METABOLIC ENGINEERING OF *ARABIDOPSIS THALIANA* AND *CARICA PAPAYA* USING NATIVE PAPAYA PROMOTERS TO CONTROL STILBENE SYNTHASE GENE EXPRESSION

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI’I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

MOLECULAR BIOSCIENCES AND BIOENGINEERING

MAY 2015

By

Luzminda Carlos Hilario

Dissertation Committee:
David Christopher, Chairperson
Yun Judy Zhu
Richard Manshardt
John Hu
Anne Alvarez

Keywords: *Carica papaya* promoters, *Stilbene synthase* expression
ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost, I want to express my deepest and sincerest gratitude to my adviser, Dr. David A. Christopher, for his encouragement, guidance and support throughout the conduct of my research. His patience and constant motivation helped me push through for the completion of this work. I am also grateful for the unwavering support and guidance of my dissertation committee, Dr. Yun Judy Zhu, Dr. Richard Manshardt, Dr. John Hu, and Dr. Anne Alvarez.

I am heartily thankful to our graduate chair Dr. Jon-Paul Bingham for his countless help and assistance for the HPLC and MS/MS analyses. I’m also thankful to constant help and support from my colleagues at Dr. Christopher’s lab: Dr. Bradley Porter, Dr. Christen Yuen, Kristie Matsumoto, Dr. Pierriden Perez, Dr. Abby Cuttriss, Dr. Eun Ju Cho, Rick Shimshock, and Cherie Ng. I am also grateful for Dr. Heather McCafferty, Mr. Ray Uchida, Mr. Mel Nishijima, Ms. Tina Carvalho and Dr. Eden Perez for their support.

I wish to express my appreciation for the funding supported by the USDA awarded to Professor David A. Christopher and for Monsanto Research Fellowship.

I am sincerely grateful to my husband Pablo, my children Rhixelou Darleen and Ian, my parents and siblings, and friends for their endless inspiration, love and support.

Finally, to our almighty God for making all these things possible.
ABSTRACT

Papaya is an important fruit crop in the tropics, but it has inadequate resistance to the devastating pathogen, *Phytophthora palmivora* that causes root and stem rot. To develop disease resistance, papaya was engineered with the gene encoding grapevine stilbene synthase (*Vst1*), the key enzyme synthesizing the antimicrobial and antifungal phytoalexin, resveratrol. However, *Vst1* expression under its own promoter is suboptimal in papaya. Thus, to improve expression, four pathogen-induced native papaya promoters highly expressed in the roots and stems were used to control *Vst1* expression for resveratrol glucoside (piceid) synthesis. The promoters were isolated from genes encoding peroxidase (*Cp9*), beta-1,3-glucanase (*Cp29*) ferulate-5-hydroxylase (*Cp35*), and hypersensitive-induced response protein (*Cp45*). These promoters and the constitutive promoter, *CaMV35S*, were fused to the *eGFP* gene and transformed into the model plant, *Arabidopsis thaliana*. This allowed an efficient and rapid evaluation of promoter functionality, strength, and tissue expression. *In silico* analysis predicted the presence of several cis-regulatory promoter elements associated with stress and defense responses. The *Arabidopsis* transcriptional machinery readily recognized the promoters at the precise transcription initiation sites as used in papaya. Qualitative and quantitative *eGFP* measurements (fluorescence, mRNA and protein levels) indicated variations in tissue expression and promoter strength. Comparison of *eGFP* mRNA and protein levels indicated post-transcriptional regulation. Predicted mRNA 5′-UTR secondary structures potentially affected the translational efficiency of the mRNAs during plant development, most notably for *Cp9* and *Cp35*. Two differentially expressed promoters (*Cp9* and *Cp29*)
were fused to \( Vst1 \) and genetically transformed into \( Arabidopsis \). They synthesized variable piceid concentrations as detected through RP-HPLC analysis and verified through tandem mass spectrometry (MS/MS). Competition for precursor substrates (\( p \)-coumaroyl CoA, malonyl CoA) between chalcone synthase and stilbene synthase resulted in reduction of anthocyanin and seed tannin levels. A new transformation protocol was developed using suspension derived-cultures facilitated \( Agrobacterium \)-mediated transformation that generated several transgenic papaya lines. PCR, Southern Blot, RT-PCR, qRT-PCR, and RP-HPLC analyses confirmed independent transgene integration and copy number, \( Vst1 \) mRNAs driven by the native promoters, and promising basal piceid levels. Transgenic papaya plants are being regenerated from the calli that have potential disease resistance against \( Phytophthora palmivora \) and other papaya diseases.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>iv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>v</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiv</td>
</tr>
<tr>
<td>List of Supplementary Tables</td>
<td>xx</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xxi</td>
</tr>
</tbody>
</table>

CHAPTER I. INTRODUCTION AND LITERATURE REVIEW .......... 1

Introduction ......................................................... 1

Literature Review .................................................... 5

Biology of Carica papaya, L. ..................................... 5

Papaya as an important fruit commodity in the local and global market ... 6

Phytophthora palmivora, a devastating disease in papaya production ..... 7

Limited natural resistance in C. papaya against P. palmivora .......... 8

Use of phytoalexin resveratrol to control P. palmivora ............... 9

Phytoalexins from the Vitaceae .................................... 12

Stilbene synthase, the key enzyme in resveratrol synthesis .......... 13

Transformation of Vsti1 for disease resistance and nutritional improvement in crops ........................................ 17

Rationale for the use of native pathogen-inducible papaya promoters .... 19

Transformation of papaya with stilbene synthase using its own inducible promoter ........................................ 19
Papaya transcriptome survey to identify tissue-regulated genes for promoter isolation ................................................. 20

Selected native papaya promoters for functional characterization ........ 23

Related function of the genes selected for promoter isolation ........... 24

a.  Cp9, peroxidase ................................................................. 24

b.  Cp29, β-1,3-glucanase ........................................................ 24

c.  Cp35, ferulate-5-hydroxylase ............................................. 25

d.  Cp45, hypersensitive-induced response ................................. 25

CHAPTER II. SIGNIFICANCE OF RESEARCH, HYPOTHESES, AND OBJECTIVES ......................................................................................... 26

Significance of Research ................................................................................................................................. 26

Hypotheses .......................................................................................................................................................... 27

Objectives ......................................................................................................................................................... 29

CHAPTER III. IDENTIFICATION AND CHARACTERIZATION OF PAPAYA (CARICA PAPAYA, L.) PROMOTERS BY HETEROLOGOUS EXPRESSION AS EGFP FUSIONS IN ARABIDOPSIS THALIANA .......... 32

Abstract ......................................................................................................................................................... 32

Introduction .................................................................................................................................................... 33

Materials and Methods ................................................................................................................................... 35

Isolation of genomic regions and cloning of the putative papaya promoters ...... 35

Construction of binary vectors for transformation ................................................. 37

Plant material and growth condition for transformation ........................................ 37

RNA Isolation and RT-PCR ............................................................................................... 39

Fluorescence microscopy .............................................................................................................................. 41

Real-Time PCR analysis ............................................................................................................................... 41
Protein isolation and immunoblot analysis ................................................................. 42
Identifying the transcriptional start sites by 5’ RLM-RACE ................................. 42

Results ........................................................................................................................................ 44

Comparison of papaya promoter sequences to the cis-regulatory element
database ........................................................................................................................................ 44
Promoter activity during the developmental stages (7 to 28 DAP) as
assessed by Epi-fluorescence microscopy ........................................................................ 54

Semi-quantitative RT-PCR and Immunoblot Analysis to measure
eGFP expression ..................................................................................................................... 59
  a. Early Developmental Stage (7-28 DAP) ................................................................. 59
  b. Reproductive Stage (1.5 months) ........................................................................ 61

5’-RLM-RACE to Identify the Transcription Initiation Site (TSS) ..................... 64

Discussion ................................................................................................................................. 67

Conclusion ...................................................................................................................................... 71

CHAPTER IV. CARICA PAPAYA NATIVE PROMOTERS DRIVING THE
EXPRESSION OF STILBENE SYNTHASE IN ARABIDOPSIS
THALIANA FOR RESVERATROL GLUCOSIDE (PICEID)
SYNTHESIS .......................................................................................................................... 72

Abstract ...................................................................................................................................... 72

Introduction ................................................................................................................................... 73

Materials and Methods .......................................................................................................... 76

Vector Construction of pCambia 1302-papaya promoter-Vst1 ......................... 76
Transformation of Arabidopsis with Carica papaya promoter:Vst1 fusions .......... 77
Southern Blot analysis ............................................................................................................ 79
RNA isolation and real-time PCR analysis ................................................................. 80
RP-HPLC and tandem Mass Spectrometry (MS/MS) analysis of resveratrol content ................................................................. 81

Quantitation of anthocyanin of 5 day-old seedlings ....................... 82

Results ................................................................................................................. 83

Transgene integration and morphological development of the transgenic plants ........................................................................ 83

Comparison of Vst1 and CHS transcript levels ................................. 86

Biochemical characterization of resveratrol levels in transgenic Arabidopsis lines expressing Vst1 under the control of different promoters ......................... 90

Anthocyanin concentration, plant height and seed pigmentation in Arabidopsis transgenic plants ....................................................... 93

Discussion ....................................................................................................... 96

Conclusion ..................................................................................................... 100

CHAPTER V. MODIFIED AGROBACTERIUM-MEDIATED TRANSFORMATION OF CARICA PAPAYA VARIETY ‘KAPOHO’ FROM SUSPENSION-DERIVED CULTURES ............... 101

Abstract ......................................................................................................... 101

Introduction .................................................................................................... 102

Materials and Methods .............................................................................. 105

Vector construction ..................................................................................... 105

Preparation of papaya tissue cultures ....................................................... 105

Utilizing cell suspension-derived culture of papaya as alternative tissues transformation ................................................................. 106

Optimization of the modified Agrobacterium-mediated transformation protocol ...................................................................................... 107

PCR to check for the presence of the transgene ........................................ 107
Results

Agrobacterium-mediated transformation using the carborundum-wounded papaya embryogenic calli .................................................................................................................................................................................. 110

Suspension cultures of non-transformed papaya cultures capable of regeneration into plants ............................................................................................................................................................................ 112

Agrobacterium-based transformation of papaya using suspension cultured cells as target tissue ........................................................................................................................................................................ 115

Discussion .......................................................................................................................................................................................... 117

Conclusion ......................................................................................................................................................................................... 123

CHAPTER VI. DEVELOPMENT OF TRANSGENIC PAPAYA PLANTS USING NATIVE PROMOTERS TO CONTROL STILBENE SYNTHASE GENE EXPRESSION FOR SYNTEHZISING PHYTOALEXIN, RESVERATROL ......................................................................................................................... 124

Abstract .......................................................................................................................................................................................... 124

Introduction ....................................................................................................................................................................................... 125

Materials and Methods ........................................................................................................................................................................ 128

Vector construction of pCAMBIA 1302-papaya promoter-Vst1 ........................................... 128

Transformation of papaya with Carica papaya promoter:Vst1 fusions .......................... 129

Southern blot analysis ............................................................................................................................................................................ 130

RNA isolation and RT-PCR analysis ................................................................................................................................................. 131

Real-Time PCR analysis (qRT-PCR) ................................................................................................................................. 132

RP-HPLC and tandem mass spectrometry (MS/MS) analysis ............................................ 133

Results .......................................................................................................................................................................................... 134
Transgene integration, RT-PCR analysis, and somatic embryo development of the transgenic papaya cultures ......................................................... 134

Comparison of VstI and CHS transcript levels ........................................... 138

Biochemical characterization of transgenic papaya expressing VstI .............. 140

In vitro culture of transgenic papaya calli for plantlet regeneration ............. 150

Discussion .................................................................................................. 151

Conclusion .................................................................................................. 155

Future Directions ....................................................................................... 156

LITERATURE CITED .................................................................................. 157
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Carica papaya genes selected from characterized root expression for promoter isolation (Porter et al. 2009)</td>
<td>23</td>
</tr>
<tr>
<td>3.1 Primers used in cloning of papaya promoter regions</td>
<td>36</td>
</tr>
<tr>
<td>3.2 A comparison of the putative cis-elements of the Cp9, Cp29, Cp35 and Cp45 promoters by the PlantCARE Promoter Analysis software</td>
<td>50</td>
</tr>
<tr>
<td>3.3 Summary of the related functions of the putative cis-elements of Cp9, Cp29, Cp35 and Cp45 promoters as determined by the PlantCARE Promoter Analysis software</td>
<td>53</td>
</tr>
<tr>
<td>3.4 A comparison of the transcription start sites (TSS) and positions of the 5’-untranslated regions (UTR) of the four native papaya promoters expressed in Arabidopsis thaliana relative to Carica papaya</td>
<td>64</td>
</tr>
<tr>
<td>3.5 Carica papaya promoters calculated minimum free energy of the 5’ UTR sequences secondary structures obtained from RNAfold Webserver (<a href="http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi">http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi</a>)</td>
<td>66</td>
</tr>
<tr>
<td>4.1 Primers used in this study</td>
<td>78</td>
</tr>
<tr>
<td>4.2 Piceid concentration of Arabidopsis thaliana homozygous T3 transgenic plants</td>
<td>84</td>
</tr>
<tr>
<td>5.1 Optimization of parameters for modified Agrobacterium-mediated transformation protocol</td>
<td>116</td>
</tr>
<tr>
<td>6.1 Primers used in this study</td>
<td>130</td>
</tr>
<tr>
<td>6.2 Piceid concentration of transgenic papaya calli</td>
<td>148</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Reaction steps catalyzed by chalcone synthase and stilbene synthase. Cinnamoyl-CoA and p-coumaroyl-CoA are the common start substrates for chalcone synthase and stilbene synthase enzymes. Chalcones are usually converted to flavones spontaneously <em>in vitro</em> (Yu et al. 2006) .................................................. 4</td>
</tr>
<tr>
<td>1.2</td>
<td>A field of papaya showing fruit, stem and root infection with <em>Phytophthora palmivora</em> on Big Island, Hawaii ......................................................... 9</td>
</tr>
<tr>
<td>1.3</td>
<td>Phytoalexin synthesis from (a) secondary metabolism shikimate acid, acetate-malonate, and acetate- mevalonate pathways; (b) enzymatic steps for resveratrol synthesis ................................................................. 11</td>
</tr>
<tr>
<td>1.4</td>
<td>Stilbene compounds in <em>Vitaceae</em> ................................................. 13</td>
</tr>
<tr>
<td>1.5</td>
<td>Common substrates malonyl CoA and p-coumaroyl CoA that are present in plant cells are utilized by enzymes CHS and STS to produce chalcone and stilbene respectively. (a) STS enzymes cyclize the resulting tetraketide intermediate product via an intramolecular C2→C7 Aldol condensation, rather than C6→C1 Claisen condensation utilized by CHS .................................................. 15</td>
</tr>
<tr>
<td>1.6</td>
<td>Figure from Parage et al. (2012) on schematic representation of the grapevine ‘PN40024’ stilbene synthase gene clusters on chromosomes 10 and 16 .......... 16</td>
</tr>
<tr>
<td>1.7</td>
<td>Figure from Parage et al. (2012) on mapping of evolutionary contrasted amino acid sites on the three-dimensional model of a typical STS protein .............. 17</td>
</tr>
<tr>
<td>1.8</td>
<td>Figure from Zhu et al. (2004) on transformation of papaya with the enzyme stilbene synthase gene showing increased resistance ......................... 19</td>
</tr>
<tr>
<td>1.9</td>
<td>Figure from Porter et al. (2008) on quantified northern blot analysis of <em>Carica papaya</em> genes regulated by <em>Phytophthora palmivora</em> ............................. 22</td>
</tr>
<tr>
<td>1.10</td>
<td>Figure from Porter et al. (2009) on time course northern blot analysis of <em>Carica papaya</em> genes for <em>Phytophthora palmivora</em> regulation ..................... 22</td>
</tr>
<tr>
<td>3.1</td>
<td>Schematic representation of the transformation vector (pCAMBIA 1302) and the corresponding papaya promoter-eGFP gene fusions .............................. 38</td>
</tr>
<tr>
<td>3.2</td>
<td>Determination of the exponential phase of semi-quantitative RT-PCR of the eGFP and <em>Actin2</em> transcripts in transgenic <em>Arabidopsis thaliana</em> (35S-eGFP) construct</td>
</tr>
</tbody>
</table>
3.3 Agarose gel (1%) electrophoresis of PCR products from the four promoter regions ................................................................. 45

3.4 Nucleotide sequence of the 2041 bp *Carica papaya* *Cp9* peroxidase gene promoter ................................................................. 46

3.5 Nucleotide sequence of the 1445 bp *Carica papaya* *Cp29* B,1-3 glucanase gene promoter .......................................................... 47

3.6 Nucleotide sequence of the 1832 bp *Carica papaya* *Cp35* ferulate-5-hydroxylase gene promoter .................................................. 48

3.7 Nucleotide sequence of the 2045 bp *Carica papaya* *Cp45* hypersensitive-induced response protein gene promoter ........................................ 49

3.8 Graphical representation of the number of regulatory motifs predicted to reside in the native papaya promoters (*Cp9, Cp29, Cp35, Cp45*) .................... 53

3.9 Expression patterns of papaya promoter-eGFP fusions (*Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP*) in the shoot apex of *Arabidopsis* at 7 and 14 DAP ........................................................................ 55

3.10 Expression patterns of papaya promoter-eGFP fusions (*Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP*) in the cotyledon (7 DAP) and cotyledon petiole (14 DAP) of *Arabidopsis* ..................................................................... 56

3.11 Expression patterns of papaya promoter-eGFP fusions (*Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP*) in the hypocotyl of Arabidopsis at 7, 14 and 21 DAP ........................................................................... 57


3.13 Semi-quantitative RT-PCR analysis (30 cycles) of eGFP and Actin2 transcript whole seedlings ........................................................................ 60

3.14 Western-blot analysis of eGFP (27 KDa) and Actin2 (loading control, 45KDa) protein levels in three independent homozygous *Arabidopsis thaliana* T3 lines (one per lane) for papaya promoter:eGFP fusion (*Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP*) in four different stages of plant development .......... 61

3.15 Quantitative comparison of transcript levels in different *Arabidopsis thaliana*
tissues (flowers, leaves, stems, roots) at the reproductive stage (1.5 months old)

3.16 Western blot analysis of eGFP (27 KDa) and Actin2 (loading control, 45KDa) protein levels in three independent homozygous Arabidopsis thaliana T3 lines of papaya promoter:eGFP fusion (Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP) in four different tissues (flower, leaf, hypocotyl, root) of flowering stage plants (1.5 months old).

3.17 Nucleotide sequence of the 5’-UTR of the different promoters and the comparison of the transcription start sites (TSS) and positions of the 5’-untranslated regions (UTR) of the four native papaya promoters expressed in Arabidopsis thaliana relative to Carica papaya.

3.18 Predicted secondary structures of the 5’ UTR of the Carica papaya promoters.

4.1 Schematic representation of the transformation vector (pCAMBIA 1302) and the corresponding Carica papaya promoter:Vst1 gene fusions.

4.2 Southern blot analysis of EcoRI-digested genomic DNA extracted from Arabidopsis transgenic lines hybridized with 1.5 kb biotinylated stilbene synthase probe.

4.3 Quantitative comparison of CaMV35S transgenic lines (a) transcript levels and (b) piceid concentration (µg per gram fresh weight) as affected by the number of transgene insertion in the genome of the plants.

4.4 CaMV35S:Vst1 homozygous T3 Arabidopsis (14 days after planting) showing normal growth and morphological development of line 37 with single insertion of the transgene; representative lines 4 and 3 showing phenotypic alterations in shoot and root parts of the plants caused my increased number of the transgene insertion compared relative to the non-transformed wild type (WT).

4.5 Growth and morphological development of homozygous T3 Arabidopsis plants (14 days after planting) transformed with (a) CaMV35S:Vst1, transgenic line 1, 7 and 37 (left to right); (b) Cp9:Vst1, transgenic line 15, 16, and 18; (c) Cp29:Vst1 transgenic line 23, 24 and 27 and (d) WT- non-transformed wild type control plants.

4.6 Morphological development of homozygous T3 Arabidopsis plants one month after planting transformed with (a) CaMV35S:Vst1, transgenic line 1, 7 and 37 (left to right); (b) Cp9:Vst1, transgenic line 15, 16, and 18; (c) Cp29:Vst1 transgenic line 23, 24 and 27 and (d) WT- non-transformed wild type control plants.
4.7 Quantitative comparison of transcript levels in wild-type and different lines of *Arabidopsis thaliana*. (a) *Vst1* transcripts (b) *CHS* transcripts ...................... 89

4.8 HPLC analysis of extracts from seedlings (28 days after planting) of transgenic *Arabidopsis* plants showing the trans-piceid peak at 7.38 minute retention time compared to the commercial standard trans-piceid (red line) also known as polydatin (red). Individual peaks of the extracts isolated from WT and lines containing different promoter-*Vst1* are shown (black lines) ....................... 91

4.9 *Trans*-piceid mass spectra verified through tandem MS-MS with the corresponding fragmentation coefficient (m/z) of the compound. (a) Sample from *Arabidopsis Cp29*-line 23 (b) Standard piceid ................................................................. 92

4.10 Anthocyanin concentrations of 5-day old seedlings measured at 530 nm absorbance (A_{530}) per gram fresh weight of tissue (a); Plant height of the transgenic *Arabidopsis* at maturity measured from rosette base to flower apex (b). ........................................................................................................... 94

4.11 Differences in the changes of seed coat pigmentation of *Arabidopsis* seeds transformed with the *Carica papaya* promoter:*Vst1* constructs with (a) CaMV35S :*Vst1*, transgenic line 1, 7 and 37 (left to right); (b) *Cp9*:*Vst1*, transgenic line 15, 16, and 18; (c) *Cp29*:*Vst1* transgenic line 23, 24 and 27 and (d) WT- non-transformed wild type control plants .................................................. 95

5.1. Schematic representation of the transformation vector (pCAMBIA 1302) and the corresponding papaya *Cp45* promoter-eGFP gene fusion ......................... 105

5.2. *Agrobacterium*-mediated transformation using carborundum-wounded calli ... 111

5.3. Establishment of suspension culture ......................................................... 113

5.4. Optimization of growth parameters for *in vitro* culture and plantlet regeneration of papaya derived from suspension culture of the non-transformed wild type (WT) plants ................................................................. 114

5.5 Development of modified *Agrobacterium*-mediated transformation protocol from suspension-derived culture .................................................. 118

5.6 *Cp45*:eGFP putative transgenic papaya cultures generated from the modified *Agrobacterium*-mediated transformation protocol with verified eGFP expression viewed under Kodak phosphorimagere ................................. 119

5.8 *Cp45:eGFP* transgenic papaya cultures generated from the modified *Agrobacterium*-mediated transformation protocol with verified *eGFP* expression showing fluorescing stems viewed under (a) white light (a); blue light (b); Kodak phosphorimager (c) ................................................................. 120

5.9 *Cp45:eGFP* transgenic papaya cultures generated from the modified *Agrobacterium*-mediated transformation protocol with verified *eGFP* expression showing fluorescing roots (a) and stem (b) viewed under hand held blue light and Kodak phosphorimager ................................................................. 121

6.1 Hygromycin-resistant putative transgenic calli growing one month after *Agrobacterium* transformation (a); Highly proliferating somatic embryos under hygromycin (50 mg/L) selection showing normal growth and development relative to the non-transformed wild-type (WT) calli (b) ........................................ 135

6.2 Molecular analyses of the transgenic papaya calli: PCR check for the presence of the *Vst1* gene and hygromycin resistance gene (*hptII*) (a); RT-PCR analysis for the detection on the differences of *Vst1* transcript expression as driven by the different native papaya promoters .......................................................... 136

6.3 Southern blot analysis of EcoRI-digested genomic DNA extracted from *Carica papaya* transgenic calli lines hybridized with 1.5 kb biotinylated stilbene synthase probe .................................................................................................................. 138

6.4 Quantitative comparison of (a) Vst1 and (b) CHS transcript levels driven by the different promoters in the transgenic papaya calli ........................................ 141

6.5 Piceid concentration (µg per gram fresh weight) of the different transgenic papaya calli lines ......................................................................................... 142

6.6 Wild-type (WT) non transformed papaya: HPLC analysis (a); normal growth and development of the proliferating somatic embryos (b) ......................... 143

6.7 Representative HPLC chromatogram peak of the transgenic *CaMV35S:Vst1* papaya calli (a); growth and development of the proliferating transgenic somatic embryos in *CaMV35S* line 1, 2 and 3 (b) ......................................................... 144

6.8 Representative HPLC chromatogram peak of the transgenic *Cp9:Vst1* papaya calli (a); growth and development of the regenerating transgenic somatic embryos showing root and shoot development in *Cp9* line 1 (b) ................................. 145

6.9 Representative HPLC chromatogram peak of the transgenic *Cp35:Vst1* papaya calli (a); growth and development of the proliferating transgenic somatic embryos in *Cp35* line 1, 4, 5 and 8 (b) ................................................................. 146
6.10 Representative HPLC chromatogram peak of the transgenic \textit{Cp45}:\textit{Vst1} papaya calli (a); growth and development of the proliferating transgenic somatic embryos in \textit{Cp45} line 1, 2, 3 and 4 (b) .......................................................... 147

6.11 \textit{Trans}-piceid mass spectra verified through tandem MS-MS with the corresponding fragmentation coefficient (m/z) of the compound. (a) Sample from \textit{Carica papaya} \textit{Cp45} line 5 (b) Standard piceid ................................. 149

6.12 Developed transgenic papaya lines transformed with \textit{Carica papaya} promoters:\textit{Vst1} fusions growing under different stages of somatic embryo development .................................................................................................................................................. 151
LIST OF SUPPLEMENTARY TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>179</td>
<td>Stilbene synthase genes and promoters used to genetically transform plants, and the resulting effects on stilbene levels, resistance to pathogens and antioxidant activities (Delaunois et al. 2009, Giovinazzo et al. 2012)</td>
</tr>
<tr>
<td>2</td>
<td>181</td>
<td>Putative cis-acting regulatory motifs in <em>Cp9</em> (peroxidase) papaya promoter. Putative motifs recognized by transcription factors (TF) were found using PlantCARE program (Selected Matrix Score for all elements is ≥ 5)</td>
</tr>
<tr>
<td>3</td>
<td>184</td>
<td>Putative cis-acting regulatory motifs in <em>Cp29</em> (B-1,3-glucanase) papaya promoter. Putative motifs recognized by transcription factors (TF) were found using PlantCARE program (Selected Matrix Score for all elements is ≥ 5)</td>
</tr>
<tr>
<td>4</td>
<td>186</td>
<td>Putative cis-acting regulatory motifs in <em>Cp35</em> (ferulate-5-hydroxylase) papaya promoter. Putative motifs recognized by transcription factors (TF) were found using PlantCARE program (Selected Matrix Score for all elements is ≥ 5)</td>
</tr>
<tr>
<td>5</td>
<td>188</td>
<td>Putative cis-acting regulatory motifs in <em>Cp45</em> (hypersensitive-induced response) papaya promoter. Putative motifs recognized by transcription factors (TF) were found using PlantCARE program (Selected Matrix Score for all elements is ≥ 5)</td>
</tr>
<tr>
<td>6</td>
<td>190</td>
<td>Comparison of the putative cis-elements of <em>Cp9</em>, <em>Cp29</em>, <em>Cp35</em> and <em>Cp45</em> promoters by PLACE Promoter Analysis</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

2,4-D - 2,4-dichlorophenoxy acetic acid

35S - Cauliflower mosaic virus (CaMV) 35S promoter

5’ RLM-RACE - 5’ RNA Ligase-Mediated-Rapid Amplification of cDNA ends

5’ UTR - 5’ Untranslated Region

aa - amino acid

*Arabidopsis* - *Arabidopsis thaliana* (var. Columbia)

ASGPB - Advanced Studies in Genomics, Proteomics and Bioinformatics

At - *Arabidopsis thaliana*

bp - base pair

CIM - Callus induction medium

cm - centimeter

cDNA - complementary DNA

*CHS* - chalcone synthase

*Cp* - *Carica papaya*

*Cp9* - Peroxidase

*Cp29* - βeta-1,3-glucanase

*Cp35* - Ferulate-5-hydroxylase

*Cp45* - Hypersensitive-induced response protein

CTAB - Cetyl trimethyl ammonium bromide

DAP - days after planting

DNA - deoxyribonucleic acid

dNTP - deoxyribonucleotide triphosphate
E. coli - Escherichia coli

eGFP - enhanced Green Fluorescent Protein
EDTA - ethylenediaminetetraacetic acid
GFP - Green Fluorescent Protein
g - gram
GFP - Green Fluorescent Protein
GMO - genetically modified organism
HPLC - High-performance liquid chromatography
hptII - hygromycin phosphotransferase
hr - hour
IPTG - Isopropyl-β-D-thiogalactoside
kb - kilo base pair
kDa - kilo Dalton
L - liter
LB - Luria Bertani medium
Mbp - mega base pair
mg - milligram
min - minute
mL - milliliter
mm - millimeter
M-MLV RT - Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT)
MS - Murashige and Skoog medium
MW - molecular weight marker
ng - nanogram
nm - nanometer
nt - nucleotide
nos - nopaline synthase
OD - optical density
ORF - Open Reading Frame
PCD - programmed cell death
PCR - polymerase chain reaction
Pfam - protein family
P. palmivora - Phytophthora palmivora
PR - pathogenesis-related
PRV - Papaya ring spot virus
qPCR - quantitative Reverse-Transcription Polymerase Chain Reaction
RNA - ribonucleic acid
RP - HPLC Reverse-Phase High Performance Liquid Chromatography
rpm - revolutions per minute
RT - room temperature
PCR - Polymerase Chain Reaction
RT-PCR - Reverse Transcription Polymerase Chain Reaction
SAM - Shoot apical meristem
SDS - Sodium dodecyl sulfate
SDS-PAGE - Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
STS - stilbene synthase
TF - Transcription Factor

TSS - Transcription start site

UTR - Untranslated Region

UV-Vis - Ultraviolet-visible spectrophotometry

W.T. or WT - wild type

µg - microgram

µL or uL - microliter

µM or uM - micrometer

Vst1 - Stilbene synthase
CHAPTER I

INTRODUCTION & LITERATURE REVIEW

Introduction

Metabolic engineering is generally defined as the direction of one or more enzymatic reactions to improve the production of existing compounds, produce new compounds, or mediate the degradation of undesirable compounds (Giovinozzo et al. 2012). Plant metabolic engineering has provided a method to improve polyphenol composition such as stilbenoids in crops where promising results have been obtained with stilbene synthase encoding genes in transgenic plants leading to novel functional food for human consumption (Delaunois et al. 2009, Giovinozzo et al. 2012). During the past two decades, stilbene synthase have been engineered to a number of crops either to improve the resistance to plant abiotic or abiotic stresses or increase its nutritional value (Appendix Table 1.1).

The choice of promoter to drive the transgene primarily controls the level of gene expression. Previous studies showed that stilbene synthase encoding genes for plant transformation have been expressed under the control of a limited number of promoters such as constitutive promoter CaMV35S, its own stress-responsive promoter Vst1, and the fungus-inducible promoter PR10.1 or the tissue-specific promoter p-nap (Coutos-Thevenot et al. 2001, Zhu et al, 2004, Fan et al. 2008, Husken et al. 2005). Traditionally, most of the genetic transformation work utilized the conventional CaMV35S promoter for transgene expression. However, as in the case of tobacco, the use of strong and
constitutive CaMV35S promoter fused to stilbene synthase triggered stilbene accumulation and caused a drastic depletion of the endogenous pools of precursors (\(p\)-coumaroyl-CoA and malonyl-CoA) leading to physiological anomalies such as male sterility (Fischer et al. 1997). To avoid this condition, the choice of optimal and regulatable promoters for metabolic engineering in plants is highly desirable.

Previously, the expression of the grapevine stilbene synthase gene in papaya enhanced resistance to Phytophthora palmivora (Zhu et al. 2004) but the expression of grapevine \(Vst1\) gene driven by its own native promoter is less than optimal in papaya. The stilbene synthase mRNA was sub-optimal in papaya. This led to the study of identification and characterization of isolated native papaya promoters from pathogen inducible genes of papaya to regulate expression of stilbene synthase in metabolic engineering of papaya for resveratrol production. The primary tissues targeted by the pathogen are roots and stems especially at the seedling stage. Therefore, we aim to express stilbene synthase gene fused to native papaya promoters to control its expression for root and stem tissue-enhanced production of the phytoalexin, resveratrol. Stilbene synthase, which is not present in papaya, is a key enzyme that controls production of potent antimicrobial and antifungal phytoalexin, resveratrol. The precursor molecules for resveratrol synthesis (one molecule of \(p\)-coumaryl-CoA and three molecules of malonyl-CoA) are ubiquitous in plant cells, and adding stilbene synthase enzyme is sufficient to divert phenylpropanoid pathway into stilbenoids instead of flavonoids synthesis (Figure 1.1) (Yu et al. 2006).

Four putative promoter regions and their 5’UTRs were isolated from the genes encoding peroxidase (Cp9), \(\beta\)-1,3-glucanase (Cp29), ferulate-5-hydroxylase (Cp35) and
hypersensitive-induced response protein (Cp45) and were fused to eGFP reporter gene to easily characterize the nature of expression and tissue localization in the model system Arabidopsis thaliana. Next, we further characterized the activity of the promoters fused to stilbene synthase in Arabidopsis thaliana to understand the morphological or metabolic effects in the transformed plants as a result of adding a new enzyme in the secondary metabolic pathway both at the developmental and reproductive stages. Finally, we transformed papaya with stilbene synthase using our own promoters to assess the regulatory activity of these promoters and evaluate the effects of newly synthesized compound resveratrol on the overall morphological growth and development of developed transgenic plants.

We aim to develop transgenic plants with controlled stilbene synthase expression thereby regulating resveratrol synthesis in targeted tissues without altering the normal phenotypic growth of the plants. We hope to provide transgenic papaya with desired levels of resveratrol that can be used for future disease resistance induction experiments for papaya crop improvement.
Figure 1.1. Reaction steps catalyzed by chalcone synthase and stilbene synthase. Cinnamoyl-CoA and \( p \)-coumaroyl-CoA are the common start substrates for chalcone synthase and stilbene synthase enzymes. Chalcones are usually converted to flavones spontaneously \textit{in vitro} (Yu et al. 2006).
Literature Review

**Biology of *Carica papaya, L.***

Papaya is usually single stemmed, semi-woody giant herb with fast indeterminate growth (1-3 m during the first year) that develops very fast (3-8 months) from seed germination to flowering (juvenile stage) and 9-15 months for harvest (Paterson et al. 2008). The plant can live up to 20 years, however due to excessive plant height and pathological constraints in the field, the commercial life of papaya orchard is 2-3 years. Papaya plants have large palmate leaves (0.6 m²) with five to nine pinnate lobes of various widths (40-60 cm in a spiral pattern and clustered in an upper section of adult individuals (Morton 1987, Ming et al. 2008).

