DISSERTATION

THE INFLUENCE OF EXTENSIN CROSS-LINKING ON BIOMASS RECALCITRANCE

Submitted by

Margaret Brigham Fleming

Department of Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2015

Doctoral Committee:

Advisor: Patricia Bedinger

Marinus Pilon
Stephen Decker
C. Kenneth Kassenbrock
Christie Peebles

ABSTRACT

THE INFLUENCE OF EXTENSIN CROSS-LINKING ON BIOMASS RECALCITRANCE

Plant cell walls are under investigation as a source for biofuel production, yet conversion of cell walls (biomass) into biofuel is currently too expensive to be competitive with gasoline. Biomass is recalcitrant; that is, it resists enzymatic degradation by cellulases into monosaccharides such as glucose. One source of recalcitrance may be the presence of extensins, covalently bound cell wall proteins that are extremely insoluble.

To determine what influence, if any, extensins have on biomass recalcitrance, I performed several experiments. I first turned to poplar biomass, which is a model source for biofuels. I found that protease treatment of poplar biomass after liquid hot water pretreatment reduced the hydroxyproline content (a proxy for extensins). The reduction in hydroxyproline content correlated with reduced recalcitrance, seen as an increase in glucose release after cellulase digestion of poplar biomass. I also tested whether *Arabidopsis* T-DNA insertional mutations in the genes encoding enzymes that perform extensin post-translational modifications could reduce extensin content or cross-linking, and whether this reduction was associated with reduced biomass recalcitrance. I found that although these mutants were hypothesized to have reduced incorporation of extensin in cell walls, no significant effects on extensin content in inflorescence stem cell walls (an analog for woody biomass), nor on glucose release from biomass, were found in any mutant line. Finally, I looked at the effects of extensin overexpression on glucose release in transgenic Arabidopsis lines containing synthetic genes encoding the complete extensin domain from *SlLRX1* or a short C-terminal region of 20 amino

acids of *SlLRX1*, fused to the red fluorescent reporter protein tdTomato. Observation of the tdTomato fluorescence in transgenic biomass after various chemical and enzymatic treatments indicated that the C-terminal 20 amino acids of *SlLRX1* are sufficient to allow a strong association with the cell wall, while the complete *SlLRX1* extensin domain leads to an even stronger, perhaps covalent linkage. Lines transformed with the complete *SlLRX1* extensin domain had more than twice the hydroxyproline content in their stems than wild-type, but this increase in hydroxyproline did not affect the amount of glucose released from stems upon cellulase digestion.

Since protease treatment reduced both hydroxyproline content and recalcitrance in poplar biomass, further experiments to assess the nature of the association between extensins and cell walls are warranted to attempt to further reduce recalcitrance. In the experiments I performed, the stems of extensin modification mutant Arabidopsis lines showed no change in extensin modification, and therefore no effect on recalcitrance was observed; stems of transgenic overexpression Arabidopsis lines showed increased extensin content, but again, no effect on recalcitrance was observed. My investigations in Arabidopsis focused on stem tissue, as this is analogous to material used in biofuel production. However, extensins are most abundantly expressed in roots in many plants, particularly in Arabidopsis. Examination of roots of both mutant and transgenic Arabidopsis may be more revealing of the interactions between extensins, cell walls, and recalcitrance.

ACKNOWLEDGEMENTS

I here acknowledge my many supporters, aiders and abettors, who contributed to the making of this PhD:

My financial supporters: the Clean Energy Supercluster who funded a seed grant (2011), the Colorado Center for Biorefining and Biofuels who funded a fellowship (#013-7), the Sustainable Biofuels Development Center; my graduate fellowships in the Program in Molecular Plant Biology and Interdisciplinary Graduate Education Research and Training-Multidisciplinary Approaches to Sustainable Bioenergy; and the CSU Biology Department.

My committee, and especially my advisor, Dr. Pat Bedinger.

The sowers, sorters and counters of seeds, runners of PCRs and gels, and pourers of plates, especially Rachel Chayer, Olivia Todd, Bronwyn Phillips, John Fleming, David Crawford, and my very first mentee who has stuck with me through it all, Eric Patterson.

The laboratories of Drs. Reddy, Stack, Garrity, Kanatous, Pilon, Bush, Kassenbrock, Peebles, and Gaines, without whose reagents, equipment, and know-how I could not have performed these experiments.

The sources of post-doc wisdom: Drs. Amanda Broz, April Randle, Bettina Broeckling, and Karl Ravet.

Barb Gibson and Dr. Todd Gaines, who have confidence in me.

The members of the Lay Down, Sister: we had a rockin' good time.

Dr. Eva Fischer, who is a good friend.

Sue and Lloyd Patterson, providers of vital baby-sitting services, beer, and skittles.

All my family.

AUTOBIOGRAPHY

I decided to apply to graduate school to study plants as I was uncharacteristically riding my bicycle to my job as an art framer at Michaels Arts and Crafts. I had recently graduated with a BA from St. John's College in philosophy and history of math and science, and had not yet found a suitable calling. On this bike ride I had the epiphany that plants were something that could be studied as a job, and that I wanted that job. I have never looked back from that moment.

I searched for graduate schools whose programs included something about chemistry of plant cells, did not require subject GREs, and were located west of the Mississippi. CSU met these criteria, so I applied and was accepted to the Program in Molecular Plant Biology.

I arrived in Dr. Pilon's lab for my first rotation: he accepted me enthusiastically although I had never heard of the central dogma, never held a pipette, and never seen an Arabidopsis plant. We cleared those hurdles in the first week, and now, through his help and the help of many others, I am pleased to call myself a scientist and a molecular biologist to boot.

DEDICATION

To the next generation: Henry, Oscar, Casper, Penelope, Harvey, John, and those yet to come. May you work your work, as I have worked mine.

To my parents, who support intellectual inquiry in all forms, including torturing Brassicas.

And to my husband, a careful and critical thinker, who knows more about extensins than he ever wanted to.

We are scientist!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: STUDY RATIONALE

Plant cell walls are a renewable source of fuel for animals (food) as well as machines (biofuel). However, to produce biofuel by converting cell wall carbohydrates into hydrocarbons, the carbohydrates (primarily glucose) must first be released from their macromolecular structures. Glucose in the plant cell wall is mainly sequestered in cellulose microfibrils, which cannot be directly converted into biofuel. However, cellulose microfibrils can be broken down into glucose monomers by digestion of biomass with cellulase enzymes. The released glucose can then be fermented into biofuel molecules such as ethanol or butanol by microorganisms[1,2]. The plant cell wall contains other components in addition to cellulose (hemicelluloses, pectins, lignins, and proteins), all of which interact to form a complex matrix that physically impedes access of cellulase enzymes to cellulose: in other words, cell walls are recalcitrant to digestion for glucose release[3]. In order to overcome this recalcitrance for biofuel production, so that cellulases work more efficiently and the yield of glucose for fermentation increases, the cell wall matrix can be partially disrupted by pretreatment before cellulases are added to the biomass. The goals of pretreatment are to remove lignin and hemicellulose, reduce cellulose crystallinity, and increase cell wall porosity, with minimal carbohydrate degradation, minimal formation of inhibitory by-products, and minimal cost[4]. Protocols for thermochemical and mechanical pretreatment of biomass have been developed to specifically address removal or modification of hemicelluloses and lignins. After pretreatment, glucose yields from cellulase digestion increase because the cellulases have improved access to cellulose[1,5]. However, even with pretreatment, cell wall digestion is still not efficient enough to make plant cell walls a cheap source of biofuel[3].

The roles of hemicellulose and of lignin in biomass recalcitrance have been extensively studied. Both are known to associate with cellulose microfibrils, and their removal in pretreatment is directly correlated with increased glucose yield after cellulase treatment[6,7]. Plant primary cell walls can also contain up to 20% protein. The most abundant cell wall proteins are the hydroxyproline-rich glycoproteins (HRGPs), which comprise a class of proteins with a spectrum of hydroxyproline content and glycosylation. Extensins are the most abundant HRGPs, with high hydroxyproline content and intermediate glycosylation[8]. Extensins are not enzymes; rather, they contribute to cell wall structure through the formation of covalent bonds within cell walls that form a cross-linked matrix[9–11]. Due to these cross-links, extensins have the potential to contribute significantly to biomass recalcitrance. My research examines the influence of extensin cross-linking on biomass recalcitrance, and I explored methods for reducing or removing cell wall extensins to potentially improve glucose yields.

REVIEW OF EXTENSIN CROSS-LINKING

The extensin post-translational modification pathway

Significant post-translational modifications are required to create mature extensins. Extensins are characterized by an abundance of serine-proline (SP_n) motifs, where n is usually between 3 and 7. These SP_n motifs are separated by short hydrophobic motifs of varying length, often including tyrosine, lysine, histidine, and valine[12]. Post-translational modifications stabilize the protein in an outstretched rather than globular conformation[13,14]. First, proline is hydroxylated to hydroxyproline in the endoplasmic reticulum, by prolyl-4-hydroxlases. Hydroxylation of proline in the HRGPs depends on the amino acid context, but most prolines found within SP_n blocks are hydroxylated[12]. Sugar "decorations", primarily chains of 3-5 arabinoses, are then added in the Golgi by glycosyltransferases. Hydroxyproline is the main site

for glycosylation with O-linked arabinose residues, but serine also is O-galactosylated[12]. Finally, fully modified extensins are secreted to the cell wall and insolubilized through the formation of cross-links by peroxidases.

Extensin cross-linking partners

Several theories for how insolubilization of extensins by crosslinking occurs have been proposed, but direct *in vivo* evidence has proved difficult to obtain. Extensin-extensin cross-links, formed inter- or intra-molecularly between tyrosines by extensin peroxidases, have been clearly shown *in vitro*, resulting in isodityrosine[15] or even di-isodityrosine[16]. *In vivo* investigations using REDOR NMR show that 25% of cell wall tyrosines in soybean become incorporated in predominantly isodityrosine cross-links between distant, likely intermolecular tyrosines (contrary to the prevailing model of isodityrosine formation between neighboring tyrosines in YXY motifs)[17]. Extensin-pectin cross-links have been shown indirectly *in vivo* by analyzing the compounds remaining in cotton cell walls after sequential extraction procedures[18], and by analyzing protein contaminants of commercially prepared sugar beet pectin[19]. Extensin-lignin cross-links between tyrosine and the aromatic residues of lignin have been proposed, but an interaction between these molecules is unlikely, as extensins are presumed to remain in primary cell walls, while lignin is found in secondary cell walls. Nevertheless, the high hydroxyproline amino acid composition of the protein component in elicitor-induced lignin formation in spruce pine has been proposed to be due to extensin-lignin complexes[20]. Furthermore, proteins in lignified spruce wood have been found that are very likely extensins, based on their antibody labeling, gel electrophoresis, and amino acid composition profiles[21]. It is possible that cross-links can occur between extensins and multiple other cell wall components. In suspension-cultured soybean cells, an extracellular complex was determined to contain pectin,

hydroxyproline-rich protein, and lignin, with strong covalent bonds between the different kinds of molecules[22]. However it occurs, extensin insolubilization can play a critical role in cell walls, as loss-of-function mutants in Arabidopsis show cell wall phenotypes ranging from abnormal root hairs to embryo lethality due to defects in cell plate formation[23,24].

Importance of cross-linking for plant function

Extensins have been implicated in many aspects of plant growth and development[25–28]. Their outstretched conformation[14], rapid insolubilization after wounding or infection[29,30], and high expression after tensile stress or weight-loading[31,32], suggest that their primary function is mechanical stabilization of the cell wall. Extensin cross-linking appears to be regulated in a cell-type-specific manner. For example, extensins in Arabidopsis are most abundantly expressed in roots and pollen[33]. In transgenic Arabidopsis overexpressing secreted GFP-tagged extensin motifs, levels of protein translation, post-translational modification, and secretion vary depending on the cell type examined[34]. These results suggest that extensins have diverse functions throughout the plant, which can be influenced by post-translational modification that in turn influences cross-linking. Multiple mechanisms may be required to explain extensin cross-linking in different cell wall environments. In summary, extensins, although simple to identify by the repetition of the SP_n motif, are difficult to analyze given their diverse expression patterns and extreme insolubility in cell walls.

Methods of solubilizing cross-linked extensin and their relevance to pretreatments

It is not currently known how pretreatment regimes affect the association of extensins with the cell wall. It is clear that extensins are difficult to release from the cell wall, since extreme treatments, such as with anhydrous hydrogen fluoride, de-glycosylate but do not solubilize extensins[35]. Extensins are also not solubilized by treatment with strong acid[36] or

base[37]. In liquid hot water pretreatments acidity is not extreme enough to cause protein hydrolysis[6]. There is, therefore, a real possibility that current pretreatment regimes do not release extensins from plant cell walls. Modification of pretreatment protocols to solubilize extensin in cell walls would likely increase glucose yields from digestion of biomass. In support of this idea, a patent application has shown that removal of extensins from woody biomass with proteases enhances biomass delignification for the production of paper[38]. Thus, removal of extensins before or during pretreatment may augment delignification; delignification is known to reduce recalcitrance[39], so increased glucose yields may result.

OVERVIEW OF EXPERIMENTAL APPROACHES

Extensin removal by protein digestion

One avenue for exploring the possible effects of extensin cross-linking on biomass recalcitrance is to decrease extensin content prior to cellulase digestion. Although extensins are thought to be protected from protease activity by their arabinose chains, which may shield the polypeptide backbone[13,40], pretreatment may be sufficient to deglycosylate extensins and make them susceptible to proteolytic cleavage. Extensin cleavage may subsequently lead to solubilization of extensin fragments, decreasing biomass reacalcitrance due to physical barriers to cellulase digestion. For example, overexpression of a tyrosine-rich extensin in woody biomass increases sugar release after protease treatment[41]. In Chapter 2, I investigate the effects of protease treatment prior to cellulose digestion on poplar biomass digestibility.

Prevention of cross-linking by mutation of genes encoding key extensin modification enzymes

A second approach to decreasing extensin content is to prevent or reduce extensin cross-linking in cell walls. A suite of Arabidopsis mutants has now been identified and partially characterized, encompassing the gene families encoding the enzymes that perform extensin post-

translational modifications. According to the accepted models of extensin insolubilization, proper post-translational modification is required for both secretion and cross-linking[23]. Therefore, mutations in the genes encoding these enzymes could affect post-translational modification of all extensins and produce cell walls that have reduced or absent extensin cross-linking and reduced recalcitrance. In Chapter 3, I analyze recalcitrance and other phenotypes of five different Arabidopsis T-DNA insertional mutants, one in each of the gene families involved in extensin post-translational modification.

Extensin overexpression in transgenic Arabidopsis

A third avenue for exploring the possible effects of extensin cross-linking on biomass recalcitrance is to increase extensin content prior to cellulase digestion. If my hypothesis is correct, increasing extensin content will increase biomass recalcitrance. In Chapter 4, I describe transgenic Arabidopsis lines transformed with tdTomato-extensin fusion constructs I designed to encode the red fluorescent reporter protein tdTomato fused with either of two different extensin domains. I qualitatively assess the strength of the association of each fluorescent reporter fusion protein with the cell wall, and I analyze recalcitrance and other phenotypes of lines independently transformed with each of the constructs.

CHAPTER TWO: INVESTIGATING THE ROLE OF EXTENSIN PROTEINS IN BIOMASS RECALCITRANCE IN POPLAR[*]

SYNOPSIS

Biological conversion of cellulosic biomass to biofuel requires a pretreatment step to maximize the monosaccharides released. Pretreatment techniques can still be improved to further increase the yield of monosaccharides. In this study we investigated extensin proteins in poplar stem biomass to determine whether their presence influences pretreatment effectiveness. We found three classical extensin genes in poplar through bioinformatic analysis of the poplar genome, which we propose to name as follows: *PtEXT1* (Potri.001G019700), *PtEXT2* (Potri.001G020100), and *PtEXT3* (Potri.018G050100). We used tissue print immunoblots to localize extensin proteins in poplar stems, finding most of the signal to be in the vascular tissue. We assessed whether common pretreatments are able to remove extensins from poplar biomass by measuring the hydroxyproline content (a proxy for extensins) in biomass after different pretreatments (liquid hot water, dilute acid, and alkaline peroxide). Liquid hot water pretreatment reduced hydroxyproline content by 20%, while over half the original hydroxyproline remained after alkaline peroxide and dilute acid pretreatments. Treatment with Fermgen protease was able to further reduce the hydroxyproline content in liquid hot water pretreated biomass by 16%. Finally, we analyzed the effect of Fermgen treatment on glucose release from pretreated poplar biomass, and found that Fermgen treatment increased the subsequent glucose yield from liquid hot water pretreated poplar by 4%. These data suggest that

---

[*] Authors: Margaret Fleming (Biology Department, Colorado State University), Steve Decker (National Renewable Energy Laboratory, Golden, CO), and Patricia Bedinger (Biology Department, Colorado State University)

structural cell wall proteins limit the glucose yield from pretreated biomass, and that one approach to increase yields may be to incorporate a protease treatment into the pretreatment protocol.

INTRODUCTION

Plant cell walls are complex structures composed of polysaccharides (cellulose, hemicelluloses, and pectin), lignin, and structural proteins[40]. Cellulose, as a polymer of glucose, is of particular interest as a source of renewable energy for biofuel production, since cellulose can be broken down to glucose by cellulases, and microorganisms can then ferment the glucose into fuel molecules[2]. However, cell wall recalcitrance, resulting from the natural crystalline structure of cellulose as well as other aspects of cell wall architecture, can limit the ability of cellulases to access and break down cellulose[42].

