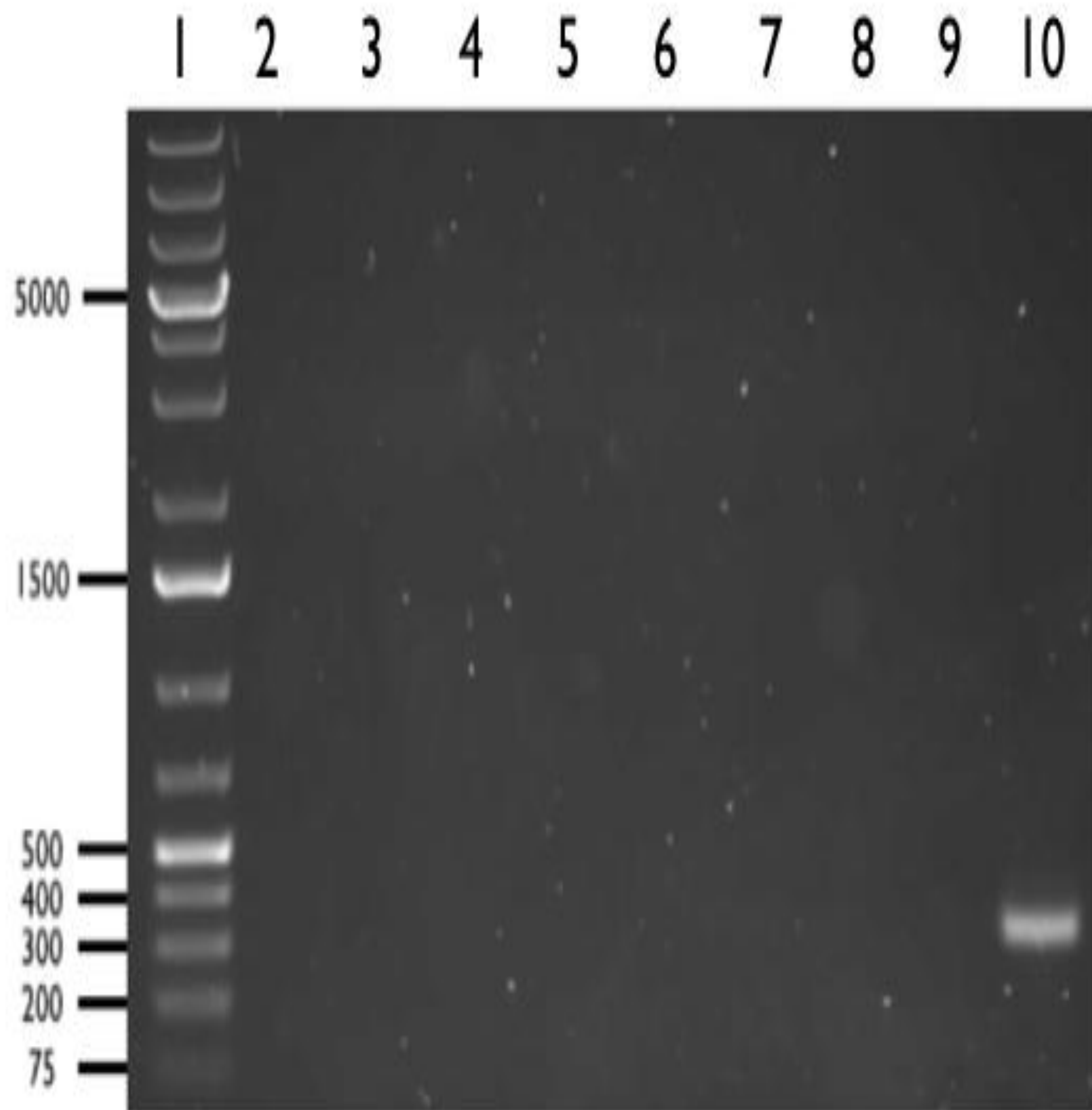


Figure 14. Gene specific detection primers using *A. thaliana* cDNA. Lanes 1, 1 kb ladder; TPG03, 1A08, 3F04, 4B07, 4H05, 6G10, 7A05, negative control, positive control (*act2*).

Discussion

Six constructs from the *C. officinalis* library were chosen and used to transform *A. thaliana* to generate homozygous transgenic lines. Although 6G10 was initially used, it was later removed from the study since it did not contain a full-length sequence. Two

constructs chosen had an insert with little known function--an auxin-induced/nodulin protein (4B07) and REF (TPG03). These two were chosen to investigate their possible role in thermotolerance. The other four inserts of each construct chosen had more well-known roles in heat stress. Although one of the main attributes sought for selection of clones from the library was the possible role in heat stress, presence of the complete coding sequence was also highly important.

Based on the sequencing results of each clone, these inserts closely matched the estimated size of full-length homologous cDNAs. Moreover, all inserts utilized showed an ATG start site at the 5' end and a polyA tail at the 3' end, suggesting presence of full length transcript. BLASTx results suggested these ATG sites were in frame.

It was unexpected that one line (3F04) did not show expression of the transgene; however, it did show expression of the *nptII* gene. The product of the transgene was a putative mitochondrial HSP. It has been suggested that in nuclear encoded mitochondrial genes, the 5' UTR and introns within the gene can play an important role in mRNA processing (Cenik et al., 2011). However, to determine the precise mechanism of why this line did not express the transgene, further investigation is necessary.

Factors Affecting Transgene Expression

Since transfection and T-DNA delivery are stochastic, the DNA can become incorporated anywhere in the genome. Therefore, variation in transgene expression between lines is observed due to the variation of insertion location. If the T-DNA is incorporated into a highly expressed region of the genome, its level of expression might also be high. In contrast, insertion of the transgene into a less active (e.g. centromeric or telomeric) region of the genome might result in lower levels of gene expression.

In addition to random insertion, variation in transgene expression is also influenced if multiple transgene copies are inserted by *A. tumefaciens*. In the case of delivering multiple copies, plant lines might show elevated levels of transgene expression, which might prove beneficial in downstream applications. For example, using a line that has multiple transgene copies in stress assays might lead to increased levels of stress tolerance. However, multiple transgene insertions have led to gene silencing and might be due to transcripts inducing the RNAi pathway, which results in transcript degradation (Pal-Bhadra, Bhadra, & Birchler, 2002).

Several other factors can influence gene expression or retention. For example, the T-DNA delivered by *A. tumefaciens* might become inserted within another gene, perhaps a gene having high importance. If the gene in which the T-DNA is inserted is responsible for regulating major metabolic pathways, then T-DNA integration might ultimately inhibit the plant from proper growth and development. In addition, a transgene might be lost or duplicated if it is incorporated within a transposable element.

Other difficulties can occur in plant transgenic studies such as variation in codon usage between different species. Codon usage bias is observed between species that rely on different tRNAs to carry amino acids to the acceptor site of the ribosome. Since species might differ in their tRNA synthases and levels of charged tRNAs, processivity of protein might not occur as efficiently.

Of additional concern, many genes that are successfully translated might require posttranslational modifications. If the cellular machinery is not able to properly modify the protein product to enable its function, then the production of the protein might actually hinder plant growth by diverting energy resources away from functional proteins.

These factors--the cellular destination of mRNA, stochastic insertion, gene silencing, number of transgene copies, codon usage, and post-translational modifications --all contribute to the uncertainty of transcription and translation of a transgene. Many researchers generate multiple plant lines that express the same transgene. In this study, only one line was obtained for each of the six constructs used. Although the binary library method generated six transgenic plant lines each harboring a distinct transgene, different lines might show different levels of expression. Therefore, future studies should focus on generating several lines.

Benefits of the Binary Library Approach

Due to its simplicity, the floral dip is perhaps the most common way to transform *A. thaliana* (Clough & Bent, 1998). Although previous research has investigated *A. thaliana* plants that have been transformed with heat-stress related genes, to the best of our knowledge no previous research has used a gene derived from *C. officinalis*. Curiously, techniques that generate a cDNA library and incorporate a plant binary vector have not been widely performed.

The binary library approach has advantages over other plant transformation techniques, e.g., allowing selection from a wide variety of genes that are already cloned into a plant expression vector. Ichikawa et al. (2006) characterized several genes based on gain-of-function of individual genes by generating a full-length *A. thaliana* cDNA library and then inserting the constructs into *A. tumefaciens* cells. Using the floral dip method, they generated more than 15,000 lines of *A. thaliana* transgenic lines and described nearly 1,500 lines with altered phenotypes. This system, called Full-length cDNA Over-eXpressing gene hunting system, or FOX hunting system, holds much

promise for large scale studies of gene function. In a separate study, Wang et al. (2009) cloned full-length cDNA from tomato (*Lycopersicon esculentum*) into *A. tumefaciens* and transformed *A. thaliana*. Transformation was done with a mixture of *A. tumefaciens* cells; thus, the transformed seeds consisted of a diverse set of lines each expressing a unique cDNA. Using this method, they generated more than 7,000 *A. thaliana* lines. After characterizing interesting phenotypes, they identified the sequences responsible for inducing the phenotypes using a forward 35S primer, which allowed for determining the transgene that was incorporated into *A. thaliana*.

While both of the methods described above incorporated a plant binary vector in library construction, no studies have generated transgenic lines using the binary library approach. The advantage is that after obtaining the sequenced 96-well plates, hundreds or even thousands of unique transcripts are immediately available for use in generating transgenic plant lines. However, in the binary library approach, more careful examination of each gene is permitted; whereas in the methods used by Ichikawa et al. (2006) and Wang et al. (2009), the procedure is not specific in choosing a transgene. The transgene that becomes inserted is identified after observing a particular phenotype. In this chapter, the binary library approach has shown to be a reliable method for selection, transformation, and expression of *C. officinalis* genes in *A. thaliana*.

CHAPTER IV

ASSESSING THERMOTOLERANCE IN TRANSGENIC LINES

Introduction

Plants are sessile organisms and are unable to relocate when subjected to stressful environments and so have evolved physiological, molecular, and genetic mechanisms that help them cope with unfavorable conditions. These mechanisms include either large-scale physiological responses (e.g., stomatal closure to limit transpiration rates) or adjustments in gene expression (e.g., up-regulation or down-regulation of gene expression).

Projected global climate temperatures indicate crop plants will have to cope with warmer temperatures in the near future (Intergovernmental Panel on Climate Change, 2007). The world's human population faces the challenge of providing food despite the negative effect of heat on crop plants (Battisti & Naylor, 2009; Porter & Semenov, 2005). Therefore, some people have suggested that it might be necessary to genetically engineer heat-tolerant crop plants (Mittler & Blumwald, 2010). An important initial step to prepare for generating heat-tolerant crop plants is to identify genes that might increase levels of thermotolerance and overexpress these genes in model plants.

Comprehending the metabolic pathways and transcriptional changes that occur in response to heat stress can be difficult. Rizhsky et al. (2004) showed that in response to a

combination of heat and drought, there are about 1,000 transcripts that show upregulation in *A. thaliana*. Dissection of each of these changes is an important goal in understanding distinct metabolic events that occur throughout the cell under heat stress conditions. Current research has moved into understanding the transcriptional machinery involved in the heat stress response and uses recent findings from molecular, genomic, and transcriptomic data (Hirayama & Shinozaki, 2010).

Several reports suggest that overexpression of heat stress-related genes can confer thermotolerance (Bhatnagar-Mathur, Vadez, & Sharma, 2008). In this study, six constructs from the *C. officinalis* library were each used to transform *A. thaliana*. The protein product of the insert of each construct was related to heat stress response and therefore might confer thermotolerance in transgenic *A. thaliana* lines that were generated (see Chapter III).

Association with Heat Stress of Transgenes

The first construct (TPG03) chosen harbored an insert that encoded a rubber elongation factor (REF). This was interesting to investigate due to its elusive role in heat stress. Gene Ontology (GO) terms associated with this transcript included *translation elongation factor activity* and *response to stress*. Chow et al. (2007) analyzed 10,040 ESTs from *H. brasiliensis* and found REF was related to stress-associated proteins. In another study, Larkindale and Vierling (2008) used a microarray to characterize heat stress-associated transcripts in *A. thaliana* and found that a homolog of REF had relatively high expression levels. Therefore, it was hypothesized overexpression of a *C. officinalis* REF might confer thermotolerance in transgenic *A. thaliana* plants.

The next construct chosen (1A08) harbored an insert that encoded a small heat shock protein (sHSP). The EST for this insert showed similarity to eight other ESTs in the *C. officinalis* library. EST 1A08 was chosen from among the other eight ESTs because more sequence data from that EST were available from the first sequencing run. The BLASTx results for 1A08 showed it had high similarity to a heat shock protein (E-value = 3.3E-70). GO results suggested this protein had a broad role in stress response indicated by terms such as *response to abiotic stimulu*s and *response to stress*. Due to the presence of eight other ESTs in the *C. officinalis* library with similar sequences, it was suspected that this small heat shock protein played a significant role in the heat stress response.

For the next construct of interest (3F04), BLASTx results (E-value = 3.19E-59) suggested that the insert encoded a low molecular weight (22 kDa) heat shock protein. GO terms associated with this sequence were *response to abiotic stimulus* and *response to stress*. Several sHSPs were overexpressed in monocots and dicots. Some studies have shown that overexpression of sHSPs conferred thermotolerance in transgenic plants (Nautiyal & Shono, 2010; Sato & Yokoya, 2007).

The next construct (4B07) harbored an insert whose translated sequence closely matched an auxin-induced protein (E-value = 5.96E-80). This also was similar to a "nodulin MtN21 family protein" (E-value = 1E-36) and a "nodulin-like protein" (E-value = 4E-29) from *A. thaliana*. Only one GO term, *cellular membrane*, was associated with this sequence, suggesting this auxin-induced/nodulin protein was localized to the cell membrane. Only recently have nodulins been suspected of playing a role in different stresses. Gollhofer, Schläwicke, Jungnick, Schmidt, and Buckhout (2011) showed that

several nodulin-like genes were expressed upon iron deficiency (a form of stress). They suggested that some nodulins were localized to the cell membrane and played an important part in regulating iron uptake. Construct 4G07 was investigated further since little is known about the role of nodulin-like proteins in heat stress.

Another construct (4H05) had an insert whose predicted protein product closely matched (E-value = 9.2E-20) a peroxidase. Other hits from NCBI suggested it was closely related to glutathione peroxidase, which catalyzes H_2O_2 into water and oxygen. Since heat stress often leads to an increase in reactive oxygen species (ROS), upregulation of a peroxidase might decrease the damage from free radicals resulting from heat stress. Previous research (Miao et al., 2006) suggested overexpression of a *C. officinalis* peroxidase in *A. thaliana* might provide increased thermotolerance.

Finally, the last construct chosen (7A05) had an insert whose protein product closely matched (E-value = 4.7E-41) allene oxide cyclase. GO terms suggested the enzyme was localized to the plastid. Allene oxide cyclase catalyzes 12,13-epoxylinolenic acid to yield cis-(+)-oxophytodienoic acid, which are both precursors of jasmonic acid, an important stress-signaling molecule. However, allene oxide cyclase is important in that it generates the first biologically active metabolite (cis-(+)-oxophytodienoic acid) in the jasmonic acid pathway (Schaller et al., 2008). Therefore, overexpression of allene oxide cyclase might confer increased levels of stress resistance.

Assessing for Thermotolerance

Thermotolerance can be assessed at various developmental stages. One method used to assess thermotolerance is by using 2.5-day-old seedlings and measuring their hypocotyl length before and after heat stress (Hong & Vierling, 2000). The hypocotyl is

the portion on the stem of the developing embryo that is located above the root and below the cotyledons. Germinating seeds have a hypocotyl that rapidly elongates while the cotyledons and shoot apical meristem are inhibited from developing. After stratification, plates are placed vertically in the dark for two or three days and then are subjected to a heat treatment. After heat treatment, the position of the cotyledons is marked and the plates are returned to the dark. After another two to three days, the change in hypocotyl is measured. Developing seedlings are largely affected by elevated temperatures because of the extensive amount of phospholipid present in dividing cells membranes and organelles. Therefore, thermotolerant plants are typified by having a longer hypocotyl when compared to control plants.

In another assay, plants were grown *in vitro*; after five days, they are subjected to a heat treatment (Clarke, Mur, Wood, & Scott, 2004). At this stage, *A. thaliana* plants have two to four true leaves. Different times and temperatures have been used for five-day-old plants. Some of these results indicate that a heat treatment of 45 minutes results in heat stress appearing in 50% or more of nontransgenic plants (Clarke et al., 2004).

Another method to assess thermotolerance is to use young, pre-flowering *A. thaliana* plants that are about 10-13 days old (Xue et al., 2010). At this stage in development, plants typically have 4-10 true leaves and can be stressed for 16 hours at about 45°C. Since plants are grown in soil and are slightly larger than five-day old plants, they are able to thermoregulate by opening their stomata. The subsequent evapotranspirational cooling helps the plants dissipate heat and, thus, can help preserve the integrity of the intracellular environment.

In the present study, a hypocotyl elongation assay, *in vitro* assay, and an *in vivo* assay were used to determine thermotolerance of plants that were 2.5 days, 5 days and 13 days old, respectively. Three hypotheses were tested, each relating to the heat stress assay performed. First, five-day-old transgenic plants would have green, healthy leaves after heat treatment. Second, germinated seedlings would have longer hypocotyls compared to controls after recovery from heat treatment. Finally, 13-day-old plants would have fewer damaged (withered and bleached) leaves after heat treatment when compared to non-transgenic plants.

Materials and Methods

Heat-shock conditions were tested in each of the six transgenic lines and control plants. Due to the wide variety of times, temperatures, and growth stages used in previous heat stress experiments, initial assays were performed to help determine what times and temperatures resulted in observable heat stress. Each assay was statistically interpreted to determine significance. All germinating seeds and plants were grown under continuous light using fluorescent bulbs (~5,500 lux). *In vitro* heat stress conditions were done using a ProBlot hybridization (Labnet International, Inc.) oven. All assays used homozygous T3 seeds derived from T2 parental lines. In all cases, seeds were subjected to cold (4˚C) conditions for at least two days for stratification.

In vitro **Thermotolerance Assay**

To determine thermotolerance in seedlings, transgenic and nontransgenic *A. thaliana* seeds were sown on MS plates and grown for five days under continuous light at 22˚C. The seedlings were then heat shocked for 30, 45, or 60 minutes at 45˚C. Following heat treatment, seedlings were incubated at 22˚C under continuous light.