Papaya has a relatively small genome of 372 Mbp with nine pairs of chromosomes (Arumuganathan and Earle 1991). Ming et al. (2007) studied the three sex forms of papaya which are female, male and hermaphrodite which is regulated by an incipient X-Y chromosome system. Papayas can either be dioecious (with male and female plants) or gynodioecious (hermaphrodite and female plants). Several studies suggest that Y chromosome contains small specific region that controls expression of male (Y) or hermaphrodite (Y^h) types while females are of XX form. All combinations among the Y and/or Y^h are lethal, therefore the male and hermaphrodite types are heterozygous (XY and XY^h) (Ming et al. 2007).

*Caricaceae* includes around 35 species in 6 genera and have a disjunct distribution between Africa and the Neotropics (Carvalho & Renner 2013). The family’s classification has been revised to compromise African *Cylicomopha* and five South and
Central American genera (*Carica, Jacaratia, Jarilla, Horovitzia and Vasconcella*) with *Carica papaya* as the only species within the genus *Carica* (Badillo 2000). A recent taxonomic revision proposed that some species formerly assigned to *Carica* were more appropriately classified in the genus *Vasconcella* (Badillo 2000) based on morphological and genetic analysis (Van Droogenbroeck et al. 2004).

The origin of papaya based on herbarium specimen record and existence of local population supports the hypothesis that South Mexico and Central America are the centers of its origin (Fuentes & Stamaria 2014). Papaya is now grown in all tropical countries and many sub-tropical regions of the world and has been grown in Hawaii for over a century (Gonzalves 1998).

**Papaya as an important fruit commodity in the local and global market**

Papaya is a widely cultivated and important tropical fruit commodity in the global market where current worldwide production has reached 10.5 metric tons with a value of 3.4 billion US dollars in 2009 (Fuentes & Santamaria 2014). Ripe papaya fruits are commonly eaten as fresh produce while the unripe fruits are rich source of papain which was first used as a commercially available meat tenderizer (FAO 2005). Papaya is also utilized in the pharmaceutical and cosmetics industries. In total world production, 52% mainly comes from India and Indonesia, 34% comes from America (Brazil and Mexico) and the remaining 14% is from Africa (Fuentes & Stamaria 2014).

In Hawaii, papaya ranks the second largest fruit industry that earned 11 million US dollars in 2011 which mainly exports its products to the US mainland. The deregulation of transgenic papaya with resistance to PRV (Papaya ringspot virus) led to
the expansion of the fresh papaya fruits to the Japanese market which significantly increased the demand for papaya production. Although papaya mainly grown in Big Island are resistant to PRV, most of the papaya solo varieties such as the solo ‘Kapoho’ (preferred commercial variety sold in the market) planted in the area are still susceptible to the devastating oomycete disease, *Phytophthora palmivora*.

**Phytophthora palmivora infection, a devastating disease in papaya production**

There are at least 8 oomycete species of *Phytophthora* that infect more than 1,000 plant species causing huge economic losses each year worldwide (Erwin and Ribeiro 1996). *Phytophthora infestans* caused the Irish potato famine and is currently responsible for multibillion-dollar losses in the production of potatoes and tomatoes (Fry and Goodwin 1997 a,b). *Phytophthora sojae* causes root and stem rot of soybean, hampering its production in several continents.

The pathogen, *Phytophthora palmivora*, threatens papaya production by inhibiting transport of water from the roots, ultimately, killing the plants, which reduces productivity and fruit quality (Nishijima 1994). *Phytophthora* species are oomycetes, which are more related to photosynthetic algae, but share many morphological features with fungi including hyphae and spore production (Lamour et al. 2007). *Phytophthora palmivora* and *Phytophthora megakarya* cause black pod of cacao, a recurring threat to chocolate production.

In papaya, the estimated annual losses range from 10-30% on most farms where growers rely on high fungicide (>100 lbs/acre/year) application for its control (Nishijima, 1994, 2002). The fungicide application leads to increased production costs and
contamination of the environment. An alternative to fungicides would significantly benefit the papaya industry and sustainable agriculture.

In the attempt to control *Phytophthora infestans*, infestation (late blight) in potato, several studies were conducted to elucidate the mechanism of resistance to this devastating plant disease (Song et al. 2003, van der Vossen et al. 2003, van der Vossen et al. 2005, Huang et al. 2005). These studies provided insight into the molecular basis of plant-*Phytophthora* interactions that lead to the understanding of plant resistance genes and pathogen avirulence genes. Recently, the *RB* gene cloned from *Solanum bulbocastanum*, a wild diploid potato species, was found to confer a broad spectrum resistance to potato late blight (Song et al. 2003). It is hoped that through genetic engineering, the current popular potato varieties will be rendered resistant to late blight using the cloned *RB* gene.

**Limited natural resistance of *C. papaya* against *P. palmivora***

*Carica papaya* is known to have limited natural resistance to a range of oomycete pathogens (Nishijima 1994). The disease is difficult to manage and breeding for resistant papaya against *Phytophthora* is limited by the absence of resistance genes in the *Carica* family. Attempts to hybridize wild relative *Vasconcellea quercifolia* to incorporate resistance genes into cultivated *Carica papaya* had also been unsuccessful due to incompatibility and production of infertile hybrids (OECD 2005).

In the early 1990s, another devastating disease, papaya ring spot virus (PRV), caused an epidemic that nearly destroyed Hawaii’s papaya industry. The viral disease was controlled when the first genetically modified virus-resistant papayas cv. ‘SunUp’
and ‘Rainbow’ were obtained (Fitch et al. 1992). However, these transgenic varieties remained susceptible to *Phytophthora palmivora* that poses serious threats of yield loss. Papaya is susceptible to *Phytophthora palmivora* at the seedling and mature stages. Severe decline and death of papaya trees due to damping off and rotting of root, fruit, and stems in poorly drained areas has been attributed to *P. palmivora* (Nishijima 1994). In heavily infected fields, papaya farmers often abandon their plants causing more infection in the neighboring fields (Figure 1.2).

![Field of papaya showing fruit, stem and root infection with Phytophthora palmivora on Big Island, Hawaii.](image)

**Figure 1.2** A field of papaya showing fruit, stem and root infection with *Phytophthora palmivora* on Big Island, Hawaii.

**Use of phytoalexin resveratrol to control Phytophthora palmivora**

Phytoalexins are pathogen induced low molecular weight compounds with antimicrobial activities derived from secondary metabolism (Figure 1.3a) (Großkinsky et
Phytoalexin precursors arise from the three major biosynthetic pathways found in all plants; the shikimate acid, acetate-malonate, and acetate-mevalonate pathways, either from a single pathway or a combination of two or three (Figure 1.3a) (Kuc, 1995). This facilitates genetic engineering of plants with novel phytoalexin genes because the precursor molecules are already synthesized by the plant and insertion of a single gene, usually a phytoalexin synthase results in production of the desired phytoalexin such as resveratrol (Figure 1.3b).

Following their identification, phytoalexins were directly incorporated into the network of plant defense responses (Jeandet et al. 2002). An excellent review of Großkinsky et al. (2012) summarizes the recent advances in the understanding of phytoalexin function, biosynthesis and regulation, in combination with novel methods of molecular engineering as well as advances in instrumental analysis of phytoalexins as a potent target for improving crop protection. Based on this, molecular approaches of modulating inducible phytoalexins such as resveratrol to improve crop protection has been the focus of attention in crop protection studies in the past two decades (Großkinsky et al. 2012, Delaunois 2009).

Zhu et al. (2004) studied the use of the phytoalexin, resveratrol, to control *P. palmivora* in papaya. Phytoalexins were shown to be important natural components in the defense of plants against fungal infection (Jeandet et al. 2002, Kuc 1995). Resveratrol has potent antimicrobial and antifungal activity (Hammerschmidt 2004). Synthesis of the stilbene resveratrol (*trans*-3,4′,5-trihydroxy stilbene) was induced when grapevine and peanut were attacked by pathogens (Langcake and Pryce 1977). In grapevine, stilbene
Figure 1.3. Phytoalexin synthesis from (a) secondary metabolism shikimate acid, acetate-malonate, and acetate- mevalonate pathway; (b) enzymatic steps for resveratrol synthesis.
synthesis is induced by inoculation with the pathogens, *Botrytis cinerea* or *Plasmopara viticola* (Blaich & Bachmann 1980; Langcake 1981). Dercks and Creasy (1989) observed that the subsequent level of resistance to *P. viticola* was positively correlated with the capacity of grapevine to synthesize stilbene. Recently, it was also shown that stilbene synthase and chitinase mRNAs were induced in the tolerant cultivars of Florida hybrid bunch grape when infected with *Elsinoe ampelina* (Vasanthaiah, et al. 2010).

**Phytoalexins from the Vitaceae**

Phytoalexins produced from *Vitaceae* belong to the stilbene family composed of 14-carbon phenolic molecules based on a *trans*-resveratrol skeleton (Dercks et al. 1995, Hart 1981, Jeandet et al. 2002). These include the predominant stilbene product resveratrol (*trans*-4,3′,5′-trihydroxystilbene), two oligomers α-viniferin and Ε-viniferin produced in the next highest amounts, and pterostilbene (3,5-dimethoxy-4′-hydroxystilbene) produced in very small quantity (Figure 1.4) (Dercks et al. 1995). Oxidation of Ε-viniferin resulted to formation of resveratrol isomer δ-viniferin, one of the major stilbenes produced in grapes (Figure 1.4) (Pezet et al. 2003). Stilbenes can exist as both *trans*- and *cis*- isomers which have different chemical characteristics and biological activities where *trans*- isomers are usually more stable and tend to dominate in plant tissues (Hart 1981). Resveratrol can be glycosylated forming piceid (5,4′-dihydroxystilbene-3-O-β-D-glucopyranoside) (Jeandet et al. 2002). Stilbenes emit a characteristic bright blue fluorescence under ultraviolet (365 nm) light (Langcake & Pryce 1976, Hart 1981, Dercks et al. 1995, Jeandet et al. 2002).
**Stilbene synthase, the key enzyme in resveratrol synthesis**

Resveratrol is synthesized by the enzyme trihydroxystilbene synthase (stilbene synthase or resveratrol synthase) (EC 2.3.1.95) using as substrates one molecule of 4-coumaroyl-CoA and three molecules of malonyl-CoA (Figure 1.5a) (Rupprich and Kindl 1978). Following the synthesis of resveratrol, conjugation to sugars or other derivatizations may occur, including methylation, prenylation, or condensation reactions to form species specific dimers, trimers, or tetramers (Sotheeswaran & Pasupathy 1993). Stilbene synthase (STS or Vst1) is related to chalcone synthases (CHS) that catalyze the first
committed step of flavonoid metabolism by the condensation of substrates through a polyketide intermediate (Figure 1.5a) (Lanz et al. 1990). STSs are members of type III polyketide synthases family and chalcone synthases (CHSs) are the most ubiquitous polyketide synthase in plants (Parage et al. 2012). STS and CHS enzymes share 75-90% amino acid sequence identity over their ~400 residues, and both enzyme families catalyze the same iterative condensation of three acetyl units (derived from decarboxylation of malonyl CoA) to a CoA-tethered phenylpropanoid starter molecule (derived from phenylalanine), most typically p-coumaroyl-CoA (Austin et al. 2004). However, STS enzymes cyclize the resulting tetraketide intermediate product via an intramolecular C2→C7 aldol condensation, rather than C6→C1 Claisen condensation utilized by CHS (Figure 1.5) (Austin et al. 2004). Comparison of STS and CHS genes revealed that STS genes do not form a separate cluster, but are instead most closely related to the CHS from the same or related plant species (Tropf et al. 1994). Site-directed mutagenesis of as few as three amino acids was sufficient to convert CHS enzyme activity to STS that suggests that STS genes most likely evolved independently in each plant species from an endogenous CHS (Tropf et al. 1994).

STS gene usually forms small families of two to five closely related paralogs. However, recent reannotation of grapevine genome (cv PN40024) by Parage et al. (2012) found a high number of STS yielding 48 STS, of which 32 are potentially functional genes encoding for proteins with STS activity (Figure 1.6). From their evolutionary analysis, they found that both STS and CHS evolution are dominated by purifying selection, with no evidence for strong selection for new functions among STS genes. They have observed that selection for increased dosage may have led to the amplification of STS family where
increasing dosage can be obtained through evolution of enhancers that will increase expression levels as well as by duplicating a gene over and over. Vannozzi et al. (2012) conducted microarray and RNA-seq analysis of STS genes in different physiological conditions and found that STS genes located on chromosome 10 were likely to be involved in constitutive and developmentally regulated stilbene biosynthesis while stress-

![Diagram](image)

**Figure 1.5.** Common substrates malonyl CoA and p-coumaroyl CoA that are present in plant cells are utilized by enzymes CHS and STS to produce chalcone and stilbene respectively. (a) STS enzymes cyclize the resulting tetraketide intermediate product via an intramolecular C2→C7 aldol condensation, rather than C6→C1 Claisen condensation utilized by CHS (Noel et al. 2004).
induced stilbene synthesis depending on the gene cluster is located on chromosome 16 (Figure 1.6).

![Chromosome 10 STS cluster (14216000-14307000)](image)

**Figure 1.6.** Figure from Parage et al. (2012) on schematic representation of the grapevine ‘PN40024’ stilbene synthase gene clusters on chromosomes 10 and 16.

The Fitmodel analysis of Parage et al. (2012) allowed the identification of one remarkable site that could be related to the evolution of STS activity in both grapevine and other species. Their analysis revealed that the site Pro-269 (Figure 1.7) exhibits a pattern of positive selection on the branch at the base of the lineage and purifying selection elsewhere and this pattern is associated with the transition from CHS to STS. Austin et al. 2004 also found that Pro-269 in grapevine corresponds to Gly-272 in pine STS, which belongs to one of the two regions, whose conformation differ substantially between CHS and STS crystal structures which suggests that this site is likely to have played a major role in STS evolution. Val-230 was found under widespread positive selection within STS family which is located at the periphery of the enzyme while other amino acids subjected to strong purifying selection pressures were Lys-14, Ser-231, and
Asn-392 (Figure 1.7). These regions could be involved in the interaction with other protein partners which have shown physical interaction and channeling of intermediates between enzymes operating sequentially in the flavonoid synthesis (Winkel 2004). It is speculated that the peripheral amino acids subjected to strong purifying selection in STS may be critical for protein-protein interactions within these complexes and for efficient stilbene synthesis interaction (Parage et al. 2012).

**Figure 1.7.** Figure from Parage et al. (2012) on mapping of evolutionary contrasted amino acid sites on the three-dimensional model of a typical STS protein. VvSTS10 protein was modeled using the structure of STS from Scots pine as a template (referred from Austin et al. 2004). In both STS monomers, remarkable amino acids are highlighted in red or blue, for amino acids subjected to positive or purifying selection, respectively. Pro-269, subjected to early positive selection in the grapevine STS family, is represented in purple. The position of the resveratrol product (R) is indicated.

**Transformation of Vst1 for disease resistance and nutritional improvement in crops**

Stilbene synthase (Vst1) genes isolated from grapevine have been transformed into various crop species that results in increased resistance to different pathogens or
improve its nutritional content which was summarized by excellent reviews of Delaunois et al. (2009) and Giovinazzo et al. (2012).

Numerous reports on the transformation with stilbene synthase or its combination (e.g. \textit{Vst1} and \textit{Vst2}) genes from \textit{V. vinifera} with either pathogen-inducible or constitutive promoters resulted in increased disease resistance in a number of crops (Appendix Table 1.1). Stilbene synthase genes have also been used to create novel phenotypes or to increase the nutritive value of a crop. Altered flower color of tobacco plants using duplicated \textit{CaMV35S} promoter was developed in tobacco indicating that stilbene synthase genes have potential for use in engineering cultivars with altered flower color (Fischer et al. 1997). Male sterility in tobacco plants was developed by directing expression of \textit{Vst1} to the anthers through use of a tapetum-specific promoter, a technique that has potential for development of a novel hybrid system (Fischer et al. 1997). Improvement of the antioxidant of tomato fruit was developed by Giovinazzo et al. (2005) using \textit{Vst1} and \textit{Vst2} under control of the \textit{CaMV355} promoter. Transgenic tomato plants were shown to accumulate resveratrol and resveratrol-glucopyranoside, as well as having increased levels of two naturally present antioxidant compounds, ascorbate and glutathione (Giovinazzo et al., 2005). Husken et al. (2005) transformed oilseed rape with the \textit{Vst1} gene under control of the seed-specific napin promoter to increase the nutrient value of the seeds. The new metabolic sink in oilseed rape with the addition of the \textit{Vst1} gene diverted the substrate from production of sinapate esters which reduce the quality of rapeseed meal to synthesis of piceid, which is considered beneficial to human health (Husken et al. 2005).
Rationale for the Use of Native Pathogen-Inducible Papaya Promoters

Transformation of papaya with stilbene synthase using its own inducible promoter

The stilbene synthase gene \( VstI \) from grapevine under the control of its own inducible promoter (Hain et al. 1993) was transformed into papaya that increased resistance of papaya to \( P.\ palmivora \) both \textit{in vitro} and \textit{in planta} (Zhu, et al. 2004) (Figure 1.8). However, the grapevine gene promoter is expressed later, and throughout the plants, which is less than optimal in papaya. It is proposed that greater resistance can be obtained by placing the stilbene synthase gene under the control of a strong papaya promoter, specifically expressed in roots and stems, which are the tissues targeted by the pathogen.

\textbf{Figure 1.8.} Figure from Zhu et al. (2004) on the transformation of papaya with the enzyme stilbene synthase gene showing increased resistance: (a & b) \textit{in vitro} leaf assay; (c) resistance shown in planta; (d) Northern blot showing stilbene synthase mRNA accumulation in leaves of a transgenic papaya line (Vst7) at 0, 5, 24, 48, and 72 hour post-inoculation with spores of \( P.\ palmivora \).
To date, most of the transformation done in papaya used the traditional promoter cauliflower mosaic virus (CaMV35S) (Odell et al. 1985). Consumers and regulatory agencies view the use of this promoter negatively because of the arbitrary increase of the transgene on the consumed portion of the crop leading to higher costs for biosafety trials and poor sales. Moreover, other periodic problems that might occur include transgene silencing (Flavell 1994, De Wilde et al. 2000, Halpin et al. 2001) and developmental abnormalities, such as male sterility (Fisher et al. 1997, Hain et al. 1993, Bornke et al. 2002). Using new native promoters in targeted tissues that allow resveratrol to be expressed before pathogen infection is hypothesized to protect the plant from pathogen establishment. In addition, using promoters that are activated and not repressed by the pathogen is hypothesized to provide an additional level of resistance. Therefore, the promoters should only be expressed in the specific tissues needed in the root and stem, and will maintain their expression upon infection to effectively increase resistance of papaya to \( P. \) palmivora, while improving public acceptance of GM papaya.

**Papaya transcriptome survey to identify tissue-regulated genes for promoter isolation**

To elucidate the differentially expressed genes from a Carica papaya root, Porter et al. (2008) conducted a transcriptome survey to identify tissue-regulated genes for promoter isolation. Their findings identified cDNAs with homology based on functional classification, for genes associated with defense, beneficial plant-microbe interaction, abiotic stress, and plant development. Homologs associated with plant defense that were identified include a hypersensitive-induced response protein (\( Cp45 \)), a pathogen-inducible tyrosine-rich hydroxyproline-rich glycoprotein (\( Cp2 \)), and two unique
peroxidases (Cp9 and Cp43) that have predominate root and hypocotyl-root transcript abundance.

The results from the previous study were utilized in evaluating the pathogen-regulation system by P. palmivora (Porter et al. 2009). Nineteen genes previously isolated from the C. papaya root cDNA library were selected and used for expression analysis following root-drench inoculation with P. palmivora zoospores. The criteria used for selection of these genes for evaluation were tissue specificity, expression level, and predicted protein function. On the basis of their results obtained from pathogen inoculations and the localization of gene expression, we utilized three PR (pathogenesis-related genes) for promoter isolation. They have the predicted identities based on protein homology as follows: peroxidase (Cp9), ß-1,3-glucanase (Cp29), hypersensitive-induced response protein (Cp45); and one abiotic stress response gene, ferulate-5-hydroxylase (Cp35) (Figure 1.9 & 1.10).

From this study, the gene Cp45 showed a low basal mRNA level and was evaluated for its response to pathogen treatment using quantitative RT-PCR (qPCR) (not shown in Figures). However, the qPCR data analysis showed that the Cp45 transcript was elevated at 6 hour post inoculation and became significantly higher than H2O-drench controls at 12, 24 (peak) and 48 hours post-inoculation. Pathogen regulation of the Cp45 transcript level was also evaluated in leaves through vacuum infiltration with zoospores. Transcript levels were significantly increased maximally at 12 hour-post inoculation and greater than H2O-drenched controls at 24-hour post inoculation.

Among these four genes, only peroxidase (Cp9) was found to be down regulated and was presumed to be the potential cross-species effector target of P. palmivora (Porter
**Figure 1.9.** Figure from Porter et al. (2008) on quantified northern blot analysis of *Carica papaya* genes regulated by *Phytophthora palmivora*. RNA was collected from *C. papaya* ‘SunUp’ seedling roots 24 hour post root-drench inoculation with *P. palmivora* zoospores (1×10⁴ zoospores ml/L) (a); and Northern blot analysis expression localization of clones *Cp9*, *Cp29*, and *Cp35* in the leaf (L), hypocotyl (H) and root (R) (b).

**Figure 1.10.** Figure from Porter et al. (2008) on the time course northern blot analysis of *Carica papaya* genes for *Phytophthora palmivora* regulation. RNA was collected from *C. papaya* ‘SunUp’ seedling roots at 1 h, 12 h, 24 h, and 48 h post root-drench inoculation with *P. palmivora* zoospores (1×10⁴ zoospores ml/L) or with water controls (H₂O). The ethidium bromide (EtBr) stained gel of rRNAs is shown at the bottom to confirm equal loading of total RNA.
et al. 2009). However, because of the high relative abundance of the $Cp9$ transcripts in the desired tissue before infection, we will evaluate the strength of the $Cp9$ gene’s promoter. The decrease in the $Cp9$ mRNA levels could be due to a decrease in $Cp9$ gene transcription or to a decrease in RNA stability. If the isolated $Cp9$ promoter would be confirmed as down regulated during pathogen inoculations, then this would further support the observation of the unique expression nature of the $Cp9$, which is presumed to be the cross-effector target of $P. palmivora$.

### Selected native papaya promoters for functional characterization

The four selected promoters that were used in this study are highly expressed in the roots, which are primary and initial tissue targeted by $P. palmivora$. This selection was on the basis of the criteria for root-expressed genes that are regulated by Phytophthora for pathogenicity (Table 1.1).

<table>
<thead>
<tr>
<th>ORF (accession number)</th>
<th>Protein homology</th>
<th>Pfam conserved domain(s)</th>
<th>Expression change</th>
<th>Tissue expression</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Cp9$ (EL784270)</td>
<td>Cationic peroxidase</td>
<td>Secretory peroxidase</td>
<td>Down</td>
<td>R &gt; L &gt; F</td>
<td>PR genes: cell wall reinforcement through lignification</td>
</tr>
<tr>
<td>$Cp29$ (EL784282)</td>
<td>Beta-1, 3-glucanase</td>
<td>Glycosyl hydro-lase</td>
<td>Up</td>
<td>F &gt; L &gt; R</td>
<td>PR genes: endoglucanases are known to release glucan oligosaccharide elicitors from the cell walls (e.g. $P. sojae$ in soybean)</td>
</tr>
<tr>
<td>$Cp35$ (EL784285)</td>
<td>Ferulate-5-hydroxylase</td>
<td>Cytochrome P450-dependent monooxygenase</td>
<td>Up</td>
<td>R &gt; L &gt; F</td>
<td>Abiotic stress gene: biosynthesis of lignin as well as other secondary metabolites; associated with broad-spectrum antifungal defense in plants</td>
</tr>
<tr>
<td>$Cp45$ (EF512303)</td>
<td>Hypersensitive-induced response protein</td>
<td>Band 7 domain of flotillin like proteins</td>
<td>Up</td>
<td>R &gt; L &gt; F</td>
<td>PR genes: regulation of cell death and disease resistance</td>
</tr>
</tbody>
</table>

Tissue expression: R - root; L - leaf; F - fruit
Related function of the genes selected for promoter isolation

a. *Cp9*, peroxidase

In Arabidopsis genome, 73 class III peroxidase genes have been identified which are implicated in quite different physiological processes such as lignification, suberization, auxin catabolism, cross-linking of cell wall proteins, defense against pathogen attack, salt tolerance and oxidative stress (Hiraga et al. 2001). Previous studies showed that one of the most important features of plant peroxidases is their capacity to react to internal or external factors, either by transcriptional or post-transcriptional regulation (Valerio et al. 2004). Gaspar et al (1985) further observed that in some cases, different plant peroxidases are activated or inactivated according to the determined temporal sequence suggesting that certain member of the family may have their own regulation and act sequentially in a particular situation.

b. *Cp29*, β-1,3-glucanase

The family of PR-2 proteins are β-1,3-glucanases are abundant, highly regulated enzymes widely distributed in seed plant species (Simons 1994). Plant β-1,3-glucanases are known for their possible role in the response to the microbial pathogens, but there is a strong evidence that these enzymes are also implicated in diverse physiological and developmental processes in the uninfected plant during cell division, microsporogenesis, pollen germination and tube growth, fertilization, embryogenesis, fruit ripening, seed germination, mobilization of storage reserves in the endosperm of cereal grains, bud dormancy and responses to wounding, cold, ozone and UV B (Leubner-Metzger and Meins 1999).
c. *Cp35* - ferulate-5-hydroxylase

In *Arabidopsis*, *FAH1* locus encodes for the enzyme ferulate-5-hydroxylase (F5H), which catalyzes the rate-limiting step in syringyl lignin biosynthesis and is required for the productions of sinnapate esters (Ruegger et al. 1999). Lignin is one of the compounds synthesized through phenylpropanoid pathway following exposure to environmental stresses such as wounding and pathogen attack. Kim et al (2005) demonstrated that ferulate-5-hydroxylase (F5H) in *Camptotheca acuminata* (CaF5H1) transcripts accumulated in the leaves in response to mechanical wounding or the application of molecules involved in the stress response, such as ethylene, ABA and hydrogen peroxide. The enhanced phenylpropanoid contents via CaF5H1 maybe a function in response to various stresses, including wounding in plants.

d. *Cp45* – hypersensitive-induced response

*Arabidopsis* has four hypersensitive-induced reaction family genes (AtHIR1-4) where the mRNA levels of all except *AtHIR4* were significantly induced by microbe-associated molecular patterns such as bacterial flagellin fragment flg22 (Qi et. al 2011). Some members of the hypersensitive-induced response (HIR) are transcriptionally induced in the cells undergoing HR (hypersensitive response) (Nadimpali et al. 2000). This family had been implicated in many functions including ion channel regulation, microdomain formation, membrane protein chaperoning, vesicle trafficking and membrane-cytoskeletal connection (Morrow & Parton 2005).
Significance of Research

This study was conducted to metabolically engineer *Arabidopsis thaliana* and *Carica papaya* with stilbene synthase (*Vst1*) gene using native papaya promoters to control its expression for basal and enhanced production of the phytoalexin, resveratrol. Stilbene synthase, which is not present in *Arabidopsis* and papaya, is a key enzyme that controls the production of potent antimicrobial and antifungal phytoalexin, resveratrol. The precursor molecules for resveratrol production consist of one molecule of *p*-coumaryl and three molecules of malonyl-CoA are ubiquitous in plants cells. Resveratrol and its derivatives are produced by adding stilbene synthase enzyme in the phenylpropanoid pathway diverting the flavonoid biosynthesis into production of stilbenes, instead of naringenin chalcones. The expression of the grapevine stilbene synthase gene in papaya enhances resistance to *Phytophthora palmivora* (Zhu et al 2004) but the expression of grapevine *Vst1* gene driven by its own native promoter is less than optimal in papaya while the conventional use of constitutive CaMV35S promoter in plant genetic engineering can potentially result to homology-dependent gene silencing or morphological abnormalities. Thus, identification and characterization of a native papaya promoter that would have basal or enhanced expression of resveratrol in papaya tissues affected by *Phytophthora palmivora* such as roots and stems without affecting the overall normal phenotypic growth and development of the transgenic plants as a result of stilbene
synthase transgene expression is highly desired.

*Arabidopsis thaliana* was used as model system for analyzing the strength, activity, and tissue localization of the native papaya promoters through eGFP gene reporter assays. *Arabidopsis* was also transformed with *Cp papaya:*Vst1 constructs to assess the phenotypic and metabolic consequences of stilbene synthase expression as affected by the native papaya promoter’s regulatory activity. Finally, we aim to develop papaya transgenic plants with modulated expression of stilbene synthase using our isolated and cloned native papaya promoters to avoid alteration of metabolic profile in the plant due to its ectopic expression. We hope to provide normal transgenic plant materials producing phytoalexin, resveratrol that can be use for future disease resistance induction experiments with improved nutritional content.

**Hypotheses**

1. The four putative promoter regions and their 5’UTRs isolated from the genes encoding peroxidase (*Cp9*), β-1,3- glucanase (*Cp29*), ferulate-5-hydroxylase (*Cp35*) and hypersensitive-induced response protein (*Cp45*) from papaya will contain *cis*-acting regulatory elements and motifs that will be recognized by the *Arabidopsis thaliana* transcriptional machinery.

2. *Arabidopsis* eGFP gene reporter expression assays will quantify and characterize the strength, activity and tissue localization of the four isolated promoters based on the eGFP fluorescence, mRNA transcript levels, and protein abundance that will help understand and identify promoter’s utility in plant genetic transformation.
3. Mapping the Transcription Start Site (TSS) will define the length of the 5’UTR and potential formation of secondary structures, which might contain elements affecting post-translational regulation.

4. Metabolic engineering of *Arabidopsis thaliana* with *Cp promoter:Vst1* fusions gene would provide useful tool to understand and characterize the amount of resveratrol being synthesized and should provide information on the overall phenotypic and metabolic consequences of its genomic integration.

5. Development of modified *Agrobacterium* transformation protocol utilizing suspension-derived cultures of papaya would serve as an alternative transformation technique that would generate fully transgenic plants for papaya transformation work.

6. Metabolic engineering of papaya with the use of our native papaya promoter would result to controlled stilbene synthase expression for production of morphologically normal transgenic papaya plants with enhanced resveratrol levels on the desired tissues.
The above hypotheses will be studied within the succeeding chapters by focusing on the following specific objectives:

**Objectives:**

**I. Isolate and characterize optimal pathogen-induced promoters from papaya that exhibit enhanced expression in *Arabidopsis* tissues such as roots and stems.**

A. Isolate and clone native papaya promoters from ‘SunUp’ cultivar.

B. Construct pCAMBIA 1302-papaya promoter:egfp transformation vector and mobilized it into *Agrobacterium tumefaciens* GV 3101.

C. Transform *Arabidopsis thaliana* with *Agrobacterium tumefaciens* GV 3101 carrying the pCAMBIA 1302-papaya promoter:egfp constructs through floral dip method.

D. Determine and characterize the strength and activity of the promoters fused to egfp in transgenic lines of *Arabidopsis* using epi-fluorescence microscopy, Reverse Transcription-Polymerase Chain Reaction (RT-PCR), Quantitative Real Time PCR (qPCR) and Western blot analyses.

E. Map the transcription start site (TSS) of the promoters to examine the length and potential secondary structures present in the 5’ UTR.

**II. Bioengineer *Arabidopsis thaliana* with *Cp* promoter:Vst1 fusions to characterize the resulting phenotypes and quantitate resveratrol production of the transgenic lines.**

A. Construct pCAMBIA 1302-papaya promoter:Vst1 transformation vector and
mobilized it to *Agrobacterium tumefaciens* GV 3101.

B. Transform *Arabidopsis thaliana* with *Agrobacterium tumefaciens* GV 3101 carrying the pCambia 1302-papaya promoter:*Vst1* constructs through floral dip method.

C. Verify transgenic *Arabidopsis* lines using molecular diagnostics through Polymerase Chain Reaction (PCR) and quantitate transcript mRNA expression of *Vst1* through Quantitative Real Time PCR (qPCR).

D. Identify and quantitate the presence of resveratrol glucoside (piceid) production in transgenic *Arabidopsis* through High Performance Liquid Chromatography (HPLC) using piceid (3,5,4'-trihydroxystilbene-3-O-β-D glucopyranoside) standard.

E. Confirm and verify the integrity of the compound form the mixture using tandem Mass Spectrophotometry (MS/MS).

F. Quantitate the anthocyanins at the 5 day-old stage of *Arabidopsis* seedlings in *Vst1* expressing lines to determine substrate competition between stilbene synthase (*Vst1*) and chalcone synthase (*CHS*).

G. Assess the overall and phenotypic traits of the transgenic plants as a result of *Vst1* genomic integration in *Arabidopsis*.

### III. Bioengineer *Carica papaya* with *Cp promoter:*Vst1 fusions to characterize the resulting phenotypes and quantitate resveratrol production of the transgenic lines.

A. Establish tissue culture from stems of germinated seedlings of papaya variety
‘Kapoho’ for establishment of cell suspension culture.

B. Utilize suspension-derived culture of papaya for optimization of modified Agrobacterium-mediated transformation protocol.

C. Transform Carica papaya with Agrobacterium tumefaciens GV 3101 carrying the pCambia 1302-papaya promoter:Vst1 constructs through the modified transformation technique.

D. Verify transgenic papaya lines through Polymerase Chain Reaction (PCR), RT-PCT to determine the relative Vst1 transcript abundance and quantitate Vst1 mRNA transcript expression through Quantitative Real Time PCR (qPCR).

E. Identify and quantitate the presence of resveratrol glucoside (piceid) production in transgenic papaya through High Performance Liquid Chromatography (HPLC) using piceid (3,5,4'-trihydroxystilbene-3-O-β-D glucopyranoside) standard.

F. Confirm and verify the integrity of the compound from the mixture using tandem Mass Spectrophotometry (MS/MS).

G. Assess the overall morphological growth and phenotypic traits of the generated transgenic papaya plants.
CHAPTER III

IDENTIFICATION AND CHARACTERIZATION OF PAPAYA (CARICA PAPAYA, L.) PROMOTERS BY HETEROLOGOUS EXPRESSION AS EGFP FUSIONS IN ARABIDOPSIS THALIANA

Abstract

Promoters are essential upstream genetic switches that activate and repress gene expression. Promoter characterization is a key prerequisite before using in downstream biotechnology applications. The model system, Arabidopsis thaliana, was utilized here to efficiently hasten the time to verify and characterize the functionality and strength of promoters from the more complex tropical tree, Carica papaya. Four putative promoter regions and their 5’UTRs were isolated from the genes encoding peroxidase (Cp9), β-1,3-glucanase (Cp29), ferulate-5-hydroxylase (Cp35) and hypersensitive-induced response protein (Cp45) and fused to eGFP. In silico analysis predicted the presence of several cis-elements associated with regulatory functions in stress and defense responses. The Arabidopsis transcriptional machinery readily recognized the promoters, as determined by qualitative and quantitative measurements of eGFP expression (fluorescence, mRNA and protein levels). The eGFP was expressed in a variety of tissues of the transgenic plants (vasculature, shoot apex, cotyledon, cotyledon petioles, hypocotyls, and root). The Cp29 and Cp45 promoters showed the highest and most promising overall expression. Comparison of eGFP mRNA and protein levels indicated post-transcriptional regulation. Identifying the precise transcription start sites (TSSs) demonstrated transcription fidelity and mapped the length of the 5’ UTR. Predicted mRNA 5’-UTR secondary structures
potentially affected the translational efficiency of the mRNAs during development, most notably for \textit{Cp9} and \textit{Cp35}. This work demonstrates the utility of using \textit{Arabidopsis} to quickly evaluate and identify useful promoters (\textit{Cp29}, \textit{Cp45}) from complex tropical plants for future biotechnology goals, and to analyze 5'UTR regulation, \textit{cis}-elements, and \textit{trans}-acting factors.