One way to address cell wall recalcitrance is to employ a pretreatment of biomass before attempting enzymatic digestion of polysaccharides. The commonly used pretreatments are thermochemical, which disrupt the cell wall structure both physically and chemically. Typical pretreatments include dilute acid, alkaline peroxide, liquid hot water, ammonia steam explosion, ammonia fiber expansion, and ionic liquid[43]. Pretreatments target lignin and hemicelluloses, since they are abundant components of the cell wall and are known contributors to recalcitrance[4]. Pretreatment can increase glucose yield, but yields must be further increased if biofuels are to be cost-competitive with conventional fuels[44]. Improving pretreatments to enhance digestive enzyme access to cellulose is one way to increase glucose yields. Structural cell wall proteins are a minor component of cell walls, but could impact cell wall digestibility. The effects of structural cell wall proteins on biomass digestibility have not yet been investigated, and structural cell wall

proteins could prove a fruitful target for reducing cell wall recalcitrance and increasing glucose yields from biomass.

Extensins are structural cell wall proteins, forming a sub-group of the hydroxyproline-rich glycoproteins (HRGPs). Extensins contain a diagnostic SPPP or SPPPP primary sequence motif, repeated at least twice[45]. The primary sequence directs substantial post-translational modifications, including hydroxylation of select prolines to hydroxyproline[46,47], the addition of an arabinosyl side-chain (containing 3-5 arabinoses) to hydroxyproline[48,49], and addition of galactose to serine[50]. The mature protein has a rigid and outstretched conformation, due to the series of proline and hydroxyproline residues creating a polyproline-2 helix, and glycosylation creating a stabilizing β-L-arabinofuranoside "sheath" around the polypeptide chain, protecting it from degradation[14,51]. Newly synthesized, fully modified extensins are secreted to the cell wall and are initially soluble and salt-elutable, but eventually become cross-linked within the cell wall due to the formation of inter-protein covalent bonds[9,15,37]. Generally, extensin cross-linking occurs as the primary cell wall matures and no longer expands, or in response to pathogen attack or mechanical wounding[31,45,52–56].

Once extensins are cross-linked within cell walls, they are extremely difficult to remove[35–37]. Therefore, extensins may pose a significant barrier to complete enzymatic breakdown of cellulose, either by limiting the complete removal of lignin, hemicelluloses, or pectin during pretreatment, or by physically blocking the activity of the processive enzymes used in cellulose degradation.

In this study we investigated extensin proteins in poplar stem biomass. We found three classical extensin genes in poplar through bioinformatic analysis of the poplar genome, and localized extensin proteins in poplar stems using tissue print immunoblots. We assessed whether

common pretreatments are able to remove extensins from poplar biomass by measuring the hydroxyproline content in biomass after different pretreatments. Finally, we analyzed the effect of protease treatment on hydroxyproline content and glucose release from pretreated poplar biomass.

MATERIALS AND METHODS

Bioinformatic analysis of poplar extensins

The extensins identified by Guo et al (2014), originally identified as containing two or more repeats of SP_3 or SP_4 as well as a signature extensin protein domain in Interpro (IPR006706, IPR006041, IPR003882, IPR003883, PR01217, or PTHR23201), were analyzed to find genes encoding "classical" extensins[57]. SignalP 4.1 (www.cbs.dtu.dk/services/SignalP/) was used with its default settings to predict the presence of a signal sequence[58]. In genes that had predicted signal sequences, protein sequence homology to annotated extensin proteins was identified with HMMER3, using a phmmer search of the SwissProt database (hmmer.janelia.org/)[59]. For the purposes of this analysis, genes with additional domains (chimeric extensins) or pollen Ole e 1 allergen/extensin domains were excluded, as well as those with similarity to other HRGPs but not extensins (such as AGPs or PRPs). Genes with a signal sequence and a predicted extensin domain, but no additional functional domains, were classified as "classical extensins".

Introns were identified in classical extensin genes using PopGenie 3.0 (popgenie.org) [60]. The organ with the highest expression for each gene was identified using the PopGenie 3.0 exPlot tool, based on expression values in the *Populus balsamifera* developmental tissue series[61] and the *P. trichocarpa* tissues series[62]. Data from Hefer et al (2015) were also considered[63].

Tissue print immunoblots

Tissue prints were made as described in Cassab and Varner (1989)[64]. Pieces of nitrocellulose paper were soaked for 30 minutes in 0.2 M $CaCl_2$ and dried. Freehand cross-sections of stems of *Populus alba x Populus tremuloides* grown in hydroponic media, between 1-3 mm thick, were made with a double-edged razor blade. The sections were rinsed in distilled H_2O for 3 seconds, dried with a KimWipe, and firmly pressed for 20 seconds onto the prepared nitrocellulose.

The nitrocellulose was air-dried, blocked in 5% non-fat milk in TBST (10 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.05% Tween-20) overnight at 4 ^{o}C and washed five times for 10 minutes in TBST. Prints were treated for 3 hours at room temperature with one of three different antibodies diluted in TBST according to the recommendations of the provider: a 1:10 dilution of mAb JIM20 (raised against pea guard cell protoplasts[65], CarboSource), a 1:10 dilution of mAb LM1 (raised against rice cell walls[66], PlantProbes), or a 1:500 dilution of polyclonal carrot extensin-1 antibody, gE-1[14] (a kind gift of Dr. L. A. Staehelin). Tissue prints were washed four times for 15 minutes in TBST at room temperature, then treated overnight at 4 ^{o}C with a 1:10,000 dilution of goat anti-rat alkaline-phosphatase conjugated secondary antibody (Sigma) for detecting mAbs JIM20 and LM1, or goat anti-rabbit alkaline-phosphatase conjugated secondary antibody (Sigma), for detecting gE-1, each diluted in TBST. Tissue prints were then washed five times for 10 minutes in TBST at room temperature and developed in 10 mL alkaline phosphatase buffer (100 mM Tris-HCl, pH9.5; 100 mM NaCl; 3 mM $MgCl_2$) containing 40 µL NBT (35 mg/mL in 70% DMSO; Sigma N6876) and 15 µL BCIP (50 mg/mL in 100% DMSO; Sigma 8503). As controls, several tissue prints were treated with secondary antibody only or with the alkaline phosphatase substrates directly after blocking; no signal was detected in either

control treatment. The developed tissue prints were imaged with an Epson Perfection V700 Photo flat-bed scanner at 12,800 dpi.

The same poplar stem cross-sections that were used for the tissue prints were subsequently dipped for 5 seconds in 0.25% Toluidine Blue O (Electron Microscopy Sciences, #22050), mounted in glycerol, and imaged with a Leica 5500 microscope, using a Leica DFC 450 color camera with Leica Application Suite V4.1 software. Images were composited with Image Composite Editor (http://research.microsoft.com/en-us/um/redmond/groups/ivm/ice/).

Pretreatment of poplar biomass

Dried, milled poplar biomass (as described in Selig, *et al*. 2010[67]) was pretreated with dilute acid or liquid hot water at the National Renewable Energy Laboratory (NREL, Golden, CO) in a 4-L ZC® vertically stirred reactor (ZipperClave®- Autoclave Engineers, Erie, PA, USA). Steam was directly injected into the bottom of the reactor through ports in a rotary-plow type of agitator and constant temperature was achieved by controlling the steam pressure in the reactor. The ZC reactor is also equipped with an electrical heating blanket set at reaction temperature to lessen steam condensation due to heat losses through the reactor wall. The contents within the ZC reactor typically reached reaction temperature within 5 to 10 seconds of starting the steam flow as measured by two thermocouples, one inserted into the bottom and one near the middle of the reactor. At the end of pretreatment, the steam pressure was slowly released through a condenser over a period of 15 to 30 seconds to lessen boil-over. For acid pretreatment, poplar was mixed with 0.5% sulfuric acid, impregnated into the biomass under vacuum with mixing, loaded into the ZipperClave® pretreatment reactor and treated at 150 ^{o}C for 20 minutes. For liquid hot water pretreatment, poplar was mixed with water, loaded into the ZipperClave® and treated at 180 ^{o}C for 40 minutes. For alkaline peroxide pretreatment, 4 g of untreated

standard BioEnergy Science Center (BESC) poplar was mixed with 100 mL of 1% hydrogen peroxide solution (pH 11.5, adjusted with 12 M NaOH), and incubated with vigorous shaking at 65 ºC for 3 hours. After all pretreatments, the biomass was extensively washed with deionized water and dried before use in further experiments. Untreated, standard BESC poplar was used as a control throughout the subsequent experiments[67].

Compositional analysis of biomass

Each type of biomass (no pretreatment, liquid hot water pretreatment, dilute acid pretreatment, or alkaline peroxide pretreatment) was analyzed using a scaled-down version of the NREL standard Laboratory Analytical Procedures for compositional analysis for structural carbohydrates, lignin, protein and ash (http://www.nrel.gov/biomass/pdfs/42618.pdf). Briefly, 100 mg of biomass was hydrolyzed in 72% (w/w) H_2SO_4 for 2 h at 30 ºC. After dilution with dH_2O to 4% (w/w) H_2SO_4, the samples were autoclaved for 60 minutes at 121 ºC to hydrolyze oligosaccharides formed in the first stage. After neutralization with Ca_2CO_3, monomeric sugars were quantified by HPLC. Structural polysaccharide content was back-calculated after adjustment for losses due to degradation against sugar recovery standards. Each sample was analyzed in duplicate using 100 mg of biomass per sample.

Hydroxyproline assay

15 mg samples of untreated, liquid hot water pretreated, dilute acid pretreated, and alkaline peroxide pretreated poplar were rehydrated in 2.1 mL 30 mM sodium citrate, pH 4.5, for 3 hours at room temperature in 2-mL screwtop tubes (Sarstedt). 4.5 µL Fermgen (Genencor), or a buffer-only control, was added to the rehydrated biomass, and the samples were incubated in a 50 ºC rotisserie oven with end-over-end rotation for 24 h. Multiple samples for each treatment and biomass type were pooled to achieve a final mass of ~100 mg. These pooled samples were

boiled overnight in distilled water, washed three times with distilled water, and dried overnight at 50 ^{o}C to remove any soluble proteins (particularly monomeric, non-covalently-bound extensin or other soluble HRGPs). Samples were re-weighed and amino acid hydrolysis was performed by adding 1.6 mL 6 M HCl and incubating in closed Sarstedt tubes at 100 ^{o}C for 18 h. The supernatant was assayed for hydroxyproline according to a modification of the methods of Kivrikko and Liesmaa[68] as follows: 500 µL were transferred to a new tube and pH-adjusted by the addition of 12 M NaOH to pH 3.0 (+/- 0.1). A higher pH led to development of a dark brown color in the samples that interfered with subsequent colorimetric analysis. The samples were centrifuged at 21,130 g for 1 minute to clarify the supernatant. 125 µL of sample supernatant was then mixed with 250 µL 50 mM sodium hypobromide and incubated at room temperature for 5 minutes. The oxidation reaction was stopped by the addition of 125 µL 6 M HCl, followed by 250 µL para-dimethylaminobenzaldehyde (DMAB, 5% in n-propanol), to all tubes. Tubes were sealed, mixed by hand, and incubated at 70 ^{o}C for 15 minutes. After cooling, the absorbance at 560 nm of each sample was measured in triplicate (200 µL each) in a 96-well plate, using a BioTek Synergy HT plate reader. Data were analyzed in Microsoft Excel 2010. Hydroxyproline standards with the same salt concentrations (4.2 M NaCl) and pH (3) as the samples were used to construct a standard curve.

Protease treatments

Fermgen

5 mg dried poplar biomass samples of four types (untreated, liquid hot water pretreated, dilute acid pretreated, and alkaline peroxide pretreated) were rehydrated in 700 µL 30 mM sodium citrate, pH 4.5, for 3 hours at room temperature in 2-mL screwtop tubes (Sarstedt). 1.5 µL Fermgen (70 mg/mL, determined with BCA assay, Pierce) (Genencor), or a buffer-only

control, was added to the rehydrated biomass, and the samples were incubated in a 50 °C rotisserie oven with end-over-end rotation for 24 h. The treatment was stopped by boiling the tubes in a water bath for 10 minutes. The biomass was then centrifuged at 21,300 g for 30 seconds and washed to remove any remaining protease activity, following Kumar and Wyman (2009)[69], using two 500 µL washes with 1 M NaCl followed by four 500 µL washes with deionized water, and finally adding 500 µL 30 mM sodium citrate buffer, pH 4.5 (final concentration 21.4 mM). The biomass was vigorously vortexed and then centrifuged for each wash, and only 500 µL of the total 700 µL was removed after each centrifugation to avoid disturbing the biomass pellet.

For a time-course experiment, 5 mg dried poplar biomass samples of all four types were rehydrated as described above. 1.5 µL Fermgen was added to the rehydrated biomass every two hours for ten hours (9 µL total) while the samples were incubated in a 50 °C rotisserie oven with end-over-end rotation for 24 h. The treatment was stopped as described above before proceeding with a digestibility assay.

Chymotrypsin

5 mg dried, liquid hot water pretreated poplar biomass samples were rehydrated in 630 µL 10 mM Tris pH 8.0, 50 mM NaCl, 10 mM $CaCl_2$ for 3 hours at room temperature. 70 µL of 1 mg/mL chymotrypsin (Sigma C-7762, dissolved in the same buffer) or buffer was added. Samples were incubated, boiled, washed and resuspended as described above.

Digestibility assay

We adapted the protocols of Santoro et al (2010) and Selig et al (2010) from high-throughput to laboratory-scale[70,67]. 50 µL of a solution containing Cellic CTEC2 (70 mg/g biomass, Novozymes), Cellic HTEC2 (2.5 mg/g biomass, Novozymes), 0.007% NaN_3 (final

concentration), and 30 mM sodium citrate, pH 4.5, was added to the samples of 5 mg biomass in 700 µL 21.4 M sodium citrate, pH 4.5 (prepared as described above). Samples were incubated in a 50°C rotisserie oven with end-over-end rotation for 1 week.

The tubes were centrifuged at 21,130 g for 1 minute. 650 µL supernatant was removed to a new tube, diluted 1:10 in water, and measured for glucose content in a GOPOD assay (Megazyme). Triplicate 20 µL aliquots of sample, glucose standard, or an enzyme-only mixture were mixed with 200 µL GOPOD reagent and incubated at 40 °C for 45 minutes in a 96-well plate. The absorbance at 510 nm was read in a BioTek Synergy HT plate reader. Data were analyzed in Microsoft Excel 2010. Glucose standards were used to construct a standard curve. The absorbance from the enzyme-only mixture was subtracted from the absorbance of the samples to correct for glucose present in Cellic CTEC2 and HTEC2.

24-hour digestibility assay

50 µL of a solution containing Accellerase 1500 (13 mg/g biomass, Dupont), 0.007% NaN3 (final concentration), and 30 mM sodium citrate, pH 4.5, was added to samples of 5 mg liquid hot water pretreated biomass in 700 µL 21.4 M sodium citrate, pH 4.5 (prepared as described above)[70]. Samples were incubated in a 50 °C rotisserie oven with end-over-end rotation for 24 hours. Glucose analysis was performed as described above. An enzyme-only (no biomass) sample was measured to correct for glucose present in Accellerase 1500.

Protein gel assay to detect proteolysis

1.5 µL of Fermgen was added to four tubes containing 20 mg Avicel (Sigma) in 700 µL 30 mM sodium citrate, pH 4.5. Two of the samples were boiled and washed to inactivate and remove Fermgen as described above. 200 µL of 2 mg/mL BSA, or 1 µL Cellic CTEC2 (237 mg/mL, determined with BCA assay, Pierce) plus 1 µL Cellic HTEC2 (206 mg/mL, determined

with BCA assay) were added to one tube with active Fermgen and one tube with Fermgen inactivated and removed by boiling and washing. All samples were incubated at 50 ^{o}C for 24 hours, and then 75 µL were removed and mixed with 25 µL 4x SDS sample buffer (200 mM Tris-HCl pH 6.8, 8% SDS, 40% glycerol, 4% β-mercaptoethanol, 50 mM EDTA, 0.08 % Bromophenol Blue), and boiled 5 minutes in a water bath. Samples were then electrophoresed on a 10% SDS-PAGE gel. The gel was then stained with Coomassie Brilliant Blue and destained with a solution of 30% methanol, 10% acetic acid.

To test Fermgen auto-proteolysis, 1.5 µL Fermgen was mixed with 700 µL 30 mM sodium citrate, pH 4.5. 20 µL were immediately removed and flash-frozen in liquid N_2. The sample was placed in a 50 ^{o}C incubator overnight, and 20 µL samples were removed and flash-frozen every hour for five hours as well as after 24 hours. Each sample was mixed with 6.5 µL 4x SDS sample buffer and boiled 5 minutes in a water bath. Samples were then loaded on a 10% SDS-PAGE gel and electrophoresed for 40 minutes at 200 V. The gel was then stained with Coomassie Brilliant Blue and destained with a solution of 30% methanol, 10% acetic acid.

