Function and Regulation of Xylem Cysteine Protease 1 and Xylem Cysteine Protease 2 in *Arabidopsis*

Ihab O. Ismail

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Horticulture

Dr. Eric P. Beers (Committee Chair)

Dr. John M. McDowell (Committee Member)

Dr. John L. Hess (Committee Member)

Dr. Jerzy Nowak (Committee Member)

Dr. Greg Welbaum (Committee Member)

May 17th, 2004

Blacksburg, Virginia

Copyright 2004, Ihab O. Ismail

Key Words: PTGS, auxin, tracheary element, auxin-response element, autolysis, IAA8

Function and Regulation of Xylem Cysteine Protease 1 and Xylem Cysteine Protease 2 in A*rabidopsis*

Ihab O. Ismail

Abstract

A functional water-conducting system, the tracheary elements of the xylem, is required to sustain plant growth and development. Tracheary element formation is dependent on many biological processes terminated by programmed cell death and cellular autolysis. The final two processes are probably dependent on the activity of hydrolytic enzymes such as XCP1 and XCP2 known to be expressed in tracheary elements during these final two processes. Thus, the transcriptional regulation of *XCP1* and the function of XCP2 were investigated. Qualitative and quantitative assessments of GUS activity as directed by various fragments of the *XCP1* promoter showed that a 237-bp internal region was able to drive GUS expression in a tracheary element-specific manner in *Arabidopsis*. A 25-bp deletion at the 3' end of this region abolished GUS expression. The 237-bp region served as bait in a yeast one-hybrid analysis. Screening of yeast colonies retrieved 109 putative positive interactions, which included a potential transcriptional regulator, indole acetic acid-induced protein 8 (IAA8). An auxin responsive element that potentially binds auxin responsive transcription factors was found within the 25-bp deletion. C*is*-elements were predicted by Genomatix and Athamap computer programs. The *cis*-elements form pyrimidine and gibberellic acid responsive elements that can potentially bind Dof and Myb transcription factors, respectively. In an independent effort, attempts to develop a mapping population to isolate upstream regulators of XCP1 expression did not succeed. Functionally, tracheary element-specific expression of *XCP2* in *Arabidopsis* suggested a specialized role for XCP2 in final phases of tracheary element differentiation. The function of XCP2 was assessed using T-DNA insertional mutants, post-transcriptional gene silencing, and through tracheary element-specific expression of the cysteine protease inhibitor, soyacystatin N in *Arabidopsis*. Our findings revealed that the absence of XCP2 expression due to T-DNA insertional mutagenesis did not affect plant growth and development in the laboratory. Soyacystatin N was an effective *in vitro* inhibitor of cysteine proteases. Plants expressing *35S*-driven cytosolic form of soyacystatin exhibited stunting and reduced apical dominance. Plants expressing *pXCP1*-driven cytosolic soyacystatin did not differ from wild type plants. Additionally, transgenic plants expressing *pXCP1*- and *35S*-directed *XCP2*-double-stranded RNA for the silencing of *XCP2* showed no unusual phenotypes compared to their wild type counterparts.

Personal Dedication

To My Wife Kerlly, My air

This is for you

Before I saw you

My words were mere mumbles

My presence was scattered

In your company

My soul is filled with Gratitude

My ego is humbled

My self is gathered

To My Nile

Whose waters cleanse the soils of my imperfection

Whose smile is my joy

Whose dream is my home

To my parents, Omar and Faida, my unending roots

To my brother Rami

To my sisters Suvana, Jumana, and Hanadi

I look for the opportunity to enrich your future

As you already have enriched mine with your memories

So know that you are

Loved, Respected, Thought of

Every day

Forever

Ihab Ismail

05/17/2004

إهداءٌ خاص

إلىَ زَوْجتي كِيرْلي،هَوَائيْ

قبل معرفتيِ فيكِ

كلماتيَ كانتْ مجرّدَ حروفٍ

ووجوديَ للوحدةِ أعَزَّ صديقْ

بوجود كِ

رُوحيَ إمتلأتْ بالعِرْفانِ وتواضعت نفسي

و معنىً أصبحَ لوجوديْ

إلىإبنيَ نائلْ،عُمْرِي

َمن غسلتْ كلمائُهُ ترا بَ نقْصيَ

مَنِ إبتسامتُهُ حلاوةُ يومي

وحلمُهُ مسكنيْ

إلىَ والديَّ العزيزين

عُمَر و فائدةْ

إلىَأخي راميْ

إلى أخواتيَ، سوفانهْ، جمانهْ، وهناديْ

أنظرإلى فرصةٍ في الغدِ لأكونَ جُزءاً من حياتكمْ

يا جذوريَ الّتي لا تذبلْ

أنتم بالفكرْ، بالقلبْ، و بالذكرْ

في كلِّ يومٍ

للأبدْ

إيهابعمر إسماعيل

١٧ \٥ \ ٢٠٠٤

Professional Acknowledgement

I would like to thank my advisor Dr. Eric P. Beers for his scientific guidance in my Ph.D. projects and his personal support in challenging situations through which I passed during my Ph.D. program. I thank him for awarding me the research assistantship through his NSF-funded research. The preparation of this dissertation would have not been possible without his critical reading and review. I wish him and his lab productive and fruitful research projects in the years to come. Finally, I hope that my various investigations will be complemented by others to produce a beneficial contribution that is shared at an international level in the very near future.

I would like to thank the department head and committee member, Dr. Jerzy E. Nowak for his advice in my research and his administrative efforts to improve the department at many levels. I extend special thanks to Dr. John L. Hess whose constant encouragement and support have been integral to my personal growth and scientific progress here at Virginia Tech. I thank Dr. Greg Welbaum for his support and Dr. John M. McDowell for his collaboration in my research. I also would like to thank Dr. Glenda Gillaspy and her lab members, and Dr. Brenda Winkel and her lab members for being always ready to share their knowledge and expertise with me.

My progress in research was possible with the help of many scientists with whom I communicated by phone or on-line or met in person. I am especially grateful to have Dr. B. Smith of Clontech as a rich resource in yeast analysis. It was his dedicated assistance over a 5-month period that enabled me to pull through my first-time effort in using yeast one-hybrid genetic screen. I thank the Virginia Bioinformatics Core Facility Laboratory in Virginia Tech and Davis Sequencing Facility in University of California at Davis for their high quality sequencing services. I thank Earl Petzold for his technical assistance through out my research. I thank my colleague Chengsong Zhao for sharing his expertise in molecular biology.

Finally, I thank all the undergraduate students with whom I interacted and who provided me with the opportunity to teach them and advise them on their research.

Table of Contents

Table of Contents

List of Figures

List of Figures

List of Tables

List of Supplemental Figures and Tables

List of Abbreviations and Definitions

Abbreviation	Definition
A	
aa	Amino acid
AFLP	Amplified Fragment Length Polymorphism
APS	Ammonium persulfate
ARE	Auxin responsive element
ARF	Auxin responsive factor
3-AT	3-amino-1,2,4-triazole, an inhibitor of leaky HIS3 expression in yeast
B	
bar	Phosphinothricin acetyltransferase
bar$^+$	Resistant to herbicide Finale (Basta equivalent)
BCA	Bicinchoninic acid
BLAST	Basic Local Alignment Search Tool
C	
cDNA	Complementary DNA
CTAB	Cetyl-trimethyl-ammonium bromide
cys N	Soyacystatin N
D	
df	Degrees of freedom
DOF	DNA one finger, a transcription factor
ds cDNA	Double-stranded complementary DNA
E	
EDTA	Ethylenediamine-tetraacetic acid
E-64	[*trans*-epoxysuccinyl-L-leucylamido-(-guanidino) butane]
F	
Finale	Glufosinate-Ammonium: botanoic acid, 2-amino-4-hydroxymethylphosphiny
G	
GA	gibberellins
gDNA	Genomic DNA
GUS	*β*-glucuronidase reporter protein
GUS	*β*-glucuronidase reporter gene
I	
IPTG	Isopropyl-*β*-D-thiogalactopyranoside
L	
LjXCP1	A putative ortholog of XCP1 in the legume *Lotus japonicus*

L (cont'd)

LPP	Leupeptin, a peptide-aldehyde reversible inhibitor of cysteine proteases (Acetyl -leucyl-leucyl-arginal)

M

MALDI	Matrix-assisted laser desorption/ionization
MXCP2	Mature protein of xylem cysteine protease 2
^{M}XCP2^{ab}	Antibody raised against the mature protein of xylem cysteine protease 2
MTF	Myb transcription factor

N

N	Number of plants tested
NCBI	The National Center for Biotechnology Information
ng	Nanogram
Ni-NTA	Nickle-nitrilotriacetic acid
NptII	Neomycin phosphotransferase II (for conferring resistance to kanamycin)
nt	Nucleotides

O

ORF	Open reading frame

P

p	Probability value
PAGE	Polyacrylamide gel electrophoresis
pCB302	A minibinary vector for plant transformation
pET-cys^{his}	Soyacystatin N-6x C-terminal HIS tagged ORF in the expression vector *pET28a*
pFGC5941	A binary vector for use in gene silencing experiments
pGEM	Cloning vector pGEMT® Eazy from Promega
pHIS2- Δ^{237}	pHIS2 containing the bidirectional deletion Δ^{237} as a bait in yeast one-hybrid
pIAA8	A putative promoter region of IAA8
pLjXCP1	A putative promoter region of LjXCP1
^{PR}XCP2	Prodomain of xylem cysteine protease 2
^{PR}XCP2^{ab}	Antibody raised against the prodomain of XCP2
^{PRM}XCP2^{ab}	Antibody raised against the pro and mature domain of XCP2
^{PRM}XCP2	Prodomain and mature domain of XCP2
PSK-α	Sulfated pentapeptide H-Tyr (SO_3)-Ile-Tyr(SO_3)-Thr-Gln-OH
p35S::cys^{his}	Soyacystatin N-6x C-terminal HIS tagged ORF driven by *p35S* promoter
p35S::GUS^{ir}	*GUS* inverted repeats driven by full length *35S* promoter
p35S::XCP2^{ir}	*XCP2* inverted repeats driven by full length *35S* promoter
PTGS	Post-transcriptional gene silencing
pXCP1::GUS^{ir}	*GUS* inverted repeats driven by full length x*ylem cysteine protease 1* promoter
pXCP1::XCP2^{ir}	*XCP2* inverted repeats driven by full length x*ylem cysteine protease 1* promoter
pXCP1::cys^{his}	Soyacystatin N-6x C-terminal HIS tagged ORF driven by *pXCP1* promoter
pXSP1::GUS	GUS coding region driven by x*ylem serine protease 1* promoter

R

RAPD	Randomly Amplified Polymorphic DNA

R (cont'd)

RFLP	Restriction Fragment Length Polymorphism
RNAi	RNA interference; a general term for gene silencing by different mechanisms
RT-PCR	Reverse transcription PCR

S

SDS	Sodium dodecyl (lauryl) sulfate
SD/-aa	Synthetically designed yeast media lacking a specific amino acid
SSLP	Simple Sequence Length Polymorphism
SSR	Simple Sequence Repeat

T

T_0	Plants treated with *Agrobacterium* culture
T_1	Plants produced from selfing of one *Agrobacterium*-treated plant
T_1 WT [A]	T_1 generation of WT [A] transformation
T_1 **[**T_1 WT [A]**]**[B]	T_1 generation produced from a sequential double transformation by binary vectors carrying construct A and B
T_{1P}	Plants produced from selfing of pool of *Agrobacterium*-treated plant
T_2	Transformed plants produced from selfing of a selected T_1 plant
T_{2P}	Transformed plants produced from selfing of a pool of T_1 plants
T-DNA	A transferred DNA fragment that can be integrated into the nuclear genome
TEs	Tracheary element(s)
TEMED	N,N,N',N'-Tetramethylethylenediamine

V

VBI	Virginia Bioinformatic Institute

W

WT	*Arabidopsis thaliana* ecotype Columbia
WT [A]	Wild type plants transformed with binary vector carrying construct A

X

χ^2	Chi-square statistical analysis
X-gluc	Cyclohexylammonium salt of glucuronide
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
X PRA/PRB	PCR reaction using X as a template and Primers A and B for amplification
X [Y]	X is transformed with Y

Y

YPDA	Yeast media containing all amino acids for optimal growth

Symbols

5'Δ^n	5' unidirectional deletion consisting of n nucleotides or base pairs
Δ^n	5'-3' bidirectional deletion consisting of n nucleotides or basepairs
IN Δ^n	An internal deletion consisting of n nucleotides or base pairs
6x HIS	A polypeptide consisting of 6 histidine amino acid residues

Kits and Reagents

Kit/Reagent	Source	Used for/as a
Avid®	Syngenta-Greensboro, NC	Miticide
BCA	Sigma- St Louis, MO	Protein quantification
BD Matchmaker™ Library Construction	D Biosciences- Palo Alto,CA	One-hybrid Screen
Cells-to-cDNA™ II kit	Ambion-Austin, TX	cDNA synthesis
Conserve®	DowAgro-Indianapolis, IN	Insecticide
Finale™	Scotts Company- Marysville, Ohio	Herbicide
Nulceospin® Extraction kit	BD Biosciences- Palo Alto,CA	Clean DNA extraction from PCR
NucleoTrap® Gel Extraction	BD Biosciences- Palo Alto, CA	Clean DNA extraction from Gel
100-bp Ladder;λ *Eco*RI-*Hind*III	Promega-Madison,WI	DNA standards
pGEM-T Eazy	Promega- Madison, WI	TA-cloning
PfuUltra™ High-Fidelity DNA Polymerase	Strategene-La Jolla, CA	Inverse long (> 3Kb) PCR DNA amplification
Quantum Prep® Freeze 'N Squeeze DNA Gel Extraction	Bio.Rad-Hercules, CA	Gel-extraction of DNA
QIA Plant® DNA Mini-kit	Qiagen- Valencia, CA	DNA prep
QIA prep® spin	Qiagen Sciences- Valencia, CA	DNA miniprep
RETROscript™	Ambion-Austin, TX	cDNA synthesis
RNase-H	Sigma- St Louis, MO	Removing RNA from DNA preparations
RNeasy® Plant Mini-kit,	Qiagen Inc, Valencia, CA	RNA miniprep
Silwet L-77	Lehle Seeds- Round Rock, TX	Wetting Agent
TOPO® TA Expression Kit	Invitrogen-San Diego, CA	Protein Purification
Wizard® plus SV	Promega- Madison, WI	DNA miniprep
X-Glu	GOLD Biotech, Inc- St Louis, MO	GUS substrate for histochemical staining

On-line Molecular Tools

ABRC	http://www.biosci.ohio-state.edu/~plantbio/Facilities/abrc/abrchome.htm
AGRIS	http://*arabidopsis*.med.ohio-state.edu/AtTFDB/index.jsp
Ambion	http://www.ambion.com/
AthaMap	http://www.athamap.de/
Bio-Rad	www.bio-rad.com
Clontech	http://www.bdbiosciences.com/clontech/
Clustal W program	http://www.ebi.ac.uk/clustalw/
Cocalico Biologicals Inc.	http://www.cocalicobiologicals.com/
Davis Sequencing	www.davissequencing.com
DNA star	http://www.dnastar.com/
Genomatix	www.genomatix.de
MEROPS	http://merops.sanger.ac.uk/
NCBI	http://www.ncbi.nlm.nih.gov/
PepIdent	http://www.expasy.org/tools/peptident.html
PLACE	http://www.dna.affrc.go.jp/PLACE/
Promega	www.promega.com
Qiagen	www.qiagen.com
RNA interference vectors	http://www.chromdb.org/plasmids/vectors2.html
Sigma	http://www.sigmaaldrich.com/
SIGNAL	http://signal.salk.edu/cgi-bin/tdnaexpress
Stratagene	www.stratagene.com
TAIR	www.*arabidopsis*.org
TIGR	http://www.tigr.org/
Virginia Bioinformatics Inst.	http://www.vbi.vt.edu

Chapter I

Function of *Arabidopsis* Cysteine Proteases with Emphasis on XCP 2

Abstract

Papain-like proteases are members of the C1A family of cysteine proteases. Xylem cysteine protease 2 (XCP2) is a putative papain ortholog. Through specific and non-specific proteolytic activity, cysteine proteases have been implicated in plant defense against insects and pathogens, intracellular protein turnover, heterochromatin degradation, and tracheary element differentiation. Cells contain cysteine protease inhibitors that regulate the activity of cysteine proteases at the post-translational level through the formation of reversible equimolar protease-inhibitor complexes. Tracheary element-specific expression of *XCP2* evidenced using *pXCP2::GUS* reporter constructs in *Arabidopsis* suggested a specialized role for XCP2 in final phases of tracheary element differentiation. The function of XCP2 in plant growth and development was assessed using T-DNA insertional mutants, post-transcriptional gene silencing, and through tracheary element-specific expression of the cysteine protease inhibitor, soyacystatin N in *Arabidopsis*. Our findings reveal that the absence of XCP2 expression due to T-DNA insertional mutagenesis does not affect plant growth and development in the laboratory. Plants expressing *p35S*-driven cytosolic form of soyacystatin exhibited stunted and reduced apical dominance. This phenotype was correlated with the presence of soyacystatin N by RT-PCR. The ability of soyacystatin to inhibit tracheary element proteases *in vivo* is under investigation. Plants expressing *pXCP1*-driven cytosolic soyacystatin did not differ from wild type plants. Soyacystatin N was an effective *in vitro* inhibitor of cysteine proteases. Additionally, transgenic plants expressing *pXCP1*- and *p35S*-directed *XCP2*-double-stranded RNA for the silencing of *XCP2* showed no unusual phenotypes compared to their wild type counterparts.

Key Words: xylem, cysteine protease, PTGS, soyacystatin N, T-DNA

Introduction

Proteases are categorized according to their catalytic site amino acid residues into serine (EC 3.4.21), cysteine (EC 3.4.22), aspartic (EC 3.4.23), and metallo proteases (EC 3.4.24) (for reviews see Beers et al., 2004; Grzonka et al., 2001). Aside from routine protein turnover, papain-like lysosomal cysteine proteases (family C1A; MEROPS peptidase database, http://merops.sanger.ac.uk/) play roles in disease mitigation, for example, in rheumatoid arthritis and muscular dystrophy, which are human diseases caused by excess proteolysis and regulated through natural or synthetic cysteine protease inhibitors (Turk et al., 2000; 2003). Functionally, this shift of cysteine proteases from catalysis of non-specific proteolysis to specific and limited intra- and extra-cellular proteolysis appears to have been facilitated in part by evolution of gene expression pattern (Nägler and Ménard, 2003). Removal of intracellular content in differentiating TEs in *zinnia* suspension culture was partially blocked by cysteine protease inhibitors suggesting that cysteine proteases are necessary for autolysis (Woffenden et al., 1998). The papain-like vacuolar-localized cysteine proteases and their mammalian lysosomal-localized counterparts share structural and functional features. First, they share three highly conserved amino acids required for catalysis, cysteine Cys^{25}, histidine His^{159}, and asparagine Asn^{175} (Wiederanders et al., 2003), using the papain numbering system. Second, they are zymogens consisting of three domains known as the predomain, prodomain, and the mature domain. The N-terminal predomain (usually 10-20 aa) is the signal peptide required for translocation to ER lumen. The prodomain (38-250 aa) contains signals for translocation through the secretory pathway. The mature protein (220-260 aa) is characterized by two domains, L and R, of comparable sizes surrounding the active site cleft. And thirdly, the prodomain functions as a specific inhibitor of its corresponding mature protein. The regulatory role of the prodomain in inhibiting the activity of its cognate mature protein is of practical and economic significance especially in pest-plant interaction. There are 30 C1A proteases in *Arabidopsis*. The focus here is on two tracheary element specific cysteine proteases, XCP1 and XCP2.

Cysteine proteases have been implicated in diverse plant processes. These include the hypersensitive response (Belenghi et al., 2003), protein turnover in germinating seeds (Cercós et al., 1999; Taneyama et al., 2001; Sutoh and Yamauchi, 2003), senescence of anthers (Xu and Chye, 1999) and leaves (Weaver et al., 1998) and petals (Wagstaff et al., 2002), and tracheary element formation (Woffenden et al., 1998; Zhao et al., 2000; Funk et al., 2002). Cysteine proteases were

also implicated in defense mechanisms against pests. Papain in latex exudates was found to be crucial in the defense mechanism of papaya (*Carica papaya*) against the larvae of the oligophagous pest *Samia ricini* and two other polyphagous pests *Mamestra brassicae* and *Spodoptera litura* (Konno et al., 2004). Konno and colleagues (2004) also found that other cysteine proteases present in latex exudates are highly toxic to the larvae of those pests. E-64, a low-molecular-mass inhibitor of cysteine proteases, also abolished latex toxicity (Konno et al., 2004). The beetle cowpea bruchid, *Callosobruchus maculates*, utilizes Cathepsin L-like cysteine proteases to invade the seeds of cowpea *Vigna unguiculata* (Moon et al., 2004). The application of dietary soyacystatin N, a proteinaceous cysteine protease inhibitor, in cowpea bruchid feeding bioassays initially led to an increase in insect mortality rate. However, increased mortaility was followed by a normal growth pattern suggesting that the cowpea bruchid utilized a defense strategy to overcome inhibition by soyacystatin N (Zhu-Salzman et al., 2003). In other feeding bioassay studies, Koiwa and colleagues (2000) showed that soyacystatin N was effective in inhibiting growth and development of the rootworm *Diabrotica vergifera vergifera*. Although engineering resistance based on cysteine protease inhibition has been attempted (Hegedus et al., 2003) with rice (Duan et al., 1996), tobacco (Johnson et al., 1989), cotton (Thomas et al., 1995), alfalfa (Thomas et al., 1994), strawberry (Graham et al., 1995), and poplar (Klopfensteine et al., 1993), insects were able to counteract the inhibitory effects of cystatins. The nature of pathogen-plant interaction is reflective of the antagonism between pest proteases and inhibitors and plant proteases and inhibitors.

At the biochemical level, cysteine protease inhibitors act on cysteine proteases by trapping them in a reversible tight equimolar complex (Barrett et al., 1998; Barrett et al., 2001; Belenghi et al., 2003). The prodomains are natural and selective inhibitors of their cognate mature protein domain. They also insure proper protein folding and stabilization in alkaline and neutral conditions (Yamamoto et al., 2002). Amino acids in the prodomain mediate the transport of the protease to the lysosomes (mammals) or vacuoles (plants). Although it is possible to design synthetic peptides homologous to prodomain sequences, establishing inhibitory interaction *in vivo* is likely to be challenging due to requirements of colocalization with cognate proteases and stabilization of synthetic prodomains which may have inherently short half-lives.

The presence, in plants, of large multi gene cysteine proteases families suggests the presence of functional redundancy (Nägler and Ménard, 2003). For example, storage proteins in the germinating seed provide a source of reduced nitrogen for the germinating seedling. The proteolytic processes by which these proteins are remobilized seem to be highly specialized; for example, there are specialized compartments named PSVs (protein storage vacuoles), specific processing mechanisms, conserved storage protein types, and specific seed-type vacuolar processing cysteine proteases, *VPE*s (Muntz, 1998). Gruis and colleagues (2002) showed that the transposon-insertional double mutant, deficient in two major VPEs, *δVPE* and *βVPE*, still showed over 80% protein-processing activity compared to wild-type plants, suggesting the presence of alternative and redundant processing activities in the seed. Subsequently, Gruis and colleagues (2004) discovered two additional vegetative VPEs, *γVPE* and *αVPE* and found that in a quadruple mutant deficient in all four *VPE*s exhibited lowered processing activity and accumulation of storage proteins. From a genetic perspective, therefore, genes of VPEs behaved as quantitative trait loci where only the collective effort impacted the overall phenotype.

The 82% identity that the XCP1 and XCP2 prodomains share with the papain prodomain suggests that papain, XCP1 and XCP2 are subject to similar trafficking and regulation (Funk et al., 2002) and perhaps perform analogous functions. The recent discovery that papain is toxic to insects and served as a deterrent to herbivory suggest that XCP1 and XCP2 may also serve in defense against insects. Papain-like enzymes are also phytotoxic as overexpression of XCP1 resulted in stunting, premature senescence and reduced fertility (Funk et al., 2002). Additionally, peptidase inhibitor studies implicated cysteine proteases as essential catalysts in TE autolysis (Woffenden et al., 1998; McCann et al., 2000; Obara and Fukuda, 2004). The ectopic expression of an endogenous cysteine protease inhibitor inihibited oxidative- and pathogen-triggered programmed cell death, two processes shown to require activation of cysteine proteases in soybean cell culture (Solomon et al., 1999). That upregulation of cysteine proteases was observed in germinating seeds implicated cysteine proteases in protein turnover. Thus, functions for XCP1 and XCP2 in regulation of TE formation, nitrogen remobilization, and defense should all be considered. As with seed VPEs, however, the presence of TE-specific cysteine proteases in addition to XCP1 and XCP2 can not at present be ruled out and may confound efforts to determine specific functions of XCP1 and XCP2.

Materials and Methods

A1.1. XCP 2 Nucleotide and Amino-acid Sequences

The nucleotide sequence and primary amino acid sequence for *Arabidopsis* XCP2 and the sequence of the putative promoter region was obtained from The *Arabidopsis* Information Resource (TAIR; www.*arabidopsis*.org). Information on the tracheary element specificity of *XCP2* transcripts and *pXCP2::GUS* expression was reported in Zhao et al. (2000) and Funk et al. (2002), respectively. Phylogenetic trees showing the relationships between XCP2 and other other cysteine proteases were produced using Clustal W program (http://www.ebi.ac.uk/clustalw/; see also Thompson et al., 1994).

A1.2. Plant Material

Wild Type *Arabidopsis thaliana* ecotype Columbia was used as the genetic background for all studies discussed here. Three T_3 SALK T-DNA collection seed samples (Alonso et al., 2003) were obtained from *Arabidopsis* Biological Resource Center (ABRC) at the Ohio State University, Columbus, Ohio. These are SALK_001772, SALK_010938, and SALK_057921. Background information on SALK lines can be obtained at http://signal.salk.edu/tdna_FAQs.html. pROK2 binary vector used for insertional mutagenesis of SALK lines (Baulcombe et al., 1986) contains *NptII* gene conferring resistance to kanamycin antibiotic. It was cautioned, however, that *NptII* is silenced in many of SALK lines (see Arciello et al., 2003). Another T-DNA line, Garlic/SAIL_1261_A12, was obtained from Torrey Mesa Research Institute whose T-DNA lines are currently being integrated into ABRC collection. For the Garlic/SAIL line, the binary vector pDAP101 was used to deliver the T-DNA into *Arabidopsis* genome. pDAP101 contains *bar* gene which is used as an herbicide selectable marker to select transformed plants. T_{2P} WT [*pXSP1::GUS*] Lines were generated in our laboratory. The T_2 WT [*p35S::XCP1*] line (Funk et al., 2002) is a XCP1 overexpressor used for testing efficacy of recombinant soyacystatin N.

A1.3. Binary vectors and cDNA Library

The RNAi (ds RNA) vector pFGC5941 was obtained from ChromDB (http://www.chromdb.org/plasmids/pFGC5941.html). The mini-binary vector pCB302 (accession # AF139061) was kindly provided by Dr. David J. Oliver (Xiang et al., 1999). The Expression Vector *pET-cys*^{his} was kindly provided by Dr. Hisasha Koiwa. The cDNA library constructed from the

Arabidopsis xylem (Zhao et al., 2000) was used as a template for RT-PCR and ds RNA vector construction.

A1.4. Integration of T-DNA into *Arabidopsis* Nuclear Genome

The DNA fragments to be analyzed *in planta* were obtained using standard recombinant DNA techniques, transferred to a suitable intermediate vector, and then cloned into its final recipient binary vector. Details regarding specific constructs are provided below. Final DNA integration into plants was mediated by *Agrobacterium tumefaciens* strain GV3101.

A1.4.1. Transformation of *Escherichia coli* strain DH5α cells and *Agrobacterium tumefaciense* strain GV3101 with plasmid DNA

Preparation and heat-shock transformation of competent DH5α cells were conducted as described in Sambrook et al. (1989). Preparation of *Agrobacterium* competent cells was done according to TAIR protocols (http://www.*arabidopsis*.org/info/). Freeze-thaw transformation of *Agrobacterium* with plasmid DNA was done according to Chen et al. (1994).

A1.4.2. *Arabidopsis* Preparation, Transformation and Selection of T_1 Transgenic Plants

Seeds of *Arabidopsis thaliana* were sown on 4"x 4" pots containing Sunshine potting media [Sun Grow Horticulture Distribution Incorporated, Bellewest, WA] covered with thin layer of moist vermiculite [Shundler, Metuchen, NJ]. Plant number per pot was between 10-15 plants. Pots were covered with a thin plastic film, vernalized at 4^{o}C for 4 days in the dark, and placed in 25^{o}C away from direct light until germination was observed. Pots were uncovered and placed under continuous light with intensity ranging from 16-20 µmol sec^{-1} m^{-1} [LI-250 meter with Quantum sensor, LI-Cor]. Seedlings were fertilized weekly as described in Beers et al. (1992). CONSERVE and AVID were used when necessary to control insects and spider mites, respectively, according to the suppliers' recommendations. A floral dip method was used for all plant transformation experiments (Clough and Bent, 1998) which was modified from Bechtold et al. (1993). Selection of transformed plants was done by spraying 10-day-old and 21-day-old T_1 plants with Finale® Herbicide (0.03% final concentration of active ingredient Glufosinate-ammonium).

A1.5. Rapid Genomic DNA extraction and Sequencing Reactions

Genomic DNA was extracted from the two oldest rosette leaves (four-leaf stage) or 3-4 floral bud clusters using the rapid CTAB method described by Murray and Thompson (1980). The only modification was the addition of 3 µl 10 mg/ml RNase-H after the tissue homogenization step to remove contaminating RNAs. DNA was quantified by spectrophotometer and its quality was assessed by running a 10-µl aliquot on an ethidium-bromide impregnated agarose gel. Between 100 and 500 ng of DNA was normally suitable for downstream PCR amplification. Sequencing of inserts was conducted by the VBI core lab and University of California Davis Sequencing facility, Davis, CA (www.davissequencing.com).

A1.6. Genotyping and Sequencing of T-DNA/gDNA Junction

A1.6.1. Garlic/SAIL 1261-A12

PR13 (TTACCTCTGACCCAGACAGA) and PR14 (GGCAAGAGCAAAACAGAGGA) were designed to flank the putative T-DNA location. PR15 (TTCATAACCAATCTCGATACAC) was specific to left-border T-DNA of pDAP101 binary vector. PCR reactions were conducted using the primer pairs PR13/PR14, PR13/PR15, PR14/PR15 on gDNA of herbicide resistant plants. The PCR product amplified by PR14/PR15 was TA-cloned into the pGEM to construct plasmid P1010. Sequencing of that plasmid was conducted using T7 and SP6 sequencing primers endogenous to pGEM.

A1.6.2. SALK_ 010938

Primers PR17 (AGACACTGTAAGATGCTGATCA) and PR18 (TGTGACAGGAAACTTGA-CAACAT) were designed to flank the putative T-DNA location. PR16 (GCGTGGACCGCTTGC-TGCAACT) was specific to left border T-DNA. PCR primer, PR19 (TCAAGATCAACCCCGCACCGC) was more compatible to PR16 than PR18. PCR reactions were conducted using the following primer pairs PR17/PR18, PR16/PR19 on gDNA of three plants of this line, designated S010938-1 through S010938-3. PCR product from PR16/PR19 reaction was TA-cloned into pGEM to construct plasmid P1013. Sequencing was conducted using SP6 and T7 sequencing primers endogenous to pGEM and PR19 endogenous to XCP2.

A1.6.3. SALK_057921 and SALK_001772

SALK_057921 PCR reactions were conducted using the primer pair PR13/PR16 on pooled gDNA from six plants. The PCR product was TA-cloned into pGEM to construct plasmid P1018. Sequencing was conducted using T7 sequencing primer endogenous to pGEM. SALK_001772 PCR reactions were conducted using the primer pairs PR18/PR17and PR18/PR16.

A1.7. Phenotypic Evaluation of T-DNA Lines

Plants grown from SALK and Garlic/SAIL seeds were compared to WT in soil and *in vitro*. Evaluations were repeated two to three times under similar conditions to see if phenotypes were reproducible. Root morphology of T_5 SALK_010938-2 seedlings was observed *in vitro* under different physiological stresses. Optimal media was prepared according to López-Bucio et al. (2002; 2003) and Linkohr et al. (2002) with necessary modification to prepare starvation media lacking nutrients to be tested.

A1.8. Development of XCP2 Antibodies and Immunoblot Analysis of The homozygous T_4 SALK_010938-2

Sequences corresponding to three regions of XCP2, PRXCP2, MXCP2, and PRMXCP2 were selected to develop PRXCP2^{ab}, MXCP2^{ab}, and PRMXCP2^{ab}, respectively. Using WT xylem cDNA library (Zhao et al., 2000) each region was amplified and cloned into pTrcHis expression vector (TOPO® TA Expression Kit). The PRXCP2 region was amplified by PR60 (TACTCCATCGTTGGATACTCC) and PR61 (AGCGAACTCTGCGTAAGATCTT). The MXCP2 region was amplified by PR62 (GACTATGCCTTTGAGTACATT) and PR63 (AAGACCAA-CCCCGCACCGC). The PRMXCP2 region was amplified by PR60 and PR63. Purification of proteins was carried out following the Qiagen Ni-NTA affinity purification manual under denaturing conditions. The purified proteins were resolved using SDS-PAGE and the bands corresponding to the expected molecular mass of the immunoreactive protein were excised, digested in-gel with modified trypsin and the trypsinized polypeptide was analyzed by MALDI-MS. XCP2 identity was confirmed by searching PepIdent (http://www.expasy.org/tools/peptident.html) program. Corresponding positively-identified polypeptides were sent to Cocalico Biologicals Inc, Reamstown, PA (http://www.cocalicobiologicals.com/) for antibody production in rabbits.

Fifty mg of T_4 SALK_010938-2 underground or above ground tissue was homogenized in protein extraction buffer [100 mM Na_2HPO_4 pH 7.2, 7 mM β-merceptaethanol] on ice using a 4:1, buffer:tissue (vol:weight). The difficult-to-homogenize woody tissues were powdered in liquid nitrogen prior to homogenization in the buffer. Quantification of proteins was conducted using the BCA assay according to Sigma's instruction manual. Immunoblots were prepared according to Woffenden et al. (1998) using PRXCP2^{ab}, MXCP2^{ab}, or PRMXCP2^{ab} as indicated in the figure legends.

A1.9. Post-transcriptional Gene Silencing

A1.9.1. Histochemical Analysis of GUS Activity

One-week-old roots and/or newly-emerged true leaves were submerged in GUS histochemical staining buffer (100 mM Na_2HPO_4 pH 7.0, 10 mM EDTA pH 8.0, 0.5 mM $K_4Fe(CN)_6$, 0.5 mM $K_3Fe(CN)_6$, 0.1% Triton X-100, 1 mM filter-sterilized X-Glu). Contents were vacuum infiltrated for 10 min at room temperature and incubated at 37^{o}C until blue stain was observed. Different staining buffers were used (Stomp, 1992; Malamy and Benfey, 1997; Oono et al., 1998). They differ in their ability to minimize leakage and/or intensity of GUS expression thus better reflecting the cell-type localization of GUS expression.

A1.9.2. Construction *of pXCP1::GUS* ^{ir} and *p35S::GUS* ^{ir}

PCR primers PR22 (TCTAGAGGCGCGCCCACCGTGGTGACGCATGT) and PR23 (GGATCCATTTAAATGCCTCTTCGCTGTACAGT) were designed to amplify a fragment in *GUS* coding region using pBI121 (Jefferson et al., 1987) as a DNA template. The PCR product was TA-cloned into pGEM to construct plasmid P1014. An *Asc*I-*Swa*I restriction digest of P1014 DNA released a 'sense' *GUS* fragment which was cloned into compatibly digested pFGC5941 to construct P1017. A *Bam*HI-*Xba*I restriction digest of P1014 released an 'antisense' *GUS* fragment which was cloned into compatibly digested P1017 to construct P1020 (i.e. *p35S::GUS* ^{ir}). pFGC5941 was also modified to include *pXCP1* instead of *p35S*. The *pXCP1* was released by an *Eco*RI-*Nco*I digest from a previously made cloning vector and was subcloned into compatibly digested pFGC5941 to construct P1001. The *GUS* sense and antisense fragments were released as described earlier from P1014. The *GUS* 'sense' fragment was cloned into P1001 to construct P1016. The *GUS* 'antisense' fragment was cloned into P1016 to construct P1021 (i.e. *pXCP1::GUS* ^{ir}).

A1.9.3. Construction of *pXCP1::XCP2* ^{ir} and *p35S::XCP2* ^{ir}

The pFGC5941 binary vector containing two 513-bp inverted repeats of of the XCP2 coding region was previously constructed (*p35S::XCP2*^{ir}) using PR59 (TCTAGAGGCGCGCCGACA-GGAAACTTGACAACATTGTC) and PR 58 (CTGGATCCATTTAAATGCCATCTTGTTG-ATTCCACAG). The 513-bp internal fragment was selected for maximum identity shared with *XCP1* so that simultaneous silencing of both XCP1 and XCP2 would be accomplished. The 2.3 kb fragment containing the two inverted repeats and CHSA intron was released as an *Xba*I-*Nco*I fragment and was subcloned into P1001 to construct P1002 (i.e. *pXCP1::XCP2*^{ir}). Both binary vectors were used to transform WT plants. Sequencing of binary vectors was conducted at the Davis Sequencing facility using CHSA specific primers PR09 (CTTTCTACCTTCCCACAATTCG) and PR10 (TACTTACACTTGCCTTGGAGT).

A1.9.4. Evaluation of GUS Silencing Using pFGC5941

Histochemical staining of GUS expression was conducted on all *bar*$^{+}$ T_1 plants that were developed from the T_{2P} *pXSP1::GUS* [*p35S::GUS* ^{ir}] and T_{2P} *pXSP1::GUS* [*pXCP1::GUS* ^{ir}]. The staining results were compared before and after transformation to assess the efficiency of silencing. WT [*pXCP1::XCP2* ^{ir}] and WT [*p35S::XCP2* ^{ir}] *bar*$^{+}$ T_1 plants were evaluated by initially comparing them to non-transformed WT plants.

A1.10. Studies on Soyacystatin N in *Arabidopsis*

A1.10.1. Purification of Soyacystatin N

The *pET-cys*^{his} (Koiwa et al., 1998) was transformed into *E. coli* DH5α cells. A 250 ml LB culture of bacterial cells was prepared. Expression was induced using 1 mM IPTG and growth was allowed to continue for 4 h at 37^{o}C after which cells were harvested by centrifugation at 4000 x*g* for 20 min. The supernatant was discarded and the pellet was stored at -80^{o}C until the next step. Cells were thawed for 15 min on ice and suspended in lysis buffer (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl). Sonication on ice was conducted using 6, 10-sec bursts at 200-300 Watts with at least 30-sec cooling time between each burst. Lysates were centrifuged at 10,000 x*g* for 30 min at 4^{o}C. The supernatant from the lysate was applied to a purification column containing Ni-NTA-agarose equilibrated with lysis buffer at pH 8.0. Three washes were conducted: wash 1, 5 ml of pH 8.0 lysis buffer; wash 2, 5

ml of pH 6.0 lysis buffer; wash 3, 5 ml of pH 5.0 lysis buffer. Purified 6x His-tagged cystatin was eluted in 2 ml of pH 4.5 lysis buffer containing 10 µM of *β*-mercaptoethanol. After elution, 100 µl of 1 M Tris pH 8.0 was added to adjust the pH to 7.0. The purified cystatin was stored at -20°C.

A1.10.2. *In vitro* activity gel assay to assess effects of soyacystatin N on activity of XCP1

Fifty mg of xylem tissue from 8-week-old root hypocotyls and 50 mg of leaves from 4-week-old WT [*p35S::XCP1*] plants were homogenized in extraction buffer [50 mM Na_2HPO_4 pH 7.2, 7 mM *β*-merceptaethanol, 20 µM LPP] using a 10:1 buffer:tissue (vol:weight) ratio. Samples were centrifuged at 12,000 rpm at 4°C for 15 min. The Supernatant was carefully transferred to a new tube and placed on ice for protein quantification. The activity gels were prepared as described in Beers and Freeman (1997).