\textbf{Introduction}

Promoters are essential upstream genetic switches that activate and repress gene expression (Buchanan et al. 2000) and are useful tools in genetic engineering. The popular constitutive promoters such as the cauliflower mosaic virus (CaMV35S) and nopaline synthase (nos) (Odell et al. 1985; Benfey and Chua, 1989) promoters have disadvantages including nonspecific increased transgene expression throughout the plant. Such constitutive expression can adversely affect plant growth and development (Potenza 2004) and, in some cases, lead to homology-based transgene silencing (Jorgensen 1992; Rocha et al. 2005). It is advantageous to produce crops with tissue-localized, temporal and developmental transgene expression to enhance disease and pest resistance in the primary infected tissues (Potenza et al. 2004). Thus, characterizing the attributes of a promoter’s expression is an important prerequisite before it can be used for crop improvement via biotechnology.

Many crop plants, especially trees, are problematic for promoter identification and verification. In the tropical tree, the complicated process of plant transformation, genotypic response, slow growth, and long regeneration times \textit{in vitro} and \textit{in vivo} hinders papaya promoter identification and characterization. At the same time, there is an
increasing need for promoters with unique expression profiles. For example, strong promoters active in papaya roots and stems are needed to genetically engineer disease resistance against the devastating pathogen, *Phytophthora palmivora*, that primarily infects these tissues.

Relatively few promoters coming from tree species, such as pine, oak, oil palm, and *Leucaena* have been isolated and characterized and even fewer have been tested in simple model systems, such as tobacco, rice, and tomato transgenic plants (No 2000, Taha 2012, Ahmadi 2003, Prashant 2012). Expression studies of promoters from these perennial tree species conferred tissue specificity at various levels with potential biotechnology applications. American chestnut (Connors et al. 2002) and aspen (Hanson et al. 2002) were found to have tissue-regulated expression in *Arabidopsis*, demonstrating the utility of the *Arabidopsis* system for tree promoter identification. However, the approaches used relied on qualitative macroscopic and microscopic observations of the reporter gene product. Quantitative measurements of gene expression at both mRNA and protein levels will provide more accurate information to enhance promoter characterization, reveal post-transcriptional effects that together, will complement the qualitative approaches.

In this study, we used the *Arabidopsis* system to efficiently hasten the time to verify and characterize the functionality and strength of promoters from the complex tropical tree, papaya. The screening parameters were used to select promoters for subsequent papaya bioengineering. The promoters were derived from four genes (encoding peroxidase, β-1,3-glucanase, hypersensitive-induced response protein, ferulate-5-hydroxylase) from the cultivar ‘SunUp’, the genome for which is completely sequenced.
(Table 3.1). These genes were chosen because they were previously found to be expressed in roots and hypocotyls (Porter et al. 2008) at the RNA level, but their individual promoters were not analyzed or characterized. Therefore, in this report, each promoter was fused to the eGFP reporter gene and compared with CaMV35S-eGFP expression during development and in a variety of mature Arabidopsis tissues. In addition to fluorescence microscopy, immunoblot and RT-PCR analyses were used to quantify transcript and protein levels and determine translational effects. In addition, 5’-RACE was used to identify the transcriptional start sites of the papaya promoters and resulting 5’UTRs expressed in Arabidopsis. Promoters Cp29 and Cp45 had the most useful expression characteristics. We demonstrated that the promoters were transcriptionally activated and stably expressed in roots and hypocotyls vasculature, which are the primary tissues targeted by the papaya pathogen, Phytophthora palmivora. The rapid characterization of these promoters and 5’UTRs will allow us to regulate the expression of pathogen resistance transgenes for future bioengineering work against this disease of the tropics.

Materials and Methods

Isolation of genomic regions and cloning of the putative papaya promoters

Total genomic DNA was extracted from the youngest fully expanded leaves of papaya cv. ‘SunUp’ using the CTAB method (Doyle and Doyle, 1990) and quantified spectrophotometrically (Beckman and Coulter DU 730 UV/Vis). Genomic sequences from the previously isolated papaya cDNAs (Porter et al. 2008) of Cp9 (peroxidase), Cp29 (beta-1,3-glucanase), Cp35 (ferulate-5-hydroxylase) and Cp45 (hypersensitive-
induced response protein) were obtained from the Papaya Genome Database (http://asgpb.mhpcc.hawaii.edu/papaya/). Primers were designed to flank the upstream ~2.0 kb regions, including the putative transcriptional start sites and the 5’ untranslated regions (5’UTR), of the four genes and had PstI and NcoI restriction sites added to their 5’-ends (Table 3.1). Using 50 ng of genomic DNA as template, the putative promoter regions were amplified by PCR using Accuzyme™ DNA Polymerase (Bioline, USA). The following PCR conditions were used: initial denaturation at 95°C for 5 min, denaturation at 95°C for 30 sec, annealing at 58°C for 30 sec, and extension at 68°C for 3 min for 35 cycles ending with a final 7 min extension at 68°C. The PCR products were purified (Qiagen, Valencia CA, USA) by agarose gel electrophoresis and sequenced at the University of Hawaii Advanced Studies of Genomics, Proteomics, and Bioinformatics (ASGPB)(http://asgpb.mhpcc.hawaii.edu/). The regions were analyzed for transcriptional cis-regulatory regions using the PlantCARE (Lescot et al. 2002) and PLACE (Higo et al. 1999) database software (http://bioinformatics.psb.ugent.be/webtools/plantcare and http://www.dna.affrc.go.jp/htdocs/PLACE/signalscan).

Table 3.1. Primers used in cloning of papaya promoter regions.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Promoter Amplicon Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cp9F</td>
<td>5’-ctgcagTCTTGAAATCTTGTAAACGACTAGATTCC-3’</td>
<td>2041</td>
</tr>
<tr>
<td>Cp9R</td>
<td>5’-ccatggTCTCTGCTAAAGCGGAGAAAGCTGAGA-3’</td>
<td></td>
</tr>
<tr>
<td>Cp29F</td>
<td>5’-ctgcagGGCCATGACTTCGAGTCTTCGTTTAG-3’</td>
<td>1445</td>
</tr>
<tr>
<td>Cp29R</td>
<td>5’-ccatggGGCCAAATATTTGAGAGATGAGAAACC-3’</td>
<td></td>
</tr>
<tr>
<td>Cp35F</td>
<td>5’-ctgcagCTACTGCTTTTTACCATGTGCTTGG-3’</td>
<td>1832</td>
</tr>
<tr>
<td>Cp35R</td>
<td>5’-ccatggGGTTACTGTTGGTTGGTGTTGGAATGGAATTTG-3’</td>
<td></td>
</tr>
<tr>
<td>Cp45F</td>
<td>5’-ctgcagGGACAGTGTGAGTTACTAACAAGAG-3’</td>
<td>2051</td>
</tr>
<tr>
<td>Cp45R</td>
<td>5’-ccatggACAAGTTAAGAACAACCAATGGCTG-3’</td>
<td></td>
</tr>
</tbody>
</table>

Sequences underlined and in lower case were introduced PstI and NcoI sites.
Construction of binary vectors for transformation

The PCR products were ligated into the \textit{PstI} and \textit{NcoI} restriction sites of the pGEM-T vector (Promega, Madison, WI, USA), propagated in \textit{E. coli} DH5\textalpha{} cells, and verified by DNA sequencing to contain the correct papaya genomic regions. The cloned PCR products were excised from the pGEM-T through double digestion with \textit{PstI} and \textit{NcoI} and were ligated in the \textit{PstI} and \textit{NcoI} sites of pCAMBIA-1302, thereby replacing the \textit{CaMV35S} promoter. The \textit{eGFP} gene (Thastrup et al. 1995), a variant of the green fluorescent protein that expresses intense green fluorescence, was used as reporter gene in place of the \textit{mgfp5}. The \textit{mgfp5} was removed from pCAMBIA-1302 by digestion with \textit{PstI} and \textit{AflII}. The \textit{eGFP} was ligated into the pCAMBIA-1302 to generate the \textit{Cp promoter:eGFP} construct. The \textit{CaMV35S:eGFP} was used as positive expression control (Figure 3.1). The orientation and reading frame of all recombinant plasmids were verified by PCR and DNA sequencing. Recombinant plasmids were mobilized onto \textit{Agrobacterium tumefaciens} strain GV3101 competent cells through the modified freeze-thaw method (Jyothishwaran et al. 2007).

Plant material and growth condition for transformation

\textit{Arabidopsis thaliana} L. Heynh (ecotype Columbia) wild type was used throughout the study. The seeds were surface sterilized, plated onto sterile $\frac{1}{2}$ MS solid medium (Murashige and Skoog 1962) supplemented with 3\% sucrose, and placed in the dark for 2 days at $4^\circ$C, then transferred to $22^\circ$C and a 16-h-light/8-h-dark cycle for seed germination. Seedlings at 10-14 days were transplanted to soil and were grown until immature unopened flower buds were formed. \textit{Arabidopsis} was transformed by the floral
**Figure 3.1.** Schematic representation of the transformation vector (pCAMBIA 1302) and the corresponding papaya promoter-eGFP gene fusions. (a) The plant selectable marker hygromycin phosphotransferase gene (hptII) is driven by the Cauliflower mosaic virus 35S promoter (CaMV35S) with a poly-A terminator. The native papaya promoter replaced the pCAMBIA 1302 CaMV35S promoter upstream of the enhanced green fluorescence protein (eGFP) gene with downstream nopaline synthase Poly-A terminator (nos Poly A) maintained; (b) The length, in base-pairs (bp) of the putative promoter includes the 5’untranslated leader region (5’UTR, gray box) relative to the translational start site (ATG) of each papaya promoter-eGFP fusion.

dip method (Clough and Bent 1998) using *Agrobacterium tumefaciens* GV3101 carrying the recombinant constructs. Putative transgenic T₁ seeds were selected *in vitro* on solidified MS medium with 50µg/ml hygromycin B (Sigma-Aldrich, Co). The presence of the papaya promoter and *eGFP* were confirmed through PCR using the previously designed primers (Table 3.1) and *eGFP*-specific primers (5’-
CCATGGTGAGCAAGGGCGAG-3’ and 5’-
CTTAAGAAACTTTATGGCCAAAATGTTTGAAAGCAG-3’ to amplify the ~1.4-2.0kb promoter regions and the 759 bp eGFP of the T-DNA insertion, respectively. Transgenic lines showing 3:1 segregation, indicating a single integration locus, were used for further analysis. Transgenic T1 seedlings were transferred to soil and self-pollinated to obtain the T2 generation, which was further screened by PCR and hygromycin. Positive plants were self-pollinated up to the T3 generation to obtain the homozygous lines in which no offspring were hygromycin sensitive.

**RNA Isolation and RT-PCR**

Total cellular RNA was isolated using the RNeasy Plant Mini kit and were treated with on column DNase treatment according to the manufacturer’s protocol (Qiagen Inc., Valencia, CA, USA). RNA was extracted from whole seedlings at different developmental stages (7, 14, 21 and 28 DAP) or from organs (flowers, leaves, stems and roots) of mature reproductive 1.5-month-old *Arabidopsis* plants. 1µg of total RNA was used to produce the first-strand cDNA with M-MLV reverse transcriptase and oligodT (Promega). The RT-PCR was carried out with Bio-X-Act Short Reaction Mix (Bioline Inc., Taunton, MA, USA) to amplify the target genes with the eGFP-specific primers (described above) or the Actin2 specific gene primers (as an internal control, primer pair 5’- TTGCAGGAGATGATGCTCCCAGG-3’ and 5’-CATTCCCACAAAACGAGGCTGAGG-3’) using 100 ng cDNA template. To quantitate the relative transcript abundance, the number of PCR cycles was determined for the linear-range amplification of the eGFP mRNA with the simultaneous RT-PCR of the internal
housekeeping gene, *Actin 2* (Figure 3.2). The PCR products were separated via agarose gel electrophoresis and stained with 1:10,000 GelRed™ (Biotium, Inc., Hayward, CA, USA). The optimized number of PCR cycles used for the quantification of eGFP expression was 30 cycles. The eGFP net PCR band intensity for each data point was normalized to the internal control *Actin2* using the Kodak Multi-Imaging System and software (Kodak Co., Rochester, NY, USA) to quantify the differences of eGFP expression of the samples.

![Figure 3.2. Determination of the exponential phase of semi-quantitative RT-PCR of the eGFP and *Actin2* transcripts in transgenic Arabidopsis thaliana (35S-eGFP construct). (a) RT-PCR products assessed via agarose gel electrophoresis of eGFP (upper) and *Actin2* (lower); (b) graphic quantitative representation of gel shown in (a) obtained by Kodak Phosphorimager from the net intensity data plotted against the number of PCR cycles.](image-url)
**Fluorescence microscopy**

Epi-fluorescence was used to assess the papaya promoter-directed *eGFP* expression in the transgenic *Arabidopsis* tissues in plants at 7 and 14 and 21 DAP, compared to the non-transformed (WT) plants and to the transgenic plants expressing the constitutive CaMV35S promoter-*eGFP*. Preparations were examined on an Olympus BX51 upright compound microscope equipped with a U-MWB2 Wide Green filter cube (465-490 nm excitation, 495 nm barrier filter, 500-530 nm long pass emission). Images were recorded with an Optronics Macrofire SP CCD camera.

**Real-Time PCR analysis**

RNA was isolated using Qiagen kit as described above and its concentration and rRNA ratio were quantified using spectrophotometer (Beckman and Coulter DU 730 UV/Vis). 1µg of total RNA was converted to cDNA using M-MLV reverse transcriptase and oligodT (Promega). eGFP PCR primers 5’-CCATGGTGAGGAAGGGCGAG-3’ and 5’-CCTCGCCGGACACGCTGAACTTGTG-3’ were designed to amplify a 100 bp product of the eGFP cDNA sequence. *Arabidopsis thaliana* Ubiquitin-Conjugating Enzyme 21 (AT5G25760) (PEROXIN4) cDNA was used as the internal standard with primers 5’-CTGCGACTCAG^GGAATCTTCTAAGGAATCTTCTAA-3’ and 5’-TTGTGCCATTGAAATTGAAACCC-3’ (Czechowski et al. 2005). 1µL (50 ng) of cDNA product was combined with forward and reverse qPCR primers to a final concentration of 500nM. This reaction was then combined with an equal volume of 2x Roche Power SYBR Green Master Mix with ROX (Applied Biosystems, Carlsbad, CA) mixed and aliquoted in 10µL triplicates into a 384-well optical qPCR plate then sealed with an
Applied Biosystems optical adhesive film. Samples were run on a 7900HT qPCR machine with the following thermocycling parameters: 2 minutes at 50°C, 10 minutes at 95°C, then the following conditions were repeated for 40 cycles: 15 seconds at 95°C, 1 minute at 60°C. A dissociation step was performed at the end of the run to ensure only a single product was formed utilizing a +1°C ramp rate per second from 60°C to 95°C.

**Protein isolation and immunoblot analysis**

Total cellular proteins were isolated from transgenic *Arabidopsis* T₃ seedlings at different developmental stages (7, 14, 21, 28 days after planting) and tissues at the reproductive stage (1.5-month-old) following the protocol of Martinez-Garcia et al. (1999). The protein levels were quantified spectrophotometrically (Beckman and Coulter DU 730 UV/Vis), separated via standard SDS-PAGE (Laemmli 1970), were electro-transferred onto a Protran® nitrocellulose membranes (PerkinElmer Life Sciences, Boston, MA, USA) and probed with an anti-GFP rabbit IgG antibody (Molecular Probe-Invitrogen Corp., Carlsbad, CA, USA). Detection was carried out using an anti-rabbit secondary antibody from the Amersham ECL Western blotting analysis system (ECL™ kit, Amersham Biosciences, Piscataway, NJ, USA). The chemiluminescent signals were developed by exposure to Kodak X-ray film.

**Identifying the transcriptional start sites by 5’ RLM-RACE**

To determine the transcriptional start sites (TSS) of the eGFP transcripts driven by the different papaya promoters expressed in *Arabidopsis* and in the papaya transgenic lines, 5’-RNA ligase-mediated rapid amplification of cDNA ends (5’-RLM-RACE) was
performed with the Ambion FirstChoice RLM- RACE kit (Ambion Inc., Austin, TX, USA) according to the manufacturer’s instructions with modifications. Total cellular RNA was isolated as described above from the two week old in vitro grown transgenic Arabidopsis plants and previously transformed somatic embryos of papaya containing the Carica papaya promoter-eGFP fusions (cultures provided from a separate study). The cDNA synthesis was primed using oligodT (Promega, Madison, WI, USA) in place of the random decamer supplied in the RLM-RACE kit. Nested PCRs were carried out using Bio-X-Act Short Reaction Mix (Bioline Inc., Taunton, MA, USA) under the following thermocycling parameters: initial denaturation at 95°C for 5 min, denaturation at 95°C for 15 sec, annealing at 60°C for 15 sec, and extension at 72°C for 30 sec for 35 cycles with a final 7 min extension at 72°C. The outer PCR consisted of 5’-RACE outer primer 5’-GCTGATGGCGATGAATGAACACTG-3’ with eGFP gene specific outer primer 5’-AAGAAGATGGTCGGCTCTGGGACG-3’. The inner PCR consisted of 1µl of the outer PCR as template with 5’-RACE inner primer 5’-CGCGGATCCGAACACTGCGTTTGCTGGCTTTTGTG-3’ and eGFP gene specific inner primer 5’-CCTCGCCGGACACGCTGAACTTGTG-3’. The PCR products were cloned into pGEM-T vector (Promega, Madison, WI, USA) and propagated in E. coli DH5α cells. Plasmids containing the inserts were extracted using Machery-Nagel NucleoSpin® Plasmid (Machery-Nagel Inc., Bethlehem, PA, USA) and transcription start sites were identified through sequencing. Three independent RACE experiments were conducted for each line. At least five plasmids isolated from bacterial colonies that were positive for the insert through PCR were submitted for sequencing to validate the results for TSS identification. Alternatively, nested PCR products purified by Machery-Nagel
NucleoSpin® Gel and PCR Clean-up were also directly submitted for sequence analysis at the ASGPB.

**Results**

**Comparison of papaya promoter sequences to the cis-regulatory element database**

The putative promoter regions consisted of ~2.0 kb upstream genomic fragments from each of the four selected genes including the entire 5’ untranslated region (5’UTR) (Figure 3.3). The predicted eukaryotic core promoter motifs, the TATA and CAAT boxes, were found (Figure 3.4–3.7). By searching against the PlantCARE databases and PLACE databases (Lescot et al. 2002; Higo et al. 1999), several cis elements related to light regulation, stress and defense response, hormone response, tissue-specific expression, enhancers and protein binding motifs were predicted for each promoter region (Supplementary Tables 1-4). A complete list of cis-elements predicted with their associated functions was presented (Supplementary Table 5). A comparison of these cis-elements and their predicted functions was summarized (Table 3.2 & 3.3, Figure 3.8).

Motifs related to abiotic stress and defense responses that were common to the four papaya promoters were the following: ARE - essential for the anaerobic induction, HSE - involved in heat stress responsiveness, MBS - MYB binding site involved in drought-induction, and TC-rich repeats - involved in defense and stress responsiveness (highlighted in Table 3.2).
Figure 3.3. Agarose gel (1%) electrophoresis of PCR products from four different promoter regions. Isolated promoters of (a) \(Cp9\) (peroxidase); (b) \(Cp29\) (beta,1-3, glucanase); (c) \(Cp35\) (ferulate-5-hydroxylase); and (d) \(Cp45\) (hypersensitive-induced response protein) are from the papaya cv. ‘SunUp’. Molecular marker (MW) is in kilobase pair (kb).

The identification of the putative regulatory elements located in the promoter regions of the three papaya pathogenesis-related genes (\(Cp9, Cp29, Cp45\)) and the abiotic response gene (\(Cp35\)) using an in silico analysis approach provides relevant preliminary information. However, a detailed characterization of expression driven by the papaya promoter regions at developmental and reproductive stages was required. We therefore sought to verify the transcriptional activity and function of these putative promoter regions by expressing them as transcriptional eGFP fusions in the efficient model system of \textit{Arabidopsis}. The resulting spatial and temporal gene expression activities were assessed using multiple qualitative and quantitative approaches.
**Figure 3.4.** Nucleotide sequence of the 2041 bp *Carica papaya* Cp9 peroxidase gene promoter. Computer-predicted *cis-elements* found related to defense and stress were highlighted in green where the names are indicated above corresponding sequence. Restriction enzymes (*Pst*I & *Nco*I) used for cloning the promoter were underlined. The nucleotide sequence of the predicted CAAT and TATA boxes are shown in box. The transcription start sites (TSS) were denoted by +1 (right red arrow) and were indicated by the single letter (red) utilized either by *Carica papaya* (*Cp*) and/or *Arabidopsis thaliana* (*At*) (blue). The first few amino acid sequence of the reporter eGFP gene was indicated (blue) after the ATG translational start codon.
Figure 3.5. Nucleotide sequence of the 1445 bp *Carica papaya* Cp29 B,1-3 glucanase gene promoter. Computer-predicted cis-elements found related to defense and stress were highlighted in green where the names are indicated above the corresponding sequence. Restriction enzymes (*Pst*I & *Nco*I) used for cloning the promoter were underlined. The nucleotide sequence of the predicted CAAT and TATA boxes are shown in box. The transcription start sites (TSS) was denoted by +1 (right red arrow) and were indicated by the single letter (red) utilized either by *Carica papaya* (Cp) and/or *Arabidopsis thaliana* (At) (blue). The first few amino acid sequence of the reporter eGFP gene was indicated (blue) after the ATG translational start codon.
Figure 3.6. Nucleotide sequence of the 1832 bp *Carica papaya* Cp35 ferulate-5-hydroxylase gene promoter. Computer-predicted cis-elements found related to defense and stress were highlighted in green where the names are indicated above the corresponding sequence. Restriction enzymes (*Pst*I & *Nco*I) used for cloning the promoter were underlined. The nucleotide sequence of the predicted CAAT and TATA boxes are shown in box. The transcription start sites (TSS) were denoted by +1 (right red arrow) and were indicated by the single letter (red) utilized either by *Carica papaya* (*Cp*) and/or *Arabidopsis thaliana* (*At*) (blue). The first few amino acid sequence of the reporter eGFP gene was indicated (blue) after the ATG translational start codon.
Figure 3.7. Nucleotide sequence of the 2045 bp *Carica papaya* Cp45 hypersensitive-induced response protein gene promoter. Computer-predicted *cis-elements* found related to defense and stress were highlighted in green where the names are indicated above and/or below the corresponding sequence. Restriction enzymes (*Pst*I & *Nco*I) used for cloning the promoter were underlined. The nucleotide sequence of the predicted CAAT and TATA boxes are shown in box. The transcription start sites (TSS) were denoted by $^{+1}$ (right red arrow) and were indicated by the single letter (red) utilized either by *Carica papaya* (*Cp*) and/or *Arabidopsis thaliana* (*At*) (blue). The first few amino acid sequence of the reporter eGFP gene was indicated (blue) after the ATG translational start codon.


Table 3.2. A comparison of the putative cis-elements of the *Cp9*, *Cp29*, *Cp35* and *Cp45* promoters by the PlantCARE Promoter Analysis software. The number of predicted motifs for each promoter is indicated.

<table>
<thead>
<tr>
<th>MOTIF</th>
<th>PROMOTER</th>
<th>FUNCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Cp9</em></td>
<td><em>Cp29</em></td>
</tr>
<tr>
<td>3-AF1 binding site</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4cl-CMA2b</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>5′UTR Py-rich stretch</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A-box</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ABRE-motif</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AACA motif</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AC-I</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AC-II</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AE-box</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ARE</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AT1-motif</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>ATC-motif</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ATCT-motif</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Element Type</td>
<td>AT</td>
<td>GC</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>AT-rich element</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Box 1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Box 4</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Box W1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CGTCA-motif</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Circadian</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Chs-CMA1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>DOTC</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ELI-box3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ERE</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>G-Box</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G-box</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GA-motif</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GAG-motif</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GARE-motif</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GATA-motif</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GCN4-motif</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>GT1-motif</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>HD-Zip3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>HSE</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>I-box</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Element</td>
<td>Predicted</td>
<td>Observed</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>L-box</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>LTR</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MBS</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MRE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Skn-1_motif</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Sp1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>TA-rich region</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>TATC-box</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TCA-element</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>TC-rich repeats</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>TCT motif</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>TGA-element</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>TGACG-motif</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WUN-motif</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3.3 Summary of the related functions of the putative cis-elements of Cp9, Cp29, Cp35 and Cp45 promoters as determined by the PlantCARE Promoter Analysis software. The number of predicted functional motifs for each promoter is indicated.

<table>
<thead>
<tr>
<th>PROMOTER</th>
<th>RELATED FUNCTION</th>
<th>Total No. of Cis-elements Found</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Light Regulation</td>
<td>Defense and Stress Response</td>
</tr>
<tr>
<td>Cp9</td>
<td>27</td>
<td>13</td>
</tr>
<tr>
<td>Cp29</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>Cp35</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>Cp45</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

Figure 3.8. Graphical representation of the number of regulatory motifs predicted to reside in the native papaya promoters (Cp9, Cp29, Cp35, Cp45). They are grouped according to their related functions analyzed by plant promoter database PlantCARE (Lescot et al. 2002).
Promoter activity during the developmental stages (7 to 28 DAP) as assessed by epi-fluorescence microscopy

The four promoter eGFP reporter constructs introduced into *Arabidopsis* were designated as follows: *Cp9*-eGFP; *Cp29*-eGFP; *Cp35*-eGFP; *Cp45*-eGFP (Figure 3.1). *Arabidopsis* transgenic plants were also produced using CaMV35S-eGFP construct to act as a positive control and the wild type (WT) untransformed plant was used as a negative control. Three independent transgenic lines per construct as verified by PCR, DNA sequencing and RT-PCR were used for the promoter-eGFP expression studies.

To investigate the promoter activity in the entire *Arabidopsis* plant at developmental stages 7, 14 and 21 DAP, the promoter-eGFP transgenic plants were visually evaluated with a fluorescence microscope. Three independent transgenic lines for each promoter fusion were evaluated and one representative transgenic line for each construct was shown in the Figures 3.9-3.12. Significant differences in the pattern of eGFP gene expression under the control of the *Cp9*, *Cp29*, *Cp35* and *Cp45* promoters were observed (Figures 3.9-3.12). Variation in expression was observed for both the strength in overall expression and for expression levels in specific tissues. Interestingly, *Cp29*-eGFP and *Cp45*-eGFP were found to be expressed at similar levels as the constitutive CaMV35S-eGFP. The basal expression of these two promoters was strong in almost all of the organs analyzed. In contrast, *Cp9*-eGFP basal expression was moderate as compared to *Cp29*-eGFP and *Cp45*-eGFP expression, while *Cp35*-eGFP expression was markedly lower in all parts of the plant (Figures 3.9-3.11). As expected, eGFP activity in plants transformed with CaMV35S-eGFP was high in all tissues (shoot apex, cotyledon, petiole, stem and root) through the stages (7-21 DAP) of plant development.
eGFP expression was present in the vasculature of the plants for Cp29, Cp45 and Cp35. The expression of Cp9-eGFP was unique from the other three promoters in that there was no clear preferential vascular tissue expression. However, the moderate-to-high eGFP fluorescence was observed in the shoot meristem, cotyledon, petiole, stem and roots for Cp9.

**Figure 3.9.** Expression patterns of papaya promoter-eGFP fusions (Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP) in the shoot apex of *Arabidopsis* at 7 and 14 DAP. The eGFP expression was compared with the corresponding tissue of wild type (WT) plants. Representative images for each promoter lines are shown. The magnification bar represents 300 µm.
Figure 3.10. Expression patterns of papaya promoter-eGFP fusions (Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP) in the cotyledon (7 DAP) and cotyledon petiole (14 DAP) of Arabidopsis. The eGFP expression was compared with corresponding tissue of wild type (WT) plants. Representative images for each promoter-eGFP line are shown. The magnification bar represents 300 µm.
**Figure 3.11.** Expression patterns of papaya promoter-eGFP fusions (Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP) in the hypocotyl of Arabidopsis at 7, 14 and 21 DAP. The eGFP expression was compared with corresponding tissue of wild type (WT) plants. Representative images for each promoter-eGFP line are shown. The magnification bar represents 300 µm.
Figure 3.12. Expression patterns of papaya promoter-eGFP fusions (Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP) in the root tissues *Arabidopsis thaliana* at 7, 14 and 21 DAP. The eGFP expression was compared with corresponding tissue of wild type (WT) plants. Representative images for each promoter-eGFP line are shown. The magnification bar represents 300 µm.

Epi-fluorescence microscopy was an efficient approach to initially detect and visually screen for the eGFP expression in *Arabidopsis* plants at different stages of plant development. However, the presence of chlorophyll can mask some of the eGFP fluorescence levels. Because of the need to quantitate discrete differences in the level of
eGFP expression to provide accurate information on the promoter activities, their regulation and usefulness for genetic engineering approaches, we conducted RT-PCR analysis followed by eGFP immunoblot analysis. This approach allowed the comparison of mRNA with protein levels in various tissues and developmental stages.

Semi-quantitative RT-PCR and Immunoblot Analysis to measure eGFP expression:

a. Early Developmental Stage (7-28 DAP)

The eGFP mRNA levels of three independent T3 transgenic lines for each of the promoter:eGFP fusions were analyzed through semi-quantitative RT-PCR (Figure. 3.13) in four developmental stages (7, 14, 21 and 28 DAP). The levels of eGFP transcripts were computed based on the RT-PCR band intensities for each of the samples, which were normalized to the internal control, Actin2. No significant differences on the basal level of eGFP mRNA expression were observed within an individual promoter from 7-28 DAP (Figure 3.13). Likewise, the level of eGFP mRNA derived from each promoter was consistent during the four developmental stages, which implies a fairly uniform expression during the early stages of development. As observed at the microscopic level, the Cp29 and Cp45 promoters consistently produced high basal eGFP transcript levels similar to eGFP mRNA levels driven by the CaMV35S promoter. The Cp9 promoter had moderate to high, while Cp35 had much higher levels of mRNA transcripts despite the low level of fluorescence previously shown in the epi-fluorescence analysis (Figs. 3.9-3.12). Thus, the lower level of eGFP fluorescence in the Cp35-eGFP lines as observed microscopically does not correlate with the higher relative eGFP mRNA levels, which were more comparable to the mRNA levels obtained with Cp29 and Cp45 and promoters.
This suggests a post-transcriptional or translational influence on the final eGFP protein levels in Cp35-eGFP lines.

<table>
<thead>
<tr>
<th>MW</th>
<th>35S</th>
<th>Cp9</th>
<th>Cp29</th>
<th>Cp35</th>
<th>Cp45</th>
<th>WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 DAP</td>
<td>600 bp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 DAP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 DAP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 DAP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.13.** Semi-quantitative RT-PCR analysis (30 cycles) of eGFP and Actin2 transcript whole seedlings. Three independent T3 homozygous Arabidopsis lines, one per lane, were evaluated for each papaya promoter:eGFP fusion (Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP). For each line, four stages of plant development (7, 14, 21 and 28 DAP) were assayed. The four papaya promoters Cp9, Cp29, Cp35, and Cp45 were compared with the CaMV35S-eGFP (positive control) and non-transgenic Wild-type (WT) was the negative control. The molecular weight (MW) is shown using a DNA size ladder in base pairs (bp).

To further substantiate and compare the data from epi-fluorescence and RT-PCR analysis, we verified the eGFP expression at the protein level using immunoblot analysis and an anti-GFP antiserum (Figure 3.14). The eGFP protein expression driven from the Cp29 and Cp45 promoters were consistently high from 7-28 DAP, which was similar to eGFP protein levels derived from the CaMV35S promoter-eGFP construct. The eGFP protein levels in Cp35:eGFP plants correspond to the lower level of fluorescence detected in the tissues examined through epi-fluorescence microscopy. As shown in Figure 3.14, the level of eGFP protein was lowest for Cp35:eGFP, despite high levels of eGFP mRNA, thus further confirming the occurrence of a post-transcriptional or translational effect.
Figure 3.14. Western-blot analysis of eGFP (27 KDa) and Actin2 (loading control, 45KDa) protein levels in three independent homozygous *Arabidopsis thaliana* T₃ lines (one per lane) for papaya promoter:eGFP fusion (*Cp9*:eGFP, *Cp29*:eGFP, *Cp35*:eGFP, *Cp45*:eGFP) in four different stages of plant development. Single, whole seedlings were used for each protein extraction at 7, 14, 21 and 28 DAP. *Arabidopsis* wild type (WT) was used as negative control.

**b. Reproductive Stage (1.5 months)**

We conducted semi-quantitative RT-PCR analysis at the reproductive stage similar to the early developmental stages (7-28 DAP), but observed significant differences in the expression from the different tissue samples (data not shown). We then decided to further subject the RNA samples from the reproductive stage (flowers, leaves, stems and roots) for analysis using real-time quantitative PCR analysis (qRT-PCR) to clearly assess the differences in the strength and activity of the papaya promoters in *Arabidopsis thaliana*. The changes in eGFP gene expression in the different organs was quantified and normalized to the stably expressed reference gene Ubiquitin-Conjugating Enzyme 21 (AT5G25760) (Czechowski et al., 2005) (Figure 3.15). As expected, CaMV35S constitutive promoter had the highest expression in all tissues analyzed. Variations in the transcript expression were observed among the promoters and among the tissues of the same promoter (Figure 3.15).
**Figure 3.15.** Quantitative comparison of transcript levels in different *Arabidopsis thaliana* tissues (flowers, leaves, stems, roots) at the reproductive stage (1.5 months old). For each *Carica papaya* promoter:eGFP fusion, three independent T₃ transgenic lines were assayed in triplicate. Levels of eGFP mRNAs were measured using real-time PCR, normalized to the internal control gene *Ubiquitin-Conjugating Enzyme 21* (AT5G25760) (*PEROXIN4*). The calibrator used for measuring the gene expression change ($2^{-\Delta\Delta CT}$) was the nontransformed *Arabidopsis* wild type (WT) Col-0 (Czechowski et. al, 2005). Three independent lines transformed with CMV35S:eGFP were used as a constitutive control. The error bars represent standard errors.