RESULTS

Three classical extensin genes in poplar

Previous work by Guo et al (2014) identified 37 genes in the *P. trichocarpa* genome, v. 3.0, with the signature SPPP or SPPP motifs of extensins, as well as extensin domains classified in InterPro, including IPR006706 (extension domain), IPR006041 (pollen Ole e 1 allergen/extensin), IPR003882 (pistil-specific extensin-like protein, an extensin chimera), IPR003883 (repetitive proline-rich cell wall protein repeat), PR01217 (proline-rich extensin), and PTHR23201 (extensin, proline-rich protein). We wanted to identify which of these 37 genes encode "classical" extensins using further diagnostic criteria. First, the gene must encode a signal

peptide for the protein to be secreted into the cell wall, in order to fulfill its role as a structural cell wall protein. Of the 37 genes, 22 encode signal sequences, according to SignalP analysis. The second criterion we considered was the protein's homology to annotated extensins. Of the 22 genes with signal sequences, those encoding proteins with homology to an annotated extensin protein were found by searching the SwissProt database with HMMER3. Thirteen genes had both signal sequences and significant homology to annotated extensins, where the sequence returned with greatest homology to the queried gene was an extensin (E-value < $3x10^{-9}$). Finally, classical extensins consist almost entirely of a single domain, comprised of repetitions of SP_{3-6} and YXY. Therefore, the final criterion was that the protein should not belong to other subgroups of HRGPs, such as arabinogalactan proteins, or contain any other domains indicative of extensin chimera proteins, such as LRX proteins. We also looked at the gene structure, since only three of the 20 classical extensins in Arabidopsis have introns, although this is not a hallmark of a classical extensin. Three poplar genes, Potri.001G019700, Potri.001G020100, and Potri.018G050100, met all of our criteria (Figure 1, Table 1). We propose the following names for these genes: *PtEXT1* (Potri.001G019700), *PtEXT2* (Potri.001G020100) and *PtEXT3* (Potri.018G050100). *PtEXT3* had the highest similarity to carrot (*Daucus carota*) extensin[71,72] based on the results of the HMMER3 search, while the other two genes had the highest similarity to Arabidopsis Extensin-2[73] (Table 1). *PtEXT1* and *PtEXT3* lack introns, while *PtEXT2* has a single intron.

We queried gene expression of the three poplar extensin genes using publicly available poplar microarray data. The organ or tissue identified as having the highest expression of each gene depends to some extent on which microarray experiment is analyzed. According to the *Populus balsamifera* developmental tissue dataset, all three genes are most highly expressed in

young leaves[61], while according to the *P. trichocarpa* tissues dataset, *PtEXT1* is most highly expressed in mature leaves and roots, *PtEXT2* is most highly expressed in roots and young leaves, and *PtEXT3* is most highly expressed in young and mature leaves[62]. Based on recent work of Hefer et al (2015), at least one poplar extensin gene (*PtEXT1*) is expressed in young woody tissue [63] (Table 1).

Localization of extensin proteins in poplar stems

In order to investigate more directly whether extensin proteins are present in poplar wood, we performed tissue prints on poplar stems to detect extensin proteins with anti-extensin antibodies (Figure 2, Figure 3).

JIM20 is an anti-extensin monoclonal antibody raised against pea guard cell protoplasts, shown to be specific for extensins[65]. Using our poplar stem tissue prints, extensins were detected primarily in secondary phloem with mAb JIM20, with some signal also present in secondary xylem and pith (Figure 2A, Figure 3A). LM1 is an anti-extensin monoclonal antibody raised to rice cell wall material, shown to be specific for extensins[66]. LM1 produced a similar pattern of extensin localization as JIM20 in poplar stems, with extensins primarily detected in secondary phloem but with some signal also in secondary xylem (Figure 2B, Figure 3B). The anti-carrot extensin antibody (gE-1) is a polyclonal antibody raised against purified glycosylated carrot extensin-1, shown to be specific for extensins[74]. The pattern of extensin localization detected with gE-1 was somewhat different from that of JIM20 and LM1, in that extensins were seen most prominently in secondary xylem, with very low signal (if any) in secondary phloem (Figure 2C, Figure 3C). All three antibodies also produced a detectable signal in the perimedullary zone at the border of the secondary xylem and the pith, which is involved in both stress and defense responses[75,76].

Effect of pretreatments and protease treatment on hydroxyproline content in poplar biomass

Having verified that extensins are expressed within woody tissues in poplar, we wished to determine first, whether standard pretreatments remove extensins from poplar stem biomass and second, whether protease treatment removes extensins from pretreated poplar stem biomass. Hydroxyproline content can be used as a proxy for extensin content[11]. We measured 193 μg hydroxyproline per gram dry weight in untreated poplar biomass (Figure 4). In poplar biomass pretreated with liquid hot water, we measured slightly less: 156 μg hydroxyproline per gram dry weight ($p = 0.009$, Student's t-test). However, we measured substantially less in poplar biomass pretreated with alkaline peroxide (126 μg hydroxyproline per gram dry weight, $p = 0.0003$) and in poplar biomass pretreated with dilute sulfuric acid (112 μg hydroxyproline per gram dry weight, $p = 0.0006$). Thus, all pretreatments removed some hydroxyproline from pretreated poplar biomass, with the amount removed depending on the type of pretreatment, and in all cases over half the original amount (measured in the untreated samples) remained. The presence of hydroxyproline after all three pretreatments indicates that none of the pretreatments completely removes extensins from poplar biomass.

In order to test whether protease treatment could remove extensins from poplar biomass, we also measured the hydroxyproline content of the four types of poplar biomass (untreated or pretreated with liquid hot water, alkaline peroxide, or dilute acid) after incubation with Fermgen. Fermgen is an acid protease designed to digest proteins in starchy biomass used for ethanol fermentation. Fermgen treatment significantly decreased hydroxyproline content in liquid hot water pretreated biomass, from 156 to 131 μg hydroxyproline per gram dry weight (a 16% decrease, $p = 0.04$, Figure 4). Fermgen treatment likewise decreased hydroxyproline content in untreated biomass, from 193 to 179 μg hydroxyproline per gram dry weight, but this difference

was not significant (a 7% decrease, $p = 0.10$). Fermgen treatment did not change the hydroxyproline content of either alkaline peroxide or dilute acid pretreated poplar biomass.

Protease treatment can increase glucose release from poplar biomass

We next investigated whether protease treatment of the four types of poplar biomass (untreated or pretreated with liquid hot water, alkaline peroxide, or dilute acid), which should degrade extensins, could increase glucose yield after subsequent treatment of the biomass with cell wall degrading enzymes. A 24-hour treatment with Fermgen caused no significant change in the glucose released from three of the four types of biomass (untreated, alkaline peroxide pretreated, or acid pretreated) after either a 24-hour (data not shown) or seven-day digestion with cell wall degrading enzymes (Figure 5). However, treatment with Fermgen caused a small but significant increase in glucose released from liquid hot water pretreated poplar after either a 24-hour (2% increase, Figure 6) or a seven-day digestion with cell wall degrading enzymes (4% increase, Figure 5). Chymotrypsin was previously reported to degrade extensins in suspension-cultured plant cells[46]. A 24-hour treatment of liquid hot water pretreated poplar biomass with chymotrypsin caused a small but significant increase in the glucose released after a 24-hour digestion with cell wall degrading enzymes (0.9% increase, Figure 6), but there was no significant difference seen after a seven-day digestion with cell wall degrading enzymes (data not shown). Papain and pronase have been reported to degrade extensins in paper pulp and in suspension-cultured plant cells, respectively[38,46]. We found that treatment of liquid hot water pretreated poplar biomass with papain, pronase, or proteinase K did not affect glucose release after a 24-hour digestion with cell wall degrading enzymes (data not shown).

Due to the possibility that Fermgen may be auto-proteolytic, we performed a time-course of Fermgen degradation (Figure 7). Fermgen is quickly degraded in the conditions we used to

treat poplar biomass. We therefore investigated whether using multiple doses of Fermgen through the same 24-hour period would have a greater effect on glucose release than a single dose. We found that poplar biomass treated with multiple doses of Fermgen before digestion with Cellic CTEC2/HTEC2 did not further increase glucose release compared to poplar biomass treated with a single dose of Fermgen (p-values > 0.05, Figure 8).

It has been reported that partial proteolysis of cellulases can stimulate cellulase activity, so we wished to confirm that this was not the cause of increased glucose release in our system, i.e. that boiling and washing biomass samples prior to cellulase treatment of biomass successfully inactivated and depleted Fermgen. Samples containing either BSA or a Cellic CTEC2/HTEC2 mixture were treated at 50 ^{o}C for 24 hours with either active or inactivated Fermgen, then analyzed using SDS-PAGE (Figure 9). We found no evidence of proteolytic degradation of either BSA or CTEC2/HTEC2 in the samples that had been boiled and washed, indicating successful inactivation and depletion of Fermgen. Therefore, the increase in glucose release from liquid hot water pretreated biomass after protease treatment is likely due to degradation of cell wall proteins (such as extensins) rather than to proteolytic activation of cellulases.

Interestingly, our results show that active Fermgen did not digest CTEC2 or HTEC2. However, active Fermgen did digest BSA, confirming it was capable of proteolytic activity. It is possible that a protease inhibitor is added to the CTEC2 and HTEC2 preparations to minimize degradation during storage, which could prevent Fermgen from digesting the enzymes.

DISCUSSION

Previous work had identified 37 extensin genes in poplar[57]. We were interested to see which of the 37 genes could be classified as classical extensins, and we found three strong

candidates. All three genes show characteristics that are shared by the classical extensins in Arabidopsis, in that they 1) encode a signal peptide, 2) consist largely of SP_n repeats in addition to showing a significant homology to annotated extensins in InterPro, and 3) do not belong to other subgroups of HRGPs or contain any other domains indicative of extensin chimeric proteins. They also have a similar gene structure to Arabidopsis classical extensins, which have one or no introns[33]: two poplar genes have no introns (*PtEXT1* and *PtEXT3*) and one has a single intron (*PtEXT2*). Since the amino acid sequences of the three poplar classical extensins have an abundance of the YXY motifs possibly involved in isodityrosine (Idt) formation and are lacking in SPSP and tri-C motifs, we classify them as Group IIb extensins[47,77]. The three proteins are dominated by repetitive motifs. After signal sequence cleavage, the PtEXT1 protein consists of 12 repeats of (S/K/L)(P/K)SPPPPY(Y/H/I/V)Y(K/S)SPPPP, with only a few non-repeat amino acids at the N- and C-termini (Figure 1). Similarly, PtEXT2 consists of 14 repeats of YHY(K/S/T)SPPPPKKSPPPP, again with only a few non-repeat amino acids at the N- and C-termini. These sequences are similar to the tomato P3 extensin repeat sequence, SPPPPSPSPPPPYYYK[78,79], a motif found across the plant kingdom, from ferns to Poaceae[47], and now identified in Populus as well. PtEXT3 has twelve SP_{3-6} motifs that make up 40% of the 167 amino acid protein, including five occurrences of $YXYXSP_{4-6}$. The amino acid sequences of the three proteins also exhibit several features thought to be important in forming a covalently linked extensin matrix in the cell wall: strong periodicity, to allow self-assembly; repetition of the YXY cross-linking motif, to allow Idt formation; and regular placement of positively charged residues (H and K), to allow electrostatic interactions with negatively charged cell wall components, such as pectins (Figure 1)[24,80].

The presence of only three classical extensin genes in poplar is somewhat surprising, since the Arabidopsis genome encodes 22 classical extensins[33]. However, based on the work of Guo et al, the Arabidopsis extensin gene family (containing 65 extensins, including short and chimeric extensins[33]) appears to be anomalous, as multiple duplication events occurred within Arabidopsis after the lineage diverged from poplar (37 extensins), papaya (18 extensins) and grape (5 extensins). There is a clear need for caution in naming genes encoding putative extensin proteins. For example, if the predicted amino acid sequence of a protein contains only two blocks of SP_n, the primary function of the protein may not relate to cell wall structure. Frequently, proteins identified as extensins with low SP_n repeat abundance lack a signal sequence, negating a function within the cell wall[33]. Furthermore, extensin chimeras and hybrids, including Leucine-rich Repeat Extensin proteins (LRXs)[81], Proline-rich Extensin-like Receptor Kinases (PERKs)[82], and extensin-arabinogalactan proteins[83], are often classified as extensin proteins. The functions of these chimeras are less well understood, although it is thought the extensin domain facilitates cell wall localization[84].

Post-translational modifications, specifically proline hydroxylation and glycosylation, are thought to be critical to extensin function[23]. Proline hydroxylation is required for subsequent glycosylation[85], but whether a particular proline is hydroxylated depends on tissue identity, growth conditions, and context within neighboring amino acid sequences[34,86,87]. Generally, however, it is thought that the "rule" for proline hydroxylation is that all prolines within a block of two or more prolines will be hydroxylated[86]. We assessed the potential level of proline hydroxylation in the three poplar extensins. Assuming maximal hydroxylation, PtEXT1 should contain 51% hydroxyproline (90/197 amino acids), PtEXT2 should contain 49% hydroxyproline (101/229 amino acids), and PtEXT3 should contain 35% hydroxyproline (54/167 amino acids)

(expressed as mol %) (Table 1). These values for hydroxyproline composition are comparable to experimentally determined values for other extensins, including runner bean (35.3%)[88], tomato P1 (31.5%) and P2 (41.8%)[89], and Douglas fir PHRGP (28.1%) and P2 (29.2%)[90]

Glycosylation of hydroxyproline has been proposed to follow the "Hyp-contiguity hypothesis": contiguous hydroxyprolines are arabinosylated with 2-5 arabinose residues per hydroxyproline, whereas clusters of discontiguous hydroxyprolines are arabinogalactosylated[86,85]. According to the Hyp-contiguity hypothesis, the poplar extensin proteins should be highly arabinosylated, since the hydroxyprolines consistently occur in contiguous blocks. Although the hydroxylation and arabinosylation "codes" have been in large part elucidated, these post-translational modifications ideally would be confirmed experimentally for each putative extensin. Determining the extent of proline hydroxylation and subsequent glycosylation of the poplar extensins will require further experimentation.

Extensins in woody species are currently not well studied, and in general the methods for detecting the presence of extensins have been indirect. For example, high concentrations of hydroxyproline have been detected in the wood of European beech (*Fagus sylvatica*)[91] and spruce (*Picea abies*)[92], as well as in callus tissue of *Pinus elliottii*[93] and suspension-cultured sycamore (*Acer pseudoplatanus*)[48,94,95] and Douglas fir (*Pseudotsuga menziesii* (*Mirbel*) *Franco*)[90,96] cells. More directly, in one gymnosperm, loblolly pine (*Pinus taeda L.*), an extensin-like protein was detected in differentiating xylem and mature wood through immunohistochemical localization[21].

In this paper we show direct evidence for extensin localization in woody tissues in poplar, a woody angiosperm, by using tissue print immunoblots probed with three different antibodies specific for extensins (Figure 2, Figure 3). One caveat about tissue prints is that they

can only show localization of soluble extensin that can be transferred to the nitrocellulose paper. Extensin that has been cross-linked into the cell wall is not detectable by this method, and presumably the bulk of extensins present in mature, woody tissues will have been cross-linked. Much of the extensin signal that we found was in the vascular tissue, with the most intense signals in the cells nearest the vascular cambium, supporting the model that soluble extensins are secreted in expanding cell walls and then cross-linked as cells mature. Our results also are supported by the recent findings of Hefer et al (2015), whose gene expression analysis showed high expression of *PtEXT1* in developing xylem tissue[63].

Even though the protein content of wood is low (0.03-0.1% by weight)[97], our detection of extensin in woody poplar tissue indicates that extensins could play a role in inhibiting the digestibility of cellulose in wood, since a cross-linked extensin matrix in the cell wall could impede the passage of processive cellulases on cellulose microfibrils. Based on earlier work in which extensins were not released through either strong acid or base digestions[36,37], we hypothesized that extensins would not be removed by the relatively mild pretreatments currently favored by industry. Indeed, we found that much of the initial hydroxyproline (81%) was still detectable in poplar biomass after liquid hot water pretreatment, although a significant amount was removed (Figure 4). More robust pretreatments (alkaline peroxide and dilute acid) were able to remove some hydroxyproline from the biomass, but even so, a substantial percentage (58% for alkaline peroxide, 65% for dilute acid) remained after pretreatment. Since extensins are still present in woody biomass after pretreatment, they could have a negative effect on glucose release by cellulases. This hypothesis is also supported by recent work in rice, which found a negative correlation between hydroxyproline content and glucose release following either liquid hot water or dilute base pretreatments[98].

Protease treatment has been used to remove extensins from cell walls. Lamport (1965) investigated the effectiveness of different proteases for releasing extensins from cell walls of suspension-cultured sycamore cells. He found chymotrypsin, pronase, papain and subtilisin each released up to 30% of the hydroxyproline from the cell walls[46]. We measured a significant reduction in hydroxyproline content after Fermgen treatment in liquid hot water pretreated poplar biomass (Figure 4). Since hydroxyproline content is a proxy for extensin content[11], and since the samples were treated to remove soluble sources of hydroxyproline, this reduction in hydroxyproline content can be attributed to the removal of extensins. It may be that extensins are bound to the cell wall in varying degrees, perhaps due to varying numbers of covalent bonds between different extensin molecules. In this case, alkaline peroxide and dilute acid pretreatments may remove all the weakly-bound extensins, while liquid hot water pretreatment is unable to remove these extensins. Therefore, Fermgen treatment is able to digest the weakly-bound or accessible extensins in liquid hot water pretreated biomass, increasing glucose release, while the extensins that remain in alkaline peroxide or dilute acid pretreated biomass may be too tightly bound or inaccessible to be removed by Fermgen treatment.