Susceptibility was scored as heat stressed based on the presence of bleached cotyledons. Three replicates of each susceptibility assay were performed using 30 seeds per plate for each transgenic line. Nontransgenic *A. thaliana* was used as a control. In addition, each line along with nontransgenic *A. thaliana* were not given a heat treatment and scored.

Hypocotyl Elongation Assay

Seeds were sown on Petri dishes (10cm^2) containing 1/2 MS salts, 1% sucrose, and 10 g L^{-1} agar and set in an upright position in the dark at 22˚C. Heat treatment was given by setting plates in a 45˚C incubator for 90 minutes. Positions of the cotyledons were marked and the plates were set in an upright position at 22˚C for 2.5 days without light. The hypocotyl of each germinated seedling was calculated by measuring the change in position of the hypocotyl three days after heat treatment. The hypocotyl elongation assay used 20 seeds per plate with three replicates for each line.

***In vivo* Thermotolerance Assay**

Transgenic and nontransgenic seeds were sown in soil and pots were placed in the dark at 4˚C. After three days, the pots were placed under continuous fluorescent light and allowed to grow for 11 days. The plants were then placed in conditions with elevated temperature (48˚C +/-2˚C) for 16 hours. Replications of the *in vivo* thermotolerance assay were performed using approximately 10 plants per pot with three replications of each line. Plants were scored as either alive and unaffected or stressed as indicated by withered or bleached leaves. *In vivo* assays were carried out using a plant growth chamber (Thermo Scientific) with heat lamps placed beneath the plants to maintain 48˚C. Plants were kept under continuous light (~3700 lux) during heat treatment.

Results

After seven days, transgenic and control plants that were subjected to 30 minutes of heat stress showed no bleaching of leaves. Two days after treatment with 45 minutes of heat stress, no bleaching was observed in either control or transgenic lines (see Figure 15). After five days of treatment with 45 minutes of heat, signs of stress became apparent in only two lines--7A08 and control (see Figure 16). Some plants from all lines showed signs of heat stress after seven days (see Figure 17). After 60 minutes of heat treatment, bleaching was observed to occur at a faster rate and also occurred in more plants among all transgenic lines.

Treatment with 45 minutes of heat stress resulted in 7% and 8% of control plants showing signs of stress five and seven days after heat treatment, respectively. Treatment of control plants with 60 minutes of heat stress resulted in 4%, 16%, and 18% of plants showing stress two, five, and seven days, respectively. Based on the results of the *in vivo* assay, about 35% of control plants showed signs of stress five days after the heat treatment (see Figure 18).

Treatment with 45 minutes of heat stress resulted in 3% of TPG03 plants showing signs of stress seven days after heat treatment. Treatment with 60 minutes of heat stress resulted in 4%, 16%, and 18% of plants showing signs of heat stress two, five, and seven days after heat stress, respectively. Based on the results of the *in vivo* assay, 23% of plants showed signs of stress after heat treatment.

Treatment with 45 minutes of heat stress resulted in 5% of 1A08 plants showing signs of stress seven days after heat treatment. Treatment with 60 minutes of heat stress resulted in 3%, 7%, and 9% of plants showing signs of heat stress two, five, and seven

days after heat stress, respectively. Based on the results of the *in vivo* assay, 13% of plants showed signs of stress after heat treatment.

Treatment with 45 minutes of heat stress resulted in 3% of 3F04 plants showing signs of stress seven days after heat treatment. Treatment with 60 minutes of heat stress resulted in 7%, 18%, and 42% of plants showing signs of heat stress two, five, and seven days after heat stress, respectively. Based on the results of the *in vivo* assay, 40% of plants showed signs of stress after heat treatment.

Treatment with 45 minutes of heat stress resulted in 3% of 4B07 plants showing signs of stress seven days after heat treatment. Treatment with 60 minutes of heat stress resulted in 1%, 13%, and 14% of plants showing signs of heat stress two, five, and seven days after heat stress, respectively. Based on the results of the *in vivo* assay, 17% of plants showed signs of stress after heat treatment.

Treatment with 45 minutes of heat stress resulted in 2% of 4H05 plants showing signs of stress seven days after heat treatment. Although no 4H05 plants showed signs of stress two days after treatment with 60 minutes of heat stress, 13% and 16% of plants showed signs of stress five and seven days later, respectively. Based on the results of the *in vivo* assay, 16% of plants showed signs of stress after heat treatment.

Treatment with 45 minutes of heat stress resulted in 8% of 7A05 plants showing signs of stress both five and seven days after heat treatment. Although no 7A05 plants showed stress two days after treatment with 60 minutes of heat stress, 10% and 18% of plants showed signs of heat stress five and seven days after heat treatment, respectfully. Based on the results of the *in vivo* assay 13% of plants showed signs of stress after heat treatment.

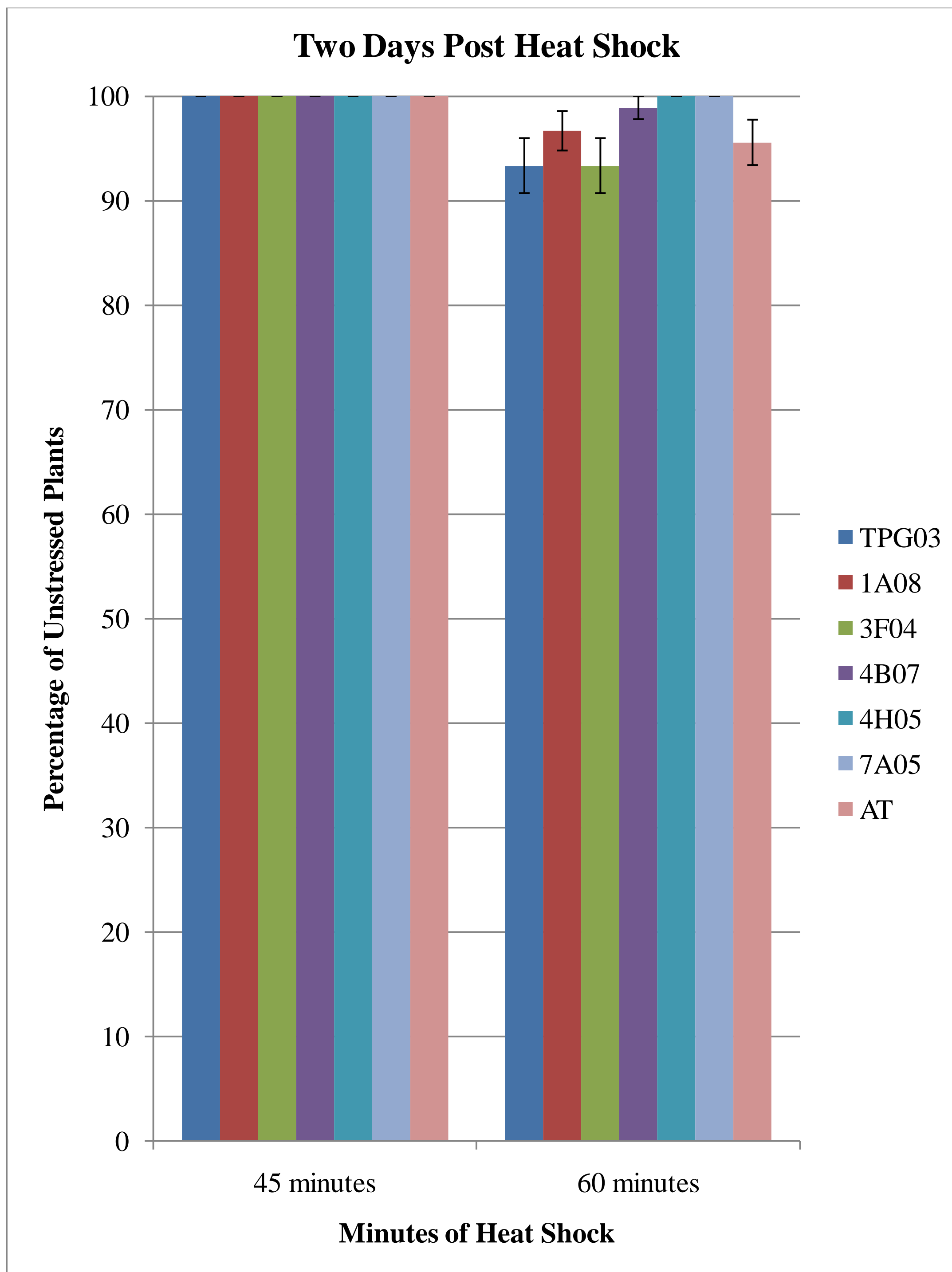


Figure 15. Results of the *in vitro* assay two days after heat shock (Nt, nontransgenic).

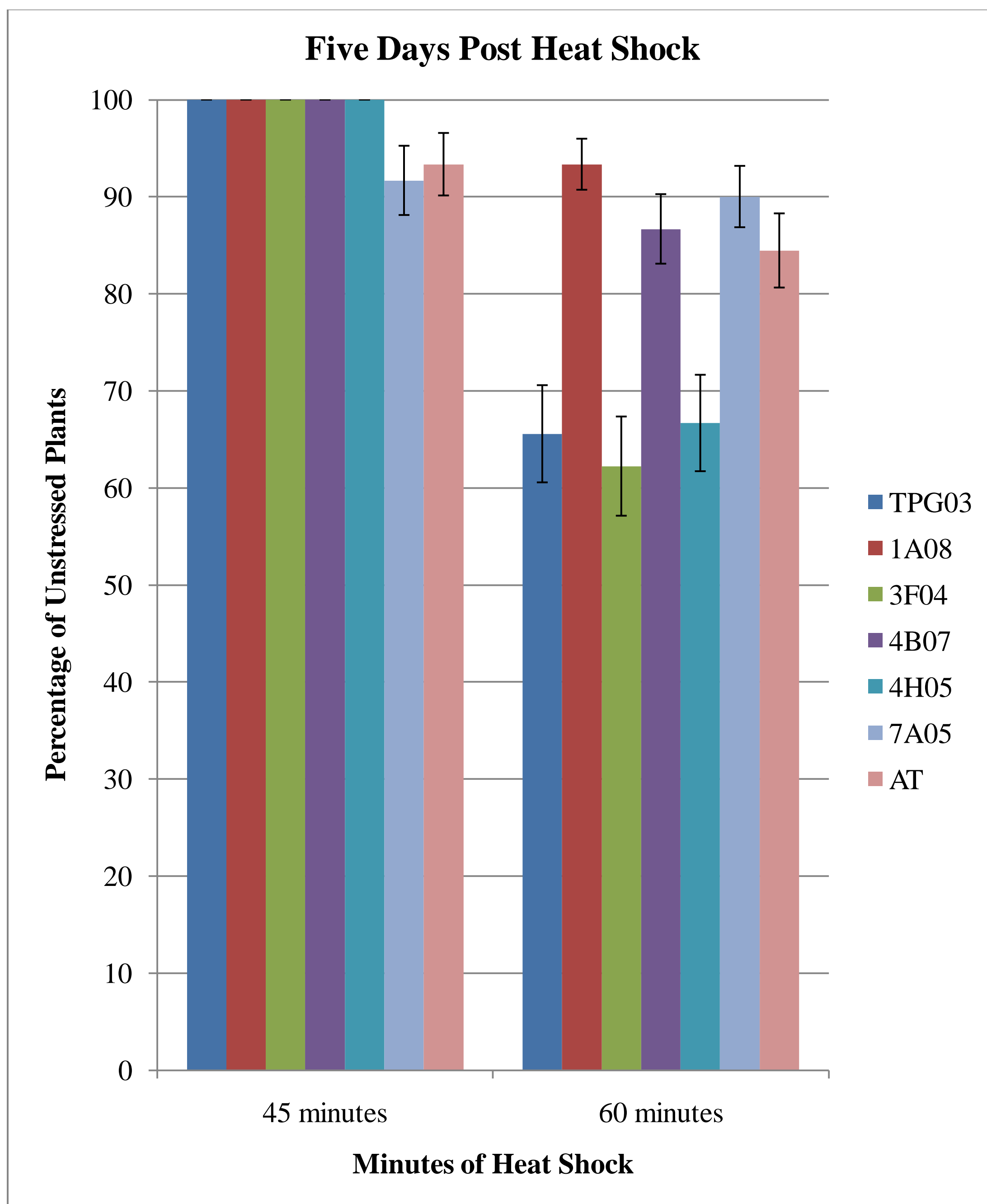


Figure 16. Results of the *in vitro* assay five days after heat shock (Nt, nontransgenic).

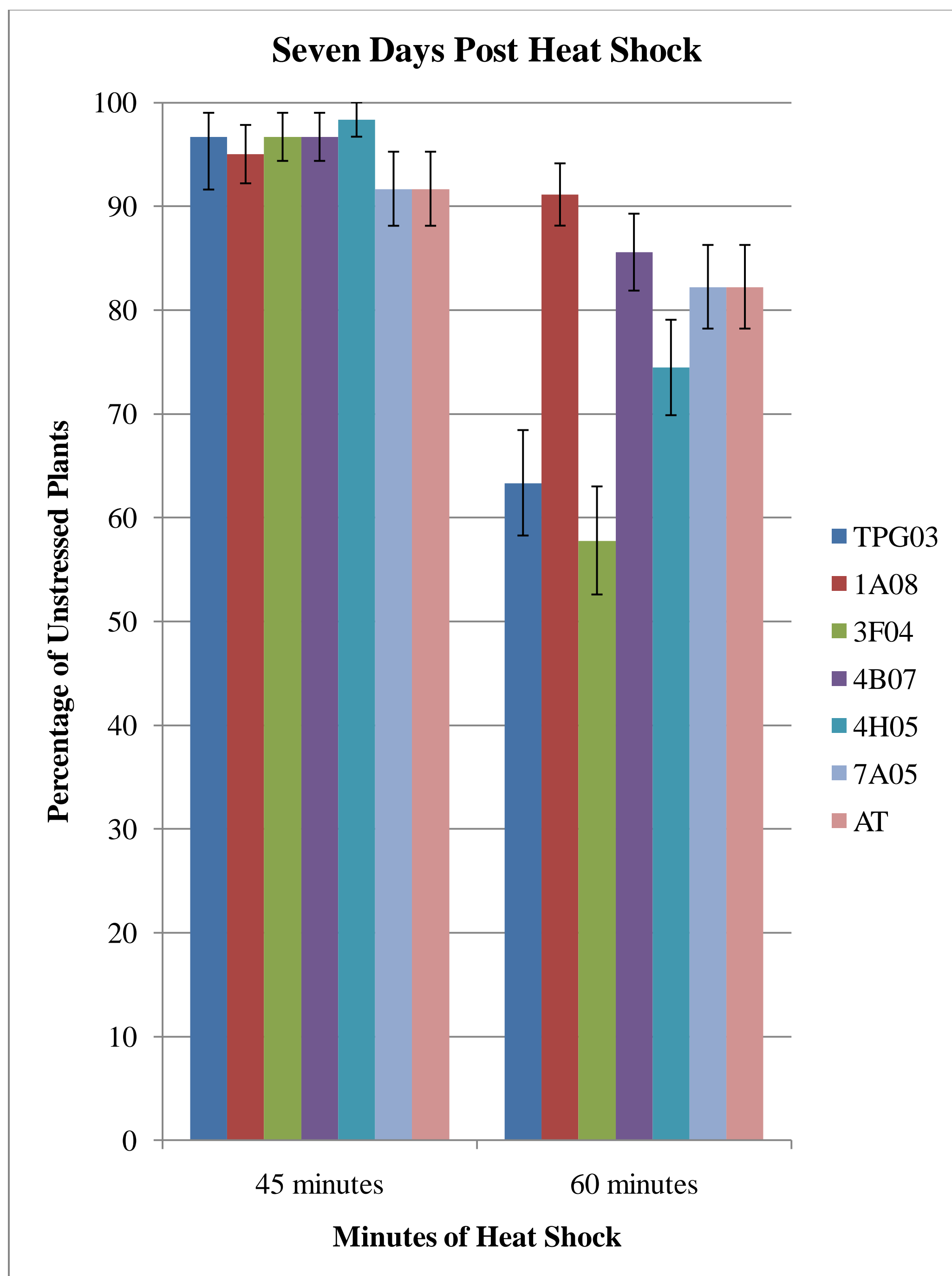


Figure 17. Results of the *in vitro* assay two days after heat shock (Nt, nontransgenic).

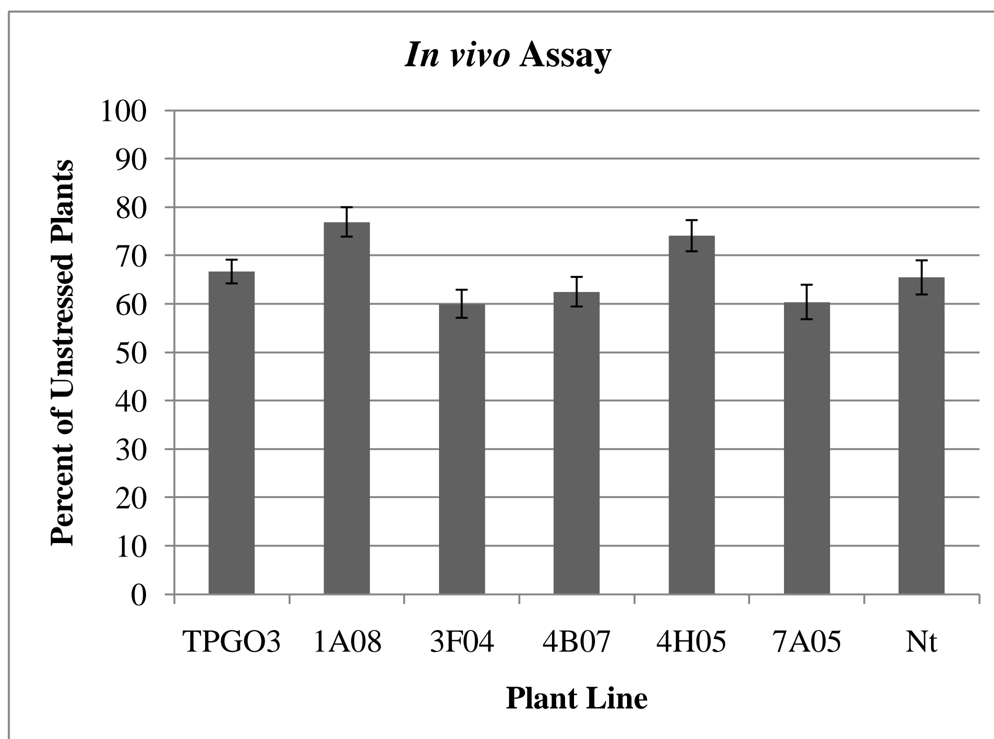


Figure 18. Results for the *in vivo* thermotolerance assay. Line 1A08 had the smallest percent of unstressed plants while 3F04 had the highest percent of stressed plants.