A1.10.3. Construction of *p35S::cys^{his}* and *pXCP1::cys^{his}* in pCB302 Minibinary Vector

The expression vector *pET-cys^{his}* served as a template for the amplification of a fragment containing *soyacystatin N* along with C-terminal 6x HIS tag (*cys^{his}*) using the PCR primers PR27 (TTAATACGACTCACTATAGGG) and PR33 (AGATCTTGTTAGCAGCCGGATCTCA). The amplified PCR product contains the endogenous *Nco*I restriction site corresponding to the first ATG of the soyacystatin N coding region. PCR product was then TA-cloned into pGEM to construct P1023. The presence of *cys^{his}* was verified by sequencing using SP6 primer endogenous to pGEM. P1023 served as a template to release an *Nco*I-*Bgl*II fragment containing *cys^{his}*. The *GFP* coding region was removed from the modified (Funk et al., 2002) pAVA120 (von Arnim et al., 1998) expression vector by *Nco*I-*Bgl*II digest. The linearized pAVA120 was used as an intermediate vector to receive the *Nco*I/*Bgl*II *cys^{his}* fragment, which led to the construction of plasmid P1025. P1025 was digested with *Pst*I to release a *p35S-cys^{his}*-Terminator fragment. This fragment was ligated to a dephosphorylated *Pst*I-linearized pCB302 to construct P1027 (*p35S::cys^{his}*). *Nco*I/*Pst*I *cys^{his}*-Terminator fragment from P1025 was inserted in *Nco*I/*Pst*I-linearized V1009, a pCB302-*pXCP1::GUS* vector (Kositsup, 2000). This led to the construction of P1026 which contains *pXCP1-cys^{his}* -Terminator (*pXCP1::cys^{his}*).

A1.10.4. Sequence assessment and analysis of the soyacystatin N fragment

pET-cys^{his} was sequenced using T7 sense primer endogenous to the vector. P1023 was sequenced using SP6 primer endogenous to pGEM. P1026 and P1027 were sequenced using PR33. Restriction digests (not shown) verified the orientation and the expected transcriptional fusions with corresponding promoter sequences. The ORF was verified through DNASTAR software. BLAST P (Altschul et al., 1990) was used to confirm homology with soyacystatin from *Glycine max*. Conserved Domain search was conducted to confirm that the sequence contains the expected domain of soyacystatin N, which inhibits papain and papain-like cysteine proteases

A1.10.5. Phenotypic analysis of T_1 and T_2 plants

The *bar*$^{+}$ T_1 plants from WT [*p35S:: cys*^{his}] and WT [*pXCP1:: cys*^{his}] were compared to wild types plants and were categorized based on any apparent mutant phenotypes. T_2 plants from each putative mutant were retested for the phenotype.

A1.10.6. Reverse transcription PCR

Total RNA was isolated from above ground tissue of T_{2P} WT [*p35S:: cys*^{his}] plants according to RNeasy® Plant kit. First strand cDNA was synthesized according to RETROscript™. Wild type gDNA was prepared using QIA Plant® DNA kit. PCR was conducted using PR54 (ATGTGCATGGAGCTGCCAAC) and PR55 (CACATTGACTGGCTTGAATTCT), which are internal to *cys*^{his} sequence. Actin coding region was amplified using the primers PR 56 (CTGACTCATGGTACTCACTC) and PR 57 (GGCCGATGGTGAGGATATTC) and was used a positive control for RT-PCR.

Results

B1.1. XCP 2

XCP 2 (*XCP2*) is located on chromosome 2 of the *Arabidopsis* genome (At2g20850). Its paralog, *XCP1*, is located on chromosome 4 (At4g35350). These proteases belong to the C1A family of papain-like cysteine proteases whose predicted function is hydrolysis of peptide bonds. At the amino acid level, XCP1, XCP2, and papain (Figure 1-1) share higher identity (82%) within their prodomains compared to their mature domains (a minimum of 44%). The apparent preservation of high identity among these prodomains suggests similar regulation and localization and perhaps function for these proteins. Although specific substrates for XCP1 and XCP2 are not known, their expression during the late stages of TE differentiation suggests they may be required in hydrolyzing TE protoplast proteins specifically during the autolysis phase (Funk et al., 2002).

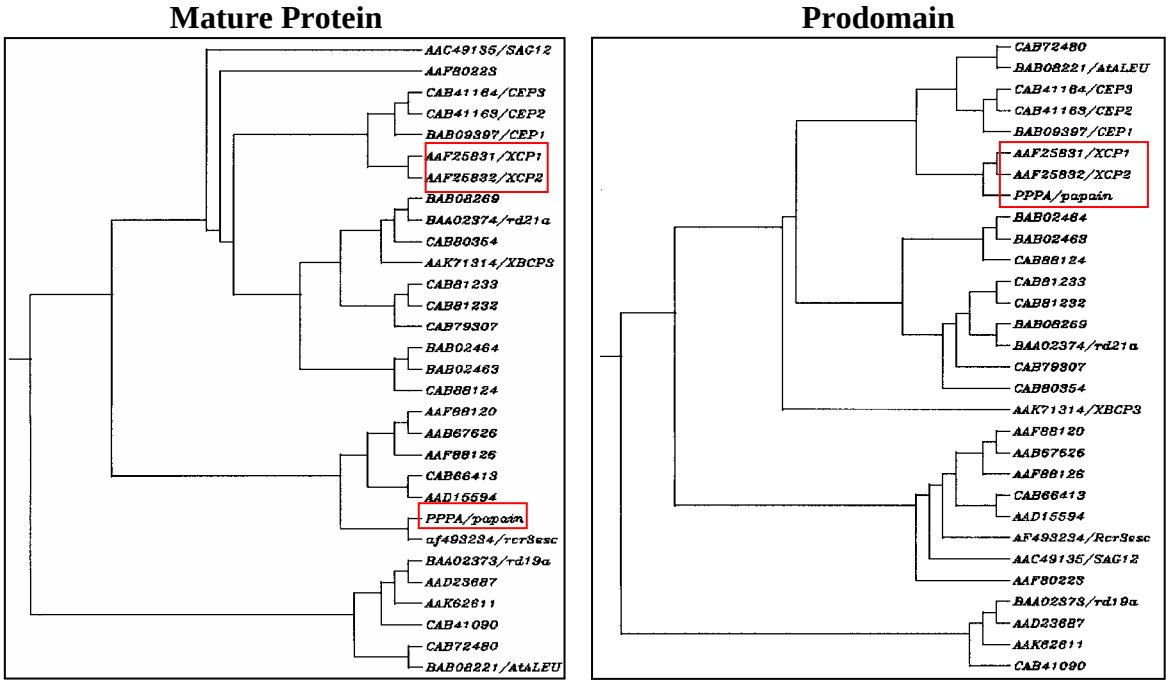


Figure 1-1. Phylogenetic trees of all *Arabidopsis* C1A proteases and the archetypical C1A protein, papain. XCP1, XCP2, and papain mature domains are more divergent (left panel) compared with the prodomains (right panel).

Gene structure of *XCP2* comprises a 1.89 kb 5' region upstream of its initiator methionine, four exons, three introns, and a 3' UTR (TAIR database; www.*arabidopsis*.org). The four exons, which extend for 1071 bp, encode a 357-amino- acid preproprotein (polypeptide molecular mass of 39.7 kD) comprised of three domains (Fig. 1-2A). The N-terminal predomain is the 28- amino- acid

signal sequence for targeting into the secretory pathway (Fig. 1-2B). The second domain, a 108-amino-acid prodomain contains the ERFNIN motif (Karrer et al., 1993) found to be conserved in papain, XCP1, and XCP2 sequences (Zhao et al., 2000). Prodomains ensure proper protein folding and stabilization of proteins in alkaline and neutral conditions (Yamamoto et al., 2002) as well as exerting inhibitory effects on the activity of their cognate mature protein (reviewed by Abrahamson et al., 2003). The third domain is the mature domain which consists of 221 amino acids (predicted molecular weight 23.8 kD) and includes the catalytic dyad Cys^{25} and His^{159} according to papain numbering (Fig. 1-2B).

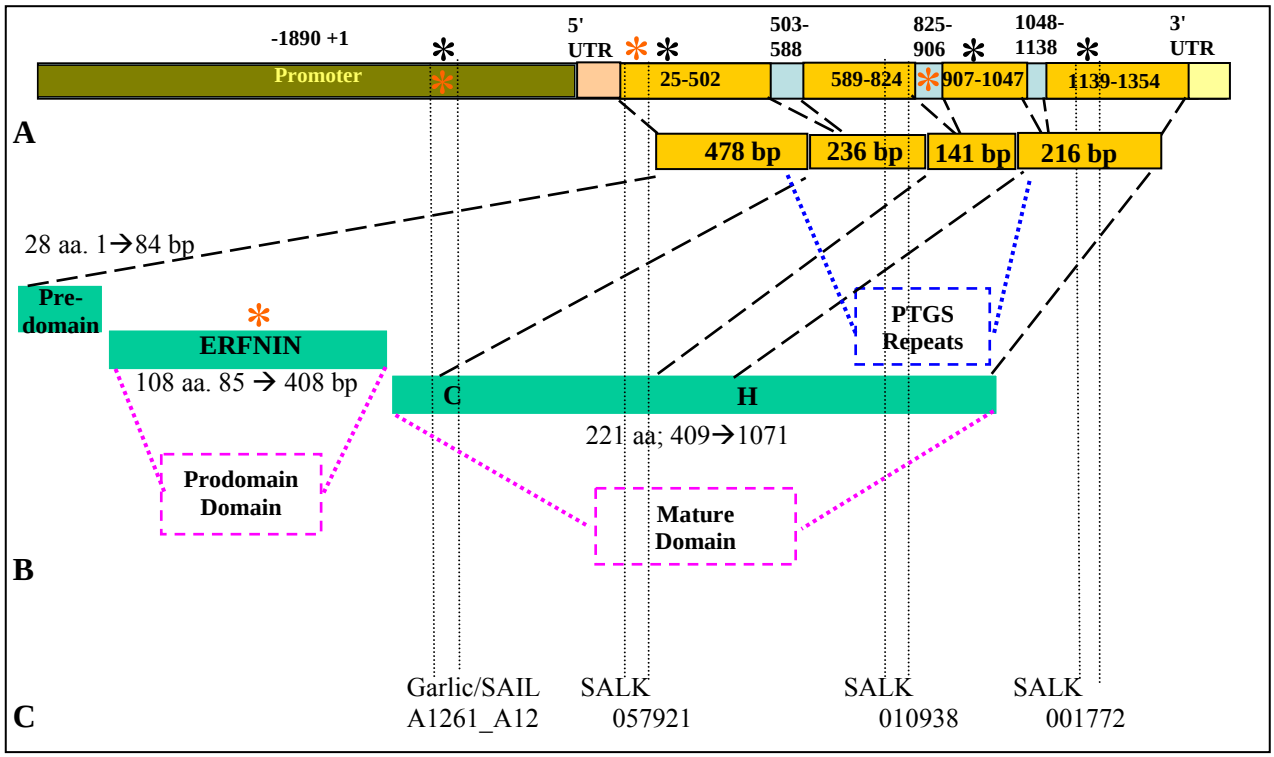


Figure 1-2. An overview of the gene and protein structures of XCP2 as they relate to T-DNA insertional mutants and post-transcriptional gene silencing analyses. **A**, Gene structure of XCP2 is represented by a 1.89 kb 5' upstream putative promoter, a 25-bp 5' UTR, a 4-exon and a 3-intron region spanning from the 25- to 1354-bp, and a 191-bp 3' UTR. **B**, The four exons encode a 357 amino-acid preproprotein (green boxes). Catalytic dyad Cys^{25} and His^{159} (papain numbering) are located in mature domain whereas the ERFNIN motif is located in the prodomain. The sequence bordered by blue lines was selected for construction of inverted repeats of XCP2 for silencing experiments. **C**, The four T-DNA lines used in T-DNA insertional mutagenesis studies. Putative locations of T-DNA inserts are represented by black asterisks whereas their verified locations based on sequenced junctions are represented by orange asterisks. T-DNA insertion in SALK_001772 was not evident from our results.

The 1.89 kb *pXCP2* was able to drive GUS expression in TE of the roots and leaves of *Arabidopsis*. Funk et al. (2002) showed the TE-localized expression of GUS at the whole leaf level (Fig. 1-3A) and in the root secondary xylem tissue (Fig. 1-3B). Figure 1-3C confirms that *pXCP2* directs GUS expression specifically in differentiating TE. GUS expression was observed at various stages of *Arabidopsis* development such as two-day-old roots, young true leaves at the two-leaf stage, young stem sections, and buds of emerging flowers (data not shown). Establishing an efficient and reproducible technique to study TE-specific GUS expression is challenging because of the discontinuous nature of TE differentiation *in vivo*. Thus, repeated testing of transgenic plants using different biological samples at different developmental stages was necessary to produce a fair representation of GUS activity in plants.

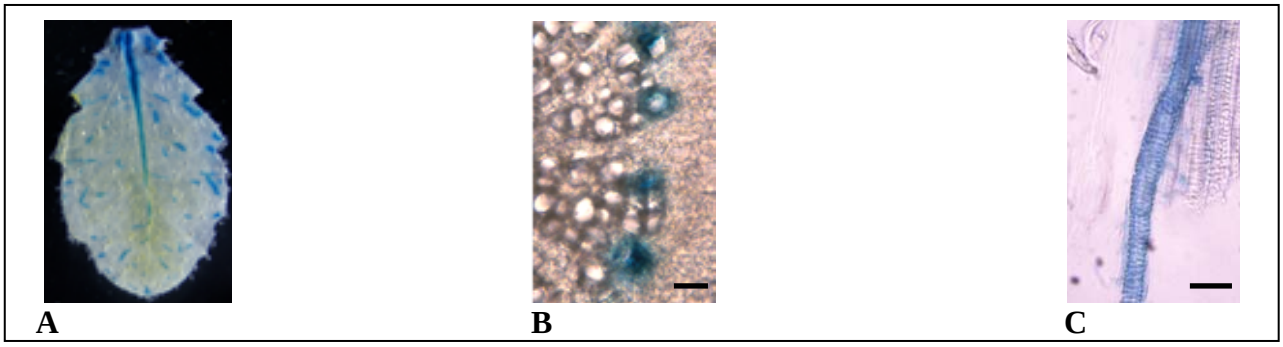


Figure 1-3. The promoter of XCP2 directs GUS expression in TE of *Arabidopsis* roots and leaves**. A**, GUS expression directed by *pXCP2* at the whole leaf level is observed at localized areas within the vascular tissue (Funk et al., 2002). **B**, A free-hand transverse section of root tissue showing that GUS expression is activated in the TE of the secondary xylem (Funk et al., 2002) by *pXCP2*. **C**, A macerated root whole mount from a 10-day-old seedlings showing GUS expression in differentiating TEs. Bars = 20 µm.

B1.2. Analysis of GARLIC/SAIL 1261_A12

B1.2.1. T-DNA Verification

According to sequences obtained from Torrey Mesa Research Institute, Garlic/SAIL line 1261_A12 contained a T-DNA insertion in the *XCP2* promoter. To identify the location of the T-DNA insertion, two T-DNA flanking primers, PR13 and PR14, and one left-border specific primer, PR15, were used on gDNA of herbicide-resistant plants (Fig. 1-4).

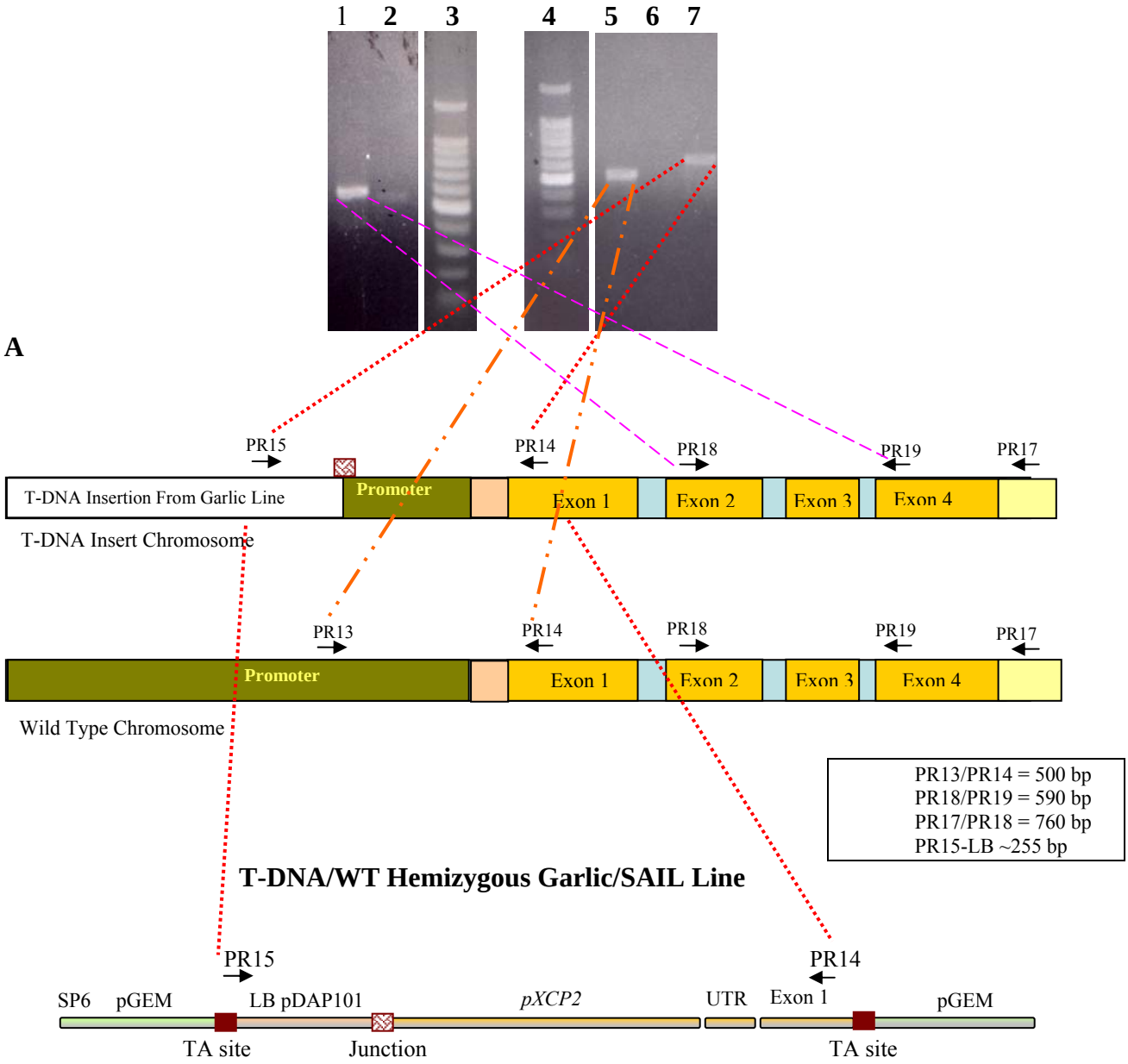


Figure 1-4. Genotyping and sequence analysis of Garlic/SAIL 1261_A12 T_{3P} plants. **A**, Gel electrophoresis of PCR products. Lanes 1, WT PR18/PR19; lane 2, WT PR13/14; lanes 3 and 4, 100-bp ladder (Promega); lane 5, Garlic PR13/PR14; lane 6, Garlic PR13/PR15; lane 7, Garlic PR15/PR14. **B**, Schematic of pGEM cloning site containing T-DNA/gDNA fragment of Garlic/SAIL 1261_A12 derived from sequence analysis using SP6 primer. Region of TA site in pGEM is represented by filled box. Region containing the junction between the T-DNA and gDNA is represented by patterned box. Box shows expected size of the PCR product when amplified by the given primers. Broken lines connect PCR primers with products visualized on ethidium-bromide stained agarose gel (A) or schematic if cloned T-DNA/gDNA junction (B).

The PCR program used a 1-min extension time sufficient to amplify up to 1000 bp fragment but not the T-DNA. Figure 1-4A shows that PR18/PR19 primer combination was able to amplify a "wild-type" XCP2 fragment of approximately 590 bp. The presence of T-DNA insertion was confirmed by the amplification of approximately 600-bp fragment (Fig. 1-4A, lane 7). The PR13/PR14 primer combination was able to amplify an XCP2 wild-type fragment (Fig. 1-4, lane 1). PR15/PR13 did not amplify any PCR fragment (Fig. 1-4A lane 6), thus confirming orientation of the T-DNA insertion in XCP2 promoter region. The Garlic/SAIL 1261_A12 [PR14/PR15] was cloned into pGEM and was subsequently sequenced using SP6 primer endogenous to pGEM (Fig. 1-4B). Sequence analysis verified that the T-DNA/gDNA junction was located in the XCP2 promoter region. This finding was in agreement with the sequence information initially provided by Torrey Mesa Research Institute (Fig. 1-2C).

B1.2.2. Garlic/SAIL 1261_A12 T-DNA line expresses XCP2 at normal levels and their initial non-wild-type phenotype was not reproducible

Selfed progeny from herbicide resistant plants were divided into two main groups. The first was a group characterized by yellowing (Y) of the first two young true leaves at the two-leaf stage of plant development. At the same developmental stage, the other group was similar to WT plants with green (G) young true leaves (Fig. 1-5A). This phenotype was repeated only once in soil but was not reproducible thereafter. Western analysis of Y, G, and WT plant protein extracts using antibody raised against the proprotein (prodomain plus the mature domain), PRMXCP2, detected three bands at the 45kD, 36kD, and less than 29kD molecular weight in all three protein samples (Fig.1-5B and 1-5C). A Reverse transcription PCR reaction conducted on Y (Fig. 1-5D), G, and WT (not shown) using the exon specific PR18/PR19 primer combination showed an expected cDNA fragment of less than 400 bp. No DNA contamination was evident in this experiment (data not shown). It is evident from this analysis that there is no mutant phenotype associated with the Y plants because the T-DNA insertion in the promoter region did not affect XCP2 at the RNA and protein levels or because the plants were not homozygous for the T-DNA. This line was not investigated any further. SALK lines with predicted T-DNA insertion in exons of XCP2 became available and they were investigated as described below.

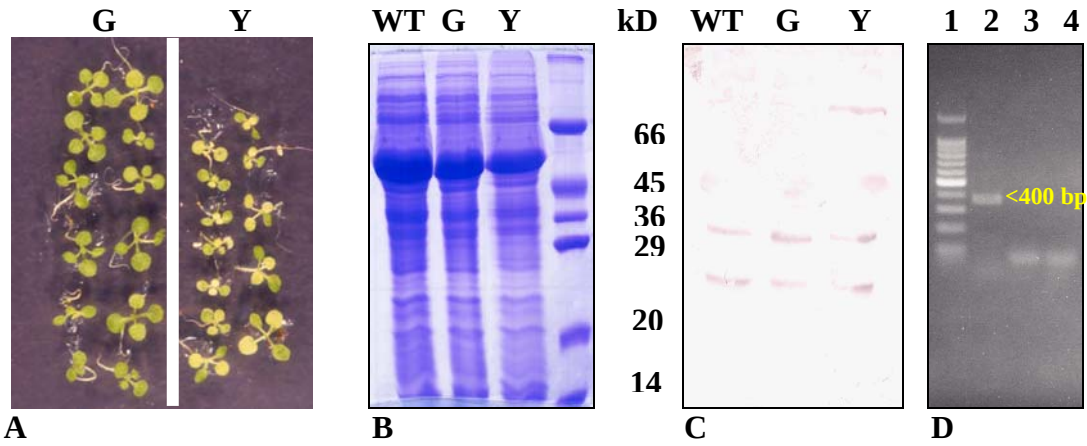


Figure 1-5. Phenotypic, Western, and RT-PCR analyses of T_{3P} Garlic/SAIL 1261_A12 Plants. **A**, T_{3P} plants derived from Garlic/SAIL 1261_A12 initially showed the yellowing (Y) phenotype represented in upper panel as well as wild type-like green (G) phenotype (lower panel). **B**, SDS-PAGE of proteins extracted from above-ground tissue of two week-old seedlings from Y, G, and WT seedlings (WT). **C**, Western analysis using PRMXCP2[ab]. D, RT PCR; lane1, 100-bp ladder (Promega); lane 2, Garlic cDNA from Y plants with PR18/PR19 showing a less than 400-bp band; lane 3, RT-PCR with no reverse transcriptase; lane 4, RT-PCR with no template.

B1.3. Analysis of SALK_010938

B1.3.1. T-DNA Verification

Genomic DNA from three plants of T_3 SALK_010938 was extracted. To identify location of T-DNA, primers flanking the putative T-DNA location, PR 18 and PR17, and the left border specific primer, PR16, were used. The PCR program used 1-min extension time sufficient to amplify up to 1000-bp fragment but not the T-DNA fragment. Three plants were analyzed. Figure 1-6A (lanes 9, 10, and 11) showed that the T-DNA insertion was present in all three individual plants since the bands were amplified using PR16/PR19 primer combination. The PCR band was cloned into pGEM and its sequence was analyzed (Fig. 1-6B). Sequence analysis confirmed the presence of T-DNA insert in intron 2 of *XCP2*. PR18/PR17 combination did not amplify a band from SALK_010938-1 and SALK_010938-2, but it did amplify a band in SALK_010938-3 (Fig. 1-6A, lanes 3, 4, and 5). In wild type plants (Fig.1-6A, lane 2), the proximity of PR18 and PR17 to each other was compatible with 1-min extension time, thus a band was amplified. Similarly, in hemizygous plants where one chromosome was a 'wild-type' a band was also amplified using

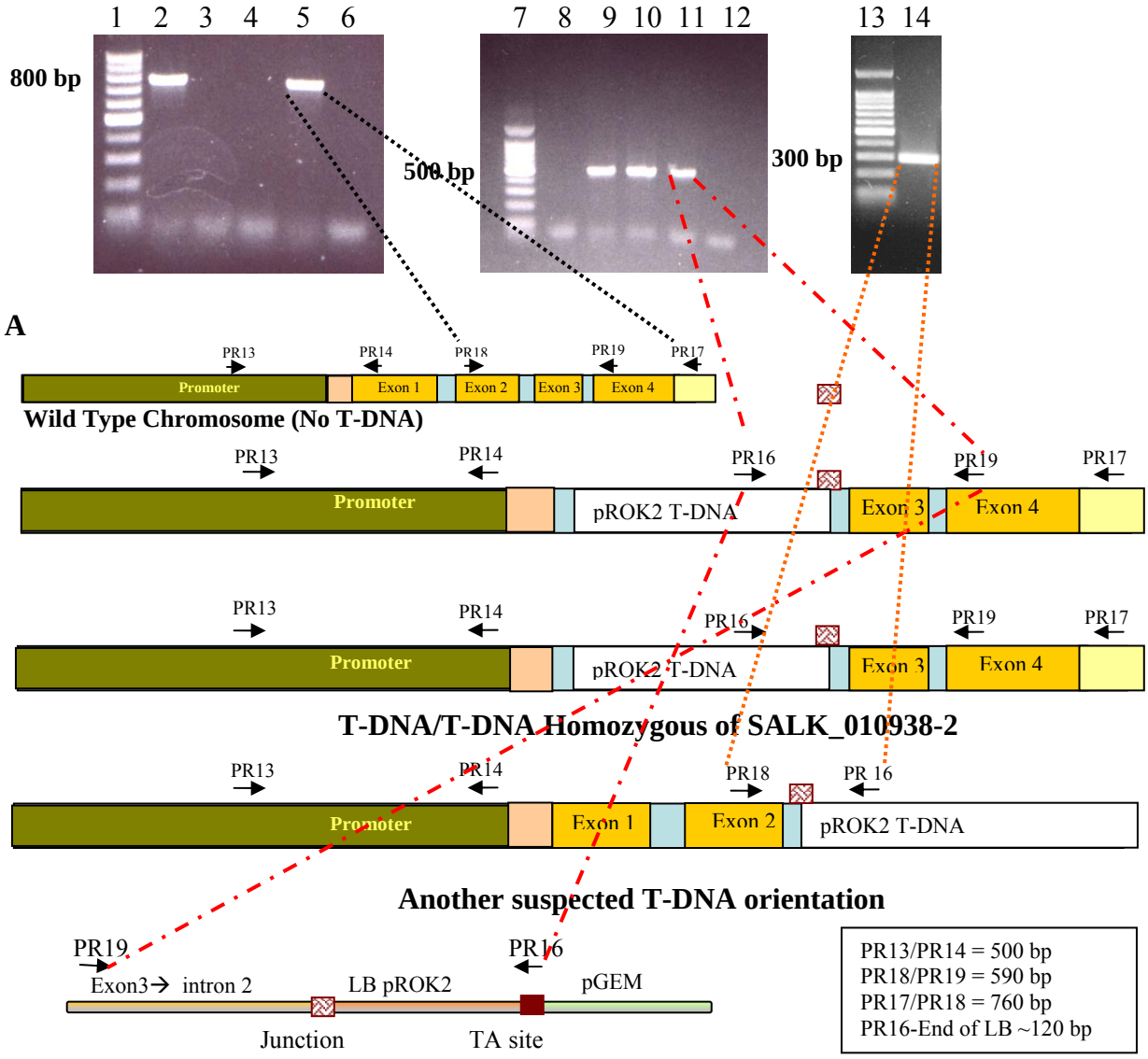


Figure 1-6. PCR genotyping on SALK_010938 T_3 plants and sequencing of T-DNA/gDNA junction. **A**, Gel electrophoresis of PCR products. Lanes 1, 7, and 13, 100-bp ladder. PR18 and PR17 were used on the following templates: 2, WT; 3, SALK_010938-1; 4, SALK_010938-2; 5, SALK_010938-3; 6, no template. PR16 and PR19 were used on the following templates: 8, WT; 9, SALK_010938-1; 10,SALK_010938-2; 11, SALK_010938-3; 12, no template. 14, PR18 and PR16 were used on SALK_010938-3. **B**, Schematic of pGEM cloning site containing the T-DNA/gDNA junction of SALK_010938-2 derived from sequence analysis using primer PR19. T-DNA was inserted in intron 2 of XCP2. An inverted T-DNA may also be present. Region of TA cloning in pGEM is represented by filled box. Region containing the junction between the T-DNA and gDNA is represented by patterned box. Numbers on side of gel show the sizes of fragments. Box shows expected size of the PCR product when amplified by the given primers. Broken lines connect PCR primers with products visualized on ethidium-bromide stained agarose gel (A) or schematic of cloned T-DNA/gDNA junction (B). Numbers on left of the gels show the size of the DNA standard band.

PR18/PR19 (Fig. 1-6A, lane 5). In T-DNA homozygous plants lacking 'wild-type' chromosomes, the T-DNA expands the distance between PR18 and PR17 far enough to prevent amplification under 1-min extension time. Thus a band was not amplified. An unexpected 300-bp fragment was amplified by PR18 and PR16 using gDNA of SALK_010938-3 (Fig.1-6A, lane 14 and Fig. 1-6B) and SALK_010938-2 (not shown). Thus the left border primer, PR16, served alternatively as a 'sense' and 'antisense' primer suggesting the presence of inverted T-DNA insertion (Fig. 1-6B) in at least these two plants. Although the PCR fragment was not sequenced, the predicted location based on the size of PCR products is likely to interrupt intron 2 as well (see Fig. 1-6B). Sequence analysis (Fig. 1-6B) of SALK_010938 [PR16/PR19] PCR fragment showed that the T-DNA/gDNA border was in intron 2 of *XCP2*. This finding was in disagreement with the sequence information provided by ABRC (Fig. 1-2C).

B1.3.2. XCP2 was not detected in the SALK_010938-2 protein extract

The results of western analysis using PRXCP2^{ab} on proteins extracts from roots and leaves of T_4 SALK_010938-2 homozygous plant (T-DNA/T-DNA) are presented in Figure 1-7. Bands of slightly less than 45 kD (PRXCP2) were detected in leaves and roots of WT plants, which were the genetic background used in all SALK T-DNA insertion lines. These bands, however, were not detected in T-DNA homozygous knockout lines. Although XCP2 expression was abolished, there were no observable differences at the phenotypic level between the WT and SALK plants (not shown) under our normal laboratory conditions.

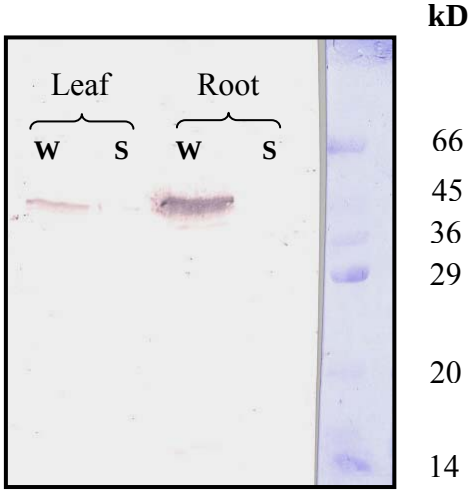


Figure 1-7. Immunoblot analysis using PRXCP2^{ab} on protein extracts from roots and leaves of T_4 SALK_010938-2 homozygous plants. W, Columbia wild type plants. S, Salk line.

B1.3.3. *In vitro* starvation experiments on SALK_010938

T_4 plants were allowed to set seed to generate the T_{5P} generation. Western analysis on T_{5P} plants (not shown) confirmed the results obtained earlier (Fig. 1-7). T_{5P} seeds were used in series of *in vitro* nutrient starvation experiments to see whether a non-wild type phenotype could be produced under various physiological stresses. Physiological stress was imposed by eliminating N source, P source, C source, or a combination of all. Pilot experiments using WT seedlings revealed signs of stress (e.g. anthocyanin production, yellowing of leaves, stunted growth) when grown on starvation media. Normal plant growth was observed in optimal media. Fourteen-day old T_{5P} (Fig. 1-8A) and WT (Fig. 1-8C) seedlings grown on optimal germination media showed no signs of stress and root and leaf development followed a typical *Arabidopsis* growth pattern. On media lacking a nitrogen source, the number and length of lateral roots appeared normal in T_{5P} seedlings (Fig. 1-8B) but drastically decreased in WT seedlings (Fig. 1-8D). Although these results were observed in all replications of this experiment, they were not reproduced in subsequent trials. Evaluation of seven-day old seedlings of WT, T_{5P}, and T_4 SALK_057921 grown on optimal medium (Fig. 1-8E) and medium lacking nitrogen (Fig. 1-8F) showed no phenotypic differences in the roots or leaves. Carbon and phosphate starvation experiments (not shown) were conducted in a similar manner to that for nitrogen starvation experiments. No phenotypic differences were observed between WT and SALK seedlings tested.

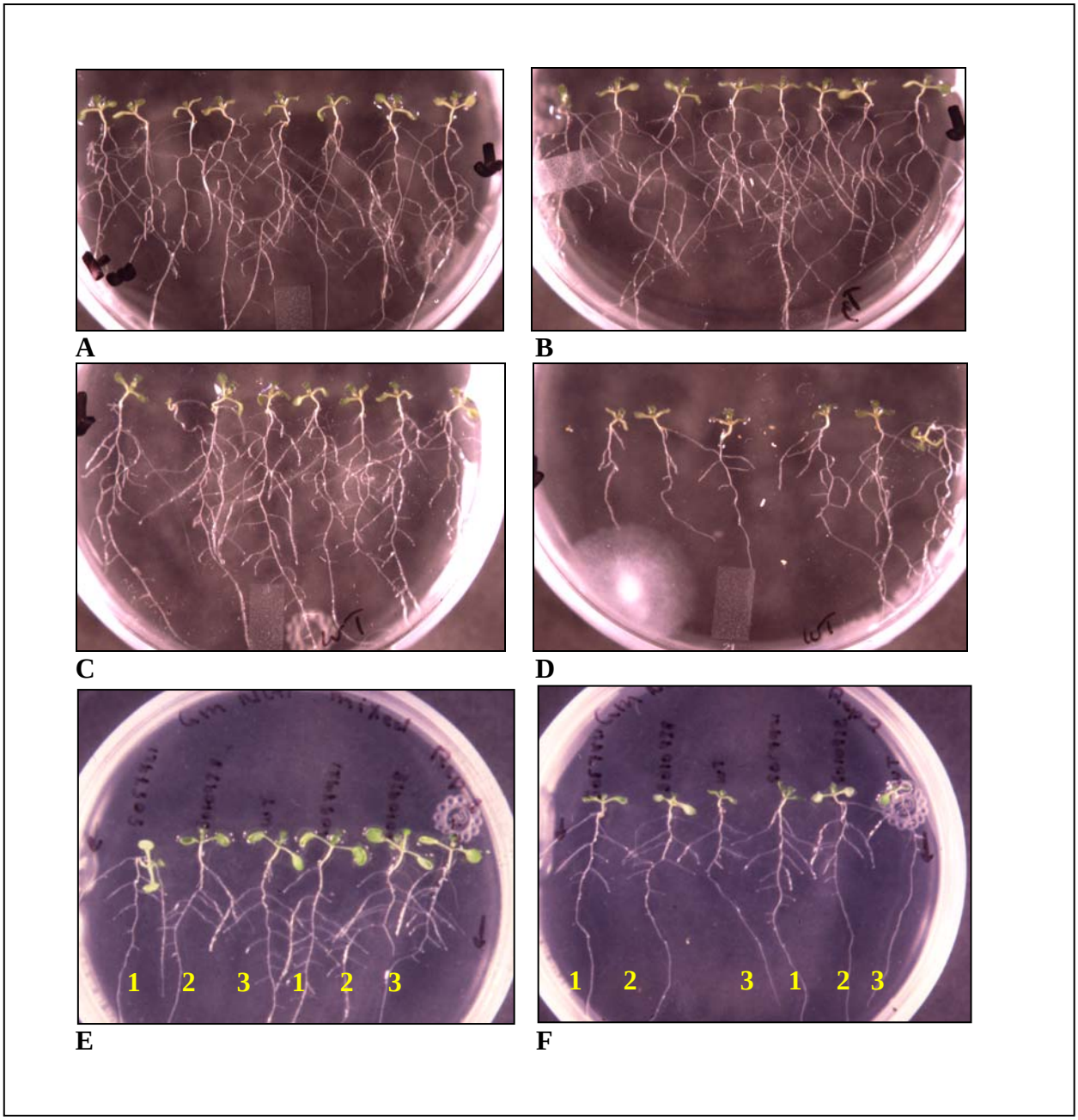


Figure 1-8. *In vitro* nitrogen starvation experiments with *Arabidopsis* T_{5P} SALK_010938-2 homozygous line compared to its wild type genetic background Columbia and another allelic T-DNA insertion T_4 SALK_057921. **A** and **C,** Effects of optimal germination media on fourteen-day old seedlings of T_5 SALK_010938-2 (**A**) and WT (**C**). **B** and **D**, Effects of germination media lacking nitrogen source on ten-day old seedlings of T_5 SALK_010938-2 (**B**) and WT (**D**) on germination media lacking nitrogen. **E** and **F**, Two representative experiments showing 10-day old seedlings of T_4 SALK_057921 (seedling 1), T_{5P} SALK_010938-2 (seedling 2), and WT (seedling 3) growing on normal germination media (**E**) and on germination media lacking nitrogen (**F**).

B1.4. Analysis of SALK_057921

B1.4.1. T-DNA Verification

Eighteen plants were grouped into three six-plant pools and gDNA was prepared from these pools. Based on PCR results, all pooled DNA contained a T-DNA insertion (Fig. 1-9A lanes 2, 3, 4). Primers flanking the T-DNA insertion were able to amplify a wild-type fragment (not shown) suggesting that T-DNA insertion is hemizygous (T-DNA/WT). Sequence analysis of the PCR product showed that T-DNA insertion interrupts the first asparagine (N) residue in the ERFNIN motif in exon 1. The results are in agreement with sequence information obtained from ABRC in that the T-DNA is in exon 1 of XCP2. None of the plants tested were homozygous for the T-DNA. T_3 plants did not reveal phenotypic differences when compared to WT plants.