It has been proposed that protease removal of extensins from woody biomass is responsible for improved paper pulp or bast fiber production[38,99]. In our studies, we used treatment with Fermgen (an acid fungal aspartate protease), chymotrypsin, papain, pronase, or proteinase K to test whether proteases could increase glycan yields from woody poplar biomass. Of these, only Fermgen treatment led to a significant increase in glycan yields from liquid hot water pretreated poplar biomass, with an increase of 4% (Figure 5). Since extensins are likely to be the most abundant proteins in woody tissue[91], extensin digestion may account for this effect. Neither alkaline peroxide nor dilute acid pretreated poplar showed any change in hydroxyproline

content or glycan release after Fermgen treatment, supporting the hypothesis that extensin digestion is the cause of increased glycan release in liquid hot water pretreated poplar biomass.

Irreversible hornification may account for the low levels of glycan release in these experiments compared to other published digestibility experiments in poplar (30% vs 80% glycan release[67]). When pretreated biomass is dried, removal of water leads to structural changes in the cell wall (hornification) that reduce permeability to small molecules and consequently reduce digestibility [100,101]. Rehydrating the biomass is insufficient to reverse these changes. In this study, since pretreatment was performed at a different facility than the subsequent analyses, drying the biomass was an unavoidable necessity to be able to handle and analyze samples accurately.

Mild biomass pretreatment with liquid hot water has advantages in biofuel production, since it is generally less expensive in terms of reagents, waste cleanup, and equipment, and fewer toxic and inhibitory byproducts are formed. Even an incremental yield increase can be significant given the scale at which industrial biofuel production takes place. Protease treatment with a protease such as Fermgen, which has already been successfully utilized in large-scale ethanol production, or other fungal proteases that specifically degrade extensins[102–104], is promising. Evidence is building that the most effective pretreatments for making cellulosic material more digestible by cell wall degrading enzymes are those that increase the enzymes' access to cellulose[5]. Protease treatment, by dismantling the extensin cross-linked matrix, may be key to improving cellulose accessibility.

> Potri.001G019700 PtEXT1

*MIYALAFCVVATSVVA*KEPYYYKSPPPPLKSSPPPSPSPPPPYHYSSPPPPKKSPPPPYIYKSP
PPPLKSPPPPYHYSSPPPPKKSPPPPYHYSSPPPPKKSPPPPYVYKSPPPPSPSPPPPYHYSSP
PPPKKSPPPPYVYKSPPPPSPSPPPPYHYSSPPPPKKSPPPPYIYKSPPPPSPSPPPPYHYSSP
PPPKK**SPHPP**YVYKSPPPPHY

> Potri.001G020100 PtEXT2

*MENRGRMGHLSPMIHAIAICLVATSVVA*YEPYYYKSPPPPSQSPPPPYHYSSPPPPKKSPPPPY
HYTSPPPPKKSPPPPYHYS**SPPQP**KKSPPPPYHYSSPPPPKKSPPPPYHYSSPPPPKK**SPLPP**Y
HYSSPPPPKKSPPPPYHYSSPPPPKK**SPPPQ**YHYTSPPPPKKSPPPPYHYTSPPPPKKSPPPPY
HYTSPPPPKKSPPPPYHYTSPPPPKKSPPPPYHYSSPPPPKKSPPPPYHYTSPPPPKKIEIVDP
W

> Potri.018G050100 PtEXT3

*MDLLHSVMLYFSLALLLSSSSEA*TDTTFSRNSALWLTYKSPPPPFHNKHKSPPPPPHKYKSPPP
PHHKCKYSPPPPVYTYRSPPPPTPMHHKSPPPSPHMFKSPPPPPYRYISPPPPPPHPPCHAYKY
LSPPPPPSYKYASPPPPKHHHHHHKHKWSPYPFITYMSPPPPHHNYPDYHYSSPPPPPPIVAY

Figure 1. Complete amino acid sequences of the three poplar classical extensin genes. Signal sequence highlighted in yellow, $SP_{3\text{-}6}$ repeats in green, H or K in blue; YXY motifs are underlined. Amino acids in boldface slightly deviate from the expected pattern of SPPPP.

Table 1. Classical extensin genes in *P. trichocarpa*.

#R, Number of times repeated motif occurs in the entire protein sequence. Amino acids, number of amino acids in mature protein after signal peptide cleavage. Max % Hyp, maximum mol percent hydroxyproline of the post-translationally modified protein, determined by assuming that all prolines within a block of 2 or more prolines are hydroxylated. #I, number of introns. Microarray expression data from Wilkins et al, 2009[61] & Yang et al, 2008[62].

Locus ID	Homologous protein (Organism)	E-value (HMMER)	Repeated motif	#R	Amino Acids	$SP_3/SP_4/SP_5/SP_6$ repeats	#YXY/H/K	Max % Hyp	#I	Highest expression (*P. balsamifera*)	Highest expression (*P. trichocarpa*)
Potri.001G019700 (PtEXT1)	Extensin-2 (*A. thaliana*)	$1.10x10^{-70}$	(S/K/L)(P/K)SPPPPY(Y/H/I/V)Y(K/S)SPPPP	12	197	0/22/0/0	12/8/21	51	0	Young leaves	Mature leaves and Roots
Potri.001G020100 (PtEXT2)	Extensin-2 (*A. thaliana*)	$6.70x10^{-75}$	YHY(K/S/T)SPPPPKKSPPPP	14	229	1/24/0/0	14/13/27	49	1	Young leaves	Young leaves and Roots
Potri.018G050100 (PtEXT3)	Extensin (*D. carota*)	$2.70x10^{-9}$	$YXYXSP_{4-6}$	5	167	1/6/3/2	5/19/14	35	0	Young leaves	Young leaves and Mature leaves

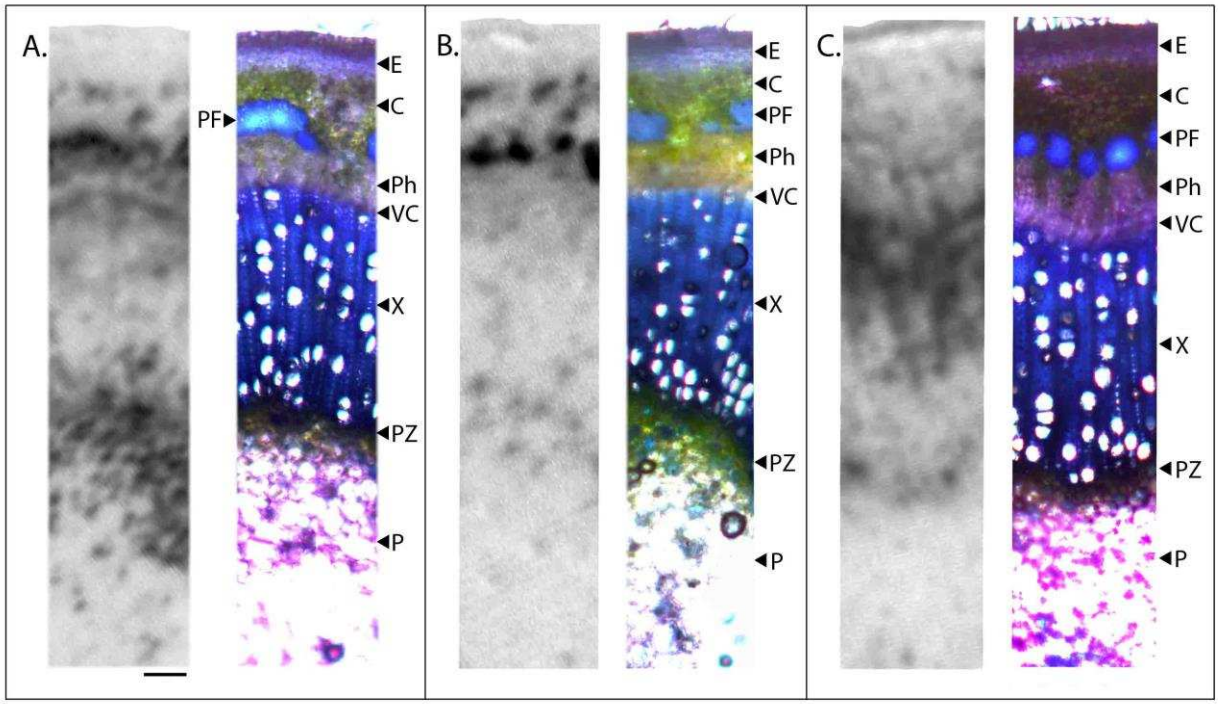

Figure 2. Tissue print immunoblots showing extensin localization in poplar stems using three different antibodies.
A. Print probed with JIM20 (left) and its corresponding stained section (right). B. Print probed with LM1 (left) and its corresponding section (right). C. Print probed with gE-1 (left) and its corresponding section (right). P, pith; PZ, perimedullary zone; X, xylem; VC, vascular cambium; Ph, phloem; PF, phloem fiber; C, cortex; E, epidermis. Bar = 100 μm.

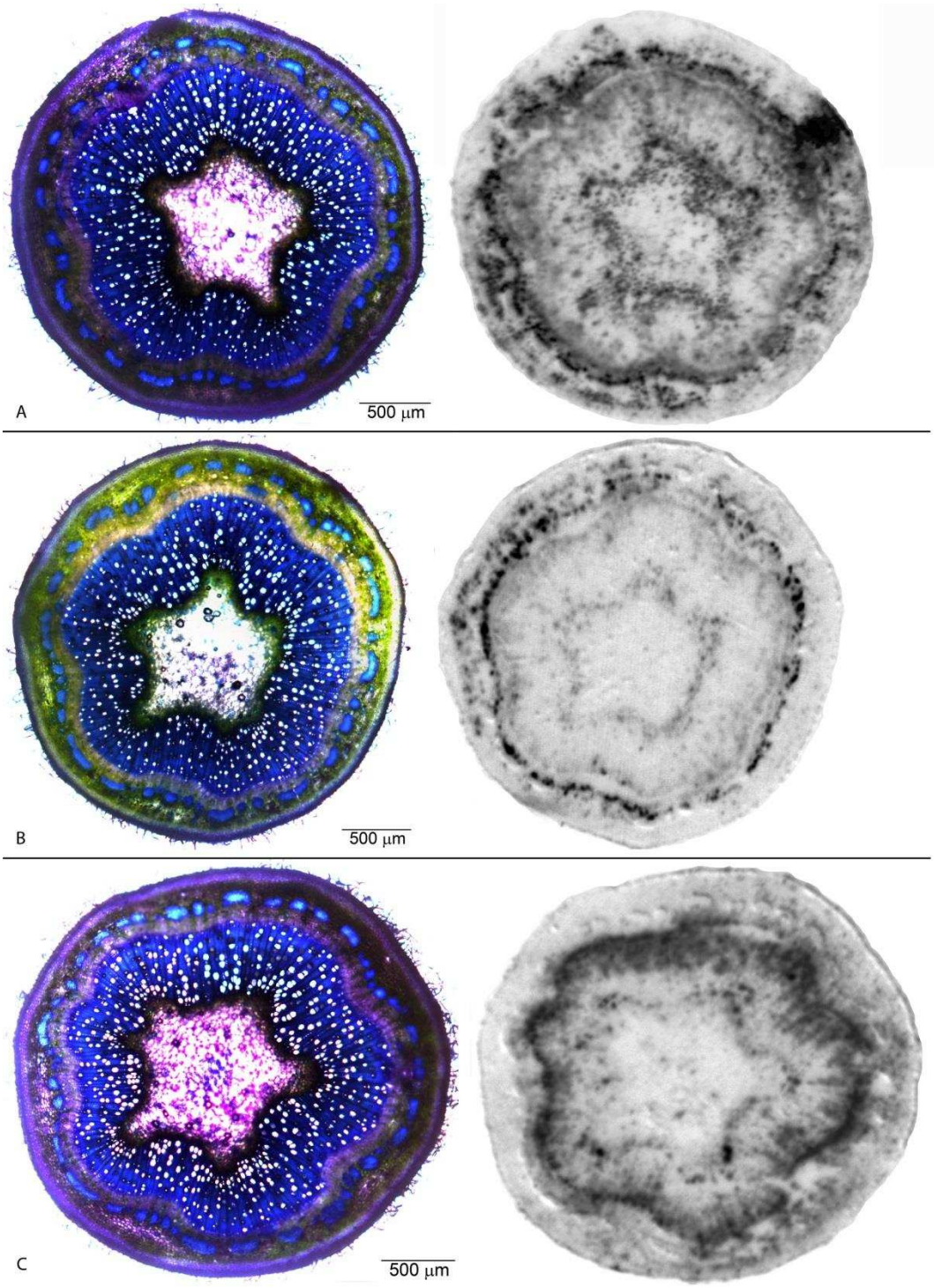

Figure 3. Full images from which Figure 2 was made.
Poplar stem sections (left) and the corresponding tissue print immunoblots (right). A, Print probed with JIM20. B, Print probed with LM1. C, Print probed with gE-1.

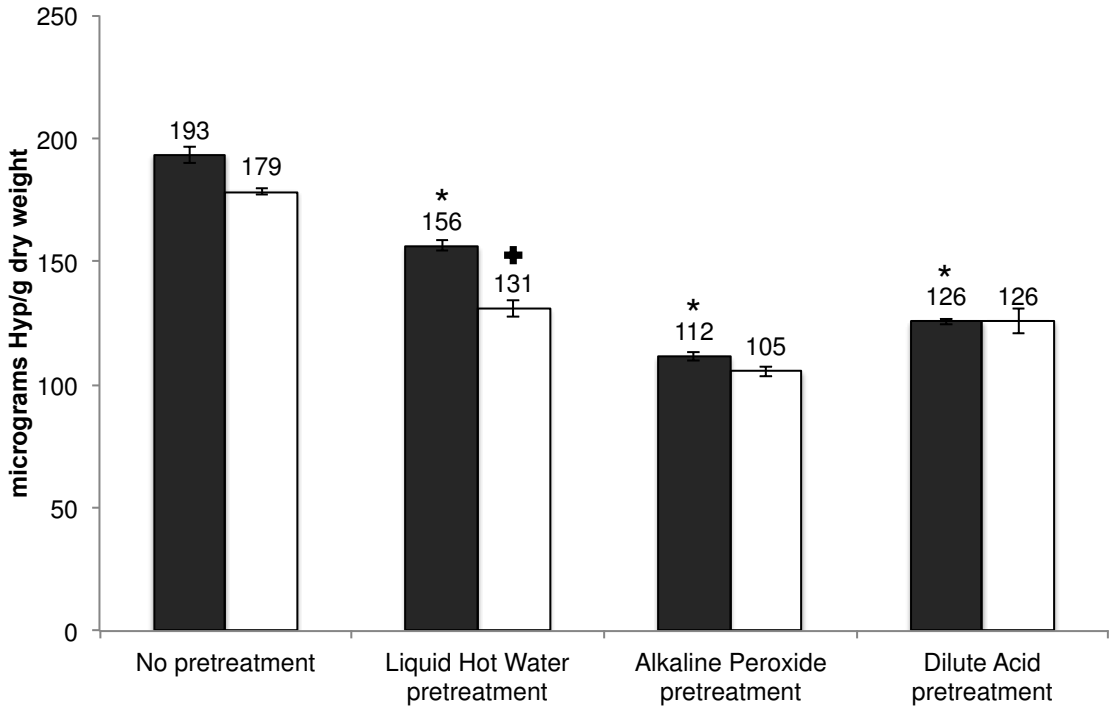

Figure 4. Hydroxyproline remaining in poplar biomass after different pretreatments and buffer-only (black bars) or Fermgen (white bars) treatment.
Bars show standard error, n = 3. *, $p < 0.05$ in comparison to no pretreatment; +, $p < 0.05$ in comparison to buffer-only treatment.

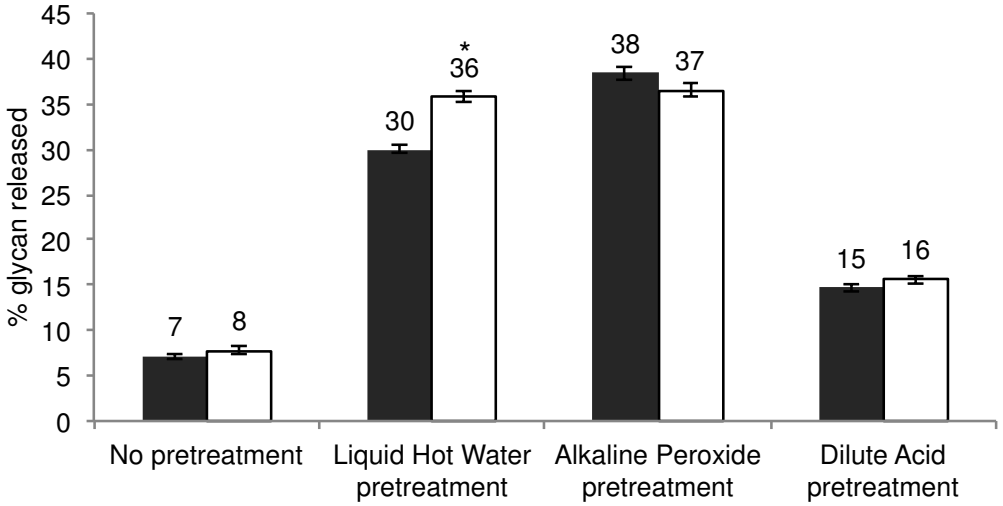

Figure 5. Effect of Fermgen treatment on glucose release from pretreated poplar biomass. Glucose released after 7-day cell wall digestion with Cellic CTEC2/HTEC2 without (black) or with (white) prior treatment with Fermgen protease. Bars show standard error, n = 6. *, $p < 0.005$ in comparison with no Fermgen treatment (Student's T-test).