The qPCR analysis indicated that *Cp9*-eGFP gave the lowest RNA levels, followed by the *Cp35* promoter, whereas eGFP RNA levels from CaMV35S promoter, *Cp29* and *Cp45* were moderately higher in all of the tissues. Although RNA levels for an individual promoter varied moderately for its respective three independent transgenic lines, the RNA levels were relatively consistent, and reflected actual differences in eGFP
RNA levels between promoters. From a tissue basis, eGFP RNA levels were highest in the leaf for all promoters followed by the stem, root and flower (Figure 3.15). Interestingly, eGFP RNA levels in roots in Arabidopsis were fairly similar and abundant between the four promoters relative to the other tissues. Because of the differences in developmental expression and possible post-transcriptional effects, we next sought to determine if eGFP protein levels were correlated with RNA levels observed.

The abundance of eGFP protein was analyzed in the same reproductive stages as for RNA analysis (Figure 3.16). The Cp29 and Cp45 promoter:eGFP fusions consistently showed high levels of eGFP protein, but no clear tissue specificity was observed. The Cp9-eGFP lines contained the lowest eGFP protein levels (flower, leaf, stem and root) as compared to eGFP present in the other three promoter lines. Cp35-eGFP protein levels were also low in all tissues except roots at the mature reproductive stage. These lower protein levels contrast with higher relative eGFP levels in the Cp9-eGFP and Cp35-eGFP lines at the younger stages of growth (Figure 3.14, Figure 3.16).

**Figure 3.16.** Western-blot analysis of eGFP (27 KDa) and Actin2 (loading control, 45KDa) protein levels in three independent homozygous Arabidopsis thaliana T3 lines of papaya promoter:eGFP fusion (Cp9:eGFP, Cp29:eGFP, Cp35:eGFP, Cp45:eGFP) in four different tissues (flower, leaf, hypocotyl, root) of flowering stage plants (1.5 months old). The CaMV35S-eGFP fusion was a positive control. Wild type (WT) Col-0 was used as negative control.
**5’-RLM-RACE to Identify the Transcription Initiation Site (TSS)**

Each of the expression constructs contained the promoter and 5’ UTR region up to the ATG codon. The question then arises whether transcription initiation occurs at the same site in *Arabidopsis* as in papaya. This is important as this further determines transcriptional and promoter accuracy, but also it determines the size and sequence of the mRNA 5’ UTR which impacts post-transcriptional regulation (Araujo et al. 2012). Therefore, we experimentally identified the transcription start sites (TSS) of the eGFP transcripts in the *Arabidopsis* promoter lines and native genes in papaya (Figure 3.17, Table 3.4).

**Table 3.4.** A comparison of the transcription start sites (TSS) and positions of the 5’-untranslated regions (UTR) of the four native papaya promoters expressed in *Arabidopsis thaliana* relative to *Carica papaya*. TSSs were determined by 5’RACE and cDNA sequencing. Nucleotide positions are indicated upstream relative to the translation start site (ATG codon).

<table>
<thead>
<tr>
<th>Native Papaya Promoter</th>
<th>Arabidopsis</th>
<th>Papaya</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSS</td>
<td>Nucleotide Position</td>
</tr>
<tr>
<td><em>Cp9</em></td>
<td>G</td>
<td>-117</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cp29</em></td>
<td>C</td>
<td>-24</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>-58</td>
</tr>
<tr>
<td><em>Cp35</em></td>
<td>A</td>
<td>-84</td>
</tr>
<tr>
<td><em>Cp45</em></td>
<td>C</td>
<td>-97</td>
</tr>
</tbody>
</table>

*From Porter et al. (2008).*
<table>
<thead>
<tr>
<th>Native Papaya Promoter</th>
<th>Plant</th>
<th>Transcription Start Site at the 5’ UTR</th>
<th>Position</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A. thaliana</td>
<td>GATTCATCAAGTCATGACCTTTTTTTACGTAGAGAGAAGAAAGAAAAATGGCTGCAAAGCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGCCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGC</td>
<td>-117</td>
<td>Experimental</td>
</tr>
<tr>
<td>Cp9</td>
<td>C. papaya</td>
<td>GATTCATCAAGTCATGACCTTTTTTTACGTAGAGAGAAGAAAGAAAAATGGCTGCAAAGCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGCCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGC</td>
<td>-117</td>
<td>Experimental</td>
</tr>
<tr>
<td></td>
<td>C. papaya</td>
<td>CTCTGAGGAAATCAGCCCTTTTTTTACGTAGAGAGAAGAAAGAAAAATGGCTGCAAAGCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGCCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGC</td>
<td>-126</td>
<td>Experimental</td>
</tr>
<tr>
<td>Cp29</td>
<td>A. thaliana</td>
<td>CTCTGAGGAAATCAGCCCTTTTTTTACGTAGAGAGAAGAAAGAAAAATGGCTGCAAAGCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGCCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGC</td>
<td>-24</td>
<td>Experimental</td>
</tr>
<tr>
<td></td>
<td>A. thaliana</td>
<td>ATCTAATCCCTTCTCTCTCTCTGGTTTCTCAGAGAGAAGAAAGAAAAATGGCTGCAAAGCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGCCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGC</td>
<td>-56</td>
<td>Experimental</td>
</tr>
<tr>
<td></td>
<td>C. papaya</td>
<td>CTCTGAGGAAATCAGCCCTTTTTTTACGTAGAGAGAAGAAAGAAAAATGGCTGCAAAGCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGCCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGC</td>
<td>-44</td>
<td>NCBI Data (Porter et al. 2008)</td>
</tr>
<tr>
<td>Cp35</td>
<td>A. thaliana</td>
<td>AGCAAGATCCTCAGAACAACATACATACATACATAAACAGCTCAAATGGATTCTCTCTCTGATTCATTCACCACCAACGATCTACCCATCTACCTTCTCAGAGAGAAGAAAGAAAAATGGCTGCAAAGCCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGC</td>
<td>-84</td>
<td>Experimental</td>
</tr>
<tr>
<td></td>
<td>C. papaya</td>
<td>AGCAAGATCCTCAGAACAACATACATACATACATAAACAGCTCAAATGGATTCTCTCTCTGATTCATTCACCACCAACGATCTACCCATCTACCTTCTCAGAGAGAAGAAAGAAAAATGGCTGCAAAGCCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGC</td>
<td>-84</td>
<td>Experimental</td>
</tr>
<tr>
<td>Cp45</td>
<td>A. thaliana</td>
<td>CTCTGAGGAAATCAGCCCTTTTTTTACGTAGAGAGAAGAAAGAAAAATGGCTGCAAAGCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGCCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGC</td>
<td>-97</td>
<td>Experimental</td>
</tr>
<tr>
<td></td>
<td>C. papaya</td>
<td>CTCTGAGGAAATCAGCCCTTTTTTTACGTAGAGAGAAGAAAGAAAAATGGCTGCAAAGCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGCCCTTTTTTTGCTGGTTTCTCAGAGAGAAGAATGGCTGCAAAGC</td>
<td>-99</td>
<td>Experimental</td>
</tr>
</tbody>
</table>

**Figure 3.17.** Nucleotide sequence of the 5’-UTR of the different promoters and the comparison of the transcription start sites (TSS) and positions of the 5’-untranslated regions (UTR) of the four native papaya promoters expressed in *Arabidopsis thaliana* relative to *Carica papaya*. TSSs were determined by 5’RACE and cDNA sequencing. Nucleotide positions are indicated upstream relative to the translation start site (ATG codon).

When expressed in Arabidopsis, two of the four papaya promoters (Cp9 and Cp35) had the exact TSS’s or were a few nts distal to the native TSS’s in papaya (Cp45). Cp9 utilizes nt G at position -117 in both papaya and *Arabidopsis*. In papaya, Cp9 utilizes an alternative minor transcription site (-126, which occurred at a low level). Cp35 utilized nt A at position -84 for both *Arabidopsis* and papaya. Cp29 utilizes two different TSS’s between *Arabidopsis* and papaya. The first Cp29 TSS in *Arabidopsis* is located at position -24 (nt C), while the second is at position -58 (nt A). Cp29 in papaya begins at position -44 (nt C), but is based on the cDNA sequence (Porter et al. 2008) and not a
Figure 3.18. Predicted secondary structures of the 5’ UTR of the *Carica papaya* promoters. Stem-loops were found from the RNAfold webserver ([http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi](http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi)). [nt – nucleotide]. The favorable free energy of formation ΔG (kcal/mol) is indicated.

Table 3.5. *Carica papaya* promoters calculated minimum free energy of the 5’ UTR sequences secondary structures obtained from RNAfold Webserver ([http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi](http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi)).

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Gene</th>
<th>5’ UTR length (nt)</th>
<th>ΔG (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cp9</em></td>
<td>Peroxidase</td>
<td>117</td>
<td>-29.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>126</td>
<td>-31.00</td>
</tr>
<tr>
<td><em>Cp29</em></td>
<td>B-1,3-glucanase</td>
<td>24</td>
<td>-0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44</td>
<td>-0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>56</td>
<td>-3.90</td>
</tr>
<tr>
<td><em>Cp35</em></td>
<td>Ferulate-5-hydroxylase</td>
<td>84</td>
<td>-4.74</td>
</tr>
<tr>
<td><em>Cp45</em></td>
<td>Hypersensitive-induced</td>
<td>97</td>
<td>-9.19</td>
</tr>
<tr>
<td></td>
<td>response</td>
<td>99</td>
<td>-9.20</td>
</tr>
</tbody>
</table>

Stable structures that can potentially inhibit translation initiation are those with ΔG < -30 kcal/mo.

precise mapping assay. The *Cp45* TSS was identified as position -97 (nt C) in *Arabidopsis*, while the papaya TSS resides only two nts upstream at position -99 (nt A).

The four promoters provide resulting mRNAs with 5’UTRs of differing length. The 5’UTR region can form secondary structures (stem loop, Figure 3.18) that might
modulate translation as indicated by their ΔG (Table 3.5). These differences could account for the altered functionality of this promoter when expressed in a system that generates variation (Tanaka et al. 2009).

**Discussion**

Here we used a simple heterologous system (*Arabidopsis*) to rapidly identify and analyze the strength and activity *in planta* of four promoters from four genes (*Cp9, Cp29, Cp35, Cp45*) of the tropical tree, papaya. We determined that the upstream papaya gene regions could efficiently drive transcription of the eGFP-reporter transgene based on analysis of eGFP fluorescence and measuring eGFP RNA levels in a variety of tissues and developmental stages. Unlike the previous studies, we also analyzed the eGFP protein levels and promoter TSSs. Identifying the TSSs demonstrated the accuracy of transcription, determined the length of 5’UTR and possible inherent translational regulatory structures. This is important consideration for deciding upstream regions to use for bioengineering purposes. Comparing eGFP protein and RNA levels allowed the identification of translational regulatory effects on final protein levels. As discussed below, such regulation could be mediated by sequences (secondary structures) in the 5’ UTR. Based on the vascular, hypocotyl and root expression, two promoters and their 5’UTRs (*Cp29* and *Cp45*) were selected for further downstream applications for bioengineering in papaya (described below).

The four promoters contained putative *cis*-regulatory elements related to light regulation, stress, defense and hormone responses, and tissue-specific expression. This is consistent with their being derived from pathogenesis-related genes, *Cp9, Cp29* and
Cp45, and an abiotic stress-response gene, Cp35 (Porter et al. 2008; Porter et al. 2009). The total number of cis-elements predicted by the bioinformatics algorithms could reflect promoter functionality and provide a basis for further testing via mutagenesis in planta in response to biotic and abiotic stresses (Table 3.3, Supplementary Table 5).

Overall, the Arabidopsis transcriptional machinery readily recognized the papaya promoters. This expression system provided information on preferential tissue-expression and a ranking of promoter strength. Most notably, the constitutive CaMV35S and Cp45 promoter were highly expressed throughout developmental stages 7-28 DAP especially in the vasculature. Vascular expression was a common feature of the promoters regardless of strength. The overall expression ranking was CaMV35S > C45 > Cp29 > Cp9 > Cp35 (Figure 3.9-3.12). Coussens et al. (2012) previously found Brachypodium promoters, pBdEF1a and pBdUBI10, were expressed stronger and constitutive in the maize.

Interestingly, there was clearly a developmental decrease in the levels of eGFP RNA and protein from different promoter-eGFP reporters between the earlier growth stages (7-28 DAP, Figure 3.13 and 3.14) and the mature reproductive stage (Figure 3.15 and 3.16). High expression in young plants lessened to various degrees depending on the promoter as plants matured. Cp29 expression lessened the least upon maturation. Analyzing gene expression in discrete aerial tissues that had differentiated in large mature plants was more straightforward as opposed to using smaller seedlings (7-28 DAP). Thus microscopic analysis of eGFP fluorescence was the primary tool to examine various tissues from 7-28 DAP seedlings. In addition, some variability between the three independent transgenic lines was evident in the mature reproductive stage, perhaps due to an effect of the transgene insertion site (Furtado et al. 2009).
The expression of the four papaya genes was primarily confined to the roots and hypocotyls (Porter et al. 2008), however they had a broader expression pattern when the promoters were expressed in *Arabidopsis*. These differences can be attributed to *cis*-element recognition, *trans*-factors, and downstream signal pathways controlling tissue development and specificity in the distantly related *Arabidopsis*. Previously, seed specificity of wheat and barley storage protein promoters was lost when they were expressed in *Oryza sativa* (Furtado et al. 2009). Such results reinforce the idea that it is difficult to predict promoter expression behavior across heterologous systems, even when evolutionarily closely related (Obertello et al. 2005; Furtado et al. 2009). Likewise, Andersson et al. (1997) observed that the cell-specific expression of the non-legume hemoglobin gene promoter changed in different plant backgrounds in response to hormone or oxygen tensions. Finally, important elements may be located within the translated region and intron (Rose, 2008) of the donor gene or upstream or downstream of the tested sequences.

It is well established that the RNA 5′- and 3′-UTRs can affect the tissue specificity, environmental response, and degree of gene expression at post-transcriptional levels, such as RNA stability and translation (Manzara et al. 1991; No et al. 2000; Kawaguchi and Bailey-Serres, 2005; Araujo et al. 2012). For example, the 5′- and 3′-UTRs of the potato Sus3 and Arabidopsis Myb genes are required for proper sucrose-inducibility and tissue-specificity (Larkin et al. 1993; Fu et al. 1995a, 1995b). Translation of Arabidopsis mRNAs is primarily affected by control elements in the 5′UTR and 3′UTR (Kawaguchi and Bailey-Serres, 2005). However, since cDNA construction proceeds from the 3′ ends of mRNA, these reverse transcripts are frequently lacking key
sequence information from the 5′ ends of the RNA, especially the TSSs. This is also reflected in the GenBank sequences. Therefore, defining the precise TSS via RACE was critical to determine the true length of the 5′UTR. It is important to note that all four genetic constructs used here contained the same 3′-UTR and eGFP coding sequence. They differed only in the promoter regions and associated RNA 5′-UTRs. For the Cp35:eGFP and Cp9: eGFP transgenes, mRNA levels were similar to the levels of Cp45:eGFP and Cp29:eGFP mRNAs in roots (Figure 3.15), yet eGFP protein levels in Cp35:eGFP and Cp9:eGFP plants were significantly lower than in the roots of plants with the Cp45:eGFP and Cp29:eGFP fusions (Figure 3.16). This post-transcriptional effect was also observed in the flowers for Cp35:eGFP.

The unique 5′ UTRs could be the basis for lower protein levels via lower translational efficiency of the Cp9 5′UTR- eGFP mRNA (Park et al. 2010; Araujo et al. 2012). Therefore, we suggest that Cp9-eGFP, and perhaps Cp35:eGFP, are undergoing translational down-regulation leading to lower eGFP. The 5′ UTRs were analyzed for the formation of the secondary structures (e.g. hairpin, stem loop), which can function as major regulatory tools in RNA stability and/or translational regulation (Table 3.5, Figure 3.18) (Araujo et al. 2012). Secondary structures situated close to the cap with a free energy of −30 kcal/mol would be sufficient to affect the pre-initiation complex binding to the mRNA (Gray and Hentze, 1994; Araujo et al. 2012), as observed in the Cp9 5′UTR (Figure 3.18).
Conclusion

In summary, four different gene promoters were isolated from papaya and were identified and expressed in Arabidopsis thaliana at all stages of plant growth. The mechanism of transcriptional and translational regulation of gene expression involves complex processes of DNA-protein, RNA-protein and protein-protein interactions that diverge in response to environmental conditions and developmental cues (Araujo et al. 2012; Manzara et al., 1991). Therefore, further investigation including functional dissection of these promoters and 5’ UTRs through mutation or deletion experiments and its response to different cues can potentially identify control processes. Two promising strong promoters with root and hypocotyl vascular expression were identified which are Cp29 and Cp45. The results also provide further evidence on the influence of the correct 5’UTR, and not just the promoter, on properly controlling transgene expression. Finally, this work provides a useful model for investigating the regulatory mechanisms of tropical plant promoters in Arabidopsis for further detection of related cis-elements, trans-acting factors and 5’-UTR regulation. Based on this work, these promoters are now being fused to pathogen resistance genes to transform papaya plants to directly assess their strength, activity, responses to pathogens, and phenotypes.
CHAPTER IV

CARICA PAPAYA NATIVE PROMOTERS CAPABLE OF DRIVING STILBENE SYNTHASE EXPRESSION IN ARABIDOPSIS THALIANA FOR RESVERATROL GLUCOSIDE (PICEID) SYNTHESIS

Abstract

The use of promoters for downstream metabolic engineering in plants requires efficiently assessing not only their transcriptional strength, but also the levels of the desired metabolite. Here we compared the effectiveness of two differentially expressed native papaya promoters (Cp9 and Cp29) with the CaMV35S promoter in driving transcription of the stilbene synthase gene (Vst1) and production of the phytoalexin trans-resveratrol glucoside (piceid) in the model system Arabidopsis thaliana (var. Columbia). Single transgene copy numbers were verified via Southern hybridization analysis of at least three independent transgenic lines for each promoter construct. The CaMV35S and Cp29 promoters consistently produced high Vst1 mRNA levels as detected by qPCR analysis, whereas these transcripts were 50-60% less in the Cp9-Vst1 lines. The transcripts for the evolutionarily-related chalcone synthase (CHS) gene were reduced by 20% and 40% relative to WT in the Cp29- and 35S-Vst1 lines, respectively, but were unchanged in the Cp9-Vst1 lines. The transgenic lines accumulated piceid as identified through HPLC and tandem mass spectrometry (MS/MS). Piceid levels were highest in the 35S-Vst1 followed by Cp29-Vst1 and Cp9-Vst1 in four week-old plants and immature green siliques (1.5 months). No overt deleterious phenotypes were observed in the Cp9- and Cp29-Vst1 lines, however, 35S-Vst1 produced smaller plants. The levels of secondary metabolites (anthocyanins) and seed pigments (tannins) were decreased in the CaMV35S-
73

VSt1 and Cp29-VSt1 lines most likely due to competition between CHS and stilbene synthase for precursors p-coumaroyl- and malonyl-CoA. The Cp29-Vst1 expression in Arabidopsis produced adequate levels of piceid for future disease resistance studies.

Introduction

Plant metabolic engineering has improved the composition of stilbenoid polyphenols in crops by introducing stilbene synthase genes into plants leading to pathogen-resistance and novel foods (Delaunois et al. 2009; Giovinazzo et al. 2012). Stilbene synthase, which is not present in Arabidopsis and papaya, catalyzes the biosynthesis of the potent anti-pathogenic phytoalexin, trans-3,5,4’ trihydroxystilbene, commonly known as resveratrol. Stilbene synthase is a member of the polyketide superfamily, which includes chalcone synthase (CHS) (Dao et al. 2011), and competes with CHS for precursor substrates. These precursors (one molecule of p-coumaryl-CoA and three molecules of malonyl-CoA) are ubiquitous in plant cells, and adding stilbene synthase is sufficient to divert the phenylpropanoid pathway (flavonoid biosynthesis) to stilbenoid synthesis (Hain et al. 1990). Once synthesized, resveratrol can be stabilized by glycosylation (Marrs et al. 1995; Regev-Shoshani et al. 2003).

Stilbene synthase has been engineered into a number of crops that synthesized resveratrol in the plant to either improve their resistance to biotic and abiotic stresses or to increase its nutritional value. Crops engineered that produced significant amount of resveratrol and its derivatives included tobacco (Hain et al 1990), alfalfa (Hipskind and Paiva 2000), poplar (Giorcelli et al. 2004), tomato (Thomzik et al. 1997; Giovinazzo et al. 2005), and strawberry (Hanhineva et al. 2009). Similarly, resveratrol treatment in non-plant organisms has realized longevity benefits in S. cerevisiae (Howitz et al. 2003),

Issues with genetic engineering ectopic resveratrol production include unbalanced expression in key tissues, post-transcriptional gene silencing, adverse phenotypes and depletion of precursors that effect phenylpropanoid metabolism (Delaunois et al. 2009). The choice of promoter to drive the transgene is critical. Previous studies expressing stilbene synthase have mainly used a narrow number of promoters, such as the strong constitutive cauliflower mosaic virus 35S promoter (*CaMV35S*) or the native grapevine *Vst1* stress-induced promoter (Fischer et al. 1997; Zhu et al. 2004; Fan et al. 2008). However, the use of the *CaMV35S* promoter triggered over-accumulation of resveratrol and caused a drastic depletion of the endogenous pools of precursors in organs such as flowers, leading to physiological anomalies such as male sterility (Fischer et al. 1997; Delaunois et al. 2009). Improvements in promoter use included the fungus-inducible promoter PR10.1 (Coutos-Thevenot et al. 2001) or the tissue-specific promoter p-nap (Husken et al. 2005). Therefore, the choice of optimal and regulated promoters for metabolic engineering in plants is highly desirable.

Previously, the expression of the grapevine stilbene synthase gene (*Vst1*) using the *Vst1* promoter in papaya enhanced resistance by 40% to the devastating wilt disease caused by *Phytophthora palmivora* (Zhu et al. 2004). However, the expression of this native grapevine promoter in papaya was sub-optimal. The stilbene synthase mRNA was
not detected immediately after inoculation, indicating that pathogen infection could be established. Therefore, to obtain more useful promoters for improving papaya disease resistance, we identified and characterized native papaya promoters with several useful features (Carlos-Hilario et al. 2014). They were derived from genes whose expression was substantially enhanced in roots and stems in young and old plants but limited in flowers (Porter et al. 2008) and were also pathogen-inducible (Porter et al. 2009).

We recently screened and tested these promoters as transcriptional fusions with the green fluorescent reporter gene (eGFP) in the model system Arabidopsis thaliana (Carlos-Hilario et al. 2014). The Arabidopsis transcriptional machinery readily recognized the papaya promoters. Qualitative and quantitative measurements of eGFP expression at both mRNA and protein levels indicated eGFP was expressed in a variety of tissues of the transgenic plants (vasculature, shoot apex, cotyledon, cotyledon petioles, hypocotyls, and root). Two promoters with strong and weak expression, Cp29 and Cp9, respectively, covered the range of expression extremes in roots and stems. In this report, the usefulness of these promoters was further characterized in comparison with expression from the CaMV35S promoter. They were transcriptionally fused to Vst1 and expressed in the efficient Arabidopsis system. Single-copy independent promoter-Vst1 transformants were characterized for stilbene synthase expression and its impact on phenotype, CHS expression, anthocyanin production, and the biochemical output of the desired resveratrol glucoside, piceid. In addition, the effects of multi-gene copy number were also examined. The research shows that Arabidopsis is an efficient system for screening promoters and Vst1 expression for subsequent use in the complex tropical tree, C. papaya to improve its disease resistance.
Materials and Methods

Vector Construction of pCambia 1302-papaya promoter-Vst1

The vector pCambia1302 containing the Cp9-eGFP and Cp29-eGFP constructs (Carlos-Hilario et al. 2014) was used as vector backbone with the Vst1 gene replacing the eGFP gene. The Vst1 gene was amplified using the primer pair (5’-GAGTTCTGATAGCTAGGCATGGCTTC-3’ and 5’-CACGTGTCAATTTGTAACCATAGGAACG-3’, the underlined bases represent added Ncol and Pmll restriction sites, respectively. The resulting 1535 bp PCR product was purified from an agarose gel using the QIAquick Gel Extraction Kit (Qiagen Inc., Valencia, CA, USA) and ligated into the TOPO-TA vector (Invitrogen, Carlsbad, CA, USA). The Vst1 gene was released by double digestion using Ncol and Pmll restriction enzymes and the correct sequence was verified by DNA sequencing. Vst1 was then cloned between the Ncol and Pmll restriction sites of the pCambia1302 vector replacing the eGFP gene to generate the construct pCambia1302-papaya promoter-Vst1 (Figure 4.1). The CaMV35S:Vst1 was used as the positive expression control. The orientation and reading frame of all recombinant plasmids were verified by PCR and DNA sequencing. Recombinant plasmids were mobilized into Agrobacterium tumefaciens strain GV3101 competent cells through the modified freeze-thaw method (Jyothishwaran et al. 2007).
Figure 4.1. Schematic representation of the transformation vector (pCAMBIA 1302) and the corresponding Carica papaya promoter:Vst1 gene fusions. (a) The plant selectable marker hygromycin phosphotransferase gene (hptII) is driven by the Cauliflower mosaic virus 35S promoter (CaMV35S) with a poly-A terminator. The native papaya promoter replaced the pCAMBIA 1302 CaMV35S promoter upstream of the stilbene synthase gene (Vst1) gene with downstream nopaline synthase Poly-A terminator (nos Poly A) maintained; (b) The length, in base-pairs (bp) of the putative promoter relative to the translational start site (ATG) of each papaya promoter:Vst1 fusion were shown.

Transformation of Arabidopsis with the C. papaya promoter:Vst1 fusions

Arabidopsis thaliana L. Heynh (ecotype Columbia) wild type was used throughout the study. The seeds were surface sterilized, plated onto sterile ½ MS solid medium (Murashige & Skoog 1962) supplemented with 3% sucrose, placed in the dark for 2 days at 4°C, then transferred to 22°C with a 16-h-light/8-h-dark cycle for germination. Seedlings at 10-14 days were transplanted to soil and were grown until immature unopened flower buds were formed. Arabidopsis was transformed by the floral
dip method (Clough & Bent 1998) using Agrobacterium tumefaciens strain GV3101 carrying the recombinant constructs. Putative transgenic T1 seeds were selected in vitro on solidified MS medium with 50µg/ml hygromycin B selection (Sigma-Aldrich, St. Louis, MO, USA). The presence of the papaya promoter and Vst1 were initially confirmed through PCR of genomic DNA using the previously designed primers (Table 1.0). Genomic DNA was extracted from Arabidopsis as described (Doyle and Doyle 1990). The PCR was carried out to amplify the target sequence using 100 ng genomic DNA with the Bio-X-Act Short Reaction Mix (Bioline Inc., Taunton, MA, USA) in a final volume of 50µl. The binary vector used in the transformation protocol served as a positive control reaction. Eppendorf Master Cycler (Eppendorf, Hamburg, Germany) conditions were as follows: denaturation at 95°C for 5 min, annealing at 55°C for 30 sec, elongation at 72°C for 1.5 min, for a total of 30 cycles, followed by a final elongation of 7 min at 72°C. The PCR products were separated via 0.8% agarose gel electrophoresis and stained with 1:10,000 GelRed™ (Biotium, Inc., Hayward, CA, USA).

Table 4.1. Primers used in this study.

<table>
<thead>
<tr>
<th>Amplification</th>
<th>Primer</th>
<th>Sequence</th>
<th>Amplicon Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>35SF</td>
<td>5'-ATGGAGTCAAAGATTTAAATAGGACCTAACACAG-3'</td>
<td>537</td>
</tr>
<tr>
<td></td>
<td>35SR</td>
<td>5'-GTCCCGGTAAGTTCACTTCACTCTTCAAGAATG-3'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cp9F</td>
<td>5'-ctcgagCTTGGAACTTCTTTGACGACTTTCC-3'</td>
<td>2041</td>
</tr>
<tr>
<td></td>
<td>Cp9R</td>
<td>5'-ccatggCTCTGCTAAAGCGGAGAAAGCTGAGA-3'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cp29F</td>
<td>5'-ctcgagGCCAAATTTTGGAGAGAGAAACAC-3'</td>
<td>1445</td>
</tr>
<tr>
<td></td>
<td>Cp29R</td>
<td>5'-ccatggGCCAAATTTTGGAGAGAGAAACAC-3'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vst1F</td>
<td>5'-GATTCTGAGCTAGTAccatggCTTC-3'</td>
<td>1535</td>
</tr>
<tr>
<td></td>
<td>Vst1R</td>
<td>5'-catggCTTAATGGAACCATAGGAACGAC-3'</td>
<td></td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>AT5G25760F</td>
<td>5'-CTTCGATCGGAAATCTTCTTAA-3'</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>AT5G25760R</td>
<td>5'-CCGCGGATGGGATTTCACTTCACC-3'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vst1F</td>
<td>5'-GAGGAAATTTAAACCTGCTCAACGCTGAAAGGTC-3'</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>Vst1R</td>
<td>5'-TCCTCACTCAGTCATGTGCTCGCTCCTTAGT-3'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CHS</td>
<td>5'-ATGGTTGATGGGTGGTCTCTTG-3'</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>CHSR</td>
<td>5'-CTCGCGGCTGACGCTGTCCTTCT-3'</td>
<td></td>
</tr>
</tbody>
</table>

Sequences underlined and in lower case were introduced PstI (F) and NcoI (R) sites for Cp9 and Cp29; NcoI (F) and PmlI site (R) for Vst1.
**Southern Blot Analysis**

Forty micrograms of genomic DNA from one-month-old whole seedlings was digested with *EcoRI* overnight and size fractionated by 0.8% agarose gel electrophoresis (Sambrook et al. 1989). The gel was denatured with 0.5 M NaOH/1.0 M NaCl (30 min) and neutralized with 0.5 M Tris-HCl/1.5 M NaCl (30 min) with gentle shaking, then transferred overnight by capillary action to a nitrocellulose N\(^+\) membrane (GE Healthcare Biosciences Corp., Piscataway, NJ, USA) using 10X SSC buffer. The DNA was covalently crosslinked to the nitrocellulose N\(^+\) membrane by exposure to UV (Stratalinker, Stratagene, La Jolla, CA). The hybridization probe was prepared by PCR-amplifying the *Vst1* fragment (1535bp) from the pCAMBIA1302 binary vector using the *Vst1* primers described above. The PCR products were separated at 0.8% agarose gel (Bioline Inc., Taunton, MA, USA) and gel-purified with the QIAquick Gel Extraction Kit (Qiagen Inc., Valencia, CA, USA). The fragment was labeled with biotin and Southern hybridizations were carried out using North2South Complete Biotin Random Prime Labeling and Detection kit (Thermo Scientific, Rockford, IL, USA). *CaMV35S-Vst1* transgenic lines were initially evaluated for transgene copy number, presence of the *Vst1* mRNA and morphological growth. Based on the initial evaluation, three independent transgenic plants containing one copy of the transgene were used for the subsequent analysis. Lines with multicopy *CaMV35S-Vst1* transgenes were also analyzed for comparison. Transgenic seedlings *T\(_1\)* were self-pollinated for two generations to obtain the *T\(_3\)* homozygous plants, for the final analyses.
RNA Isolation and qPCR Analysis

Total cellular RNA was isolated from four week-old WT and transgenic Arabidopsis plants using the RNeasy Plant Mini kit and was treated with an on-column DNase treatment according to the manufacturer (Qiagen Inc., Valencia, CA, USA). RNA was extracted from four week-old whole seedlings of transgenic Arabidopsis plants. RNA concentration and rRNA ratio were quantified using spectrophotometer (Beckman and Coulter DU 730 UV/Vis). 1µg of total RNA was converted to cDNA using M-MLV reverse transcriptase and oligodT (Promega, Madison, Wisconsin, USA). Vst1 PCR primers 5’-GAGGAAATTAGAAACGCTCAACGTGCCAAGG-GTC-3’ and 5’-CTTCAACTCAGTCATGTGCTCGCTCTTAGT-3’ (Table 1.0) were designed to amplify a 159 bp Vst1 gene sub-fragment. Arabidopsis Ubiquitin-Conjugating Enzyme 21 (AT5G25760) (PEROXIN4) cDNA was used as the internal standard with primers 5’-CTGCGACTCAGGGAATCTTCTA-3’ and 5’-TTGTGCCATTGAATTGACCC-3’ (Czechowski et al. 2005). 1µL (50 ng) of cDNA product was combined with forward and reverse qPCR primers to a final concentration of 500nM. This reaction was then combined with an equal volume of 2x Roche Power SYBR Green Master Mix with ROX (Applied Biosystems, Carlsbad, CA), mixed and aliquoted in 10µL triplicates into a 384-well optical qPCR plate, then sealed with an Applied Biosystems optical adhesive film. Samples were run on a 7900HT qPCR machine with the following thermocycling parameters: 2 minutes at 50°C, 10 minutes at 95°C, then 40 cycles of 15 seconds at 95°C, 1 minute at 60°C. A dissociation step was performed at the end of the run to ensure only a single product was formed utilizing a +1°C ramp rate per second from 60°C to 95°C.
HPLC and Tandem Mass Spectrometry (MS/MS) Analysis of Resveratrol Content

Four week-old transgenic Arabidopsis plants were used for HPLC analysis along with the WT control. Metabolites were extracted from whole seedlings with 80% methanol. The samples were ground, vortexed vigorously for 1-2 minutes, and then incubated overnight with shaking at 100 rpm. Samples were centrifuged to remove insoluble debris. The supernatant was concentrated to dryness and the residue dissolved in 100µl of 100% methanol and 900µl of 2% acetic acid. The samples were centrifuged twice at 10,000 g for 10 minutes and 500 µl aliquots were stored in -20°C until ready to use for HPLC analysis.

The standard resveratrol glucoside (piceid) (Sigma-Aldrich, St. Louis, MO, USA) and samples of tissue extracts containing piceid were resolved by Reverse-Phase High Performance Liquid Chromatography (RP-HPLC), which consisted of a Waters 2695 automated HPLC system interfaced with a Waters 996 Photodiode Array Detection (PDA), scanning 210-600 nm. Peaks were resolved using a Kinetex C18 column (Phenomenex; 2.6µm; 100A; 100 x 4.6 mm) at a flow rate of 0.8mL/min and eluted with a linear 8%/min gradient of Solvent A (2% v/v acetic acid/Aq.) against Solvent B (2% v/v acetic acid/MeOH), with a 3 min re-equilibration period between sample injections. Resulting chromatographic profiles were extracted at 300nm, and the target peak area integrated using Waters Empower Pro software. Standards of piceid were analyzed in triplicate, allowing for the precise establishment of piceid concentrations reported as µg/fresh tissue weight.

Mass spectrometry experiments were undertaken using AB/MDS-Sciex API 3000 triple quadrupole mass spectrometer (Thornhill, Ontario, Canada). The ESI-MS system
was calibrated manually in positive mode with PPG 3000 (AB/MDS-Sciex) to achieve <5-ppm mass accuracy, as per manufacturer’s protocol. The mass spectrometry was interfaced with a PE ABI 140D Solvent Delivery System. Ultra-high purify N₂ was delivered by a Peak Scientific generator. RP-HPLC purified analyte fraction was injected as a 20uL bolus, at a flow rate 25µL/min, in a solvent of 70% MecN/30% 0.1% v/v formic acid/H₂O. Initially full ion spectra were taken from m/z 100-600 Da to provide identification of the parent molecular mass using Q-3.