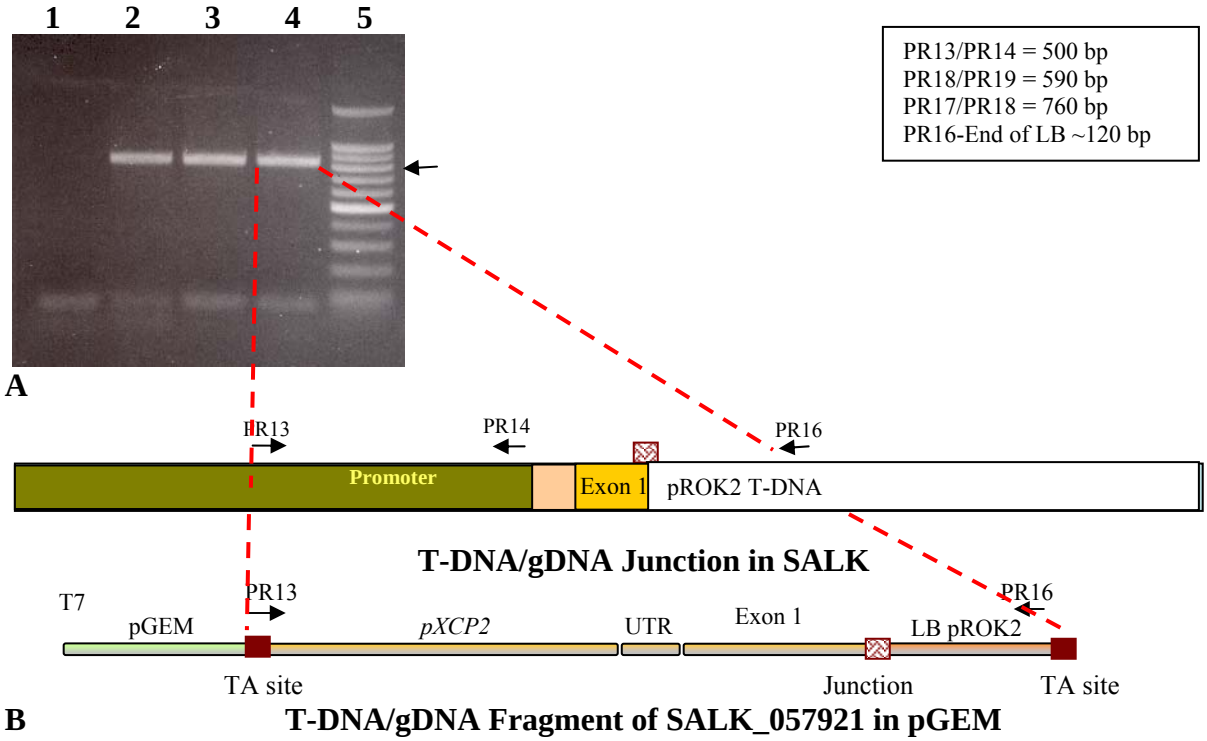


Figure 1-9. PCR genotyping of T_{3P} SALK_057921 plants. **A**, Gel electrophoresis of PCR products. Lanes 1, no template PR13/PR16; 2, Plant pool I gDNA PR13/PR16; 3, Plant pool II gDNA PR13/PR16; 4, Plant pool III gDNA PR13/PR16; 5, The 100-bp ladder. **B**, Schematic of gDNA/T-DNA junction in SALK_057921. Region of TA cloning in pGEM is represented by filled box. Region containing the junction between the T-DNA and gDNA is represented by patterned box. Sequencing using T7 primer showed that the T-DNA derived from pROK2 binary vector was inserted in exon 1 interrupting the ERFNIN motif. Box shows expected size of the PCR product when amplified by the given primers. Broken lines connect PCR primers with products visualized on ethidium-bromide stained agarose gel (A) or schematic of cloned T-DNA/gDNA junction (B).Number on right of gel is the size of DNA standard band.

B1.5. Analysis of SALK_001772

B1.5.1. T-DNA Verification

PCR amplification revealed only the presence of wild-type fragments (Fig. 1-10). Primers PR18 and PR17 amplified a 'wild-type' fragments of ~760 bp (lane 1) using WT gDNA. The same fragment was obtained from SALK pools I and II (lanes 6 and 7, respectively). Primers PR16 and PR18 failed to amplify a T-DNA junction from SALK pools I and II (lanes 4 and 5). SALK_010938-3 was intended for use as a negative control. The significance of this amplification was discussed earlier (1.3.). PCR reactions using PR16 and PR17 did not amplify any bands in neither pools (data not shown).

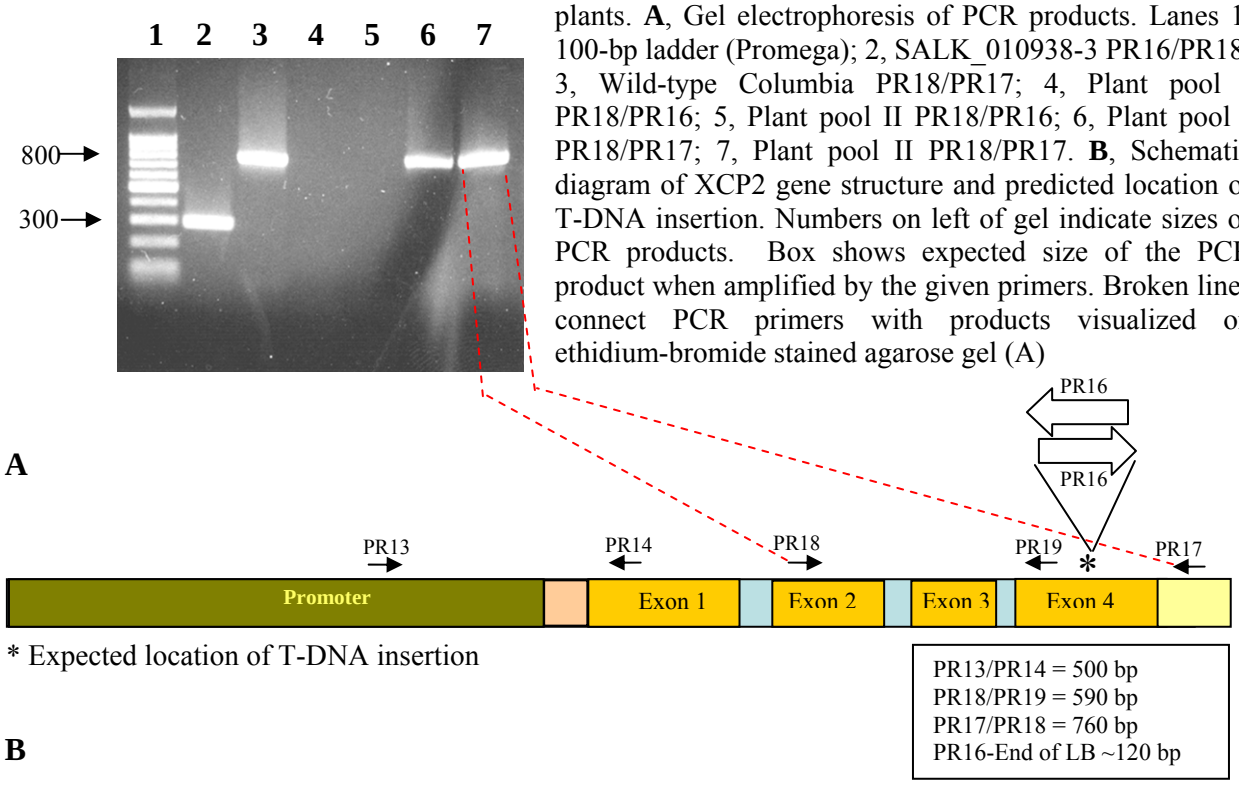


Figure 1-10. PCR genotyping of SALK_001772 T_{3P} plants. **A**, Gel electrophoresis of PCR products. Lanes 1, 100-bp ladder (Promega); 2, SALK_010938-3 PR16/PR18; 3, Wild-type Columbia PR18/PR17; 4, Plant pool I PR18/PR16; 5, Plant pool II PR18/PR16; 6, Plant pool I PR18/PR17; 7, Plant pool II PR18/PR17. **B**, Schematic diagram of XCP2 gene structure and predicted location of T-DNA insertion. Numbers on left of gel indicate sizes of PCR products. Box shows expected size of the PCR product when amplified by the given primers. Broken lines connect PCR primers with products visualized on ethidium-bromide stained agarose gel (A)

B1.5.2. Phenotypic Evaluation

The T_3 seeds obtained from SALK_001772 showed a 3% germination rate (i.e. 1 seed out of 30 seeds germinated). A second batch of seeds was requested and seeds were germinated *in vitro* and in soil. However, germination rate did not exceed 5%. Based on PCR results, the germinating plants were wild type plants and do not contain the insert. Whether this low germination rate/lethality was correlated to T-DNA insertion, thus, could not be investigated.

B1.6. Post-transcriptional Gene Silencing

B1.6.1. *p35S::GUS*^{ir} and *pXCP1::GUS*^{ir} effectively and equally silenced GUS expression in WT [*pXSP1::GUS*]

Tracheary element-specific GUS expression directed by *pXSP1* was used as a phenotypic marker to test the efficiency of pFGC5941 in cell-type specific silencing. This experiment served as a pilot experiment upon which XCP2 silencing experiment was designed. The plant material used for this study was T_{2P} WT [*pXSP1::GUS*]. Segregation analysis of GUS expression in T_{2P} population was conducted and the 1 GUS^+:1 GUS^- ratio was not rejected (χ^2 =0.93; N=53; df =1, p = 0.05). This ratio was statistically maintained in a T_{1P} **[**T_{2P} [*pXSP1::GUS*]**]** [GV3101] population (χ^2 =0.1, N=11; df =1, p= 0.05). The pFGC5941-based constructs contain the *bar* gene conferring resistance to Finale® Herbicide. Twenty four T_1 $herb^+$ plants from each experiment were tested for GUS expression at the flower stage. Herbicide resistant plants exhibiting GUS activity suggested that the plants are double-transformed. The empty vectors *p35S:: empty* and *pXCP1:: empty* (see Fig. 1-11) did not silence GUS expression as evidenced by the high percentage of GUS-positive plants, which resembled GUS expression in the control plants; transformed with GV3101 *Agrobacterium* strain only. On the other hand, *p35S::GUS*^{ir} and *pXCP1::GUS*^{ir} caused a near complete silencing of GUS expression.

B1.6.2. Phenotypic observations of T_1 WT [*p35S::XCP2*^{ir}]and [*pXCP1::XCP2*^{ir}] plants

Following the same concept presented above, *p35S:: empty* and *pXCP1:: empty* served as empty vectors. Experimental vectors *p35S::XCP2*^{ir} and *pXCP1::XCP2*^{ir} were used to silence the endogenous XCP2. There were no observed differences between WT and the putatively XCP2-silenced plants.

First Transformation T_{2P} WT [*pXSP1::GUS*]		**After Double Transformation:** T_{1P} **[**T_{2P} WT [*pXSP1::GUS*]**] and**				
		[*p35S::empty*]	[*pXCP1::empty*]	[*p35S::GUS^{ir}*]	[*pXCP1::GUS^{ir}*]	[*GV3101*] 'CONTROL'
		bar^{+}	*bar*$^{+}$	*bar*$^{+}$	*bar*$^{+}$	*bar*$^{+}$
	1	positive	positive	negative	negative	positive
	2	positive	positive	negative	negative	positive
	3	positive	negative	negative	negative	positive
	4	positive	negative	negative	negative	negative
	5	positive	negative	negative	negative	negative
	6	positive	positive	----------	negative	positive
	7	positive	positive	positive	negative	positive
	8	positive	positive	negative	negative	negative
	9	positive	positive	negative	negative	positive
	10	positive	negative	negative	negative	negative
	11	positive	positive	negative	negative	negative
	12	positive	positive	negative	negative	
	13	negative	positive	negative	negative	
	14	positive	positive	negative	negative	
	15	positive	positive	negative	negative	
	16	positive	negative	positive	negative	
	17	negative	positive	negative	negative	
	18	positive	negative	negative	negative	
	19	positive	positive	negative	negative	
	20	positive	negative	negative	---------	
	21	negative	negative	negative	negative	
	22	positive	positive	negative	positive	
	23	positive	negative	negative	negative	
	24	negative	positive	negative	negative	
30/53	**GUS+/total**	**20/24**	**15/24**	**2/23**	**1/23**	**6/11**
56	**GUS + %**	**83**	**62**	**9**	**4**	**54**

Figure 1-11. Assessment of silencing efficiency of ds RNA vectors using GUS as a marker. Herbicide resistant plants carrying the ds RNA vector were randomly selected from each population. Per plant, newly opened flower clusters were used for GUS staining prior to and after the transformation. The T_{2P} WT [*pXSP1::GUS*] and T_{3P} WT [*pXSP1::GUS*] showed GUS segregation of 1:1 ratio.

B1.7. Soyacyctatin N

B1.7.1. The soyacystatin N domain was successfully integrated into the minibinary vector pCB302

Translated soyacystatin N sequences were aligned (Fig. 1-12A). The absence of several residues of his tag sequence was likely due to the sequencing reaction. For example, reliable reading of nucleotide sequences normally begins after 30 bases away from sequencing primer. PR33 sequencing primer was proximal to his tag sequence which has resulted in skipping the reading few of the 6x HIS residues. A BLAST P search using the ORF sequence as a query returned a 100% identity score to *Glycine max* soyacystatin. There is no signal sequence included in this recombinant protein. Thus, the protein is predicted to be directed to the cytosol. Restriction digests and PCR reactions verified that the soyacyctatin N fragment was fused to *p35S* in P1027 and to *pXCP1* in P1026 (data not shown). Taken together, the soyacystatin N fragment did not contain a mutation in its sequence and was cloned directly downstream of *p35S* and *pXCP1* sequences.

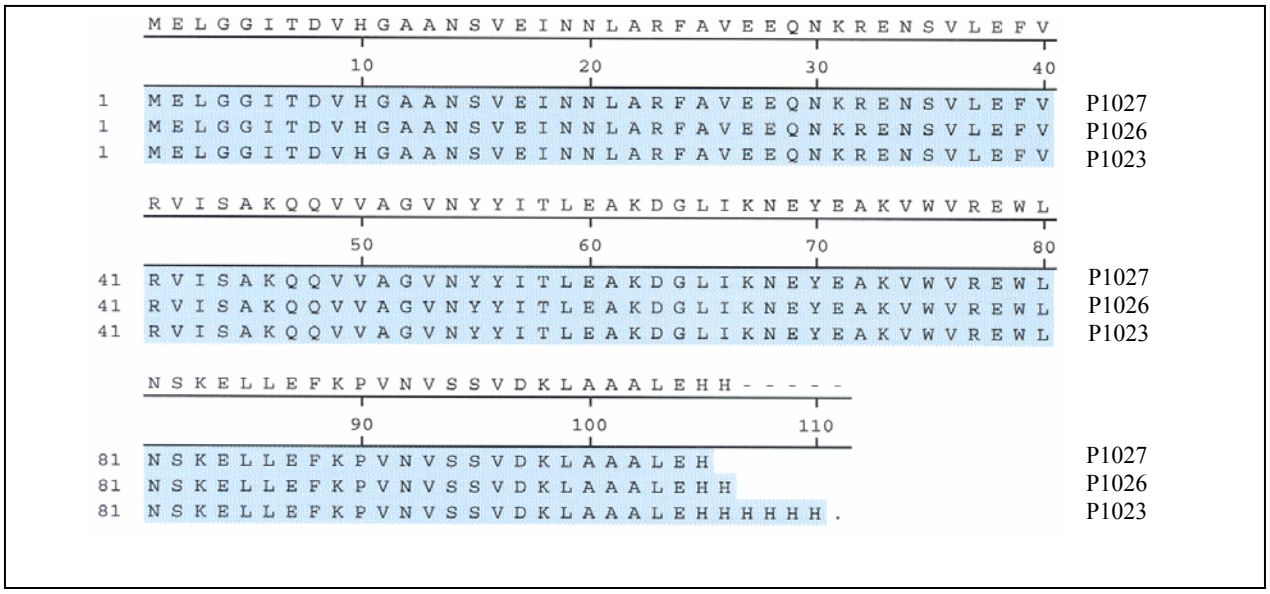


Figure 1-12. Soya*cystatin*^{his} translated DNA sequence of binary vectors P1026 and P1027 and cloning vector P1023.

The amino acid sequence was used as a query to search for a conserved domain (Marchler-Bauer et al., 2003). Figure 1-13A shows a single conserved domain that was shared among many members of cystatin families (Fig. 1-13B). Structural homologs, such as human cystatin stefin B,whose crystal structure was resolved, showed that the conserved N-terminal residue G and QVVAG were essential in binding the proteinase target. Belenghi and colleagues (2003) showed that an *Arabidopsis* cystatin

Figure 1-13. Computer-based analysis of soyacystatin N ORF. **A**, Conserved Domain Database search (Marchler-Bauer et al., 2003) using *cys*^{his} translated amino acid from P1027 as a template. **B**, Conserved domain alignments using the same template. Conserved residues (G)Gly[11], (Q)Gln[55], (V)Val[56], (V)Val[57], (A)Ala[58], and (G)Gly[59] [human cystatin C numbering] and semi-conserved motif (W)Trp[78] and (E)Glu[79] [Query numbering] are boxed. Human cystatin C, 1G96_A, (Janowski et al., 2001). Oryzacystatin-I, 1EQK_A [Nagata et al., 2000]. Cowpea cystatin, gi# 1169196 (Fernandes et al., 1993). *Purys communis* cystatin, gi# 2317876 (Technologie des Produits Vegetaux, France). Ragweed *Ambrosia artemisiifolia* cystatin gi# 437312, (Rogers et al., 1993). *Carica papaya* cystatin, gi# 311505, (RC Bateman, University of Southern Mississippi). *Glycine max* cystatin, gi#1277166 (Botella et al., 1996; Zhao et al., 1996; Xiaomu Niu, Horticulture, Purdue University). *Arabidopsis thaliana* cytatin, gi# 15226769. 1STF-I, gi#494619, chain I of papain complexed with stefin B cystatin (Jerala et al., 1990; Stubbs et al., 1990).

1 protein, homologous at the amino acid level to soyacystatin N (Fig. 1-13 B) was structurally similar to human stefin B which binds papain (Fig, 1-13 D). This sequence comparison and experimental evidence presented before (Koiwa et al., 2001) suggested that the soyacystatin N sequence used in our study should function in inhibiting papain-like cysteine proteases *in planta*.

B1.7.2. Soyacystatin N inhibits cysteine proteases *in vitro*

To provide experimental evidence that soyacystatin N can inhibit the activity of cysteine proteases XCP1 and XCP2; an activity gel assay was conducted. Protein extracts from secondary xylem of the roots (Fig. 1-14A, lane 1) and leaves (Fig.1-14A, lane 2) of XCP1 overexpressor line T_{2P} WT [*p35S::XCP1*] contained protease activity at approximately 26 kD shown previously to be due to XCP1 (Funk et al., 2002) capable of hydrolyzing gelatin (Fig 1-14B, lanes) in the absence of the proteinaceous inhibitor soyacystatin N. The addition of soyacystatin N blocked the 26-kD proteolytic activity (Fig 1-14C). Similar experiments were conducted using LPP as control. LPP also blocked the activity of the 26 kD protein (data not shown)

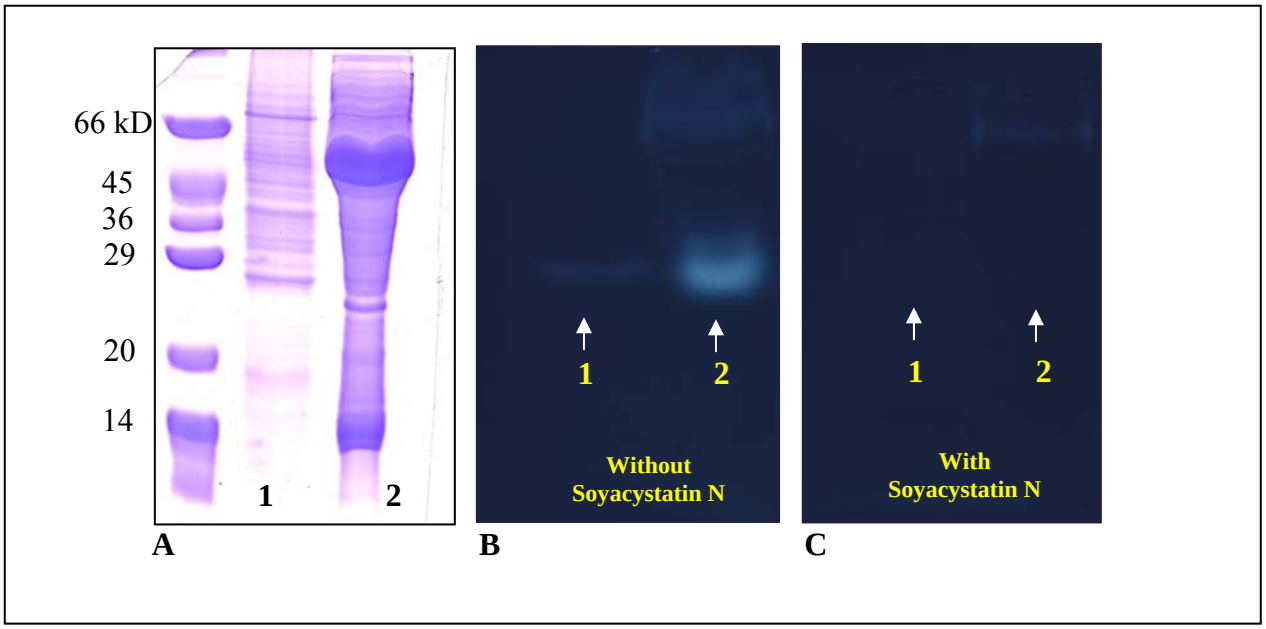


Figure 1-14. Cysteine proteases from root secondary xylem and leaves of the T_2 WT [*p35S*::*XCP1*] were inhibited by Soyacystatin N *in vitro*. **A**, SDS-PAGE of coomassie-stained proteins extracted from root of secondary xylem (lane1) and leaves (lane 2) of 4-week-old T_2 WT [*p35S*::*XCP1*] plants. **B**, Evidence of activity of at 26-kD was identified previously as due to XCP1 protein (Funk et al., 2002) using gelatin-impregnated gel as a substrate in the absence of purified soyacystatin N. **C**, Inhibition of proteolytic activity by purified soyacystatin N.

B1.7.3. Soybean cystatin mRNA transcripts were correlated with reproducible phenotypes

A reproducible phenotype consisting of stunted and reduced apical dominance was recovered only from WT [*p35S::cys*^{his}] in T_1 and T_2 generations compared to WT [*pXCP1::cys*^{his}](data not shown). Reverse transcription PCR on five putative mutants was conducted and compared to wild-type-like and WT plants. Figure 1-15 shows that the expected 250-bp soyacystatin fragment was amplified in all five phenotypes whereas it was not amplified in wild-type-like plants. Attempts to purify soyacystatin N using Ni-NTA affinity purification did not succeed, however.

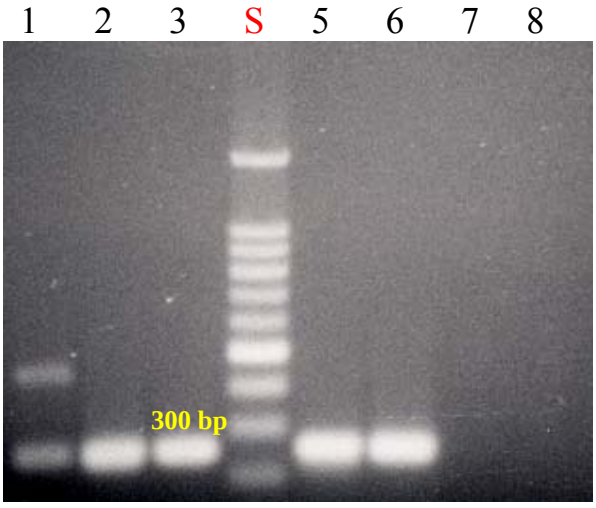


Figure 1-15. RT-PCR of soyacystatin N. 1st cDNA from above-ground tissue of five 4-week-old individual T_2 WT [*p35S::cys*^{his}] plants (lanes 1,2,3,5, and 6) showing the mutant phenotype and a representative of wild-type-like plants (lane 7) was used as a template for the RT PCR reactions using PR54 and PR55 internal to *cys*^{his} coding region. Lane S, the 100-bp DNA ladder. Lane 8, RT-PCR without template. Necessary positive and negative control reactions for PCR and RT-PCR and assessing genomic contamination were also conducted at the time (not shown). The number in the gel points to the 300-bp band of the molecular marker.

Discussion

The proteolytic activity of xylem cysteine protease 2 could be involved in routine protein turnover in tracheary elements, in defense against insects that feed specifically on xylem, or in producing intracellular molecular signals that initiate or assist in maintaining a developmental process in *Arabidopsis*. These functions can be envisioned for XCP2 because other cysteine proteases are involved in protein turnover in germinating seeds of rice, or due to its high homology to papain that was found to be toxic to insects, or due to the potential for proteases to initiate events leading to tracheary element cell death.

In the current investigation, we have utilized T-DNA insertional mutagenesis to recover a loss-of-function mutant of XCP2 from which XCP2 function could be inferred. T-DNA insertion in the coding region and sometimes in intronic regions of any gene is likely to abolish the expression of that gene. That was true in the homozygous line that was derived from SALK_010938 where the T-DNA insertion was mapped to intron 2 of *XCP2* (Fig. 1-6). XCP2 protein was not detected by western analysis (Fig. 1-7). However, the absence of XCP2 expression yielded phenotype similar to WT plants in the laboratory. Further, growing the homozygous line SALK_010938-2 under stress conditions did not induce any aberrant phenotypes compared to WT. Initially, it was thought that XCP2 could be involved in protein turnover in the roots where it is highly expressed (C. Zhao and E. Beers, unpublished) where the accumulation of protein degradation products (organic nitrogen in form of NH_4^+) could modulate developmental processes in plants, as it is known that gene expression can be controlled by endogenous levels of nitrate (Wang et al., 2000).

Another approach to recover a loss-of-function mutant was through the silencing of XCP2 at the post-transcriptional level. We have first demonstrated that the ds RNA vector pFGC5941 carrying the 500-bp inverted repeat of GUS effectively silenced GUS expression that was directed by a different binary vector. Utilizing pFGC5941 carrying a 513-bp inverted repeat of XCP2 produced transgenic plants that were similar to WT, although RT-PCR was not performed to confirm absence of *XCP2* transcripts. The reverse biochemical approach utilizing soyacystatin N to inhibit cysteine proteases *in planta* including XCP2 yielded phenotypes for plants expressing the transcript of soyacystatin N (Fig. 1-15) but only for T2 WT [*p35S::cys*^{his}] plants. The role of cysteine protease inihibtors is critical to plant growth and its defense against pathogens in general. Further analysis of the cystatin N-expressing plants may reveal a function of XCP2 or other cysteine proteases in *Arabidopsis*.

Chapter II

Transcriptional Regulation of *Arabidopsis* XCP 1

Abstract

A functional water-conducting system, the xylem, is required to sustain plant growth and development. In the secondary tissues of vascular plants, the differentiation of water-conducting tracheary elements is the result of orchestrated regulation of vascular cambium formation, initiation of xylem differentiation, cell expansion, secondary wall thickening, programmed cell death (PCD) and cellular autolysis. Tracheary element PCD and autolysis are probably dependent on the activity of hydrolytic enzymes such as XCP 1 known to be expressed in tracheary elements during the latter stages of differentiation. Information on molecular signals triggering the expression of genes involved in developmentally programmed cell death is scarce and mutants that affect the PCD and autolytic phases of tracheary element differentiation are lacking. Identification of regulatory regions in the *XCP1* promoter would lead to a better understanding of the regulation of final phases of tracheary element differentiation. In stable transgenic lines of *Arabidopsis*, a 0.578 kb upstream region of *XCP1* (-533 bp/+45) drives expression of the *β*-glucuronidase reporter enzyme, GUS, in tracheary elements. Transcriptional fusions of 5', 5'-3', and internal deletions of the 0.578 kb with the *GUS* coding region were stably integrated in *Arabidopsis* via *Agrobacterium*-mediated transformation. Qualitative and quantitative assessments of GUS activity show that a 237-bp internal region is able to drive GUS in a tracheary element-specific manner in *Arabidopsis*. A 25-bp deletion at 3' end of this region abolishes GUS expression, indicating that this region is necessary for tracheary element-specific GUS expression. The 237-bp region served as bait in a yeast one-hybrid analysis to identify *Arabidopsis* proteins that could interact with DNA elements in the *XCP1* promoter. The ds cDNA molecules generated from four-week old *Arabidopsis* root-hypocotyls served as the library for the one-hybrid screen. Screening of $\approx 10^6$ yeast colonies retrieved 109 putative positive interactions, which included a potential transcriptional regulator, indole aceticacid-induced protein 8 (IAA8). Common, predicted *cis*-elements in *pXCP1* and *pLjXCP1* were predicted by Genomatix and Athamap. These elements form pyrimidine and GARE (gibberellic acid responsive element) motifs that can potentially bind Dof and Myb transcription factors, respectively. Further manual inspection of the *pXCP1* retrieved an auxin responsive element that potentially binds auxin responsive transcription factors. In an independent effort, attempts to develop a mapping population to isolate upstream regulators of XCP1 expression have not succeeded. A brief discussion on the strategy used is presented.

Key Words: promoter *cis*-elements, *trans*-acting factors, yeast one-hybrid, GA, tracheary elements

Introduction

Plants produce many cell types throughout their life cycle, among which are the water-conducting TEs of the xylem. Having undergone developmentally programmed cell death (PCD), functional TEs are in essence cell walls that no longer contain protoplasts. Initiation, progression, and termination of TE formation are influenced by hormones especially auxins and cytokinins (Kuriyama and Fukuda, 2001), tissue-specific expression of *trans*-acting factors (Karlsson, 2003), and the activity of cysteine proteases, and ubiquitin-dependent proteolysis (Woffenden et al., 1998). TEs also provide mechanical support to vascular plants due to secondary cell wall thickening and cell wall lignification. Lignin is essential for mechanical support because of its rigidity. It also facilitates water transport because of its hydrophobicity and impermeability (Karlsson, 2003). Understanding cellular mechanisms leading to xylem differentiation is important because it will facilitate genetic engineering approaches for introducing quantitative and qualitative changes in wood (secondary xylem) and provide valuable insights on the complex networks leading to the development of vascular tissues.

By definition, hormones are organic compounds produced by one tissue and transported into another tissue where they induce downstream physiological responses (Karlsson, 2003). In their review, Kuriyama and Fukuda (2001) discussed the effects of several hormones and their mode of regulatory action on the induction and progression of TEs. Auxin is an inducer of TE differentiation and the manipulation of its endogenous levels affects the amount of TEs and xylem tissues (Sachs, 2000). Cytokinin is also implicated in promoting vascular tissue development in *Arabidopsis* (Rupp et al., 1999) and inducing TE formation in *zinnia* suspension culture (Aloni, 1987). While auxin and cytokinin are well-established promoters of xylem differentiation, roles played by other hormones are not well-described. Gibberellins (GAs), for example, appear to be involved in elongation of TE cells but not in their induction. However, a gain-of-function study (Eriksson et al., 2000) using transgenic hybrid aspen overexpressing a gene encoding GA_{20} oxidase, a key enzyme in GA biosynthesis, showed that GAs promote xylem cell differentiation. Ethylene-dependent TE formation has not been reported despite the abundance of ethylene-signaling mutants. It is known, however, that ethylene has an impact on lignification by initiating a cascade of events that eventually induce expression of lignification genes (Hennion et al., 1992). Söderman et al. (1999) linked ABA to the modulation of the activity of *Arabidopsis thaliana* homeobox 6, ATHB6, which is expressed

in the vascular bundles. Brassinosteroids have been implicated in TE differentiation in genetic and biochemical experiments. While mutants defective in BRs synthesis show xylem defects (e.g. decreased xylem cells compared to wild type), biochemical inhibition of BR synthesis by BR synthesis specific inhibitor such as uniconazole suppressed TE formation (Yamamoto et al., 1997) and treatment with brassinazole resulted in increased phloem differentiation and reduced xylem differentiation in cress *Lepidium sativum* (Nagata et al., 2001; review Müssig and Altman, 2003). In *zinnia*, PSK-α, a novel cell proliferation agent isolated from cultured asparagus cells, induced TE formation in media suppressive to TE-formation (Matsubayashi et al., 1999).

In vivo transcriptional regulation involves the collaboration of factors which facilitate chromatin remodeling, DNA unwinding and reforming, initiation of basal transcriptional machinery, and recruitment of specific DNA-binding proteins to DNA regulatory regions. More specifically, gene expression at the transcriptional level is mainly facilitated by DNA regions located 5' upstream of gene coding regions and less likely by non-coding intronic or untranslated regions. DNA-binding proteins can act as activators or repressors of gene expression. They can be altered through spatial and temporal multisite phosphorylation of their active domains. Thus, a given transcription factor can be involved in multiple cellular processes in several tissues. DNA-protein interactions can also be restricted through the differential expression of transcription factors and the genes they control. Specificity of DNA-protein interaction depends on core nucleotide sequences in the DNA regulatory region of the gene and the positions of particular amino acids in the protein active domain (for a review see Neidle, 2002).

There are four major functional categories of transcription factors. These are helix-loop-helix, basic leucine zipper, high-mobility group, and zinc fingers. Helix-loop-helix and zinc fingers will be emphasized here because of their reported involvement in various stages of vascular differentiation. Myb transcription factors (MTFs) belong to the helix-loop-helix functional group of transcription factors. Their DNA binding domain is conserved among the group members and is located in the N-terminal region. The MTF is further divided into three subgroups based on the number of imperfect repeats of tryptophan residues found in their DNA-binding domain: Myb1R, MybR2R3, and Myb3R (Stracke et al., 2001). The R2R3 subgroup is the largest of the three with approximately 125 members in *Arabidopsis* alone. Members of this group have been correlated with

regulation of phenylpropanoid pathway activity (Borevitz et al., 2000), cell fate determination (Lee and Schiefelbein., 1999), and cell death in the hypersensitive response (Lee et al., 2001). Zinc finger transcription factors comprise several subclasses such as Dof and WRKY transcription factors. Dof factors have been implicated in binding to pyrimidine boxes such as CTTT/AAAG to modulate gibberellin-regulated cysteine protease expression in rice (Washio, 2001). Auxin responsive factors, ARFs, contain domains for facilitating protein-protein and DNA-protein interactions. ARFs belong to the helix-loop-helix family of transcription factors and they dimerize specifically with members of the Aux/IAA family of short-lived auxin responsive proteins. There are currently 22 known members of ARFs or ARF-like protein in *Arabidopsis* (AGRIS). Through their conserved domains III and IV, ARFs and Aux/IAA can form homo- and hetero-dimers. Given the number of known ARFs and Aux/IAAs in *Arabidopsis*, the combinatorial potential for unique dimerization pairs is very large. ARFs are likely to exert their regulatory function through the TGTCTC DNA *cis*-element (Ulmasov et al., 1997) found common in most genes regulated by auxin.

Identification of novel genes that affect the regulation of vascular differentiation can be accomplished using reverse and forward genetic approaches augmented with *cis*-element prediction computer programs. A reverse genetic approach includes selecting a DNA sequence and mutating it with the intent of inducing a phenotype whereas a forward genetic approach aims at identifying a sequence that underlies a specific mutant phenotype (Peters et al., 2003). Promoter deletion analysis where upstream sequences of gene coding regions are mutated have been a classical reverse genetic approach aiming to identify DNA sequences that alter the expression of the corresponding gene. By combining deletion analysis with the well-defined yeast genetic system (review by Forsburg, 2001), DNA-protein interactions can be defined. The concept of the yeast one-hybrid screen used in the study is presented in the supplementary Figure (S4).

Map-based cloning (MBC) or positional cloning is a forward genetic approach that many years ago was regarded as a labor intensive and time-consuming approach to cloning genes. The advantages of MBC is its universality (i.e. species-independent), generation of large spectrum of mutations compared to T-DNA tagging, and the possibility of obtaining second site mutations which allows complementation tests to take place (Peters et al., 2003). Currently, molecular marker-based technology makes MBC a rapid method for isolating genes from a segregating population of mutant

phenotypes. MBC in *Arabidopsis* is especially useful because of the availability of reverse genetic resources such as transcript profiling and the increasing number of T-DNA mutations covering, eventually, the entire genome. Furthermore, molecular markers such as SSLP (Lukowitz et al., 2000), RFLP, RAPD, SSR, and AFLP are saturating the genome of *Arabidopsis* allowing rapid identification and cloning of the gene of interest. All of the above resources can be very useful once a mutant phenotype is identified.

Physical and computer-based sciences have expanded their investigations to include biology leading to the fields of biophysics and bioinformatics, respectively. Bioinformatics has facilitated reliable predictions of putative DNA-binding elements and their associated DNA-binding proteins. Programs were developed based on common *cis*-elements clusters (Frith et al., 2001; Frith et al., 2003), transcription factor synergism (Hannenhalli and Levy, 2002), evolutionary relationships (Sumiyama et al., 2001), and orthologous gene clustering (Jegga et al., 2002). The OrthParaMap program (Cannon and Young, 2003) was developed and tested successfully on various organisms including *Arabidopsis*. The online programs Genomatix (www.genomatix.de) and the *Arabidopsis*-specific online server AthaMap (Steffens et al., 2004) are used in our analysis. The Genomatix resource is based on datasets developed formerly by promoter trapping experiments in *Arabidopsis*. It is currently integrating invaluable tools such as the ability to use DNA microarray results directly for promoter mining and *cis*-element predictions. AthaMap is used because the program is likely to retrieve common elements that are located in putative regulatory regions such as introns and promoters but not exons and predicted untranslated regions. The Ohio State University developed AGRIS, *Arabidopsis* Gene Regulatory Information Server, which categorizes all known *Arabidopsis trans*-acting factors into distinct families (Davuluri et al., 2003). Complete or near-complete eukaryotic genome sequences of several models such as *Arabidopsis thaliana*, yeast (*Saccharomyces cerevisiae*), humans (*Homo sapiens*), nematode (*Caenorhabditis elegans*), fruit fly (*Drosophila melanogaster*), rice (*Oryza sativa*), and the model legume (*Lotus japonicus*) open the door for gaining insights not only into the complex nature of a particular organism but also into how one organism can relate to other organisms evolutionarily (for a recent review on comparative genomics see Venter and Botha, 2004).

The development of the *in vitro* zinnia suspension culture marked a breakthrough for investigating TE formation. Based on results from the zinnia system, progression of TE formation can be divided into four major phases (Oh et al. 2003):1, the division phase where xylem cell precursors continue to divide; 2, the expansion phase where cells expand to their final size; 3, the maturation phase when lignification and secondary cell wall synthesis occur; and 4, death and autolysis phase of protoplast degeneration. Due to the *zinnia* system's efficiency and simplicity, numerous genes involved in TE formation have been identified and characterized. However, identification of genes involved in TE formation in a genetically tractable model such as *Arabidopsis* would facilitate forward and reverse genetic investigations for the deciphering of TE regulatory networks. Lignification and cellulose synthesis have received considerable attention from researchers due to the economic importance of these two compounds in the forest and paper industries; early and late stages of TE formation have received less attention. As a result, most of the genes involved in lignin and cellulose biosynthesis (Karlsson, 2003) have been identified and partially characterized whereas genes involved in the final induction of TE formation and in critical autolytic phase of TE formation have not. Information on how lignin and cellulose biosynthesis genes are regulated may yield links to conserved mechanisms by which other genes in TE formation are regulated.

Cellulose and Hemicellulose Genes

The *Arabidopsis* Irregular Xylem (IRX) genes, IRX1, IRX3, and IRX5, have been cloned and were found to correspond to the cellulose synthases CesA8 (Turner and Somerville, 1997), CesA7 (Taylor et al., 2000; Taylor et al., 1999; Taylor et al., 2003), and CesA4 (Gardiner et al., 2003), respectively. The *irx* mutant phenotypes were first identified through a mutant screen where defects in the xylem formation were correlated with an eight fold decrease in cellulose content in mature inflorescence stems (Turner and Somerville, 1997). To our knowledge, regulation of these genes has not been reported.

Two *Pinus taeda* xylem genes, *PtX3H6* and P*tX14A9* are expressed in differentiating xylem. Based on their predicted amino acid sequence, they belong to arabinogalactan proteins which play roles in cell-cell communications and cell wall synthesis (Loopstra and Sederoff, 1995). The

presence of homologous sequences in *Arabidopsis*, rice, pine, and poplar suggests that the function of arabinogalactas is conserved across monocots, dicots, gymnosperms, and angiosperms. Although they have been implicated in secondary cell wall formation in loblolly pine, their regulation remains elusive (Zhang et al., 2003). Regulatory DNA regions were suspected in the 3' untranslated regions of these genes. However, *cis*-element identification was not attempted (No et al., 2000).