An analysis of variance (see Table 3) was used to compare the hypocotyl length between nontransgenic and transgenic lines that were not given a heat treatment or those that were given a heat treatment (see Figure 19). In the control group (no heat stress), transgenic lines had a longer average hypocotyl length compared to nontransgenic seedlings (see Figure 17), which was expected. In the no stressed seedlings, the average hypocotyl length was 10.4 mm while in the stressed seedlings the average hypocotyl length was 4.0 mm (see Figure 17).

Within the heat stressed group, several significant differences in hypocotyl length were observed. In four transgenic lines, the hypocotyl elongation was significantly shorter compared to the nontransgenic line, which had a length of 5.02 mm (SE = 0.22).

TPG03 ($p < 0.01$) had an average length of 4.2 mm (SE = 0.41), 3F04 ($p < 0.01$) had an average length of 3.5 mm (SE = 0.32), 4B07 ($p < 0.01$) had an average length of 4.72 mm (SE = 0.47), and 4H05 ($p < 0.01$) had an average length of 3.7 mm (SE = 0.40). No significant difference was found between the nontransgenic line and 1A08 or 7A05.

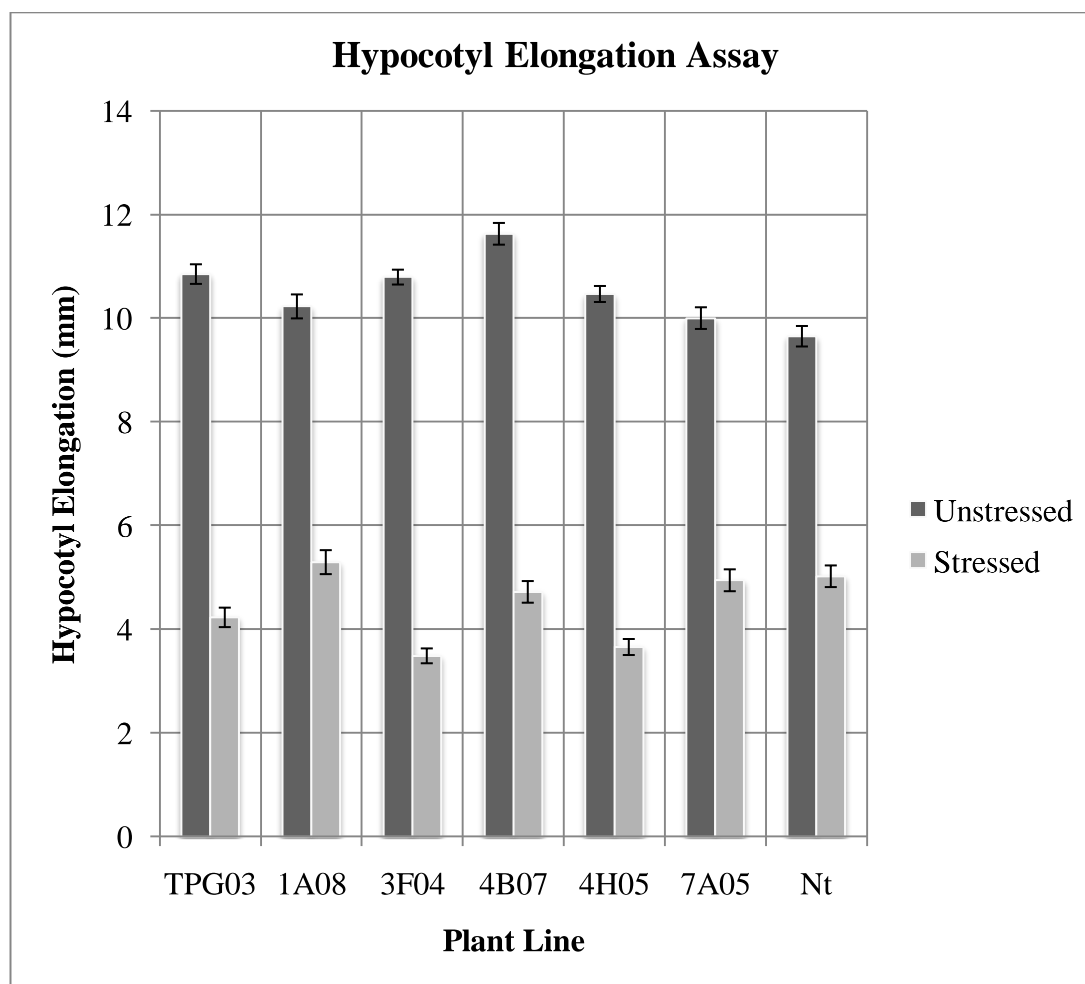


Figure 19. Comparison of hypocotyl elongation between transgenic and nontransgenic (Nt) plants that were either unstressed (dark gray) or given a heat treatment (light gray) 2.5 days after removal from 4°C.

Table 3

ANOVA Results for Hypocotyl Elongation Assay

Source	Type III Sum of Squares	df	Mean square error	F	*p*
Seed type	69.3	6	11.6	2.69	.014
Treatment	6603.7	1	6603.7	1540.1	< 0.01
Seed type * treatment	210.2	6	35.0	1540.1	< 0.01
Error	3074.4	5	717		

* Seeds were germinated in the dark and were either given a heat treatment or not given a heat treatment (control).

Discussion

Transgenic plants varied by their ability to tolerate each heat stress treatment. This could have been dependent on the function of the protein product of each transgene. Line TPG03 overexpressed a rubber elongation factor (REF), which has an unknown role in heat stress. It is known to function by binding small rubber particles and facilitates assimilation of isoprene units to polymerize larger (polyisoprene) molecules (Cornish, 2001). In our study, REF conferred thermotolerance at 45 minutes but not at 60 minutes of heat stress in five-day-old plants. This might have been due to an upper thermal limit for the function of REF; however, further research is needed to better characterize the role of REF in heat stress. Measurements of expression levels of the REF gene in *Hevea brasiliensis* (rubber tree) have been performed by Priya et al. (2006). In this study, REF was found to have high expression in the plant vascular tissue when expressed in tobacco. This gene was also overexpressed in response to heat stress in a study performed by Larkindale and Vierling (2008).

In this study, two different HSPs were overexpressed in *A. thaliana*--lines 1A08 and 3F04. Line 1A08 showed less signs of heat stress compared to 3F04 five and seven days after being given 45 minutes of heat treatment. While 3F04 had a significantly shorter hypocotyl compared to the nontransgenic line, 1A08 had a slightly longer hypocotyl compared to the nontransgenic line. The reason 1A08 showed less signs of heat stress in all stress experiments might have been due to functional differences in the two types of sHSP. For example, line 1A08 had a sHSP thought to be localized to the cytoplasm; whereas line 3F04 overexpressed a sHSP predicted to localize to the mitochondria.

Several nodulin proteins that play a role in nodule formation also function in establishing a relationship with root-associated organisms (Stougaard, 2000). Less well studied has been the role of nodulins in heat stress. In the 4B07 line used in this study, the protective mechanism of this protein after 45 minutes and 60 minutes of heat stress might have been due to enhancement of stress signaling; however, further research is needed In one study, Fu, Li, and Yu (2010) heat stressed control and transgenic *A. thaliana* lines overexpressing the nodulin-related protein 1 (*NRP1*) gene. They found transgenic lines had lower levels of *NRP1* expression after heat stress but levels increased after cold stress. In addition, there were decreased levels of abscisic acid compared to control lines, suggesting NRP1 might function in stress signal transduction.

When water is unavailable for evaporational cooling, the plant relies on mature cells that have established macrostructural heat stress protective mechanisms. These can include trichomes, well-developed cytoskeletal structures, and cognate heat shock proteins. In this study, 4H05 plants overexpressed a peroxidase, which has been

suggested to help in heat stress by reducing oxidative compounds such as H_2O_2. Based on the results, the increased thermotolerance in five-day-old plants in this study might be due to the activity of the peroxidase. Miao et al. (2006) investigated the effects of ROS damage on *A. thaliana* lines lacking glutathione peroxidase 3 (*ATGPX3*) and lines overexpressing *ATGPX3*. While knockout lines were more susceptible to H_2O_2 and drought stress, overexpressing lines showed increased levels of drought tolerance.

The enzyme overexpressed in line 7A05 was allene oxide cyclase (AOC). Upregulation was suspected to lead to an increase in jasmonic acid signaling and thus increase signaling that in turn would lead to expression of stress-related genes. Compared to control plants, line 7A05 showed higher thermotolerance two and five days after heat treatment with 60 minutes of heat treatment. However, on day 7, 705 and control lines showed the same percentage of plants displaying heat stress. This suggests that the benefits of AOC might be beneficial for only five days after heat stress. There was also no difference in hypocotyl length between 7A05 and control plants, which suggests similar thermotolerance in these seedlings.

In the *in vitro* study, five of the six transgenic lines showed significant levels of thermotolerance compared to control plants, supporting the hypothesis that transgenic lines would be less affected by heat stress. An interesting finding was that control plants showed a decrease in heat stress from day 3 to day 4. Two transgenic lines--3F04 and 4H05 also showed this same phenomenon. This is difficult to explain but might have been due to recovery and repair mechanisms within the plant.

The second hypothesis, that transgenic plants would have longer hypocotyls compared to control plants, was not supported. Previous studies indicated that elongating

seedlings exposed to heat stress damages crucial membrane networks within the developing seedlings (Hong & Vierling, 2000). Heat stress-associated genes are thought to protect the membrane system and thus allow for continuous elongation (Hong & Vierling, 2000). One explanation for this result might be related to transgenic lines producing less indole-3-acetic acid (IAA), an important plant stress hormone. In a study by Gray, Östin, Sandberg, Romano, and Estelle (1998), nine-day-old seedlings had higher levels of auxin and also had longer hypocotyls than transgenic plants. This suggests transgenic plants overexpressing heat stress-related genes had shorter hypocotyls because less IAA was produced.

Taken together, the results in these experiments indicated that some of the transgenic lines were more thermotolerant than control plants depending on the heat treatment. However, determining which plant lines were more thermotolerant depends on the assay. For example, the *in vitro* assay suggested that 1A08, 4B07 and 7A08 lines were more thermotolerant than control plants after heat treatment of 45 and 60 minutes. The *in vivo* experiment showed two lines, 1A08 and 4H05, had less heat stress damage five days after heat treatment. Results from the hypocotyl elongation assay showed that 1A08 had the greatest hypocotyl elongation. Line 1A08 was the only plant line that consistently showed less heat stress in all three assays and, therefore, might be the best candidate to either investigate further or transform into crop plants. Future studies should also focus on using *in vivo* conditions, since this is more closely in line with what plants might be exposed to in the field.

CHAPTER V

CONCLUSIONS AND FUTURE PERSPECTIVES

The purpose of this project was to express *C. officinalis* heat stress-associated genes in *A. thaliana* and assess transgenic lines for heat tolerance. It is important to recapitulate the major conclusions from the *C. officinalis* library, transformation, and heat stress experiments along with their implications. In addition to summarizing the major findings from this research, the final chapter of this dissertation discusses several other facets. First, discussion of the benefits and drawbacks to constructing a binary library for downstream applications is discussed. Other methods to determine thermotolerance are also discussed. Potential pitfalls with heterologous expression and methods to analyze transcription of a transgene and translational production of a protein are also given. The chapter concludes with future research that might use other constructs from the *C. officinalis* library to generate plants with increased thermotolerance.

Summary of the *C. officinalis* Expressed Sequence Tag Library

In this project, cDNA that was reverse transcribed from total mRNA was cloned into a plant binary vector and transformed into *E. coli*, which is similar to the process of cDNA library construction. However, following the sequencing reactions, vector-insert constructs were obtained on the 96-well sequencing plates. This yielded a binary library in which more than 1,000 different constructs were available for plant transformation

studies. Such an approach was beneficial in that it bypassed the need to isolate and amplify a gene of interest and avoided independent ligation reactions. Several genes from *C. officinalis* were then transformed into *A. thaliana* to generate multiple transgenic lines; whereas traditional methods of transformation are often limited to a single gene.

Since the genes used in this study had not been previously characterized, this was an effective approach to cloning novel genes. Often, when attempting to isolate genes without nucleotide sequence data, degenerate primers can be used. However, this approach can be quite cumbersome and does not always lead to successful amplification of a gene of interest.

Although multiple transgenic lines were generated, some drawbacks to the binary library approach became apparent. First, the inserts generated from the mRNA were not full length; thus, not all genes identified in the library were able to be used for plant transformation. When performing first strand synthesis for library construction, reverse transcriptase might not have reverse transcribed the entire length of the mRNA, which led to several instances of a truncated insert. Therefore, additional screening of the candidate genes had to be performed by looking at the sequence data from the library.

Some inserts contained an AUG initiation codon and, when compared to BLASTx, the full length of the protein was suggested to be present. This suggested the binary library approach might facilitate not only gene characterization but also could be an important step in facilitating plant biotechnology experiments.

***C. officinalis* Library Compared to Other Libraries**

The number of ESTs in a library varies. For example, Natarajan et al. (2010) analyzed 12,084 ESTs generated from the developing seeds of jatropha (*Jatropha curcas*)

to better understand oil production. In contrast Lindqvist et al. (2006) sequenced only 942 ESTs from a cosmopolitan mint in Hawaii known as Maohiohi (*Stenogyne rugosa*). In this case, so few sequence reads were needed because they were only interested in ESTs from immature fruits of the mint plant. These were sufficient data to show associations with a wide range of Gene Ontology (GO) terms, perhaps suggesting fruit development is a complex process. These experiments demonstrate that while only a few hundred or a few thousand sequences can be included in EST libraries, useful information can still be gained. Similarly, results from this dissertation suggest that the C. officinalis library in this study were useful to gain insight into physiological responses to heat stress.

While EST library analysis can be used to identify novel genes, it can also help characterize transcripts expressed in tissues (e.g., fruits or leaves), at developmental times, or that might give rise to distinct traits (e.g., weediness). To understand weediness, Broz et al. (2007) analyzed 4,969 ESTs generated from spotted knapweed (*Centaurea maculosa*). EST libraries can also help determine the transcriptional state of the organism to help identify which genes are upregulated. For example, Lee et al. (2004) analyzed 8,525 ESTs from a leaves pepper plant (*Capsicum annum*) after infection with *Xanthomonsa axonopodis*, a causative agent of soybean pustules. They suggested 136 of the transcripts were involved in pathogen response and might be useful for future research to protect plants from *X. axonopodis*. These examples help demonstrate the diversity of purposes for generating EST data. Arguably, in each of these examples, more sequence data would have resulted in a better characterization of the transcriptional state and identification of more genes. The EST library in this project

consisted of 912 sequences and provided enough data to determine that *C. officinalis* was expressing several genes associated with heat stress.

Summary of the Transformation Experiments

The selection of genes from the binary library was an important consideration in generating thermotolerant plant lines. Two genes, one encoding a rubber elongation factor (TPG03) and another encoding a nodulin-like protein (4B07), were overexpressed in *A. thaliana* plants. These two genes were used to investigate their possible role in thermotolerance. All plant lines showed expression of the *nptII* gene and all but one line (3F04) did not show expression of the *C. officinalis* transgene.

In this study, six different transgenic lines each with a unique *C. officinalis* gene were generated. Generating so many different lines is atypical in plant biotechnology experiments. Many plant transgenic studies express a single gene in several different plant lines, which limits the diversity of genes expressed in the transgenic plants. Since the present study used a binary vector in library construction, several lines of *A. thaliana* expressing different genes were generated. The binary library approach bypassed several complicated molecular biology techniques; this reduced the time and expenses required to express *C. officinalis* genes into *A. thaliana*. However, the binary library approach was limited to using only full-length genes present in the library. In addition, although it allowed for generating several different transgenic plants, the work became somewhat difficult due to limited lab space.

Summary of Heat Stress Assays

Results from the *in vitro* experiments suggested a few transgenic lines were significantly different than control plants. Less bleaching of leaves on some transgenic

lines was observed but this difference was only between 10 or fewer plants. Although no differences were seen from treatments at 45˚C for 30 minutes, other researchers found 30 minutes was enough time for heat stress to become evident (Clarke et al., 2004). In this experiment, 45 minutes of heat treatment resulted in differences between control and transgenic plants. At 60 minutes, a greater difference was observed.

The hypocotyl elongation assay showed several differences between control and transgenic plants. Unexpectedly, the hypocotyls of several transgenic lines were much shorter than control plants; the opposite case was expected. It was hypothesized that this might be partly due to decreased stress signaling. Levels of auxin hormones are increased when plants are under stress, which causes hypocotyl elongation; however, some researchers reported transgenic lines to have shorter hypocotyls (Gray et al., 1998). Indeed, the results from the hypocotyl elongation assay in this study suggested there might have been a decrease in auxin levels.

Based on the results from the *in vivo* heat stress assays, control plants and transgenic *A. thaliana* were not significantly different. This could have been due to several reasons; however, it is likely that a single gene might not have been effective in conferring thermotolerance to an entire plant even though the gene was constitutively expressed. Although some researchers have shown one gene can confer increased heat tolerance, this did not seem to be the case for the *in vivo* experiments performed for the genes in this study.

Based on the three heat stress assays, some lines showed greater levels of thermotolerance compared to non-transgenic plants depending on the heat treatment. The results suggested that 1A08, which overexpressed a small heat shock protein, might be of

interest to pursue further because 1A08 showed longer hypocotyls and less leaf damage compared to the nontransgenic plants.

Future Studies

Currently, there is interest in understanding how plants respond to stress. A major goal of plant stress biology is to characterize the master regulators. Such regulators are thought to be transcription factors that upregulate several other stress-associated genes involved in the stress response. For example, Mishra et al. (2002) generated several transgenic tomato plants that either overexpressed or suppressed various heat shock factors. They suggested that an important heat shock factor (HsfA1) acted as a master regulator in coordinating the heat stress response in tomato by inducing expression of other transcription factors and heat shock proteins.