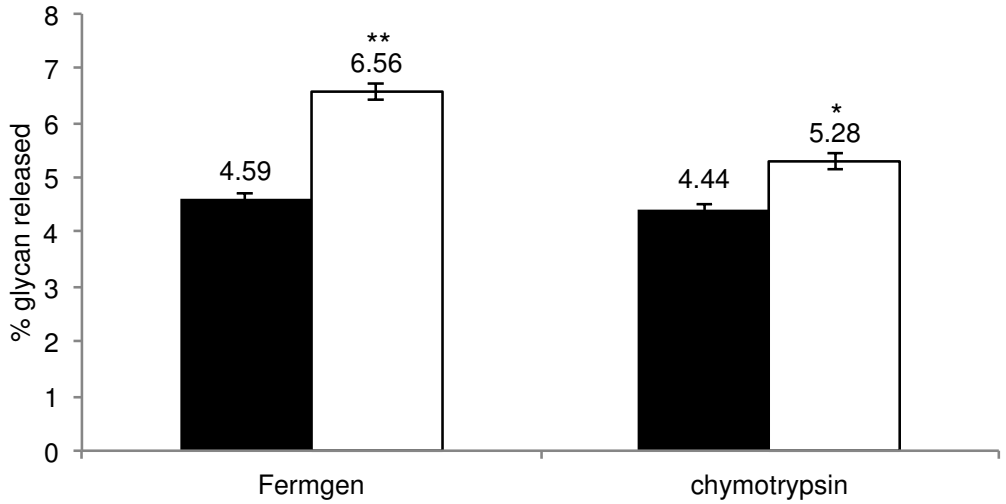

Figure 6. Effect of protease treatment on glucose release from liquid hot water pretreated poplar after 24 h cell wall digestion with Accellerase 1500.
Black, no protease treatment; white, with protease treatment. Bars show standard error, n = 12.
**, $P<0.0001$; *, $P<0.05$ in comparison with no protease treatment (Student's T-test).

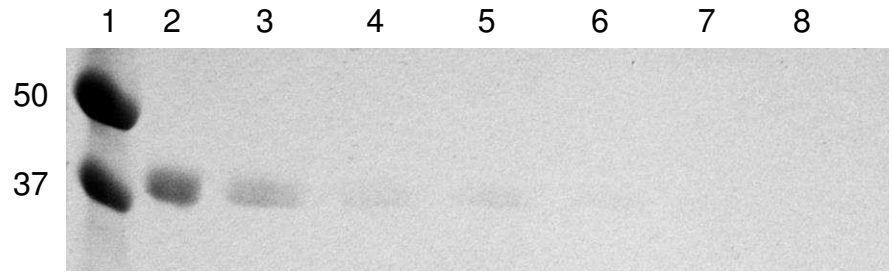

Figure 7. Time course of Fermgen auto-digestion.
SDS-PAGE analysis with Coomassie Brilliant Blue staining of Fermgen during the course of a 50 ºC incubation.
Lane 1, Molecular weight markers (sizes shown in kD); Lane 2, 1.75 µg Fermgen after 0 h incubation at 50 ºC; Lane 3, after 1 h incubation; Lane 4, after 2 h incubation; Lane 5, after 3 h incubation; Lane 6, after 4 h incubation; Lane 7, after 5 h incubation; Lane 8, after 24 h incubation.

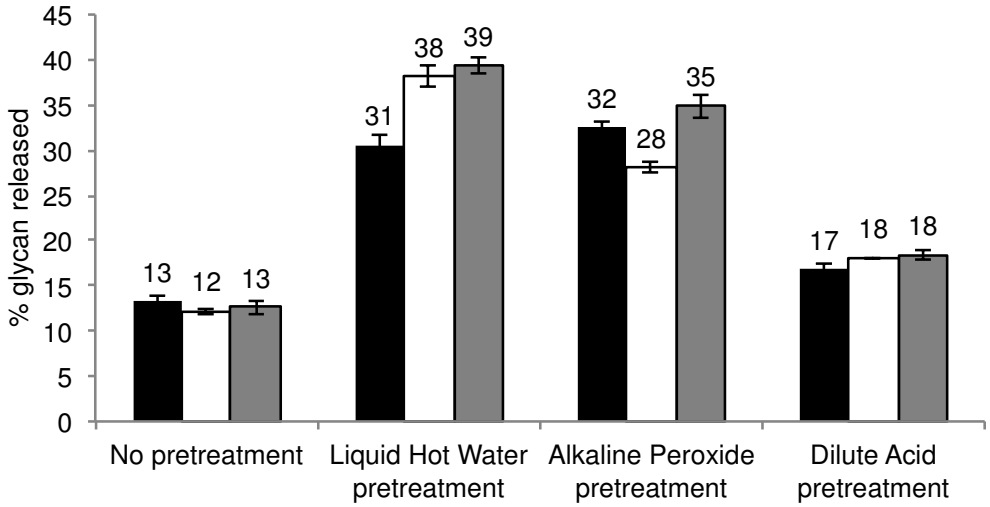

Figure 8. Effect of multiple doses of Fermgen on glucose release.
Glucose released after 7-day cell wall digestion with Cellic CTEC2/HTEC2 after prior treatment with different doses of Fermgen protease. Black, no Fermgen; white, single dose of Fermgen; grey, six doses of Fermgen over 24 hours. Bars show standard error, n = 3.

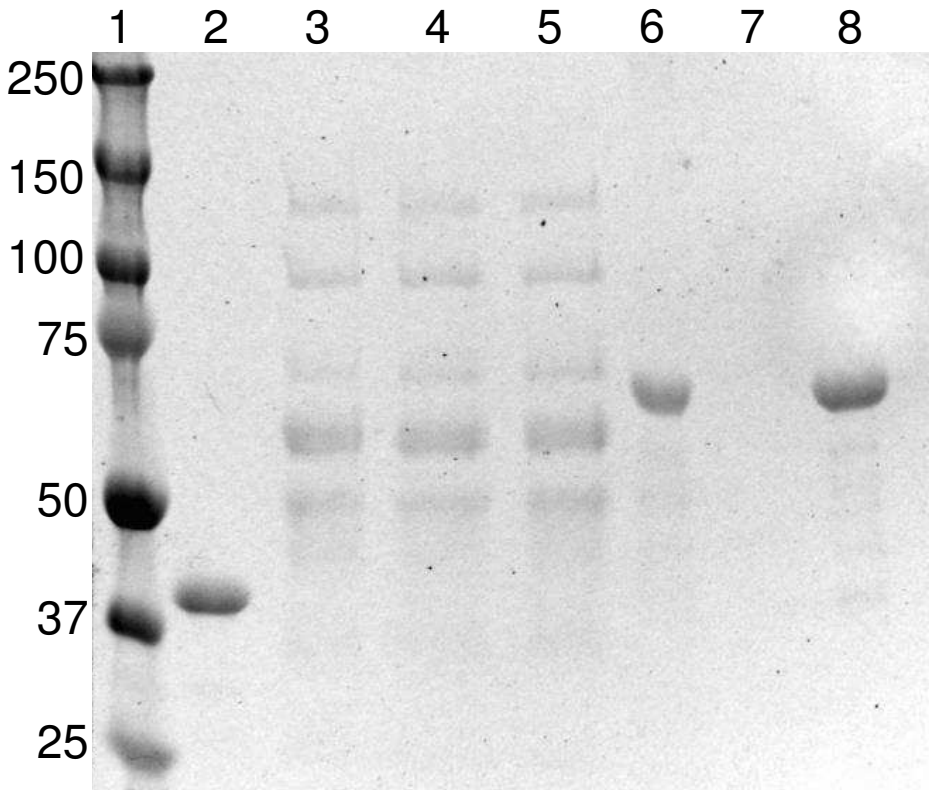

Figure 9. Analysis of Fermgen deactivation protocol.
SDS-PAGE analysis with Coomassie Brilliant Blue staining of CTEC2/HTEC2 (lanes 3-5) or BSA (lanes 6-8) after Fermgen treatment. Lane 1, Molecular weight markers (sizes shown in kD); Lane 2, 1.75 µg Fermgen; Lane 3, 6.22 µg CTEC2 with 3.09 µg HTEC2; Lane 4, 6.22 µg CTEC2 with 3.09 µg HTEC2, treated with active Fermgen; Lane 5, 6.22 µg CTEC2 with 3.09 µg HTEC2, treated with boiled and washed Fermgen; Lane 6, 2 µg BSA; Lane 7, 2 µg BSA, treated with active Fermgen; Lane 8, 2 µg BSA, treated with boiled and washed Fermgen.

CHAPTER THREE: THE EFFECT OF PREVENTING EXTENSIN CROSS-LINKING ON GLUCOSE YIELD IN ARABIDOPSIS

SYNOPSIS

Using another approach to investigate whether biofuel production from biomass is reduced by extensin cross-linking, I examined a suite of previously characterized Arabidopsis T-DNA insertional mutants predicted to have reduced extensin cross-linking. The mutated genes (At2g17720, *P4H5*; At5g25265, *HPAT1*; At1g19360, *RRA3*; At2g35610, *XEG113*; At3g01720, *SGT1*) each encode a different enzyme in the recently elucidated extensin post-translational modification pathway. I performed qRT-PCR on each line, and verified that the wild-type allele was not transcribed in every line except *sgt1*, which encodes a premature stop codon due to the T-DNA insertion, and which was previously shown to have reduced SGT activity. I grew wild-type and homozygous mutant plants together in a growth chamber and measured developmental traits (*e.g.*, time to bolting, time to flowering, inflorescence stem thickness and length, length of the inflorescence stem between the rosette and first silique [cauline stem], cauline stem biomass). Both *p4h5* and *hpat1-1* produced less biomass, with shorter, thinner inflorescence stems than wild-type plants ($p < 0.05$); the other mutant lines were not significantly different from wild-type. I measured hydroxyproline content in inflorescence stems to verify reduced extensin cross-linking in the mutants, and found no significant differences in hydroxyproline content between mutant and wild-type plants (~0.05% w/w). Consequently, I measured no differences in glucose release between mutant and wild-type plants after pretreatment with dilute base and digestion with Cellic CTEC2/HTEC2. In future experiments, this suite of mutants could be used to determine whether extensin cross-linking contributes to biomass recalcitrance in root tissue,

which contains high levels of extensin, even though roots are not commonly considered a source of biomass for biofuel production.

INTRODUCTION

In the previous chapter, I described experiments to test the effect of enzymatic removal of extensins on glucose release from poplar biomass. Another approach to reducing the amount of extensins in cell walls prior to pretreatment and cell wall digestion is to use genetic tools to disrupt either extensin production or extensin post-translational modification. For this approach, I used previously characterized T-DNA insertional mutants in Arabidopsis, as Arabidopsis is a more tractable system for genetic manipulation than poplar, and the Arabidopsis inflorescence stem is an accepted model for woody tissue[105].

Extensins contain a hallmark repeated sequence of $SP_{3\text{-}6}$. This sequence is highly post-translationally modified: serine is galactosylated, all adjacent prolines are hydroxylated, and finally a chain of 3-5 arabinoses is added to each hydroxyproline (Figure 10). Proline hydroxylation and subsequent arabinosylation are necessary to stabilize extensin in its extended linear conformation. Both electron microscopy[14,106] and molecular modeling[23,107] of glycosylated and deglycosylated extensins demonstrate that without these modifications, extensins adopt a more compact, globular form, and are unlikely to integrate correctly into the cell wall or possibly even be secreted.

The Arabidopsis genome, unlike that of poplar, includes a large extensin gene family, comprising 65 extensin genes. 22 of these encode classical extensins consisting of a single domain with an abundance of SP_n repeats[33]. Because of the size of the gene family in Arabidopsis, a mutation in a single extensin gene would not necessarily be expected to result in a mutant phenotype, yet mutant phenotypes have been observed in certain single extensin gene

mutants in Arabidopsis. For example, a mutation of *EXT3* (*RSH*) is embryo lethal in the homozygous state. In particular, embryos display defects in cell plate formation, preventing effective cytokinesis[24,108]. In addition, homozygous single mutations in each of five classical extensin genes (*EXT6*, *EXT7*, *EXT10*[†], *EXT11*, and *EXT12*) result in defective, shorter root hairs[23], and the homozygous mutant of an extensin chimera, LRX1, also has defective, shorter root hairs[109,110]. Although root hair phenotypes were observed in these single extensin gene mutants, it may be that no corresponding mutant phenotype occurs in inflorescence stems, perhaps due to genetic redundancy in the extensin gene family, or perhaps due to the different growth habits of root hairs and inflorescence stems (inflorescence stems were not examined in the published study). Therefore, rather than looking at extensin gene mutations, I examined mutants with defects in genes encoding the enzymes that perform post-translational modifications of extensins to determine the phenotypic effects of reduced extensin content on inflorescence stem cell walls. Reducing the activity of post-translational modification enzymes could affect the modification of all extensins, and improperly modified extensins may not be as abundantly cross-linked into cell walls.

The genes of the extensin post-translational modification pathway in Arabidopsis have recently been elucidated, as described below (Figure 10). A different enzyme performs each modification, and the enzymes of this pathway are each encoded by small multi-gene families or by single genes. Mutant analysis of these genes is therefore a more promising pathway for

---

[†] The mutant line of *EXT10* (At5g06640) that was investigated (SALK_099527) has two T-DNA insertions on chromosome 5. In the published study in which a root hair phenotype for *ext10* was reported[23], the homozygous line was bred using PCR primers designed for the secondinsertion, which is upstream of the promoter for At5g55900 (encoding a sucrase/ferredoxin-like family protein). Conclusions concerning root hair phenotypes in extensin gene mutants are questionable.

finding a phenotype than mutant analysis of extensin genes. Arabidopsis lines with T-DNA insertional mutations in the genes encoding each enzyme have been characterized. These mutants were found to have growth phenotypes, including root hair growth (Table 2). No previous studies have examined cell wall digestibility or extensin (hydroxyproline) content.

Post-translational modification of extensins begins with proline hydroxylation (Figure 10, Table 3). Proline is hydroxylated by prolyl-4-hydroxylases (P4H). Thirteen *P4H* genes have been identified in Arabidopsis[111], and one, *P4H5* (At2g17720), has been specifically shown to hydroxylate extensin prolines[112]. Arabinose is then added to hydroxyproline by a hydroxyproline arabinosyltransferase (HPAT). Three genes encoding this kind of enzyme have been identified in Arabidopsis (*HPAT1*, At5g25265; *HPAT2*, At2g25260; *HPAT3*, At5g13500)[113]. Mutation of any one of the three genes produces under-arabinosylated extensins; the *hpat1* mutant has the most significant difference in arabinosylation compared to wild-type[113]. The second arabinose of the arabinose chain is added by an arabinosyltransferase. Three genes encoding the proposed arabinosyltransferase have been identified in Arabidopsis (*RRA1*, At1g75120; *RRA2*, At1g75110; *RRA3*, At1g19360)[114]. Mutation of any one of the three genes leads to under-arabinosylated extensin, and mutation of *RRA3* causes the most significant difference in arabinosylation compared to wild-type[23]. The third arabinose of the arabinose chain is added by a different arabinosyltransferase. A single gene encoding this proposed arabinosyltransferase has been identified in Arabidopsis, *XEG113* (At2g35610)[115]. The arabinosyltransferases for the fourth and fifth arabinoses of the arabinose chain have yet to be identified. Finally, the serine in the SP_n repeat is galactosylated by a serine galactosyltransferase. Again, a single gene encoding this enzyme has been identified in Arabidopsis, *SGT1* (At3g01720)[116]. Interestingly, the SGT1

enzyme prefers a hydroxylated acceptor peptide[116], so a *P4H* mutation may prevent or reduce all further post-translational modifications including galactosylation of the adjacent serine.

I obtained Arabidopsis lines with T-DNA insertions in *P4H5*, *HPAT1*, *RRA3*, *XEG113*, and *SGT1* from the Arabidopsis Biological Resource Center (ABRC) and Nottingham Arabidopsis Stock Centre (NASC) (Table 3). I assessed whether mutation of the genes encoding the post-translational modification enzymes produced any significant physiological phenotype relevant to biofuel production, including inflorescence stem height and diameter. Since extensin cross-linking adds stability and marks the end of cell expansion, reduction of cross-linking may therefore permit cells to expand more than they normally would, resulting in larger cells or organs. Because cell expansion is involved in inflorescence development (bolting) and flower development it may be that reducing extensin cross-linking affects these developmental stages.

Several of these mutants have already been shown to have decreased cell wall extensin content (*p4h5*, *rra3*, and *xeg113-2*[23]). Therefore, if extensins are involved in biomass recalcitrance, the mutants should have increased cell wall digestibility compared to wild-type plants. I measured the hydroxyproline content of inflorescence stems of all the mutants as a proxy for the extensin content, and then compared the glucose release from cell walls (digestibility) of the inflorescence stem of each mutant to the digestibility of wild-type stems.