Phenylpropanoid pathway genes for lignification and their possible regulation by MTFs

A conserved AC-rich region has been found in the promoters of many phenylpropanoid pathway genes leading to lignification suggesting that the genes are regulated by MTFs with specific binding affinity for the AC-region. The *Eucalyptus gunnii* cinnamyl coA reductase, CCR, is expressed in a vascular tissue undergoing lignification. Based on deletion analysis and electrophilic mobility shift assay, Lacombe et al. (2000) showed an interaction between a critical AC-element and a myb transcription factor. Lauvergeat et al. (2002) also showed that the promoter of cinnamyl alcohol dehydrogenase, which is involved in lignin biosynthesis, contained a critical AC-rich region. An AC-rich region, CCAACCCC, in the 4-coumarate: coenzyme A ligase, 4-CL, was implicated in *4-CL* xylem-specific expression (Hauffe et al., 1993). The *4-CL* promoter was shown to contain positive *cis*-elements that directed *4-CL* expression in the xylem and negative *cis*-elements that suppressed its expression in the phloem (Hauffe et al., 1993). Similarly, *Phenylalanine Ammonia Lyase 2*, *PAL2*, promoter contains positive and negative regulatory *cis*-elements. Among these elements is the positive AC-rich element which specified the expression of *PAL2* in xylem and suppressed its expression in the phloem. Gel retardation assays showed that the AC-rich region interacted with a xylem-specific AC-binding factor and that a heptamer of the AC-rich region fused to a minimal 35S promoter specified xylem expression in transgenic tobacco (Seguin et al., 1997). *Pinus taeda* MYB4 is a pine MTF that was cloned from a cDNA library derived from differentiating cells undergoing lignification. Yeast-one hybrid assays showed that PtMyb4 was able to activate transcription using an AC-rich element as bait (Patzlaff et al., 2003).

Vascular differentiation genes and their possible regulation by ARFs and MTFs

Whereas lignification and cellulose biosynthesis events occur midway through TE formation, early events are initiated by hormones and other signaling molecules (discussed earlier) that impinge

upon transcription factors specific to TE differentiation. *IAA8*, an indole acetic acid responsive gene, is a transcriptional regulator involved in early events of TE differentiation (Groover et al., 2003). *IAA8* belongs to Aux/IAA family whose members are known to bind auxin responsive factors, ARFs (Guilfoyle et al., 1998a; 1998b). Genomatix analysis of the putative promoter of IAA8 revealed AC-rich regions distributed throughout the sequence suggesting possible regulation by MTFs. Another group of transcription factors that are implicated in vascular differentiation are *Arabidopsis thaliana* homeobox proteins, ATHBs, which belong to the leucine-zipper type of transcription factors. In a recent report, several ATHBs were found to be highly expressed in the xylem compared to the bark (Oh et al., 2003). Promoters of *ATHB8* (Baima et al., 1995), *ATHB-15* (Ohashi-Ito et al., 2003), *ATHB-9* and *ATHB-14* (Oh et al., 2003) were analyzed using PLACE, a cis-element database. Analysis showed that myb-binding elements are present in the myb-promoters implicating autoregulation and myb-myb interactions as possible mechanisms to direct specific gene expression. The cysteine proteseses XCP1, the subject of this study, and XCP2 have been cloned from the xylem of *Arabidopsis* (Zhao et al., 2000) and were found to be expressed in the final stages of tracheary element differentiation (Funk et al., 2002). Their promoters are predicted to contain the TGTCTC and AC-rich elements for binding to ARFs and MTFs, respectively. In summary, consdering the current information about the promoters of IAA8, the ATHBs, and XCP1/XCP2, auxin- and myb-mediated transcriptional regulation may cooperate to regulate *XCP1* and XCP2 expression reflecting a well-documented theme fir transcriptional regulationof TE genes.

Materials and Methods

A2.1. Plant Material

Arabidopsis thaliana ecotype Columbia maintained in our laboratory was used for all transformation experiments with deletion constructs. Seeds of *Lotus japonicus* accession # Gifu B-129 were kindly provided by Dr. Shusei Sato (Kazusa DNA Research Institute. 2-6-7 Kazusa-Kamatari, Kisarazy, Chiba 292-0818, Japan]

A2.2. PCR-based Mutagenesis of *XCP1* Promoter Region

A2.2.1. Construction of Recipient Binary Vector P2043

To clone and identify transgenic plants with 5' unidirectional deletions and 5'-3' bidirectional deletions and their appropriate controls, pBI121 binary vector (Chen et al., 2003) was modified to contain the herbicide resistant marker to facilitate selection of primary plant transformants. Pnos-*bar*-Tnos fragment was released from the minibinary vector pCB203 (Xiang et al., 1999) by *Hind*III digestion followed by a filling reaction via Klenow large fragment and then digestion with *Bam*HI. This fragment was subcloned into a compatibly digested pGEM. This new plasmid was then digested at the unique *Bam*HI site and the linearized vector was then digested with *Eco*RI releasing the *Eco*RI-*Eco*RI Pnos-*bar*-Tnos fragment for cloning into pBI121. Restrictions digests verified the orientation of the herbicide resistant gene. This binary vector contained *Hind*III and *Bam*HI restriction sites suitable for downstream cloning.

A2.2.2. Construction of Recipient Binary Vector P2071

P2043 was further modified as follows. The binary vector pFGC5941 was used as a donor of a *Hind*III/*Bam*HI fragment containing restrictions sites *Hind*III, *Sph*I, *Pst*I, *Apa*I, *Xma*I, *Pac*I, *Xba*I, *Avr*II, and *Bam*HI, which was subcloned into *Hind*III-*Bam*HI digested P2043. The new construct was digested with *Hind*III-*Spe*I followed by filling in reaction by Klenow large fragment and intramolecular religation of the vector. The new binary vector contained the *Xma*I-*Pac*I-*Xba*I and *Bam*HI restriction sites.

A2.2.3. Cloning of full length *pXCP1* in P2071 (Full^{578}/P1036)

Full length *XCP1* promoter was defined as the 578 bp upstream region of *XCP1* gene (AT4g35350). PR07 (GCACTAGTGTGTTTGCACTTTGCAGG) and PR05 (GGATCCCCAAATTTGTTCACT) PCR primers were designed to amplify this region from WT genomic DNA. The PCR product was TA-cloned into pGEM vector designated P1003. The full length promoter *XCP1* was released via *Spe*I-*Bam*HI and subcloned into *Xba*I-*Bam*HI digested P2071. Colony PCR and sequencing verified the presence of Full^{578}.

A2.2.4. Cloning of 5' Unidirectional Deletions in P2043 (**5'Δ^{455}**/P1028, **5'Δ^{357}**/P1029, **5'Δ^{313}** P1033, **5'Δ^{181}**/P1031, and **5'Δ^{70}**/P1032) P2043

PR01 (AAGCTTCTCATTTGTACG), PR02 (AAGCTTCAGAAATTTTAAGCCTCTACG), PR03 (AAGCTTGATCCAACCGTGAAGACTCG), PR04 (AAGCTTGCTCTATTCCTCTA-CTCTG) primers were individually used with the antisense PR05 to amplify the 5' unidirectional deletions of *XCP1* promoter using P1003 as a template. All PCR products were TA-cloned into pGEM vector to construct P1004, P1005, P1006, and P1030, respectively. *Hind*III/*Bam*HI digestion on P1003, P1004, P1005, P1006, and P1030 released the **5'Δ^{455}**/P1028, **5'Δ^{357}**/P1029, **5'Δ^{313}** P1033, **5'Δ^{181}**/P1031, and **5'Δ^{70}**/P1032 deletion fragments, respectively. All fragment deletions were successfully cloned into P2043 as *Hind*III-*Bam*HI fragments and their presence was verified by PCR and sequence analysis.

A2.2.5. Cloning of bidirectional deletion in P2043 (Δ^{131}/P1045)

P1036 served as a template to amplify Δ^{131} fragment using the primers PR44 (AAGCTTCAGAAATTTTAAGCCTCTA) and PR45 (GGATCCTTTTTGGTTAAGCCTTTT) to construct plasmid P1043. P1043 was digested with *Hind*III-*Bam*HI to release Δ^{131}. The Δ^{131} was inserted into a compatibly P1028 replacing **5'Δ^{455}** fragment.

A2.2.6. Cloning of 35S minimal promoter-driven five bidirectional deletions in P2043 (Δ^{131}/P1041; Δ^{164}/P1060, Δ^{184}/P1059; Δ^{213} /P1058; Δ ^{237}/P1057)

The binary vector pFGC5941 containing the *p35S* served as a template in PCR reactions to amplify a minimal 35S promoter fragment using primers PR46 (GGATCCCGCAAGACCCTTCCTCTA) and

PR47 (CCCGGGCGTGTTCTCTCCAAATGA). The primers amplified a -46/+8 35S minimal region containing transcription initiation site +1 and an upstream TATA box sufficient for initiating transcription according to Benfey et al. (1990). The PCR product was TA-cloned into pGEM to construct P1040. The 35S minimal region was released as a *Bam*HI-*Xma*I fragment and cloned into compatibly digested P1045 to construct Δ^{131}/P1041. Colony PCR analysis and sequencing confirmed the presence of 35S minimal promoter sequence and the Δ^{131} sequence in the binary vector P2043. Four bidirectional deletions were also constructed. Δ^{164} fragment was amplified using PR44 and PR64 (GGATCCACGCCGGTCTCCGAGTCTTCA) to construct pGEM-based plasmid P1053. Δ^{184} fragment was amplified using PR44 and PR65 (GGATCCGATTAATTTAAAAACTAAGTA) to construct pGEM-based plasmid P1052. Δ^{213} fragment was amplified using PR44 and PR66 (GGATCCCATAGG-ATTGGCTTTGAAGCA) to construct pGEM-based plasmid P1051. Δ^{237} fragment was amplified using PR44 and PR67 (GGATCCCAAGTTGGAGACAAGACA) to construct pGEM-based plasmid P1050. Deletion fragments were released as *Hind*III/*Bam*HI fragments from P1050, P1051, P1052, and P1053 and cloned into *Hind*III-*Bam*HI linearized P1041 to construct Δ^{164}/P1060, Δ^{184}/P1059, Δ^{213} /P1058, and Δ^{237}/P1057, respectively.

A2.2.7. Construction of 35S minimal promoter region in P2043 ($35S^{MP}$)

P1041 served as a template for the release of Δ^{131} as *Hind*III-*Bam*HI fragment. The ends of the binary vector were filled in by Klenow large fragment and religated. Colony PCR and sequence analysis verified the presence of intact minimal 35S promoter region.

A2.2.8. Cloning of two internal deletions in P2071 (IN Δ^{446} /P1046 and INΔ^{385}/P1047)

P1003 served as a template for an inverse PCR reaction using primers PR49 (TGATCCAACCGTGAAGACTCGGA) and PR50 (AAATTGTAACTATTCTTGATAAGTTTT). PCR was conducted using a high fidelity PfuUltra™ High-Fidelity DNA Polymerase (Stratagene, CA). Gradient PCR program was set as follows using Strategene Gradient 40 Robocylcer: 94°C 5' [94°C 1' 51°C-65°C 1' 72°C 3' 28"] x25 72°C 5'. The successful PCR reactions included the following modifications: the increase of [Mg^{++}] by 0.5 mM and the addition of DMSO to a final concentration of 2%, and the annealing temperature to 51°C using only 100 ng of DNA template. Several PCR products resulted from the PCR reaction. The expected size band was gel purified using Quantum Prep® Freeze 'N Squeeze DNA Gel Extraction Spin Column (Bio-Rad) and

quantified on-gel by comparing with 100-bp molecular standards. Phosphorylation of the PCR product was conducted using T4 Polynucloetide Kinase (Stratagene) following Promega's technical bulletin No. 519 to enhance religation of blunt ends generated by Pfu DNA polymerase. Sequencing of several colonies verified the presence of two internal deletions from the same reaction. The first type corresponds to the desired deletion area (IN Δ^{446}) and the second was a larger internal deletion (INΔ^{385}) at 5' end of the PCR product. Both internal deletions were released as SpeI-*Bam*HI fragments and subcloned into compatibly digested P2071.

A2.2.9. Cloning of *Lotus japonicus* promoter in P2043 (*pLjXCP1* / P1039)

A 100 kb genomic sequence of chromosome 1 (gi# 21908002; accession # AP004984.1) of the model legume plant *Lotus japonicus* (Handberg and Stougaard, 1992) was found to contain sequences similar to those of *XCP1*/*XCP2*. Comparative alignment against mRNA/cDNA sequences (Asamizu et al., 2000; Endo et al., 2000; Endo et al., 2002; Perry et al., 2003) revealed three high-score alignments with AV427014.1, AV411315.1, and BU494482.1 in NCBI nucleotide database. The latter was shown to be similar to a cysteine proteinase mRNA. Three exon regions similar to exons 2,3, and 4 of *XCP1*/*XCP2* were mapped. A manual annotation for that region identified putatively the first ATG, exon-intron boundaries, and stop codon. A final putative protein sequence was constructed and compared to XCP1 amino acid sequence. A 2kb upstream region, designated *pLjXCP1*, was amplified by PR42 (GGATCCGGCAATTGTTGTTATTGGAT) and PR43 (AAGCTTAAGGTGGGTTTAAAGCGGTG) using genomic DNA of *Lotus japonicus* accession Gifu B-129 as a template. A *Hind*III-*Bam*HI digest released the 2kb fragment which was gel-purified, quantified, and cloned into compatibly digested P2043 to construct plasmid *pLjXCP1*.

A2.3. Map-based Cloning

A2.3.1 Whole-genome based Chemical Mutagenesis and Development of Mutant Lines

Full length promoter of *XCP1* was previously fused to GUS coding region in pCB302 minibinary vector and the successful integration in *Arabidopsis* nuclear genome was verified through the histochemical staining of TE-specific GUS expression (Kositsup, 2000; Funk et al., 2002). Primary transformants (T_1 generation) were allowed to set seeds and T_2 seeds were collected. Segregation analysis of GUS expression was conducted on T_2, T_3, and T_4 generations from a single plant

descendent to select for single copy number of transgene as well as stable transgenic line with 100% GUS expression. T_4 4-4 was selected due to its 3:1 (GUS^+: GUS^-) segregation ratio at T_3 level. Chemical mutagenesis pilot experiment was conducted to select for optimal concentration of the mutagen ethyl methane sulfonate, EMS. A 0.3% of the mutagen resulted in ~50% mortality rate among treated seeds. Based on weight, approximately 20,000 seeds from T_4 4-4 line were treated with 0.3% EMS for a period of 8 h. Seeds were sown in 100 pots (also referred to as pools) to generate first mutagenized generation, M_1. M_2 seeds were harvested from each pool individually. M_2 seeds were stored in microfuge tubes according to TAIR recommendations. Over 14,000 M_2 plants were screened using GUS histochemical stain (section 1.8.1) to look for altered GUS expression. Positional cloning strategy took into consideration practical recommendations described in Lukowitz et al. (2000) and Malberg (1994).

A2.3.2. Sequence Analysis of Mutant M_3 11-2 Generated by Chemical Mutagenesis

A putative mutant M_2 11-2 was retrieved as a null mutant because of its inability to drive GUS expression. Genomic DNA was isolated from plant F_1-2 and PCR reactions were designed to amplify the GUS coding region as well as a region spanning 5' end of GUS coding region fused to the full length promoter of *XCP1*. PCR primers PR06 and PR07 were used to amplify full length promoter of *XCP1* and approximately 100 bp of 5' GUS coding region whereas PCR primers PR11 and PR12 were used to amplify almost full length GUS coding region with the exception of several base pairs at the 3' end of the fragment. Both PCR products were TA-cloned into pGEM to construct plasmids P1022 and P1019, respectively. Plasmid P1022 was sequenced once from one orientation where as plasmid P1019 was sequenced on both orientation using three sense, PR11, PR28, PR29, and three antisense, PR30, PR31, PR32 primers to sequence over lapping fragments on on both positive and negative strands of GUS coding region. The internal primers used for P1019 sequences were designed based on sequences obtained from a sequencing reaction covering the preceding region (i.e. primer walking). Sequence analysis through multiple alignments and sequence management-based approach using DNA Star program showed that no mutations occured within those regions.

A2.4. Yeast One-hybrid Analysis

A2.4.1. Construction of Bait Reporter Vector, pHIS2-Δ^{237}

PR51 (CCCGGGTTCAGAAATTTTAAGCCT) and PR52 (GAGCTCCAAGTTGGAGACAA-GACA) were designed to amplify a 237-bp DNA bait fragment from *XCP1* promoter using plasmid P1003 as a template. PCR product was quantified on gel and gel-purified using Quantum Prep® Squeeze and Freeze Columns. One hundred ng of insert and vector in a ratio of 3:1 was used to construct P1048, a pGEM-plasmid containing the bait DNA. Sequencing using SP6 and T7 was conducted to verify the presence of the bait DNA fragment and to assess the fidelity of DNA amplification. The insert was subsequently released as *Sac*I-*Xma*I fragment and ligated (10 ng) into 90 ng of compatibly digested 7.2 kb pHIS2 reporter vector to construct P1049. Sequencing reactions using PR51, PR52 and PR53 were conducted to verify the presence of the 237-bp bait DNA in the pHIS2 background and absence of mutations in the DNA.

A2.4.2. Construction of 1st cDNA and double-stranded cDNA Library

Total RNA was extracted according to Qiagen Plant RNA minikit using 50 mg fresh weight of roots of 4-week-old *Arabidopsis* plants grown as described by Lev-Yadun, (1994). Total RNA quality, quantity, and stability were assessed according to Qiagen's specification. First strand cDNA was synthesized using 2 µg of total RNA by Advantage 2 PCR kit using Oligo (dT) primer from Clontech. To construct ds cDNA compatible for *in vivo* homologous recombination in yeast, 2 µl of 1st strand cDNA served as a template for Long Distance PCR using the following program: [94°C 1' 1x; [94°C 30' 68°C 6'*] 25x; 68°C 5'] where * indicates a 5-sec extension time per cycle. Advantage 2 PCR kit reagents (Clontech) were used for this purpose. The ds cDNA was purified with BD CHROMASPIN™ TE-400 Columns and quality was assessed by running a 7-µl aliquot on 1.2% ethidium-bromide impregnated agarose gels.

A2.4.3. Assessment, Preparation of Yeast Cells for Yeast One-hybrid Screen, and the Triple Transformation of Yeast Competent Cells

Y187 strain nutritional requirements for optimal growth, preparation of yeast competent cells, and pilot transformation experiments were conducted as described by Clontech BD Matchmaker™ Library Construction and Screening Kit. Modified library-scale yeast triple transformation with ds

cDNA molecules, P1049, and the linearized pGADT7-rec2 was carried out following the Clontech's catalog C5003-1. Thirteen 150-mm SD/-leu/-his/-trp/50 µM 3-AT Petri plates were each streaked with 150-µl aliquot of transformed Y187 cells. Plates were incubated in the dark at 30°C in an inverted position and colony formation and growth was observed for 10 days when the experiment was terminated.

A2.4.4. Colony PCR and Sequencing of ds cDNA Inserts in pGADT7-Rec Plasmid

Colonies grown on SD/-leu/-his/-trp supplemented with 50 µM 3-AT were retested on fresh plates. A single colony was used as a template by suspending a few cells in a 5-µl ddH2O aliquot and incubating the suspension in 95°C thermocycler for 5 min. Primers PR70 (CTATTCG-ATGATGAAGATACCCCACCAAACCC) and PR71 (GTGAACTTGCGGGGTTTTTCAGT-ATCTACGAT) at a final concentration of 0.2 µM, dNTPs at a final concentration of 125 µM were added to a final 50-µl Red Taq-based reaction using the following PCR program: 94°C 5' x1; [94°C 1' 62°C 1' 72°C 2.5'] x32; 72°C 10' x1; 15°C 15' x1. Approximately 7 µl were run on 1.2 % ethidium-bromide impregnated agarose gel. PCR products were further purified using NucleoSpin and NucleoTrap Extraction kits. PCR products similar in size were digested by the frequent-cutter *Alu*I to verify their restriction pattern. Purified PCR products were submitted to VBI for sequencing using the internal T7 primer.

A2.5. Quantification of *β*-Glucuronidase Activity in *Arabidopsis*

The quantitative fluorometric GUS assay is based on the hydrolysis of MUG substrate, 4-methylumbelliferyl *β*-D-glucuronide, by *β*-glucuronidase, GUS, present in protein extracts of GUS-transformed plants to produce 4-methylumbelliferone (MU) whose fluorescence is measured at wave length 460nm. A given protein sample was processed as described in Gallagher (1992). Readings were compiled using TKO100 Mini-Fluorometer [Hoefer Scientific Instruments] according to the manual. All readings were converted to pmol MU/minute/µg protein using the calculations presented in Gallagher (1992). Protein quantification was carried using BCA assay as described in the Sigma technical manual.

A2.6. Effects of GA_3 and ABA on T_2 WT [*pXCP1::GUS*]

Preliminary experiments were carried out to test whether the *pXCP1* is responsive to hormonal treatments of GA_3 (50 µM) and ABA (100 µM). Hormonal applications were delivered through an apical drenching method described in Little and MacDonald (2003) of T_2 WT [*pXCP1::GUS*] with 20 µl of hormonal solution mixed with 0.2 µl/ml Silwet L-77. Histochemical GUS stain was conducted 12 h after the application began. Efficiency of drenching was determined through observing typical growth characteristic of plants treated with GA_3 and ABA. For example, ABA applications are expected to produce a stunted growth and delayed flowering compared to non-treated plants and to GA_3-treated plants, which should flower prematurely.

A2.7. Computer-Assisted Analysis and Presentation

The web-based interface softwares Genomatix and AthaMap were used to identify putative *cis*-elements within the promoter region. On-line resources used for functional genomics were the National Center for Biotechnology Information (NCBI), The *Arabidopsis* Information Resource (TAIR), MEROPS peptidase database, *Arabidopsis* Gene Regulatory Information Server (AGRIS), The Institute for Genomic Research (TIGR), Salk Institute Genomic Analysis Laboratory (SIGnAL). DNASTAR program was used for sequence management. Digital images were captured by Sony MPEGMOVIE*EX* DSC-S85 camera and DNASTAR sequence maps were digitally scanned by hpScanjet and stored as TIFF images. Microscopic images were photographed using Kodak T-160 film via Steroscope SR (ZEISS) microscope, developed into slides and digitized using DIMAGE Scan Dual III AF-2840 (MINOLTA). All digital conversions to TIFF images were done through AdopePhotoshop (Adobe Systems Inc., San Jose, CA). All images are exported through Microsoft Windows XP PC system programs, MS Power Point, MS Word, and MS Excel.

Results

B2.1. Gene Structure of *Arabidopsis* XCP 1

XCP 1 (*XCP1*) is located on the acrocentric chromosome 4 of *Arabidopsis thaliana* nuclear genome (AT4g 35350.1; 16810534-16811882) and encodes a 356 amino-acid papain-like cysteine peptidase that was localized to the vacuoles of TEs (Funk et al., 2002). An approximately 700-bp region upstream of the *XCP1* initiator Met comprises an intergenic region that contains putative *cis*-elements required for its TE-specific expression. However, just 578-bp was found to be sufficient to drive TE-specific expression of GUS. This region (-533 to +45) consists of 533 bp upstream of the predicted transcription initiation site (ACCAT) and continues through a 45-bp 5' UTR region. This region terminates just 2 bp upstream of the first endogenous ATG. The predicted TATATAA covers a region from the -32 to -26. Thus, *XCP1* is a typical eukaryotic gene with classic TATA box and transcription initiation site (Fig. 2-1).

B2.2. Specificity of Tracheary Element Expression of Full Length Promoter *XCP1* in *Arabidopsis*

The 578-bp region of *XCP1* promoter (*pXCP1*) was fused with the reporter gene, *GUS,* and stably integrated into *Arabidopsis* plants using two binary vectors, pCB302 (Xiang et al., 1999; Funk et al., 2002) and pBI121 (Chen et al., 2003). The binary vector pBI121 was modified to include different cloning sites and the *bar* herbicide resistance gene to facilitate subcloning of the DNA fragments and subsequent selection of transgenic plants, respectively. In independent studies using the two vectors, TE-specific expression was observed in T_1 primary transformants and T_2 independent transformants containing *pXCP1::GUS* constructs. All of the deletion constructs in this study were cloned into modified pBI121 and transgenic plants were selected by virtue of their ability to survive two applications of the herbicide FINALE. Fig. 3A shows a confocal image depicting GUS activity in a root sample of a *pXCP1::GUS* plant using a non-destructive ImaGene Green fluorescent dye. The image demonstrates clearly the restricted expression of GUS in differentiating TEs.

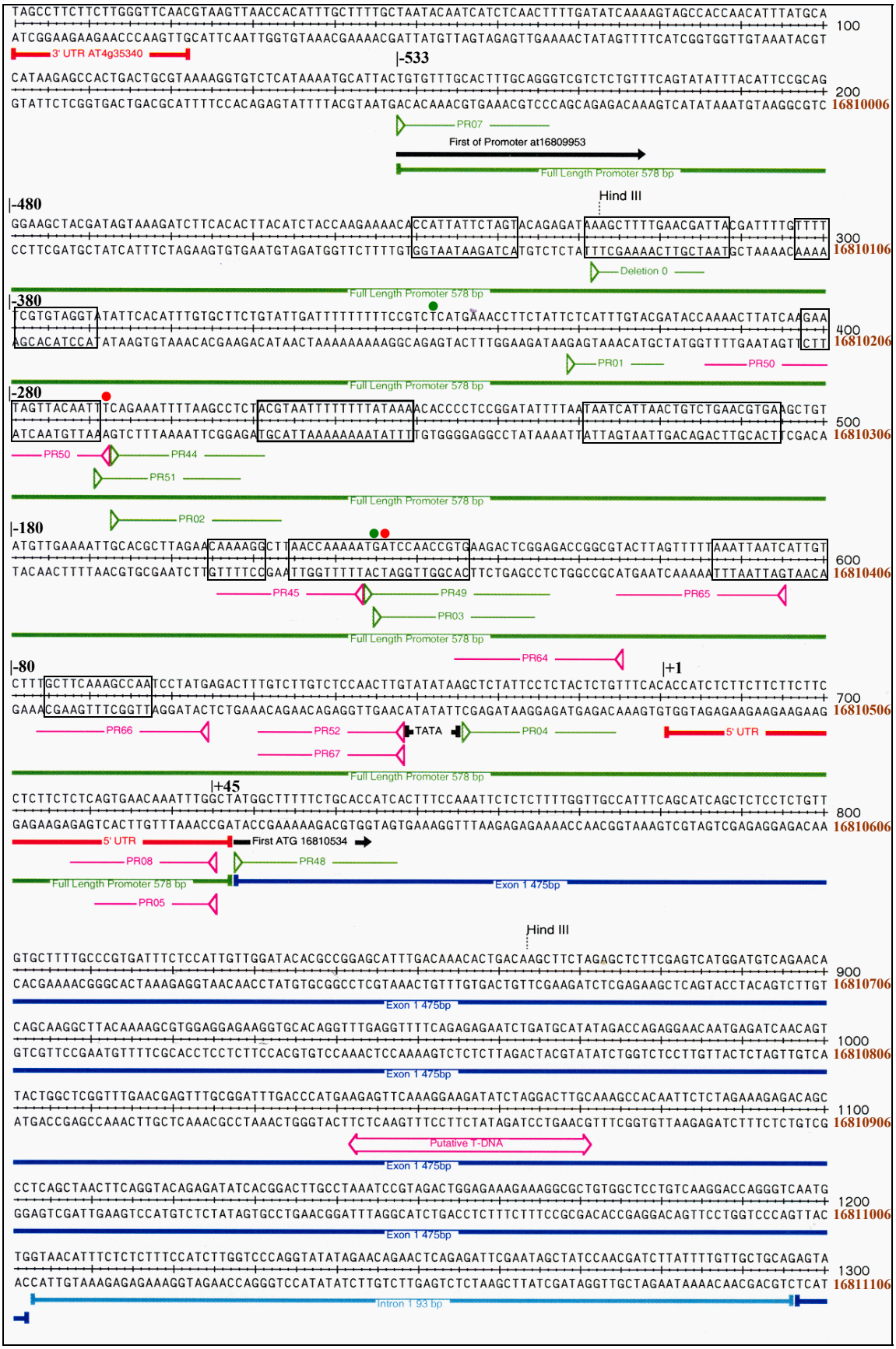

TAGCCTTCTTCTTGGGTTCAACGTAAGTTAACCACATTTGCTTTTGCTAATACAATCATCTCAACTTTTGATATCAAAAGTAGCCACCAACATTTATGCA
ATCGGAAGAAGAACCCAAGTTGCATTCAATTGGTGTAAACGAAAACGATTATGTTAGTAGAGTTGAAAACTATAGTTTTCATCGGTGGTTGTAAATACGT
100
3' UTR AT4g35340
|-533
CATAAGAGCCACTGACTGCGTAAAAGGTGTCTCATAAAATGCATTACTGTGTTTGCACTTTGCAGGGTCGTCTCTGTTTCAGTATATTTACATTCCGCAG
GTATTCTCGGTGACTGACGCATTTTCCACAGAGTATTTTACGTAATGACACAAACGTGAAACGTCCCAGCAGAGACAAAGTCATATAAATGTAAGGCGTC
200
16810006
PR07
First of Promoter at16809953
Full Length Promoter 578 bp
Hind III
|-480
GGAAGCTACGATAGTAAAGATCTTCACACTTACATCTACCAAGAAAACACCATTATTCTAGTACAGAGATAAAGCTTTTGAACGATTACGATTTTGTTTT
CCTTCGATGCTATCATTTCTAGAAGTGTGAATGTAGATGGTTCTTTTGTGGTAATAAGATCATGTCTCTATTTCGAAAACTTGCTAATGCTAAAACAAAA
300
16810106
Deletion 0
Full Length Promoter 578 bp
|-380
TCGTGTAGGTATATTCACATTTGTGCTTCTGTATTGATTTTTTTTTCCGTCTCATGAAACCTTCTATTCTCATTTGTACGATACCAAAACTTATCAAGAA
AGCACATCCATATAAGTGTAAACACGAAGACATAACTAAAAAAAAAGGCAGAGTACTTTGGAAGATAAGAGTAAACATGCTATGGTTTTGAATAGTTCTT
400
16810206
PR01
PR50
Full Length Promoter 578 bp
|-280
TAGTTACAATTTCAGAAATTTTAAGCCTCTACGTAATTTTTTTTATAAAACACCCCTCCGGATATTTTAATAATCATTAACTGTCTGAACGTGAAGCTGT
ATCAATGTTAAAGTCTTTAAAATTCGGAGATGCATTAAAAAAAATATTTTGTGGGGAGGCCTATAAAATTATTAGTAATTGACAGACTTGCACTTCGACA
500
16810306
PR50
PR44
PR51
Full Length Promoter 578 bp
PR02
|-180
ATGTTGAAAATTGCACGCTTAGAACAAAAGGCTTAACCAAAAATGATCCAACCGTGAAGACTCGGAGACCGGCGTACTTAGTTTTTAAATTAATCATTGT
TACAACTTTTAACGTGCGAATCTTGTTTTCCGAATTGGTTTTTACTAGGTTGGCACTTCTGAGCCTCTGGCCGCATGAATCAAAAATTTAATTAGTAACA
600
16810406
PR45
PR49
PR65
PR03
Full Length Promoter 578 bp
PR64
|-80
|+1
CTTTGCTTCAAAGCCAATCCTATGAGACTTTGTCTTGTCTCCAACTTGTATATAAGCTCTATTCCTCTACTCTGTTTCACACCATCTCTTCTTCTTCTTC
GAAACGAAGTTTCGGTTAGGATACTCTGAAACAGAACAGAGGTTGAACATATATTCGAGATAAGGAGATGAGACAAAGTGTGGTAGAGAAGAAGAAGAAG
700
16810506
PR66
PR52
TATA
PR04
5' UTR
PR67
Full Length Promoter 578 bp
|+45
CTCTTCTCTCAGTGAACAAATTTGGCTATGGCTTTTTCTGCACCATCACTTTCCAAATTCTCTCTTTTGGTTGCCATTTCAGCATCAGCTCTCCTCTGTT
GAGAAGAGAGTCACTTGTTTAAACCGATACCGAAAAAGACGTGGTAGTGAAAGGTTTAAGAGAGAAAACCAACGGTAAAGTCGTAGTCGAGAGGAGACAA
800
16810606
5' UTR
First ATG 16810534
PR08
PR48
Full Length Promoter 578 bp
Exon 1 475bp
PR05
Hind III
GTGCTTTTGCCCGTGATTTCTCCATTGTTGGATACACGCCGGAGCATTTGACAAACACTGACAAGCTTCTAGAGCTCTTCGAGTCATGGATGTCAGAACA
CACGAAAACGGGCACTAAAGAGGTAACAACCTATGTGCGGCCTCGTAAACTGTTTGTGACTGTTCGAAGATCTCGAGAAGCTCAGTACCTACAGTCTTGT
900
16810706
Exon 1 475bp
CAGCAAGGCTTACAAAAGCGTGGAGGAGAAGGTGCACAGGTTTGAGGTTTTCAGAGAGAATCTGATGCATATAGACCAGAGGAACAATGAGATCAACAGT
GTCGTTCCGAATGTTTTCGCACCTCCTCTTCCACGTGTCCAAACTCCAAAAGTCTCTCTTAGACTACGTATATCTGGTCTCCTTGTTACTCTAGTTGTCA
1000
16810806
Exon 1 475bp
TACTGGCTCGGTTTGAACGAGTTTGCGGATTTGACCCATGAAGAGTTCAAAGGAAGATATCTAGGACTTGCAAAGCCACAATTCTCTAGAAAGAGACAGC
ATGACCGAGCCAAACTTGCTCAAACGCCTAAACTGGGTACTTCTCAAGTTTCCTTCTATAGATCCTGAACGTTTCGGTGTTAAGAGATCTTTCTCTGTCG
1100
16810906
Putative T-DNA
Exon 1 475bp
CCTCAGCTAACTTCAGGTACAGAGATATCACGGACTTGCCTAAATCCGTAGACTGGAGAAAGAAAGGCGCTGTGGCTCCTGTCAAGGACCAGGGTCAATG
GGAGTCGATTGAAGTCCATGTCTCTATAGTGCCTGAACGGATTTAGGCATCTGACCTCTTTCTTTCCGCGACACCGAGGACAGTTCCTGGTCCCAGTTAC
1200
16811006
Exon 1 475bp
TGGTAACATTTCTCTCTTTCCATCTTGGTCCCAGGTATATAGAACAGAACTCAGAGATTCGAATAGCTATCCAACGATCTTATTTTGTTGCTGCAGAGTA
ACCATTGTAAAGAGAGAAAGGTAGAACCAGGGTCCATATATCTTGTCTTGAGTCTCTAAGCTTATCGATAGGTTGCTAGAATAAAACAACGACGTCTCAT
1300
16811106
Intron 1 93 bp

Figure 2-1. *XCP1* Genomic Map. Intergenic region upstream of first ATG of *XCP1* (AT4g35350.1) comprises the 578-bp region under analysis. Bold numbers on the right pertain to *XCP1* actual location on the *Arabidopsis* genome. Numbers on the left pertain specifically to the promoter region which uses the putative transcription initiation site as +1 position (i.e., -533 to +45). Exons, introns, UTRs, and promoter region are indicated by dark blue, light blue, red, and green bars, respectively. PCR and sequencing primers are represented by green (sense) and pink (antisense) sided arrows, respectively. Engineered restriction sites are not presented as a part of this sequence (see materials and methods). Boxes indicate regions predicted to contain regulatory *cis*-elements predicted by Genomatix and Athamap software programs. The regions between the two green dots and between the two red dots were deleted in internal deletions constructs (see materials and methods).

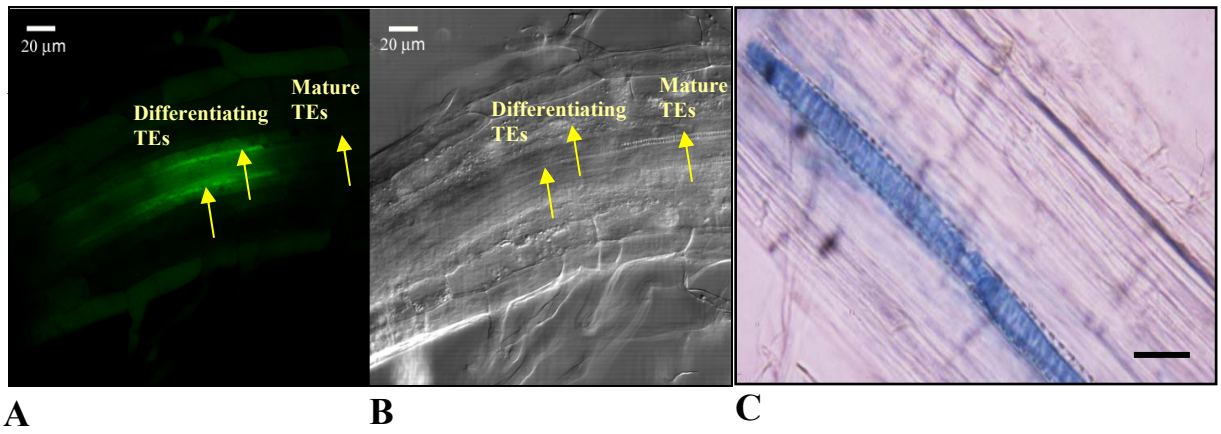
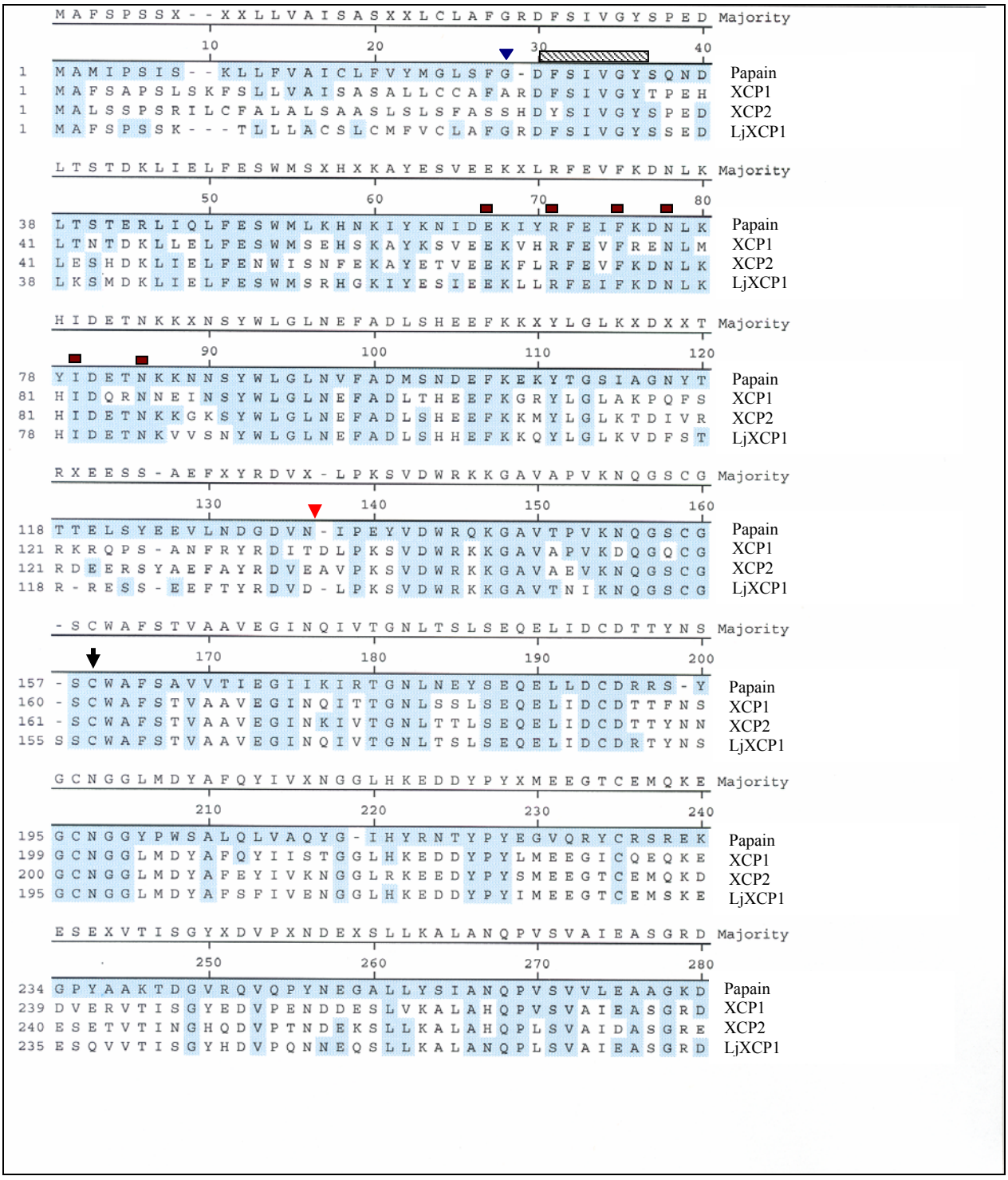


Figure 2-2. Tracheary element-specific GUS expression in *Arabidopsis roots.* **A** and **B,** Confocal images of *pXCP1*-GUS expression using the non-destructive ImaGene Green fluorescent dye (**A**) and showing tracheary elements using transmitted light (**B**). Image is kindly provided by Dr. Candace Haigler at North Carolina State University. GUS activity is limited to differentiating TEs (**A**). **C.** Transmitted light micrograph of a seedling root whole mount showing GUS expression activated by the 2.0 kb upstream region of the putative *XCP1* ortholog from *Lotus japonicas.*Bar = 20 µm

B2.3. Heterologous Expression of a *Lotus japonicus* Putative Promoter in *Arabidopsis*

Orthologous genes that share specific expression patterns may have conserved DNA elements within their promoter regions. A 100 kb region of the model legume *Lotus japonicus* (accession # AP004984.1) containing the suspected *XCP1* ortholog was manually annotated. Based on available expressed sequence tags, intron-exon junctions were predicted and exon splicing was conducted to construct an ORF consisting of 351 amino acids. At the amino acid level, the identity between LjXCP1 and XCP1 was estimated at 72% and is higher than XCP1 homology to XCP2 (68%). LjXCP1, XCP1, XCP2, and papain share the structural features conserved in papain-like cysteine proteases (Fig.2-3): the catalytic dyad His^{25} and Asn^{159} in the mature domain and the ERFNIN motif and the signature DFSIVGY in the prodomain (Funk et al., 2002). Consistent with this high degree of identity at the amino acid level, the 2.0 kb 5' *pLjXCP1* directed GUS expression in TEs (Fig. 2-2C). The level of expression was much higher than that of *pXCP1::GUS* fusions and more comparable to that of *pXCP2::GUS* (data not shown).