The first step in incorporating transcription factors in future transgenic experiments was to identify vector-insert constructs in the *C. officinalis* binary library that encode for stress-associated transcription factors. Based on the BLASTx and Gene Ontology (GO) results, more than 20 sequences from the library were associated with transcriptional regulation. Three members of the WRKY family of transcription factors were identified, which are named after their heptapeptide DNA-binding domain (WRKYGQK) at the N-terminus (Eulgem, Rushton, Robatzek, & Somssich, 2000). A majority of the WRKY members are involved in several different abiotic stresses (Guo et al., 2010) and some of these might be master regulators for upregulating several stress-associated genes. In addition to WRKY members, other transcription factors (i.e., MADS domain-containing transcription factor, LHY transcription factor, multiple stress-responsive zinc-finger protein) are of interest. At least 15 transcripts with the GO terms

transcription factor activity were found. Future experiments should focus on generating transgenic plants overexpressing these transcriptional regulators to help determine the master control mechanisms for inducing a genome-wide heat stress response.

Measuring Heat Stress

Different methods to measure heat stress can also be applied to future studies; these include an electrolyte leakage assay and a chlorophyll fluorescence assay. Since membranes are a primary site where heat stress is initially detected by the plant, the electrolyte leakage assay has been a hallmark of research investigating the effects of heat on membrane destabilization (Martineau, Williams, & Sullivan, 1978). The procedure relies on using conductivity measurements to estimate the amount of ions leaked from a plant cell after inducing heat stress. Kang et al. (2011) measured electrolyte leakage in 10-day-old transgenic *A. thaliana* that overexpressed a transcription factor (AP2) that was previously suggested to play a role in freezing tolerance. Plants were heat stressed at 45˚C for one hour; the transgenic plants were estimated to have lost only 36% of their ions while nontransgenic lines lost 80% of their ions 150 minutes after the heat stress. Nearly all intracellular ions were lost from the plant cells after 3.5 hours of heat treatment in control and transgenic lines. Comparatively, Li, Zhou, Chen, and Yu (2010) measured leakage from mutant *A. thaliana* plants overexpressing *wrky39* and knockout mutants lacking *wrky39*. Leaves and stems from three-week-old plants were collected, weighed, and placed in a test tube containing 5 mL of water. The plants were given a heat treatment of 42˚C while conductivity measurements of the water were recorded and resulted in *wrky39* overexpressing lines showing reduced leakage compared to knockout

mutants and nontransgenic controls. These data strongly suggest that the electrolyte leakage assay can be useful in quantifying heat stress damage.

Another interesting and reliable assay to measure stress in plants is to measure the amount of chlorophyll fluorescence in stressed leaves (Lichtenthaler, 1987; Willits & Peet, 2001). This is done by growing plants until the desired developmental stage is reached and then subjecting the plants to heat stress. An equal amount of biomass is subsequently harvested from control (wild type) and transgenic plants followed with extraction of chlorophyll using acetone. The absorbance is measured using a standard spectrophotometer. Plants with reduced chlorophyll content are also typified by displaying etiolated leaves, a decrease in turgor, and a delay in production of new leaves.

While heat tolerant plants are expected to contain more chlorophyll, this is not the case in every experiment that has attempted to overexpress thermotolerance genes into *A. thaliana*. Li, Fu, Chen, Huang, and Yu (2011) generated *A. thaliana* plants constitutively expressing different *wrky* genes and measured the chlorophyll content of 25-day-old plants after heat stress for six hours at 48˚C. They found that transgenic plants overexpressing *wrky25* and *wrky26* had less chlorophyll content than wild type plants. However, *A. thaliana* plants overexpressing *wrky33* showed slightly higher levels of chlorophyll. In another experiment using a less invasive technique, Saidi et al. (2009) measured photosynthetic electron quantum yield content by measuring the Fv/Fm (variable and maximum fluorescence of chlorophyll, respectively) ratio to estimate damage from heat stress in the moss, *Physcomitrella patens*. Their primary interest was to help determine how plants sense changes in ambient temperatures. After subjecting plants to heat stress for one hour at 38˚C fluorescence, measurements over 10 minute

intervals over a time course of one hour gradually decreased, suggesting a loss of thermotolerance. These studies provide examples of techniques useful to assay for thermotolerance that future research using the *C. officinalis* library should incorporate.

Transcript Expression and Expression Levels

A variety of methods can be employed when assaying for gene expression: RT-PCR, northern blot, and quantitative real time PCR (qRT-PCR). Although more sensitive assays for detecting transcripts exist, northern blot continues to be used to estimate size and relative abundance of transcripts. Dafny-Yelin, Tzfira, Vainstein, and Adam (2008) used northern blot to determine transcripts for sHSPs in *A. thaliana*. Flowers were more abundant than plant leaves after exposure to heat stress.

More frequently, researchers rely on PCR in place of northern blot due to the sensitivity and ability to work with cDNA instead of RNA. Other experiments have included both northern blot and PCR to further confirm gene expression. For example, while Dafny-Yelin et al. (2008) used northern blot to show relative abundance of sHSP transcripts, they also used PCR to show different classes of sHSPs were expressed in both flowers and leaves of *A. thaliana*.

qRT-PCR relies on using a reporter dye that can detect formation of double stranded DNA as it is amplified under PCR conditions. Montero-Barrientos et al. (2010) generated transgenic *A. thaliana* plants over-expressing *T. harzianum* HSP70. qRT-PCR of transgenic plants showed a down regulation of other stress-related HSP genes (HSP101, HSP90, HSP18) and a heat shock transcription factor. Ascorbate peroxidase (*APX1*) and superoxide dismutase (*SOS1*) showed a lower expression under non-heat stress conditions but were upregulated after heat treatment. Overall, the qRT-PCR results

suggested that *HSP70* could be used to help confer heat and other abiotic stress tolerance in plants. Arguably, there is much potential for using qRT-PCR to analyze transcript levels in *C. officinalis* genes expressed in transformed plants.

Transcriptional Influences on Transgene Expression

Generating multiple transgenic lines might also be an important step in developing heat tolerant plants. When *A. tumefaciens* delivers the transgene, its location is not site-specific within the nucleus. The transgene might be incorporated into an area of the genome with either high or low transcriptional activity. Each line is derived from individual transformation events acquired from the initial infection with *A. tumefaciens*. Upon screening of T1 seeds for transgenic plants, several of the transgenic plants, i.e., if grown in the presence of kanamycin, are rescued, transferred to soil, and allowed to mature.

In this study, the expression of six transgenes was confirmed by PCR. Although a transgene might be transcribed, it might not result in translation. This might be due to several reasons. A well-studied phenomenon is codon usage bias. This phenomenon occurs because amino acids are encoded by multiple codons, which can be present in unequal amounts in transcripts from two different species. While one species might rely on a particular charged tRNA corresponding to a particular codon, another species might have this codon in lesser amount and therefore less of the necessary tRNA. Codon usage bias has been studied in both monocots and dicots. Although some have suggested the differences are subtle (Kawabe & Miyashita, 2003), more recent findings suggest that GC content, especially at the third position, is higher in rice compared to *A. thaliana* (Wang & Hickey, 2007).

Although variations in codon usage in Fabaceae and Brassicaceae have not been systematically compared in domains or kingdoms, several differences in codon usage became apparent. For this reason, expression of animal genes is often engineered to contain codons that are more frequently used in plants (Geyer et al., 2010). Altering codons within the transcript can help ensure inclusion of the correct amino acid while considering the charged tRNAs available carrying the required amino acid and thus enhancing translational efficiency.

An important consideration in heterologous gene expression in plants is that mRNA stability is highly dependent upon the 5' and 3' untranslated regions (UTRs). It has been observed that although the AUG initiation site in a transcript is necessary for translation, there are rare cases in which AUG is not required for initiating translation. Depeiges, Degroote, Espagnol, and Picard (2006) showed that replacement of the AUG start site with AUC, GUG, ACG or CUG in *A. thaliana* led to transcription of a phosphinothricin gene, albeit at 5-10% levels compared to when AUG was the initiation codon. However, this instance was a rare exception; nearly all genes require an AUG start site for translation initiation. The Kozak sequence--which consists of the four bases before the initiation codon, the initiation codon, and the two bases after the initiation codon--also influences translation. In plants, the Kozak consensus sequence is 5'-AACAAUGGC-3' (Lütcke, 1987). In addition, hairpin loops and the internal ribosome entry site present in the 5'UTR contribute to the mRNA secondary structure and, thus, influence translation (Pesole et al., 2001).

Detecting Translation

Various proteomics techniques have been developed to detect the presence of a protein or determine differential expression patterns. Neilson, Gammulla, Mirzaei, Imin, and Haynes (2010) suggested we should apply 1-D or 2-D electrophoresis and other proteomic techniques in heat stressed cereal crops such as barley, rice and wheat to help understand how the proteins accumulate and behave. To detect a protein, total protein is separated by size using gel electrophoresis. After transferring the protein bands to a positively charged membrane, the next step requires hybridization using a protein-specific antibody. The subsequent antibody-protein complex is detected using a fluorescently- or radioactively-labeled secondary antibody. Based on these techniques, protein assays can be performed to determine if translation of the transgene has occurred.

**Constructs from the *C. officinalis*
Library in Other Plants**

An interesting feature of the library constructed in this project was that any vector-insert had the potential to be used in transforming plants other than *A. thaliana*. For example, tissue culture relied on *Agrobacterium*-mediated transformation; thus, nearly any plant that was cultured *in vitro* could be transformed. Transformation of model monocot plants (e.g., *Brachypodium distachyon*) might be of great interest since a great majority of crop plants are monocots.

The larger goal of this project was to identify heat stress-related genes derived from *C. officinalis* and overexpress these genes in *A. thaliana*. Transgenic lines showing thermotolerance when a particular gene was overexpressed suggested that this gene might be a good candidate to use in crop plants to lessen the predicted affects of heat stress in reducing grain output and thus avoid hunger in a growing world population. However, it

could be that genetically modified crops are not needed and that food access is the problem, not a food shortage.

The aim of this project was to identify heat stress-associated genes and express these in the model plant, *A. thaliana*. It should be noted that while some genes can be upregulated and show promise for conferring thermotolerance in *A. thaliana*, real-world applications such as using these genes in crop plants might not confer thermotolerance.

A field setting is markedly different than a lab setting. For example, there is a complex interaction between soil microbes and plant root systems. Temperatures also fluctuate more wildly than the temperatures used in these experiments. Drought or abiotic factors might also play more of a role in heat stress in some situations. Therefore, the results from these experiments only represented a model of what one might observe in the field.

In a recent review by Mittler and Blumwald (2010), they suggested using genes whose protein products were adapted to function in desert or saline environments. In addition to desert plants, attention should also be given to tropical plants for heat stress-associated genes. The work presented here helped show that tropical plants could also have several heat stress-associated genes that might be useful for expression in transgenic plants. Determining which of these genes to overexpress and implementing the various heat stress assays is perhaps a large goal that might be better carried out by multiple investigators.

Ainsworth and Ort (2010) suggested that climate change could bring larger impacts than most realize. For example, while corn needs to be grown at more northern latitudes, these soils are poorer in nutrients than their prairie counterparts, which might

lead to lower yields. Thus, generating heat tolerant crop plants should be a major goal of plant breeding and plant genetic engineering. Although plant breeding and the green revolution has provided significant gains in food production, for people restricted to living in climate change-prone regions, a second surge in food production might not be possible. Determining genes that provide thermotolerance in model species and expressing these in crop plants might lend support to sustaining a global world food supply.

REFERENCES

Adamska, I., Kloppstech, K., & Ohad, I. (1992). UV light stress induces the synthesis of the early light-inducible protein and prevents its degradation. *Journal of Biological Chemistry*, *267*(34), 24732.

Ahuja, I., de Vos, R. C. H., Bones, A. M., & Hall, R. D. (2010). Plant molecular stress responses face climate change. *Trends in Plant Science, 15*(12), 664-674.

Ainsworth, E. A., & Ort, D. R. (2010). How do we improve crop production in a warming world? *Plant Physiology*, *154*(2), 526 -530.

Åkerfelt, M., Morimoto, R. I., & Sistonen, L. (2010). Heat shock factors: Integrators of cell stress, development, and lifespan. *Nature Reviews Molecular Cell Biology*, *11*(8), 545-555.

Alia, Hayashi, H., Sakamoto, A., & Murata, N. (1998). Enhancement of the tolerance of *Arabidopsis* to high temperatures by genetic engineering of the synthesis of glycinebetaine. *The Plant Journal: For Cell and Molecular Biology*, *16*(2), 155-161.

Allakhverdiev, S. I., Kreslavski, V. D., Klimov, V. V., Los, D. A., Carpentier, R., & Mohanty, P. (2008). Heat stress: an overview of molecular responses in photosynthesis. *Photosynthesis Research*, *98*(1), 541–550.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410.

Anderson, J. V., Horvath, D. P., Chao, W. S., Foley, M. E., Hernandez, A. G., Thimmapuram, J., …Mikel, M. A. (2007). Characterization of an EST database for the perennial weed leafy spurge: An important resource for weed biology research. *Weed Science*, *55*(3) 193–203.

Arrhenius, S. P., & Langenheim, J. H. (1983). Inhibitory effects of Hymenaea and Copaifera leaf resins on the leaf fungus, *Pestalotia subcuticularis*. *Biochemical Systematics and Ecology*, *11*(4), 361–366.

Asada, K. (2006). Production and scavenging of reactive oxygen species in chloroplasts and their functions. *Plant Physiology*, *141*(2), 391-396.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., …Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25.

Ashraf, M., & Foolad, M. R. (2007). Roles of glycine betaine and proline in improving plant abiotic stress resistance. *Environmental and Experimental Botany*, *59*(2), 206-216.

Baniwal, S. K., Bharti, K., Chan, K. Y., Fauth, M., Ganguli, A., Kotak, S., … von Koskull-Döring, P. (2004). Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. *Journal of Biosciences*, *29*(4), 471-487.

Battisti, D., & Naylor, R. (2009). Historical warnings of future food insecurity with unprecedented seasonal heat. *Science*, *323*(5911), 240.

Bevan, M. W., & Chilton, M. (1982). T-DNA of the *Agrobacterium* Ti and Ri plasmids. *Annual Review of Genetics*, *16*(1), 357–384.

Bhatnagar-Mathur, P., Vadez, V., & Sharma, K. K. (2008). Transgenic approaches for abiotic stress tolerance in plants: Retrospect and prospects. *Plant Cell Reports*, *27*(3), 411–424.

Binelli, G., & Mascarenhas, J. P. (1990). *Arabidopsis*: Sensitivity of growth to high temperature. *Developmental Genetics*, *11*(4), 294-298.

Blüthgen, N., Brand, K., Čajavec, B., Swat, M., Herzel, H., & Beule, D. (2008). Biological profiling of gene groups utilizing gene ontology. Retrieved from http://arxiv.org/abs/q-bio/0407034

Boguski, M. S., Lowe, T. M., & Tolstoshev, C. M. (1993). dbEST--database for "expressed sequence tags." *Nature Genetics*, *4*(4), 332-333.

Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., & Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *The Plant Cell Online*, *12*(12), 2383-2394.

Bowler, C., & Fluhr, R. (2000). The role of calcium and activated oxygens as signals for controlling cross-tolerance. *Trends in Plant Science*, *5*(6), 241-246.

Boyko, A., & Kovalchuk, I. (2011). Genome instability and epigenetic modification–heritable responses to environmental stress? *Current Opinion in Plant Biology*, *14*(3), 260-266.

Broz, A. K, Broeckling, C. D, He, J., Dai, X., Zhao, P. X, & Vivanco, J. M. (2007). A first step in understanding an invasive weed through its genes: An EST analysis of invasive *Centaurea maculosa*. *BMC Plant Biology*, *7*(1), 25-35.

Bruneau, A., Forest, F., Herendeen, P. S., Klitgaard, B. B., & Lewis, G. P. (2001). Phylogenetic relationships in the Caesalpinioideae (Leguminosae) as inferred from chloroplast trnL intron sequences. *Systematic Botany*, *26*(3), 487–514.

Carlson, J. E., Leebens-Mack, J. H, Wall, P. K., Zahn, L. M., Mueller, L. A., Landherr, L. L., … Depamphilis, C. W. (2006). EST database for early flower development in California poppy (*Eschscholzia californica* Cham., Papaveraceae) tags over 6000 genes from a basal eudicot. *Plant Molecular Biology*, *62*(3), 351–369.

Cenik, C., Chua, H. N., Zhang, H., Tarnawsky, S. P., Akef, A., Derti, A., …Palazzo, A. F. (2011). Genome analysis reveals interplay between 5′ UTR introns and nuclear mRNA export for secretory and mitochondrial genes. *PLoS Genetics*, *7*(4), e1001366.

Chen, F., Al-Ahmad, H., Joyce, B., Zhao, N., Köllner, T. G., Degenhardt, J., & Stewart, C. N., Jr. (2009). Within-plant distribution and emission of sesquiterpenes from *Copaifera officinalis*. *Plant Physiology and Biochemistry*, *47*(11-12), 1017-1023.

Chinnusamy, V., & Zhu, J. K. (2009). Epigenetic regulation of stress responses in plants. *Current Opinion in Plant Biology*, *12*(2), 133–139.

Chow, K. S., Wan, K. L., Isa, M., Bahari, A, Tan, S-H, Harikrishna, K, & Yeang, H-Y. (2007). Insights into rubber biosynthesis from transcriptome analysis of *Hevea brasiliensis* latex. *Journal of Experimental Botany*, *58*(10), 2429-2440.

Christie, P. J., & Cascales, E. (2005). Structural and dynamic properties of bacterial type IV secretion systems. *Molecular Membrane Biology*, *22*(1-2), 51-61.

Clarke, S. M., Mur, L. A. J., Wood, J. E., & Scott, I. M. (2004). Salicylic acid dependent signaling promotes basal thermotolerance but is not essential for acquired thermotolerance in *Arabidopsis thaliana*. *The Plant Journal*, *38*(3), 432-447.

Clough, S. J, & Bent, A. F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal: For Cell and Molecular Biology*, *16*(6), 735-743.

Cornish, K. (2001). Similarities and differences in rubber biochemistry among plant species. *Phytochemistry*, *57*(7), 1123–1134.

Dafny-Yelin, M., Tzfira, T., Vainstein, A., & Adam, Z. (2008). Non-redundant functions of sHSP-CIs in acquired thermotolerance and their role in early seed development in *Arabidopsis*. *Plant Molecular Biology*, *67*(4), 363-373.

da Silva Medeiros, R., & Vieira, G. (2008). Sustainability of extraction and production of copaiba (*Copaifera multijuga* Hayne) oleoresin in Manaus, AM, Brazil. *Forest Ecology and Management*, *256*(3), 282–288.