For tandem mass spectrometry (referred to as MS/MS) experiments, collision-induced dissociation (CID) of the identified parent molecular mass was effected by bombardment with nitrogen. Bombardment was confined to quadrupole-2 (Q-2) with a collision cell gas thickness of 3 x10¹⁴ atoms/cm² and a collision energy (Q-0 to Q-2 rod offset voltage) typically set at ~20–40 eV. The resulting CID (daughter ion) spectra were obtained by scanning quadrupole-3 (Q-3) from m/z 100–400 Da. MS/MS data analysis was assisted with the use of Analyst Software (v1.4.1; AB/MDS-Sciex) and ACD ChemSketch (ACD/Labs).

**Quantitation of Anthocyanin of 5 day-old Seedlings**

To quantitate the anthocyanin levels in Cp promoter:Vst1 transgenic Arabidopsis lines, seeds were surfaced sterilized and plated on Murashige and Skoog (MS) basal medium without a nitrogen source (Caisson Laboratories, North Logan, UT, USA). The seeds were grown for 5 days under a 16 hr light/8 hr dark cycle. Accumulation of anthocyanins in seedlings was observed and quantitation was done as described previously (Shirley et al. 1995) with a few modifications. In brief, 25 Arabidopsis seedlings were harvested, weighed and soaked in 250 µl of 1% HCL (v/v) in EtOH
overnight at 4°C. Samples were ground and centrifuged briefly and the supernatant volume was measured. Anthocyanin levels in the supernatants were determined spectrophotometrically (A₅₃₀ nm) and total gram fresh weight of the seedlings was computed. Mean values were taken from three independent transgenic lines of each promoter-Vst1 line with three technical replicates.

Results

Transgene integration and morphological development of the transgenic plants

Independent transgenic Arabidopsis lines surviving in hygromycin (50ug/L) were screened for the presence of the Vst1 transgene and the respective promoter via PCR analysis. Then Southern hybridization was used to identify Arabidopsis lines (T₃ generation) harboring the CaMV35S:Vst1, Cp9:Vst1 and Cp29:Vst1 transgenes as a single copy (Figure 4.2). CaMV35S:Vst1 transgenic lines also were identified that had multiple copies of the transgene. Lines were also evaluated for morphological growth and phenotypes relative to WT.

We observed that the transgenic lines with more than one copy of the CaMV35S:Vst1 transgene have lower Vst1 expression than with the single insertion only (Figure 4.3). The Vst1 transcript levels decreased as the number of transgene increased. Moreover, the growth and development of the plants were affected (Figure 4.4). To avoid the adverse phenotypic effects of multiple transgene insertions, we utilized three independent transgenic lines containing one copy of the transgene (T₃ homozygous) for each promoter fusion for all subsequent analyses.
Figure 4.2. Southern blot analysis of EcoRI-digested genomic DNA extracted from *Arabidopsis* transgenic lines hybridized with 1.5 kb biotinylated stilbene synthase probe. The three promoters CaMV35S, Cp9 and Cp29 fused to the Vst1 gene are indicated above the panel. Transgenic plants with single insertion site (blue) utilized for subsequent analysis. Lanes: WT-Wild Type DNA from non-transformed control plant, P-plasmid DNA used as positive control. Number above the blot indicates numbered lines. Transgene copy number denoted below the lanes in the blot.

The selected transgenic plants with a single promoter- Vst1 insertion were initially evaluated at the two-week stage of development. Most of the plants have normal root and shoot development. CaMV35S-Vst1 line 1 had shorter roots and shoots relative to WT (Figure 4.5). After one month, the plants grown in soil exhibited minor variations in growth in the CaMV35S- and Cp29-Vst1 lines as compared to the WT control. CaMV35S-Vst1 line 1 had relatively smaller plants, although no dramatic phenotypic effects were observed (Figure 4.6). Cp29-Vst1 plants at one month stage were also slightly smaller, but grew vigorously at the reproductive stage to produce a significant amount of seed. On the other hand, Cp9-Vst1 resembled the WT at all stages of plant development.
Figure 4.3. Quantitative comparison of CaMV35S transgenic lines (a) transcript levels and (b) piceid concentration (µg per gram fresh weight) as affected by the number of transgene insertion in the genome of the plants. The levels of mRNAs were measured using real-time qPCR, normalized to the internal control gene Ubiquitin-Conjugating Enzyme 21 (AT5G25760) (PEROXIN4). The calibrator used for measuring the gene expression change (2−ΔΔCT) was the non-transformed Arabidopsis wild type (WT) (Czechowski et. al, 2005).
Figure 4.4 CaMV35S:Vst1 homozygous T₃ Arabidopsis (14 days after planting) showing normal growth and morphological development of line 37 with single insertion of the transgene; representative lines 4 and 3 showing phenotypic alterations in shoot and root parts of the plants caused by increased number of the transgene insertion compared relative to the non-transformed wild type (WT).

Comparison of Vst1 and CHS transcript levels

Quantitative Real-Time PCR (qPCR) analysis was performed to measure the relative Vst1 transcript levels among the three lines. In Figure 4.7a, the highest levels of Vst1 transcripts were found in CaMV35S-Vst1, followed by Cp29-Vst1 and Cp9-Vst1 lines. The differences in Vst1 transcript expression profiles are indicative of the strength and activity of the promoters in the transgenic plants. As expected, no Vst1 transcripts were detected in the wild type control plants. Arabidopsis plants contain significant levels of flavonoids and isoflavonoids (Viet et al. 1999, Lapcik et al. 2006) and the evolutionarily-related enzyme, CHS, can compete directly with stilbene synthase for the shared metabolic precursor substrates p-coumaroyl CoA and malonyl CoA. It is possible that the expression of Vst1 could affect Chs gene expression. Therefore, Chs mRNA
levels were also quantified in the same plants. The levels of Chs transcripts decreased in the CaMV35S-Vst1 and Cp29-Vst1 lines relative to WT (Fig. 4.7b), but Chs mRNA levels were unaffected in Cp9-Vst1 plants.
Figure 4.6. Morphological development of homozygous T3 *Arabidopsis* plants one month after planting transformed with (a) CaMV35S:*Vst1*, transgenic line 1, 7 and 37 (left to right); (b) *Cp9:Vst1*, transgenic line 15, 16, and 18; (c) *Cp29:Vst1* transgenic line 23, 24 and 27 and (d) WT- non-transformed wild type control plants.
Figure 4.7. Quantitative comparison of transcript levels in wild-type and different lines of *Arabidopsis thaliana*. (a) *Vst1* transcripts (b) *CHS* transcripts. Three independent T3 transgenic lines were assayed in triplicate from each promoter. The levels of mRNAs were measured using real-time qPCR, normalized to the internal control gene *Ubiquitin-Conjugating Enzyme 21* (AT5G25760) (*PEROXIN4*). The calibrator used for measuring the gene expression change ($2^{-\Delta \Delta CT}$) was the non-transformed *Arabidopsis* wild type (WT) (Czechowski et al. 2005). *CaMV35S:Vst1* were used as a constitutive control. The error bar represents the standard error.
Biochemical characterization of resveratrol levels in transgenic Arabidopsis lines expressing Vst1 under the control of different promoters

Four week-old transgenic lines were analyzed by HPLC for the accumulation of trans-piceid resveratrol (Fig. 4.8). The presence of trans-piceid was detected at significant levels in CaMV35S-Vst1 and Cp29-Vst1 lines, with a trace amount detected in Cp9-Vst1 (Fig. 4.8). The UV spectra of the piceid standard (red) and the transgenic lines match the peak at the 7.38 min retention time (A₃₀₀ nm). The biochemical outputs of piceid were 17.25 µg/gfw in CaMV35S-Vst1 followed by 15.38 µg/gfw in Cp29-Vst1 and 2.49 µg/gfw in Cp9-Vst1 (Table 4.2). The very low levels of piceid in Cp9-Vst1 is correlated with Cp9 being the weakest promoter used.

Piceid accumulation was also analyzed in immature green siliques (1.5 months). The presence of piceid in various quantities was detected in green siliques from all of the transgenic lines driven by CaMV35S, Cp9 and Cp29 promoters (Table 4.2). The amount of piceid recovered was positively correlated with the levels detected in the whole plant at the 4-week stage. Green siliques from CaMV35S promoter had the highest amount of piceid (24.89 µg/g fresh weight), followed by Cp29 promoter (14.57 µg/g fresh weight) and the least was in plants with the Cp9 promoter (4.20 µg/g fresh weight) (Table 4.2).

To verify the identity of the new compound in the tissues, samples of methanolic subjected to tandem MS/MS analysis after the HPLC peaks co-migrating with piceid were collected. The newly formed compound was verified to be resveratrol glucoside (Fig. 4.9), also known as trans-piceid or polydatin (Liu et al. 2011, Mattivi et al. 1995; Teguo et al. 1996). As shown in Figure 4.9a, the total ion chromatogram indicated major
Figure 4.8. HPLC analysis of extracts from seedlings (28 days after planting) of transgenic Arabidopsis plants showing the trans-piceid peak at 7.38 minute retention time compared to the commercial standard trans-piceid (red line) also known as polydatin (red). Individual peaks of the extracts isolated from WT and lines containing different promoter-Vstl are shown (black lines).
Figure 4.9. Trans-piceid mass spectra verified through tandem MS-MS with the corresponding fragmentation coefficient (m/z) of the compound. (a) Sample from Arabidopsis Cp29-line 23 (b) Standard piceid.

peak from transgenic Arabidopsis of m/z 149.2 [glucose] and the minor peaks m/z 229.4 [resveratrol], and m/z 167.2 which co-eluted with trans-piceid standard (Figure 4.9b). These peaks were not present in extracts of WT controls (data not shown).
Table 4.2. Piceid concentration of *Arabidopsis thaliana* homozygous T₃ transgenic plants.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Transgenic Line Number</th>
<th>Piceid (g/gram fresh weight)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4-week stage</td>
<td>Average</td>
<td>Green Siliques</td>
</tr>
<tr>
<td>CaMV35S</td>
<td>1</td>
<td>15.38 ± 1.89</td>
<td>17.25 ± 3.27</td>
<td>23.69 ± 6.94</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>16.51 ± 3.21</td>
<td></td>
<td>22.60 ± 7.014</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>19.87 ± 4.72</td>
<td></td>
<td>28.94 ± 3.484</td>
</tr>
<tr>
<td>Cp9</td>
<td>15</td>
<td>2.586 ± 0.40</td>
<td>2.49 ± 0.59</td>
<td>5.20 ± 1.39</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>3.19 ± 0.95</td>
<td></td>
<td>3.14 ± 0.430</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>1.72 ± 0.42</td>
<td></td>
<td>4.27 ± 0.712</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>11.589 ± 2.14</td>
<td></td>
<td>16.17 ± 1.820</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>14.61 ± 4.65</td>
<td></td>
<td>14.43 ± 3.55</td>
</tr>
</tbody>
</table>

**Anthocyanin concentration, plant height and seed pigmentation in Arabidopsis transgenic plants**

The observed decrease in *Chs* transcripts and the competition between CHS and stilbene synthase for common substrates may disrupt downstream metabolism. Anthocyanins are pigments in the phenylpropanoid pathway that may be impacted. Therefore, the amounts of anthocyanins in the transgenic plants were analyzed from the 5-day-old grown seeds without a nitrogen source. Anthocyanin levels were ~50% lower in the CaMV35S-Vst1 and Cp29-Vst1 lines relative to Cp9-Vst1 and WT (Fig. 4.1a). Despite the reduced anthocyanin levels in the plants, the overall height, and growth and development was not drastically affected (Fig. 10b).

The transgenic plants produced significant amounts of seeds (Fig. 4.11). However, seed pigmentation was visibly altered in CaMV35S-Vst1 and Cp29-Vst1 lines to a pale brown coat. This implies that elevated expression of stilbene synthase will compete for precursors in the flavonoid pathway affecting the formation of seed-coat tannins. The Cp29-Vst1 seeds have lesser reduction in seed pigmentation as compared to CaMV35S-7 and -37 lines.
Figure 4.10. Anthocyanin concentrations of 5-day old seedlings measured at 530 nm absorbance (A_{530}) per gram fresh weight of tissue (a); Plant height of the transgenic *Arabidopsis* at maturity measured from rosette base to flower apex (b). Three biological replicates with six technical replicates of each *Carica papaya* promoter:Vst1 constructs were used. *CaMV35S* constitutive promoter was used as positive control, WT- wild type non-transformed plants were used as negative control. The error bar represents the standard error.
Figure 4.11. Differences in the changes of seed coat pigmentation of *Arabidopsis* seeds transformed with the *Carica papaya promoter: Vst1* constructs with (a) *CaMV35S : Vst1*, transgenic line 1, 7 and 37 (left to right); (b) *Cp9:Vst1*, transgenic line 15, 16, and 18; (c) *Cp29:Vst1* transgenic line 23, 24 and 27 and (d) WT- non-transformed wild type control plants.
DISCUSSION

The present study efficiently utilized the simple *Arabidopsis* system to evaluate the expression and downstream metabolic output of heterologous transcriptional fusions of promoters derived from a complex tropical tree. Two native papaya promoters differing in strength were transcriptionally fused to the *Vst1* gene encoding stilbene synthase for the biosynthesis of the anti-microbial *trans*-resveratrol glucoside (piceid). The fusions were introduced into the *Arabidopsis* genome, independent transgenic lines were verified and inbred to homozygosity. Three independent lines were selected that contained single transgene insertions from each promoter-*Vst1* fusion. The *Vst1* gene expression at the mRNA level and the biochemical output of piceid were quantitatively determined. The work lays the foundation for using these promoters and the *Vst1* gene to genetically engineer papaya for resistance to the tropical wilt disease caused by *Phytophthora palmivora* and to improve nutraceutical antioxidant content.

Southern blot hybridization analysis determined transgene copy numbers from these lines. *Vst1* transcript levels were highest in the lines with a single copy of the *CaMV35S-Vst1* transgene, while *Vst1* transcript levels were lower in the lines with two or more copies (Lines 3, 4, 5, 6, 8, 10 and 11). Our results were consistent with the previous findings in white pine (Tang et al. 2007) and poplar (Giorcelli et al. 2004) that suggested homology dependent gene silencing. Recently, Liu et al. (2011) observed that resveratrol synthase (*PcRS*) transcripts were low or absent in *Polygonum cuspidatum* that have a high copy number of the transgene. In our study here, however, we also observed the multiple *CaMV35S-Vst1* transgene insertions had morphological effects (altered root and shoots) on the growth of the seedlings. This could be due to epistatic effects or insertion
of the transgene into important genes. To eliminate potential gene silencing and genome position effects due to multiple T-DNA insertions, we selected transgenic lines with one T-DNA insertion for further analysis of the promoter-\textit{Vst1} gene fusions.

Most of the transformations with stilbene and resveratrol synthases in various plant species were driven by the \textit{CaMV35S} promoter or its own homologous promoter (Hain et al. 1993, Thomzik et al. 1997, Stark-Lorenzen et al. 1997, Zhu et al. 2004). This study reports the first accumulation of resveratrol glucoside, piceid, using the heterologous native promoters isolated from papaya expressed in \textit{Arabidopsis}. RP-HPLC and mass spectrometry verified the biosynthesis of piceid in these plants. The biosynthesis of piceid has been shown in \textit{Polygonum cuspidatum} (Liu et al. 2011), white poplar (Giorcelli et al. 2004), kiwi (Kobayashi et al. 2000), \textit{Brassica napus} seeds (Husken et al. 2005) and papaya (Zhu et al. 2004). The conjugation of glucose to resveratrol is known to protect plant cells from any toxic effect while also protecting resveratrol from oxidation and degradation (Regev-Shoshani et al. 2003). For other secondary metabolites, such as flavonoids, the addition of sugar groups is well documented to enhance their solubility and stability, which is critical for transport and storage in the vacuole or cell wall (Marrs et al. 1995; Winkel-Shirley 2001a).

\textit{Vst1} transcript levels correlated with promoter strength as previously determined using GFP fusions (Carlos-Hilario et al. 2014). The significant amount of piceid in \textit{CaMV35S-Vst1} and \textit{Cp29-Vst1} lines (strongest promoters) contrasts with low amounts in the \textit{Cp9-Vst1} plants with the weak promoter. More importantly, metabolic output of piceid produced by the transgene-encoded enzyme, stilbene synthase, also correlated with promoter strength. Taken together, this shows that the \textit{Arabidopsis} system can be used to
faithfully assess heterologous promoter-gene combinations for biotechnology applications in other crops. This can save time when the original promoter is from complex plants, such as papaya, that are difficult to transform and have prolonged regeneration times. These transgenic Arabidopsis now represent important experimental material for disease resistance studies. In addition, the expression of piceid in the immature siliques as fruit tissue is important information for evaluating Vst1 expression in the edible part of the plant to improve its nutritional content.

Recent reports on plant transformation with stilbene synthase yielded differences in concentrations of resveratrol and its derivatives depending on the plant species and the extraction procedure used. The quantities of piceid detected here in four week-old plants using the CaMV35S and Cp29 promoters (15.4-19.9 µg/gfw) were similar to the range detected in transgenic alfalfa (HipSkind & Paiva 2000), kiwi (Kobayashi et al. 2000) and tomato (Giovanizzo 2005), but less than those in transgenic Arabidopsis (Liu et al. 2011), poplar (Giorcelli et al. 2004) and tobacco (Hain et al. 1993). Piceid accumulation in the four week-old plants was positively correlated with the amount of piceid found later in the reproductive stage. Some of the highest piceid concentrations reported (93.31-183.73 µg/gfw) were in Arabidopsis over-expressing the resveratrol synthase (PcRS) gene from Polygonum cuspidatum (Liu et al. 2011). However, this expression level resulted to drastic reduction of anthocyanin and seed pigmentation. In our transgenic CaMV35S and Cp29 plants, anthocyanin levels were reduced less and the seed pigmentation was qualitatively lower than when using the Polygonum gene.

The reduction in anthocyanins and seed tannins was direct evidence that there is competition for precursor substrates (p-coumaroyl CoA and malonyl CoA) between CHS
and ectopically expressed stilbene synthase. The reduction in these pigments was not as
dramatic as in the *Arabidopsis* tt4 mutants which resulted to yellow seeds and cotyledons
(Shirley et al 1995). The competition between *CHS* and stilbene synthase did not have
any apparent adverse or deleterious effects in growth and development of the transgenic
*Arabidopsis* lines. Tannins are known to have roles in seed development and other
physiological processes. Our results for *Cp29-Vst1* were consistent with the previous
report in transgenic alfalfa (Hipskind & Paiva 2004) and *Arabidopsis* (Liu et al. 2011)
where normal plants were generated. However, a shorter phenotype was observed in *Cp9-
Vst1* (Figure 10b) and *CaMV35S-Vst1* lines although it did not affect the reproductive
ability of the plants, which set ample seed. Stilbene accumulation does not affect plant
development, growth, fertility, morphology and other secondary metabolite pools
(Giovinazzo et al.2005, Schewendiek et al. 2007, Delaunois et al. 2009). In our
experiments, *Arabidopsis* lines that accumulated the highest levels of piceid exhibited
normal flower color, seed set, and germination relative to WT. In contrast, tobacco lines
constitutively expressing resveratrol synthase resulted in male-sterility and the loss of
flower pigment (Fischer et al. 1997). However, fertility was restored by the exogenous
application of flavonols, suggesting that resveratrol synthase depleted metabolic
intermediates required by *CHS*, arresting pollen development (Hipskind and Paiva 2000).

A more perplexing result was the decrease in expression of endogenous *CHS* in
*CaMV35S* and *Cp29 Vst1*-expressing *Arabidopsis* lines. These results suggest that the
significant reduction in flavonoid levels in transgenic plants may be due to suppression of
endogenous *CHS* expression. We speculate that the decrease in precursor substrates for
CHS lead to a feedback down-regulation of CHS expression. This may be a process to
attenuate the pathway when precursors are inadequate. Similarly, flavonoids and the endogenous CHS mRNA levels significantly decreased in strawberry (Hanhineva et al. (2009) and tomato plants (Giovinazzo et al. 2005) transformed with the grapevine stilbene synthase gene. Hanhineva et al. (2009) suggested that the decrease in CHS expression depended on regulatory control of the phenylpropanoid pathway. In contrast to these results, Liu et al. (2011) did not observe CHS down-regulation in Arabidopsis expressing resveratrol synthase. Determining the nature of the impact of stilbene synthase on CHS expression will require further experiments.

Conclusion

The present study efficiently utilized the simple Arabidopsis system to evaluate the expression and downstream metabolic output of heterologous transcriptional fusions of promoters derived from a complex tropical tree. This work lays the foundation for using these promoters and the Vst1 gene to genetically engineer papaya for resistance to the tropical wilt disease caused by Phytophthora palmivora and to improve nutraceutical antioxidant content.
CHAPTER V

MODIFIED AGROBACTERIUM-MEDIATED TRANSFORMATION OF CARICA PAPAYA VARIETY ‘KAPOHO’ FROM SUSPENSION-DERIVED CULTURES

Abstract

Alternative transformation protocol was developed for papaya using suspension-derived cultures. The ability of the tissues for plantlet regeneration using suspension cultures were demonstrated in the control non-transformed and transformed papaya tissues. Suspension-derived cultures maintained at globular young calli stage characterized by yellowish/creamy embryos which are identified to be highly receptive for Agrobacterium infection. The optimum parameters selected to obtain putative transgenic calli were the following: 1 day co-cultivation of Agrobacterium; cell density used for transformation was 0.005 at OD₆₀₀; and elimination of Agrobacterium post-transformation was done in combination of 250 mg/L carbenicillin and 250 mg/L cefotaxime. Putative transgenic calli were all confirmed positive for the presence of eGFP transgene, Cp45 promoter sequence, and hygromycin resistance gene (hptII) using PCR. eGFP mRNA transcripts expression and protein expression were detected using RT-PCR and Western blot analysis, respectively. eGFP fluorescence viewed under hand-held blue light and Kodak phosphorimager further confirmed the expression of Cp45:eGFP in the transformed papaya plantlets as shown in the fluorescence of root and stem tissues. This technique serves as an alternative way to generate transgenic plants in a simple laboratory setup that facilitates Agrobacterium-mediated transformation of papaya.
Introduction

Papaya is a widely cultivated and important tropical fruit commodity in the global market where current worldwide production has reached 10.5 metric tons with a value of 3.4 billion US dollars in 2009 (Fuentes & Santamaria 2014). Ripe papaya fruits are commonly eaten as fresh produce while the unripe fruits are rich source of papain which was first used as a commercially available meat tenderizer (FAO 2005). Papaya is also utilized in the pharmaceutical and cosmetics industries. In total world production, 52% mainly comes from India and Indonesia, 34% comes from America (Brazil and Mexico) and the remaining 14% is from Africa (Fuentes & Stamaria 2014). Here in Hawaii, papaya is the second largest fruit crop industry that exports fresh papaya fruits to US mainland and Japan.

Most of the improvements of important quality and horticultural traits in papaya were done through conventional (e.g. crosses resulting to hybrids) or through non-conventional breeding technique such as in genetic engineering. Transgenic papaya modified with Papaya ringspot virus (PRV) resistance was the first genetically modified tree and fruit crop, and the first transgenic crop developed by a public institution (Gonzalves 1998). The development of transgenic papaya with resistance to Phytophthora palmivora (Zhu et al. 2007), resistance to mites (McCafferty et al. 2006), long shelf life by inhibiting ethylene production (Laurena et al. 2002), and tolerance to herbicide (Cabrera-Ponze et al. 1995) or aluminum toxicity (Dela Fuente et al. 1997) were among the efforts in papaya genetic trait improvement. In addition, there had been attempts to use transgenic technology to produce vaccine against tuberculosis (Zhang et al. 2003) and cysticercosis (Hernandez et al. 2007). These recent advances in papaya
genetic engineering were successfully done through improvement of more suitable systems of transformation.

Successful plant transformation system highly rely on several factors such as the use of promoters, selection markers, tissue culture system, and the use of efficient delivery systems for transgene insertion in the plant’s genome. *Agrobacterium*-mediated and particle bombardment technology are two common methods for generating transgenic plants. Most of the transformation done in papaya genetic engineering utilized particle or biolistic bombardment technology that requires procurement of the sophisticated equipment such as the gene gun. Biolistic transformation which was first reported by Sanford et al. (1987) is a method that introduces the foreign genes into intact plant cells and tissues via bombardment with high velocity microprojectiles (Klein et al. 1988 a,b). In biolistics, efficient transformation requires introduction of genes with minimal tissue damage and several physical parameters should be optimized in transient expression assay before stable transformation can be done. Moreover, the difficulty in procurement of the gene gun in other laboratories limits the use of this technology. Biolistics or particle bombardment method seems to be more cumbersome than *Agrobacterium*-mediated transformation which is more simple and straightforward method which only requires competent *Agrobacterium* cultures carrying the T-DNA plasmid for co-cultivation. *Agrobacterium* method of transformation integrates only few copies of the disarmed T-DNA carrying the genes into the host genome (Smith & Hood 1995, Raineri et al. 1990) that leads to fewer problems with transgene cosuppression and instability (Hansen et al. 1997). In addition, *Agrobacterium* single-cell transformation system avoids forming mosaic plants, which are more frequent when direct

The first experiments done in papaya involved *Agrobacterium* infection of leaf discs and recovery of transgenic callus but there was no regeneration of whole plants (Pang & Sanford 1988). The first successful transformation in papaya using *Agrobacterium* was reported by Fitch et al. (1993). Since the last decade, various research laboratories in different countries have successfully used *Agrobacterium*-mediated transformation in generating transgenic papayas (Mendoza et al. 2008). Although there had been successes in *Agrobacterium*-mediated transformation, the adoption of the developed protocol seems to be more difficult in other varieties of papaya. There had been problems with the differences in the genotypic response where other variety e.g ‘Kapoho’ is harder to transform that variety ‘Sunset’ (Fitch et al. 1992). We have tried utilizing the previous *Agrobacterium* transformation technique (Zhu et al. 2006) for our tissue cultures but we have difficulty obtaining hygromycin-resistant calli. Carborundum treatment that creates wounding in the somatic embryos resulted in chalky white or bleached cultures that eventually died. Thus we tried to modify the *Agrobacterium* transformation system in our laboratory using suspension-derived cultures. We determined the optimum co-cultivation time, cell density for co-cultivation, and identified the right antibiotic combination for latent *Agrobacterium* elimination post-transformation. Here, we demonstrated the ability to obtain independent transgenic lines from the simple system optimized which could be used as an alternative technique in conducting papaya genetic trait improvement through genetic engineering.
Materials and Methods

Vector construction

Previously constructed $Cp45:eGFP$ vector in pCAMBIA 1302 vector backbone (Chapter III) was utilized for the transformation work (Figure 5.1). Recombinant plasmids were mobilized onto $Agrobacterium tumefaciens$ strain GV3101 competent cells through the modified freeze-thaw method (Jyothishwaran et al. 2007) for use in $Agrobacterium$ co-cultivation.

**Figure 5.1.** Schematic representation of the transformation vector (pCAMBIA 1302) and the corresponding papaya $Cp45$ promoter-eGFP gene fusion. The plant selectable marker hygromycin phosphotransferase gene ($hptII$) is driven by the Cauliflower mosaic virus 35S promoter (CaMV35S) with a poly-A terminator. The native papaya $Cp45$ promoter replaced the pCAMBIA 1302 CaMV35S promoter upstream of the enhanced green fluorescence protein (eGFP) gene with downstream nopaline synthase Poly-A terminator ($nos$ Poly A) maintained.

Preparation of papaya tissue cultures

The commercial solo papaya cv. ‘Kapoho’ was used for the study. Previously published procedures with a few modifications (Fitch et. al 1990) were used for producing papaya seedlings for hypocotyl tissue excision to produce primary callus for suspension culture establishment. The mucilage of papaya seeds was removed by mechanical abrasion while washing with soap and water. The seeds were surface sterilized using distilled water with 20% Clorox plus Tween 20 ($40\mu l/L$) for 1 h with...
shaking (100 rpm at 25°C). Seeds were rinsed three times in sterile distilled water and then soaked in 1.0 M KNO₃ for 24 h (100 rpm at 25°C). The following day, seeds were rinsed three times in sterile distilled water and were transferred sterile petri dish with moist filter paper under the sterile fume hood. The seeds were observed until the seed coats started to crack allowing embryo germination. The seeds were plated onto ½ MS basal medium supplemented with 30g/liter sucrose and 100mg/L myo-inositol solidified with 3.2 g/L gelrite at pH 5.6 to 5.8. Culture bottles were incubated at ~22–25°C with a 16-h light/ 8-h dark photoperiod for seed germination. Hypocotyl tissues of the germinated seeds from 3-4 weeks were dissected into 1-2 mm sections and were inoculated onto somatic embryo induction medium consisting of ½ MS supplemented with 10 mg/L 2,4-D (CIM), 100 mg/L myo-inositol, 70g/L sucrose, solidified with 3.2 g/L gelrite at pH 5.6 to 5.8.

**Utilizing cell suspension-derived culture of papaya as alternative tissues for transformation**

The papaya proembryogenic mass of cells cultured in callus induction medium (CIM) after one month were used for initiation of suspension culture. The initial material is characterized by whitish friable calli. These calli were then transferred onto ½ MS liquid medium supplemented with 2 mg /L 2,4-D, 100mg/L myo-inositol and 70g/L sucrose at pH 5.6 to 5.8. The suspension cultures are regularly maintained on a 50-ml fresh medium every 3-4 weeks. Highly proliferating cells in the suspension cultures are maintained for Agrobacterium-mediated transformation.

Another set of cultures from suspension-derived cultures were subcultured to
regular CIM solid medium (with reduced 2,4-D – 2 mg/L) for continued somatic embryo proliferation and germination. The regeneration of this type of tissues was optimized where the embryogenic potential of suspension-derived cells for plantlet regeneration was conducted.

**Optimization of the modified Agrobacterium-mediated transformation protocol**

The different levels of cell densities for co-cultivation, the co-cultivation period, and the concentration of antibiotics were optimized for use in the transformation of suspension-derived cultures. The following parameters were used for three initial plate cultures to determine which condition would be optimal for the development of putative transgenic calli: co-cultivation period (1 or 2 days); concentration of *Agrobacterium* inoculum with optical density measured in spectrophotometer at OD$_{600}$ (0.1 or 0.005); and the different combination levels of antibiotics to eliminate *Agrobacterium* after co-cultivation (carbenicillin and/or cefotaxime) were determined.

**PCR to check for the presence of the transgene**

The presence of the papaya promoter, hygromycin resistance, and *eGFP* gene contained in the T-DNA of the pCambia 1302 vector were confirmed through PCR. The presence of the *Cp45* papaya promoter and *eGFP* were confirmed through PCR using the previously designed primer pairs (5’-ctgcagGGACAGTTTGAGTATTACAACAGAG-3’ and 5’-ccatggACAAAGTTAACAACACACAAATGGCTG-3’) (listed in Chapter III Table 3.1) and *eGFP*-specific primers (5’-CCATGGTGAGCAAGGGCGAG-3’ and 5’-
CTTAAGAAACTTTATTGCCAAATGTGGTAACG-3’) to amplify the ~2051 bp promoter regions and the 759 bp eGFP of the T-DNA insertion, respectively. The presence of the hptII gene for the hygromycin resistance used as a selectable marker in the T-DNA was amplified using primer pair 5’-GCGTCGGTTTCCACTATCGGCGAGTACTTC-3’ and 5’-CCGCGACGTCGACTGAGAAGTTTCTGATCG-3’. The PCR was carried out to amplify the target sequence using 100 ng genomic DNA with Bio-X-Act Short Reaction Mix (Bioline Inc., Taunton, MA, USA) adjusted to a final volume of 50µl. The binary vector construct used in the transformation protocol served as a positive control reaction. Eppendorf Master Cycler (Eppendorf, Hamburg, Germany) conditions used were as follows: denaturation at 95°C for 5 min, annealing at 55°C for 30 sec, elongation at 72°C for 1.5 min, for a total of 30 cycles, followed by a final elongation of 7 min at 72°C. The PCR products were separated via 0.8% agarose gel electrophoresis and stained with 1:10,000 GelRed™ (Biotium, Inc., Hayward, CA, USA).

**RNA Isolation and RT-PCR**

Total cellular RNA was isolated using the RNeasy Plant Mini kit and were treated with on column DNase treatment according to the manufacturer’s protocol (Qiagen Inc., Valencia, CA, USA). RNA was extracted from Cp45:eGFP transgenic calli. 1µg of total RNA was used to produce the first-strand cDNA with M-MLV reverse transcriptase and oligodT (Promega). The RT-PCR was carried out with Bio-X-Act Short Reaction Mix (Bioline Inc., Taunton, MA, USA) to amplify the target genes with the eGFP-specific primers (described above) using 100 ng cDNA template. The PCR products were
separated via agarose gel electrophoresis and stained with 1:10,000 GelRed™ (Biotium, Inc., Hayward, CA, USA). The optimized number of PCR cycles used for the quantification of eGFP expression was 30 cycles. The eGFP PCR band intensity was captured using the Kodak Multi-Imaging System and software (Kodak Co., Rochester, NY, USA).

**Protein isolation and immunoblot analysis**

Total cellular proteins were isolated from transgenic papaya calli following the protocol of Martinez-Garcia et al. (1999). The protein levels were quantified spectrophotometrically (Beckman and Coulter DU 730 UV/Vis), separated via standard SDS-PAGE (Laemmli 1970), were electro-transferred onto a Protran® nitrocellulose membranes (PerkinElmer Life Sciences, Boston, MA, USA) and probed with an anti-GFP rabbit IgG antibody (Molecular Probe-Invitrogen Corp., Carlsbad, CA, USA). Detection was carried out using an anti-rabbit secondary antibody from the Amersham ECL Western blotting analysis system (ECL™ kit, Amersham Biosciences, Piscataway, NJ, USA). The chemiluminescent signals were developed by exposure to Kodak X-ray film.

**Detection of fluorescence in the transformed papaya cultures**

Fluorescence was used to assess the papaya promoter-directed eGFP expression in the transgenic papaya cultures compared to the non-transformed (WT) plants. Cultures in plate were viewed under the Kodak Multi-Imaging System and software (Kodak Co., Rochester, NY, USA) to check for the eGFP fluorescence of the transgenic lines. The
plate cultures were laid on the Kodak imager platform and excitation filters were set at 395nm. The fluorescence of the cultures were also viewed using hand-held blue lamp.

Results

**Agrobacterium-mediated transformation using the carborundum-wounded papaya embryogenic calli**

Previously published procedures (Fitch et. al 1990) were used for producing papaya somatic embryo cultures from seedling hypocotyls of papaya cv. ‘Kapoho’. The embryogenic calli were obtained two to four weeks after induction in CIM (callus induction medium). Asynchronous somatic embryo formation was observed where different stages (e.g. globular, heart, torpedo) of developing somatic embryos were simultaneously observed. Highly proliferating somatic embryos are regularly subcultured onto CIM media every four weeks and maintained in the dark. The established somatic embryo cultures after 3-4 months were source tissue for *Agrobacterium*-mediated transformation using the carborundum-wounded treatment for papaya embryogenic calli (Zhu et al. 2006) (Figure 5.2).