MAFSPSSX--XXLLVAISASXXLCLAFGRDFSIVGYSPED Majority
10 20 30 40
1 MAMIPSIS--KLLFVAICLFVYMGLSFG-DFSIVGYSQND Papain
1 MAFSAPSLSKFSLLVAISASALLCCAFARDFSIVGYTPEH XCP1
1 MALSSPSRILCFALALSAASLSLSFASSHDYSIVGYSPED XCP2
1 MAFSPSSK---TLLLACSLCMFVCLAFGRDFSIVGYSSED LjXCP1
LTSTDKLIELFESWMSXHXKAYESVEEKXLRFEVFKDNLK Majority
50 60 70 80
38 LTSTERLIQLFESWMLKHNKIYKNIDEKIYRFEIFKDNLK Papain
41 LTNTDKLLELFESWMSEHSKAYKSVEEKVHRFEVFRENLM XCP1
41 LESHDKLIELFENWISNFEKAYETVEEKFLRFEVFKDNLK XCP2
38 LKSMDKLIELFESWMSRHGKIYESIEEKLLRFEIFKDNLK LjXCP1
HIDETNKKXNSYWLGLNEFADLSHEEFKKXYLGLKXDXXT Majority
90 100 110 120
78 YIDETNKKNNSYWLGLNVFADMSNDEFKEKYTGSIAGNYT Papain
81 HIDQRNNEINSYWLGLNEFADLTHEEFKGRYLGLAKPQFS XCP1
81 HIDETNKKGKSYWLGLNEFADLSHEEFKKMYLGLKTDIVR XCP2
78 HIDETNKVVSNYWLGLNEFADLSHHEFKKQYLGLKVDFST LjXCP1
RXEESS-AEFXYRDVX-LPKSVDWRKKGAVAPVKNQGSCG Majority
130 140 150 160
118 TTELSYEEVLNDGDVN-IPEYVDWRQKGAVTPVKNQGSCG Papain
121 RKRQPS-ANFRYRDITDLPKSVDWRKKGAVAPVKDQGQCG XCP1
121 RDEERSYAEFAYRDVEAVPKSVDWRKKGAVAEVKNQGSCG XCP2
118 R-RESS-EEFTYRDVD-LPKSVDWRKKGAVTNIKNQGSCG LjXCP1
-SCWAFSTVAAVEGINQIVTGNLTSLSEQELIDCDTTYNS Majority
170 180 190 200
157 -SCWAFSAVVTIEGIIKIRTGNLNEYSEQELLDCDRRS-Y Papain
160 -SCWAFSTVAAVEGINQITTGNLSSLSEQELIDCDTTFNS XCP1
161 -SCWAFSTVAAVEGINKIVTGNLTTLSEQELIDCDTTYNN XCP2
155 SSCWAFSTVAAVEGINQIVTGNLTSLSEQELIDCDRTYNS LjXCP1
GCNGGLMDYAFQYIVXNGGLHKEDDYPYXMEEGTCEMQKE Majority
210 220 230 240
195 GCNGGYPWSALQLVAQYG-IHYRNTYPYEGVQRYCRSREK Papain
199 GCNGGLMDYAFQYIISTGGLHKEDDYPYLMEEGICQEQKE XCP1
200 GCNGGLMDYAFEYIVKNGGLRKEEDYPYSMEEGTCEMQKD XCP2
195 GCNGGLMDYAFSFIVENGGLHKEDDYPYIMEEGTCEMSKE LjXCP1
ESEXVTISGYXDVPXNDEXSLLKALANQPVSVAIEASGRD Majority
250 260 270 280
234 GPYAAKTDGVRQVQPYNEGALLYSIANQPVSVVLEAAGKD Papain
239 DVERVTISGYEDVPENDDESLVKALAHQPVSVAIEASGRD XCP1
240 ESETVTINGHQDVPTNDEKSLLKALAHQPLSVAIDASGRE XCP2
235 ESQVVTISGYHDVPQNNEQSLLKALANQPLSVAIEASGRD LjXCP1

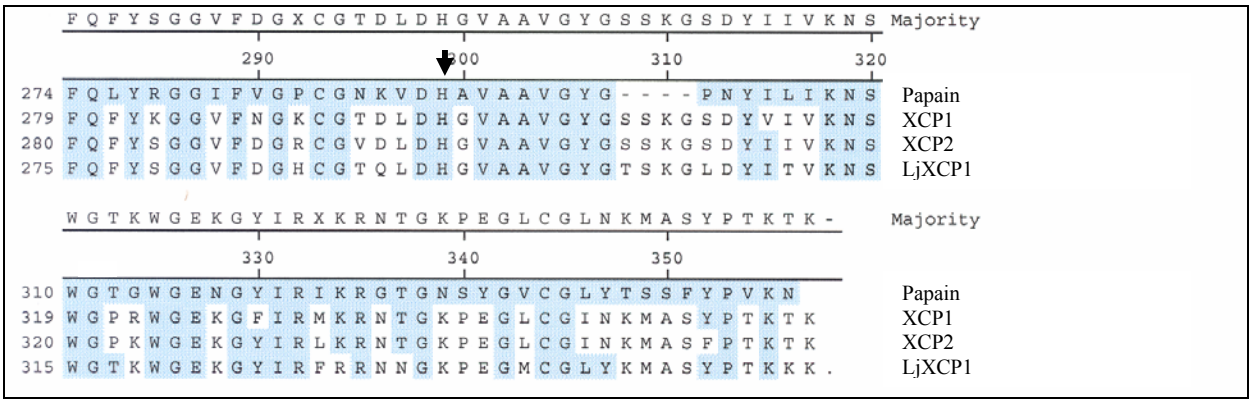


Figure 2-3. LjXCP1 is an ortholog of XCP1 and XCP2. LjXCP1, XCP1, and XCP2 are 351, 355, and 356 amino acid proteins that belong to C1A papain family of cysteine proteases (Beers et al., 2004). Conserved catalytic dyad, cysteine and histidine residues (black arrows), the prodomain ERFNIN motif (dark squares), and papain signature peptide in the prodomain (shaded rectangle) are identified. Predicted cleavage site (blue arrow) between predomain and prodomain and between the prodomain and the mature protein (red arrow) based on the papain sequence are shown. Blue-shaded areas indicate identity. Dashes indicate gaps in alignment.

B2.4. Computer-based Analysis of Promoter Region of *XCP1* and Its Ortholog from *Lotus japonicus*

Since both the 578-bp *pXCP1* and the 2.0 kb *pLjXCP1* drive GUS expression in a TE-specific manner in *Arabidopsis,* both sequences are likely to contain conserved *cis*-elements required for TE-specific expression. A comparative analysis of both sequences using Genomatix and AthaMap programs was conducted to identify putative conserved motifs. Both programs yielded similar prediction of elements and as well as their distribution *pXCP1* sequences. The sequence of *L. japonicus*, on the other hand, was only analyzed with Genomatix since Athamap is an *Arabidopsis*-specific program. Table 2-1 lists DNA elements found conserved between the 237-bp region of *pXCP1* and the 2 kb region of *pLjXCP1*. Fig. 5 illustrates the distribution of conserved DNA motifs between the full length *pXCP1* and *pLjXCP1* using *XCP1* promoter sequences as a template. Tables 2-2 and 2-3 provide a description of transcription factors that are predicted to bind these DNA elements. Manual inspection of the *pXCP1* region also reveals an auxin responsive element (ARE), TGTCTC (Fig.2-4), which is predicted to bind ARFs. Dut to its presence as TGTCTA in *pLjXCP1* (5' 798-803 3'), it was not retrieved in the Genomatix analysis. The linear distribution of the conserved elements (Table 2-1) on *pXCP1* and *pLjXCP1* is different (Fig. 2-4B) for all elements except for the AC-elements where they appear in proximity of one another (Fig. 2-4B and Table 2-1)

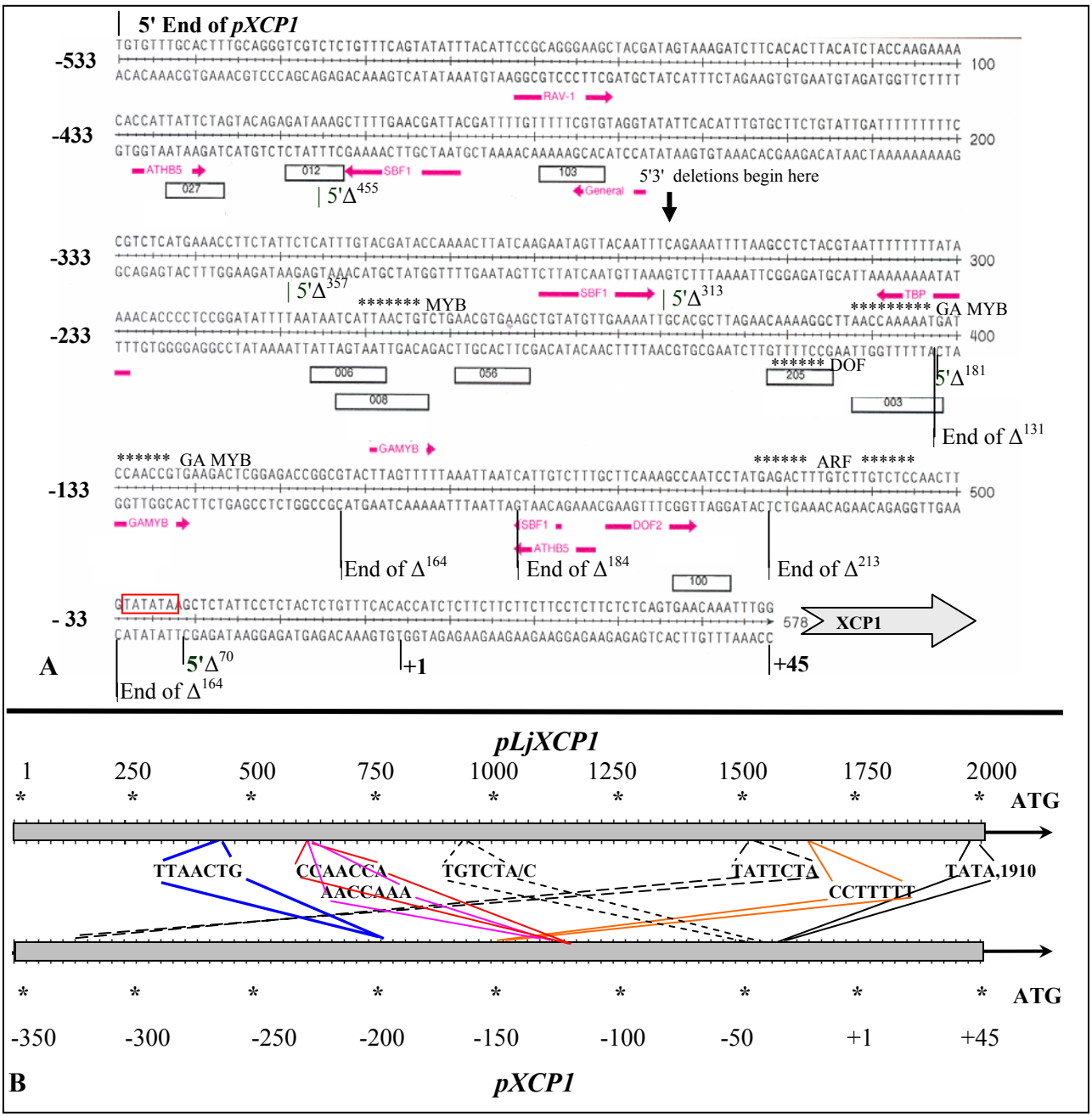


Figure 2-4. Distribution of putative *cis*-elements and corresponding transcription factors on the sequence of Full length[578] *XCP1* promoter. **A**, Predictions of Athamap (pink arrows) and Genomatix (black squares) using the 578-bp region of *pXCP1* as a template are displayed below sequences. Abbreviations are codes used by pfam and transfac databases to describe transcription factors (Tables 2-2 and 2-3). The 5' unidirectional deletions and the 5'3' bidirectional deletions are dispayed. Length of *pXCP1* is presented as a negative sequence based on the +1 transcription initiation site (left) and as positive sequence beginning with 1 at the 5' end. Stars indicate Myb-, Dof-, and ARF-binding elements common to *pLjXCP1* and *pXCP1* sequences and located within the internal 237-bp region.**B**, A schematic showing the linear distribution of the *cis*-elements (Table 2-1) on *pLjXCP1* and *pXCP1*. TGTCTA, present in *pLjXCP1*. TGTCTC, present in *pXCP1*.

578-bp Region of *XCP1* Full Length Promoter				**2000-bp Region of *Lotus Japonicus***	
Genomatix		**Athamap**		**Genomatix**	
Sequence	**Position/Anchor**	**Sequence**	**Position**	**Sequence**	**Position/Anchor**
TATTCTA	107-113/110	CCAT**TATTCT**	103-111	**TATTCT**A	1497-1503/1500
TCAT**TAACTGT**	327-337/332	**TAACTGT**C	331-338	GTTT**TAACTG**C	483-493/488
CCTTTTG	378-384/381	**CCTTTT**		**CCTTTTT**	1608-1614/1611
AACCAAAAATG	388-398/393			**AACCAAA**CTAA	585-595/590
		T**CCAACC**GTG	401-410	**CCAACC**A	583-589/586

Table 2-1. Computer-based prediction of *cis*-regulatory elements in the 237-bp region found common to *pXCP1 and pLjXCP1* using Genomatix and AthaMap programs. Bold sequences represent core elements conserved found in each DNA module. TATTCTA, light regulation; TTAACTG, Myb-binding; CCTTTT, Dof-binding; AACCAAA, another Myb-binding. Numbering, 1 begin at 5' end of *pXCP1* and *pLjXCP1*.

Factor	**Family**	**Reference**
RAV1	AP2	Kagaya et al., 1999
DOF2	DOF	Yanagisawa and Schmidt, 1999
SBF1	GT	Predicted
ATHB5	HD-ZIP	Johannesson et al., 2001
GAMYB	MYB	Gubler et al., 1999
TBP	TBP	Shahmuradov et al., 2003

Table 2-2. AthaMap list of transcription factors predicted to bind *cis*-elements in *pXCP1* and *pLjXCP1*.AP2, APETALA 2 domain; DOF, DNA one-finger; GT, trihelix binding domain; HD-ZIP, homeo-domain leucine zipper; MYB, helix loop helix; TBP, general DNA binding.

IUPAC Name	**Possible Role**	**Pfam**	**Reference**
MYBPLANT	Myb; regulation of phenyl propanoid/lignin biosynthesis	003	Tamagnone et al., 1998
HDZIP2ATATHB2	Myb; negative autoregulation of homoebox gene ATHB2	006	Ohgishi et al., 2001
MYB2AT	Myb; water stress response and autoregulation	008	Urao et al., 1993
IBOX	Myb; binding to I-box	012	Rose et al., 1999
10PEHVPSBD	Light regulation of chloroplast specific gene in barley	027	Thum et al., 2001
ABRELATERD1	Dehydration response and dark-induced senescence	056	Simpson et al., 2003
CCAATBOX1	Hormone regulation	100	Predicted
CELLCYCLESC	Cell cycle control	103	Predicted
PYRIMIDINEBOX	Gibberellin-response of TAACAAA and CCTTTT	205	Morita et al., 1998

Table 2-3. Genomatix list of transcription factors predicted to bind *cis*-elements in *pXCP1* and *pLjXCP1*.

The above information suggests the presence of three regions that may be required to mediate a TE-specific expression in *Arabidopsis*. The first region includes the auxin responsive element, TGTCTC, that potentially binds auxin responsive transcription factors (Ulmasov et al., 1997). The second region includes the GA Myb binding element CCAACC. The third region includes AACCAAAA, CCTTTT, and TTAACTGTC DNA elements that potentially bind Myb, Dof a Myb transcription factors. These elements are conserved between *pXCP1* and *pLjXCP1* and are located in within the 0.237 kb region (-32 to -268) of *pXCP1* and within ~1 kb region (483-1503) of *pLjXCP1*.

B2.5. Experimental Deletions of *XCP1* Promoter

The 578-bp region of the *XCP1* promoter was subjected to 5' unidirectional, 5'-3' bidirectional deletions, and internal deletion analysis (Fig. 2-5) to define region(s) that play regulatory roles in *XCP1* expression. These deletions have been independently and stably integrated into the *Arabidopsis* nuclear genome as transcriptional fusions with the *GUS* reporter gene using a modified binary vector pBI121. Figure 2-5 shows the constructs on a common scale. Fig. 2-6 demonstrates a PCR verification of the presence of the deletion construct fused with the reporter gene using *Agrobacterium* culture as a template. In addition to their resistance to the herbicide Finale and the presence of Kanamycin *NptII* gene, T_1 and T_2 plants not expressing GUS also contained the deletion construct fused with the reporter gene.

B2.5.1. 5' Unidirectional Deletions of *XCP1* Promoter

Figure 2-5 shows that elimination of sequences upstream of -268 does not alter the specificity of TE-specific GUS expression. Histochemical staining of T_1 plants (N=20-35) and their selfed progeny showed levels of TE-specific GUS expression similar to those of *pXCP1*. Elimination of a 132-bp fragment in $5'\Delta^{181}$ (Fig. 2-5) downstream of -268 completely abolished GUS expression indicating that the 132-bp fragment contains necessary DNA elements.

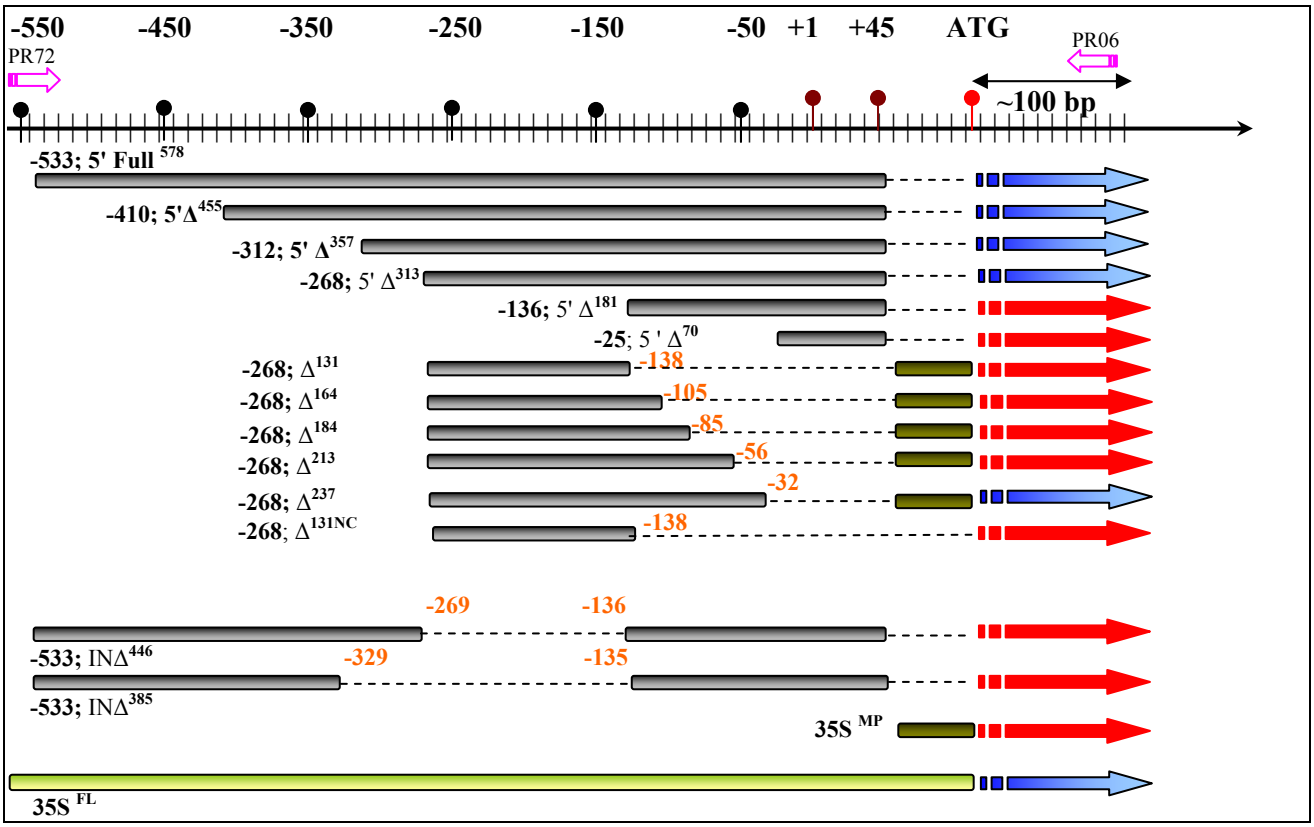


Figure 2-5. A schematic diagram of transcriptional fusion deletion constructs used for *XCP1* promoter analysis. **Bar color codes**; Gray, Deletion constructs; Green, 35S minimal promoter; Yellow, Full length 35S promoter. **Arrow color codes**, Blue, GUS is expressed; Red, GUS is not expressed. **Double arrow**, Distance of primer PR06 from first ATG is approximately 100 bp. PR06 is primer used for sequencing of all deletion constructs. PR06/PR72 is primer pair used for PCR verification of *pXCP1::GUS* construct and subsequent deletions. **Naming of constructs**; 35S ^{FL}, 35S full length promoter; 35S ^{MP}, 35S minimal promoter; 5'Δ, deletions of the 5' end of *pXCP1*; Δ, deletions of 5' and 3' ends of *pXCP1*, INΔ, removing of an internal fragment followed by religating the flanking regions of *pXCP1* maintaining the natural 5'→3' orientation; Δ^{131NC}, Negative control lacking the 35S minimal promoter**. Superscript numbers,** Length of *pXCP1* sequence analyzed. **Scale,** +1, Transcription initiation site. Sequences down stream are positive. Sequences upstream are negative. **5' UTR**, +1 to +45. **Negative numbers,** End positions of deletion constructs numbered according to the +1 transcription initiation site.

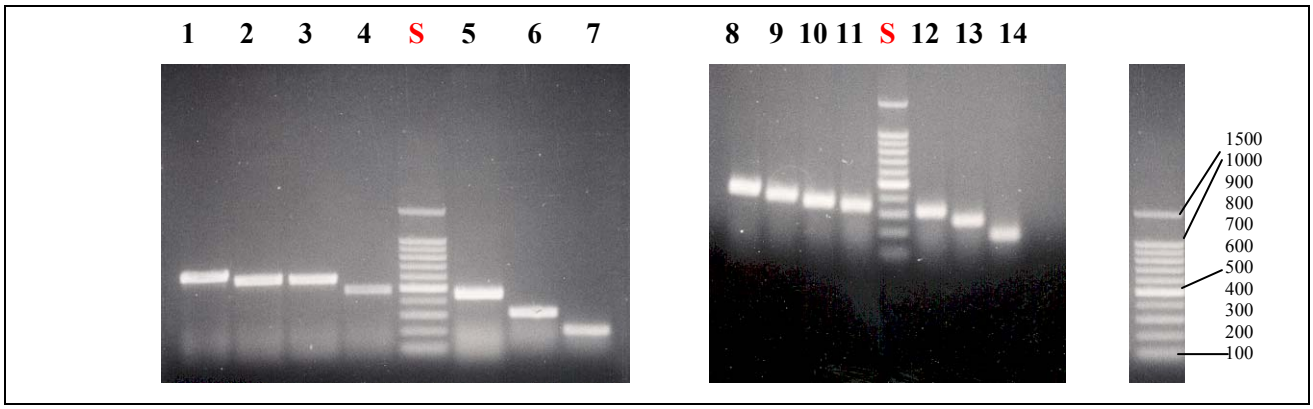


Figure 2-6. PCR amplification of *pXCP1* deletion series using transformed *Agrobacterium* strain GV3101 culture as a template and PR72/PR06 as primers flanking the inserts. Lanes 1, Full ^{578}; 2, INΔ^{446}; 3, 5'Δ^{455}; 4, 5'Δ^{357}; 5, 5'Δ^{313}; 6, 5'Δ^{181}; 7, 5'Δ^{70}; 8, Δ^{237}; 9, Δ^{213}; 10, Δ^{184}; 11, Δ^{164}; 12, Δ^{131}; 13, Δ^{131NC}; 14, 35S ^{MP}. S, 100-bp ladder. Δ^{131NC}, The internal deletion without the 35S ^{MP} thus serving as a negative control (NC).

Quantitative measurement of GUS activity of the five 5' unidirectional deletions was conducted (fig. 2.7). Per deletion construct, two biological samples were collected from herb$^+$ T_2 plants whose T_1 descendent plant was characterized by its resistance to Finale and the presence of the deletion in fusion with *GUS* coding region as tested by PCR (data not shown). To achieve a fair representation of GUS activity, two biological samples were also collected from herb$^+$ T_2 plants derived from another T_1 plant characterized as above. Assessment of GUS activity was conducted as described by Gallagher (1992) using two replications per biological sample. Figure 2-7 shows that proteins from plants carrying the 5'Δ^{181} and 5'Δ^{70} constructs hydrolyzed MUG substrate at rates 1.6 fold and 1.8 fold higher than the proteins extracted from WT plants, respectively. Compared to the 5'Δ^{181} and 5'Δ^{70}, similar amounts of proteins from plants carrying the 5'Δ^{455}, 5'Δ^{357}, 5'Δ^{313} constructs hydrolyzed MUG at rates 13 fold, 17 fold, and 16 fold higher than proteins from WT plants, respectively. Thus, quantitative assessment of GUS activity supports the histochemical results reported earlier.

B2.5.2. Internal Deletions of XCP1 Promoter and 5'-3' Bidirectional Deletion

Since results from 5' deletion experiments suggest the presence of DNA elements in the 132-bp fragment located in the -136/-268 region, this fragment was deleted from the promoter region

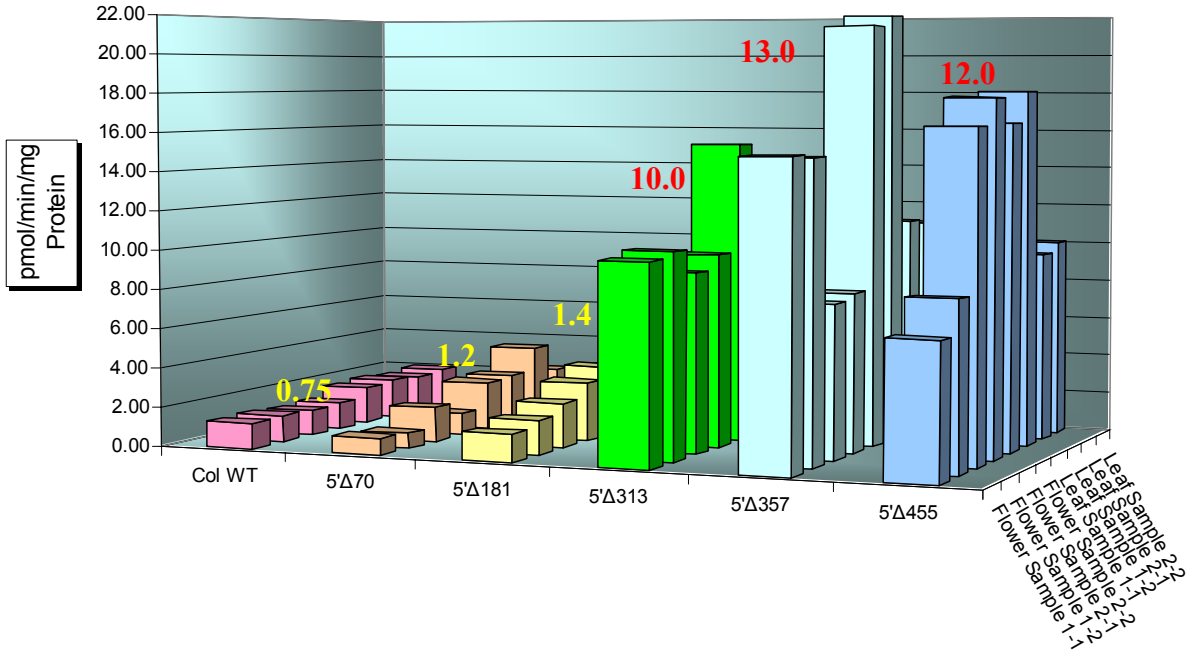


Figure 2-7. Quantitative analysis of GUS activity in T_2 WT [5' unidirectional deletions]. GUS activity was quantified as pmole of hydrolyzed 4-methylumbelliferyl β-D-glucuronide (MUG) per min per mg plant protein. Amount of GUS-hydrolyzed MUG was determined by converting the fluorescent reading of the product 4-ethylumbelliferone measured at 460nm into pmole as described in Gallagher (1992). *X* axis shows the five unidirectional deletions compared to wild type plants. *Y* axis shows amount of hydrolyzed MUG by GUS present in the protein extracts. *Z* axis shows the two biological samples (leaf and flower) and their two replications from two T_1 transgenic lines. Results were based on a 240-min assay. Average rate of GUS activity (pmole/min/mg protein) per deletion compared to WT is presented in yellow-coded and red-coded numbers on the graph.

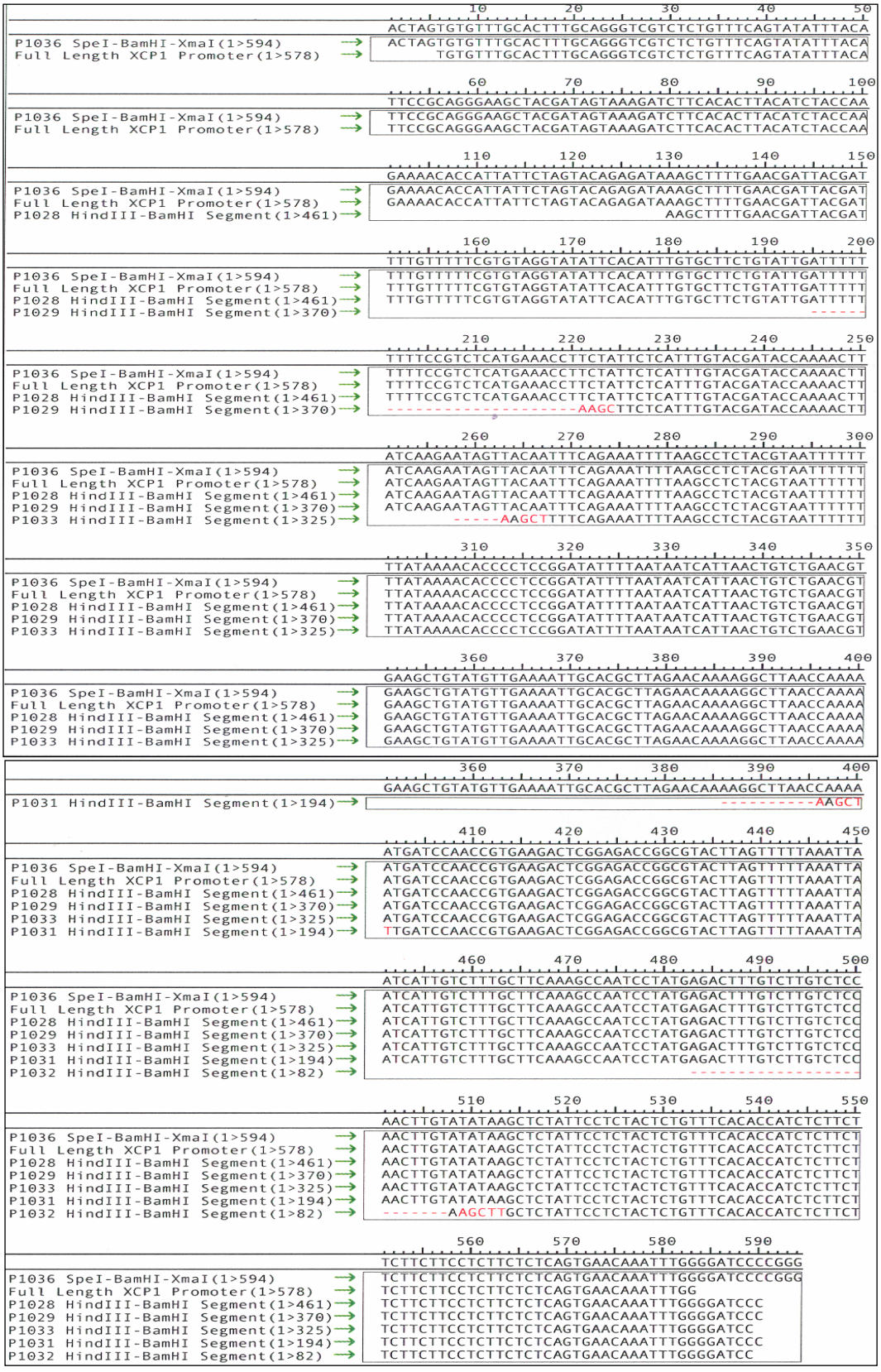
and the 5' and 3' flanking regions were fused together to construct a 446 bp fragment ($IN\Delta^{446}$). GUS expression was not observed in any of the T_1 plants tested (N=14). A second but larger internal deletion was also recovered as a by-product of the PCR reaction. The 385-bp fragment ($IN\Delta^{385}$) abolished GUS expression in T_1 plants tested (N=3).The 132-bp fragment was subsequently fused to a minimal 35S promoter sequence (Benfey et al., 1990) upstream of GUS coding region to test whether this fragment alone was sufficient for activating GUS in TEs. T_1 plants containing this bidirectional deletion also failed to activate GUS expression (N=10). Hence, although necessary, the 132-bp fragment is not sufficient to drive GUS expression.

B2.5.3. 5'-3' Bidirectional Deletion Series

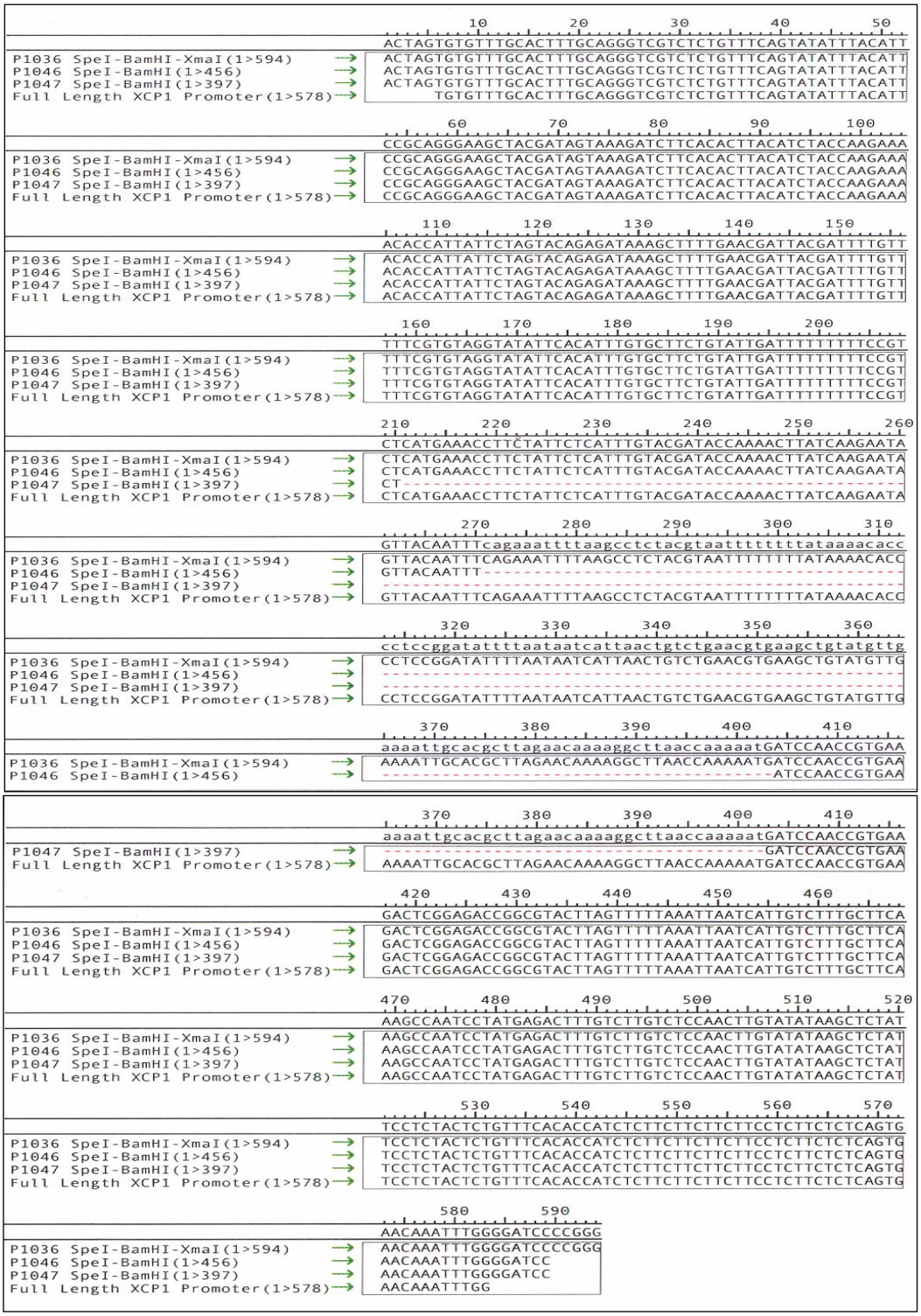
Building on results from 5' unidirectional and 5'-3' bidirectional and internal deletion, the 132-bp fragment was step-wise extended at the 3' end (Fig. 2-5) to eventually cover a 237-bp region that terminates 1 bp upstream of predicted TATA box located downstream of -32 position. Sequences downstream of the -32 position are known to comprise common binding sites for the basal transcription machinery in eukaryotes and for the 3' extensions, these were provided by the minimal 35S promoter. The reason for the incremental 3' extension of the 132-bp region was to attempt to reconstitute the DNA elements that can reactivate GUS expression. The 237-bp region -32/-268, the longest 3' extension, was the only 3' extension construct able to drive GUS expression in TEs (N=6). All shorter 3' extensions failed to reactivate GUS expression. Thus, the additional 25-bp fragment located at the 3' end of the 237-bp appears to contain DNA elements that complemented upstream elements and facilitated the reactivation of GUS expression. Interestingly, this 25-bp fragment includes the known ARF binding element, TGTCTC.