Daugaard, M., Rohde, M., & J\ä\ättel\ä, M. (2007). The heat shock protein 70 family: highly homologous proteins with overlapping and distinct functions. *FEBS Letters*, *581*(19), 3702–3710.

Dennis, M. S., & Light, D. R. (1989). Rubber elongation factor from *Hevea brasiliensis*: Identification, characterization, and role in rubber biosynthesis. *Journal of Biological Chemistry*, *264*(31), 18608-18617.

Depeiges, A., Degroote, F., Espagnol, M. C., & Picard, G. (2006). Translation initiation by non-AUG codons in *Arabidopsis thaliana* transgenic plants. *Plant Cell Reports*, *25*(1), 55-61.

Desfeux, C., Clough, S. J., & Bent, A. F. (2000). Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the *Arabidopsis* floral-dip method. *Plant Physiology*, *123*(3), 895 -904.

Dreher, K., & Callis, J. (2007). Ubiquitin, hormones and biotic stress in plants. *Annals of Botany*, *99*(5), 787-822.

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

Feibert, E. B., & Langenheim, J. H. (1988). Leaf resin variation in *Copaifera langsdorfii*: Relation to irradiance and herbivory. *Phytochemistry*, *27*(8), 2527–2532.

Frank, G., Pressman, E., Ophir, R., Althan, L., Shaked, R., Freedman, M., …Firon, N. (2009). Transcriptional profiling of maturing tomato (*Solanum lycopersicum* L.) microspores reveals the involvement of heat shock proteins, ROS scavengers, hormones, and sugars in the heat stress response. *Journal of Experimental Botany*, *60*(13), 3891-3908.

Fu, Q., Li, S., & Yu, D. (2010). Identification of an Arabidopsis Nodulin-related protein in heat stress. *Molecules and Cells*, *29*(1), 77–84.

Geiler-Samerotte, K. A., Dion, M. F., Budnik, B. A., Wang, S. M., Hartl, D. L., & Drummond, D. A. (2011). Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(2), 680-685.

Geyer, B. C., Kannan, L., Cherni, I., Woods, R. R., Soreq, H., & Mor, T. S. (2010). Transgenic plants as a source for the bioscavenging enzyme, human butyrylcholinesterase. *Plant Biotechnology Journal*, *8*(8), 873–886.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., & Robles, M. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420-3435.

Gollhofer, J., Schläwicke, C., Jungnick, N., Schmidt, W., & Buckhout, T. J. (2011). Members of a small family of nodulin-like genes are regulated under iron deficiency in roots of *Arabidopsis thaliana*. *Plant Physiology and Biochemistry*, *49*(5), 557-564.

Gong, M., van der Luit, A. H., Knight, M. R., & Trewavas, A. J. (1998). Heat-shock-induced changes in intracellular Ca^{2+} level in tobacco seedlings in relation to thermotolerance. *Plant Physiology*, *116*(1), 429-437.

Goyvaerts, E., Dennis, M., Light, D., & Chua, N. H. (1991). Cloning and sequencing of the cDNA encoding the rubber elongation factor of *Hevea brasiliensis*. *Plant Physiology*, *97*(1), 317-321.

Gramosa, N. V., & Silveira, E. R. (2005). Volatile constituents of *Copaifera langsdorffii* from the Brazilian northeast. *The Journal of Essential Oil Research*, *17*(2), 130–132.

Gray, W. M., Östin, A., Sandberg, G., Romano, C. P., & Estelle, M. (1998). High temperature promotes auxin-mediated hypocotyl elongation in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(12), 7197.

Grefen, C., & Blatt, M. R. (2008). SNAREs--molecular governors in signaling and development. *Current Opinion in Plant Biology*, *11*(6), 600-609.

Grigorova, B., Vaseva, I., Demirevska, K., & Feller, U. (2011). Combined drought and heat stress in wheat: Changes in some heat shock proteins. *Biologia Plantarum*, *55*(1), 105–111.

Guo, R., Yu, F., Gao, Z., An, H., Cao, X., & Guo, X. (2011). GhWRKY3, a novel cotton (*Gossypium hirsutum* L.) WRKY gene, is involved in diverse stress responses. *Molecular Biology Reports*, *38*(1), 49-58.

Halford, N. G. (2009). New insights on the effects of heat stress on crops. *Journal of Experimental Botany, 60*(15), 4215.

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., …White, R. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, *32*(Database issue), D258-261.

Hashida, S.-N., Uchiyama, T., Martin, C., Kishima, Y., Sano, Y., & Mikami, T. (2006). The temperature-dependent change in methylation of the Antirrhinum transposon Tam3 is controlled by the activity of its transposase. *The Plant Cell*, *18*(1), 104-118.

Hatey, F., Tosser-Klopp, G., Clouscard-Martinato, C., Mulsant, P., & Gasser, F. (1998). Expressed sequence tags for genes: A review. *Genetics, Selection, Evolution*, *30*(6), 521-541.

Hessling, M., Richter, K., & Buchner, J. (2009). Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. *Nature Structural & Molecular Biology*, *16*(3), 287–293.

Hey, S. J., Byrne, E., & Halford, N. G. (2010). The interface between metabolic and stress signaling. *Annals of Botany*, *105*(2), 197-203.

Hirayama, T., & Shinozaki, K. (2010). Research on plant abiotic stress responses in the post-genome era: Past, present and future. *The Plant Journal*, *61*(6), 1041–1052.

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

Hong, S. W., & Vierling, E. (2000). Mutants of Arabidopsis thaliana defective in the acquisition of tolerance to high temperature stress. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(8), 4392-4397.

Hong, S., & Vierling, E. (2001). Hsp101 is necessary for heat tolerance but dispensable for development and germination in the absence of stress. *The Plant Journal*, *27*(1), 25-35.

Hua, J. (2009). From freezing to scorching: Transcriptional responses to temperature variations in plants. *Current Opinion in Plant Biology*, *12*(5), 568-573.

Ichikawa, T., Nakazawa, M., Kawashima, M., Iizumi, H., Kuroda, H., Kondou, Y., ... Matsui, M. (2006). The FOX hunting system: An alternative gain-of-function gene hunting technique. *The Plant Journal*, *48*(6), 974-985.

Intergovernmental Panel on Climate Change. (2007). The physical science basis: Contribution of working group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller (Eds.)]. Cambridge, United Kingdom and New York, NY: Cambridge University Press.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, *36*(suppl 2), W5.

Kang, H.-G., Kim, J., Kim, B., Jeong, H., Choi, S. H., Kim, E. K., … Lim, P. O. (2011). Overexpression of FTL1/DDF1, an AP2 transcription factor, enhances tolerance to cold, drought, and heat stresses in *Arabidopsis thaliana*. *Plant Science*, *180*(4), 634-641.

Kantorow, M., & Piatigorsky, J. (1998). Phosphorylations of [alpha]A- and [alpha]B-crystallin. *International Journal of Biological Macromolecules*, *22*(3-4), 307-314.

Kawabe, A., & Miyashita, N. T. (2003). Patterns of codon usage bias in three dicot and four monocot plant species. *Genes & Genetic Systems*, *78*(5), 343-352.

Knight, H., & Knight, M. R. (2001). Abiotic stress signalling pathways: Specificity and cross-talk. *Trends in Plant Science*, *6*(6), 262-267.

Kochetov, A. V., Sirnik, O. A., Rogosin, I. B., Glazko, G. V., Komarova, M. L., & Shumny, V. K. (2002). Contextual features of higher plant mRNA 5"-untranslated regions. *Molecular Biology*, *36*(4), 510–516.

Kotak, S., Larkindale, J., Lee, U., von Koskull-Döring, P., Vierling, E., & Scharf, K.-D. (2007). Complexity of the heat stress response in plants. *Current Opinion in Plant Biology*, *10*(3), 310-316.

Kriehuber, T., Rattei, T., Weinmaier, T., Bepperling, A., Haslbeck, M., & Buchner, Johannes. (2010). Independent evolution of the core domain and its flanking sequences in small heat shock proteins. *The FASEB Journal*, *24*(10), 3633-3642.

Krishna, P., & Gloor, G. (2001). The Hsp90 family of proteins in *Arabidopsis thaliana*. *Cell Stress & Chaperones*, *6*(3), 238-246.

Langenheim, J. H., Convis, C. L., Macedo, C. A., & Stubblebine, W. H. (1986). *Hymenaea* and *Copaifera* leaf sesquiterpenes in relation to lepidopteran herbivory in southeastern Brazil. *Biochemical Systematics and Ecology*, *14*(1), 41–49.

Langridge, J. (1968). Thermal responses of mutant enzymes and temperature limits to growth. *Molecular and General Genetics MGG*, *103*(2), 116–126.

Larcher, W. (1987). Stress bei pflanzen. *Naturwissenschaften, 74*,158-167.

Larkindale, J., & Vierling, E. (2008). Core genome responses involved in acclimation to high temperature. *Plant Physiology*, *146*(2), 748-761.

Latijnhouwers, M., Xu, X. M., & Moller, S. G. (2010). *Arabidopsis* stromal 70-kDa heat shock proteins are essential for chloroplast development. *Planta*, *232*(3), 567-578.

Lee, S., Kim, S.-Y., Chung, E., Joung, Y.-H., Pai, H.-S., Hur, C.-G., & Choi, D. (2004). EST and microarray analyses of pathogen-responsive genes in hot pepper (Capsicum annuum L.) non-host resistance against soybean pustule pathogen (Xanthomonas axonopodis pv. glycines). *Functional & Integrative Genomics*, *4*(3), 196-205.

Lenne, C., Block, M. A., Garin, J., & Douce, R. (1995). Sequence and expression of the mRNA encoding HSP22: The mitochondrial small heat-shock protein in pea leaves. *Biochemical Journal*, *311*(3), 805-813.

Lester, G. E. (2008). Antioxidant, sugar, mineral, and phytonutrient concentrations across edible fruit tissues of orange-fleshed honeydew melon (Cucumis melo L.). *Journal of Agricultural and Food Chemistry*, *56*(10), 3694–3698.

Leyman, B., Geelen, D., Quintero, F. J., & Blatt, M. R. (1999). A tobacco syntaxin with a role in hormonal control of guard cell ion channels. *Science*, *283*(5401), 537-540.

Li, S., Fu, Q., Chen, L., Huang, W., & Yu, D. (2011). *Arabidopsis thaliana* WRKY25, WRKY26, and WRKY33 coordinate induction of plant thermotolerance. *Planta*, *233*(6), 1237-1252.

Li, S., Zhou, X., Chen, L., Huang, W., & Yu, D. (2010). Functional characterization of *Arabidopsis thaliana* WRKY39 in heat stress. *Molecules and Cells*, *29*(5), 475-483.

Lichtenthaler, H. K. (1987). Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Plant Cell Membranes* 148, 350-382).

Lindquist, S. (1986). The heat-shock response. *Annual Review of Biochemistry*, *55*(1), 1151–1191.

Lindquist, S., & Craig, E. A. (1988). The heat-shock proteins. *Annual Review of Genetics, 22*(1), 631-677. doi:10.1146/annurev.ge.22.120188.003215

Lindqvist, C., Scheen, A.-C., Yoo, M.-J., Grey, P., Oppenheimer, D. G., Leebens-Mack, … Albert, A. A. (2006). An expressed sequence tag (EST) library from developing fruits of an Hawaiian endemic mint (*Stenogyne rugosa*, Lamiaceae): characterization and microsatellite markers. *BMC Plant Biology*, *6*(16), 1-16.

Lippincott, J. A., & Lippincott, B. B. (1975). The genus *Agrobacterium* and plant tumorigenesis. *Annual Review of Microbiology*, *29*(1), 377-405.

Liu, C. J., Blount, J. W., Steele, C. L., & Dixon, R. A. (2002). Bottlenecks for metabolic engineering of isoflavone glycoconjugates in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America, 99*(22), 14578-14583.

Liu, X., Wang, Z., Wang, L., Wu, R., Phillips, J., & Deng, X. (2009). LEA 4 group genes from the resurrection plant Boea hygrometrica confer dehydration tolerance in transgenic tobacco. *Plant Science*, *176*(1), 90–98.

Long, S. P., & Ort, D. R. (2010). More than taking the heat: crops and global change. *Current Opinion in Plant Biology, 13*(3), 240–247.

Lovelock, C. E., Virgo, A., Popp, M., & Winter, K. (1999). Effects of elevated CO_2 concentrations on photosynthesis, growth and reproduction of branches of the tropical canopy tree species, *Luehea seemannii* Tr. & Planch. *Plant, Cell & Environment*, *22*(1), 49-59.

Lütcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F., & Scheele, G. A. (1987). Selection of AUG initiation codons differs in plants and animals. *The EMBO Journal*, *6*(1), 43-48.

Machuka, J., Bashiardes, S., Ruben, E., Spooner, K., Cuming, A., Knight, C., & Cove, D. (1999). Sequence analysis of expressed sequence tags from an ABA-treated cDNA library identifies stress response genes in the moss Physcomitrella patens. *Plant & Cell Physiology*, *40*(4), 378-387.

Maestri, E., Klueva, N., Perrotta, C., Gulli, M., Nguyen, H. T., & Marmiroli, N. (2002). Molecular genetics of heat tolerance and heat shock proteins in cereals. *Plant Molecular Biology*, *48*(5), 667–681.

Mahan, J. R., McMichael, B. L., & Wanjura, D. F. (1995). Methods for reducing the adverse effects of temperature stress on plants: A review. *Environmental and Experimental Botany*, *35*(3), 251–258.

Martin, J., Horwich, A. L., & Hartl, F. U. (1992). Prevention of protein denaturation under heat stress by the chaperonin Hsp60. *Science*, *258*(5084), 995-998.

Martineau, J. R., Williams, J. E., & Sullivan, J. H. (1978). Temperature tolerance in soybeans. I. Evaluation of a technique for assessing cellular membrane thermostability. *Crop Science*, *19*(1), 75-78.

May, M. J., & Leaver, C. J. (1994). *Arabidopsis thaliana* γ-glutamylcysteine synthetase is structurally unrelated to mammalian, yeast, and *Escherichia coli* homologs. *Proceedings of the National Academy of Sciences, 91*, 10059–10063.

May, M. J., Vernoux, T., Leaver, C., Montagu, M. V, & Inzé, D. (1998). Glutathione homeostasis in plants: implications for environmental sensing and plant development. *Journal of Experimental Botany*, *49*(321), 649-667.

McGinnis, S., & Madden, T. L. (2004). BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, *32*(suppl 2), W20.

Meinke, D. W., Cherry, J., Michael, D. C., Rounsley, S. D., & Koornneef, M. (1998). *Arabidopsis thaliana*: A model plant for genome analysis. *Science*, *282*(5389), 662 -682.

Miao, Y., Lv, D., Wang, P., Wang, X. C., Chen, J., Miao, C., & Song, C. P. (2006). An *Arabidopsis* glutathione peroxidase functions as both a redox transducer and a scavenger in abscisic acid and drought stress responses. *The Plant Cell Online*, *18*(10), 2749-2766.

Mishra, S. K., Tripp, J., Winkelhaus, S., Tschiersch, B., Theres, K., Nover, L., & Scharf, K.-D. (2002). In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. *Genes & Development*, *16*(12), 1555-1567.

Mittermeier, R. A., Myers, N., Thomsen, J. B., Da Fonseca, G. A. B., & Olivieri, S. (1998). Biodiversity hotspots and major tropical wilderness areas: Approaches to setting conservation priorities. *Conservation Biology*, *12*(3), 516–520.

Mittler, R., & Blumwald, E. (2010). Genetic engineering for modern agriculture: challenges and perspectives. *Annual Review of Plant Biology*, *61*, 443–462.

Mohammad, A., Miranda-Ros, J., Estrada Navarrete, G., Quinto, C., Olivares, J. E., Garca-Ponce, B., & Sanchez, F. (2004). Nodulin22 from *Phaseolus vulgaris* protects *Escherichia coli* cells from oxidative stress. *Planta*, *219*(6), 993-1002.

Molinari, H. B. C., Marur, C. J., Daros, E., de Campos, M. K. F., de Carvalho, J. F. R. P., Filho, J. C. B., … Vieira, L. G. E. (2007). Evaluation of the stress-inducible production of proline in transgenic sugarcane (*Saccharum* spp.): Osmotic adjustment, chlorophyll fluorescence, and oxidative stress. *Physiologia Plantarum, 130,* 218-229.

Montero-Barrientos, M., Hermosa, R., Cardoza, R. E., Gutiérrez, S., Nicolás, C., & Monte, E. (2010). Transgenic expression of the *Trichoderma harzianum hsp70* gene increases *Arabidopsis* resistance to heat and other abiotic stresses. *Journal of Plant Physiology*, *167*(8), 659-665.

Müller-Moulé, P. (2008). An expression analysis of the ascorbate biosynthesis enzyme VTC2. *Plant Molecular Biology*, *68*(1-2), 31-41.

Murata, N., & Los, D. A. (1997). Membrane fluidity and temperature perception. *Plant Physiology*, *115*(3), 875.

Nascimento, J. È. C., & Langenheim, J. H. (1986). Leaf sesquiterpenes and phenolics in *Copaifera multijuga* on contrasting soil types in a central Amazonian rain forest. *Biochemical Systematics and Ecology*, *14*(6), 615–624.

Natarajan, P., Kanagasabapathy, D., Gunadayalan, G., Panchalingam, J., Shree, N., Sugantham, P., … Madasamy, P. (2010). Gene discovery from Jatropha curcas by sequencing of ESTs from normalized and full-length enriched cDNA library from developing seeds. *BMC Genomics*, *11*, 606-620.

Nautiyal, P. C., & Shono, M. (2010). Analysis of the role of mitochondrial and endoplasmic reticulum localized small heat shock proteins in tomato. *Biologia Plantarum, 54*(4), 715-719.

Neilson, K. A., Gammulla, C. G., Mirzaei, M., Imin, N., & Haynes, P. A. (2010). Proteomic analysis of temperature stress in plants. *Proteomics*, *10*(4), 828–845.

Pahlich, E. (1993). Larcher's definition of plant stress: A valuable principle for metabolic adaptibility research. *Revista Brasileira de Fisiologia Vegetal*, *5*(2), 209–216.

Pal-Bhadra, M., Bhadra, U., & Birchler, J. A. (2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Molecular Cell*, *9*(2), 315-327.