MATERIALS AND METHODS

Plant material

T-DNA insertional mutants (in ecotype Columbia (Col-0) background) in At3g01720 (*sgt1*, SALK_059879C), At5g25265 (*hpat1-2*, SALK_120066, and *hpat1-3*, SALK_048143) and At2g17720 (*p4h5*, SALK_152869) were obtained from ABRC[117]. T-DNA insertional mutants (Col-0 background) in At5g25265 (*hpat1-1*, GABI_298B03, NASC stock ID N428527) and

At1g19360 (*rra3*, GABI_223B05 [not GABI_233B05, as frequently reported[23,107]], NASC stock ID N421329) were obtained from NASC[118,119]. A homozygous T-DNA insertional mutant in At2g35610 (*xeg113-2*, SALK_066991) was a kind gift of Dr. Markus Pauly (UC Berkeley).

Genotyping

Plants homozygous for the T-DNA insertion were identified for each line by PCR, using two pairs of primers. One pair is gene-specific, positioned on either side of the site of the T-DNA insertion, resulting in DNA amplification if there is no T-DNA insertion present. The other pair consists of a gene-specific forward primer and a T-DNA specific reverse primer, resulting in DNA amplification only if the T-DNA insertion is present. Homozygous mutant plants show amplification only with the second set of primers. The primers used are shown in Table 4 and were designed with SIGnAL T-DNA express primer design (signal.salk.edu/tdnaprimers.2.html).

DNA was extracted from young leaves with the Shorty prep method[120] as follows: leaves were frozen at -80 ºC, then ground with a micropestle for 15 seconds over dry ice. 500 µL of Shorty buffer (0.2 M Tris/HCl, pH 9.0, 0.4 M LiCl [a substitution for the 0.25 M NaCl of the original protocol], 25 mM EDTA, 1% SDS) was added and the tissue was ground for another 15 seconds. Samples were centrifuged for 5 minutes at 21,130 g. 350 µL of supernatant was mixed with 350 µL of isopropanol and centrifuged for 10 minutes at 21,130 g. The supernatant was decanted and the pellet was dried at room temperature for 30 minutes. Samples were resuspended in 200 µL of TE buffer (10 mM Tris, pH 8.0, 0.1 mM EDTA).

PCR was performed on extracted DNA using a reaction mixture of 2 µL of DNA, 10 µL of EconoTaq PLUS GREEN (Lucigen, #30033-1), and 12.5 pmol of each primer in 20 µL total, in a BioRad C1000 Thermal Cycler with the following standard program: 95 ºC for 3 minutes, then 34 cycles of denaturation at 95 ºC for 30 seconds, annealing at 55 ºC for 45 seconds, and

elongation for l minute at 72 ºC, followed by 5 minutes at 72 ºC, holding at 4 ºC. 10 µL of each PCR product were analyzed on a 1% agarose gel. The remaining PCR product from reactions with LB+RP (amplifying the mutant allele) was purified with the Qiagen QIAquick PCR Purification Kit (Qiagen, 28104) and submitted to the Proteomics and Metabolomics Facility at CSU for sequencing on an ABI 3130xL Genetic Analyzer using the same primers (LB and RP). The resulting sequence data were analyzed with A plasmid Editor (ApE, available at biologylabs.utah.edu/jorgensen/wayned/ape/) to determine the precise location of the T-DNA insertion.

Effects of the T-DNA insertion on transcription of the mutated gene for each line were tested with quantitative reverse transcription PCR (qRT-PCR). Primers were designed to include an exon-exon junction in the amplicon to distinguish between genomic DNA or cDNA. Primers for Arabidopsis Actin-8 (At1g49240) were used as a positive control. Primer sequences are shown in Table 5 and were designed using NCBI Primer-BLAST (ncbi.nlm.nih.gov/tools/primer-blast/) to match only the target template of the gene of interest.

RNA was extracted from young leaves of three homozygous mutant and wild-type (Col-0) plants of each line with the RNeasy Plant Mini kit (Qiagen, 74904). RNA concentrations were quantified with the NanoDrop-ND-1000. 1 µL of DNase (Invitrogen, 18068-15) was added to 200 ng of RNA in 8 µL total, and samples incubated at room temperature for 15 minutes. 1 µL of 25 mM EDTA was added and samples incubated at 65 ºC for 10 minutes. cDNA was synthesized using a reaction mixture of 10 µL of DNase-treated RNA and 14 µL of qScript cDNA SuperMix (Quanta, 95048-500) in a final volume of 20 µL, in a BioRad C1000 Thermal Cycler with the following program: 25 ºC for 5 minutes, 42 ºC for 30 minutes, and 85 ºC for 5 minutes. QPCR was performed on cDNA using a reaction mixture of 1 µL of cDNA, 12.5 µL of PerfeCTa

SYBRgreen supermix (Quanta, 95068-500), and 5 pmol of each primer in 20 μL total, in a BioRad CFX-Connect Thermal Cycler with the following program: denaturation at 95 ^{o}C for 3 minutes, followed by 32 cycles of denaturation at 95 ^{o}C for 10 seconds and annealing and extension at 55 ^{o}C for 1 minute 15 seconds. A dissociation curve between 60-95 ^{o}C was performed on the products and analyzed with BioRad CFX manager. Expression levels were analysed in Microsoft Excel 2011 using the $2^{-\Delta\Delta CT}$ method[121] to determine relative levels of gene expression, using Actin-8 expression as a standard. 10 μL of each qPCR product were analyzed on a 1% agarose gel.

Phenotyping

Homozygous mutant and wild-type (Col-0) plants were grown in soil (Fafard 4P mix) in 4-inch pots in a temperature-controlled growth chamber (Percival Scientific) under long-day conditions (16 h light/8 h dark) at 22 ^{o}C. 3 pots per line were randomly distributed in each of two flats (6 pots total).

Plants were photographed 5 weeks and 7 weeks after germination. A number of parameters were measured, based on the growth stages defined by Boyes et al (2001), that could have relevance to extensin function in inflorescence stems[122]. These parameters included stem diameter at the inflorescence base, inflorescence stem length, cauline stem length (from the base of the rosette to the first silique), number of stems, number of branches off the main stem, wet weight of the entire inflorescence and wet weight of the cauline stem only.

Inflorescence stems were harvested when at least two siliques had shattered, between 43 and 61 days after germination. The cauline stems were flash-frozen in liquid nitrogen and freeze-dried for three days in a Labconco FreeZone[1] at -50 ^{o}C. The stems were re-weighed to determine

dry weight and ground in a TissueLyser II (Qiagen) twice for 1 minute at 30 oscillations/second each time. Ground samples were stored at -20 ºC.

Hydroxyproline assay

Three 150 mg pools of freeze-dried, ground cauline stem biomass from 3-5 plants were made for each line. Three 50 mg samples from each pool were weighed into 2 mL Sarstedt screwtop tubes and 1.4 mL of 6 M HCl was added to each tube. Samples were hydrolyzed at 100 ºC for 18 hours. The supernatant was assayed for hydroxyproline according to a modification of the methods of Kivrikko and Liesmaa[68] as follows: 500 µL were transferred to a new tube and pH-adjusted by the addition of 12 M NaOH to pH 3.0 (+/- 0.1). A higher pH led to development of a dark brown color in the samples that interfered with subsequent colorimetric analysis. The samples were centrifuged at 21,130 g for 1 minute to clarify the supernatant. 125 µL of sample supernatant was then mixed with 250 µL of 50 mM sodium hypobromide and incubated at room temperature for 5 minutes. The oxidation reaction was stopped by the addition of 125 µL of 6 M HCl, followed by 250 µL of para-dimethylaminobenzaldehyde (DMAB, 5% in n-propanol) to all tubes. Tubes were sealed, mixed by hand, and incubated at 70 ºC for 15 minutes. After cooling, the absorbance at 560 nm of each sample was measured in triplicate (200 µL each) in a 96-well plate, using a BioTek Synergy HT plate reader. Data were analyzed in Microsoft Excel 2011. Hydroxyproline standards with the same salt concentrations (4.2 M NaCl) and pH (3) as the samples were used to construct a standard curve.

Digestibility assay

Glucose release (digestibility) was analyzed following an adaptation of the protocol of Santoro *et al*[70]. Triplicate aliquots of freeze-dried, ground biomass for each plant of 5 mg +/- 0.3 mg were weighed into Sarstedt 2 mL screwtop tubes. 700 µL of 6.25 mM NaOH was added and

the tubes were vortexed and placed in a 90 ºC heat block for 3 hours. At the end of the pretreatment, the samples were placed on ice, then centrifuged at 21,130 g for 30 seconds, and 50 µL of cell wall digestion mix was added, consisting of Cellic CTEC2 (70 mg/g biomass, Novozymes), Cellic HTEC2 (2.5 mg/g biomass, Novozymes), 0.45 M sodium citrate, pH 4.5 (final concentration 30 mM), and 0.1% NaN_3 (final concentration 0.007%). Samples were incubated in a 50 ºC rotisserie oven with end-over-end rotation for 20 hours.

The tubes were centrifuged at 21,130 g for 1 minute. 650 µL of supernatant was removed to a new tube, diluted 1:10 in water, and measured for glucose content in a glucose oxidase/peroxidase (GOPOD) assay (Megazyme). Triplicate 20 µL aliquots of sample, glucose standard, or an enzyme-only mixture were mixed with 200 µL of GOPOD reagent and incubated at 40 ºC for 45 minutes in a 96-well plate. The absorbance at 510 nm was read in a BioTek Synergy HT plate reader. Data were analyzed in Microsoft Excel 2011. Glucose standards were used to construct a standard curve. The absorbance from the enzyme-only mixture was subtracted from the absorbance of the samples to correct for glucose present in Cellic CTEC2 and HTEC2 enzyme preparations.

RESULTS

Genotyping of mutants and establishment of insert site

Homozygous mutants were identified by agarose gel analysis of two genotyping PCRs. In the first PCR, a pair of gene-specific primers could amplify the wild-type allele, but not the mutant allele containing a T-DNA insert. In the second PCR, a T-DNA-specific and a gene-specific primer could amplify the mutant allele but not the wild-type allele (Figure 11 – Figure 15). Homozygous mutants, producing bands only from the second PCR, were identified for all lines (*xeg113*-2 and *sgt1* were received as homozygous mutants and were also confirmed with

genotyping PCRs) (Figure 14B, Figure 15B). Products of the second PCR (amplifying the mutant allele) from homozygous mutant plants were sequenced to determine the exact location of the T-DNA insert (Figure 11A – Figure 15A).

The sequence data for each mutant were compared to the genomic and T-DNA left border sequence (obtained from arabidopsis.org [SALK lines] or gabi-kat.de/ [GABI lines]) to determine the exact location of the T-DNA inserts. All T-DNA inserts were within exons (Table 4). For *p4h5*, T-DNA sequence begins at bp 892 (counting from the A of the start codon as base 1), within exon 4. Prior to the T-DNA insertion, bases 774 – 874 were deleted, which includes most of intron 3. Translation of the resulting genomic sequence results in a premature stop codon at amino acid 169. For *hpat1-1*, T-DNA sequence begins at bp 748, within exon 3 (the PCR products for *hpat1-2* and *hpat1-3* were not sequenced). For *rra3*, T-DNA sequence begins at bp 1307, within exon 2. For *xeg113-2*, T-DNA sequence begins at bp 2621, within exon 9. The direction of the T-DNA insert is such that the existence of premature stop codons could not be determined in *hpat1-1*, *rra3*, and *xeg113-2*. For *sgt1*, T-DNA sequence begins at bp 294, two bases before the end of exon 1. Translation of the resulting genomic sequence results in a premature stop codon at amino acid 111. The position of the T-DNA within an exon rather than an intron, as well as the creation of premature stop codons in *p4h5* and *sgt1*, make it much more likely that transcription and protein function will be affected in these mutant lines.

Transcript abundance in homozygous mutants

I extracted RNA from leaves of homozygous mutants for each line and subjected the RNA to reverse transcription followed by qPCR, as well as agarose gel electrophoresis, in order to quantify transcript abundance for the gene of interest. I analyzed all seven mutant lines using two primer sets for each gene. Generally, one primer set was positioned 5' to the T-DNA insert

("5' primer set 1") and the other set flanked the T-DNA insert ("3' primer set 2"). Exceptions were primers for *p4h5*, where primer set 2 is 3' to the T-DNA insert rather than flanking, and *sgt1*, where both primer sets are 3' to the T-DNA insert. Actin-8 primers were used as a positive control, and products from actin-8 amplification were detected in wild-type and mutant plants to the same degree. Delta-delta Ct values were calculated to compare relative expression of the gene of interest in mutant and wild-type plants[121] (Figure 16A). In all cases where a primer set amplified mutant cDNA, the relative expression was less in mutant than in wild-type plants, except for *sgt1*, in which expression was much greater (600% with 5' primer set 1 and 800% with 3' primer set 2, Figure 16B).

Several mutants (*p4h5*, *hpat1-1*, *xeg113-2*) were previously analyzed by performing PCR on cDNA followed by agarose gel electrophoresis, and were found to have no detectable transcript compared to wild-type[23,113,115]. However, in my experiments, transcript could be detected, depending on the primer set. For *p4h5*, the 5' primer set 1 did not amplify a product from cDNA, while the 3' primer set 2 did, although at reduced levels compared to wild-type (Figure 16A, Figure 17). Neither primer set flanks the T-DNA insertion in *p4h5*, so the absence of amplification from 5' primer set 1 was not due to the increased size of the transcript from the T-DNA insertion. This result indicates that transcription of the mRNA was disrupted by the T-DNA insertion, so that the wild-type allele was not transcribed; considering also that the genomic DNA sequence was rearranged and that the putative amino acid sequence in the mutant contains premature stop codons, a functional protein product is extremely unlikely to be made, confirming the earlier results of Velasquez *et al*[23]. Amplification of cDNA at the 3' end but not the 5' end of the gene may be explained by multiple T-DNA insertions at the same locus. A second T-DNA may have inserted in the reverse orientation of the first, so that the 35S promoter

in the T-DNA insert drives transcription of the 3' fragment of the gene. Further sequencing data will be necessary to resolve this anomaly.

The three *hpat1* mutants had different degrees of reduction in *HPAT* transcripts. For all three *hpat1* mutant lines, products were detected with the 5' primer set 1. The 3' primer set 2 also amplified products in both *hpat1-2* (SALK_120066) and *hpat1-3* (SALK_048143), although at reduced levels compared to wild-type (Figure 16A, Figure 18). However, the 3' primer set 2 did not amplify any detectable product in *hpat1-1* (GABI_298B03), confirming the previous results of Ogawa-Ohnishi *et al.*[113]. This result indicates that in the *hpat1-1* mutant line, either transcription of the mRNA was disrupted by the T-DNA insertion, or the insertion is present in the mRNA, making the amplicon too large to be amplified with the experimental conditions. In either case, the wild-type allele would not be translated and a functional protein product is unlikely to be made. For the *hpat1-2* and *hpat1-3* lines, further analysis would be required to determine whether the mRNA or protein are disrupted due to the location of the qPCR primers relative to the T-DNA insert. Therefore, for further studies, only *hpat1-1* was used.

For *rra3*, slightly less product than wild-type was detected with the 5' primer set 1, while the 3' primer set 2 did not amplify a detectable product (Figure 16A, Figure 19). Similarly, for *xeg113-2*, less product than wild-type was detected with the 5' primer set 1, while the 3' primer set 2 did not amplify a detectable product, confirming the previous results of Gille *et al*[115] (Figure 16A, Figure 20). As the 3' primer set 2 flanks the T-DNA insertion in *rra3* and *xeg113-2*, these results indicate, as in *hpat1-1*, that either transcription of the mRNA was disrupted by the T-DNA insertion, or the insertion is present in the mRNA so that the region between the primers is too large to be amplified with the given reaction conditions.

For *sgt1*, both primer sets were 3' to the T-DNA insertion, and more product was detected from each primer set in *sgt1* than in wild-type, indicating greater SGT1 transcript abundance (Figure 16B, Figure 21). The T-DNA insertion may be up-regulating the transcriptional machinery, as it occurs two bases before the end of exon 1. Even though transcript abundance is not reduced in *sgt1*, my sequencing data indicate that a premature stop codon should be present in the transcript. Furthermore, Saito *et al.* confirmed that *sgt1* has negligible SGT1 activity in microsomal preparations, so wild-type function of the *sgt1* allele is disrupted, and a mutant phenotype in the inflorescence is still expected with this line[116].

Effect of mutant genotype on developmental phenotypes

Mutations in each of the genes encoding extensin post-translational modification enzymes have been reported to have various phenotypes (Table 2). Interestingly, reduced growth is only reported in cells undergoing tip growth (root hairs and pollen tubes). Otherwise, either increased growth is observed in roots or leaves, or there is no difference between the mutant and wild-type.

I measured a suite of developmental characteristics in which reduced extensin cross-linking might produce a phenotype, based on the growth stages of Arabidopsis described by Boyes *et al.*[122] (Table 6). In my experiments, the mutants had no obvious or striking phenotype, appearing comparable to wild-type plants at all stages of growth (Figure 22, Figure 23). I found no differences in bolting, flowering, or silique ripening between mutant and wild-type plants, as they bolted, flowered, and matured on approximately the same day after germination. However, it should be noted that all plants in my experiments developed more rapidly than those of Boyes *et al.* by approximately 5 days for each stage.