B2.5.4. Sequence Analysis of Deletions

In addition to restriction digest and PCR-based verification of deletion inserts in binary vector, all deletion constructs were sequenced in cloning vector pGEM and in the modified pBI121 binary vector (i.e., P2043 and P2071). Figures 2-8 A, B, and C show a comparative sequence analysis between the deletions and the wild-type promoter as presented in TAIR database. Sequencing was conducted using the antisense primer PR06 specific to GUS coding region which is positioned approximately 100 bp downstream of its initiator Met codon, ATG. Prior to conducting sequence comparisons, sequence ambiguities were first resolved using chromatograms aided by comparing the sequences that were derived from both pBI121 and pGEM carrying the respective deletion constructs. The sequence alignment was conducted using the Sequence Management Program in DNASTAR software. Parameters were set to allow for large gaps since large regions of dissimilarity are present among the deletion sequences. There were no observable base mutations in these sequences when compared to the WT.


P1036 SpeI-BamHI-XmaI(1>594)
Full Length XCP1 Promoter(1>578)
P1028 HindIII-BamHI Segment(1>461)
P1029 HindIII-BamHI Segment(1>370)
P1033 HindIII-BamHI Segment(1>325)
P1031 HindIII-BamHI Segment(1>194)
P1032 HindIII-BamHI Segment(1>82)

A


P1036 SpeI-BamHI-XmaI(1>594)
P1046 SpeI-BamHI(1>456)
P1047 SpeI-BamHI(1>397)
Full Length XCP1 Promoter(1>578)

B

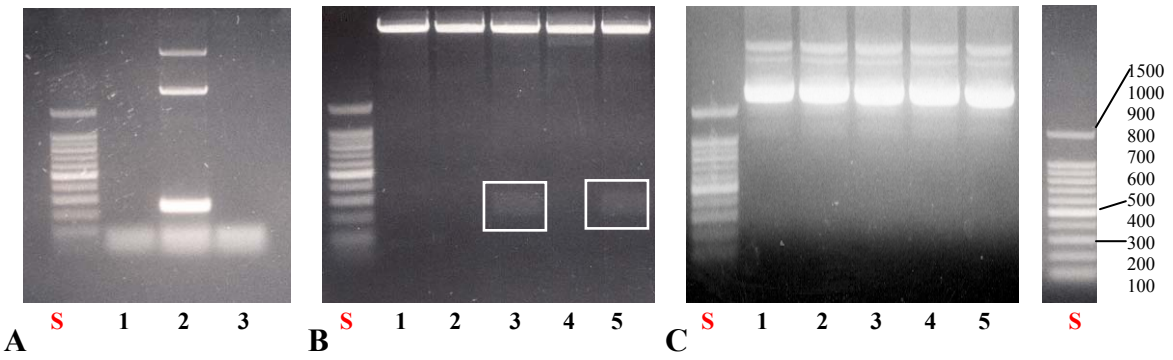
C

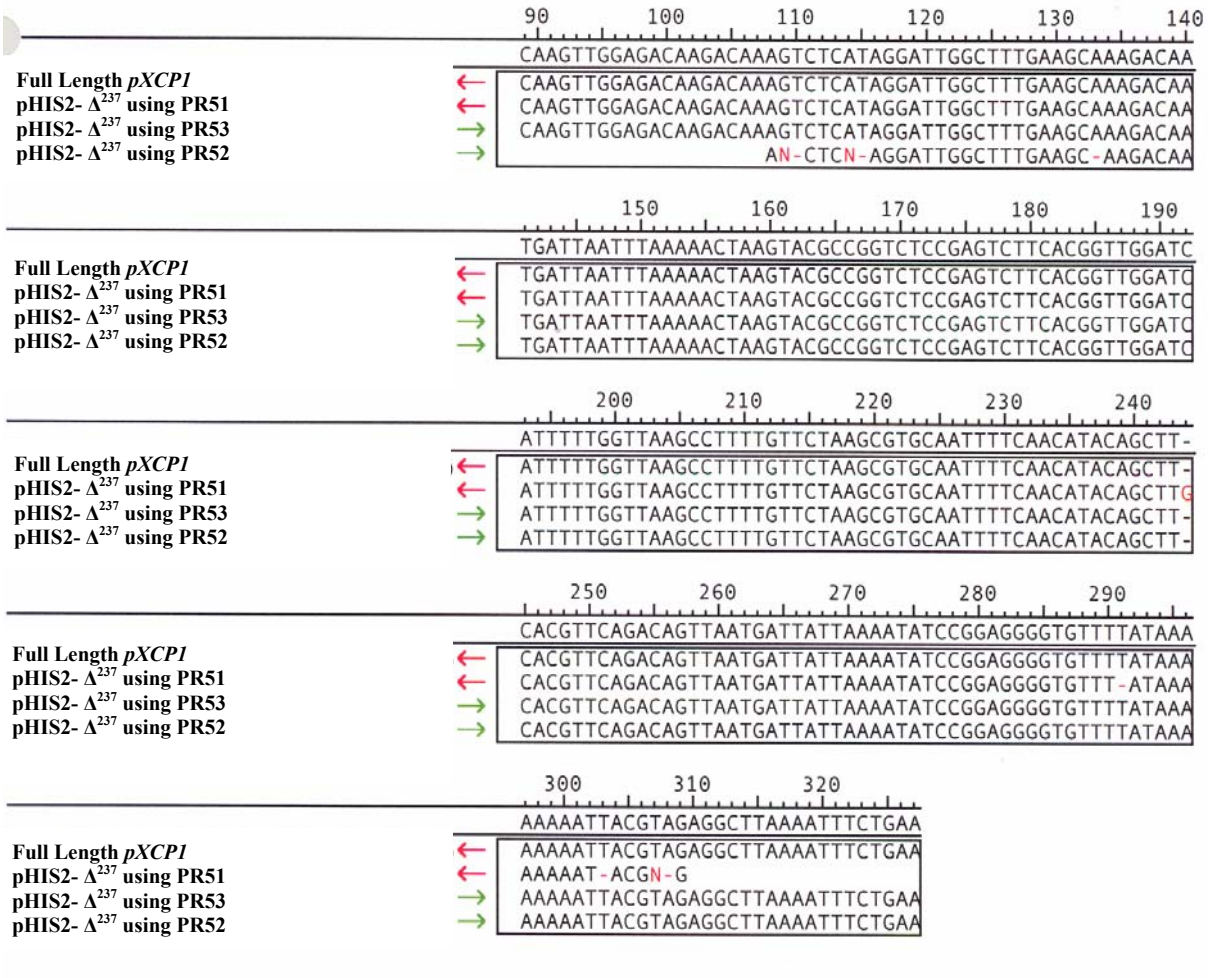
Figure 2-8. Sequence analysis of *pXCP1*deletions. **A**, Sequence analysis of the *XCP1* promoter sequence from TAIR compared to Full length^{578} (P1036), 5'Δ^{455} (P1028), 5'Δ^{357} (P1029), 5'Δ^{313} (P1033), 5'Δ^{181} (P1031), and 5'Δ^{70} (P1032). **B**, *XCP1* promoter sequence from TAIR compared to Full Length^{578}, INΔ^{446} (P1046), INΔ^{385} (P1047). **C**, *XCP1* promoter from TAIR compared to Full Length^{578}, Δ^{237} (P1057), Δ^{213} (P1058), Δ^{184} (P1059), Δ^{164} (P1060), Δ^{131} (P1041), Δ^{131NC} (P1045), 35S ^{MP} (P1042). **Dashes**, Dissimilarities in the alignments.

B2.6. Yeast One-Hybrid Analysis

B2.6.1. Construction of *XCP1* Bait Sequences for One-Hybrid Analysis

The 237-bp (-268 to -32) region shown by deletion analysis (Fig. 2-5) to be necessary and sufficient to direct TE-specific GUS expression was selected to serve as a bait sequence for the yeast one hybrid screen. The 237-bp bait fragment was faithfully amplified and integrated into pHIS2 to construct P1049 (Fig. 2-9 C).





D

Figure 2-9. Construction and sequencing of bait promoter for yeast screen. **A,** PCR amplification of the 237-bp fragment. Lane 2, Primers PR51/PR52 only; 2 pGEM-*pXCP1* DNA plasmid with PR51/PR52; 3 P1003 only. **B,** Restriction analysis of pHIS2- Δ^{237} putative positive plasmids. Lanes 1-5, *Sac*I/*Xma*I digested DNA plasmids 1 to 5. **C,** Lanes 1-5, undigested DNA plasmids 1 to 5. **D**, The *pXCP1* sequence compared to Δ^{237} sequence using PR53 primer (vector specific), PR51 and PR52 (PCR fragment-specific). Boxes in B show the *Xma*I-*Sac*I Δ^{237} released fragment. Initial red-coded nucleotide mismatches were manually corrected using chromatograms. S, 100-bp ladder.

B2.6.2. Defining Yeast Strain Y187 Nutritional Requirements

Y187 has the following genotype: MATα, *ura 3-52*, *his 3-200*, *ade 2-101*, *trp 1-901*, *leu 2-3, 112*, *gal4*Δ, *met*$^-$, *gal80*Δ, *URA3::GAL1*$_{UAS}$*-GAL1*$_{TATA}$*-lacZ,MEL1*. Relevant to our study are the following mutations, *his 3-200*, *trp 1-901*, and *leu 2-3, 112* which inhibit yeast growth on media lacking histidine, tryptophan, and leucine, respectively. The mutation *ura 3-52* is compensated for by the preengineered *URA3::GAL1*$_{UAS}$*-GAL1*$_{TATA}$*-lacZ, MEL1* construct (Clontech).

Y187 was streaked on YPDA and SD/-lue, SD/-trp, SD/-his, SD/-ura, SD/-met, SD/-ade 150-mm plates according to the Clontech manual to confirm the nutritional requirements of the yeast strain Y187. As expected, non-transformed yeast grows on the non-inhibitory media YPDA (Fig. 2-10 A) and on synthetically designed media lacking uracil (SD/-ura; Figure 2-10 B). The yeast did not grow on all other media as it lacks the genes that can compensate for lack of leucine, tryptophan, histidine, methionine, and adenine (not shown). Yeast Y187 transformed with pHIS2 containing 237-bp bait sequences of *XCP1* alone (i.e., Y187 [P1049]) was able to grow on SD/-trp (Fig. 2-10C). pHIS2 contains two TATA boxes one of which is constitutively inducing the expression of *HIS3* gene and producing background growth. The effect of this TATA must be suppressed so that the other TATA box can induce *HIS3* where it reflects putative positive DNA-protein interactions. The effect of the constitutive TATA box is evident in Figure 2-10 D where background growth is observed on SD/-trp/-his. Supplementing the media with 40 mM 3-AT (Fig. 2-10E) dramatically decreased background growth. A complete suppression of background growth was observed at 50 mM 3-AT (data not shown). A small scale transformation experiment was conducted to test transformation efficiency and ability of transformed Y187 to grow on prohibitive media. This is a necessary step to insure that Y187 strain as well as the vectors carrying the bait and ds cDNA sequences function as expected *in vivo*. Y187 [P1049 + ds cDNA + linearized pGADT7-Rec2] was able to grow on SD/-leu (Fig. 2-10F). This demonstrates that Y187 was successfully transformed with functional pGADT7-Rec2 (i.e., effective homologous recombination between ds cDNA molecules and the linearize vector). Growth of Y187 [P1049 + ds cDNA + linearized pGADT7-Rec2] was observed on SD/-trp/-leu/-his without 3-AT (Fig. 2-10G). A 40-mM 3-AT supplement dramatically decreased the growth (Fig. 2-10H) and at 50 mM 3-AT growth was completely suppressed.

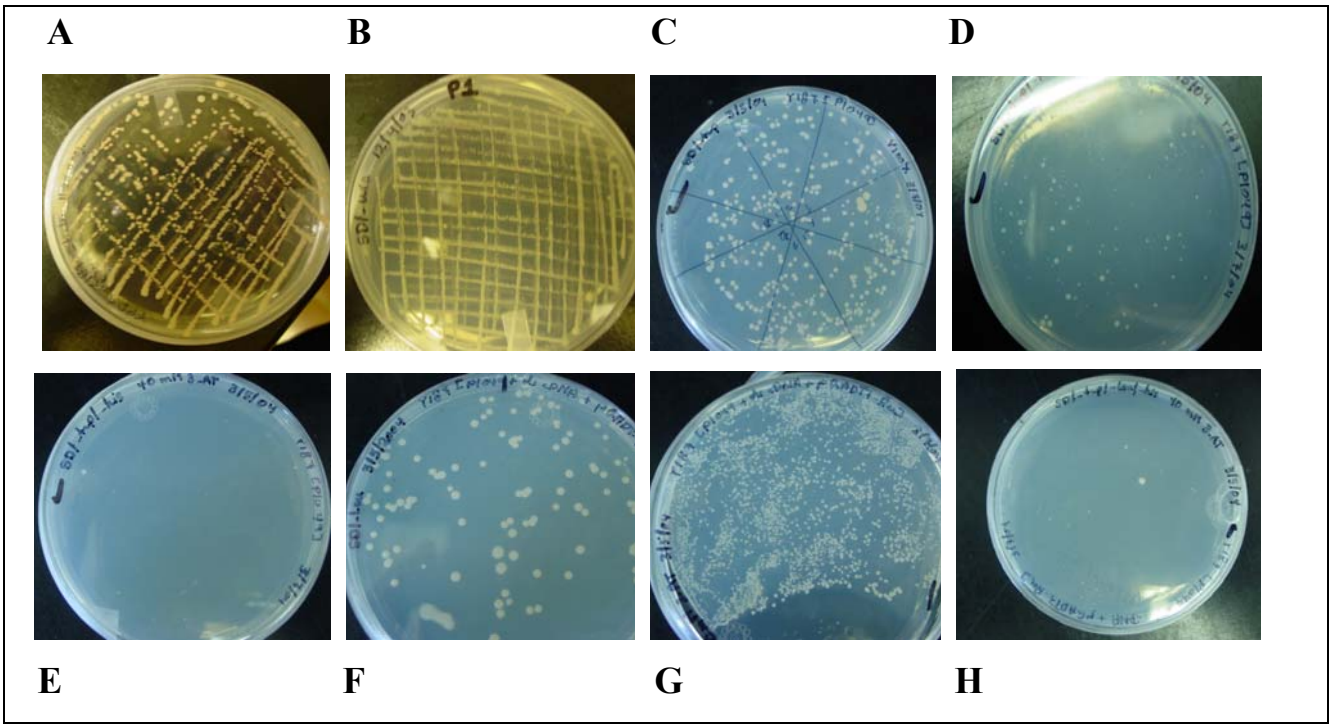


Figure 2-10. Nutritional requirements for yeast Y187 strain. Non-transformed Yeast 187 strain colonies on YPDA agar medium (**A**) and on SD/-ura (**B**). Yeast 187 [P1049] is able to grow on SD/-trp (**C**) and on SD/-trp/-his (**D**) and very minimally on SD/-trp/his supplemented with 40 mM 3-AT (**E**). Yeast 187 [P1049+ ds cDNA + pGADT7-Rec2 *Sma*I-linearized] was able to grow on SD/-lue (**F**), and with a dramatic decrease with increasing 3-AT concentration from 0 mM (**G**) to 40 mM (H) on SD/-trp/-leu/-his.

B2.6.3. Construction of *Arabidopsis* Root RNA ds cDNA Library as a Prey for Yeast One-Hybrid Analysis

XCP1, and thus *XCP1*-associated transcription factors, are expressed at higher level in roots and hypocotyls than in young leaves due to the larger amount of secondary vascular tissue naturally present in roots. *Arabidopsis* plants were grown under low population density and regularly pruned to remove inflorescences (Zhao et al., 2000). Together, these conditions delay senescence and promote secondary growth. Root-hypocotyls from 4-5 week old plants were used as starting material to isolate total RNA (Fig. 2-11A). RNA was incubated at 37^0C for 0-30 min to test its stability. No degradation of RNA was observed (Fig. 2-11B). The 1st strand cDNA was synthesized from total RNA via a CDIII Oligo dT-based, dC/BD SMART III coupled reverse transcription reaction followed by template switching and extension. Using this 1st cDNA as a template, the 5' sense and 3' antisense primers amplify a ds cDNA library with CDIII and BD SMART III anchors that would facilitate cloning a single ds cDNA molecule into pGADT7-Rec2 via *in vivo* homologous

recombination. Quality of 1st strand cDNA was tested by amplifying full length sequences of *XCP1* and actin using exon specific primers (Fig. 2-11C). Expected bands correspond to cDNA and not genomic DNA showing that RNA sample was not contaminated with genomic DNA. The ds cDNA library was subsequently purified and ds cDNA molecules were resolved on agarose gel (Fig.2-11D). Quality of ds cDNA was also tested in a similar manner to 1st strand cDNA and identical results were obtained (data not shown) indicating absence of genomic DNA contamination in the ds cDNA molecules.

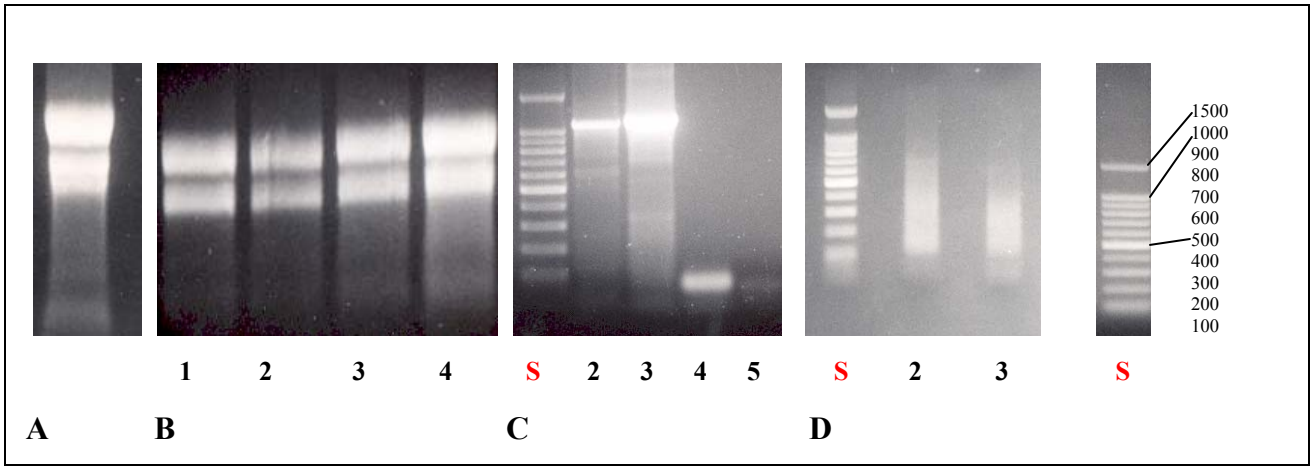


Figure 2-11. Contruction of prey ds cDNA library for yeast screen. **A**, Total RNA from *Arabidopsis* root-hypocotyls was resolved on an ethidium bromide impregnated formaldehyde agarose gel. **B,** Lanes 1-4, Total RNA was heated at 37°C for 30, 20, 10, 0 min, respectively, and 2 µg was run per lane was run on ethidium-promide impregnated formaldehyde gel to assess RNA stability. **C,** Quality assessment of 1st cDNA template by PCR. Lane 1, 100 bp ladder; Lane 2, 1st strand with primers PR41/PR48; Lane 3, 1st strand with primers PR56/PR57; Lane 4, primers alone PR41/PR48; Lane 5, primers alone PR56/PR57. **D,** Construction of ds cDNA from 1st strand cDNA. Lane 1, 100 bp ladder; Lane 2 and 3 are ds cDNA molecules ranging from >1500 bp to 100 bp. S, 100-bp ladder.

B2.6.4. Yeast One-Hybrid Analysis Using *Arabidopsis* ds cDNA and P1049

Library scale analysis was conducted using the Matchmaker One-Hybrid System. Y187 was transformed with both plasmids. The pHIS2 vector was circularized by cloning the 237-bp promoter region of *XCP1* into its *Xma*I-*Sac*I sites (pHIS2-Δ^{237}). Although it was argued against using a long DNA sequence as bait in yeast one-hybrid screem, Wei et al., (1999) and Santi et al., (2003) were successful in using a 181-bp and a 305-bp DNA sequence, respectively, to clone cDNAs that encoded relevant transcription factors. The pGADT7-Rec2 is expected to circularize *in vivo* via

homologous recombination with a ds cDNA molecule. After cotrnsforming yeast Y187 with pHIS2-Δ^{237}, pGADT7-rec2, and ds cDNA molecules and plating the cells on SD/-leu/-his/-trp + 50 mM 3-AT, a total of 109 colonies were observed growing on media lacking all three amino acids and containing 50 mM 3-AT, an inhibitory level of 3-AT that was determined experimentally under our conditions prior to the actual analysis. Such colonies are putative positive colonies that suggest an interaction between specific ds cDNA-encoded amino acids and the 237-bp promoter. Of the 109 colonies, only 46 were observed within 4-5 days of incubation suggesting a strong interaction. The remaining colonies were observed after 9 days of incubation while stored at 4^{o}C. According to the manufacturer's description of the system and Dr. B. Smith (yeast scientist at Clontech, personal communication), a 7-day incubation is a recommended time where strong DNA-protein interactions are typically manifested in colony growth. Forty three colonies (the other three remained small) were retested on SD/-leu/-trp/-his supplemented with 50 mM 3-AT for the dual purpose of segregating the colonies for PCR analysis and confirming the phenotype. All 43 colonies were found positive under the second test conditions. Subsequently, colony PCR was conducted using gene specific primers PR71 and PR72. These primers should amplify the ds cDNA insert, an approximately 180-bp 5' flanking region containing T7 sequences followed by HA epitop tag, and about 80-bp 3' flanking region. Only 34 colonies contained predominantly single-band products (Fig. 2-12) whereas 5 showed multiple bands and the remaining 4 produced no products at all (not shown). Restriction digestion using the frequent cutter *Alu* I (AGCT) was conducted on purified PCR products of similar sizes to determine if a similar restriction pattern was observed (data not shown). Such analysis showed that the restriction patterns were different for all clones tested. Three colonies were excluded from analysis due to insufficient quality and quantity of purified PCR product (Fig. 2-12, lanes 6, 7, 33). Thirty one colonies were submitted for sequencing using T7 sequencing primer internal to the PCR product.

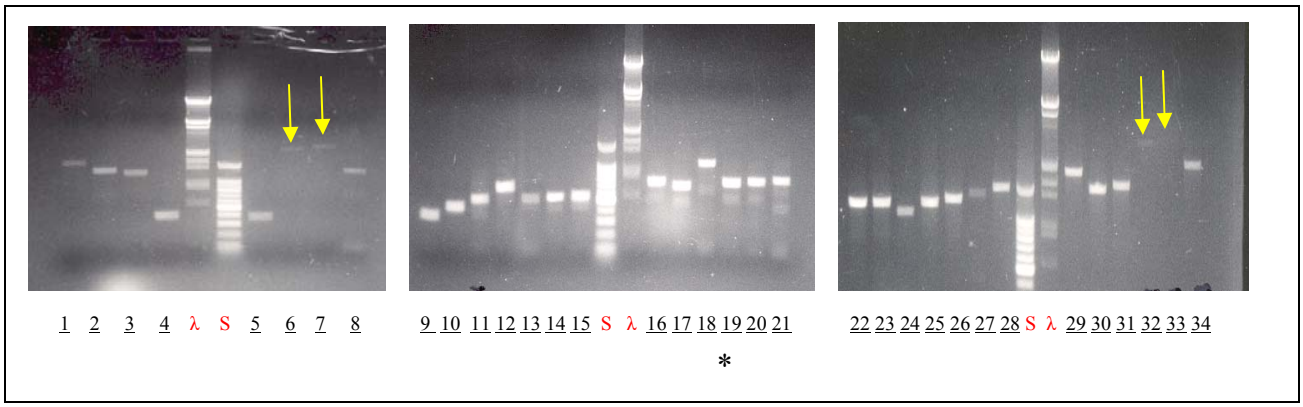


Figure 2-12. Colony PCR of His+/Leu+/Trp+ transformed yeast Y187 colonies. Thirty four colonies amplified by primers PR70/PR71 and showing predominantly single-band PCR product are shown. Arrows, Faint bands. Star, PCR product from colony P3C4. S, 100-bp ladder. λ, lambda *EcoR*I-*Hind*III DNA standard.

ID	Product	Best Hit	Identity Score	Description from TAIR database	Molecular Function derived from TAIR database
P1/C3	1200	At5g13450	664/692	ATP Synthase subunit	ATP biosynthesis
P1/C5	500	At4g24990	161/163	Geranylgeranylated protein	Uknown function
P1/C6	800	At5g59880	378/419	Actin-depolymerizing factor	Actin-binding/depolymerization
P1/C7	1200	At4g15780	212/223	Synaptobrevin-related protein	Unknown
P2/C3	1900	At3g20870	290/292	Metal transcporter protein	Zinc ion transport activity
P3/C4	800	At2g22670	118/144	Auxin-responsive protein, IAA8	Transcription factor activity
P3/C7	900	At3g16640	502/507	Similar to a tumor protein	Protein translocation into organelles; DnaJ-type
P4/C1	800	At2g28910	198/199	CAX-interacting protein, CXIP4	Nucleic-acid binding activity
P5/C1	1200	At4g35300	468/491	Similarity to a hexose tranporter	Transporter activity
P7/C1	1100	At4g31480	472/514	Similar to coat protein beta subunit	Clathrin-binding activity
P7/C2	1300	At2g17930	215/226	Similar to inositol kinases	Inositol phosphatydilinositol kinase activity
P7/C3	1900	At5g06560	837/866	Not known	Unknown
P8/C3	2100	At1g55630	415/418	Protein containing PPR repeat	ATP-binding activity
P10/C2	900	At2g34680	502/504	Leucine-rich repeat protein	Hydrolase activity
P11/C2	1200	At5g52920	741/759	Similar to pyruvate kinase isozyme	Pyruvate kinase activity
P11/C4	900	At2g30440	438/440	Thylakoidal processing enzyme	Serine-type peptidase activity

Table 2-4. BLAST results and inferred functions of 16 sequenced PCR products showing high identity scores against *Arabidopsis* transcript database.

B2.7. Phenotypic characterization of two mutant lines generated by chemical mutagenesis

Figure 2-13 represents the steps in developing the M_2 screening population and the screening process of two putative mutants that initially showed GUS expression different from the TE specific GUS expression in non-mutagenized *pXCP1::GUS* plants. A non-segregating herb$^+$ *pXCP1::GUS Arabidopsis* population (T_4 4-4) was developed from a GUS$^+$ plant selected from a population that exhibited a 3:1 GUS$^+$ bar$^+$: GUS$^-$ bar $^-$ segregation ratio. Seeds of T_4 4-4 were mutagenized by ethyl methane sulfonate and 100 M_2 pools were generated (Fig. 2-13). M_1 plants were not evaluated. The first mutant, namely M_2 11-2, was selected due to its negative GUS expression. Flowering of this mutant was evident 12 weeks after germination was first observed (Figure 2-14 A). Backcrossing the mutant with non-mutagenized WT would facilitate eventual positional cloning attempts by segregating away irrevelant mutations from the mutant of interest. Reciprocal crosses of M_2 11-2 with WT were not possible and very few selfed M_3 seeds were obtained of which only one GUS negative M_3 11-2-1 plant germinated. The backcross was attempted again and the ♀ WT X ♂ M_3 11-2-1 was the only successful cross suggesting female sterility of the M_3 plant. The 38 F_1 plants resulting from the above cross showed a 1:1 segregation ratio in female sterility as characterized by silique development and seed set of the plants tested. GUS expression, based on histochemical staining results, was not observed in any of the 38 F_1 plants. PCR reactions for the amplification of *pXCP1* and *pXCP1::GUS* fragments were conductd using gDNA from randomly selected F_1 plants. An F_1 plant designated, F_1-2, was bar$^+$, GUS$^-$, able to set selfed seeds, and contained *pXCP1* region in fusion with *GUS* coding region. Initial sequencing of *pXCP1* and *GUS* coding region did not show 100% alignment with the endogenous *pXCP1* sequences from TAIR or GUS coding sequences derived from pBI121 sequences, respectively. However, manual chromatogram analysis resovled 99% of the ambiguities (data no shown). To develop a mapping population, attempted reciprocal crosses of F_1-2 with *Landsber erecta* (*Ler*), a genetically different *Arabidopsis* ecotype, did not result in silique development. Subsequently, F_2-2 populations were developed. Prior to reattempting any crosses with *Ler*, segregation analysis of *GUS* and *bar* expression in selected F2 and F3 plants was conducted (summarized in Fig. 2-13). The results suggested that the transgene inheritance pattern was not following a mendalian pattern and epigenetic factors may confound our ability to use GUS reporter gene in subsequent steps of positional cloning.

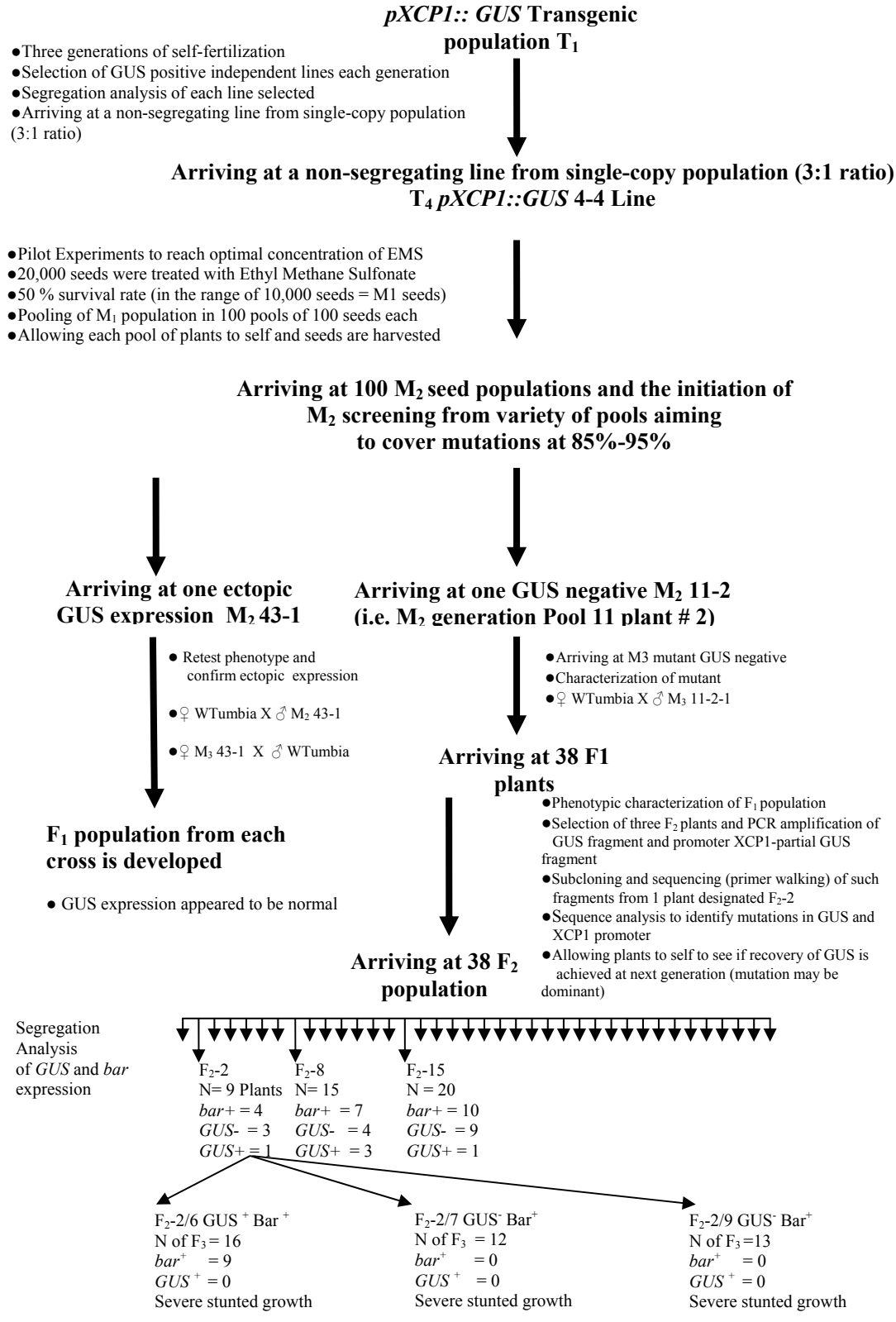


Figure 2-13. A Schematic of technical steps in the mutant screen and analysis of *pXCP1::GUS* mutant lines M_2 11-2 and M_2 43-1.

A **B** **C**

Figure 2-14. Mutant phenotypes produced by chemical mutagenesis of *pXCP1::GUS Arabidopsis*. **A**, A non-flowering GUS-negative 11-week-old M_2 11-2 mutant. **B**, F_2 plants of the ♀ WT X ♂ M_2 43-1 cross containing the mutant phenotype of F_1 M_2 43-1 that initially showed an ectopic expression of GUS. **C**, Enlarged M_2 43-1.

The phenotype itself could in principle be used in mapping. However, due to the extremely low fecundity and slow growth rate exhibited by the mutant, it was decided that efforts to continue mapping the mutant would cease in favor of the yeast one-hybrid approach to identify upstream regulators of *XCP1* expression. The second mutant line (Fig. 2-13 and Figure 2-14B), namely M_2 43-1, initially showed ectopic expression of GUS. Although apparently sterile, reciprocal crosses with WT were eventually successful and F_2 segregating population was developed (Figure 2-14B). However, GUS expression appeared to be normal in differentiating TEs. Similar to M_2 11-2 mutant, it was decided that efforts to continue mapping the mutant would cease.

Discussion

The ability of the 237-bp internal region located -32/-268 fused with 35S minimal promoter to activate GUS expression in TEs (N= 6) provides evidence that critical DNA elements are present within that region. When fused together without the internal 237-bp region, the 3' flanking region -32/+45 and the 5' flanking region -268/-533 failed to activate GUS expression suggesting lack of elements sufficient to activate TE-specific GUS expression. Additionally, the inability of the -136/+45 region to activate GUS suggested that the region -136/-32 does not contain elements that are sufficient to activate the reporter gene. The four 5'-3' bidirectional deletions, -268/-138 region (N=15), -268/-105 region (N=13), -268/-85 region (N=8), and -268/-56 region (N=2) regions failed to activate GUS expression. The fifth 5'-3' bidirectional deletion constitunting the -268/-32 (N= 6) region activated GUS expression suggesting that the 3' 25-bp region located -32/-56 is a critical region. Based on deletion analysis, predicted conserved *cis*-elements between *pXCP1* and *pLjXCP1* sequences, and independent experimental evidence that cysteine proteases are regulated by hormones, a model for distribution of DNA regulatory elements in the *pXCP1* can be proposed (Fig. 2-15). The model hypothesizes that there are at least two DNA regulatory regions in the *pXCP1*. Region one is located within the -32/-57 fragment. This region contains two core elements

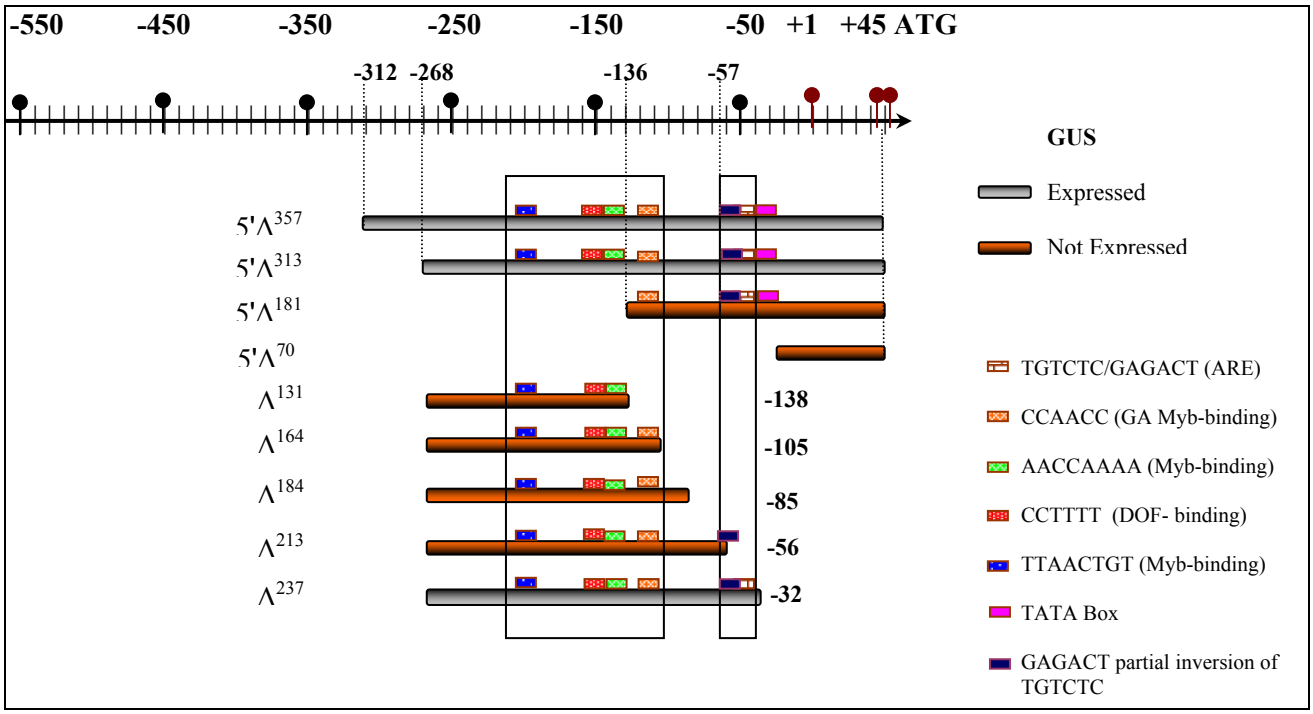


Figure 2-15. Distribution of possible critical *cis*-elements in *XCP1* promoter.

that may bind an ARF transcription factor. Specifically, the ARE element, TGTCTC, and the inverted repeat GAGACA bind ARF1 (Ulmasov et al., 1997; Park et al., 2002). When fused to a *GUS* reporter gene, the synthetic composite ARE structure GAGACAxxxTGTCTC further increased GUS expression in response to auxin application. Substitution such as GAGcCAxxxTGgCTC dramatically decreased GUS expression (Ulmasov et al., 1997). The TGTCTC element is present in *pXCP1* and located in the 25-bp region required to reconstitute GUS expression in TEs (Fig. 2-15). A partial inverted repeat, GAGACT, located in the -53/-57 region is interrupted in the GUS-negative plants carrying the Δ^{213} but was complemented in the GUS-positive plants carrying the Δ^{237} fragment. This complementation may have reactivated GUS expression in the Δ^{237}-plants.

The identification of the two putative AREs correlated with a region required for TE-specific expression suggests that XCP1 expression is regulated by auxin. Whether or not the TGTCTC element present in *pXCP1* is a simple element or part of a composite ARE, it cannot activate GUS expression alone. For example, in the plants carrying the $5'\Delta^{181}$ fragment, the TGTCTC element could not direct GUS expression. A partial TGTCTC element was found as TGTCTA in *pLjXCP1*. Constituting the second region of the model are four DNA elements found common to the 2.0 kb *pLjXCP1* and 0.237 kb *pXCP1* (section 2.4); First, the TTAACTG (-193/-203) which is predicted to bind A Myb transcription factor; Second, the CCTTTT (-149/-154), a pyrimidine box predicted to bind DOF or DOF-like transcription factors; Third, the AACCAAA (-138/-147), an element that is predicted to bind a Myb transcription factor; Fourth, the CCAACC element (-123/-134) which is predicted to also bind a GA Myb factor. Region two appears to be necessary to direct GUS in TEs. The interaction of *cis*-elements in both regions is sufficient and necessary.

The distribution of the conserved elements in *pXCP1* and *pLjXCP1* is different (Fig. 2-4). The clustering of the two AC-rich elements appears to be similar, however (Fig. 2-4B). Although important, the linear distribution of the *cis*-elements is not by itself detrimental to the regulation of transcription as structural features of the promoter region may bring together two *cis*-elements in proximity to on another regardless of their linear distance from each other. AC-rich elements bind members of the myb family of transcription factors, a large family that regulates large numbers of plant processes. Members of this family are also known to autoregulate their expression whose promoter region contains AC-rich elements.