Paterson, A. H., Felker, P., Hubbell, S. P., & Ming, R. (2008). The fruits of tropical plant genomics. *Tropical Plant Biology*, *1*(1), 3-19.

Pearl, L. H., & Prodromou, C. (2006). Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annual Review of Biochemistry*, *75*, 271-294.

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., & Liuni, S. (2001). Structural and functional features of eukaryotic mRNA untranslated regions. *Gene*, *276*(1-2), 73-81.

Plowden, C. (2003). Production ecology of Copaíba (*Copaifera* spp.) oleoresin in the Eastern Brazilian Amazon. *Economic Botany*, *57*(4), 491–501.

Porter, J. R., & Semenov, M. A. (2005). Crop responses to climatic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1463), 2021-2035.

Priya, P., Venkatachalam, P., & Thulaseedharan, A. (2006). Molecular cloning and characterization of the rubber elongation factor gene and its promoter sequence from rubber tree (*Hevea brasiliensis*): A gene involved in rubber biosynthesis. *Plant Science*, *171*(4), 470–480.

Ream, W. (1989). *Agrobacterium tumefaciens* and interkingdom genetic exchange. *Annual Review of Phytopathology*, *27*(1), 583-618.

Ristic, Z., Momčilović, I., Bukovnik, U., Prasad, P. V., Fu, J., DeRidder, B. P., …Mladenov, N. (2009). Rubisco activase and wheat productivity under heat-stress conditions. *Journal of Experimental Botany*, *60*(14), 4003-4014.

Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in drosophila. *Experientia*, *18*(12), 571-573.

Ritossa, F. (1996). Discovery of the heat shock response. *Cell Stress & Chaperones*, *1*(2), 97.

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., & Mittler, R. (2004). When defense pathways collide: the response of *Arabidopsis* to a combination of drought and heat stress. *Plant Physiology*, *134*(4), 1683-1696.

Ruelland, E., & Zachowski, A. (2010). How plants sense temperature. *Environmental and Experimental Botany*, *69*(3), 225-232.

Saidi, Y., Finka, A., Muriset, M., Bromberg, Z., Weiss, Y. G., Maathuis, F. J. M., & Goloubinoff, P. (2009). The heat shock response in moss plants is regulated by specific calcium-permeable channels in the plasma membrane. *The Plant Cell Online*, *21*(9), 2829 -2843.

Sakamoto, T., & Murata, N. (2002). Regulation of the desaturation of fatty acids and its role in tolerance to cold and salt stress. *Current Opinion in Microbiology*, *5*(2), 208-210.

Sakurai, H., & Enoki, Y. (2010). Novel aspects of heat shock factors: DNA recognition, chromatin modulation, and gene expression. *FEBS Journal*, *277*(20), 4140-4149.

Salvucci, M. E. (2008). Association of Rubisco activase with chaperonin-60beta: A possible mechanism for protecting photosynthesis during heat stress. *Journal of Experimental Botany*, *59*(7), 1923-1933.

Salvucci, M. E., & Crafts-Brandner, S. J. (2004). Relationship between the heat tolerance of photosynthesis and the thermal stability of rubisco activase in plants from contrasting thermal environments. *Plant Physiology*, *134*(4), 1460-1470.

Sanford, J. C. (1990). Biolistic plant transformation. *Physiologia Plantarum*, *79*(1), 206-209.

Sato, Y., & Yokoya, S. (2007). Enhanced tolerance to drought stress in transgenic rice plants overexpressing a small heat-shock protein, sHSP17.7. *Plant Cell Reports, 27*(2), 329-334.

Schäfer, G., & Brown G. D. (2009). Generation of retroviral macrophage cDNA expression libraries and functional screening for surface receptors. *Methods in Molecualar Biology, 531,* 1-15. doi: 10.1007/978-1-59745-396-7_1

Schaller, F. Biesgen, C., Mussig, C., Altmann, T., & Weiler, E. W. (2002). 12-Oxophytodienoate reductase 3 (OPR3) is the isoenzyme involved in jasmonate biosynthesis. *Planta, 210,* 979–984.

Schaller, F., Zerbe, P., Reinbothe, S., Reinbothe, C., Hofmann, E., & Pollmann, S. (2008). The allene oxide cyclase family of *Arabidopsis thaliana*–Localization and cyclization. *Febs Journal*, *275*(10), 2428–2441.

Schaupp, A., Marcinowski, M., Grimminger, V., Bösl, B., & Walter, S. (2007). Processing of proteins by the molecular chaperone Hsp104. *Journal of Molecular Biology*, *370*(4), 674–686.

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, *270*(5235), 467 -470.

Schöffl, F., Prändl, R., & Reindl, A. (1998). Regulation of the heat-shock response. *Plant Physiology*, *117*(4), 153-158.

Schrader, E. K., Harstad, K. G., & Matouschek, A. (2009). Targeting proteins for degradation. *Nature Chemical Biology*, *5*(11), 815–822.

Shi, L.-X., & Theg, S. M. (2010). A stromal heat shock protein 70 system functions in protein import into chloroplasts in the moss *Physcomitrella patens*. *The Plant Cell Online*, *22*(1), 205-220.

Siddique, M., Gernhard, S., von Koskull-Döring, P., Vierling, E., & Scharf, K. D. (2008). The plant sHSP superfamily: Five new members in Arabidopsis thaliana with unexpected properties. *Cell Stress and Chaperones*, *13*(2), 183–197.

Simoes-Araujo, J. L., Alves-Ferreira, M., Rumjanek, N. G., & Margis-Pinheiro, M. (2008). VuNIP1 (NOD26-like) and VuHSP17.7 gene expression are regulated in response to heat stress in cowpea nodule. *Environmental and Experimental Botany*, *63*(1-3), 256-265.

Skipsey, M., Andrews, C. J., Townson, J. K., Jepson, I., & Edwards, R. (1997). Substrate and thiol specificity of a stress-inducible glutathione transferase from soybean. *FEBS Letters*, *409*(3), 370-374.

Slabas, A. R., & Fawcett, T. (1992). The biochemistry and molecular biology of plant lipid biosynthesis. *Plant Molecular Biology*, *19*(1), 169–191.

Stougaard, J. (2000). Regulators and regulation of legume root nodule development. *Plant Physiology*, *124*(2), 531-540.

Su, P.-H., & Li, H.-min. (2010). Stromal Hsp70 is important for protein translocation into pea and *Arabidopsis* chloroplasts. *The Plant Cell Online*, *22*(5), 1516 -1531.

Sun, W., Van Montagu, M., & Verbruggen, N. (2002). Small heat shock proteins and stress tolerance in plants. *Biochimica Et Biophysica Acta*, *1577*(1), 1-9.

Sung, D., Kaplan, F., & Guy, C. L. (2001). Plant Hsp70 molecular chaperones: Protein structure, gene family, expression and function. *Physiologia Plantarum*, *113*(4), 443-451.

Taiz, L., & Zeiger, E. (2010). *Plant physiology* (5th Ed.). Sunderland, MA: Sinauer Associates Inc.

Tian, L., & Chen, Z. J. (2001). Blocking histone deacetylation in *Arabidopsis* induces pleiotropic effects on plant gene regulation and development. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(1), 200-205.

Valvekens, D., Montagu, M. V., & Lijsebettens, M. V. (1988). *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(15), 5536-5540.

van Montfort, R. L. M., Basha, E., Friedrich, K. L., Slingsby, C., & Vierling, E. (2001). Crystal structure and assembly of a eukaryotic small heat shock protein. *Nature Structural & Molecular Biology*, *8*(12), 1025–1030.

Vierling, E. (1991). The roles of heat shock proteins in plants. *Annual Review of Plant Biology*, *42*(1), 579–620.

von Koskull-Döring, P., Scharf, K. D., & Nover, L. (2007). The diversity of plant heat stress transcription factors. *Trends in Plant Science*, *12*(10), 452–457.

Walther, G. R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., … Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, *416*(6879), 389–395.

Wang, H.-C., & Hickey, D. A. (2007). Rapid divergence of codon usage patterns within the rice genome. *BMC Evolutionary Biology*, *7 Suppl 1*, S6.

Wang, M., Liu, M., Li, D., Wu, J., Li, X., & Yang, Y. (2010). Overexpression of FAD2 promotes seed germination and hypocotyl elongation in *Brassica napus*. *Plant Cell, Tissue and Organ Culture*, *102*(2), 205-211.

Wang, W., Vinocur, B., & Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. *Planta*, *218*(1), 1–14.

Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*, *9*(5), 244–252.

Wang, Y. C., Yang, C. P., Liu, G. F., Jiang, J., & Wu, J. H. (2006). Generation and analysis of expressed sequence tags from a cDNA library of *Tamarix androssowii*. *Plant Science*, *170*(1), 28–36.

Wang, Y. H., Mosebach, C. M, Kibbey, A. S, Ryhal, M. K, Jones, A. D, Palmer, J. A, & Kochian, L. V. (2009). Generation of *Arabidopsis* mutants by heterologous expression of a full-length cDNA library from tomato fruits. *Plant Molecular Biology Reporter*, *27*(4), 454–461.

Waters, E. R. (1995). The molecular evolution of the small heat-shock proteins in plants. *Genetics*, *141*(2), 785-795.

Waters, E. R., & Vierling, E. (1999). Chloroplast small heat shock proteins: Evidence for atypical evolution of an organelle-localized protein. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(25), 14394 -14399.

Willits, D. H., & Peet, M. M. (2001). Measurement of chlorophyll fluorescence as a heat stress indicator in tomato: Laboratory and greenhouse comparisons. *American Society for Horticultural Science, 126*(2), 188–194.

Wood, A. J., Duff, R. J., & Oliver, M. J. (1999). Expressed sequence tags (ESTs) from desiccated Tortula ruralis identify a large number of novel plant genes. *Plant and Cell Physiology*, *40*(4), 361.

Xue, Y., Peng, R., Xiong, A., Li, X., Zha, D., & Yao, Q. (2010). Over-expression of heat shock protein gene *hsp26* in *Arabidopsis thaliana* enhances heat tolerance. *Biologia Plantarum*, *54*(1), 105-111.

Ye, J., McGinnis, S., & Madden, T. L. (2006). BLAST: improvements for better sequence analysis. *Nucleic Acids Research*, *34*(Web Server issue), W6-9.

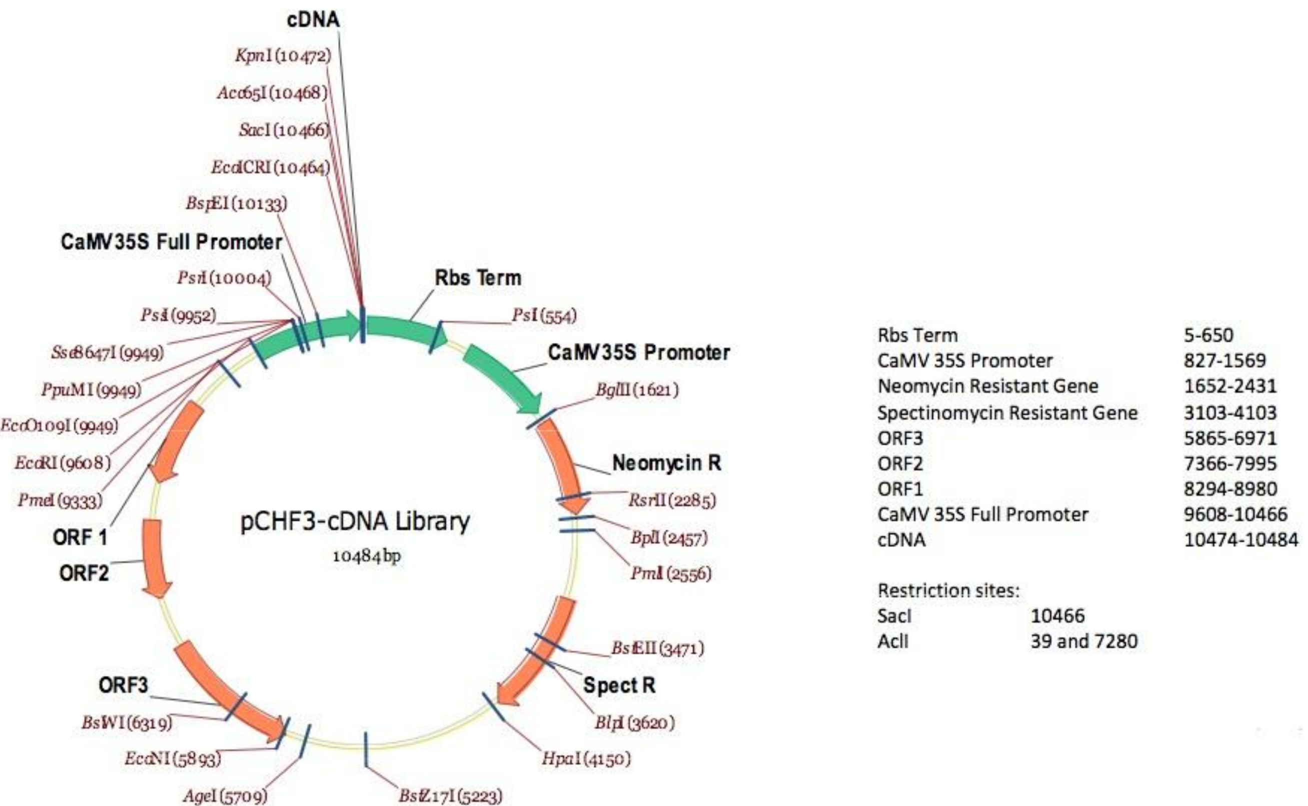
Yoshimura, K., Miyao, K., Gaber, A., Takeda, T., Kanaboshi, H., Miyasaka, H., & Shigeoka, S. (2004). Enhancement of stress tolerance in transgenic tobacco plants overexpressing *Chlamydomonas* glutathione peroxidase in chloroplasts or cytosol. *The Plant Journal*, *37*(1), 21–33.

Zhang, J., Liu, T., Fu, J., Zhu, Y., Jia, J., Zheng, J., ….Wang, G. (2007). Construction and application of EST library from *Setaria italica* in response to dehydration stress. *Genomics*, *90*(1), 121–131.

Ziegler, J., Stenzel, I., Hause, B., Maucher, H., Hamberg, M., Grimm, R., …Wasternack, C. (2000). Molecular cloning of allene oxide cyclase. *Journal of Biological Chemistry*, *275*(25), 19132 -19138.

APPENDIX A

pCHF3 VECTOR MAP



Rbs Term	5-650
CaMV 35S Promoter	827-1569
Neomycin Resistant Gene	1652-2431
Spectinomycin Resistant Gene	3103-4103
ORF3	5865-6971
ORF2	7366-7995
ORF1	8294-8980
CaMV 35S Full Promoter	9608-10466
cDNA	10474-10484

Restriction sites:

SacI	10466
AclI	39 and 7280

The flanking sequences of the cDNA insert

SacI KpnI

5′ AGAACACGGG GGACGAGCTC GGTACCC - cDNA - CTGGAGCTTT CGTTCGTATC ATCGGTTTCG ACAACGTTCG TCAAGTTCAA TGC

APPENDIX B

pCHF3 SEQUENCE

CTGGAGCTTTCGTTCGTATCATCGGTTTCGACAACGTTCGTCAAGTTCAATGC
ATCAGTTTCATTGCGCACACACCAGAATCCTACTGAGTTCGAGTATTATGGCA
TTGGGAAACATGTTTTTCTTGTACCATTTGTTGTGCTTGTAATTTACTGTGTTT
TTTATTCGGTTTTCGCTATCGAACTGTGAAATGGAAATGGATGGAGAAGAGTT
AATGAATGATATGGTCCTTTTGTTCATTCTCAAATTAATATTATTTGTTTTTTC
TCTTATTTGTTGTGTGTTGAATTTGAAAATATAAGAGATATGCAAACATTTTG
TTTTGAGTAAAAATGTGTCAAATCGTGGCCTCTAATGACCGAAGTTAATATGA
GGAGTAAAACACTTGTAGTTGTACCATTATGCTTATTCACTAGGCAACAAATA
TATTTTCAGACCTAGAAAAGCTGCAAATGTTACTGAATACAAGTATGTCCTCT
TGTGTTTTAGACATTTATGAACTTTCCTTTATGTAATTTTCCAGAATCCTTGTC
AGATTCTAATCATTGCTTTATAATTATAGTTATACTCATGGATTTGTAGTTGA
GTATGAAAATATTTTTTAATGCATTTTATGACTTGCCAATTGATTGACAACAT
GCATCAAGCTAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAA
AACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAG
CTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGC
AGCCTGAATGGCGAATGCTAGAGCAGCTTGCATGGTGGAGCACGACACTCTC
GTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTG
AGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCC
AGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTAC
AAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCG
ACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAG
AAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATAACATGGT
GGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAA
GACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCC
TCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAA
GGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTT
CAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGA
GCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTG
ATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGC
AAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACACGCTGA
AATCACCAGTCTCTCTCTACAAATCTATCTCTCTCGATTCGCAGATCTGTCGA
TCGACCATGGGGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTT
GGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTC
TGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCA
AGACCGACCTGTCCGGTGCCCTGAATGAACTCCAGGACGAGGCAGCGCGGCT
ATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCA
CTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCT
CCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAA
TGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGC
GAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGAT
CAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCG
CCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACACATGG
CGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCA
TCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGC
TACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTC
GTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCT

TCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGGATCGATCCTCTAGCT
AGAGTCGATCGACATCGAGTTTCTCCATAATAATGTGTGAGTAGTTCCCAGAT
AAGGGAATTAGGGTTCTTATAGGGTTTCGCTCACGTGTTGAGCATATAAGAA
ACCCTTAGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAAT
TCCTAAAACCAAAATCCAGTACTAAAATCCAGATCACCTAAAGTCCCTATAG
ATCCCCCGAATTAATTCGGCGTTAATTCAGTACATTAAAAACGTCCGCAATGT
GTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACCACAATATATCCTGCCAC
CAGCCAGCCAACAGCTCCCCGACCGGCAGCTCGGCACAAAATCACCACTCGA
TACAGGCAGCCCATCAGTCCGGGACGGCGTCAGCGGGAGAGCCGTTGTAAG
GCGGCAGACTTTGCTCATGTTACCGATGCTATTCGGAAGAACGGCAACTAAG
CTGCCGGGTTTGAAACACGGATGATCTCGCGGAGGGTAGCATGTTGATTGTA
ACGATGACAGAGCGTTGCTGCCTGTGATCAATTCGGGCACGAACCCAGTGGA
CATAAGCCTCGTTCGGTTCGTAAGCTGTAATGCAAGTAGCGTAACTGCCGTCA
CGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGG
CGGTTTTCATGGCTTCTTGTTATGACATGTTTTTTTGGGGTACAGTCTATGCCT
CGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTAT
GGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAA
ACATCATGGGGGAAGCGGTGATCGCCGAAGTATCGACTCAACTATCAGAGGT
AGTTGGCGTCATCGAGCGCCATCTCGAACCGACGTTGCTGGCCGTACATTTGT
ACGGCTCCGCAGTGGATGGCGGCCTGAAGCCACACAGTGATATTGATTTGCT
GGTTACGGTGACCGTAAGGCTTGATGAAACAACGCGGCGAGCTTTGATCAAC
GACCTTTTGGAAACTTCGGCTTCCCCTGGAGAGAGCGAGATTCTCCGCGCTGT
AGAAGTCACCATTGTTGTGCACGACGACATCATTCCGTGGCGTTATCCAGCTA
AGCGCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTTGCAGGTAT
CTTCGAGCCAGCCACGATCGACATTGATCTGGCTATCTTGCTGACAAAAGCA
AGAGAACATAGCGTTGCCTTGGTAGGTCCAGCGGCGGAGGAACTCTTTGATC
CGGTTCCTGAACAGGATCTATTTGAGGCGCTAAATGAAACCTTAACGCTATG
GAACTCGCCGCCCGACTGGGCTGGCGATGAGCGAAATGTAGTGCTTACGTTG
TCCCGCATTTGGTACAGCGCAGTAACCGGCAAAATCGCGCCGAAGGATGTCG
CTGCCGACTGGGCAATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCATACT
TGAAGCTAGACAGGCTTATCTTGGACAAGAAGAAGATCGCTTGGCCTCGCGC
GCAGATCAGTTGGAAGAATTTGTCCACTACGTGAAAGGCGAGATCACCAAGG
TAGTCGGCAAATAATGTCTAGCTAGAAATTCGTTCAAGCCGACGCCGCTTCG
CCGGCGTTAACTCAAGCGATTAGATGCACTAAGCACATAATTGCTCACAGCC
AAACTATCAGGTCAAGTCTGCTTTTATTATTTTTAAGCGTGCATAATAAGCCC
TACACAAATTGGGAGATATATCATGCATGACCAAAATCCCTTAACGTGAGTT
TTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGA
GATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCT
ACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGG
TAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCC
GTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTC
TGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACC
GGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAA
CGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACT
GAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAC

GAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC
CTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAG
CGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTC
TCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACA
ATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGT
GACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCT
GACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCC
GGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGC
AGGGTGCCTTGATGTGGGCGCCGGCGGTCGAGTGGCGACGGCGCGGCTTGTC
CGCGCCCTGGTAGATTGCCTGGCCGTAGGCCAGCCATTTTTGAGCGGCCAGC
GGCCGCGATAGGCCGACGCGAAGCGGCGGGGCGTAGGGAGCGCAGCGACCG
AAGGGTAGGCGCTTTTTGCAGCTCTTCGGCTGTGCGCTGGCCAGACAGTTATG
CACAGGCCAGGCGGGTTTTAAGAGTTTTAATAAGTTTTAAAGAGTTTTAGGC
GGAAAAATCGCCTTTTTTCTCTTTTATATCAGTCACTTACATGTGTGACCGGTT
CCCAATGTACGGCTTTGGGTTCCCAATGTACGGGTTCCGGTTCCCAATGTACG
GCTTTGGGTTCCCAATGTACGTGCTATCCACAGGAAAGAGACCTTTTCGACCT
TTTTCCCCTGCTAGGGCAATTTGCCCTAGCATCTGCTCCGTACATTAGGAACC
GGCGGATGCTTCGCCCTCGATCAGGTTGCGGTAGCGCATGACTAGGATCGGG
CCAGCCTGCCCCGCCTCCTCCTTCAAATCGTACTCCGGCAGGTCATTTGACCC
GATCAGCTTGCGCACGGTGAAACAGAACTTCTTGAACTCTCCGGCGCTGCCA
CTGCGTTCGTAGATCGTCTTGAACAACCATCTGGCTTCTGCCTTGCCTGCGGC
GCGGCGTGCCAGGCGGTAGAGAAAACGGCCGATGCCGGGATCGATCAAAAA
GTAATCGGGGTGAACCGTCAGCACGTCCGGGTTCTTGCCTTCTGTGATCTCGC
GGTACATCCAATCAGCTAGCTCGATCTCGATGTACTCCGGCCGCCCGGTTTCG
CTCTTTACGATCTTGTAGCGGCTAATCAAGGCTTCACCCTCGGATACCGTCAC
CAGGCGGCCGTTCTTGGCCTTCTTCGTACGCTGCATGGCAACGTGCGTGGTGT
TTAACCGAATGCAGGTTTCTACCAGGTCGTCTTTCTGCTTTCCGCCATCGGCT
CGCCGGCAGAACTTGAGTACGTCCGCAACGTGTGGACGGAACACGCGGCCG
GGCTTGTCTCCCTTCCCTTCCCGGTATCGGTTCATGGATTCGGTTAGATGGGA
AACCGCCATCAGTACCAGGTCGTAATCCCACACACTGGCCATGCCGGCCGGC
CCTGCGGAAACCTCTACGTGCCCGTCTGGAAGCTCGTAGCGGATCACCTCGC
CAGCTCGTCGGTCACGCTTCGACAGACGGAAAACGGCCACGTCCATGATGCT
GCGACTATCGCGGGTGCCCACGTCATAGAGCATCGGAACGAAAAAATCTGGT
TGCTCGTCGCCCTTGGGCGGCTTCCTAATCGACGGCGCACCGGCTGCCGGCG
GTTGCCGGGATTCTTTGCGGATTCGATCAGCGGCCGCTTGCCACGATTCACCG
GGGCGTGCTTCTGCCTCGATGCGTTGCCGCTGGGCGGCCTGCGCGGCCTTCAA
CTTCTCCACCAGGTCATCACCCAGCGCCGCGCCGATTTGTACCGGGCCGGAT
GGTTTGCGACCGTCACGCCGATTCCTCGGGCTTGGGGGTTCCAGTGCCATTGC
AGGGCCGGCAGACAACCCAGCCGCTTACGCCTGGCCAACCGCCCGTTCCTCC
ACACATGGGGCATTCCACGGCGTCGGTGCCTGGTTGTTCTTGATTTTCCATGC
CGCCTCCTTTAGCCGCTAAAATTCATCTACTCATTTATTCATTTGCTCATTTAC
TCTGGTAGCTGCGCGATGTATTCAGATAGCAGCTCGGTAATGGTCTTGCCTTG
GCGTACCGCGTACATCTTCAGCTTGGTGTGATCCTCCGCCGGCAACTGAAAGT

TGACCCGCTTCATGGCTGGCGTGTCTGCCAGGCTGGCCAACGTTGCAGCCTTG
CTGCTGCGTGCGCTCGGACGGCCGGCACTTAGCGTGTTTGTGCTTTTGCTCAT
TTTCTCTTTACCTCATTAACTCAAATGAGTTTTGATTTAATTTCAGCGGCCAGC
GCCTGGACCTCGCGGGCAGCGTCGCCCTCGGGTTCTGATTCAAGAACGGTTG
TGCCGGCGGCGGCAGTGCCTGGGTAGCTCACGCGCTGCGTGATACGGGACTC
AAGAATGGGCAGCTCGTACCCGGCCAGCGCCTCGGCAACCTCACCGCCGATG
CGCGTGCCTTTGATCGCCCGCGACACGACAAAGGCCGCTTGTAGCCTTCCATC
CGTGACCTCAATGCGCTGCTTAACCAGCTCCACCAGGTCGGCGGTGGCCCAT
ATGTCGTAAGGGCTTGGCTGCACCGGAATCAGCACGAAGTCGGCTGCCTTGA
TCGCGGACACAGCCAAGTCCGCCGCCTGGGGCGCTCCGTCGATCACTACGAA
GTCGCGCCGGCCGATGGCCTTCACGTCGCGGTCAATCGTCGGGCGGTCGATG
CCGACAACGGTTAGCGGTTGATCTTCCCGCACGGCCGCCCAATCGCGGGCAC
TGCCCTGGGGATCGGAATCGACTAACAGAACATCGGCCCCGGCGAGTTGCAG
GGCGCGGGCTAGATGGGTTGCGATGGTCGTCTTGCCTGACCCGCCTTTCTGGT
TAAGTACAGCGATAACCTTCATGCGTTCCCCTTGCGTATTTGTTTATTTACTCA
TCGCATCATATACGCAGCGACCGCATGACGCAAGCTGTTTTACTCAAATACA
CATCACCTTTTTAGACGGCGGCGCTCGGTTTCTTCAGCGGCCAAGCTGGCCGG
CCAGGCCGCCAGCTTGGCATCAGACAAACCGGCCAGGATTTCATGCAGCCGC
ACGGTTGAGACGTGCGCGGGCGGCTCGAACACGTACCCGGCCGCGATCATCT
CCGCCTCGATCTCTTCGGTAATGAAAAACGGTTCGTCCTGGCCGTCCTGGTGC
GGTTTCATGCTTGTTCCTCTTGGCGTTCATTCTCGGCGGCCGCCAGGGCGTCG
GCCTCGGTCAATGCGTCCTCACGGAAGGCACCGCGCCGCCTGGCCTCGGTGG
GCGTCACTTCCTCGCTGCGCTCAAGTGCGCGGTACAGGGTCGAGCGATGCAC
GCCAAGCAGTGCAGCCGCCTCTTTCACGGTGCGGCCTTCCTGGTCGATCAGCT
CGCGGGCGTGCGCGATCTGTGCCGGGGTGAGGGTAGGGCGGGGGCCAAACTT
CACGCCTCGGGCCTTGGCGGCCTCGCGCCCGCTCCGGGTGCGGTCGATGATT
AGGGAACGCTCGAACTCGGCAATGCCGGCGAACACGGTCAACACCATGCGG
CCGGCCGGCGTGGTGGTGTCGGCCCACGGCTCTGCCAGGCTACGCAGGCCCG
CGCCGGCCTCCTGGATGCGCTCGGCAATGTCCAGTAGGTCGCGGGTGCTGCG
GGCCAGGCGGTCTAGCCTGGTCACTGTCACAACGTCGCCAGGGCGTAGGTGG
TCAAGCATCCTGGCCAGCTCCGGGCGGTCGCGCCTGGTGCCGGTGATCTTCTC
GGAAAACAGCTTGGTGCAGCCGGCCGCGTGCAGTTCGGCCCGTTGGTTGGTC
AAGTCCTGGTCGTCGGTGCTGACGCGGGCATAGCCCAGCAGGCCAGCGGCGG
CGCTCTTGTTCATGGCGTAATGTCTCCGGTTCTAGTCGCAAGTATTCTACTTTA
TGCGACTAAAACACGCGACAAGAAAACGCCAGGAAAAGGGCAGGGCGGCAG
CCTGTCGCGTAACTTAGGACTTGTGCGACATGTCGTTTTCAGAAGACGGCTGC
ACTGAACGTCAGAAGCCGACTGCACTATAGCAGCGGAGGGGTTGGATCAAA
GTACTTGATCCCGAGGGGAACCCTGTGGTTGGCATGCACATACAAATGGACG
AACGGATAAACCTTTTCACGCCCTTTTAAATATCCGTTATTCTAATAAACGCT
CTTTTCTCTTAGGTTTACCCGCCAATATATCCTGTCAAACACTGATAGTTTAA
ACTGAAGGCGGGAAACGACAATCTGATCCAAGCTCAAGCTGCTCTAGCCAAT
ACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCAC
GACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGA
GTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGT
ATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATG
ACCATGATTACGAATTCGGTCCCCAGATTAGCCTTTTCAATTTCAGAAAGAAT

GCTAACCCACAGATGGTTAGAGAGGCTTACGCAGCAGGTCTCATCAAGACGA
TCTACCCGAGCAATAATCTCCAGGAAATCAAATACCTTCCCAAGAAGGTTAA
AGATGCAGTCAAAAGATTCAGGACTAACTGCATCAAGAACACAGAGAAAGA
TATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATTCAAGGCTTGC
TTCACAAACCAAGGCAAGTAATAGAGATTGGAGTCTCTAAAAAGGTAGTTCC
CACTGAATCAAAGGCCATGGAGTCAAAGATTCAAATAGAGGACCTAACAGA
ACTCGCCGTAAAGACTGGCGAACAGTTCATACAGAGTCTCTTACGACTCAAT
GACAAGAAGAAAATCTTCGTCAACATGGTGGAGCACGACACACTTGTCTACT
CCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGACTTT
TCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCAGCTATCT
GTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCA
TCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGT
CCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTT
CCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAA
GGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGG
AAGTTCATTTCATTTGGAGAGAACACGGGGGACGAGCTCGGTACCCNNNNNN
NNNNN

APPENDIX C

GO TERMS OVERREPRESENTED IN THE EXPRESSED SEQUENCE TAG LIBRARY

Fisher's exact test results to determine over- or underrepresentation of Gene Ontology (GO) terms for the set of *C. officinalis* terms (test set) comapred against terms for the *A. thaliana* set (reference set) of terms. Gene Ontology Identification (GO Term) are provided along with their name (Name) using the controlled vocabulary. The FDR (false discovery rate) and FWER (family wise error rate) are adjustments to the p-value due to multiple testing (Blüthgen et al., 2008). Each number (#) of terms is given for the nonannotated (nonannot) test group and reference group.

GO ID	GO TERM	FDR	FWER	single test p-value	# in test group	# in reference group	# nonannot test	# nonannot reference group
0008152	metabolic process	7.96E-09	2.36E-08	0	240	7822	191	14736
0044444	cytoplasmic part	7.96E-09	2.36E-08	0	196	5919	235	16639
0043229	intracellular organelle	7.96E-09	2.36E-08	0	238	7808	193	14750
0043226	organelle	7.96E-09	2.36E-08	0	238	7809	193	14749
0043231	intracellular membrane-bounded organelle	7.96E-09	2.36E-08	0	224	7396	207	15162
0006350	transcription	7.96E-09	2.36E-08	0	23	143	408	22415
0006091	generation of precursor metabolites and energy	7.96E-09	2.36E-08	0	28	301	403	22257
0044424	intracellular part	7.96E-09	2.36E-08	0	272	8595	159	13963
0005739	mitochondrion	7.96E-09	2.36E-08	0	58	1113	373	21445
0005737	cytoplasm	7.96E-09	2.36E-08	0	240	6349	191	16209
0005622	intracellular	7.96E-09	2.45E-08	1.14E-11	279	8971	152	13587
0043227	membrane-bounded organelle	7.96E-09	2.55E-08	1.85E-11	224	7402	207	15156
0009987	cellular process	7.96E-09	2.59E-08	2.26E-11	248	9408	183	13150
0044238	primary metabolic process	9.13E-09	3.20E-08	6.51E-11	187	6456	244	16102
0016020	membrane	3.94E-08	1.48E-07	6.15E-10	143	4606	288	17952
0009628	response to abiotic stimulus	5.32E-08	2.13E-07	1.01E-09	57	1217	374	21341

0016043	cellular component organization	1.53E-07	6.75E-07	2.56E-09	45	856	386	21702
0006950	response to stress	1.53E-07	6.87E-07	2.67E-09	80	2086	351	20472
0015979	photosynthesis	2.22E-07	1.11E-06	4.83E-09	18	155	413	22403
0009536	plastid	2.22E-07	1.11E-06	4.93E-09	107	3195	324	19363
0009058	biosynthetic process	9.83E-07	5.16E-06	2.13E-08	98	2911	333	19647
0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1.04E-06	5.69E-06	2.52E-08	53	1197	378	21361
0009579	thylakoid	1.07E-06	6.14E-06	2.90E-08	29	445	402	22113
0005840	ribosome	4.19E-06	2.51E-05	1.13E-07	29	476	402	22082
0005198	structural molecule activity	6.73E-06	4.21E-05	1.34E-07	30	509	401	22049
0008135	translation factor activity, nucleic acid binding	8.66E-06	6.06E-05	2.43E-07	15	140	416	22418
0043283	biopolymer metabolic process	8.66E-06	6.14E-05	2.54E-07	132	4553	299	18005
0034961	cellular biopolymer biosynthetic process	8.66E-06	6.22E-05	2.55E-07	62	1623	369	20935
0045182	translation regulator activity	8.66E-06	6.27E-05	2.64E-07	15	141	416	22417
0043284	biopolymer biosynthetic process	1.03E-05	7.74E-05	2.88E-07	62	1629	369	20929
0005488	binding	1.43E-05	1.11E-04	4.97E-07	246	10170	185	12388
0019538	protein metabolic process	2.66E-05	2.13E-04	7.91E-07	108	3578	323	18980
0034645	cellular macromolecule biosynthetic process	3.63E-05	3.00E-04	1.23E-06	62	1703	369	20855
0043170	macromolecule metabolic process	3.65E-05	3.10E-04	1.27E-06	132	4687	299	17871
0009059	macromolecule biosynthetic process	3.96E-05	3.47E-04	1.50E-06	62	1714	369	20844
0009056	catabolic process	6.11E-05	5.49E-04	2.16E-06	35	748	396	21810
0003824	catalytic activity	7.96E-05	7.36E-04	3.04E-06	193	7677	238	14881
0010467	gene expression	9.41E-05	8.94E-04	3.90E-06	62	1767	369	20791
0000166	nucleotide binding	2.63E-04	0.00256293	7.45E-06	71	2169	360	20389
0043228	non-membrane-	3.77E-04	0.00385939	1.25E-05	41	1028	390	21530

	bounded organelle							
0043232	intracellular non-membrane-bounded organelle	3.77E-04	0.00385939	1.25E-05	41	1028	390	21530
0030529	ribonucleoprotein complex	5.76E-04	0.00602818	2.25E-05	29	631	402	21927
0005783	endoplasmic reticulum	0.00122102	0.0130402	4.49E-05	20	365	411	22193
0050896	response to stimulus	0.00199078	0.0216607	6.89E-05	103	3734	328	18824
0005886	plasma membrane	0.0043364	0.0476142	1.65E-04	63	2055	368	20503
0005773	vacuole	0.00434263	0.0487143	1.74E-04	25	572	406	21986
0005576	extracellular region	0.0157943	0.169385	4.63E-04	20	439	411	22119