We have tried the carborundum-treatment protocol to generate putative transgenic calli but we failed to obtain hygromycin-resistant calli from several plate cultures we transformed. The calli did not show any signs of growth after the treatment. The embryogenic potential and regenerability was lost given the series of infection and decontamination processes. Carborundum wounding affected the capacity of the tissues
to regrow even under the recovery medium (CIM without selectable antibiotic) where somatic embryos became chalky white, brown and unresponsive to the callus induction media and eventually died. Several cultures are also overgrown by the Agrobacterium which are not killed after co-cultivation. We then sought to look for alternative protocol to generate transgenic papaya using suspension-derived calli.

**Figure 5.2.** Agrobacterium-mediated transformation using carborundum-wounded calli (a) highly proliferating somatic embryos generated from hypocotyl sections using CIM media; (b) tissues were wounded with carborundum (600 mesh) by vortexing for 1 min; (c) washed calli with liquid induction media and were submerged in Agrobacterium culture for 10 min; (d) tissues were blotted dry on a tissue paper to remove excess Agrobacterium; (e) co-cultivation for 24 hours at dark condition (protocol adopted from Zhu et al. 2006).

**Suspension cultures of non-transformed papaya cultures capable of regeneration into plants**

The protocol described by Fitch et al. (1993) for production of proembryogenic callus using hypocotyl sections of germinated plants were utilized as starting material for use in the suspension culture establishment with a few modifications. The seeds were
placed in the moist filter paper until the seeds started to crack and were transferred to agar-water to germinate the seeds. The hypocotyls of the germinated plantlets were dissected (1-2 mm sections) and were plated onto CIM media to induce primary callus formation. The whitish friable clump of calli developed on the hypocotyl tissues after a month were utilized to start the suspension cultures (Figure 5.3). The liquid ½ MS media consisted of the basic CIM components but with reduced 2,4-D (2 mg/L) was used for suspension culture establishment.

To determine the capacity of suspension-derived culture to undergo normal somatic embryogenesis and to test whether these tissues have the ability to regenerate plants *in vitro* and in soil, we used the non-transformed cultures as a control set-up for the suspension-culture system (Figure 5.4). The papaya proembryogenic mass of cells cultured in CIM after one month were used for initiation of suspension cultures. The system utilized for somatic embryogenesis and plantlet regeneration system for these wild type control papaya plants was as described in the previous protocol (Zhu et. al 2006). Plantlets generated from *in vitro* culture of suspension culture were observed to have similar patterns of developmental stages as compared to the somatic embryos cultured under the normal CIM solid medium (Figure 5.4a). The plantlets generated *in vitro* acclimatized during transfer to soil and successful establishment of healthy and normal plants were generated from the suspension cultures (Figure 5.4b).

Since we have demonstrated that suspension culture produces good quality somatic embryos that regenerated into plants, we then optimized the alternative protocol of transformation using suspension-derived tissues as an alternative approach to transform papaya plant using *Agrobacterium*. 
Figure 5.3. Establishment of suspension culture: [1] mature fruit variety ‘Kapoho’ as the seed source; [2] washing the seeds with soap and water (3-4x); [3] seeds with removed mucilage are laid on a moist filter paper to induce seed germination; [4] germinated plant for hypocotyl excision (1mm disc) for CIM (callus induction medium) inoculation; [5] primary calli formation in CIM after 1-2 months; [6] suspension culture after 1 month; [7] proliferating suspension culture after 2 months.
Figure 5.4. (a) Optimization of growth parameters for *in vitro* culture and plantlet regeneration of papaya derived from suspension culture of the non-transformed wild type (WT) plants: [1] Primary calli development in CIM medium; [2] proliferation of somatic embryos under CIM with reduced 2,4-D (2 mg/L); [3] exposure of the somatic embryos at the lighted condition for cotyledonary leaf formation at ½ MS medium without 2,4-D; [4-7] growth and development of plantlet with complete root and shoots; (b) Plant acclimatization and establishment to soil: [1] plants with complete root and shoot transferred to soil covered with saran wrap plastic; [2] plantlet recovery and removal of plastic cover; [3] development of plants and subsequent transfer to bigger pots; [4-5] full-grown plants and successful establishment in the soil.
Agrobacterium-based transformation of papaya using suspension cultured cells as target tissue.

Different transformation parameters were optimized to select for the best condition to obtain putative transgenic calli from suspension-derived cultures (Table 5.1). The parameters for cell suspension culture co-cultivation period, Agrobacterium cell density, and decontamination using antibiotic combination procedure were determined. Overnight grown cultures of Agrobacterium tumefaciens (strain GV 3101) in LB liquid medium (with 100µg/ml of kanamycin) adjusted to cell density OD600 at 0.1 or 0.005 were used as transformation buffer for papaya. The virulence of the Agrobacterium tumefaciens was induced through incubation for 1-3 hours of 20µM acetylsyringone. 2-ml of Agrobacterium was added to the suspension-culture-derived culture growing on the CIM (with 2 mg/L 2,4-D) on top of filter paper (Figure 5.5 [3]). The excess Agrobacterium buffer was aspirated and was allowed to co-cultivate from 1-2 days. After the co-cultivation period, this is followed by 2-3x washing to remove the excess Agrobacterium using liquid MS medium with 250 mg/L cefotaxime. The papaya tissues laid on a filter paper were transferred onto CIM (2 mg/L 2,4-D) for somatic embryogenesis with different antibiotic treatments that serves as the recovery medium (Figure 5.5 [4]). After two weeks, the tissues are transferred onto selection medium with 50 μg/L hygromycin to select for putative transgenic papaya calli (Figure 5.5[5]. Putative transgenic papaya calli lines and regularly subcultured on the same selection media from 3-4 weeks interval for two months (Figure 5.5 [5]). Thereafter, somatic embryos are subcultured regularly without hygromycin to induce maturation in ½ MS₀ medium (CIM without 2,4-D, with 30g/L sucrose) that generated plantlets on MBN medium as described by Zhu et al.
Table 5.1. Optimization of parameters for modified Agrobacterium-mediated transformation protocol.

<table>
<thead>
<tr>
<th>Co-cultivation Period</th>
<th>Agrobacterium Cell Density (OD600)</th>
<th>Antibiotics Used</th>
<th>No. of putative transgenic calli / growth condition (post-co-cultivation with Agrobacterium)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Carb</td>
<td>Cef</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>250</td>
<td>0</td>
</tr>
</tbody>
</table>

Carb-carbenicillin; Cef-cefotaxime

(2006) (Figure 5.5 [7]).

From the parameters tested, more putative transgenic calli were obtained using combination of 250 mg/L carbenicillin plus 250 mg/L cefotaxime after co-cultivation for 1 day using Agrobacterium cell density of 0.005 (highlighted in gray). This system is therefore utilized in the subsequent transformation done in papaya with Vst1 which will be discussed in the succeeding Chapter (VI) that generated 75 transgenic lines from the four Cp promoter:Vst1 transformation constructs.

The putative transgenic calli transformed with Cp45-eGFP were viewed under the Kodak phosphorimager to detect eGFP fluorescence in proliferating somatic embryos (Figure 5.6). The two representative lines were subjected to PCR analysis to check for the presence of the eGFP transgene, papaya promoter sequence, and hygromycin resistance gene (hptII) (Figure 5.7 a). RT-PCR and Western Blot analyses further confirmed the presence of the mRNA transcripts produced in the cells and the protein expressed in the
transgenic tissue cultures, respectively (Figure 5.7 b&c). Plantlets in vitro showing eGFP expression in the different tissues such as stem and roots were visualized using hand-held blue light and Kodak phosphoimager which were not shown in non-transformed papaya plants (Figure 5.8 & 5.9)

**Discussion**

The potential use of suspension-derived cultures as material for *Agrobacterium* transformation was conducted using the pCAMBIA 1302-*Cp45-eGFP* construct. This promoter was known to express in *Arabidopsis* promoter assay (Chapter III) and was initially used to determine the transformation parameters for papaya using suspension cultures. This technique was optimized as an alternative complementary approach for transformation of papaya using *Agrobacterium*.

From the different parameters tested, we found that 1 day co-cultivation with OD$_{600}$ at 0.005 of *Agrobacterium*, and a combination of 250 mg/L cefotaxime and carbenicillin were the optimized condition to obtain significant number of putative calli lines. It was observed that *Agrobacterium* easily persists in the cultures when the co-cultivation is longer. One day co-cultivation regime is enough to allow the *Agrobacterium* infection and T-DNA transfer but maintained the cell density that would be easier to eliminate post-transformation. Most of the cultures can be easily overgrown when excessive initial *Agrobacterium* is added to transform the tissue. The reduction of the *Agrobacterium* cell density to 0.005 allowed us to control further growth of the *Agrobacterium* with the combination of two antibiotics which are optimal for the growth of the plants. In the previous transformation work in papaya, one of the major challenging
Figure 5.5. Development of modified Agrobacterium-mediated transformation protocol from suspension-derived culture: [1] highly proliferating suspension culture [2] 2 ml of suspension culture transferred onto CIM (with reduced 2,4-D at 2 mg/L); [3] proliferating calli for 24 hour inoculation of Agrobacterium tumefaciens containing the construct; [4] recovery medium with antibiotics eliminate Agrobacterium in the tissues [5] growth of the putative transgenic calli under recovery medium with 50 mg/l hygromycin; [6] somatic embryo maturation under CIM with reduced 2,4-D (2 mg/L) concentration; [7] somatic embryo germination exposed to lighted condition for cotyledonary leaf formation; [8] plantlet regeneration under ½ MS without 2,4-D at the lighted condition.
Figure 5.6. *Cp45:eGFP* putative transgenic papaya cultures generated from the modified *Agrobacterium*-mediated transformation protocol (a-d) with verified eGFP expression viewed under Kodak phosphorimager (e).

**Figure 5.7** Molecular analyses confirming the presence of the: [a] transgene *eGFP*, *Cp45* promoter and hygromycin resistance in the T-DNA vector construct; [b] mRNA transcript analysis using RT-PCR; [c] Western blot for *eGFP* protein expression of the *Cp45:eGFP* transgenic calli.
Figure 5.8 *Cp45:eGFP* transgenic papaya cultures generated from the modified *Agrobacterium*-mediated transformation protocol with verified eGFP expression showing fluorescing stems viewed under (a) white light (a); blue light (b); Kodak phosphorimager (c).
Figure 5.9 *Cp45:eGFP* transgenic papaya cultures generated from the modified *Agrobacterium*-mediated transformation protocol with verified *eGFP* expression showing fluorescing roots (a) and stem (b) viewed under hand held blue light and Kodak phosphorimager.

aspects of transformation was to control latent *Agrobacterium* growth especially when super virulent strain is used (Fitch et al. 1993). The ability of the antibiotic in killing *Agrobacterium* relies on how cells or tissues can access antibiotic treatment which is affected by amount of tissues being treated. The thin layer of suspension-derived cells allowed the ample amount of papaya tissues for *Agrobacterium* treatment as well as allowed for more contact with the antibiotics in the media. These simple parameters are
very important in obtaining successful transformants *via Agrobacterium* co-cultivation. This system also eliminates the step of carborundum-wounding treatment that decreased the stress on the tissues that aided the cells to recover from *Agrobacterium* infection.

In our observation, it is important to identify the most suitable stage of cells or tissues that would be used prior to *Agrobacterium* infection. The selection of the most appropriate tissue will highly affect the efficiency of transformation. The quality of cultures used from suspension-derived culture is characterized by yellowish/creamy colored young globular somatic embryos. In this case, tissue stages at around 2-3 month under liquid suspension cultures are appropriate for *Agrobacterium* infection. A suspension culture, which is on the earlier stage characterized by a whitish color of the cells, was observed to be an unsuitable stage for transformation. The use of suspension-derived culture maintained at the young globular stage gave mostly uniform types of tissues for transformation. The ability to obtain fully transgenic cells is possible in this system.

The putative transgenic hygromycin-resistant calli were all verified positive through PCR, RT-PCR and Western blot analyses. In addition, fluorescence signals of *eGFP* detected under hand-held blue light and Kodak phosphorimager further confirmed the activity of *Cp45* driving the *eGFP* expression in the transgenic plantlets.

The optimized transformation protocol from this system can be used in future genomic transformation of papaya. The ability to obtain fully transgenic plants without the use of the sophisticated equipment such as gene gun facilitates transformation of papaya in a simple laboratory setup.

122
Conclusion

The suspension-derived calli were proven to be suitable tissues for *Agrobacterium* transformation in papaya. The tissues from this system which are maintained at globular young calli stage characterized by yellowish/creamy embryos are identified to be highly receptive for *Agrobacterium* infection. The optimum parameters selected to obtain putative transgenic calli were the following: 1 day co-cultivation of *Agrobacterium*; cell density used for transformation was 0.005 at OD$_{600}$; and *Agrobacterium* post-transformation was eliminated using combination of 250 mg/L carbenicillin and cefotaxime. Putative transgenic calli were all confirmed positive for transgene integration using PCR. *eGFP* mRNA transcripts were detected using RT-PCR and protein expression were determined using Western blot analysis. *eGFP* fluorescence under hand-held blue light and Kodak phosphorimager further confirmed the expression of *Cp4:eGFP* in the transformed plantlets. The utility of this technique facilitates transformation of papaya using simple laboratory setup. This system is therefore utilized in the subsequent transformation done in papaya with four *Carica papaya* promoter:*Vst1* constructs which will be discussed in the succeeding Chapter (VI) that generated 75 independent transgenic papaya lines.
CHAPTER VI

DEVELOPMENT OF TRANSGENIC PAPAYA PLANTS WITH STILBENE SYNTHASE CAPABLE OF SYNTHESIZING PHYTOALEXIN, RESVERATROL

Abstract

Papaya is one of the most important fruit crops in the tropics that have limited resistance to the devastating pathogen, Phytophthora palmivora. Previously, the expression of the grapevine stilbene synthase gene in papaya enhanced resistance to Phytophthora palmivora (Zhu et al. 2004) but the expression of Vst1 gene driven by its own native promoter is less than optimal in papaya. This led to the identification and characterization of four new pathogen-induced native papaya promoters expressed in the roots and stems to modulate Vst1 expression for resveratrol synthesis. The promoters previously isolated from genes encoding peroxidase (Cp9), beta-1,3-glucanase (Cp29) ferulate-5-hydroxylase (Cp35), and hypersensitive-induced response protein (Cp45) were fused to stilbene Vst1 gene. The constitutive promoter CaMV35S-Vst1 construct served as a positive control. PCR, RT-PCR, qRT-PCR and Southern Blot analyses confirmed the transgene integration and Vst1 mRNA transcripts expression in the transgenic papaya calli lines. Normal patterns of somatic embryogenesis were observed from transgenic papaya lines developed. The synthesis of a new phytoalexin, resveratrol glucoside (piceid) was detected through RP-HPLC analysis and further verified through tandem mass spectrometry (MS/MS). The concentration of piceid differed as follows: CaMV35S (25.5 µg g⁻¹ f.wt), Cp45 (11.90 µg g⁻¹ f.wt), Cp35 (2.76 µg g⁻¹ f.wt), Cp9 (2.46 µg g⁻¹ f.wt). Promising basal piceid levels detected in transgenic papaya using the native promoters can be used as materials for disease induction experiments in the future. The media
requirements for somatic embryo maturation, germination and plantlet regeneration for transgenic calli are currently being optimized to obtain normal plants. Transgenic papaya plants will be regenerated from these transgenic calli provides future source of lines that has improved nutritional content with potential disease resistance against *Phytophthora palmivora* and other papaya diseases.

**Introduction**

Papaya (*Carica papaya, L.*) is a widely cultivated and important tropical fruit commodity in the global market as well as source of nutrition for subsistence farmers (FAO, 2005). In Hawaii, it is the second largest fruit industry with $11 million (USD) revenue in 2011. The deregulation of PRV resistant of solo variety ‘Rainbow’ papaya for shipment to Japanese market has increased the demand for local and export market. Despite the fact that Hawaii can grow PRV resistant papaya, a devastating disease *Phythophthora palmivora* greatly limits papaya production. This oomycete pathogen seriously threatens papaya production by inhibiting transport of water from the roots, ultimately, killing the plants, which reduces productivity and fruit quality (Nishijima 1994). It was known that papaya has limited natural resistance to a range of oomycete pathogens (Nishijima a 1994). The disease is difficult to manage and breeding for resistant papaya against *Phytophthora* is limited by the absence of resistance genes in the *Carica* family. Attempts to hybridize wild relative *Vasconcellea quercifolia* to incorporate resistance genes into cultivated *Carica papaya* had also been unsuccessful due to incompatibility and production of infertile hybrids (OECD, 2005). The only option used by farmers to
control *P. palmivora* and other diseases is to maintain farm management practices that include fungicide application which leads to increased production costs and contamination of the environment. An alternative to fungicides such as development of genetically modified papaya capable of producing phytoalexins that can resist *Phytophthora palmivora* infection would significantly alleviate the problems in papaya production.

Phytoalexins are important natural components in the defense of plants against fungal infection (Jeandet et al. 2002, Kuc 1995). Resveratrol has potent antimicrobial and antifungal activity (Hammerschmidt 2004). Synthesis of the stilbene resveratrol (*trans*-3,4,5-trihydroxystilbene) was induced when grapevine and peanut were attacked by pathogens (Langcake and Pryce 1977). In grapevine, stilbene synthesis is induced by inoculation with the pathogens, *Botrytis cinerea* or *Plasmopara viticola* (Blaich & Bachmann 1980, Langcake 1981). Dercks and Creasy (1989) showed that the subsequent level of resistance to *P. viticola* was positively correlated with the capacity of grapevine to synthesize stilbene. Recently, it was also shown that stilbene synthase and chitinase mRNAs were induced in the tolerant cultivars of Florida hybrid bunch grape when infected with *Elsinoe ampelina* (Vasanthaiah et al. 2010).

Resveratrol is synthesized by the enzyme trihydroxystilbene synthase (stilbene synthase or resveratrol synthase) (EC 2.3.1.95) using as substrates one molecule of 4-coumaroyl-CoA and three molecules of malonyl-CoA (Rupprich & Kindl 1978). These two substrates are commonly present in plants. Therefore, introduction of a single gene encoding stilbene synthase may be sufficient to synthesize resveratrol in heterologous plant species for increased resistance against fungal pathogens. This process is a form of
metabolic engineering in which the biosynthetic capacity of the plants is altered by diverting substrates for other purposes. *Stilbene synthase* genes isolated from grapevine have been transformed into various crop species that results in increased resistance to different pathogens as well as increased its nutritional content (Appendix Table 1).

Previously, the expression of the grapevine *stilbene synthase* gene in papaya enhanced resistance to *Phytophthora palmivora* (Zhu et al. 2004) but the expression of grapevine *Vst1* gene driven by its own native promoter is less than optimal in papaya. The *stilbene synthase* mRNA was not detectable immediately after inoculation that indicates little or no constitutive expression. This led to the study of identification and characterization of isolated native papaya promoters from pathogen inducible genes of papaya to regulate expression of stilbene synthase for use in metabolic engineering of papaya for resveratrol production. To date, most of the transformation done in papaya used the traditional promoter, cauliflower mosaic virus (*CaMV35S*) (Odell et al. 1985). However, the use of this promoter is viewed negatively by consumers and regulatory agencies because of the arbitrary increase of the transgene on the consumed portion of the crop, leading to higher costs for biosafety trials and poor sales. Moreover, other periodic problems include transgene silencing (Flavell 1994, De Wilde et al. 2000, Halpin et al. 2001) and developmental abnormalities, such as male sterility (Bornke et al. 2002, Hain et al. 1993).

Four putative promoter regions and their 5’UTRs were isolated from the genes encoding peroxidase (*Cp9*), β-1,3-glucanase (*Cp29*), ferulate-5-hydroxylase (*Cp35*) and hypersensitive-induced response protein (*Cp45*) and were fused to stilbene synthase gene (*Vst1*) from grapevine. We transformed papaya with stilbene synthase using our own
promoters to assess the regulatory activity of these promoters and evaluate the effects of newly synthesized compound resveratrol on the overall morphological growth and development of transgenic papaya. We aim to develop transgenic papaya plants with controlled stilbene synthase expression thereby regulating resveratrol synthesis without altering the normal phenotypic growth of the plants. We hope to provide transgenic papaya with desired levels of resveratrol that can be used for future disease resistance induction experiments for papaya crop improvement.

Materials and Methods

Vector construction of pCAMBIA 1302-papaya promoter-\textit{Vst1} 

The vector pCAMBIA 1302 containing the native papaya promoter-\textit{eGFP} constructs (Carlos-Hilario et al. 2014) previously used for \textit{Arabidopsis} transformation was used as vector backbone. The stilbene synthase gene (\textit{Vst1}) gene was amplified using the primer pair (5’-GAGTTCGTAGCTAGGCCATGGCTTC-3’ and 5’-CACGTGTCAATTGTAAACCATAGGAACG-3’, the sequence underlined represents added \textit{NcoI} and \textit{PmII} restriction sites respectively. The stilbene synthase 1535 bp PCR product was purified from agarose gel using QIAquick Gel Extraction Kit (Qiagen Inc., Valencia, CA, USA) and ligated into the TOPO-TA vector (Invitrogen, Carlsbad, CA, USA). The \textit{Vst1} gene was released by double digestion using the restriction enzymes \textit{NcoI} and \textit{PmII} and correct sequence was verified through DNA sequencing. \textit{Vst1} was cloned between the \textit{NcoI} and \textit{PmII} restriction sites of the pCAMBIA1302 vector replacing the \textit{eGFP} gene with the \textit{Vst1} gene to generate the construct pCAMBIA1302-papaya promoter-\textit{Vst1} (see Chapter IV, Figure 4.1). The \textit{CaMV35S:Vst1} was used as
positive expression control. The orientation and reading frame of all recombinant plasmids were verified by PCR and DNA sequencing. Recombinant plasmids were mobilized onto *Agrobacterium tumefaciens* strain GV3101 competent cells through the modified freeze-thaw method (Jyothishwaran et al. 2007).

**Transformation of papaya with *Carica papaya* promoter: *Vst1* fusions**

Papaya suspension culture cv. ‘Kapoho’ was used for papaya transformation. Papaya was transformed by the modified *Agrobacterium*-mediated transformation protocol developed as described above (Chapter V, see Materials and Methods). Putative transgenic calli were selected *in vitro* on solidified CIM medium with 50µg/ml hygromycin B (Sigma-Aldrich, St. Louis, MO, USA). The presence of the papaya promoter and *Vst1* were confirmed through PCR using the previously designed primers (Table 5.1). The PCR was carried out to amplify the target sequence using 100 ng genomic DNA with Bio-X-Act Short Reaction Mix (Bioline Inc., Taunton, MA, USA) adjusted to a final volume of 50µl. The binary vector construct used in the transformation protocol served as a positive control reaction. Eppendorf Master Cycler (Eppendorf, Hamburg, Germany) conditions used were as follows: denaturation at 95°C for 5 min, annealing at 55°C for 30 sec, elongation at 72°C for 1.5 min, for a total of 30 cycles, followed by a final elongation of 7 min at 72°C. The PCR products were separated via 0.8% agarose gel electrophoresis and stained with 1:10,000 GelRed™ (Biotium, Inc., Hayward, CA, USA).
Table 6.1. Primers used in this study.

<table>
<thead>
<tr>
<th>Amplification</th>
<th>Primer</th>
<th>Sequence</th>
<th>Amplicon Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>35SF</td>
<td>5’-ATGGAGTCAAGAGATTCAAATAGGGACCTAACAG-3’</td>
<td>537</td>
</tr>
<tr>
<td>PCR</td>
<td>35SR</td>
<td>5’-GTCCCCGGTGTTCTTCCTCAAAGTTGAAATG-3’</td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>Cpf9F</td>
<td>5’-ctgcagTCTTGAATCATGTGAGAATGTAACC-3’</td>
<td>2041</td>
</tr>
<tr>
<td>PCR</td>
<td>Cpf9R</td>
<td>5’-ctgcagGGCCATGACTTCACTGCTTCCGGTAG-3’</td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>Cp5F</td>
<td>5’-ctgcagGCTTGAATCAGATGAGAAACCC-3’</td>
<td>1445</td>
</tr>
<tr>
<td>PCR</td>
<td>Cp5R</td>
<td>5’-ctgcagGGTTACTGGTTGGTGGTGAATATG-3’</td>
<td>1832</td>
</tr>
<tr>
<td>PCR</td>
<td>Cp3F</td>
<td>5’-ctgcagGGACAGTTTAGTACAAACAGAAG-3’</td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>Cp3R</td>
<td>5’-ctgcagGGTTACTGGTTGGTGGTGAATATG-3’</td>
<td>2051</td>
</tr>
<tr>
<td>PCR</td>
<td>Vst1F</td>
<td>5’-GAGGAAATTAGAAACGCTCAACGGAAGGGTGC-3’</td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>Vst1R</td>
<td>5’-GAGGAAATTAGAAACGCTCAACGGAAGGGTGC-3’</td>
<td></td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>ActinF</td>
<td>5’-TCCCGGAGGAGTTTCTCCTAGTATGT-3’</td>
<td>1535</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>ActinR</td>
<td>5’-CCCTCTTATGTTCTTCATCAACCA-3’</td>
<td>100</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Vst1F</td>
<td>5’-GAGGAAATTAGAAACGCTCAACGGAAGGGTGC-3’</td>
<td></td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Vst1R</td>
<td>5’-CTTCAACTCAGTACGGTGTCGTCGGTTTGAATG-3’</td>
<td>159</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>CHSF</td>
<td>5’-ATGGCGGGTGGCGGTATTGTAACCGTG-3’</td>
<td>100</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>CHSR</td>
<td>5’-CAGTGGTGGGAAGGGTGGCGGT-3’</td>
<td></td>
</tr>
</tbody>
</table>

Sequences underlined and in lower case were introduced PstI (F) and NcoI (R) sites for Cp9 and Cp29, NcoI (F) and PmlI site (R) for Vst1.

**Southern blot analysis**

Genomic DNA was extracted from putative transgenic papaya calli (Doyle and Doyle 1990). Forty micrograms of DNA was digested with EcoRI-HI-HF (New Englands Biolabs, Inc.) overnight and size fractionated by 0.8% agarose gel electrophoresis (Sambrook et al. 1989). The gel was denatured with 0.5 M NaOH/1.0 M NaCl (30 min) and neutralized with 0.5 M Tris-HCl/1.5 M NaCl (30 min) with gentle shaking, then transferred overnight by capillary action to a nitrocellulose N+ membrane (GE Healthcare Biosciences Corp., Piscataway, NJ, USA) using 10X SSC buffer. The DNA was covalently crosslinked to the nitrocellulose N+ membrane by ultraviolet light (Stratalinker, Stratagene, La Jolla, CA). The hybridization probe was prepared by amplifying the Vst1 (1535 bp) from the pCAMBIA 1302 binary vector through PCR.
using the \textit{Vst1} primers described above. The PCR products were separated at 0.8% agarose gel (Bioline Inc., Taunton, MA, USA) and gel-purified with QIAquick Gel Extraction Kit (Qiagen Inc., Valencia, CA, USA). The fragment was labeled with biotin and Southern hybridizations were carried out using North2South Complete Biotin Random Prime Labeling and Detection kit (Thermo Scientific, Rockford, IL, USA). Several independent transgenic plants were developed containing variable copy number of the transgene from each of the \textit{Carica papaya} promoter that were used for the subsequent analysis.

\textbf{RNA isolation and RT-PCR}

Total cellular RNA was isolated using the RNeasy Plant Mini kit and were treated with on column DNase treatment according to the manufacturer’s protocol (Qiagen Inc., Valencia, CA, USA). RNA was extracted from \textit{Carica papaya} transgenic calli. 1\textmu g of total RNA was used to produce the first-strand cDNA with M-MLV reverse transcriptase and oligodT (Promega). The RT-PCR was carried out with Bio-X-Act Short Reaction Mix (Bioline Inc., Taunton, MA, USA) to amplify the \textit{Vst1} genes with the \textit{Vst1}-specific primers (described above) using 100 ng cDNA template. The PCR products were separated via agarose gel electrophoresis and stained with 1:10,000 GelRed\textsuperscript{TM} (Biotium, Inc., Hayward, CA, USA). The optimized number of PCR cycles used for the quantification of \textit{Vst1} expression was 30 cycles. The \textit{Vst1} PCR band intensity for each of the promoter was visualized using the Kodak Multi-Imaging System and software (Kodak Co., Rochester, NY, USA).
Real-Time PCR analysis

Previously isolated total cellular RNA described above was utilized to convert RNA to cDNA using M-MLV reverse transcriptase and oligodT (Promega, Madison, Wisconsin, USA). *Vst* primers 5’-GAGGAAATTAGAAACGCTCAACGTGCCAAGGGTC-3’ and 5’-CTTCAACTCTCAGTCATGTGCTCGCTCTTTAGT-3’ (Table 6.1) were designed to amplify a 159 bp product of the *Vst* sequence. PCR primers 5’-ATGGCGGTGGACGGTG-3’ and 5’-CAC AGT TGG AGGGAGTGGCGGT-3’ were used to amplify 100 bp product of *CHS*. *Carica papaya* actin cDNA was used as the internal standard with primers 5’-TCCCAGGGCAGTTTTCCCTAGTATTGT-3’ and 5’-CCCTCTTTAGATTGTGCTTTCATCACC-3’. 1µL (50 ng) of cDNA product was combined with forward and reverse qPCR primers to a final concentration of 500nM. This reaction was then combined with an equal volume of 2x Roche Power SYBR Green Master Mix with ROX (Applied Biosystems, Carlsbad, CA) mixed and aliquoted in 10µL triplicates into a 384-well optical qPCR plate then sealed with an Applied Biosystems optical adhesive film. Samples were run on a 7900HT qPCR machine with the following thermocycling parameters: 2 minutes at 50°C, 10 minutes at 95°C, then the following conditions were repeated for 40 cycles: 15 seconds at 95°C, 1 minute at 60°C. A dissociation step was performed at the end of the run to ensure only a single product was formed utilizing a +1°C ramp rate per second from 60°C to 95°C.
**HPLC and tandem mass spectrophotometry (MS/MS) analysis**

Proliferating calli of transgenic papaya were used for HPLC analysis along with the control wild type non-transformed calli. Metabolites were extracted with 80% methanol. The samples were ground, vortexed (vigorously for 1-2 minutes), and then incubated overnight with shaking at 100 rpm. Samples were centrifuged to remove insoluble debris. The supernatant was concentrated to dryness and were dissolved in 100µl of 100% methanol added with 900µl of 2% acetic acid. The samples were centrifuged twice at 10,000 g for 10 minutes where 500 µl aliquot is finally stored in -20°C until ready to use for HPLC analysis.

The standard piceid (Sigma-Aldrich, St. Louis, MO, USA) and samples of tissue extracts containing piceid were resolved by Reverse-Phase High Performance Liquid Chromatography (RP-HPLC), which consisted of a Waters 2695 automated HPLC system interfaced with a Waters 996 Photodiode Array Detection (PDA), scanning 210-600 nm. Peaks were resolved using a Kinetex C18 column (Phenomenex; 2.6µm; 100A; 100 x 4.6 mm) at a flow rate of 0.8mL/min. and eluted with a linear 8%/min. gradient of Solvent A (2% v/v acetic acid/Aq.) against Solvent B (2% v/v acetic acid/MeOH), with a 3 min. re-equilibration period between sample injections. Resulting chromatographic profiles were extracted at 300nm, and the target peak area integrated using Waters Empower Pro software. Standards of piceid where undertaken in triplicate, allowing for the precise establishment of piceid concentrations reported as µg/fresh tissue weight.

Mass spectrometry experiments were undertaken using AB/MDS-Sciex API 3000 triple quadrupole mass spectrometer (Thornhill, Ontario, Canada). The ESI-MS system was calibrated manually in positive mode with PPG 3000 (AB/MDS-Sciex) to achieve
<5-ppm mass accuracy, as per manufacturer’s protocol. The mass spectrometry was interfaced with a PE ABI 140D Solvent Delivery System. Ultra-high purify N₂ was delivered by a Peak Scientific generator. RP-HPLC purified analyte fraction was injected as a 20uL bolus, at a flow rate 25uL/min, in a solvent of 70% MecN/30% 0.1% v/v formic acid/H₂O. Initially full ion spectra were taken from m/z 100-600 Da to provide identification of the parent molecular mass using Q-3.

For tandem mass spectrometry (referred to as MS/MS) experiments, collision-induced dissociation (CID) of the identified parent molecular mass was effected by bombardment with nitrogen. Bombardment was confined to quadrupole-2 (Q-2) with a collision cell gas thickness of 3 x10^{14} atoms/cm² and a collision energy (Q-0 to Q-2 rod offset voltage) typically set at ~20–40 eV. The resulting CID (daughter ion) spectra were obtained by scanning quadrupole-3 (Q-3) from m/z 100–400 Da. MS/MS data analysis was assisted with the use of Analyst Software (v1.4.1; AB/MDS-Sciex) and ACD ChemSketch (ACD/Labs).

Results

Transgene integration, RT-PCR analysis, and somatic embryo development of the transgenic papaya cultures

Putative transgenic initial calli surviving in hygromycin (50 mg/L) were formed one month after Agrobacterium transformation (Figure 6.1a). The initial calli were subcultured at a regular basis at CIM media with reduced 2.4-D (4mg/L) where highly proliferating somatic embryos maintained under hygromycin selection has shown normal
Figure 6.1. Hygromycin-resistant putative transgenic calli growing one month after *Agrobacterium* transformation (a); Highly proliferating somatic embryos under hygromycin (50 mg/L) selection showing normal growth and development relative to the non-transformed wild-type (WT) calli (b).
somatic embryo development relative to the non-transformed wild-type (WT) calli (Figure 6.1b).

Several putative transgenic papaya calli transformed with the different \( Cp \) promoter: \( Vst1 \) constructs were screened for PCR analysis to check for the presence of \( Vst1 \) transgene and hygromycin resistance gene of the inserted T-DNA region from the transformation pCAMBIA 1302 vector (Figure 6.2a). All of the independent transgenic lines of each of the papaya promoter: \( Vst1 \) lines were shown positive for PCR analysis. However, we failed to develop \( Cp29 \) transgenic papaya lines and we suspect that \( Cp29 \)

![Figure 6.2](image.png)

**Figure 6.2.** Molecular analyses of the transgenic papaya calli: PCR check for the presence of the \( Vst1 \) gene and hygromycin resistance gene \( (hptII) \) (a); RT-PCR analysis for the detection on the differences of \( Vst1 \) transcript expression as driven by the different native papaya promoters.
elevated activity as shown in promoter analysis in transgenic *Arabidopsis* lines transformed with *Vst1* (Chapter IV) might be toxic to the initial callus development in papaya.

Transgenic lines developed were evaluated on the *Vst1* mRNA transcript expression as driven by the different papaya promoters (Figure 6.2b). The transcript level of *Vst1* was highest in three *CaMV35S* lines, which was comparable to the banding intensity shown in six *Cp45* lines. Most of the transgenic lines in *Cp9* and *Cp35* lines were shown to have lower *Vst1* transcript expression. The banding intensity patterns in the RT-PCR analysis had similar strength and activity in the promoter assessment conducted using the *eGFP* reporter assay in *Arabidopsis thaliana* at the reproductive stage as discussed in the previous (Chapter III). The presence of *Vst1* transcripts indicates that mRNA is produced in the transgenic papaya, which is expressed at varying levels as driven by the native papaya promoters. *Vst1* could function as an active enzyme for resveratrol synthesis, a branched pathway in the flavonoid biosynthesis.