Only two mutants lines, *p4h5* and *hpat1-1*, showed a significant growth phenotype compared to wild-type. Both produced thinner and shorter stems (Figure 24), with fewer branches off the main bolt, and less biomass (Figure 25) ($p < 0.04$, Student's T-test). The stems of *p4h5* were on average 1.2 mm in diameter (wt = 1.4 mm, $p = 0.01$) and 319 mm long (wt = 362 mm, $p = 0.004$), with a cauline stem of 114 mm (wt = 128 mm, $p = 0.04$). The stems of *hpat1-1* were on average 1.3 mm in diameter ($p = 0.15$) and 315 mm long ($p = 0.02$), with a cauline stem of 114 mm ($p = 0.04$) (Figure 24). *p4h5* had on average 2.9 branches off the main bolt (wt = 3.7, $p = 0.02$), while *hpat1-1* had 3.2 branches ($p = 0.07$). The total inflorescence biomass was on average 1.9084 g for *p4h5* (wt = 2.540 g, $p = 0.004$) and 1.9675 g for *hpat1-1* ($p = 0.01$). The cauline stem wet and dry weights were similarly reduced compared to wild-type (0.8989 g wet, 92.6 mg dry), as *p4h5* was 0.6826 g wet ($p = 0.006$), 63.9 mg dry ($p = 0.003$) and *hpat1-1* was 0.6832 g wet ($p = 0.01$), 64.1 g dry ($p = 0.003$) (Figure 25). The other mutants had no significant differences in growth compared to wild-type ($p > 0.3$), except for cauline stem length, which was shorter than wild-type in the case of *xeg113-2* (average *xeg113-2* cauline stem length = 107 mm, $p = 0.01$) (Figure 24).

Effect of mutant genotype on hydroxyproline content

The amount of hydroxyproline in inflorescence stems of mutant and wild-type lines was measured in order to determine whether the mutations affected the abundance of extensins bound in cell walls (Figure 26). Surprisingly, the hydroxyproline content in stems was not significantly different from wild-type in any mutant lines ($p > 0.2$, Student's T-test). I measured an average of 0.048% Hyp/mg dry weight in wild-type plants, and a comparable amount in each mutant line: *p4h5*, 0.050%; *hpat1-1*, 0.054%; *rra3*, 0.044%; *xeg113-2*, 0.044%; *sgt1*, 0.046%. The stem hydroxyproline content I measured is comparable to what has been reported in tomato stems

(0.05%)[123], as well as different Arabidopsis ecotypes (0.06% in ecotype Landsberg *erecta* and 0.03% in Wassilewskija[124]).

Effect of mutant genotype on digestibility

Since each mutant line was reported to have a mutant phenotype, presumably due to disrupted post-translational modification of extensins, I anticipated that inflorescence stem extensins in the mutants would also be affected, resulting in more digestible inflorescence stem biomass from the mutants compared to wild-type. I determined how much glucose was released after pretreatment and cell wall digestion depending on the plant's genotype by measuring glucose release from 9 – 12 plants from each line in triplicate (Figure 27). No line had significant differences in glucose released compared to wild-type ($p > 0.4$, Student's T-test). The median value measured for wild-type plants was 8.74% glucose/mg dry weight, and the mutants were comparable: *p4h5*, 9.22%; *hpat1-1*, 9.00%; *rra3*, 8.91%; *xeg113-2*, 9.02%; *sgt1*, 9.12%. The average amount of glucose released was similar to that reported in another study on Arabidopsis using similar methods (9.15 ± 0.81% glucose/mg dry weight [this study] versus 12.81 ± 1.48% glucose/mg dry weight[70]).

DISCUSSION

Extensin post-translational modification mutants and plant development

A cross-linked extensin network is related to cessation of primary cell growth[27]. It may be that cross-linking of extensins contributes to developmental signals required for ending primary cell growth. If cross-linking is reduced or absent, those developmental signals may be delayed, leading to larger cells. A second hypothesis is that in the absence of cross-linking, a physical barrier to further cell expansion is not introduced, again leading to larger cells. These hypotheses are supported by work in tomato, in which silencing of *SlP4H7* in mature plants by

virus-induced gene silencing has been shown to cause a small but significant increase in cell expansion, leading to larger leaves[123]. Furthermore, mutation of *HPAT1*, *XEG113*, or *SGT1*, which presumably causes decreased extensin cross-linking, results in increased growth phenotypes in Arabidopsis, including longer hypocotyls (*hpat1-1* and *xeg113-2*), larger rosettes (*xeg113-2*) and longer roots and larger leaves (*sgt1*).

Results supporting the hypothesis that decreased extensin post-translational modification can lead to increased growth have also been obtained using chemical treatments to prevent prolyl-4-hydroxylase activity. Treatment of Arabidopsis with 3,4-dehydro-L-proline (DHP, a P4H inhibitor) makes root cells grow longer than without treatment[125]. Treatment of tobacco mesophyll protoplasts with DHP leads to strikingly abnormal cell morphologies. The plastids cluster around the nucleus, cell division is inhibited, and protoplasts develop into "monster" cells of exceptional size and irregular shape[126].

Rather than a phenotype of increased growth, in my experiments, mutant plants were developmentally identical to wild-type plants except for *p4h5* and *hpat1-1*, which were smaller than wild-type plants, producing shorter, thinner stems and less biomass. The appearance of a mutant phenotype in *p4h5* and *hpat1-1* may be due to the known drastic defects in root hair growth in these mutants[23,107]. Root hairs are essential for nutrient uptake, so defective root hairs may lead to reduced plant growth from nutrient deficiency or other stress, although symptoms such as chlorosis were not observed. The difference in the observed phenotype between *SlP4H7*-silenced tomato plants and *p4h5* Arabidopsis mutants could also be because VIGS silenced *SlP4H7* in adult plants that were able to develop in the presence of functional SlP4H7; *p4h5* mutants, however, never had functional P4H5, possibly creating a more dramatic effect on overall plant growth.

Effectiveness of mutant lines for associating hydroxyproline content with digestibility

The current model of extensin incorporation in the cell wall postulates that the full suite of post-translational modifications is required for formation of a functional extensin cross-linked network[23]. I analyzed hydroxyproline content in the inflorescence stems of each mutant, and expected to see a reduction in hydroxyproline, especially in *p4h5*, due to decreased extensin cross-linking resulting from incomplete post-translational modification. Several of the mutants (*p4h5*, *rra3*, and *xeg113-2*) had previously been shown to have reduced extensin content or epitope signal in root tissue[23].

I found no significant differences in hydroxyproline content in stems between any of the mutant and wild-type Arabidopsis plants. Since many of the enzymes of the extensin post-translational modification pathway are encoded by multi-gene families, genetic redundancy or tissue-specificity of gene expression may have reduced or nullified any phenotypic effects. Although I selected the mutant line from each gene family with the strongest phenotype (Table 2, Table 3), other members of the gene family may have had sufficient activity to adequately modify extensins, so that extensin post-translational modification and cross-linking, and therefore hydroxyproline content, were unaffected. The XEG and SGT enzymes are each thought to be encoded by single genes rather than gene families. However, other genes encoding these enzymes may exist that lack strong sequence homology to the known genes, so functional rather than genetic redundancy may be a factor in the unchanged hydroxyproline content of *xeg113-2* and *sgt1*.

Other analysis of extensin content in these mutant lines has been confined to the roots, particularly root hairs. Expression of P4H and extensin genes is much higher in roots than stems in Arabidopsis[33], and extensin content in Arabidopsis, as with most dicots, is much higher in

roots than in stems (0.84% in roots[23] vs. 0.05% in stems[this study]). I only examined hydroxyproline content in stems, since Arabidopsis inflorescence stems are an analogue for woody tissue[105], used in second-generation biofuel production. Therefore, tissue differences in gene expression, or differences in inherent hydroxyproline levels in different tissues, could explain the unchanged hydroxyproline content in the cauline stems of the mutant lines.

Another explanation is that perhaps complete post-translational modification is not in fact necessary for extensin secretion, cell wall association, or cross-linking. Root hair phenotypes in *rra3* and *xeg113-2* were presumed to be a result of reduced extensin cross-linking, determined by a reduction in extensin epitope signal in roots, using the JIM20 antibody[23]. However, this antibody detects glycosylated extensin[65], so the reduced signal could have been because of the absence of the second (*rra3*) or third (*xeg113-2*) arabinose rather than absence of extensin. In that case, if extensin modification by RRA3 and XEG113 is not necessary for extensin secretion or possibly for cell wall association, *rra3* and *xeg113-2* mutants would be expected to have normal levels of hydroxyproline. Other studies suggest that even extensin proline hydroxylation is unnecessary for extensin-cell wall interactions. Treatment with α,α'-dipyridyl prevents proline hydroxylation. When used on aerated carrot root disks, carrot extensin is not hydroxylated or glycosylated, since glycosylation occurs only on hydroxyproline residues. However, α,α'-dipyridyl treatment does not prevent secretion or tight binding of proline-containing proteins to the cell wall[127]. Deglycosylation of extensins clearly alters the conformation from linear to globular[14,106], so the non-glycosylated extensins resulting from α,α'-dipyridyl treatment are likely to be globular, but perhaps the globular form retains some functionality. In that case, mutation of any of the genes encoding extensin post-translational modification enzymes would not be adequate for suppressing extensin-cell wall interactions.

Because extensin cross-linking is predicted to be reduced in all the mutants, I expected stems from mutant plants to be more digestible than stems from wild-type plants. However, since extensin (hydroxyproline) content was unchanged in the mutant stems, this system was not capable of answering the question I posed: does extensin cross-linking increase recalcitrance? Further options include addressing genetic redundancy by creating lines with reduced function of all the genes in a particular family (either by crossing multiple single mutants or by using RNA-interference to reduce translation of transcripts from multiple genes simultaneously), and addressing tissue specificity of extensin cross-linking by examining hydroxyproline content and digestibility of roots of all five mutant lines. Extensin content must be reduced before drawing any conclusions about the influence of extensin cross-linking on biomass recalcitrance.

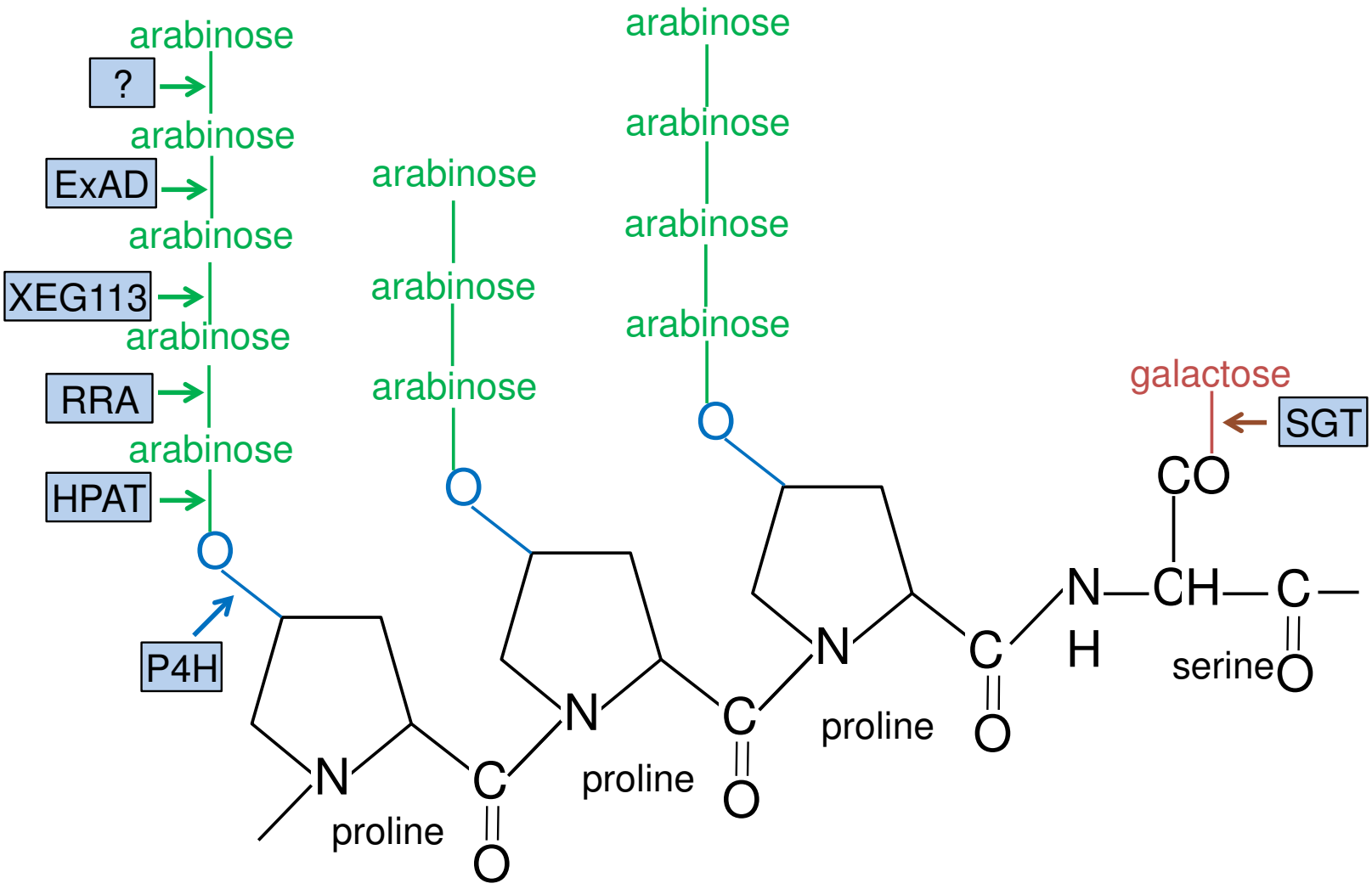

Figure 10. Diagram of extensin post-translational modification pathway.

Table 2. Summary of previous findings regarding enzymes known to be involved in extensin post-translational modification.

Enzyme name Loci	Sub-cellular localization	T-DNA insertional mutants	Absence of transcript?	Mutant phenotypes	Extensin content
Prolyl-4-hydroxylase					
At2g43080 (P4H1) At3g06300 (P4H2) At2g17720 (P4H5)	ER and Golgi[23,112]	SALK_152869 (*p4h5*)	Yes[23]	Longer roots and shoots, larger leaves[123] Shorter root hairs[23,112]	0.55% Hyp w/w (in roots; wt = 0.84%)[23] Reduced signal in root hairs (JIM20)[107]
Arabinosyltransferase					
At5g25265 (HPAT1) At2g25260 (HPAT2) At5g13500 (HPAT3)	Trans-membrane, cis-Golgi[113]	GABI_298B03 (*hpat 1-1*) SALK_120066 (*hpat1-2*) SALK_048143 (*hpat 1-3*)	Yes (*hpat 1-1*)[113] ND (hpat 1-2, hpat 1-3)	Longer hypocotyls, thinner cell walls, early flowering, early leaf senescence, shorter pollen tubes[113] Shorter root hairs[107]	ND
At1g75120 (RRA1) At1g75110 (RRA2) At1g19360 (RRA3)	Golgi[23]	GABI_223B05 (*rra3*)	ND	Shorter root hairs[23]	30% of wild-type signal in dot immunoblot of roots (JIM12)[23] No signal in immunolabeled root hairs (JIM20)[107]
At2g35610 (XEG113)	Trans-Golgi[23]	SALK_066991 (xeg113-2)	Yes[115]	Longer hypocotyls, larger rosette diameter, earlier bolting[115] Shorter root hairs[23]	0.49% Hyp w/w in seedlings (wt = 0.50%)[115] 40% of wild-type signal in dot immunoblot of roots (JIM12)[23]
Serine galactosyltransferase					
At3g01720 (SGT1)	ER membrane, cis-Golgi[116]	SALK_059879C (*sgt1-1*)	ND	Longer roots, larger leaves[116] Shorter root hairs[107]	ND

Table 3. Rationale for selection of genes for investigation in this study.

Gene	Function	# of genes in family/ # characterized	Gene selected for this study	Gene locus Mutant line	Rationale
Prolyl-4-hydroxylase (*P4H*)	Converts proline to hydroxyproline	13/4 (*P4H1*, *2*, *5*, *13*)	*P4H5*	At2g17720 SALK_152869	Mutant has strongest phenotype (shortest root hairs) of mutants in three characterized *P4H* genes. *P4H5* is the most highly expressed of P4H genes (with highest expression in roots).[23]
Hydroxyproline arabinosyltransferase (*HPAT*)	Adds first arabinose to hydroxyproline	3/3	*HPAT1*	At5g25265 GABI_298B03	Mutant has greatest reduction of activity in microsomal fractions.[113]
Reduced Residual Arabinose (*RRA*)	Adds second arabinose to chain (function not confirmed)	3/3	*RRA3*	At1g19360 GABI_223B05	Mutant has most reduced extensin epitope signal in roots.[23]
Xyloglucanase 113 (*XEG*)	Adds third arabinose to chain (function not confirmed)	1/1	*XEG113*	At2g35610 SALK_066991	Single gene.[115]
Serine Galactosyltransferase (*SGT*)	Adds galactose to serine; requires hydroxylation of adjacent proline	1/1	*SGT1*	At3g01720 SALK_059879C	Single gene.[116]

Table 4. Primers used to genotype Arabidopsis.