APPENDIX D

GENES AND PREDICTED PROTEIN SEQUENCES USED IN THE TRANSFORMATION STUDIES

TPG03

CGCTTTTCATTTTTGTGATTTTTAGGGAATCAACCAATCGAAAGAGTTC
TGTGCAGCGATTATGGCCGAAGCGAATCCCCAAGAGCAGCAACCCAT
GGAGAACAAGGAAGAGCAACAACGACTCAAGTACCTGGAATTCGTGC
ATGTTGCTATCATTCAAGCCCTTCTGCGTTTCTCCATGCTCTATGACTT
CGCCAAGGAACGAACTGGGCCTCTCAAACCTGGCGTTAGGCCGTCGA
AAAAAACTGTCAAAACTGTCGTCGGACCTGTTTATGACAAGTTTCACG
AAGTCCCCGCCGAGGTTCTCCGATACGTTGATCGCAAGGTCGATGAAT
CTATGACCGAGTTGGATCGCCGCGTGCCCACCGATGTAAAGAAGGCGT
CATCACAAGCCCTCTCGGCGGCTCAAAAGGCTCCTGAAGCGGCTCGAA
CCGTGGTATCTGAAGTGAGGCGCGCTGGTGTGGTCGATACTGCATCTG
GGTTGGCGAAGTCTGTTTACGGCAAGTGCGAGCCCACCGCCAAGGATT
TGTATTCTAAGTATGAACCGAAGGCCGAGAAGTGTGCGGTGTCGGCAT
GGCGCAAGCTGAACAAGCTTCCACTCTTCCCTCAAATGGCTAATGTTA
TTGTACCAACTGCCGCATATTGTACTGAGAAGTACAACGAGACTGTTG
TTTCTACTGCTGAGAAAGGTTACAAGGTTTCAGCGTACCTCCCGTTGGT
GCCTACAGAGAAGATAGCTAAGATTTTCTGTGAAGGTGAAGCTGCAA
AAGAGCCATCGGTCTCTTGCGTTGAAGCTCATTGATTTTTTTTTATTTTT
TTTAATGTTGTTTTTCTTAATCTTTGGAAAATATTTCAATACTCCGGTA
GAACTGAAATTTTTATGATCACTGCCTATTCTGTTTTGTGGGGGGAAA
AAAACCTCTGCAGAATTTTACCACTCACAAGTCACAATATTAAAATAG
CATTGAAATTTCAATTCTGCTCTGGAGTTAGGGGTAGAAATATTTTCCG
AGGGTTGTCTGGATTTGCCTTGAAACTTTTTTTAATATGAAAATTAAAG
ACGATTGTTTGTTTTGTATATTCTGGAATCACTTGTGATATGGATTTGC
TGTGTAATATAAATGTTTTATTTTCTGAAAAAAAAAAAAAAAAAAAAA
AAA

MAEANPQEQQPMENKEEQQRLKYLEFVHVAIIQALLRFSMLYDFAKERT
GPLKPGVRPSKKTVKTVVGPVYDKFHEVPAEVLRYVDRKVDESMTELDR
RVPTDVKKASSQALSAAQKAPEARTVVSEVRRAGVVDTASGLAKSVYG
KCEPTAKDLYSKYEPKAEKCAVSAWRKLNKLPLFPQMANVIVPTAAYCT
EKYNETVVSTAEKGYKVSAYLPLVPTEKIAKIFCEGEAAKEPSV SCVEAH

1A08

GAACTATTTGCAATTAGTCGCTATGGCTCAAGCGCTGTCTTTTTCAACA
TGTTGTCCCTCACCTCTGCTATCCCACAGAGGAGTAAGTTTTTCAGCCA
ACCCCAGAGACAAGAGCAACGTTGCAGCCCCGCCGTGCTCGGCCAGC
TTTCCGTCGAGGGGGCCGATGCGGATGCGACGGCTGGGAGTGGTGAG
GGCTCAGGCCGCCGGTGAGAACAAAGACAGTTCTTCAGTGGATGTGC
ATGTTAGTAAGGATAATCAACAAGGCACTGCTGTTGAGAGAAGGCCTC
GTACGACGGCCATTGACATTTCTCCGTTTGGTTTATTGGACCCTTGGTC
GCCTATGAGAACTATGCGCCAGATGCTGGACACAATGGACAGAATCTT
CGAAGACACCATGGCATTCCCCGGAAGAAGCAGAGGCGGCGTTAGCG
GCGGCGAAATCCGAGCACCGTGGGACATCAAGCACGAAGAGGACGAA
ATCAAGATGCGGTTCGACATGCCTGGTCTTTCCAAGGACGATGTTAAG
GTTTCTGTGGAGGACGATGTTCTTGTCATCAGAGGCGAGCATAGGAAG
GAAGAAGGCGATGATTCTTGGATGAGCAGGAGCCATAGCTCGTACGA
TACAAGATTGCAGCTTCCTGAAAACTGCGAGAAGGACAAAGTTAAAG
CAGAGTTGAAAAACGGTGTTCTTTATATTACTGTTCCTAAGACCAAAG
TCGAACGCAAAGTCACTGATGTAGAAATTCAGTGAGACCTTGGTAGCT
GCCGAGTAAGACCNTTTGCAGTTTGTGNCCTNGTTCTTCCCGGGACTCT
CTCTCTCTGCTTTTTGCCATGTTGTGNAATACTGGNTTAAAAAGGTTGA
AAGTGTGAAAGTGATGGAAAAAATATTTTGCAATCTTATTGCTNCNAA
AAATTGTCATTTTNCTTCNAAAAAAAAAAAAAAAAAAAAAAAA

MAQALSFSTCCPSPLLSHRGVSFSANPRDKSNVAAPPCSASFPSRGPMRM
RRLGVVRAQAAGENKDSSSVDVHVSKDNQQGTAVERRPRTTAIDISPFGL
LDPWSPMRTMRQMLDTMDRIFEDTMAFPGRSRGGVSGGEIRAPWDIKHE
EDEIKMRFDMPGLSKDDVKVSVEDDVLVIRGEHRKEEGDDSWMSRSHSS
YDTRLQLPENCEKDKVKAELKNGVLYITVPKTKVERKVTDVEIQ

3F04

ATCAGATTCAATCTCGACTCGAAGCTTCTTTTGCAGCTGCAAGCTACG
ATGGCGTCGGCTTCACTTGCTCTGAAGCGTCTCCTGTACTCTGGTCCTG
CCATTTCTTCAAGGTCTCTGATACGTACTGTGCCTCGGATGGTGTCTGC
TTCACGCTCCTTCAACACCAACGCCATGCGAGATTATGATGATGGAAA
TGGCCGCAGCCTTGACGTTGATCGTCGCTCTGATCGCTCTGTCCCACGC
GGTGATCGTCGTGATGATCTCTTCCCAGATGTGTTTGATCCATTTTCTC
CGACTCGGAGCTTAAGCCAGGTTCTGAATCTGATGGACCAATTTGTGG
ACAACCCATTTCTCTCTGCTTCGCGAGGGATTGGAGCTGGAGGGCTTC
GTCGTGGCTGGGATGCCAAGGAAACTGAAGATGCTCTGTTTCTTCGAG
TTGACATGCCGGGGCTTGGCAAAGACGATGTGAAGGTCTCTGTGGAGC
AGAACACTCTGATCATTAAGGGTGAAGGTGGAAAAGAAGAAGGGGAA
GAGGATAGTGCACGTCGGTACTCCAGCAGGATTGACCTACCAGAGAA
GCTTTACAAGATTGATCAAATCAAAGCTGAGATGAAGAATGGTGTGCT
GAAGGTGGTTGTGCCCAAAATGAAAGAGGAGGACAGGTCGGATGTGT
ACCAAGTTAAAGTTGATTAGGATGAGTTTTCGGTTTTTGAGTGATTGG
AAGGTTTTAGTCTGGCTTAGTCTTTTNCGGAATGCANATAGNGATCTG
ATGTTCTGTCTAATGANTGTTCTGACTTGAANTNGTCTGTNTCATTGAA
AGTTAATAAGGNCAGANTTCCNTGGAATNCAATNCTGTNNTTGCTTGC
NNTNGGAANNTGAAAAAAAAAAAAAAAAAAAAAAAAA

MASASLALKRLLYSGPAISSRSLIRTVPRMVSASRSFNTNAMRDYDDGNG
RSLDVDRRSDRSVPRGDRRDDLFPDVFDPFSPTRSLSQVLNLMDQFVDNP
FLSASRGIGAGGLRRGWDAKETEDALFLRVDMPGLGKDDVKVSVEQNTL
IIKGEGGKEEGEEDSARRYSSRIDLPEKLYKIDQIKAEMKNGVLKVVVPK
MKEEDRSDVYQVKVD

4B07

CTGAAAAGGTGAAAATCCCCCATACCAATCAATCAGGAGAGAGAAAA
ATGCAGGGTGTGGGAGTAACTTCAGTGATGGTTGCAGTCCAGTTTGCA
GAAGTGGGCGTGAACACTATGATTAAATCAGCAAACACCAATGGAAT
GAGCAACTTTGTTTACGTTGTTTATTCAAACTTCTTGGCCTTGTGTATC
CTCGTACCATCTTCCCTCCTCTATCACAGAAAGAAAGCTGCTCCTCCAC
TCACAGTCTCAATTGTGTGCAGAATGTTCCTTGTTGGCTTTTTCGCTTC
TTCTGCGCAGACTTTGATGTATACTGGAATTGGTTACAGTTCTCCAACT
CTGTCTTGTTCGCAGGTGGATCTTATTCCAGCTTTTACCTTCTTGATTGC
TGTTGTATCCAGGATGGAAAATCTGAATCCAAAACTAAAAGGCAGTCA
GGCAAAAATGGTTGGCACCGTGGTGTCTATCACAGGGGCATTAATAGT
GACATTATACAAAGGCCTGCCACTTACCGGTGCTTACCAGCTGCATGA
AACTTTGCGCTCTCCAAAATCAAATTGGCTGCTTGGAGGCCTTTTGCTT
GCAGTTGCTAGCTTAGGCAATTCATTTGTGCTCGTCATTCAGACTTGGA
TTATCAAGGAGTACCCTACAGAGTTAGTGGTAACGACCATTAGTTGTG
GTTTTGCGGTCATCGTCTCAACCATAATAGCAGTCATTGCTGAGAAAA
ATCCAAAAGCATGGTCACTCCATCTTGATATGGAGTTATTAGCCATTA
TATTTTGTGCGATTTTTGTGGTATCAATTCGAAGCGTTGTTCAAACATG
GGCCTGCCGCAAGAAAGGGCCTTTGTATGTCGCCATGTTTAATCCTCTT
GGGATGGTGATTGCAGTTGGCATGGGGTTCATATTTTTAGGGGATGCT
CTTTATCTTGGAAGCATTATTGGAGCTGGCGTAATAGCTGTTGGGTTTT
ATGCTGTGTTGTGGGGGAAGGCTCAAGAGGAAAAGATGTTAGATGAG
AAAAATGGAATGGGTAGTCTTCATTCATCTTCTTCTTCAACCCCCTTGT
TGCAGAACAAAAACATCGATGTGTAGAGAATTGTCAAGTTTTATTTTT
GGGACAATACAAATAAGTCCCCTGTAGTGTGTCTTAGTATCACATAAA
TACCCTTAGAATTTTAAAACCTGCAAAATAATCCTCTTAAAATTTGTAA
AAAAAAAAAAAAAAAAAAAAAAA

MQGVGVTSVMVAVQFAEVGVNTMIKSANTNGMSNFVYVVYSNFLALCI
LVPSSLLYHRKKAAPPLTVSIVCRMFLVGFFASSAQTLMYTGIGYSSPTLS
CSQVDLIPAFTFLIAVVSRMENLNPKLKGSQAKMVGTVVSITGALIVTLYK
GLPLTGAYQLHETLRSPKSNWLLGGLLLAVASLGNSFVLVIQTWIIKEYPT
ELVVTTISCGFAVIVSTIIAVIAEKNPKAWSLHLDMELLAIIFCAIFVVSIRSV
VQTWACRKKGPLYVAMFNPLGMVIAVGMGFIFLGDALYLGSIIGAGVIA
VGFYAVLWGKAQEEKMLDEKNGMGSLHSSSSSTPLLQNKNIDV

4H05

CATGGCTGCTGCATCGAGTAGAACTCGCGTTTGTAGCTGGTTGCTTCTT
TTCTCTGCAATTCTGGTTGTTTCTCACTTGCTTGTCACCCAGGCCGCAG
CAAGCGACGTTCCAATAGTAAATGGACTCTCATTTAGTTTCTATGACA
AAAGCTGTCCTAAGCTCGAATCCATAGTCAGAACCGAGCTCAAGAAA
ATTTTCAAGAAAGACATCGGCCAAACTGCTGGCTTGTTGAGGCTTCAC
TTCCATGACTGCTTTGTCCAGGGTTGTGATGGTTCGGTATTGTGTGGAT
GGATCTGCAAGTGGACCGAGTGAGAAAGATGCACCTCCAAACTTAAG
CCTAAGGGCTCAGGCCTTTAAGATCATCGATGATCTTCGGGCTCGAGT
TCACAAGCGCTGCGGAAGGATCGTCTCCTGTGCTGACATTACTGCCCT
TGCTGCTCGGGATTCTGTTTTCCTGTCAGGAGGGCCTGAGTACGACATT
CCACTAGGAAGGCGAGATGGGCTAACCTTTGCCACAAGAAACGTAAC
TTTGGCCAACCTTCCAGCACCATCCAGCACCGCCGCCGAAATCCTCGA
CTCTCTCGCCACCAAAAACTTGAACCCCACCGACGTCGTGGCCCTCTC
CGGCGGCCACACCATCGGAATCGGCCACTGCTCCTCCTTCACCAACAG
GCTCTTCCCACAAGATCCCGTCATGGACAAAACCTTCGCCAAAAATCT
CAAGCTAACTTGCCCCCACCAACACCACCGACAACACCACCGTGTTGG
ACATCCGCAGCCCCAACAAGTTCGACAACAAGTACTACGTTGATCTCA
TGAACCGACAGGGATTGTTCACTTCAGACCAAGATCTGTACACTGATA
AAAAGACGAGGGGAATTGTGACCAGCTTTGCTGTGAACCAGAGCTTGT
TCTTCGAGAAGTTCGTAGATGCCATGCTCAAGATGGGTCAGTTGAGCG
TGTTGACCGGAAGTCAAGGTGAAATTCGTGCCAATTGCTCCGTGAGAA
ACACTGATAATAAATCCATATTGACTTCAATAGTGGTAGATGTAGTGG
AAGAAATTCTGGAAATGTGAATCTGGAGCGTTTTCAAACCTAACACTT
TAATTGTATGCCTTTCAGCTTTTTGTTTCCAGTATTTGTGTTTGTAAAAT
GTATCTTTCACTTCAGATATCATAAATGAAAATAAAATAAATTTGATG
AAAATCAAGTTAGCGCAGAAAATCCTTGGGCATCCCACCTGGGCAGG
ATTGACTGGCAAGTGCCATTTTCCGTGGGCAGTAAAAAAAAAAAAAA
AAAAAAAAAA

MAAASSRTRVCSWLLLFSAILVVSHLLVTQAAASDVPIVNGLSFSFYDKS
CPKLESIVRTELKKIFKKDIGQTAGLLRLHFHDCFVQGCDGSVLCGWICK
WTEERCTSKLKPKGSGLDHRSSGSSSQALRKDRLLCHYCPCCSGFCFPVR
RAVRHSTRKARWANLCHKKRNFGQPSSTIQHRRRNPRLSRHQKLEPHRR
RGPLRRPHHRNRPLLLLHQQALPTRSRHGQNLRQKSQANLPPPTPPTTPPC
WTSAAPTSSTTSTTLIS

7A05

ATCAATGGCCTCTACAAGCTTCTCGAAAATGATCTCTTCTTCCCCTAAT
CTCTTCCATCCAAAATCAATCCCTTCTTCCCAAACTCCAAAGCAACTG
GGATCTACTTTCTTTCAGCCCTTCCGAACTCATGGTGAAAATTTCTCAA
GTACCCGTCAATTATCCACCACCAGAAGTTCCAAGAGTAGCACAACCA
CAGCATTCTTCTTCAATAAGCCCAAGGAGCTTCAAGATTCTCAAAAGC
CAGTTCAAGAATTGAGTGTCTATGAAATCAACGAACGCGATCGGGGA
AGCCCTGCATACTTGAGGCTGAGTCAAAAATCTACGAACTCTCTGGGA
GATCTAGTGCCATTCAGCAACAAGCTTTACTCAGGAGATTTAGAGAAG
AGGGTAGGTATAACAGCAGGCATATGTGTGCTCATCCAACACGTGCCT
GAGAAAAAAGGAGACAGATATGAAGCAATTTACAGCTTCTACTTTGG
AGACTATGGCCACATATCAGTGCAAGGACCTTATCTAACATACCAAGA
CACATATCTGGCAGTGACTGGAGGATCTGGAATCTTTGAAGGTGTTTA
TGGGAAGGTCAAACTCCAGGCAGATTGTCTTTTCCTTTCAAGTTGTTCT
ACACTTTTCTACTTGAAGGGGTATTCCTGATTTGCCTCCTGAGCTTACT
TGGAAAGCTAGTGGAGCCCATCNCCCAAGTTGCTGGAGCCCATCTGAA
GCTGGCCAAGGGGCNCCNCAGCCCTCCATGCCNGGCATCTCTTAAATT
TTCCCCGGAAANGAATTTGAANGGCNTTCCCATGGGTACNAGAATTTT
TGATNNNCAAAAATTAAAAAGGTTTTGGCGGGAAAAACTTTCANGTTT
TNGAATTNANCAATTCCGGGAAGAAGAAACAAGNAAATTTTCANAAT
TTCCTTGGGNTTTNCAAAAAAAAAAAAAAAA

MASTSFSKMISSSPNLFHPKSIPSSQTPKQLGSTFFQPFRTHGENFSSTRQLS
TTRSSKSSTTTAFFFNKPKELQDSQKPVQELSVYEINERDRGSPAYLRLSQ
KSTNSLGDLVPFSNKLYSGDLEKRVGITAGICVLIQHVPEKKGDRYEAIYS
FYFGDYGHISVQGPYLTYQDTYLAVTGGSGIFEGVYGKVKLQADCLFLSS
CSTLFYLKGYS