Since the 237-bp region of *pXCP1* contains *cis*-elements sufficient and necessary for the TE-directed GUS expression, it was used as bait in yeast one-hybrid screen. The yeast screen yielded 46 colonies within 4-6 days of incubation under appropriate growth conditions. These colonies represent putative positive interactions between the ds cDNA products and the 237-bp region of *XCP1* promoter. Of the 46 colonies, three ceased growth and were excluded from the further analysis.

Of the remaining 43 colonies only 34 contained a single PCR product whereas 4 colonies released no PCR product and 5 colonies released multiple bands. Absence of PCR amplification was likely due to PCR conditions and multiple PCR products resulted from having more than one pGADT7-rec2 plasmid in yeast. Of the 34 column-purified single-band PCR products (Fig. 2-12), 31 produced sufficient DNA quantities and were selected for sequencing. Of the 31 unique sequenced PCR products, 16 showed significant alignments with *Arabidopsis* transcripts (Table 2-4) among which was *IAA8* (At2g22670). *IAA8* received a special attention for five reasons; First, its involvement in TE differentiation in *zinnia* and *Arabidopsis* has been documented (Groover et al., 2003); Second, IAA8 belongs to Aux/IAA family of short-lived auxin response transcriptional regulators lending further support to our experimental deletion results which implicated the ARE, TGTCTC, in TE-directed GUS expression; Third, there is a potential DNA-binding activity of Aux/IAA members due to the partial homology between their conserved third domain and the amphipathic $\beta\alpha_1\alpha_2$ DNA-binding fold of the Arc and MetJ repressors in prokaryotes (Guilfoyle et al., 1998b); Fourth, transcript profiles (C. Zhao and E. Beers, unpublished) of all 16 sequences (Table 2-4) distinguished IAA8 as a xylem-biased gene (Fig. 2-16) resembling the transcript profile of *XCP1*; Fifth, *pIAA8* is saturated with AC-rich regions on both stands (TAIR) suggesting the presence of hormon-related myb-based transcriptional regulation, an apparently common theme for TE gene expression.

Manual analysis of the chromatogram (Fig. 2-17) resolved sequence ambiguities of the P3C4 PCR sequence and showed that the inserted cDNA sequence in pGADT7-rec2 first aligned with IAA8 exon 3 at position 9 where GGT codon begins as an in-frame fusion with the yeast activation domain. The sequence extends through exon 3/4 and exon 4/5 splice sites and terminates with the TGA terminator codon. In IAAs, domain IV is constructed via the correct splicing of exon 3 and

exon 4. Sequences after TGA codon were difficult to resolve. The sequence contains domain IV found conserved in many IAA members. The 5' end of the cDNA sequence encodes 24 amino acids upstream of domain IV and partially includes the α_2 portion of the $\beta\alpha_1\ \alpha_2$ region of domain III. The cDNA sequence includes the invariant amino acids in the α_2 region. Although $\beta\alpha_1\alpha_2$ secondary structure in IAA8 is precidted to bind DNA, our data suggest that domain IV of IAA8 has the DNA-binding activity, which, if true, is a novel finding. Typically, the highly conserved domain IV is responsible for IAA/ARF hetero- or IAA/IAA homo-dimerization (Abel et al., 1994). Aux/IAA family members were shown experimentally to bind ARFs through a dimerization of the conserved domain III and IV found in both ARFs and most of IAA proteins (Guilfoyle et al., 1998a, b). Ulmasov et al. (1997) showed that ARF1 interacted with specificity to TGTCTC element using carrot suspension cultures.

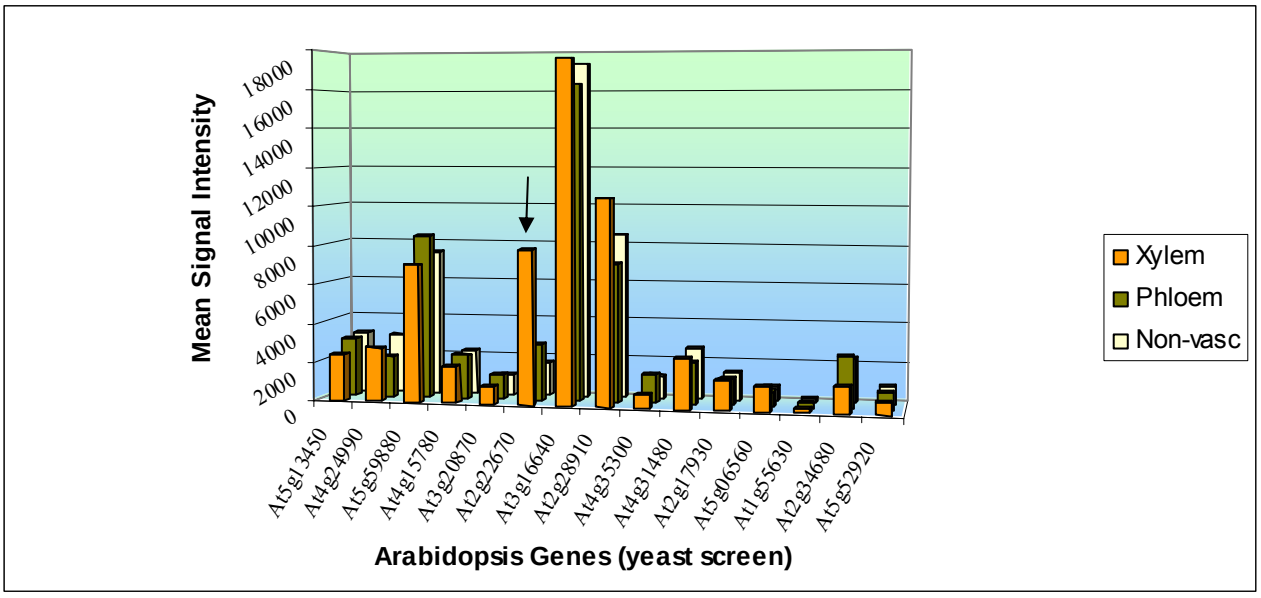


Figure 2-16. Tissue-specific expression level for *Arabidopsis* genes retrieved from yeast one-hybrid screen. Arrow points to the transcriptional regulator IAA8 (At2g22670).

In an independent study, Ulmasov et al. (1997) showed by mobility shift assay that ARF1 binds to TGTCTC and by yeast two-hybrid screen that ARF1 binds to Aux/IAA proteins. Thus IAA8, by homology to other IAA members, contains domain III and domain IV and is likely to bind an ARF in a similar manner. In addition to exploring IAA8 potential for interaction with TE-specific ARF proteins, perhaps in yeast two-hybrid screen, it will be interesting to see whether IAA8 can bind *pXCP1* at the TGTCTC site or other sites within the 237-bp region.

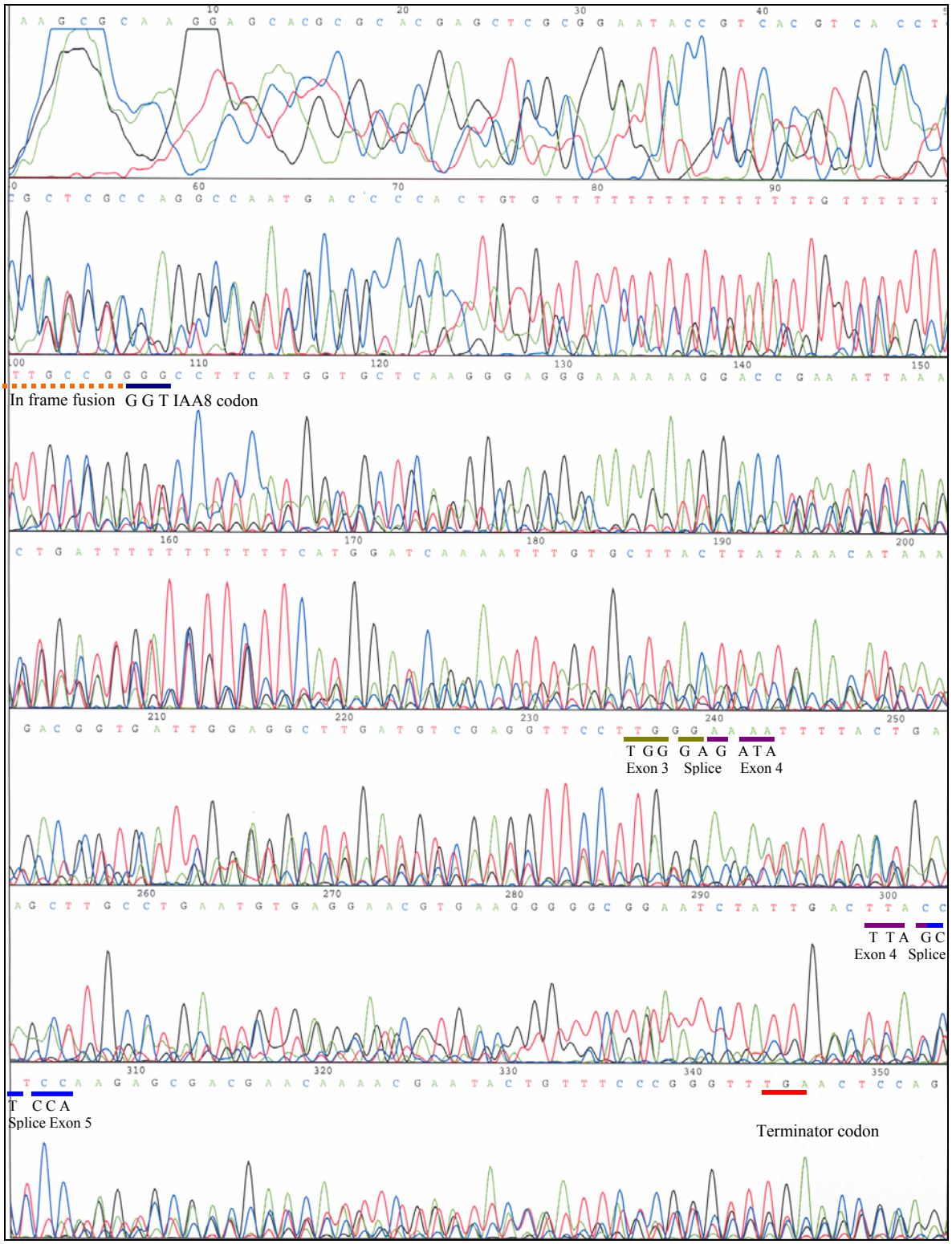


Figure 2-17. Sequence analysis of the PCR product amplified from P3C4 colony. Mismatches between the cDNA sequence and IAA8 (At2g22670) are displayed for first codon and splicing regions. In-frame fusion between activation domain and the IAA8 sequence need further verification.

Using upstream region of *IAA8* as a template, Genomatix predicts a large number of AC-rich DNA elements similar to AACCAAAA and CCAACC elements found within the 237-bp region of *pXCP1*. As mentioned earlier these elements were shown to bind MTF. Thus, expression of both genes may be linked to myb-transcription factors. Collectively, these AC-rich regions shared between the *pXCP1* and *pIAA8* point to the possibility of cross talk between auxin-mediated and myb-mediated expression of these two xylem genes.

It is conceivable that GA contributes to the regulation of TE differentiation, and in particular TE cysteine protease gene expression, for at least four reasons; First, recent work with transgenic poplar has implicated gibberellins as regulators of xylem cell elongation (Eriksson et al., 2000); Second, like aleurone cells of small grains (see Cercós et al., 1999; Fath et al., 2000; Gubler et al., 2002) where protease gene expression is promoted by GAs, TEs express high levels of papain-like enzymes and undergo PCD; Third, a GA 3-β-hydroxylase gene known as *GA4* is a phloem- and/or cambium-specific gene (C. Zhao and E. Beers, unpublished) suggesting that active GAs are synthesized in tissues adjacent to the xylem, but apparently not in the xylem, thereby providing an effective mechanism for cambium or phloem cell-mediated, GA-dependent xylem differentiation; Fourth, GA effects on plant growth and development can be modulated by auxin (Fu and Harberd, 2003). The presense of GA Myb binding element, DOF-binding elements, and an ARF-bidning element within a region required for TE-specific expression suggests a possible interaction between MTF and ARF and DOF proteins that could direct auxin-mediated GA-regulated *XCP1* expression. Based on deletion analysis, the $5'\Delta^{181}$ which interrupted the AC-rich elements (Fig. 2.15) abolished GUS expression whereas the $5'\Delta^{313}$ preserved TE-specific GUS expression.

In preliminary testing, ABA (50µM) and GA_3 (50 µM) were independently applied to the apical meristems of four-week old *pXCP1::GUS* plants and GUS histochemical staining was conducted before and after treatment using newly emerging leaves four-day post treatment. There were no observable differences in GUS staining pattern or intensity correlated with GA_3 or ABA treatments although plant growth did respond as expected to treatments confirming efficacy of GA_3 and ABA applications. These observations suggested that *pXCP1* was not regulated by GA_3 or ABA under our laboratory conditions. Further testing is required to test the effects of GA_3 and ABA on XCP1 expression.

Conclusion

The restricted expression of XCP1 and its paralog XCP2 in differentiating TEs suggests that XCP1 and XCP2 perform TE-specific functions and that the expression of *XCP1* and *XCP2* is regulated by transcription factors that are expressed during TE formation. That the transcriptional regulation of *XCP1* may be mediated by auxin via interactions involving IAA8 is consistent with the well-established requirement for auxin in TE differentiation. The identification of new transcriptional regulators of XCP1 expression will ultimately lead to a broader understanding of other features of the regulation of TE differentiation. As cysteine proteases, XCP1 and XCP2 presumably degrade cellular proteins. Based on information from the current study, however, it does not appear that XCP2 is essential for normal differentiation of TEs. XCP2 may play a role that is redundant with other TE proteases, such as XCP1.

List of Supplemental Figures and Tables

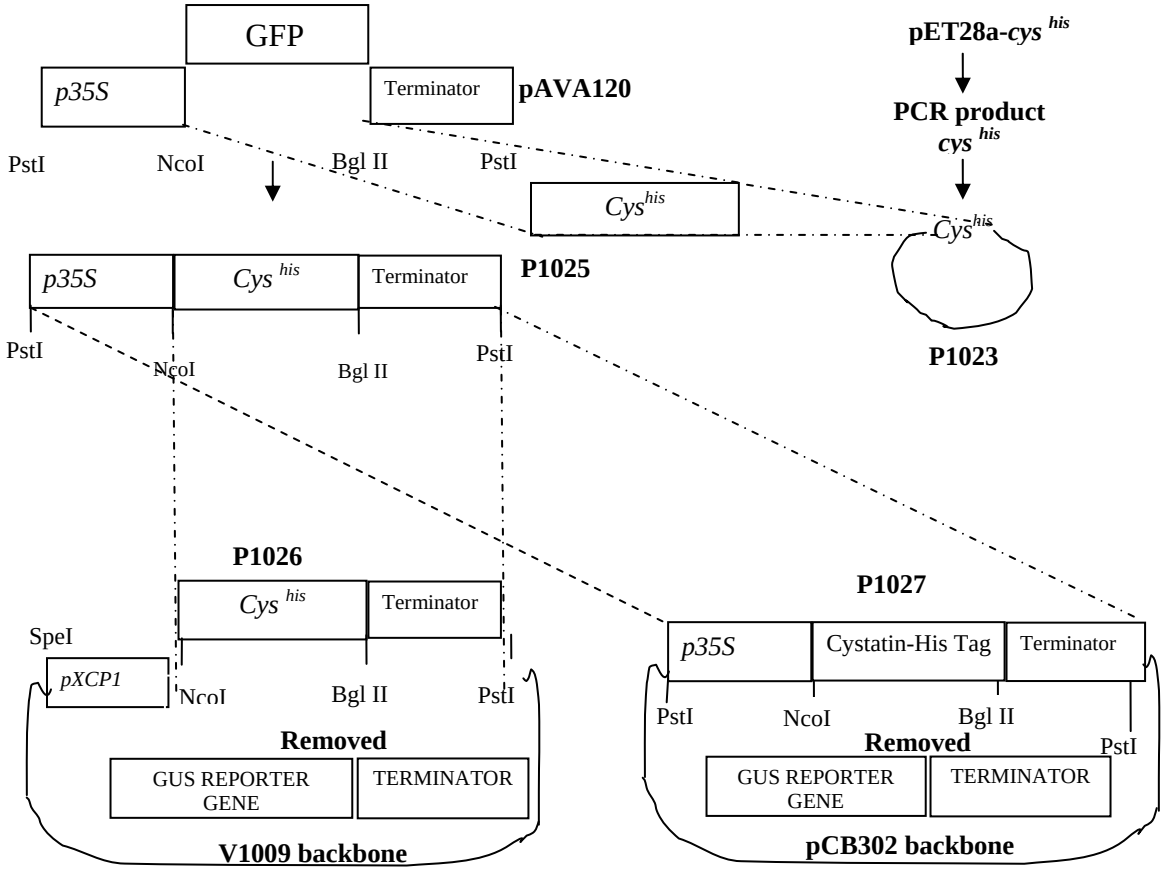


Figure S1. Construction of P1026 and P1027. The expression vector *pET-cys*^{his} served as a template for the amplification of a fragment containing *soyacystatin N* along with C-terminal 6x HIS tag (*cys*^{his}) using PR27 (TTAATACGACTCACTATAGGG) and PR33 (AGATCTTGTTAGCAGCCGGATCTCA). The amplified PCR product contains the endogenous *Nco*I restriction site corresponding to the first ATG of the soyacystatin N coding region. PCR product was then TA-cloned into pGEM to construct P1023. The presence of *cys*^{his} was verified by sequencing using SP6 primer endogenous to pGEM. P1023 served as a template to release an *Nco*I-*Bgl*II fragment containing *cys*^{his}. The *GFP*, green fluorescent protein, coding region was removed from the modified (Funk et al., 2002) pAVA120 (von Arnim et al., 1998) expression vector by *Nco*I-*Bgl*II digest. The linearized pAVA120 was used as an intermediate vector to receive the *Nco*I/*Bgl*II *cys*^{his} fragment, which led to the construction of plasmid P1025. P1025 was digested with *Pst*I to release a *p35S-cys*^{his}-Terminator fragment. This fragment was ligated to a dephosphorylated *Pst*I-linearized pCB302 to construct P1027 (*p35S::cys*^{his}). *Nco*I/*Pst*I *cys*^{his}-Terminator fragment from P1025 was inserted in *Nco*I/*Pst*I-linearized V1009, a pCB302-*pXCP1::GUS* vector (Kositsup, 2000). This led to the construction of P1026 which contains *pXCP1- cys*^{his} -Terminator (*pXCP1::cys*^{his}).

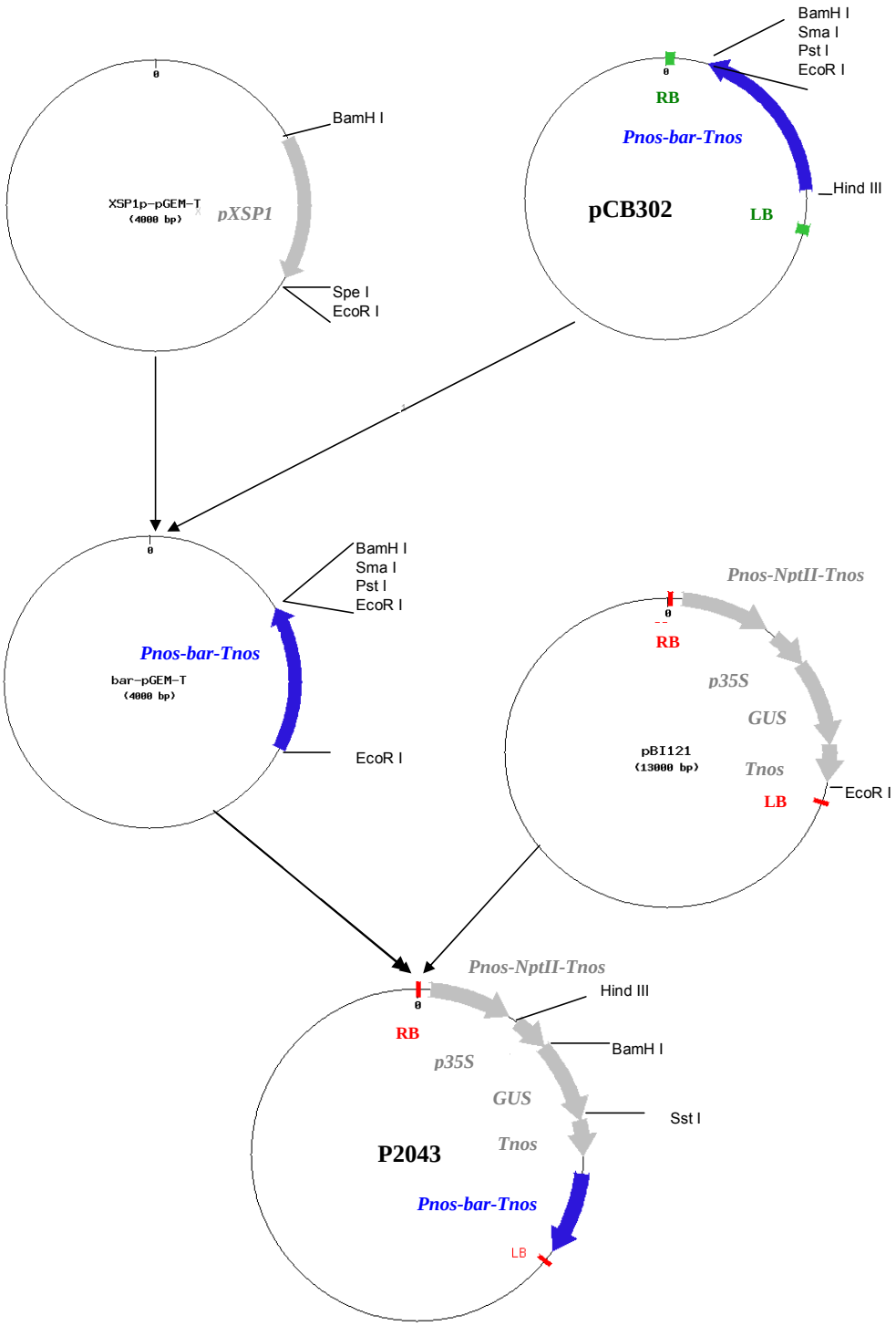


Figure S2. Construction of the recipient binary vector P2043.

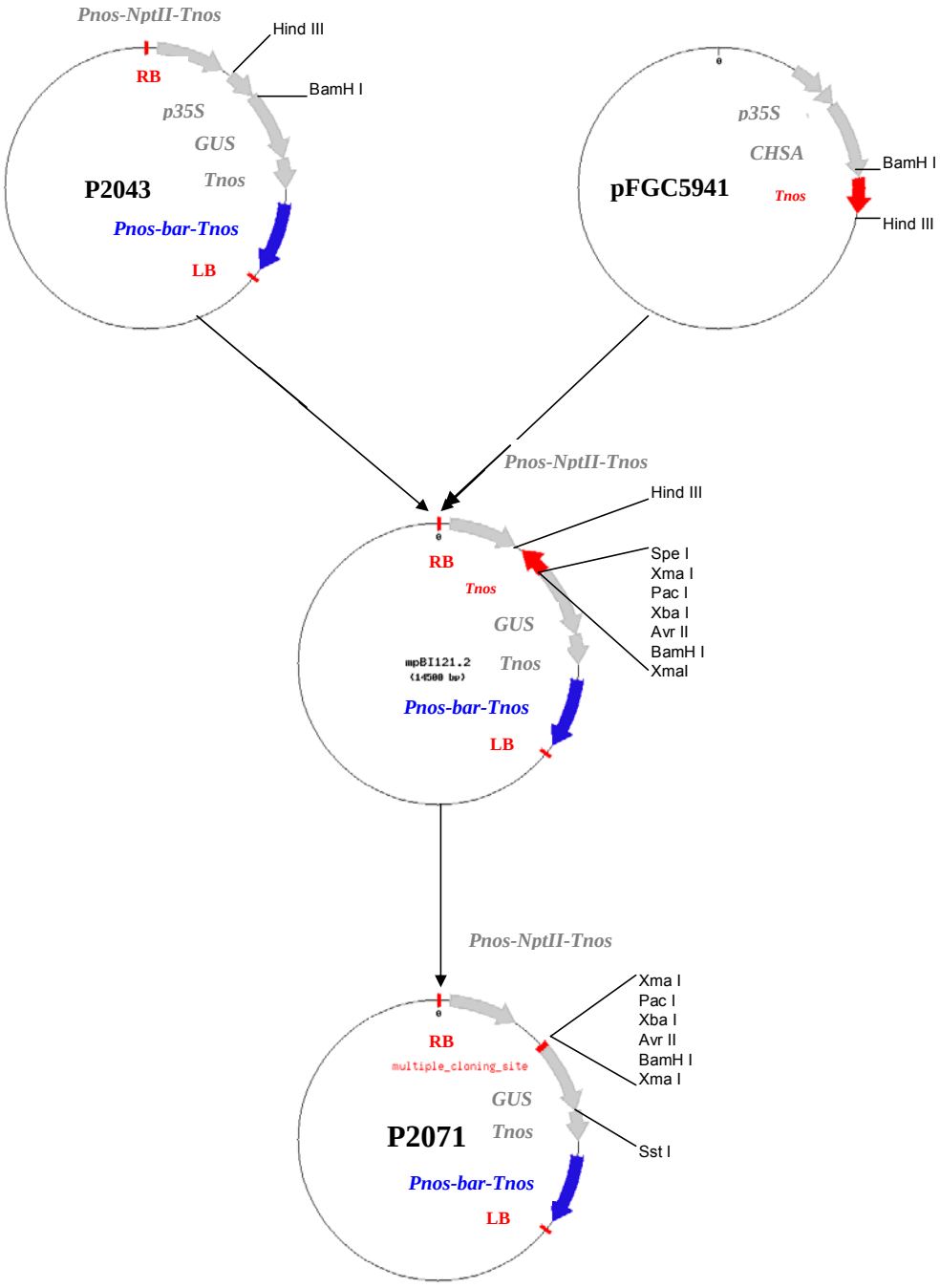


Figure S3. Construction of the recipient binary vector P2071.

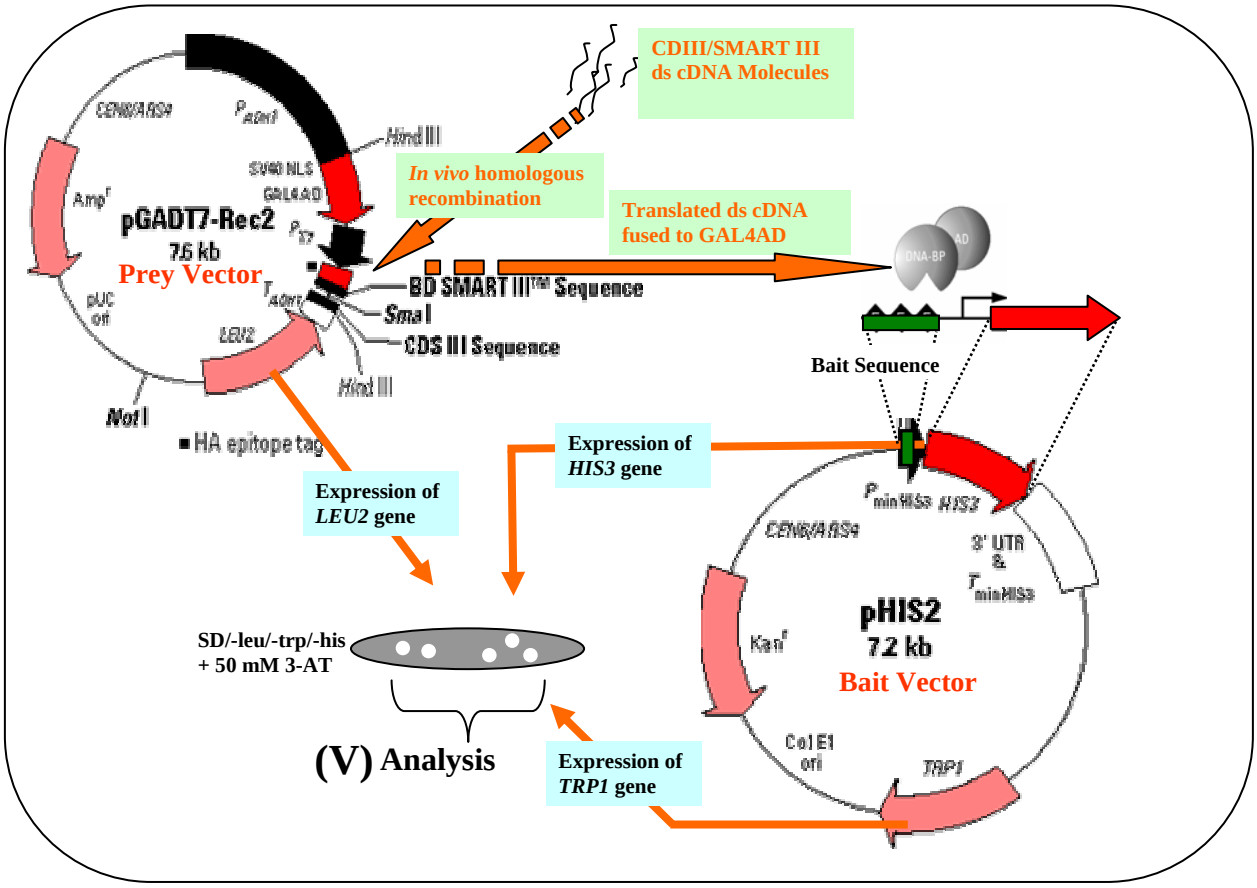


Figure S4. Identifying positive protein-DNA interactions using Match Maker yeast one-hybrid screen (Clontech). Yeast strain Y187 is engineered to be auxotrophic for histidine, leucine, and tryptophan. Y187 also has a leaky expression of *HIS3* gene which is inhibited by predetermined concentration of 3-AT. The DNA region of interest, referred to as bait, is cloned into the vector pHIS2 directly upstream of a minimal promoter-*HIS3* gene construct. When circularized by the bait DNA sequence, pHIS2 will facilitate yeast growth on media lacking tryptophan via the activation of *TRP1* expression. A cDNA library from the desired tissue is constructed in form of free double stranded cDNA molecules with ends designed to serve as prime sites for *in vivo* homolgous rescombination with the recipient vector pGADT7-rec2. When circularized by a ds cDNA molecule, pGADT7-Rec2 will facilitate yeast growth on media lacking leucine via the activation of *LEU2* expression. The pGADT-T7-Rec2 is constructed so that it translates the ds cDNA molecule fused in-frame with GAL4 activation domain (GAL4AD). The translated amino acid with binding affinity to bait DNA brings the fused GAL4AD into proximity to the minimal promoter of *HIS3* gene which leads to the transactivation of *HIS3* gene. Expression of HIS3 facilitates yeast growth on media lacking histidine. Yeast colonies growing on yeast synthetic media lacking histidine, tryptophan, and leucine and supplemented with inhibitory concentration of 3-AT putatively contain a polypeptide that specifically interacts with the bait DNA sequence. The cDNA sequence in pGADT7-rec2 can be amplified by specific primer pairs via PCR, sequenced by T7 primer, identified by BLAST, and further studied using downstream applications such as gel retardation assays or mutagenesis of the protein domain.

Table S1. PCR Primer Reference.

Primer	Sequence	Template	Orientation	5' Site
PR01	AAGCTTCTCATTTGTACG	*XCP1 Promoter*	Sense	*Hind*III
PR02	AAGCTTCAGAAATTTTAAGCCTCTACG	*XCP1 Promoter*	Sense	*Hind*III
PR03	AAGCTTGATCCAACCGTGAAGACTCG	*XCP1 Promoter*	Sense	*Hind*III
PR04	AAGCTTGCTCTATTCCTCTACTCTG	*XCP1 Promoter*	Sense	*Hind*III
PR05	GGATCCCCAAATTTGTTCACT	*XCP1 Promoter*	Antisense	*Bam*HI
PR06	ATCCAGACTGAATGCCCACA	*GUS*	Antisense	No
PR07	GCACTAGTGTGTTTGCACTTTGCAGG	*XCP1 Promoter*	Sense	*Spe*I
PR08	GCCATGGCCAAATTTGTTCACTGAG	*XCP1 Promoter*	Antisense	*Nco*I
PR09	CTTTCTACCTTCCCACAATTCG	pFGC5941	Antisense	No
PR10	TACTTACACTTGCCTTGGAGT	pFGC5941	Sense	No
PR11	TCACCGAAGTTCATGCCAGT	*GUS*	Antisense	No
PR12	ACGTCCTGTAGAAACCCCAA	*GUS*	Sense	No
PR13	TTACCTCTGACCGAGACAGA	XCP2	Sense	No
PR14	GGCAAGAGCAAAACAGAGGA	*XCP2*	Antisence	No
PR15	TTCATAACCAATCTCGATACAC	LB Garlic	Either	No
PR16	GCGTGGACCGCTTGCTGCAACT	LB Salk	Either	No
PR17	AGACACTGTAAGATGCTGATCA	*XCP2*	Antisense	No
PR18	TGTGACAGGAAACTTGACAACAT	*XCP2*	Sense	No
PR19	TCAAGATCAACCCCGCACCGC	*XCP2*	Antisense	No
PR20	GGTCTAGAGGCGCGCCATGTTACGTCCTGTAGAA	*GUS*	Sense	*Xba*I-*Asc*I
PR21	CAGGATCCATTTAAATTCATTGTTTGCCTCCCTG	*GUS*	Antisense	*Bam*HI/*Swa*I
PR22	TCTAGAGGCGCGCCCACCGTGGTGACGCATGT	*GUS*	Sense	*Xba*I/*Asc*I
PR23	GGATCCATTTAAATGCCTCTTCGCTGTACAGT	*GUS*	Antisense	*Bam*HI/*Swa*I
PR24	TCTAGAACCGTCGTCGCAGGTTAATA	*XCP2*	Sense	*Xba*I/*Asc*I
PR25	GGATCCAAAGAGCCGTTTGAGTACGT	*XCP2*	Antisense	*Bam*HI
PR26	ATTTAGGTGACACTATAGAA	pGEM	Antisense	No
PR27	TTAATACGACTCACTATAGGG	pGEM	Sense	No
PR28	TTATGCTCTACACCACGCCGAA	*GUS*	Sense	No
PR29	GGCATCGTGGTGATTGATGAAACTG	*GUS*	Sense	No
PR30	CTGTTCGGCGTGGTGTAGAGCATTA	*GUS*	Antisense	No
PR31	CAGTTTCATCAATCACCACGATGCC	*GUS*	Antisense	No
PR32	ATCTCTTCAGCGTAAGGGTAAT	*GUS*	Antisense	No
PR33	AGATCTTGTTAGCAGCCGGATCTCA	pET28b	Antisense	*Bgl* II
PR34	TCCCCTCGACAGCTGCAA	*XCP1 Gene*	Antisense	No
PR35	GTAGAGGATCGAGATCTCGA	pET28b	Sense	No
PR36	ATGATTGAACAAGATGGATTGCA	*NPT II*	Sense	No
PR37	TCAGAAGAACTCGTCAAGAAGG	*NPT II*	Antisense	No
PR38	GTCAACATGGTGGAGCACGACACA	*35S Promoter*	Sense	No
PR39	CCATGGAATTAGGTGGCATCACCGA	Cystatin	Sense	*Nco*I
PR40	TTGCCATGTCCTACACGCCGAA	pFGC5941	Sense	No
PR41	TCACTTGGTCTTGGTAGGAT	*XCP1 Gene*	Antisense	No
PR42	GGATCCGGCAATTGTTGTTATTGGAT	*Lotus japonicus*	Antisense	*Bam*HI
PR43	AAGCTTAAGGTGGGTTTAAAGCGGTG	*Lotus japonicus*	Sense	*Hind*III
PR44	AAGCTTCAGAAATTTTAAGCCTCTA	*XCP1 Promoter*	Sense	*Hind*III
PR45	GGATCCTTTTTGGTTAAGCCTTTT	*XCP1 Promoter*	Antisense	*Bam*HI
PR46	GGATCCCGCAAGACCCTTCCTCTA	*35S Promoter*	Sense	*Bam*HI
PR47	CCCGGGCGTGTTCTCTCCAAATGA	*35S Promoter*	Antisense	*Sma*I/*Xma*I
PR48	ATGGCTTTTTCTGCACCATC	*XCP1* Gene	Sense	No
PR49	TGATCCAACCGTGAAGACTCGGA	*XCP1* Promoter	Sense	No
PR50	AAATTGTAACTATTCTTGATAAGTTTT	*XCP1* Promoter	Antisense	No
PR51	CCCGGGTTCAGAAATTTTAAGCCT	*XCP1* Promoter	Sense	*Sma*I/*Xma*I

Table S1 (continued). PCR Primer Reference.

Primer	Sequence	Template	Orientation	5' Site
PR52	GAGCTCCAAGTTGGAGACAAGACA	*XCP1* Promoter	Antisense	*Sac*I
PR53	ATATTCTTCGAAGAAATCACAT	pHIS2	Antisense	No
PR54	ATGTGCATGGAGCTGCCAAC	Soybean Cystatin	Sense	No
PR55	CACATTGACTGGCTTGAATTCT	Soybean Cystatin	Antisense	No
PR47	CCCGGGCGTGTTCTCTCCAAATGA	35S Promoter	Antisense	*Sma*I/*Xma*I
PR48	ATGGCTTTTTCTGCACCATC	XCP1 Gene	Sense	No
PR49	TGATCCAACCGTGAAGACTCGGA	*XCP1* Promoter	Sense	No
PR50	AAATTGTAACTATTCTTGATAAGTTTT	*XCP1* Promoter	Antisense	No
PR51	CCCGGGTTCAGAAATTTTAAGCCT	*XCP1* Promoter	Sense	*Sma*I/*Xma*I
PR52	GAGCTCCAAGTTGGAGACAAGACA	*XCP1* Promoter	Antisense	*Sac*I
PR53	ATATTCTTCGAAGAAATCACAT	pHIS2	Antisense	No
PR54	ATGTGCATGGAGCTGCCAAC	Soybean Cystatin	Sense	No
PR55	CACATTGACTGGCTTGAATTCT	Soybean Cystatin	Antisense	No
PR56	CTGACTCATGGTACTCACTC	Actin	Antisense	No
PR57	GGCCGATGGTGAGGATATTC	Actin	Sense	No
PR58	CTGGATCCATTAAATGCCATCTTGTTGATTCCACAG	*XCP2* coding	Antisense	*Xba*I/*Asc*I
PR59	TCTAGAGGCGCGCGACAGGAAACTTGACAACATTGTC	*XCP2* coding	Sense	*Bam*HI/*Swa*I
PR60	TACTCCATCGTTGGATACTCC	*XCP2* coding	Sense	No
PR61	AGCGAACTCTGCGTAAGATCTT	*XCP2* coding	Antisense	No
PR62	GACTATGCCTTTGAGTACATT	*XCP2* coding	Sense	No
PR63	AAGACCAACCCCGCACCGC	XCP2 coding	Antisense	No
PR64	GGATCCACGCCGGTCTCCGAGTCTTCA	*XCP1* Promoter	Antisense	*Bam*HI
PR65	GGATCCGATTAATTTAAAAACTAAGTA	*XCP1* Promoter	Antisense	*Bam*HI
PR66	GGATCCCATAGGATTGGCTTTGAAGCA	*XCP1* Promoter	Antisense	*Bam*HI
PR67	GGATCCCAAGTTGGAGACAAGACA	*XCP1* Promoter	Antisense	*Bam*HI
PR68	AAGCTTCCCGGGAACAAAAGGCTT	*XCP1* Promoter	Sense	*Hind*III/*Xma*I
PR69	GGATCCGAGCTCTCACGGTTGGAT	*XCP1* Promoter	Antisense	*Bam*HI/*Sac*I
PR70	CTATTCGATGATGAAGATACCCCACCAAACCC	pGADT7-Rec2	Sense	No
PR71	GTGAACTTGCGGGGTTTTTCAGTATCTACGAT	pGADT7-Rec2	Antisense	No
PR72	ACAGGAAACAGCTATGACCA	pBI121	Sense	No

MAIN BIBLIOGRAPHY

Abel, S., Oeller, P.W., and Theologis, A. (1994). Early auxin-induced genes encode short-lived nuclear proteins. Proceedings of the National Academy of Sciences of the United States of America **91,** 326-330.