Southern blot analysis further verified transgene integration of *Vst1* in papaya genome harboring the *CaMV35S:*Vst1, *Cp9:*Vst, *Cp35:*Vst1 and *Cp45:*Vst1 constructs (Figure 6.3). However, Southern blot for the other papaya transgenic lines did not show genomic integration of the *Vst1* although all of the lines were PCR and RT-PCR-positive lines. There could be potentially proteins or carbohydrates which might be present in the calli tissues that interfered with the genomic DNA restriction enzyme digestion that resulted to poorly resolved fragments in the agarose gels and poor nitrocellulose membrane transfer. It would be worth to wait for the plants to develop green leaf tissues to obtain ample amount and good quality genomic DNA for Southern blot analysis.
Figure 6.3. Southern blot analysis of EcoRI-digested genomic DNA extracted from Carica papaya transgenic calli lines hybridized with 1.5 kb biotinylated stilbene synthase probe. The four promoters CaMV35S, Cp9, Cp35 and Cp45 fused to the Vst1 gene are indicated above the panel. Lanes: WT-Wild Type DNA from non-transformed control plant, P-plasmid DNA used as positive control. Number above the blot indicates numbered transgenic lines. Transgene copy number denoted below the lanes in the blot.

Comparison of Vst1 and CHS transcript levels

Carica papaya is known to contain significant levels of flavonoids (Canini et al. 2007) and Vst1 can compete directly CHS with for the metabolic intermediates p-coumaroyl CoA and malonyl CoA. RT-PCR analysis detected Vst1 transcripts in the different transgenic calli but the difference in the banding intensity pattern does not effectively quantitate the level of expression based on this qualitative data alone. Therefore, Quantitative Real-Time PCR (qPCR) analysis was performed to directly
assess and compare the relative transcript levels of \textit{Vst1} and \textit{CHS} in transgenic papaya calli (Figure 6.4). Interestingly, our data showed that the highest level of \textit{Vst1} transcripts are found in \textit{Cp45}, followed by the constitutive \textit{CaMV35S}, the weaker promoters \textit{Cp35}, and the weakest \textit{Cp9} promoter. The differences in \textit{Vst1} transcript expression profiles are indicative of the strength and activity of the promoters in the transgenic papaya. Previously, \textit{Cp45} was found to be highly expressed in \textit{Carica papaya} roots and is pathogen-inducible (Porter et al 2009). \textit{Cp45} was also shown to be highly expressed in the previous \textit{eGFP} assay conducted in the previous discussion (Chapter III). It would be very promising to further assess the activity of \textit{Cp45} promoter when plants will be grown from these transgenic calli. As expected, no \textit{Vst1} transcripts were detected in the wild type control plants.

Among the several transgenic lines develop, only \textit{Cp9} had lower levels of \textit{CHS} which could be due to down regulation of \textit{CHS} due to \textit{Vst1} activity being favored over the \textit{CHS} (Figure 2b). In contrast, \textit{CHS} in \textit{CaMV35S} and \textit{Cp45} were slightly increased in transgenic calli lines. Another interesting observation is the \textit{Cp35} elevated activity that is threefold greater than the WT. This is a very unique and interesting expression pattern of \textit{CHS} that implies \textit{Cp35} promoter regulatory elements is weak enough to induce \textit{Vst1} activity favoring increased expression of \textit{CHS} over \textit{Vst1}. This observation was inconsistent with the result we obtained in \textit{Arabidopsis} where \textit{Vst1} expression caused \textit{CHS} down regulation in the transgenic plants (Chapter IV discussion). The dynamics of \textit{CHS} versus \textit{Vst1} expression was variable in the four different promoters in transgenic papaya calli. This could imply that the nature and signals contain in the promoter regions regulates the \textit{Vst1} ectopic expression affecting \textit{CHS} activity and might be variable from a
species to another more complex species. We could not detect down-regulation activity in *Cp45* and *CaMV35S* promoter as a result of *Vst1* activity except for *Cp9*. Inconsistencies on the *Vst1/CHS* dynamic expression in the transgenic papaya are potentially due to the type and stage of transgenic tissue we analyzed, or potentially environmental trigger effect as being detected by the promoters e.g response to cues such as light, temperature etc. The variable regulatory activity of the promoters resulted to differences in the balance of expressed transcript levels of *Vst1* versus *CHS*. It is definite to say that *Vst1* has activity in all the transgenic calli at variable levels depending on the promoter used but we cannot generalize if there is direct or indirect correlation to the *CHS* activity. It would be early to conclude if this pattern of expression was due to developmental regulation or to any environmental condition that led to the differential *Vst1/CHS* activity in the young transgenic papaya calli tissues. It would be interesting to see if the fully grown plantlets developed from these transgenic calli would correspond to the same transcript expression which we hope to assess once plantlets are obtained.

**Biochemical characterization of transgenic *Carica papaya* expressing *Vst1***

The individual transgenic calli lines were analyzed by HPLC for the accumulation of resveratrol (Figure 6.5). As expected, quantification of piceid accumulation concentration on the transgenic papaya calli was found highest in *CaMV35S* promoter with the average of 25.50 µg/ gram fresh weight from the transgenic lines analyzed. The samples from *Cp45* yielded 11.90 µg/ gram fresh weight piceid which is a very interesting basal expression in papaya. *Cp45* piceid more than 50% lower in piceid concentration as compared to the constitutive *CaMV35S* promoter which indicates a
Figure 6.4. Quantitative comparison of (a) \textit{Vst1} and (b) \textit{CHS transcript} levels driven by the different promoters in the transgenic papaya calli. The levels of mRNAs were measured using real-time qPCR, normalized to the internal control gene \textit{Ubiquitin-Conjugating Enzyme 21} (AT5G25760) (\textit{PEROXIN4}). The calibrator used for measuring the gene expression change (\(2^{-\Delta\Delta CT}\)) was the nontransformed \textit{Arabidopsis} wild type (WT) (Czechowski et al, 2005).
modulated activity of 

modulated activity of $Cp45$ in papaya transgenic calli. The weak promoters $Cp35$ and $Cp9$ had 2.76 and 2.49 µg/gram fresh weight respectively. $CaMV35S$, $Cp9$, $Cp29$ and $Cp45$ transgenic calli had variable piceid concentration as represented by the HPLC peaks detected from the different promoter activity (Figures 6.6 to 6.10). The UV spectra of piceid standard (red) and the transgenic lines from each of the promoters matched the peak at 7.38 minute retention time at $A_{300}$ nm absorbance. The transgenic calli with variable piceid accumulation was shown to have normal somatic embryo development relative to the WT somatic embryos. It would very promising if the lines with strong basal expression in the calli stage would produce the similar significant
Figure 6.6. Wild-type (WT) non transformed papaya: HPLC analysis (a); normal growth and development of the proliferating somatic embryos (b).
Figure 6.7. Representative HPLC chromatogram peak of the transgenic CaMV35S:Vst1 papaya calli (a); growth and development of the proliferating transgenic somatic embryos in CaMV35S line 1, 2 and 3 (b).
Figure 6.8. Representative HPLC chromatogram peak of the transgenic \textit{Cp9}:\textit{Vst1} papaya calli (a); growth and development of the regenerating transgenic somatic embryos showing root and shoot development in \textit{Cp9} line 1 (b).
Figure 6.9. Representative HPLC chromatogram peak of the transgenic *Cp35:*Vst1 papaya calli (a); growth and development of the proliferating transgenic somatic embryos in *Cp35* line 1, 4, 5 and 8 (b).
Figure 6.10. Representative HPLC chromatogram peak of the transgenic *Cp45:Vst1* papaya calli (a); growth and development of the proliferating transgenic somatic embryos in *Cp45* line 1, 2, 3 and 4 (b).
Table 6.2. Piceid concentration of transgenic papaya calli.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Transgenic Line Number</th>
<th>Piceid (µg/gram fresh weight)</th>
<th>Calli Stage</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>CaMV35S</td>
<td>1</td>
<td>15.43 ± 0.488</td>
<td></td>
<td>25.50 ± 1.907</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>37.46 ± 5.228</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>23.62 ± 0.930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cp9</td>
<td>1</td>
<td>3.22 ± 0.107</td>
<td></td>
<td>2.49 ± 0.174</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.38 ± 0.656</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.72 ± 0.097</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>8.86 ± 0.308</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.75 ± 0.082</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1.48 ± 0.029</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>2.15 ± 0.499</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>1.66 ± 0.033</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>0.419 ± 0.139</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>0.781 ± 0.468</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cp35</td>
<td>1</td>
<td>1.15 ± 0.009</td>
<td></td>
<td>2.76 ± 0.901</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.86 ± 0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5.74 ± 6.115</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.10 ± 0.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>1.95 ± 0.107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cp45</td>
<td>1</td>
<td>1.42 ± 0.883</td>
<td></td>
<td>11.90 ± 1.518</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9.37 ± 1.401</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>14.95 ± 1.084</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8.23 ± 2.979</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>25.56 ± 2.243</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

amount of piceid once the normal papaya plants are developed.

To verify the identity of the new compound in the tissues, HPLC purified samples were collected at the expected retention time for piceid and subjected to tandem MS/MS analysis. The newly formed compound was verified to be resveratrol glucoside (Fig. 6.11a), also known as trans-piceid or polydatin (Liu et al. 2011, Teguo et al. 1996). Verification involved both comparisons to a standard piceid MS/MS ion spectrum and fragment analysis using ACD ChemSketch freeware (ACD/Labs). As shown in Figure 6.11a, the total ion chromatogram indicated a major peak from transgenic Arabidopsis at m/z 391.0, which corresponds to the parent ion (piceid) and the minor peak at m/z 229.5.
Figure 6.11. *Trans*-piceid mass spectra verified through tandem MS-MS with the corresponding fragmentation coefficient (m/z) of the compound. (a) Sample from *Carica papaya* *Cp45* line 5 (b) Standard piceid.
which corresponds to the unglycosylated free resveratrol. Other fragments were also observed, all of which were also found in the ion spectrum of the trans-piceid standard (Fig. 6.11b). These peaks were not present in extracts of WT controls (not shown).

**In vitro culture of transgenic papaya calli for plantlet regeneration.**

Several transgenic calli lines are currently being maintained for normal somatic embryo formation to regenerate plantlets for further characterization (Figure 6.12). Somatic embryos that are highly proliferating stage are regularly subcultured in CIM with reduced 2,4-D (0.5-1.0) to induce more somatic embryo maturation and germination which could be later transferred to the lighted condition in the growth chamber for plant shoot and root formation. ½ MS devoid of any plant growth regulator exposed to the lighted condition was used as a transition from germinating somatic embryos to plantlet formation. Moreover, elimination of 2,4-D as a plant hormone at this stage would eliminate its residual effects in the somatic embryos. Currently, the most advance stages of growth were observed on *Cp9* plants that started forming new leaf and roots in the few plate culture lines. We hope to obtain fully developed papaya plants from these transgenic somatic embryos as a potential source of plants with improved resveratrol content and potentially with enhance disease resistance in the future plant pathogen assays.
Figure 6.12. Developed transgenic papaya lines transformed with *Carica papaya* promoters: *Vst1* fusions growing under different stages of somatic embryo development.

**Discussion**

We successfully transformed and expressed a grapevine stilbene synthase gene *Vst1*, resulting in the accumulation of a novel foreign phytoalexin, piceid, in transgenic *Carica papaya* driven by the native promoters from papaya *Cp9*, *Cp35* and *Cp45*. The constitutive *CaMV35S* promoter was used as the positive control. Most of the transformations with *Vst1* in various crop species were driven by the *CaMV35S* promoter (Lim et al. 2005, Giorcelli et al. 2004, Pan et. al 2012). In other previous studies, *Vst1* and its own pathogen-inducible promoter were utilized where *Vst1* accumulation was
triggered following inoculation of pathogenic fungi (Hain et al. 1993, Thomzik et al. 1997, Stark-Lorenzen et al. 1997, Zhu et al. 2004). The study of transformation of Vst1 in papaya (Zhu et al. 2004) describes the transient nature of Vst1 expression. Initially, Vst1 driven by its native promoter does not accumulate mRNA at time zero, indicating that Vst1 has to be induced either by UV irradiation or wounding (Thomzik et al. 1997). Our study demonstrated the first report that basal Vst1 accumulation is possible using the native papaya promoters. Cp45 lines have highest Vst1 expression levels based on the qRT-PCR analysis as compared to CaMV35S, Cp9 and Cp35 promoters. However, the highest Vst1 transcript did not correlate to the highest concentration of piceid synthesized in the papaya transgenic calli lines as in Cp45 lines compared to the CaMV35S. This modulated expression is favorable for use in papaya genetic engineering since Cp45 maintained piceid levels at an optimal range prior to exposure to pathogen infection. The piceid produced in Cp45 provides some sort of initial protection for the plants. In addition, low basal piceid synthesis observed in Cp9 and Cp35 lines indicates that these lines are good candidate to disease induction experiments. We want to identify ideal promoters that have basal expression that does not alter plant physiological development, but would be highly induced in the presence of the pathogen. This means that piceid would return to pre-inoculation base level after its activity of controlling the disease infection. This regulated activity prevents imbalance of metabolites in the plants due to substrate competition of Vst1 in the flavonoid biosynthesis pathway.

The fact that these promoters are isolated from the same species where it was originally found, the promoter cis-elements were readily recognized by papaya producing the basal expression of Vst1 that resulted to piceid synthesis. The possible involvement of
these different genes in papaya physiological processes renders the \( Vstl \) expression detectable even without pathogen inoculation or environmental stress that can trigger promoter activity. These promoters isolated contain the modulatory activity in expressing a transgene back in the papaya plant. The advantage of using native papaya promoters over the \( Vstl \) native promoter from grapevine is to refine the control point of expression of the new transgene inserted into the plant’s genome. In our case, the papaya genes selected are pathogen-regulated (Porter et. al 2009) and therefore detecting a basal expression from these promoters is a very interesting and promising to pursue further characterization in the adult plants for disease resistance assay in the future.

Reverse-Phase High performance liquid chromatography (RP-HPLC) was employed to detect the presence of piceid in the transgenic plants. Methanolic extracts from calli tissues of \( CaMV35S \) had the highest piceid, followed by \( Cp45 \) while \( Cp35 \) and \( Cp9 \) consistently shown to be the weak promoters. Recent reports on plant transformation with stilbene synthase yielded differences in concentration of resveratrol and its derivatives depending on the species and the extraction procedure conducted. The identity of the compound synthesized was verified in the MS/MS analysis, and thus further confirmed that we obtained the same biochemical output as a result of \( Vstl \) expression in the transgenic papaya tissues similar to the compound synthesized in transgenic papaya that resulted to resistance against \( P. palmivora \) infection (Zhu et al 2004).

The possible competition for metabolic intermediates, \( p \)-coumaroyl CoA and malonyl CoA, by high levels of apparent \( Vstl \) activity does not have any adverse or deleterious effects in growth and development as observed in the developed transgenic \( Carica papaya \) calli lines. The basal piceid synthesis for \( Cp45 \) were comparable to the
range of piceid yield with the previous report in transgenic alfalfa (Hipskind & Paiva 2004) but a little lower in pathogen-induced papaya (Zhu et al. 2004) and did not result to any abnormality in the transgenic plants. Similarly, no visible changes on the overall growth and development of somatic embryos were observed on piceid producing calli driven by the native papaya promoters. This implies that Vst1 does not interfere in the calli stage thereby producing similar patterns of somatic embryo growth relative to the WT non-transformed calli. It would be early to say that the promoter regulatory activity can be perfectly demonstrated at the calli stage. We wanted to characterize the final output of Cp promoter:Vst1 fusions in the full grown regenerated papaya plants in vitro and finally in the soil to fully characterize them in the adult plant stages which is currently underway.

The data in the piceid yield is indicative of the strength and regulatory activity of native papaya promoters. It is interesting to note that Cp9 initially was detected to be high basal expression in the roots of papaya (Porter et al 2008). However, series of analysis in the eGFP and Vst1 fusions of Cp9 in Arabidopsis thaliana and in papaya did not show the strong basal expression in the analyses conducted as discussed in the previous chapters (III & IV). We could only assume that Cp9 promoter sequence requires other regulatory elements that are not located on the promoter region we have isolated. The other regions such as introns, 3’UTR sequences, or other enhancer sequences located farther away from the promoter region, or the availability of the trans-acting elements could all play in addition or in concert to express the right spatial and temporal promoter activity driving the gene expression.

The study helps us understand the intricacies of the process of transcription of
Vst1 as modulated by plant-derived promoters in papaya. In general, Cp45 showed strongest Vst1 transcript activity than in CaMV35S promoter. However, the final biochemical output as piceid was shown highest in the CaMV35S promoters. This non-correlative and regulated expression of Vst1 in piceid production for Cp45 lines is potentially due to the mRNA stability of Cp45 Vst1 transcripts that resulted to regulated enzyme activity after transcription. The activity of Cp45 with good basal expression looks very promising for papaya bioengineering. The fact that the gene of this promoter is pathogen inducible makes it a candidate for P. palmivora assays suppose we generate plants from these lines.

The papaya promoter’s activity in the transgenic calli yielding considerable amount of new phytoalexin in the form of resveratrol glucoside is a very promising result. The information on the preliminary levels of phytoalexin synthesis provides a guide to carefully assess the biochemical output in every stage of plant development and to finally relate that to an increased level of resistance in the future disease resistance assay.
Conclusion

This study provides basic information on the different promoter strength and activity in driving stilbene synthase expression that resulted to variable concentrations of resveratrol glucoside (piceid) in the papaya transgenic calli. We have utilized Cp9, Cp35, and Cp45 native papaya promoters that controlled the activity of stilbene synthase for production of piceid. The Vst1 transcripts levels and the biochemical output piceid was compared to the traditional CaMV35S promoter. To date, this is the first report of using native papaya promoters from the complex tree species with basal and regulated Vst1 activity in transgenic papaya calli. There were no differences on the growth patterns of the somatic embryos generated from several transgenic lines as compared to the non-transformed plants indicating that Vst1 expression did not cause altered and adverse physiological effects at the calli stage. The developed transgenic papaya lines generated can provide materials for future studies in abiotic and abiotic stress induction.

We want to continuously monitor the growth and development of these transgenic papaya calli lines to produce normal plants for further morphological and biochemical characterization in planta to finally assess its utility in papaya genetic improvement.
LITERATURE CITED


Benfey PN and Chua NH (1989) Regulated genes in transgenic plants. Science 244:174-


Fu H, Kim SY, Park WD (1995a) High-level tuber expression and sucrose inducibility of a potato Sus4 sucrose synthase gene require 5’ and 3’ flanking sequences and the


164


McCafferty HRK, Moore PH and Zhu JY (2006) Improved *Carica papaya* tolerance to


Schwekendiek A, Spring O, Heyerick A, Pickel B, Pitsch NT, Peschke F, de Keukeleire


Phytochemistry 42:1591-1593.


Supplementary Table 1. Stilbene synthase genes and promoters used to genetically transform plants, and the resulting effects on stilbene levels, resistance to pathogens and antioxidant activities (Delaunois et al. 2009, Giovinazzo et al. 2012).

<table>
<thead>
<tr>
<th>Plant Species</th>
<th>Gene</th>
<th>Promoter</th>
<th>Biochemical output</th>
<th>Stilbene concentration (µg/g FW)</th>
<th>Biological activity</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nicotiana tabacum L.</td>
<td>Arachis hypogea STS gene</td>
<td>Stress-induced promoter</td>
<td>trans-Resveratrol</td>
<td>-</td>
<td>-</td>
<td>Hain et al. (1990)</td>
</tr>
<tr>
<td></td>
<td>Vst1 and Vst2</td>
<td>Vst1</td>
<td>trans-Resveratrol</td>
<td>400</td>
<td>Resistance to Botrytis cinerea</td>
<td>Hain et al. (1993)</td>
</tr>
<tr>
<td></td>
<td>Chimeric STS gene</td>
<td>CaMV35S</td>
<td>Resveratrol</td>
<td>50-290</td>
<td>Altered flower morphology, male sterility</td>
<td>Fischer et al. (1997)</td>
</tr>
<tr>
<td>Triticum aestivum L.</td>
<td>Vst1</td>
<td>Vst1 (+35S-4 fold)</td>
<td>-</td>
<td>-</td>
<td>Resistance to Botrytis cinerea</td>
<td>Leckband and Lorz (1998)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Liang et al. (2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chimeric STS gene</td>
<td>Maize ubiquitin</td>
<td>2</td>
<td>-</td>
<td>Fettig and Hess (1999)</td>
</tr>
<tr>
<td></td>
<td>Vst1 and Vst2</td>
<td>Vst1 (+35S-4 fold)</td>
<td>Unknown derivative stilbene compounds</td>
<td>35-190</td>
<td>Resistance to Puccinia recondita and Septoria nodorum</td>
<td>Serazetdinova et al. (2005)</td>
</tr>
<tr>
<td>Hordeum vulgare L.</td>
<td>Vst1</td>
<td>Vst1 (+35S-4 fold)</td>
<td>-</td>
<td>-</td>
<td>Resistance to Botrytis cinerea</td>
<td>Leckband and Lorz (1998)</td>
</tr>
<tr>
<td>Medicago sativa L.</td>
<td>Arachis hypogea STS gene (AhRS)</td>
<td>CaMV35S</td>
<td>trans-Piceid</td>
<td>0.5-20</td>
<td>Resistance to Phoma medicaginis</td>
<td>Hipskin and Paiva (2000)</td>
</tr>
<tr>
<td>Arabidopsis thaliana L.</td>
<td>SbSTS1</td>
<td>CaMV35S</td>
<td>trans- and cis- Piceid</td>
<td>584</td>
<td>-</td>
<td>Yu et al. (2005, 2006)</td>
</tr>
<tr>
<td>Actinidia delicosa</td>
<td>pSV25</td>
<td>CaMV35S</td>
<td>trans-Piceid</td>
<td>20-182</td>
<td>No resistance to Botrytis cinerea</td>
<td>Kobayashi et al. (2000)</td>
</tr>
<tr>
<td>Vitis vinifera L.</td>
<td>Vst1</td>
<td>ms PR10.1</td>
<td>Resveratrol</td>
<td>In vitro resistance to Botrytis cinerea</td>
<td></td>
<td>Coutos-Thévenot et al. (2001)</td>
</tr>
<tr>
<td>Vitis pseudoreticulata STS</td>
<td></td>
<td>CaMV35S</td>
<td>Resveratrol</td>
<td>2.6</td>
<td>Under investigation</td>
<td>Fan et al. (2008)</td>
</tr>
<tr>
<td>Malus domestica Borkh.</td>
<td>Vst1</td>
<td>Vst1</td>
<td>Unknown resveratrol glycoside</td>
<td>-</td>
<td>-</td>
<td>Szankowski et al. (2003)</td>
</tr>
<tr>
<td></td>
<td>Vst1</td>
<td>Vst1</td>
<td>trans-Piceid</td>
<td>3-7 for non-UV-irradiated fruit; 23-62 for UV-irradiated fruit</td>
<td>No influence on other phenolic compounds</td>
<td>Rühmann et al. (2006)</td>
</tr>
</tbody>
</table>
### Supplementary Table 1. Continued.

<table>
<thead>
<tr>
<th>Plant Species</th>
<th>Gene</th>
<th>Promoter</th>
<th>Biochemical output</th>
<th>Stilbene concentration (µg/g FW)</th>
<th>Biological activity</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lycopersicon esculentum Mill.</td>
<td>Vst1 and Vst2</td>
<td>Vst1</td>
<td>Resveratrol</td>
<td>-</td>
<td>Resistance to Phytophthora infestans; No resistance to Botrytis cinerea and Alternaria solani</td>
<td>Thomzik et al. (1997)</td>
</tr>
<tr>
<td></td>
<td>StSy</td>
<td>CaMV35S</td>
<td>trans-Resveratrol and trans-Piceid</td>
<td>4.53</td>
<td>Antioxidant primary metabolism and increase in total antioxidant activity</td>
<td>Giovinazzo et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>StSy</td>
<td>CaMV35S</td>
<td>trans-Resveratrol and trans-Piceid</td>
<td>0.1-1.2</td>
<td>Enhancement of natural antiradical properties</td>
<td>Morelli et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>StSy</td>
<td>CaMV35S</td>
<td>trans-, cis- Resveratrol and trans-, cis-piceid</td>
<td>0.42-126 depending on the stage of ripening and fruit</td>
<td>Differences in rutin, naringenin and chlorogenic acid contents</td>
<td>Nicoletti et al. (2007)</td>
</tr>
<tr>
<td>Oryza sativa L.</td>
<td>Vst1</td>
<td>Vst1</td>
<td>-</td>
<td>-</td>
<td>Resistance to Pyricularia oryzae?</td>
<td>Stark-Lorenzen et al. (1997)</td>
</tr>
<tr>
<td>Rehmannia glutinosa Libosch.</td>
<td>AhRS3</td>
<td>CaMV35S</td>
<td>Resveratrol and piceid</td>
<td>22-116 Up to 650 with stress treatment</td>
<td>Antioxidant capabilities Resistance to Fusarium oxysporum</td>
<td>Lim et al. (2005)</td>
</tr>
<tr>
<td>Lactuca sativa L.</td>
<td>Parthenocissus henryana STS</td>
<td>CaMV35S</td>
<td>trans-Resveratrol</td>
<td>56.4</td>
<td>Effect on Hela cell morphology</td>
<td>Liu et al. (2006)</td>
</tr>
<tr>
<td>Pisum. Sativum L.</td>
<td>Vst1</td>
<td>Vst1</td>
<td>Occurrence of two resveratrol glucoside compounds</td>
<td>0.53-5.2</td>
<td>-</td>
<td>Richter et al. (2006)</td>
</tr>
<tr>
<td>Carica papaya L.</td>
<td>Vst1</td>
<td>Vst1</td>
<td>Resveratrol glucoside</td>
<td>54</td>
<td>Resistance to Phytophthora palmivora</td>
<td>Zhu et al. (2004)</td>
</tr>
<tr>
<td>Brassica napus L.</td>
<td>Vst1</td>
<td>p-nap</td>
<td>Resveratrol glucoside</td>
<td>361-616</td>
<td>Food quality improvement: high piceid rate content and reduction of sinapate esters</td>
<td>Hüsken et al. (2005)</td>
</tr>
<tr>
<td>Humulus lupulus L.</td>
<td>Vst1</td>
<td>CaMV35S</td>
<td>trans- and cis- Piceid, unknown stilbene cis-isomer, trans-astringin, trans- and cis-resveratrol</td>
<td>490-560</td>
<td>Higher amounts of flavonoids and acids</td>
<td>Schwekendiek et al. (2007)</td>
</tr>
<tr>
<td>Ipomoea batatas</td>
<td>Arachis hypogaea RS</td>
<td>CaMV35S</td>
<td>Resveratrol glucoside</td>
<td>52-340 (µg/g dry weight)</td>
<td>-</td>
<td>Pan et al. (2012)</td>
</tr>
</tbody>
</table>
**Supplementary Table 2.** Putative cis-acting regulatory motifs in *Cp9* (peroxidase) papaya promoter. Putative motifs recognized by transcription factors (TF) were found using PlantCARE program (Selected Matrix Score for all elements is ≥ 5). The site name, consensus sequence and the first organism where it was described were indicated. The strand positions (+) and (-) indicate sense or antisense DNA strands.

<table>
<thead>
<tr>
<th>Function</th>
<th>Motif</th>
<th>Species</th>
<th>Sequence</th>
<th>Sequence Found on Motif</th>
<th>Matrix Score</th>
<th>Position</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaerobic Induction</td>
<td>ARE</td>
<td><em>Zea mays</em></td>
<td>TGGTTT</td>
<td>AAACCA</td>
<td>6</td>
<td>1831-1836</td>
<td>-</td>
</tr>
<tr>
<td>Auxin</td>
<td>TGA-element</td>
<td><em>Brassica oleracea</em></td>
<td>AACGAC</td>
<td>AACGAC</td>
<td>6</td>
<td>20-25</td>
<td>+</td>
</tr>
<tr>
<td>Defense and Stress</td>
<td>TC-rich repeats</td>
<td><em>Nicotiana tabacum</em></td>
<td>ATTTTCTTCA</td>
<td>TGAAGAAAAT</td>
<td>10</td>
<td>1645-1654</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ATTTTCTTCA</td>
<td>TGAATAAAAAT</td>
<td>9</td>
<td>52-61</td>
<td>-</td>
</tr>
<tr>
<td>Drought</td>
<td>MBS</td>
<td><em>Arabidopsis thaliana</em></td>
<td>CAACTG</td>
<td>CAACTG</td>
<td>6</td>
<td>535-540</td>
<td>+</td>
</tr>
<tr>
<td>Endosperm Expression</td>
<td>GCN4 motif</td>
<td><em>Oryza sativa</em></td>
<td>TGAGTCA</td>
<td>TGAGTCA</td>
<td>7</td>
<td>501-507</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Skn-1 motif</td>
<td><em>Oryza sativa</em></td>
<td>GTCAT</td>
<td>GTCAT</td>
<td>5</td>
<td>504-508</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GTCAT</td>
<td>ATGAC</td>
<td>5</td>
<td>269-273</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GTCAT</td>
<td>GTCAT</td>
<td>5</td>
<td>1123-1127</td>
<td>+</td>
</tr>
<tr>
<td>Enhancer</td>
<td>5UTR Py-rich stretch</td>
<td><em>Lycopersicon esculentum</em></td>
<td>TTTCTTCTCT</td>
<td>AGAAAAGAAA</td>
<td>9</td>
<td>1957-1966</td>
<td>-</td>
</tr>
<tr>
<td>Ethylene</td>
<td>ERE</td>
<td><em>Diathus caryophyllus</em></td>
<td>ATTTCAAA</td>
<td>TTTGAAAT</td>
<td>8</td>
<td>1556-1563</td>
<td>-</td>
</tr>
<tr>
<td>Heat Stress</td>
<td>HSE</td>
<td><em>Brassica oleracea</em></td>
<td>AAAAATTTTC</td>
<td>TAAATTTTTT</td>
<td>9</td>
<td>219-228</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAAAATTTTC</td>
<td>AAAAAGTTTC</td>
<td>9</td>
<td>964-973</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAAAATTTTC</td>
<td>TAAATTTTTT</td>
<td>9</td>
<td>455-464</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAAAATTTTC</td>
<td>GAAAATTTTT</td>
<td>9</td>
<td>921-930</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAAAATTTTC</td>
<td>TAAATTTTTC</td>
<td>9</td>
<td>233-241</td>
<td>-</td>
</tr>
<tr>
<td>Light Responsiveness</td>
<td>AE-box</td>
<td><em>Arabidopsis thaliana</em></td>
<td>AGAAACTT</td>
<td>AGTTTCT</td>
<td>8</td>
<td>968-974</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>AT1-motif</td>
<td><em>Solanum tuberosum</em></td>
<td>ATTAATTTTACA</td>
<td>ATTTAAATTTA</td>
<td>11</td>
<td>140-151</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ATTAATTTTACA</td>
<td>ATTAATTTTACT</td>
<td>11</td>
<td>778-789</td>
<td>+</td>
</tr>
<tr>
<td>ATCT motif</td>
<td>Arabidopsis thaliana</td>
<td>AATCTAATCT</td>
<td>AGATTATAT</td>
<td>9</td>
<td>1383-1391</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------</td>
<td>------------</td>
<td>-----------</td>
<td>------</td>
<td>-----------</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td>Petroselinum crispum</td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>147-151</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>155-160</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>163-168</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>181-186</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>276-281</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>381-386</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>496-501</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>788-783</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>1174-1179</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>1210-1215</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>1246-1252</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>1335-1340</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box I</td>
<td>Pisum sativum</td>
<td>TTTCAAA</td>
<td>TTTGAAA</td>
<td>7</td>
<td>918-924</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTTCAAA</td>
<td>TTTGAAA</td>
<td>7</td>
<td>1556-1562</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>GAG motif</td>
<td>Arabidopsis thaliana</td>
<td>AGAGAGT</td>
<td>AGAGAGT</td>
<td>7</td>
<td>1905-1911</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>GT1 motif</td>
<td>Arabidopsis thaliana</td>
<td>GGTTAA</td>
<td>GGTTAA</td>
<td>6</td>
<td>1223-1229</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Avena sativa</td>
<td>GGTTAA</td>
<td>TTAACC</td>
<td>6</td>
<td>1730-1735</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGTTAA</td>
<td>TTAACC</td>
<td>6</td>
<td>1729-1735</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>I-box</td>
<td>Flaveria trinervia</td>
<td>GATATGG</td>
<td>GATATGG</td>
<td>7</td>
<td>587-593</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Sp1</td>
<td>Zea mays</td>
<td>CC(G/A)CCC</td>
<td>CCACCCGATA</td>
<td>5.5</td>
<td>581-590</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>TCT motif</td>
<td>Arabidopsis thaliana</td>
<td>TCTTAC</td>
<td>TCTTAC</td>
<td>6</td>
<td>1780-1785</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>chs-CMA1a</td>
<td>Daucus carota</td>
<td>TTACTTAA</td>
<td>TTACTTAA</td>
<td>8</td>
<td>1324-1332</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTACTTAA</td>
<td>TTACTTAA</td>
<td>8</td>
<td>475-482</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>HD-Zip 3</td>
<td>GTAAT(G/C)ATTAC</td>
<td>GTAATGATTAA</td>
<td>9.5</td>
<td>490-499</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>binding site</td>
<td>Arabidopsis thaliana</td>
<td>GTAAT(G/C)ATTAC</td>
<td>GTAATGATTAA</td>
<td>9.5</td>
<td>490-499</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Unnamed 1</td>
<td>Glycine max</td>
<td>GAATTTAATTAA</td>
<td>GAATTTAATTAA</td>
<td>11</td>
<td>658-669</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Salicylic</td>
<td>TCA- element</td>
<td>CAGAAAAGGA</td>
<td>CAAAAAAGGA</td>
<td>9</td>
<td>1032-1041</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>acid</td>
<td>Wound Responsive</td>
<td>WUN-motif</td>
<td>Brassica oleracea</td>
<td>TCATTACGAA</td>
<td>TCATTATGAA</td>
<td>9</td>
<td>1016-1025</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------</td>
<td>-----------</td>
<td>-------------------</td>
<td>------------</td>
<td>------------</td>
<td>-----</td>
<td>-----------</td>
</tr>
<tr>
<td>GAGAAGAATA</td>
<td>GAGAAGAAGA</td>
<td>1609-1618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAATTTCCT</td>
<td>AAATTTCCT</td>
<td>1838-1846</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Supplementary Table 3.** Putative cis-acting regulatory motifs in *Cp29 (B-1,3-glucanase)* papaya promoter. Putative motifs recognized by transcription factors (TF) were found using PlantCARE program (Selected Matrix Score for all elements is ≥ 5). The site name, consensus sequence and the first organism where it was described were indicated. The strand positions (+) and (-) indicate sense or antisense DNA strands.

<table>
<thead>
<tr>
<th>Function</th>
<th>Motif</th>
<th>Species</th>
<th>Sequence</th>
<th>Motif Sequence Found</th>
<th>Matrix Score</th>
<th>Position</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaerobic Induction</td>
<td>ARE</td>
<td>Zea mays</td>
<td>TGGTTT</td>
<td>TGGTTT</td>
<td>6</td>
<td>1412-1417</td>
<td>+</td>
</tr>
<tr>
<td>Binding Protein</td>
<td>AT-rich element</td>
<td>Glycine max</td>
<td>ATAGAAATCAA</td>
<td>ATAGAAATCAT</td>
<td>10</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Circadian control</td>
<td>circadian</td>
<td>Lycopersicon esculentum</td>
<td>CAAAGATATC</td>
<td>GATGTCTTTTG</td>
<td>9</td>
<td>86-95</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CAANNNNATC</td>
<td>CAATATAATC</td>
<td>6</td>
<td>904-913</td>
<td>+</td>
</tr>
<tr>
<td>Defense and Stress</td>
<td>TC-rich repeats</td>
<td>Nicotiana tabacum</td>
<td>ATTTTCTCCA</td>
<td>ATTTTATCCA</td>
<td>9</td>
<td>120-129</td>
<td>+</td>
</tr>
<tr>
<td>Drought</td>
<td>MBS</td>
<td>Arabidopsis thaliana</td>
<td>CAACTG</td>
<td>CAGTTG</td>
<td>6</td>
<td>935-940</td>
<td>-</td>
</tr>
<tr>
<td>Endosperm Expression</td>
<td>Skn-1 motif</td>
<td>Oryza sativa</td>
<td>GTCAT</td>
<td>ATGAC</td>
<td>5</td>
<td>11-15</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GTCAT</td>
<td>ATGAC</td>
<td>5</td>
<td>137-141</td>
<td></td>
</tr>
<tr>
<td>Enhancer</td>
<td>5UTR Py-rich stretch</td>
<td>Lycopersicon esculentum</td>
<td>TTTCTTCTCT</td>
<td>TTTCTTCTCT</td>
<td>9</td>
<td>786-795</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>TA-rich region</td>
<td>Nicotiana tabacum</td>
<td>TATATATATATATATATA</td>
<td>TATATATATATATATATA</td>
<td>21</td>
<td>431-452</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TATATATATATATATA</td>
<td>TATATATATATATATA</td>
<td>21</td>
<td>439-460</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TATATATATATATATA</td>
<td>TATATATATATATATA</td>
<td>22</td>
<td>435-456</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TATATATATATATATA</td>
<td>TATATATATATATATA</td>
<td>22</td>
<td>433-454</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TATATATATATATATA</td>
<td>TATATATATATATATA</td>
<td>22</td>
<td>437-458</td>
<td>+</td>
</tr>
<tr>
<td>Heat</td>
<td>HSE</td>
<td>Brassica</td>
<td>AAAAAATTTTC</td>
<td>GAATTTTTTT</td>
<td>9</td>
<td>779-788</td>
<td>-</td>
</tr>
<tr>
<td>Light Responsiveness</td>
<td>CAT box</td>
<td>Arabidopsis thaliana</td>
<td>GCCACT</td>
<td>AGTGGCT</td>
<td>6</td>
<td>1245-1251</td>
<td>-</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------</td>
<td>----------------------</td>
<td>--------</td>
<td>----------</td>
<td>-----</td>
<td>-----------</td>
<td>----</td>
</tr>
<tr>
<td>3-AF1 binding site</td>
<td>Solanum tuberosum</td>
<td>TAAGAGAGGAAG</td>
<td>TAAGAGAGGAC</td>
<td>10</td>
<td>1190-1200</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>4cl-CMA2b</td>
<td>Solanum tuberosum</td>
<td>TCTCAACCAACCAG</td>
<td>TGTGGTTGGAGA</td>
<td>13</td>
<td>1265-1275</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>ATCT motif</td>
<td>Arabidopsis thaliana</td>
<td>AATCTAATCT</td>
<td>AATATAATCC</td>
<td>9</td>
<td>905-914</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AATCTAATCC</td>
<td>TATCTAATCC</td>
<td>9</td>
<td>1385-1394</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AATCTAATCT</td>
<td>AGATTAGAGT</td>
<td>9</td>
<td>1018-1027</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td>Petroselinum crispum</td>
<td>ATTAATATTAATATTAAT</td>
<td>ATTAATATTAATATTAAT</td>
<td>6</td>
<td>464-469</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAATATTAAT</td>
<td>ATTAATATTAAT</td>
<td>6</td>
<td>649-654</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAATATTAAT</td>
<td>ATTAATATTAAT</td>
<td>6</td>
<td>824-829</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAATATTAAT</td>
<td>ATTAATATTAAT</td>
<td>6</td>
<td>978-983</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAATATTAAT</td>
<td>ATTAATATTAAT</td>
<td>6</td>
<td>989-994</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAATATTAAT</td>
<td>ATTAATATTAAT</td>
<td>6</td>
<td>993-998</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAATATTAAT</td>
<td>ATTAATATTAAT</td>
<td>6</td>
<td>1072-1077</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTAATATTAAT</td>
<td>ATTAATATTAAT</td>
<td>6</td>
<td>1320-1325</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>GAG motif</td>
<td>Spinacea oleracea</td>
<td>AGAGATGG</td>
<td>CATCTCT</td>
<td>7</td>
<td>1420-1426</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>GT1 motif</td>
<td>Avena sativa</td>
<td>GGTATAAT</td>
<td>ATTAACC</td>
<td>7</td>
<td>580-586</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arabidopsis thaliana</td>
<td>GGTATAAT</td>
<td>TTAACC</td>
<td>6</td>
<td>581-586</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>L-box</td>
<td>Petroselinum crispum</td>
<td>TCTCAACCAACC</td>
<td>GGTGGAGAGAGA</td>
<td>10</td>
<td>1268-1279</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>chs-CMA1a</td>
<td>Daucus carota</td>
<td>TTACTTAA</td>
<td>TTAAGTAA</td>
<td>8</td>
<td>302-309</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
**Supplementary Table 4.** Putative cis-acting regulatory motifs in *Cp35 (ferulate-5-hydroxylase)* papaya promoter. Putative motifs recognized by transcription factors (TF) were found using PlantCARE program (Selected Matrix Score for all elements is ≥ 5). The site name, consensus sequence and the first organism where it was described were indicated. The strand positions (+) and (-) indicate sense or antisense DNA strands.

<table>
<thead>
<tr>
<th>Function</th>
<th>Motif</th>
<th>Species</th>
<th>Sequence</th>
<th>Motif Sequence Found</th>
<th>Matrix Score</th>
<th>Position</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abscisic acid</td>
<td>ABRE</td>
<td><em>Oryza sativa</em></td>
<td>AGTACGTGCC</td>
<td>AGTACGTGCC</td>
<td>9</td>
<td>774-782</td>
<td>+</td>
</tr>
<tr>
<td>Anaerobic Induction</td>
<td>ARE</td>
<td><em>Zea mays</em></td>
<td>TGGTTT</td>
<td>TGGTTT</td>
<td>6</td>
<td>603-608</td>
<td>+</td>
</tr>
<tr>
<td>Defense and Stress TC-rich repeats</td>
<td>TC-rich repeats</td>
<td><em>Nicotiana tabacum</em></td>
<td>ATTTCTTCTCA</td>
<td>ATTTCTTCTCA</td>
<td>9</td>
<td>149-158</td>
<td>+</td>
</tr>
<tr>
<td>Drought</td>
<td>MBS</td>
<td><em>Arabidopsis thaliana</em></td>
<td>TAACTG</td>
<td>TAACTG</td>
<td>6</td>
<td>160-165</td>
<td>+</td>
</tr>
<tr>
<td>Endosperm Expression GCN4 motif</td>
<td>Oryza sativa</td>
<td>CAAGCCA</td>
<td>AGTAGGTGG</td>
<td>7</td>
<td>774-782</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Endosperm Expression Skn-1 motif</td>
<td>Oryza sativa</td>
<td>GTCAT</td>
<td>ATGAC</td>
<td>5</td>
<td>993-997</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Enhancer 5UTR Py-rich stretch</td>
<td><em>Lycopersicon esculentum</em></td>
<td>TTTCTTCTCT</td>
<td>TTTCTTCTAT</td>
<td>9</td>
<td>151-160</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Enhancer TA-rich region</td>
<td><em>Nicotiana tabacum</em></td>
<td>TATATATATATATAATATAA</td>
<td>TATATATATATATAATATAA</td>
<td>21</td>
<td>1081-1102</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Fungal Elicitor</td>
<td>Box-W1</td>
<td><em>Petroselinum crispum</em></td>
<td>TTGACC</td>
<td>GGTCAA</td>
<td>6</td>
<td>1502-1507</td>
<td>-</td>
</tr>
<tr>
<td>Gibberellin</td>
<td>TATC-box</td>
<td><em>Oryza sativa</em></td>
<td>TATCCCA</td>
<td>TATCCCA</td>
<td>7</td>
<td>1727-1733</td>
<td>+</td>
</tr>
<tr>
<td>Heat Stress</td>
<td>HSE</td>
<td><em>Brassica oleracea</em></td>
<td>AAAAATTTTC</td>
<td>GTAATTTTTT</td>
<td>9</td>
<td>1405-1414</td>
<td>-</td>
</tr>
<tr>
<td>Light Responsive ACE-box</td>
<td><em>Petroselinum crispum</em></td>
<td>AAAACGTTTA</td>
<td>AAAAAGTTTA</td>
<td>9</td>
<td>1462-1471</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Light Responsive ATCT motif</td>
<td><em>Pisum sativum</em></td>
<td>AATCTAATCC</td>
<td>AATTTAATCT</td>
<td>9</td>
<td>1574-1583</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Motif</td>
<td>Species</td>
<td>Sequence</td>
<td>Length</td>
<td>Position</td>
<td>Methyl jasmonate</td>
<td>Salicylic acid</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>----------</td>
<td>--------</td>
<td>----------</td>
<td>------------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>Box 1</td>
<td><em>Arabidopsis thaliana</em></td>
<td>AATCTAATCT</td>
<td>CATCTAATCC</td>
<td>9</td>
<td>1003-1012</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box 1</td>
<td><em>Pisum sativum</em></td>
<td>TTTAAAAA</td>
<td>TTTGAAAA</td>
<td>7</td>
<td>49-55</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td><em>Petroselinum crispum</em></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>416-421</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td><em>Pisum sativum</em></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>528-533</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td><em>Petroselinum crispum</em></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>637-642</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td><em>Petroselinum crispum</em></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>654-659</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td><em>Petroselinum crispum</em></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>879-884</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td><em>Petroselinum crispum</em></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>1212-1217</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td><em>Petroselinum crispum</em></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>1553-1558</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td><em>Petroselinum crispum</em></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>1588-1593</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td><em>Petroselinum crispum</em></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>799-804</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td><em>Antirrhinum majus</em></td>
<td>CACGTA</td>
<td>TACGTG</td>
<td>6</td>
<td>1481-1486</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G-box</td>
<td><em>Solanum tuberosum</em></td>
<td>CACATGG</td>
<td>CCATGTG</td>
<td>7</td>
<td>765-771</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G-box</td>
<td><em>Daucus carota</em></td>
<td>TACGTG</td>
<td>TACGTG</td>
<td>6</td>
<td>1481-1486</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>GA-motif</td>
<td><em>Helianthus annuus</em></td>
<td>AAAGATGA</td>
<td>TCACTCTTT</td>
<td>8</td>
<td>267-274</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>GATA-motif</td>
<td><em>Arabidopsis thaliana</em></td>
<td>GATAGGA</td>
<td>TCCTATC</td>
<td>7</td>
<td>1327-1333</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>GT1-motif</td>
<td><em>Avena sativa</em></td>
<td>GGTTAAT</td>
<td>ATTAACC</td>
<td>7</td>
<td>321-327</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>TCT motif</td>
<td><em>Arabidopsis thaliana</em></td>
<td>TCTTAC</td>
<td>GTAAGA</td>
<td>6</td>
<td>1141-1146</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Methyl jasmonate</td>
<td>CGTCA-motif</td>
<td>CGTCA</td>
<td>TGACG</td>
<td>5</td>
<td>994-998</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Methyl jasmonate</td>
<td>TGACG-motif</td>
<td>TGACG</td>
<td>TGACG</td>
<td>5</td>
<td>994-998</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Salicylic acid</td>
<td>TCA-element</td>
<td>CCATCTTTTT</td>
<td>AAAATATGG</td>
<td>9</td>
<td>80-89</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
**Supplementary Table 5.** Putative cis-acting regulatory motifs in *Cp45* (hypersensitive-induced response) papaya promoter. Putative motifs recognized by transcription factors (TF) were found using PlantCARE program (Selected Matrix Score for all elements is ≥ 5). The site name, consensus sequence and the first organism where it was described were indicated. The strand positions (+) and (-) indicate sense or antisense DNA strands.

<table>
<thead>
<tr>
<th>Function</th>
<th>Motif</th>
<th>Species</th>
<th>Sequence</th>
<th>Motif Sequence Found</th>
<th>Matrix Score</th>
<th>Position</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>α Amylase</td>
<td>A-Box</td>
<td><em>Oryza sativa</em></td>
<td>AATAACAAACTCC</td>
<td>GAGTTTTTTATT</td>
<td>12</td>
<td>1306-1317</td>
<td>-</td>
</tr>
<tr>
<td>Abscisic acid</td>
<td>ABRE</td>
<td><em>Arabidopsis thaliana</em></td>
<td>TACGGTC</td>
<td>GACCGTA</td>
<td>7</td>
<td>902-908</td>
<td>-</td>
</tr>
<tr>
<td>Anaerobic Induction</td>
<td>ARE</td>
<td><em>Zea mays</em></td>
<td>TGGTTT</td>
<td>AAACCA</td>
<td>6</td>
<td>655-660</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TGGTTT</td>
<td>TGGTTT</td>
<td>6</td>
<td>1467-1472</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TGGTTT</td>
<td>AAACCA</td>
<td>6</td>
<td>1232-1237</td>
<td>-</td>
</tr>
<tr>
<td>AT-rich DNA binding protein</td>
<td>AT-rich</td>
<td><em>Glycine max</em></td>
<td>ATAGAAATCAA</td>
<td>TTGATTTCCAT</td>
<td>10</td>
<td>1358-1368</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>element</td>
<td></td>
<td>ATAGAAATCAA</td>
<td>ATAGAAACTAA</td>
<td>10</td>
<td>1512-1522</td>
<td>+</td>
</tr>
<tr>
<td>Defense and Stress</td>
<td>TC-rich repeats</td>
<td><em>Nicotiana tabacum</em></td>
<td>ATTTTCTTCA</td>
<td>TCAAGAAAATT</td>
<td>9</td>
<td>581-591</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ATTTTCTTCA</td>
<td>TGGCTGAAAAT</td>
<td>9</td>
<td>1884-1893</td>
<td>-</td>
</tr>
<tr>
<td>Drought</td>
<td>MBS</td>
<td><em>Arabidopsis thaliana</em></td>
<td>TAACTG</td>
<td>CAGTTA</td>
<td>6</td>
<td>1491-1496</td>
<td>-</td>
</tr>
<tr>
<td>Endosperm Expression</td>
<td>AACA motif</td>
<td><em>Arabidopsis thaliana</em></td>
<td>TAACCA</td>
<td>TCGAGAAAAT</td>
<td>9</td>
<td>69-74</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>GCN4 motif</td>
<td></td>
<td>TAACCA</td>
<td>TGGCTTGTGA</td>
<td>7</td>
<td>1240-1246</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Skn-1 motif</td>
<td></td>
<td>GTCA</td>
<td>GTCA</td>
<td>5</td>
<td>608-612</td>
<td>+</td>
</tr>
<tr>
<td>Elicitor Responsive</td>
<td>ELI-box3</td>
<td><em>Brassica oleracea</em></td>
<td>AAACCAATT</td>
<td>AAACCAATT</td>
<td>9</td>
<td>655-663</td>
<td>+</td>
</tr>
<tr>
<td>Enhancer</td>
<td>TA-rich region</td>
<td><em>Nicotiana tabacum</em></td>
<td>TATATATATATAT</td>
<td>ATATATATATAT</td>
<td>20</td>
<td>709-721</td>
<td>+</td>
</tr>
<tr>
<td>Gibberellin</td>
<td>GARE-motif</td>
<td><em>Brassica oleracea</em></td>
<td>AAACAGA</td>
<td>TCTGTGTG</td>
<td>7</td>
<td>807-813</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TCTGTGTG</td>
<td>CAACATA</td>
<td>7</td>
<td>24-30</td>
<td>-</td>
</tr>
<tr>
<td>Heat Stress</td>
<td>HSE</td>
<td><em>Brassica</em></td>
<td>AAAAATTTTC</td>
<td>GAAAATTTTT</td>
<td>9</td>
<td>1170-1179</td>
<td>+</td>
</tr>
<tr>
<td>Light Response</td>
<td>3-AF1 binding site</td>
<td>oleracea</td>
<td>Solanum tuberosum</td>
<td>AAAACGTTTA</td>
<td>AAATATCTCTA</td>
<td>10</td>
<td>1889-1899</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------</td>
<td>----------</td>
<td>-------------------</td>
<td>-------------</td>
<td>--------------</td>
<td>----</td>
<td>-----------</td>
</tr>
<tr>
<td>ATC- motif</td>
<td>Arabidopsis thaliana</td>
<td></td>
<td>GCCAATCC</td>
<td>GGACTTTT</td>
<td>8</td>
<td>844-851</td>
<td>-</td>
</tr>
<tr>
<td>ATCT motif</td>
<td>Arabidopsis thaliana</td>
<td></td>
<td>AATCTAATCT</td>
<td>AATCTAATCA</td>
<td>9</td>
<td>1499-1510</td>
<td>+</td>
</tr>
<tr>
<td>Box I</td>
<td>Pisum sativum</td>
<td></td>
<td>TTTAAAA</td>
<td>TTTAAAA</td>
<td>7</td>
<td>949-956</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TTTAAAA</td>
<td>TTTGAAAA</td>
<td>7</td>
<td>1016-1022</td>
<td>-</td>
</tr>
<tr>
<td>Box 4</td>
<td>Petroselinum crispum</td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>302-307</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>1496-1501</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>1545-1550</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ATTAAT</td>
<td>ATTAAT</td>
<td>6</td>
<td>1799-1804</td>
<td>+</td>
</tr>
<tr>
<td>G-Box</td>
<td>Antirrhinum majus</td>
<td></td>
<td>CACGTA</td>
<td>TACGTG</td>
<td>6</td>
<td>461-466</td>
<td>-</td>
</tr>
<tr>
<td>G-box</td>
<td>Nicotiana tabacum</td>
<td></td>
<td>CAGACGTGGCA</td>
<td>CAGACGTGGCA</td>
<td>10</td>
<td>733-743</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Daucus carota</td>
<td></td>
<td>TACGTG</td>
<td>TACGTG</td>
<td>6</td>
<td>461-466</td>
<td>+</td>
</tr>
<tr>
<td>GAG-motif</td>
<td>Hordeum vulgare</td>
<td></td>
<td>GGAGATG</td>
<td>CATCTCC</td>
<td>7</td>
<td>140-146</td>
<td>-</td>
</tr>
<tr>
<td>MRE</td>
<td>Petroselinum crispum</td>
<td></td>
<td>AACCTAA</td>
<td>AACCTAA</td>
<td>7</td>
<td>329-335</td>
<td>+</td>
</tr>
<tr>
<td>Low Temperature Response</td>
<td>LTR</td>
<td>Hordeum vulgare</td>
<td>CCGAAA</td>
<td>CCGAAA</td>
<td>6</td>
<td>680-685</td>
<td>+</td>
</tr>
<tr>
<td>Methyl jasmonate</td>
<td>CGTCA-motif</td>
<td>Hordeum vulgare</td>
<td>CGTCA</td>
<td>CGTCA</td>
<td>5</td>
<td>781-785</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>TGACG-motif</td>
<td>Hordeum vulgare</td>
<td>TGACG</td>
<td>CGTCA</td>
<td>5</td>
<td>781-785</td>
<td>-</td>
</tr>
<tr>
<td>Meristem specific activation</td>
<td>dOCT</td>
<td>Arabidopsis thaliana</td>
<td>CaCGGATC</td>
<td>CACGGATC</td>
<td>8</td>
<td>771-778</td>
<td>+</td>
</tr>
<tr>
<td>Salicylic acid</td>
<td>TCA-element</td>
<td>Brassica oleracea</td>
<td>CAGAAAAGGA</td>
<td>TCCTTTTCTA</td>
<td>9</td>
<td>1067-1076</td>
<td>-</td>
</tr>
</tbody>
</table>
**Supplementary Table 6.** Comparison of the putative cis-elements of Cp9, Cp29, Cp35 and Cp45 promoters by PLACE Promoter Analysis.

<table>
<thead>
<tr>
<th>MOTIF</th>
<th>Cp9</th>
<th>Cp29</th>
<th>Cp35</th>
<th>Cp45</th>
<th>FUNCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10PEHVPSBD</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Chloroplast plastid gene (psbD) promoter</td>
</tr>
<tr>
<td>-300ELEMENT</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>Present upstream of the promoter from the B-hordein gene</td>
</tr>
<tr>
<td>-300MOTIFZMZEIN</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Motif in -300 elements of alpha-zein genes of maize</td>
</tr>
<tr>
<td>2SSEEDPROTBANAPA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Conserved in many storage-protein gene promoter</td>
</tr>
<tr>
<td>AACACOREOSGLUB1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Involved in controlling the endosperm-specific expression</td>
</tr>
<tr>
<td>ABRELATERD1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>Etiolation-induced expression of erd1 (early responsive to dehydration) in Arabidopsis</td>
</tr>
<tr>
<td>ABRERATCAL</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Upstream regions of 162 Ca2+-responsive upregulated genes</td>
</tr>
<tr>
<td>ACGTABOX</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>Responsible for sugar repression</td>
</tr>
<tr>
<td>ACGTATERD1</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>Etiolation-induced expression of erd1 (early responsive to dehydration) in Arabidopsis</td>
</tr>
<tr>
<td>AGCBOXNPGLB</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Conserved in most PR-protein genes</td>
</tr>
<tr>
<td>AMYBOX1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>Conserved sequence found in 5'-upstream region of alpha-amylose gene</td>
</tr>
<tr>
<td>AMYBOX2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Conserved sequence found in 5'-upstream region of alpha-amylose gene</td>
</tr>
<tr>
<td>ANAERO1CONSENSUS</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>In silico in promoters of 13 anaerobic genes involved in the fermentative pathway</td>
</tr>
<tr>
<td>ANAERO3CONSENSUS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Motifs in promoters of anaerobically induced genes of different plant species.</td>
</tr>
<tr>
<td>ARFAT</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>Auxin response factor (ARF) binding site in the</td>
</tr>
</tbody>
</table>
promoters of auxin response genes

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Frequency</th>
<th>Conserved</th>
<th>Consensus</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARR1AT</td>
<td>22 14 22 29</td>
<td>&quot;ARR1-binding element&quot; found in Arabidopsis; ARR1 is a response regulator; AGATT is found in the promoter of rice non-symbiotic haemoglobin-2 (NSHB) gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASF1MOTIFCAMV</td>
<td>0 0 1 1</td>
<td>ASF-1 binding site in CaMV 35S promoter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIHD1OS</td>
<td>4 2 0 0</td>
<td>Binding site of OsBIHD1, a rice BELL homeodomain transcription factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOXIIINTPATPB</td>
<td>4 2 4 2</td>
<td>Conserved in several NCII (nonconsensus type II) promoters of plastid genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOXLCOREDCPAL</td>
<td>0 0 1 0</td>
<td>&quot;CAAT promoter consensus sequence&quot; found in legA gene of pea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAATBOX1</td>
<td>19 13 11 21</td>
<td>&quot;CAAT promoter consensus sequence&quot; found in legA gene of pea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CACTFTPPCA1</td>
<td>34 20 34 21</td>
<td>Component of Mem1 (mesophyll expression module 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CANBNNAPA</td>
<td>0 0 0 2</td>
<td>Seed specificity; activator and repressor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAREOSREP1</td>
<td>0 2 0 0</td>
<td>Promoter region of a cystein proteinase (REP-1) gene in rice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CARGATCONSENSUS</td>
<td>2 0 0 0</td>
<td>Promoter found in the flowering-time gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CARGCW8GAT</td>
<td>6 6 6 8</td>
<td>A variant of CArG motif with a longer A/T-rich core; Binding site for AGL15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CATATGGMSAUR</td>
<td>0 0 2 2</td>
<td>Involved in auxin responsiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBFHV</td>
<td>1 0 0 1</td>
<td>Binding site of dehydration-responsive binding proteins (DREBs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCA1ATLHCB1</td>
<td>0 2 0 2</td>
<td>Related to regulation by phytochrome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCAATBOX1</td>
<td>1 0 1 6</td>
<td>Common sequence found in the 5'-non-coding regions of eukaryotic genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CELLCYCLESC</td>
<td>0 1 0 0</td>
<td>Cell-cycle-specific activation of transcription;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEREGLUBOX2PSLEGA</td>
<td>1 0 0 0</td>
<td>Sequence homologous to the cereal glutenin gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Element</td>
<td>Score 1</td>
<td>Score 2</td>
<td>Score 3</td>
<td>Score 4</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>CGACGOSAMY3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>CGCBOXAT</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>CPBCSPOR</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>CURECORECR</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>DOFCOREZM</td>
<td>25</td>
<td>14</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>DPBCOREDCDC3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>DRE2COREZMRAB17</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DRECRTCOREAT</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>EBOXBNNAPA</td>
<td>12</td>
<td>2</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>EECCRCAH1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>ELRECOREPCRP1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ERELEE4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>EVENINGAT</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GARE2OSREP1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>GAREAT</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>GATABOX</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>Expression Pattern</td>
<td>Core/Box</td>
<td>Sequence Information</td>
<td>Details</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>----------</td>
<td>----------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>GCCCORE</td>
<td>1 0 0 0 0</td>
<td>Core of GCC-box found in many pathogen-responsive genes such as PDF1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCN4OSGLUB1</td>
<td>1 0 0 0 0</td>
<td>Required for endosperm-specific expression;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLMHVCHORD</td>
<td>2 0 0 0 0</td>
<td>Involved in the nitrogen response of c-hordein promoter;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GT1CONSENSUS</td>
<td>20 12 20 33</td>
<td>Consensus GT-1 binding site in many light-regulated genes; GT-1 can stabilize the TFIIA-TBP-DNA (TATA box) complex; The activation mechanism of GT-1 may be achieved through direct interaction between TFIIA and GT-1; Binding of GT-1-like factors to the PR-1a promoter influences the level of SA-inducible gene expression;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GT1CORE</td>
<td>2 1 2 0</td>
<td>Critical for GT-1 binding to box II of rbcS;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GT1GMSCAM4</td>
<td>4 2 4 2</td>
<td>Plays a role in pathogen- and salt-induced SCaM-4 gene expression;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTGANTG10</td>
<td>9 7 9 8</td>
<td>&quot;GTGA motif&quot; in tobacco late pollen gene g10 which shows homology to pectate lyase;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDZIP2ATATHB2</td>
<td>2 4 2 0</td>
<td>Binding site of the ATHB-2 regulated by light signals;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBOX</td>
<td>0 0 0 2</td>
<td>Conserved sequence upstream of light-regulated genes;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBOXCORE</td>
<td>8 5 0 10</td>
<td>Conserved sequence upstream of light-regulated genes of monocots and dicots;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INRNTPSADB</td>
<td>2 0 2 2</td>
<td>Light-responsive transcription of psaDb depends on Inr, but not TATA box;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRONLOWER</td>
<td>1 1 0 1</td>
<td>Consensus sequence for plant introns;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1BOXATPDF1</td>
<td>1 2 0 1</td>
<td>Involved in L1 layer-specific expression;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Position</td>
<td>Score</td>
<td>Consensus</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
<td>-------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>LECPLEACS2</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>Core element in LeCp (tomato Cys protease) binding cis-element</td>
</tr>
<tr>
<td>LTREHVBLT49</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Low-temperature-responsive element in barley</td>
</tr>
<tr>
<td>LTREATLT178</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Putative low temperature responsive element (LTRE)</td>
</tr>
<tr>
<td>LTRECOREATCOR15</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Core of low temperature responsive element (LTRE) of cor15a gene in Arabidopsis</td>
</tr>
<tr>
<td>MARABOX1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>&quot;A-box&quot; found in SAR (scaffold attachment region; or matrix attachment region, MAR)</td>
</tr>
<tr>
<td>MARARS</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>&quot;ARS element&quot;; Motif found in SAR (MAR)</td>
</tr>
<tr>
<td>MARTBOX</td>
<td>16</td>
<td>3</td>
<td>16</td>
<td>&quot;T-Box&quot;; Motif found in SAR (scaffold attachment region; or MAR)</td>
</tr>
<tr>
<td>MYB1AT</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>MYB recognition site found in the promoters of the dehydration-responsive gene rd22</td>
</tr>
<tr>
<td>MYB1LEPR</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>Regulates defence-related gene expression via GCC box and non-GCC box cis elements</td>
</tr>
<tr>
<td>MYB26PS</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Myb26 binding site</td>
</tr>
<tr>
<td>MYB2AT</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Binding site for ATMYB2 (involved in regulation of genes responsive to water stress)</td>
</tr>
<tr>
<td>MYB2CONSENSUSAT</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>MYB recognition site of the dehydration-responsive gene rd22</td>
</tr>
<tr>
<td>MYBATRD22</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Binding site for MYB (ATMYB2) in dehydration-responsive gene, rd22</td>
</tr>
<tr>
<td>MYBCORE</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>Binding site of ATMYB2 responsive to water stress in Arabidopsis</td>
</tr>
<tr>
<td>MYBCOREATCYCB1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>&quot;Myb core&quot; activates reporter gene without leading to M-phase-specific expression</td>
</tr>
<tr>
<td>MYBGAHV</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>Central element of gibberellin (GA) response complex (GARC)</td>
</tr>
<tr>
<td>MYBPLANT</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>Plant MYB binding site</td>
</tr>
<tr>
<td><strong>MYBPZM</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>MYBST1</strong></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><strong>MYCATERD1</strong></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>MYCATERD22</strong></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>MYCCONSENSUSAT</strong></td>
<td>12</td>
<td>2</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td><strong>NAPINMOTIFBN</strong></td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td><strong>NODCON1GM</strong></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>NODCON2GM</strong></td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><strong>NTBBF1ARROLB</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>OSE1ROOTNODULE</strong></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>OSE2ROOTNODULE</strong></td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><strong>PIBS</strong></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>POLASIG1</strong></td>
<td>20</td>
<td>12</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td><strong>POLASIG2</strong></td>
<td>14</td>
<td>4</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td><strong>POLASIG3</strong></td>
<td>20</td>
<td>7</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>Sequence</td>
<td>Start</td>
<td>End</td>
<td>Length</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------</td>
<td>-----</td>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>POLLEN1LELAT52</td>
<td>11</td>
<td>5</td>
<td>11</td>
<td>Regulatory elements responsible for pollen specific activation of tomato</td>
</tr>
<tr>
<td>PREATPRODH</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>bZIP transcription factors function as activator for hypo osmolarity-inducible ProDH</td>
</tr>
<tr>
<td>PRECONSCRHSP70A</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Consensus sequence of plastid response element in the promoters of HSP70A</td>
</tr>
<tr>
<td>PYRIMIDINEBOXHEPB1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>&quot;Pyrimidine box&quot; found in the barley EPB-1 (cysteine proteinase) gene promoter</td>
</tr>
<tr>
<td>PYRIMIDINEBOXOSRAMY1A</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>Gibberellin-response cis-element</td>
</tr>
<tr>
<td>QELEMENTZMIZ13</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>Involved in expression enhancing activity</td>
</tr>
<tr>
<td>RAV1AAT</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Binding consensus sequence of Arabidopsis (A.t.) transcription factor, RAV1; expression level of RAV1 were relatively high in rosette leaves and roots</td>
</tr>
<tr>
<td>RBCSCONSENSUS</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>rbcS general consensus sequence</td>
</tr>
<tr>
<td>REALPHALGLHCB21</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Required for phytochrome regulation</td>
</tr>
<tr>
<td>REBETALGLHCB21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Required for phytochrome regulation</td>
</tr>
<tr>
<td>RHREPAEXPA7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&quot;Right part of RHEs (Root Hair-specific cis-Elements)&quot; Arabidopsis thaliana</td>
</tr>
<tr>
<td>ROOTMOTIFTAPOX1</td>
<td>35</td>
<td>21</td>
<td>36</td>
<td>Motif found both in promoters of rolD</td>
</tr>
<tr>
<td>RYREPEATBNNAPA</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>B2 domain of ABI3 is necessary for ABA-independent and ABA-dependent activation</td>
</tr>
<tr>
<td>RYREPEATGMGY2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>&quot;RY repeat motif (CATGCAT)&quot;</td>
</tr>
<tr>
<td>RYREPEATLEGUMINBOX</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>&quot;RY repeat” or legumin box found in seed-storage protein genes like soybean</td>
</tr>
<tr>
<td>S1FBOXSORPS1L21</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Might play a role in downregulating RPS1 and RPL21 promoter activity</td>
</tr>
<tr>
<td>S1FSORPL21</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>&quot;S1F binding site&quot; (&quot;S1 site&quot;) in spinach (S.o.) RPL21 gene</td>
</tr>
<tr>
<td>SEBFCONSSTPR10A</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Binding site of the potato silencing element binding</td>
</tr>
<tr>
<td>Motif Code</td>
<td>Logs2</td>
<td>Logs2</td>
<td>Pvals2</td>
<td>Pvals5</td>
</tr>
<tr>
<td>---------------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>SEF1MOTIF</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>SEF3MOTIFGM</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SEF4MOTIFGM7S</td>
<td>11</td>
<td>9</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>SORL1P1AT</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>SORL1P2AT</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SORLREP3AT</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SORLREP4AT</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SP8BFIBSP8BIB</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>SPCOREZMC1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>SREATATMSD</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>SURE1STPAT21</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SURECOREATSULTR11</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>SV40COREENHAN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TAAAGSTKST1</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>TATABOX2</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>TATABOX3</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>TATABOX4</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>TATABOX5</td>
<td>19</td>
<td>11</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Sequence</td>
<td>Count</td>
<td>Count</td>
<td>Count</td>
<td>Count</td>
</tr>
<tr>
<td>---------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>TATABOXOSPAL</td>
<td>9</td>
<td>5</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>TATAPVTRNALEU</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>TATCCAOSAMY</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TATCCAYMOTIFOSRAMY3D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TBOXATGAPB</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>TEOBOXATEEF1AA1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UP2ATMSD</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>WBBOXPCWRKY1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>WBOXATNPR1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>WBOXATNPR1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>WBOXHVISO1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>WBOXNTCHN48</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WBOXNTERF3</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>WRKY71OS</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>WUSATAg</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ZDNAFORMINGATCAB1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>