Abrahamson, M., Alvarez-Fernandez, M., and Nathanson, C.-M. (2003). Cystatins. In Proteases and the Regulation of Biological Processes, J. Saklatvala, H. Nagase, and G. Salvesen, eds (London, UK: Portland Press), pp. 179-199.

Aloni, R. (1987). Differentiation of Vascular Tissues. Annual Review of Plant Physiology and Plant Molecular Biology **38,** 179-204.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301,** 653-657.

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

Arciello, S., Corrado, G., and Rao, R. (2003). *Npt II* gene silencing in transgenic tobacco plants. Advances in Horticulture Science **17,** 231-235.

Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S. (2000). Generation of 7137 non-redundant expressed sequence tags from a legume, *Lotus japonicus*. DNA Research **7,** 127-130.

Baima, S., Possenti, M., Matteucci, A., Wisman, E., Altamura, M.M., Ruberti, I., and Morelli, G. (2001). The *Arabidopsis* ATHB-8 HD-zip protein acts as a differentiation-promoting transcription factor of the vascular meristems. Plant Physiology **126,** 643-655.

Barrett, A., Rawlings, N., and Woessner, J. (1998). Handbook of Proteolytic Enzymes. (London and San Diego: Acadmic Press).

Barrett, A.J., Rawlings, N.D., and O'Brien, E.A. (2001). The *MEROPS* database as a protease information system. Journal of Structural Biology **134,** 95-102.

Baulcombe, D., Saunders, G., Bevan, M., Mayo, M., and Harrison, B.M. (1986). Expression of biologically-active viral satellite RNA from the nuclear genome of transformed plants. Nature **321,** 446-449.

Bechtold, N., Ellis, J., and Pelletier, G. (1993). In-planta *Agrobacterium*-mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences **316,** 1194-1199.

Beers, E.P., and Freeman, T.B. (1997). Proteinase activity during tracheary element differentiation in zinnia mesophyll cultures. Plant Physiology **113,** 873-880.

Beers, E.P., Jones, A.M., and Dickerman, A.W. (2004). The S8 serine, C1A cysteine and Al aspartic protease families in *Arabidopsis*. Phytochemistry **65,** 43-58.

Beers, E.P., Moreno, T.N., and Callis, J. (1992). Subcellular-localization of ubiquitin and ubiquitinated proteins in *Arabidopsis thaliana*. Journal of Biological Chemistry **267,** 15432-15439.

Belenghi, B., Acconcia, F., Trovato, M., Perazzolli, M., Bocedi, A., Polticelli, F., Ascenzi, P., and Delledonne, M. (2003). AtCYS1, a cystatin from *Arabidopsis thaliana*, suppresses hypersensitive cell death. European Journal of Biochemistry **270,** 2593-2604.

Benfey, P.N., Ren, L., and Chua, N.H. (1990). Tissue-specific expression from Camv 35S-enhancer subdomains in early stages of plant development. European Molecular Biology Organization Journal **9,** 1677-1684.

Borevitz, J.O., Xia, Y.J., Blount, J., Dixon, R.A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell **12,** 2383-2393.

Botella, M.A., Xu, Y., Prabha, T.N., Zhao, Y., Narasimhan, M.L., Wilson, K.A., Nielsen, S.S., Bressan, R.A., and Hasegawa, P.M. (1996). Differential expression of soybean cysteine proteinase inhibitor genes during development and in response to wounding and methyl jasmonate. Plant Physiology **112,** 1201-1210.

Cannon, S.B., and Young, N.D. (2003). OrthoParaMap: Distinguishing orthologs from paralogs by integrating comparative genome data and gene phylogenies. BioMed Central Bioinformatics **4**, 35-40.

Cercós, M., Gomez-Cadenas, A., and Ho, T.H.D. (1999). Hormonal regulation of a cysteine proteinase gene, EPB-1, in barley aleurone layers: *cis*- and *trans*-acting elements involved in the co-ordinated gene expression regulated by gibberellins and abscisic acid. Plant Journal **19,** 107-118.

Chen, H., Nelson, R.S., and Sherwood, J.L. (1994). Enhanced recovery of transformants of *Agrobacterium tumefaciens* after freeze-thaw transformation and drug selection. Biotechniques **16,** 664-670.

Chen, P.Y., Wang, C.K., Soong, S.C., and To, K.Y. (2003). Complete sequence of the binary vector pBI121 and its application in cloning T-DNA insertion from transgenic plants. Molecular Breeding **11,** 287-293.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant Journal **16,** 735-743.

Davuluri, R.V., Sun, H., Palaniswamy, S.K., Matthews, N., Molina, C., Kurtz, M., and Grotewold, E. (2003). AGRIS: *Arabidopsis* Gene Regulatory Information Server, an information resource of *Arabidopsis cis*-regulatory elements and transcription factors. BioMed Central Bioinformatics **4,** 25-35.

Duan, X.L., Li, X.G., Xue, Q.Z., AboElSaad, M., Xu, D.P., and Wu, R. (1996). Transgenic rice plants harboring an introduced potato proteinase inhibitor II gene are insect resistant. Nature Biotechnology **14,** 494-498.

Endo, M., Hakozaki, H., Kokubun, T., Masuko, H., Takahata, Y., Tsuchiya, T., Higashitani, A., Tabata, S., and Watanabe, M. (2002). Generation of 919 expressed sequence tags from immature flower buds and gene expression analysis using expressed sequence tags in the model plant *Lotus japonicus*. Genes & Genetic Systems **77,** 277-282.

Endo, M., Kokubun, T., Takahata, Y., Higashitani, A., Tabata, S., and Watanabe, M. (2000). Analysis of expressed sequence tags of flower buds in *Lotus japonicus*. DNA Research **7,** 213-216.

Eriksson, M.E., Israelsson, M., Olsson, O., and Moritz, T. (2000). Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. Nature Biotechnology **18,** 784-788.

Fath, A., Bethke, P., Lonsdale, J., Meza-Romero, R., and Jones, R. (2000). Programmed cell death in cereal aleurone. Plant Molecular Biology **44,** 255-266.

Fernandes, K., Sabelli, P., Barratt, D., Richardson, M., Xavier-Filho, J., and Shewry, P. (1993). The resistance of cowpea seeds to bruchid beetles is not related to levels of cysteine proteinase inhibitors. Plant Molecular Biology **23,** 215-219.

Forsburg, S.L. (2001). The art and design of genetic screens: Yeast. Nature Reviews Genetics **2,** 659-668.

Frith, M.C., Hansen, U., and Weng, Z.P. (2001). Detection of *cis*-element clusters in higher eukaryotic DNA. Bioinformatics **17,** 878-889.

Frith, M.C., Li, M.C., and Weng, Z.P. (2003). Cluster-Buster: finding dense clusters of motifs in DNA sequences. Nucleic Acids Research **31,** 3666-3668.

Fu, X.D., and Harberd, N.P. (2003). Auxin promotes *Arabidopsis* root growth by modulating gibberellin response. Nature **421,** 740-743.

Funk, V., Kositsup, B., Zhao, C.S., and Beers, E.P. (2002). The *Arabidopsis* xylem peptidase XCP1 is a tracheary element vacuolar protein that may be a papain ortholog. Plant Physiology **128,** 84-94.

Gallagher, S. (1992). Quantitation of GUS Activity by Fluorometry. In GUS Protocols: Using the GUS Gene as a Reporter of Gene Expression, S.R. Gallagher, ed (Academic Press, Inc.), pp. 47-60.

Gardiner, J.C., Taylor, N.G., and Turner, S.R. (2003). Control of cellulose synthase complex localization in developing xylem. Plant Cell **15,** 1740-1748.

Graham, J., McNicol, R.J., and Greig, K. (1995). Towards genetic based insect resistance in strawberry using the Cowpea trypsin inhibitor gene. Annals of Applied Biology **127,** 163-173.

Groover, A.T., Pattishall, A., and Jones, A.M. (2003). IAA8 expression during vascular cell differentiation. Plant Molecular Biology **51,** 427-435.

Gruis, D., Selinger, D.A., Curran, J.M., and Jung, R. (2002). Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases. Plant Cell **14,** 2863-2882.

Gruis, D.F., Schulze, J., and Jung, R. (2004). Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases. Plant Cell **16,** 270-290.

Grzonka, Z., Jankowska, E., Kasprzykowski, F., Kasprzykowska, R., Lamkiewicz, L., Wiczk, W., Wieczerzak, E., Ciarkowski, J., Drabik, P., Janowski, R., Kozak, M., Jaskolski, M., and Grubb, A. (2001). Structural studies of cysteine proteases and their inhibitors. Acta Biochimica Polonica **48,** 1-20.

Gubler, F., Chandler, P.M., White, R.G., Llewellyn, D.J., and Jacobsen, J.V. (2002). Gibberellin signaling in barley aleurone cells. Control of SLN1 and GAMYB expression. Plant Physiology **129,** 191-200.

Gubler, F., Raventos, N., Keys, M., Watts, R., Mundy, J., and Jacobsen, J.V. (1999). Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. Plant Journal **17,** 1-9.

Guilfoyle, T., Hagen, G., Ulmasov, T., and Murfett, J. (1998a). How does auxin turn on genes? Plant Physiology **118,** 341-347.

Guilfoyle, T.J., Ulmasov, T., and Hagen, G. (1998b). The ARF family of transcription factors and their role in plant hormone-responsive transcription. Cellular and Molecular Life Sciences **54,** 619-627.

Handberg, K., and Stougaard, J. (1992). *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. Plant Journal **2,** 487-496.

Hannenhalli, S., and Levy, S. (2002). Predicting transcription factor synergism. Nucleic Acids Research **30,** 4278-4284.

Hauffe, K.D., Lee, S.P., Subramaniam, R., and Douglas, C.J. (1993). Combinatorial interactions between positive and negative *cis*-acting elements control spatial patterns of 4cl-1 expression in transgenic tobacco. Plant Journal **4,** 235-253.

Hegedus, D., Baldwin, D., O'Grady, M., Braun, L., Gleddie, S., Sharp, A., Lydiate, D., and Erlandson, M. (2003). Midgut proteases from *Mamestra configurata* (Lepidoptera: Noctuidae) larvae: Characterization, cDNA cloning, and exressed sequence tag analysis. Archives of Insect biochemistry and Physiology **53,** 30-47.

Hennion, S., Little, C.H.A., and Hartmann, C. (1992). Activities of enzymes involved in lignification during the postharvest storage of etiolated asparagus spears. Physiologia Plantarum **86,** 474-478.

Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M., and Jaskolski, M. (2001). Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. Nature Structural Biology **8,** 316-320.

Jefferson, R.A., Klass, M., Wolf, N., and Hirsh, D. (1987). Expression of chimeric genes in *Caenorhabditis elegans*. Journal of Molecular Biology **193,** 41-46.

Jegga, A.G., Sherwood, S.P., Carman, J.W., Pinski, A.T., Phillips, J.L., Pestian, J.P., and Aronow, B.J. (2002). Detection and visualization of compositionally similar *cis*-regulatory element clusters in orthologous and coordinately controlled genes. Genome Research **12,** 1408-1417.

Jerala, R., Trstenjakprebanda, M., Kroonzitko, L., Lenarcic, B., and Turk, V. (1990). Mutations in the QVVAG region of the cysteine proteinase inhibitor Stefin-B. Biological Chemistry Hoppe-Seyler **371,** 157-160.

Johannesson, H., Wang, Y., and Engstrom, P. (2001). DNA-binding and dimerization preferences of *Arabidopsis* homeodomain-leucine zipper transcription factors *in vitro*. Plant Molecular Biology **45,** 63-73.

Johnson, R., Narvaez, J., An, G., and Ryan, C. (1989). Expression of proteinase inhibitors I and II in transgenic tobacco plants: Effects on natural defense against *Manduca sexta* larvae. Proceedings of the National Academy of Sciences of the United States of America **86,** 9871-9875.

Kagaya, Y., Ohmiya, K., and Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. Nucleic Acids Research **27,** 470-478.

Karlsson, M. (2003). Molecular factors involved in the formation of secondary vascular tissues and lignification in higher Plants. Studies on the CuZn-SOD and members of MYB and Zinc-finger transcription factors families. Ph.D. Dissertation (Umea: Swedish University of Agriculture Sciences). pp 50.

Karrer, K.M., Peiffer, S.L., and Ditomas, M.E. (1993). Two distinct gene subfamilies within the family of cysteine protease genes. Proceedings of the National Academy of Sciences of the United States of America **90,** 3063-3067.

Klopfenstein, N.B., McNabb, H.S., Hart, E.R., Hall, R.B., Hanna, R.D., Heuchelin, S.A., Allen, K.K., Shi, N.Q., and Thornburg, R.W. (1993). Transformation of populus hybrids to study and improve pest resistance. Silvae Genetica **42,** 86-90.

Koiwa, H., D'Urzo, M.P., Assfalg-Machleidt, I., Zhu-Salzman, K., Shade, R.E., An, H.J., Murdock, L.L., Machleidt, W., Bressan, R.A., and Hasegawa, P.M. (2001). Phage display selection of hairpin loop soyacystatin variants that mediate high affinity inhibition of a cysteine proteinase. Plant Journal **27,** 383-391.

Koiwa, H., Shade, R.E., Zhu-Salzman, K., D'Urzo, M.P., Murdock, L.L., Bressan, R.A., and Hasegawa, P.M. (2000). A plant defensive cystatin (soyacystatin) targets cathepsin L-like digestive cysteine proteinases (DvCALs) in the larval midgut of western corn rootworm (*Diabrotica virgifera virgifera*). FEBS Letters **471,** 67-70.

Koiwa, H., Shade, R.E., Zhu-Salzman, K., Subramanian, L., Murdock, L.L., Nielsen, S.S., Bressan, R.A., and Hasegawa, P.M. (1998). Phage display selection can differentiate insecticidal activity of soybean cystatins. Plant Journal **14,** 371-379.

Konno, K., Hirayama, C., Nakamura, M., Tateishi, K., Tamura, Y., and Hattori, M. (2004). Papain protects papaya trees from herbivorous insects: role of cysteine proteases in latex. Plant Journal **37,** 370-378.

Kositsup, B. (2000). Tissue and cell-type localization and partial characterization of a xylem papain-type cysteine protease from *Arabidopsis*. Master Thesis (Blacksburg: Virginia Polytechnic Institute and State University), pp. 49.

Kuriyama, H., and Fukuda, H. (2001). Regulation of tracheary element differentiation. Journal of Plant Growth Regulation **20,** 35-51.

Lacombe, E., Van Doorsselaere, J., Boerjan, W., Boudet, A.M., and Grima-Pettenati, J. (2000). Characterization of *cis*-elements required for vascular expression of the Cinnamoyl CoA Reductase gene and for protein-DNA complex formation. Plant Journal **23,** 663-676.

Lauvergeat, V., Rech, P., Jauneau, A., Guez, C., Coutos-Thevenot, P., and Grima-Pettenati, J. (2002). The vascular expression pattern directed by the *Eucalyptus gunnii* cinnamyl alcohol dehydrogenase EgCAD2 promoter is conserved among woody and herbaceous plant species. Plant Molecular Biology **50,** 497-509.

Lee, M.M., and Schiefelbein, J. (1999). WEREWOLF, a MYB-related protein in *Arabidopsis*, is a position-dependent regulator of epidermal cell patterning. Cell **99,** 473-483.

Lee, M.W., Qi, M., and Yang, Y.O. (2001). A novel jasmonic acid-inducible rice myb gene associates with fungal infection and host cell death. Molecular Plant-Microbe Interactions **14,** 527-535.

Lev-Yadun, S. (1994). Induction of schlereid differentiation in the pith of *Arabidopsis thaliana* (L.) Heynh. Journal of Experimental Botany **45,** 1845-1849.

Linkohr, B.I., Williamson, L.C., Fitter, A.H., and Leyser, H.M.O. (2002). Nitrate and phosphate availability and distribution have different effects on root system architecture of *Arabidopsis*. Plant Journal **29,** 751-760.

Little, C., and MacDonald, J. (2004). Effects of exogenous gibberellin and auxin on shoot elongation and vegetative bud development in seedlings of *Pinus sylvestris* and *Picea glauca*. Tree Physiology **23,** 73-83.

Loopstra, C.A., and Sederoff, R.R. (1995). Xylem-specific gene expression in loblolly pine. Plant Molecular Biology **27,** 277-291.

López-Bucio, J., Cruz-Ramirez, A., and Herrera-Estrella, L. (2003). The role of nutrient availability in regulating root architecture. Current Opinion in Plant Biology **6,** 280-287.

López-Bucio, J., Hernandez-Abreu, E., Sanchez-Calderon, L., Nieto-Jacobo, M.F., Simpson, J., and Herrera-Estrella, L. (2002). Phosphate availability alters architecture and causes changes in hormone sensitivity in the *Arabidopsis* root system. Plant Physiology **129,** 244-256.

Lukowitz, W., Gillmor, C.S., and Scheible, W.R. (2000). Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. Plant Physiology **123,** 795-805.

Malamy, J.E., and Benfey, P.N. (1997). Organization and cell differentiation in lateral roots of *Arabidopsis thaliana*. Development **124,** 33-44.

Malberg, R.L. (1993). Production and Analysis of Plant Mutants, Emphasizing *Arabidopsis thaliana*. In Methods in Plant Molecular Biology and Biotechnology, B.R. Glick and J.E. Thompson, eds (CRC Press, Inc.), pp. 11-28.

Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S.Q., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., Liebert, C.A., Liu, C.L., Madej, T., Marchler, G.H., Mazumder, R., et al. (2003). CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Research **31,** 383-387.

Matsubayashi, Y., Takagi, L., Omura, N., Morita, A., and Sakagami, Y. (1999). The endogenous sulfated pentapeptide phytosulfokine-alpha stimulates tracheary element differentiation of isolated mesophyll cells of zinnia. Plant Physiology **120,** 1043-1048.

McCann, M., Stacy, N., and Roberts, K. (2000). Targeted cell death in xylogenesis. In Programmed Cell Death in Animals and Plants, J. Bryant, S. Huges, and J. Garland, eds (BIOS Scientific Publishers), pp. 193-201.

Moon, J., Salzman, R., Ahn, J.-E., Koiwa, H., and Zhue-Salzman, K. (2004). Transcriptional regulation in cowpea bruchid guts during adaptation to a plant defense protease inhibitor. Insect Molecular Biology **13,** 283-291.

Morita, A., Umemura, T., Kuroyanagi, M., Futsuhara, Y., Perata, P., and Yamaguchi, J. (1998). Functional dissection of a sugar-repressed alpha-amylase gene (RAmylA) promoter in rice embryos. FEBS Letters **423,** 81-85.

Muntz, K. (1998). Deposition of storage proteins. Plant Molecular Biology **38,** 77-99.

Murray, M., and Thompson, W. (1980). Rapid isolation of high-molecular-weight plant DNA. Nucleic Acids Research **8,** 4321-4325.

Müssig, C., and Altmann, T. (2003). Changes in gene expression in response to altered SHL transcript levels. Plant Molecular Biology **53,** 805-820.

Nagata, K., Kudo, N., Abe, K., Arai, S., and Tanokura, M. (2000). Three-dimensional solution structure of oryzacystatin-I, a cysteine proteinase inhibitor of the rice, *Oryza sativa* L. *japonica*. Biochemistry **39,** 14753-14760.

Nagata, N., Asami, T., and Yoshida, S. (2001). Brassinazole, an inhibitor of brassinosteroid biosynthesis, inhibits development of secondary xylem in cress plants (*Lepidium sativum*). Plant and Cell Physiology **42,** 1006-1011.

Nägler, D.K., and Mènard, R. (2003). Family C1 cysteine proteases: Biological diveristy or redundancy? Biological Chemistry **384,** 837-843.

Neidle, S. (2002). Principles of DNA-Protein Interaction. In Nucleic Acid Structure and Recognition, S. Neidle, ed (Oxford University Press, Inc.).

No, E.G., Zhou, Y.X., and Loopstra, C.A. (2000). Sequences upstream and downstream of two xylem-specific pine genes influence their expression. Plant Science **160,** 77-86.

Obara, K., and Fukuda, H. (2004). Programmed cell death in xylem differentiation. In Programmed Cell Death in Plants, J. Gray, ed (Blackwell Publishing), pp. 131-154.

Oh, S., Park, S., and Han, K.H. (2003). Transcriptional regulation of secondary growth in *Arabidopsis thaliana*. Journal of Experimental Botany **54,** 2709-2722.

Ohashi-Ito, K., Demura, T., and Fukuda, H. (2003). Analysis of the HD-Zip III homeobox gene ZeHB-13 in relation to vascular differentiation in zinnia. Plant and Cell Physiology **44,** S170-S170.

Ohgishi, M., Oka, A., Morelli, G., Ruberti, I., and Aoyama, T. (2001). Negative autoregulation of the *Arabidopsis* homeobox gene ATHB-2. Plant Journal **25,** 389-398.

Oono, Y., Chen, Q.G., Overvoorde, P.J., Kohler, C., and Theologis, A. (1998). age Mutants of *Arabidopsis* exhibit altered auxin-regulated gene expression. Plant Cell **10,** 1649-1662.

Park, J.Y., Kim, H.J., and Kim, J. (2002). Mutation in domain II of IAA1 confers diverse auxin-related phenotypes and represses auxin-activated expression of Aux/IAA genes in steroid regulator-inducible system. Plant Journal **32,** 669-683.

Patzlaff, A., McInnis, S., Courtenay, A., Surman, C., Newman, L.J., Smith, C., Bevan, M.W., Mansfield, S., Whetten, R.W., Sederoff, R.R., and Campbell, M.M. (2003). Characterisation of a pine MYB that regulates lignification. Plant Journal **36,** 743-754.

Perry, J.A., Wang, T.L., Welham, T.J., Gardner, S., Pike, J.M., Yoshida, S., and Parniske, M. (2003). A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. Plant Physiology **131,** 866-871.

Peters, J.L., Cnudde, F., and Gerats, T. (2003). Forward genetics and map-based cloning approaches. Trends in Plant Science **8,** 484-491.

Rogers, B.L., Pollock, J., Klapper, D.G., and Griffith, I.J. (1993). Sequence of the proteinase inhibitor cystatin homolog from the pollen of *Ambrosia artemisiifolia* (Short Ragweed). Gene **133,** 219-221.

Rose, A., Meier, I., and Wienand, U. (1999). The tomato I-box binding factor LeMYBI is a member of a novel class of Myb-like proteins. Plant Journal **20,** 641-652.

Rupp, H.M., Frank, M., Werner, T., Strnad, M., and Schmulling, T. (1999). Increased steady state mRNA levels of the STM and KNAT1 homeobox genes in cytokinin overproducing *Arabidopsis thaliana* indicate a role for cytokinins in the shoot apical meristem. Plant Journal **18,** 557-563.

Sachs, T. (2000). Integrating cellular and organismic aspects of vascular differentiation. Plant and Cell Physiology **41,** 649-656.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor Laboratory Press, New York, NY).

Santi, L., Wang, Y.M., Stile, M.R., Berendzen, K., Wanke, D., Roig, C., Pozzi, C., Muller, K., Muller, J., Rohde, W., and Salamini, F. (2003). The GA octodinucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene Bkn3. Plant Journal **34,** 813-826.

Seguin, A., Laible, G., Leyva, A., Dixon, R.A., and Lamb, C.J. (1997). Characterization of a gene encoding a DNA-binding protein that interacts *in vitro* with vascular specific *cis* elements of the phenylalanine ammonia-lyase promoter. Plant Molecular Biology **35,** 281-291.

Shahmuradov, I.A., Gammerman, A.J., Hancock, J.M., Bramley, P.M., and Solovyev, V.V. (2003). PlantProm: a database of plant promoter sequences. Nucleic Acids Research **31,** 114-117.

Simpson, S.D., Nakashima, K., Narusaka, Y., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Two different novel *cis*-acting elements of erd1, a clpA homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. Plant Journal **33,** 259-270.

Söderman, E., Hjellstrom, M., Fahleson, J., and Engstrom, P. (1999). The HD-Zip gene ATHB6 in *Arabidopsis* is expressed in developing leaves, roots and carpels and up-regulated by water deficit conditions. Plant Molecular Biology **40,** 1073-1083.

Solomon, M., Belenghi, B., Delledonne, M., Menachem, E., and Levine, A. (1999). The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. Plant Cell **11,** 431-443.

Steffens, N.O., Galuschka, C., Schindler, M., Bulow, L., and Hehl, R. (2004). AthaMap: an online resource for *in silico* transcription factor binding sites in the *Arabidopsis thaliana* genome. Nucleic Acids Research **32,** D368-D372.

Stomp, A.-M. (1992). Histochemical Localization of *β*-Glucuronidase. In GUS Protocols: Using GUS Gene as a Reporter of Gene Expression, S.R. Gallagher, ed (Academic Press, Inc), pp. 103-115.

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. Current Opinion in Plant Biology **4,** 447-456.

Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990). The refined 2.4Å X-ray crystal-Structure of recombinant human stefin-B in complex with the cysteine proteinase papain- a novel type of proteinase-inhibitor interaction. European Molecular Biology Organization Journal **9,** 1939-1947.

Sumiyama, K., Kim, C.B., and Ruddle, F.H. (2001). An efficient *cis*-element discovery method using multiple sequence comparisons based on evolutionary relationships. Genomics **71,** 260-262.

Sutoh, K., and Yamauchi, D. (2003). Two *cis*-acting elements necessary and sufficient for gibberellin-upregulated proteinase expression in rice seeds. Plant Journal **34,** 635-645.

Tamagnone, L., Merida, A., Parr, A., Mackay, S., Culianez-Macia, F.A., Roberts, K., and Martin, C. (1998). The AmMYB308 and AmMYB330 transcription factors from antirrhinum regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. Plant Cell **10,** 135-154.

Taneyama, M., Okamoto, T., Yamane, H., and Minamikawa, T. (2001). Involvement of gibberellins in expression of a cysteine proteinase (SH-EP) in cotyledons of *Vigna mungo* seedlings. Plant and Cell Physiology **42,** 1290-1293.

Taylor, N.G., Howells, R.M., Huttly, A.K., Vickers, K., and Turner, S.R. (2003). Interactions among three distinct CesA proteins essential for cellulose synthesis. Proceedings of the National Academy of Sciences of the United States of America **100,** 1450-1455.

Taylor, N.G., Laurie, S., and Turner, S.R. (2000). Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell **12,** 2529-2540.

Taylor, N.G., Scheible, W.R., Cutler, S., Somerville, C.R., and Turner, S.R. (1999). The irregular xylem3 locus of arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell **11,** 769-779.

Thomas, J.C., Adams, D.G., Keppenne, V.D., Wasmann, C.C., Brown, J.K., Kanost, M.R., and Bohnert, H.J. (1995). Protease inhibitors of *Manduca sexta* expressed in transgenic cotton. Plant Cell Reports **14,** 758-762.

Thomas, J.C., Wasmann, C.C., Echt, C., Dunn, R.L., Bohnert, H.J., and McCoy, T.J. (1994). Introduction and expression of an insect proteinase inhibitor in Alfalfa (*Medicago Sativa* L). Plant Cell Reports **14,** 31-36.

Thompson, J., Higgins, D., and Gibson, T. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weigthing, position-specific gap penalities and weight matrix choice. Nucleic Acids Research **22,** 4673-4680.

Thum, K.E., Kim, M., Morishige, D.T., Eibl, C., Koop, H.U., and Mullet, J.E. (2001). Analysis of barley chloroplast psbD light-responsive promoter elements in transplastomic tobacco. Plant Molecular Biology **47,** 353-366.

Turk, B., Turk, D., and Turk, V. (2000). Lysosomal cysteine proteases: more than scavengers. Biochimica Biophysica Acta-Protein Structure and Molecular Enzymology **1477,** 98-111.

Turk, D., Turk, B., and Turk, V. (2003). Papain-like lysosomal cysteine proteases and their inhibitors: drug discivery targets? In Proteases and regulation of biological processes, J. Saklatvala, H. Nagase, and G. Salvesen, eds (London: Portland Press), pp. 15-30.

Turner, S.R., and Somerville, C.R. (1997). Collapsed xylem phenotype of *Arabidopsis* identifies mutants deficient in cellulose deposition in the secondary cell wall. Plant Cell **9,** 689-701.

Ulmasov, T., Hagen, G., and Guilfoyle, T.J. (1997). ARF1, a transcription factor that binds to auxin response elements. Science **276,** 1865-1868.

Urao, T., Yamaguchi-Shinozaki, K., Urao, S., and Shinozaki, K. (1993). An *Arabidopsis* Myb homolog is induced by dehydration stress and its gene product binds to the conserved Myb recognition sequence. Plant Cell **5,** 1529-1539.

Venter, M., and Botha, F.C. (2004). Promoter analysis and transcription profiling: Integration of genetic data enhances understanding of gene expression. Physiologia Plantarum **120,** 74-83.

von Arnim, A.G., Deng, X.W., and Stacey, M.G. (1998). Cloning vectors for the expression of green fluorescent protein fusion proteins in transgenic plants. Gene **221,** 35-43.

Wang, R.C., Guegler, K., LaBrie, S.T., and Crawford, N.M. (2000). Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. Plant Cell **12,** 1491-1509.

Wagstaff, C., Leverentz, M.K., Griffiths, G., Thomas, B., Chanasut, U., Stead, A.D., and Rogers, H.J. (2002). Cysteine protease gene expression and proteolytic activity during senescence of *Alstroemeria* petals. Journal of Experimental Botany **53,** 233-240.

Washio, K. (2001). Identification of Dof proteins with implication in the gibberellin-regulated expression of a peptidase gene following the germination of rice grains. Biochimica Biophysica Acta-Gene Structure and Expression **1520,** 54-62.

Weaver, L.M., Gan, S.S., Quirino, B., and Amasino, R.M. (1998). A comparison of the expression patterns of several senescence-associated genes in response to stress and hormone treatment. Plant Molecular Biology **37,** 455-469.

Wei, Z., Angerer, R.C., and Angerer, L.M. (1999). Identification of a new sea urchin Ets protein, SpEts4, by yeast one-hybrid screening with the hatching enzyme promoter. Molecular and Cellular Biology **19,** 1271-1278.

Wiederanders, B., Kaulmann, G., and Schilling, K. (2003). Functions of propeptide parts in cysteine proteases. Current Protein & Peptide Science **4,** 309-326.

Woffenden, B.J., Freeman, T.B., and Beers, E.P. (1998). Proteasome inhibitors prevent tracheary element differentiation in zinnia mesophyll cell cultures. Plant Physiology **118,** 419-429.

Xiang, C.B., Han, P., Lutziger, I., Wang, K., and Oliver, D.J. (1999). A mini binary vector series for plant transformation. Plant Molecular Biology **40,** 711-717.

Xu, F.X., and Chye, M.L. (1999). Expression of cysteine proteinase during developmental events associated with programmed cell death in brinjal. Plant Journal **17,** 321-327.

Yamamoto, R., Demura, T., and Fukuda, H. (1997). Brassinosteroids induce entry into the final stage of tracheary element differentiation in cultured zinnia cells. Plant and Cell Physiology **38,** 980-983.

Yamamoto, Y., Kurata, M., Watabe, S., Murakami, R., and Takahashi, S.Y. (2002). Novel cysteine proteinase inhibitors homologous to the proregions of cysteine proteinases. Current Protein & Peptide Science **3,** 231-238.

Yanagisawa, S., and Schmidt, R.J. (1999). Diversity and similarity among recognition sequences of Dof transcription factors. Plant Journal **17,** 209-214.

Zhang, Y., Brown, G., Whetten, R., Loopstra, C.A., Neale, D., Kieliszewski, M.J., and Sederoff, R.R. (2003). An arabinogalactan protein associated with secondary cell wall formation in differentiating xylem of loblolly pine. Plant Molecular Biology **52,** 91-102.

Zhao, C.S., Johnson, B.J., Kositsup, B., and Beers, E.P. (2000). Exploiting secondary growth in *Arabidopsis*. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. Plant Physiology **123,** 1185-1196.

Zhao, Y., Botella, M.A., Subramanian, L., Niu, X.M., Nielsen, S.S., Bressan, R.A., and Hasegawa, P.M. (1996). Two wound-inducible soybean cysteine proteinase inhibitors have greater insect digestive proteinase inhibitory activities than a constitutive homolog. Plant Physiology **111,** 1299-1306.

Zhu-Salzman, K., Koiwa, H., Salzman, R.A., Shade, R.E., and Ahn, J.E. (2003). Cowpea bruchid *Callosobruchus maculatus* uses a three-component strategy to overcome a plant defensive cysteine protease inhibitor. Insect Molecular Biology **12,** 135-145.

Curriculum Vitae

3517 Bent Tree Lane, #310
Stow, Ohio 44224

E-mail: is52mih@vt.edu
Residence: 330-922-0761

<http://myprofile.cos.com/ihismail>
-US Permanent Resident-

Qualifications

- ♦ Education and research in molecular genetics, tissue culture, and plant production
- ♦ Advising, administrative, teaching, and management skills
- ♦ Application of computer in research, teaching, presentation, and office work
- ♦ Fluent in English and Arabic languages

Education

Ph.D. Horticulture/Molecular Cell Biology and Biotechnology. 01/1999-6/2004. GPA 3.86/4.00. Virginia Polytechnic Institute & State University, Blacksburg, VA

M.Sc. Horticulture/Genetics and Tissue Culture, 08/1991-06/1994.
GPA 3.00/4.00. The Ohio State University, Columbus, Ohio

B.Sc. Horticulture/Production and Management. 06/1987-06/1991 GPA 4.68/5.00 (Academic Honor). King Saűd University, Riyadh, Saudi Arabia

Contributions

1. Ismail, I.O. 2004. Function and Regulation of Xylem Cysteine Protease 1 and 2 in *Arabidopsis*.Ph.D. dissertation. Virginia Tech Press (submitted).

2. Ismail, I.O., Zhao, C.Z., and E. P. Beers. 2004. Identification of transacting factors that interact with *cis*-elements required for tracheary elements specific expression of *XCP1*. In preparation for Plant Molecular Biology.

3. Ismail, I. O., E.P. Beers, et al. 2004. Effects of over-expression, silencing, and inhibition of the tracheary element-specific proteases XCP1 and XCP2 on plant growth and subcellular structure. In preparation for Plant Physiology.

4. Ismail I.O. Towards Understanding Transcriptional Regulation of Xylem Cysteine Protease 1 in *Arabidopsis*. Poster, 20th Annual Research Symposium and Exposition. March 23, 2004. Blacksburg, Virginia.

5. Ismail, I.O., Zhao, C., and E. P. Beers. Deletion analysis of two promoters involved in tracheary elements differentiation in *Arabidopsis*. Poster, 14th International Conference *Arabidopsis* Research. June 20-June 24, 2003. Madison, Wisconsin.

6. Ismail, I.O. (ed.) 1997. Plant Tissue Culture from Physiological, Genetic, and Practical Perspective. Dar Al-Fikr Press. Academic text (Arabic). p189. Amman, Jordan.

7. Ismail, I.O. Protoplast fusion: is it superior to conventional hybridization? Poster, the College of Agriculture Minisymposium. June 1996. Amman, Jordan.

8. Ismail, I. O. 1994. Analysis and preliminary comparison of selfed progeny from *Brassica napus* resynthesized by both somatic & sexual hybridization of the progenitor species. Master Thesis. The Ohio State University Press. Columbus, Ohio.

Contributions (continued)

Future direct on-line submission of:

1. DNA Sequence, T-DNA/gDNA junction in SALK_010938 insertional mutant
2. DNA Sequence, T-DNA/gDNA junction in SALK_057921 insertional mutant
3. DNA Sequence, T-DNA/gDNA junction in SAIL A1261_A12 insertional mutant

Technical Proficiency

- ♦ Analysis at the DNA, RNA, Protein, and plant levels
- ♦ DNA-protein interactions in yeast
- ♦ Chemical- and PCR-based mutagenesis
- ♦ Molecular marker-assisted breeding
- ♦ Field Experimental design and greenhouse management

Advanced Courses

- ♦ Eukaryotic Molecular Gene Expression
- ♦ Biotechnology; Biochemistry; Biostatistics; Bioinformatics
- ♦ Genetics; Breeding; Cytology; Cell Tissue Culture; Plant Pathology

Intramural Funding

- ♦ Partial funding for a Ph.D. project, October 2003. Awarded by the Graduate Research and Development Program at Virginia Tech. $375.00
- ♦ Travel for 14th International Conference *Arabidopsis* Research, June 2003. Awarded by the department of horticulture at Virginia Tech. $600.00
- ♦ Travel for *Arabidopsis* Minisymposium, April 2002. Awarded by the department of horticulture at Virginia Tech. $100.00
- ♦ Tuition scholarship from Phyllis G. & Reginald H. Nelson, Phyllis G. & Reginald Foundation, 2001/2002 fiscal year. $ 5,000.00

Extracurricular Activities

- ♦ Served as a departmental delegate to Graduate Student Assembly-Virginia Tech
- ♦ Prepared proposals for NSF (doctoral), and Sigma Xi (research funding)
- ♦ Mentored and advised summer interns in molecular biology research

Teaching

- ♦ Developed course material and laboratory sessions in tissue culture and physiology
- ♦ Taught undergraduate students (US and abroad) and trained summer interns
- ♦ Acquired teaching skills through workshops at Virginia Tech
- ♦ Utilized MS word, Excel, Power Point, and on-line genomic databases

Recognition

- ♦ Nominated for Johnson Scholarship in Advanced Research and Teaching, Virginia Tech, Spring 2004.
- ♦ Invited to the 5th Annual Henry J. Decker Leadership Conference, Virginia Tech, March 3rd, 2004.
- ♦ Nomination for membership based on academic achievement from Gamma Sigma Delta, the Honor Society of Agriculture-Virginia Tech Institute Chapter, Spring 2004.
- ♦ Named an Associate Member at Sigma Xi Honor Society-Virginia Tech Institute Chapter, 2001.

Employment

♦ 09/2000- 06/2004: Ph.D. Graduate Research Assistant. Horticulture, Virginia Tech.

I have investigated the function and regulation of two xylem cysteine proteases, *XCP1 and XCP2* in the model eukaryote *Arabidopsis*. Proteins that potentially regulate XCP1 were identified via the yeast one-hybrid screen. Critical *cis*-elements were identified via XCP1 promoter mutagenesis. Evidence showing that XCP2 does not play a major role in *Arabidopsis* development was presented through T-DNA insertional mutagenesis, post-transcriptional gene silencing, and post-translational inhibition of cysteine proteases.

♦ 1/2001- 06/2001: Ph.D. Graduate Teaching Assistant. Horticulture, Virginia Tech

I took full responsibility of lecture and laboratory presentations, exam preparation, and grading for one undergraduate course in horticulture. My teaching skills were evaluated by the department and students and I was recommended for extension the following semester.

♦ 06/1994- 12/1998: Industrial experience in horticulture in USA and abroad

Jordan (1996): I was appointed the assistant manager of Celtics Laboratories where I managed 9 employees, 6 greenhouses and 1 tissue culture facility to produce and market ornamental crops. As a self-employed agricultural coordinator, I developed marketing reports and mediated trade communications for local small businesses.

Ohio, USA (1994/95; 1997/98): I worked as horticulturist in Boulevard Gardens Nursery and Fairlawn Country Club where I performed tasks in residential and commercial landscaping, ornamental propagation and care, and gained customer service experience.

Additional Research

♦ 01/1999-06/2000: Ph.D. level, laboratory of Dr. R.E. Veilleux, Virginia Tech

I produced androgenic dihaploids to develop resistance to late blight in potato and participated in molecular marker-assisted polymorphism analysis. I gained experience in potato genetics and breeding and pathology of *phytophthora infestans.* I worked in laboratory, field, and greenhouse settings.

♦ 08/1991-06/1994: Master level, laboratory of Dr. P.S. Jourdan, Ohio State University

I worked and reported on the genetic assessment of protoplast fusion as a source of germplasm enhancement compared to sexual hybridization in *Brassica napus*. Experience was gained in cell tissue culture, breeding and genetics of allopolyploidy in *Brassica spp*.

References

Contact information will be provided upon request

1. Dr. Eric P. Beers, advisor, Horticulture
2. Dr. Jerzy Nowak, committee member and department head, Horticulture
3. Dr. John L. Hess, committee member and department head, Biochemistry
4. Dr. Greg Welbaum, committee member, Horticulture
5. Dr. Glenda Gillaspy, Biochemistry
6. Dr. Jerry Williams, teaching coordinator, Horticulture