Locus Gene name Mutant line	Gene-specific left primer (LP, 5′ → 3′)	Gene-specific right primer (RP, 5′ → 3′)	T-DNA left border primer (LB, see end of table)	Gene-specific PCR product size (in bp) (LP + RP)	T-DNA specific PCR product size (in bp) (LB + RP)	Distance (in bp) of T-DNA from start codon (exon of insertion, exon location in bp)
At2g17720 P4H5 SALK_152869	CATTTTGAGAGCTCGTTCCAC	AGTTATTTCTTGGGAGCCTCG	LbB1.3	1065	537-837	892 (exon 4, 848-945)
At5g25265 HPAT1-1 GABI_298B03	ATCTCCGCTAATGCTCCTCTC	TCATGAAAATAGGCATGGATGATA	GABI T-DNA	1216	700-800	748 (exon 3, 647-849)
At5g25265 HPAT1-2 SALK_120066	TCCATGCCTATTTTCATGAGC	TTTCAGGAATTTGCAGAGACC	LbB1.3	1107	535-835	ND
At5g25265 HPAT1-3 SALK_048143	AAAAGCGAAGAGTGGAGAAGC	ATGCAAGATCCGAGTAAACCC	LbB1.3	1255	581-881	ND
At1g19360 RRA3 GABI_223B05	TTATGTAAAGAGAACGATGTTGCG	AGAAGGCATGAAGCAGTAAACATT	GABI T-DNA	801	650-900	1307 (exon 2, 424-1540)
At2g35610 XEG113 SALK_066991	ACAATGCAGGAGGTTTCATTG	AATCTTTCTTCTCGCTCCTGC	LbB1.3	1116	504-804	2625 (exon 9, 2561-2675)
At3g01720 SGT1 SALK_059879C	GTGAGCTGTATCTTGGCGAAC	CGCCACTACCTAATCATAACC	LbB1.3	1060	488-788	294 (exon 1, 1-296)
T-DNA left border primer	LB, 5’ → 3’					
LbB1.3	ATTTTGCCGATTTCGGAAC					
GABI T-DNA	ATAATAACGCTGCGGACATCTACATTTT					

Table 5. Primers used to determine expression levels of mutant genes in Arabidopsis.

Locus Gene name	Primer set	Forward primer (5' → 3')	Reverse primer (5' → 3')	gDNA product size (bp)	cDNA product size (bp)
At2g17720 P4H5	5' primer set 1	GCATCTTCGTTATCAGCCGC	TCTTTGCTCCCACCGGTTTT	741	360
	3' primer set 2	ATGATGGTGGCGAGACTGTG	GTGCAAGCTCGAAGGGTCTA	176	176
At5g25265 HPAT1	5' primer set 1	TTCAAGCTTCTGCTGGACCC	GGGATGGATCAGGGTTATGTTG	213	141
	3' primer set 2	GGGCTTGGAGCTGCGTTTCC	TCCCATGGTGGCTGAATCATGAAG	471	317
At1g19360 RRA3	5' primer set 1	TGATCGCGTGGCAGATAGAC	ATCGGGGAAGCTATCAAGCG	307	307
	3' primer set 2	TGAGCTTCTGGGTCTTTGGC	ACTCGTTCTGGCGATTCACA	517	264
At2g35610 XEG113	5' primer set 1	TTTGGTTTGGGCATCCTGGT	CTTCATTGCTCCGCTTTGGG	458	294
	3' primer set 2	TGACGATCTACGCGACTGTG	GGAATTGCCTTGAGCCTCCT	753	236
At3g01720 SGT1	5' primer set 1	TCCACGAGAAGGTGTTGAGC	GACGCAGTATGAGTCCCTCG	322	242
	3' primer set 2	TGGGAGTTTAAGGCAGCTCG	TTCCTCCAGTTCGCCTTGTC	608	434
At1g49240 ACT8	Reference gene	ATGAAGATTAAGGTCGTGGCA	CCGAGTTTGAAGAGGCTAC	524	417

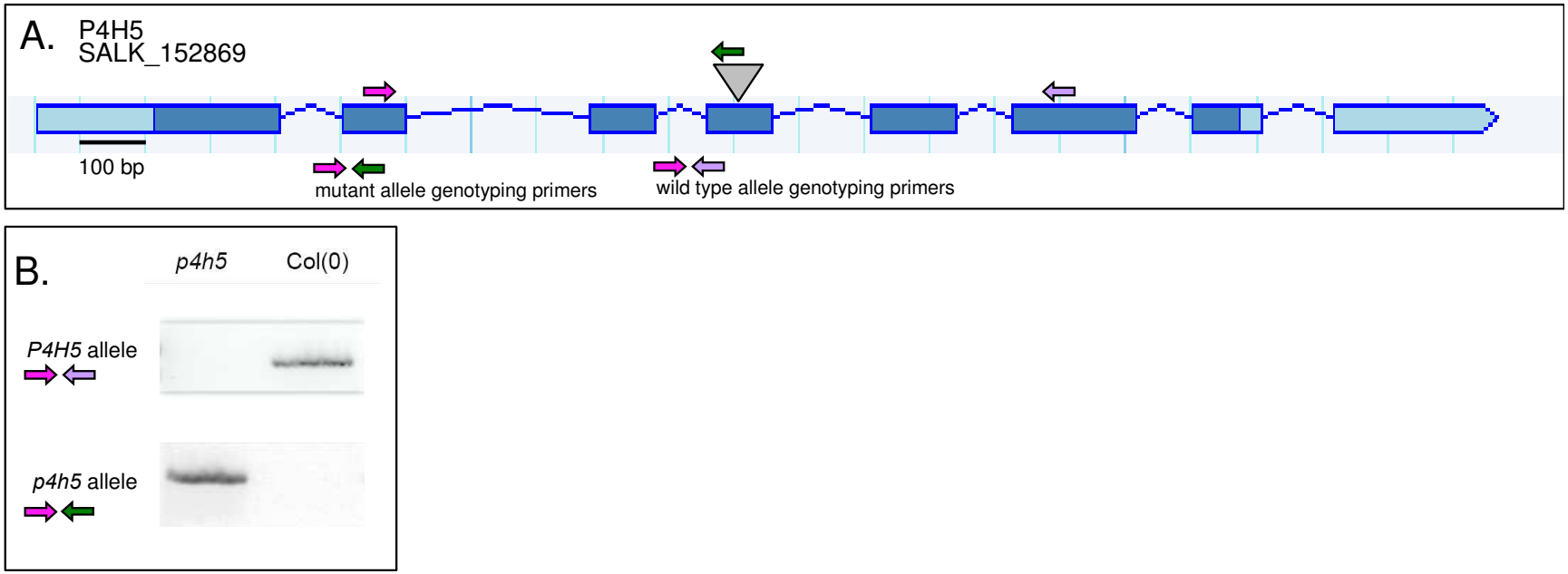

Figure 11. Genotyping of *p4h5*.
A. Gene model with locations of genotyping primers and T-DNA insert (locations of primers and T-DNA insert [grey triangle], shown above gene model, to scale). Gene model diagram from arabidopsis.org. Location of T-DNA insert determined by sequencing *p4h5* allele PCR product from B, below. B. Agarose gel electrophoresis results of genotyping PCR with wild-type (top row) and mutant (bottom row) allele-specific primers on mutant (left lane) and wild-type (right lane) plants.

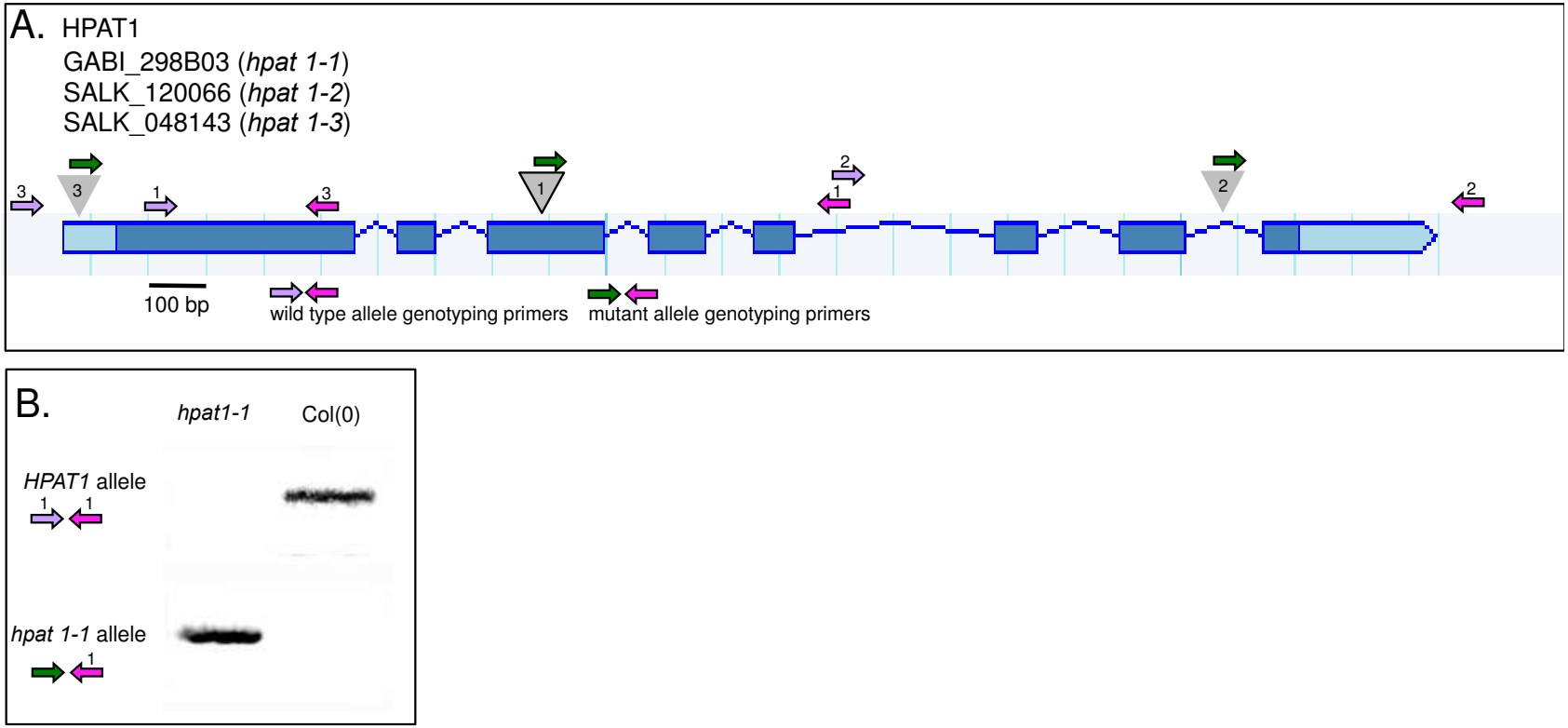

Figure 12. Genotyping of *hpat1-1*, *hpat1-2*, and *hpat1-3*.
A. Gene model with locations of genotyping primers and T-DNA insert (locations of primers and T-DNA insert [grey triangles], shown above gene model, to scale). Gene model diagram from arabidopsis.org. Location of T-DNA insert determined by sequencing *hpat1-1* allele PCR product from B, below. B. Agarose gel electrophoresis results of genotyping PCR with wild-type (top row) and mutant (bottom row) allele-specific primers on mutant (left lane) and wild-type (right lane) plants.

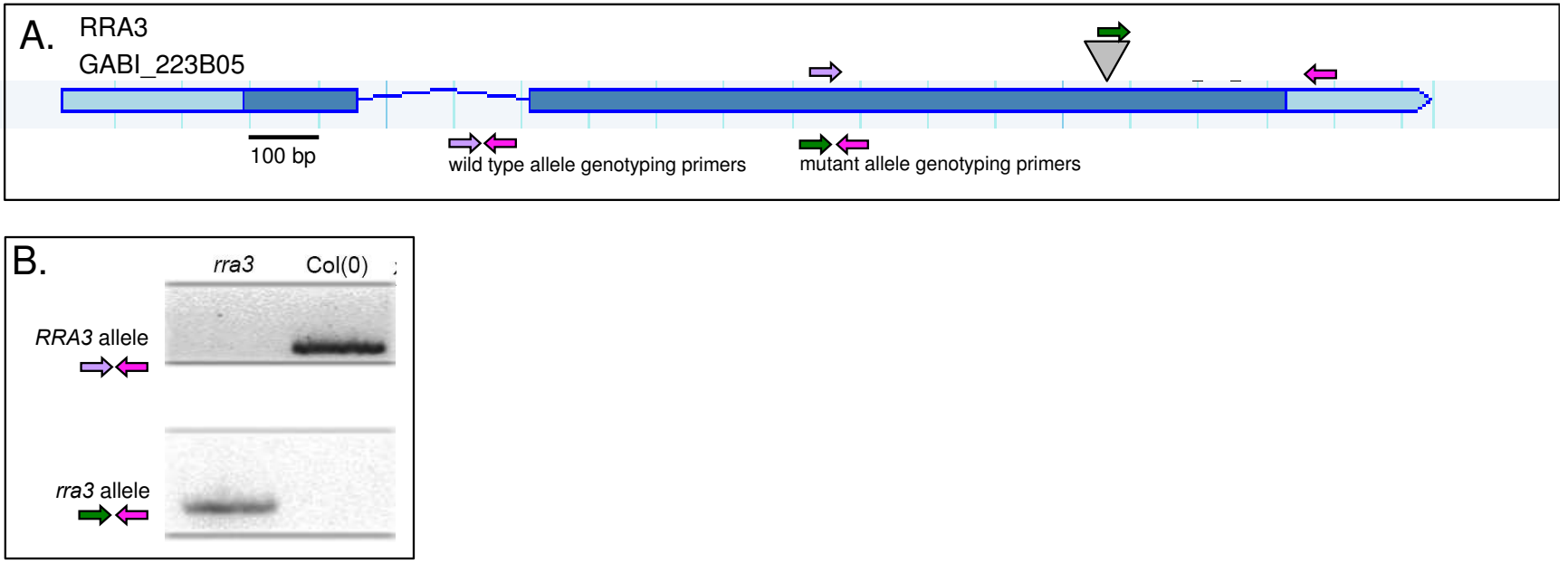

Figure 13. Genotyping of *rra3*.
A. Gene model with locations of genotyping primers and T-DNA insert (locations of primers and T-DNA insert [grey triangle], shown above gene model, to scale). Gene model diagram from arabidopsis.org. Location of T-DNA insert determined by sequencing *rra3* allele PCR product from B, below. B. Agarose gel electrophoresis results of genotyping PCR with wild-type (top row) and mutant (bottom row) allele-specific primers on mutant (left lane) and wild-type (right lane) plants.

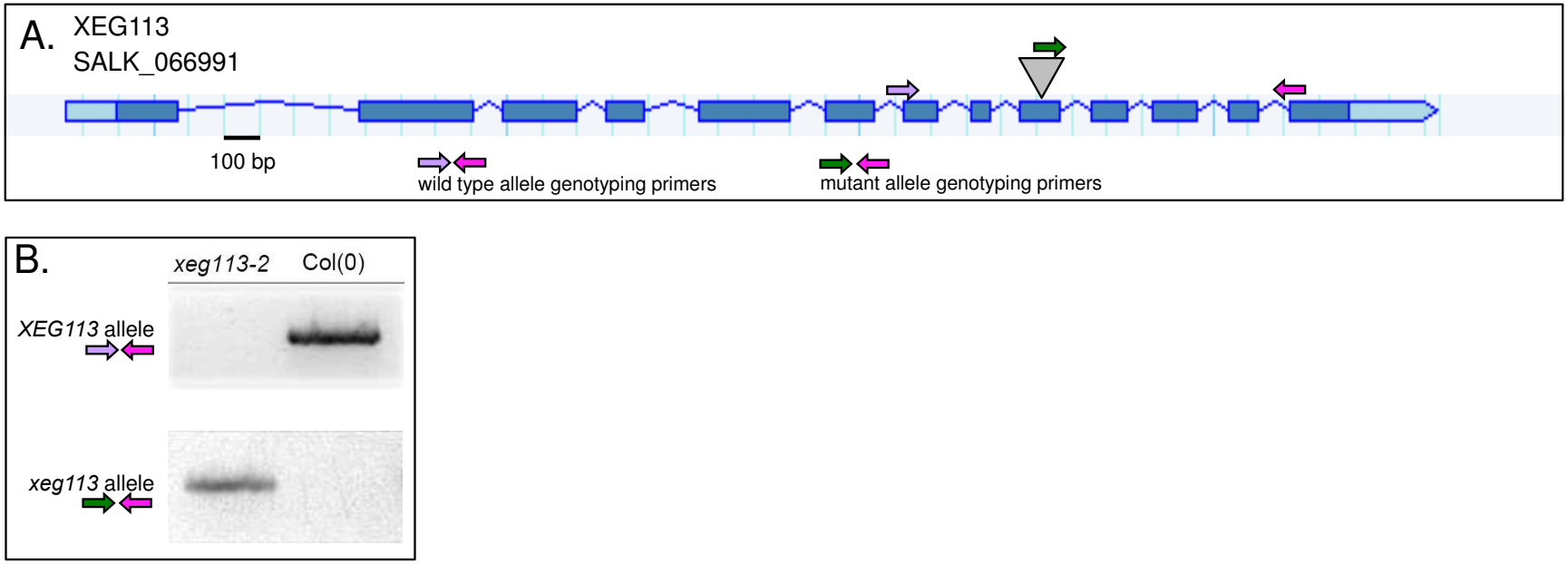

Figure 14. Genotyping of *xeg113-2*.
A. Gene model with locations of genotyping primers and T-DNA insert (locations of primers and T-DNA insert [grey triangle], shown above gene model, to scale). Gene model diagram from arabidopsis.org. Location of T-DNA insert determined by sequencing *xeg113-2* allele PCR product from B, below. B. Agarose gel electrophoresis results of genotyping PCR with wild-type (top row) and mutant (bottom row) allele-specific primers on mutant (left lane) and wild-type (right lane) plants.

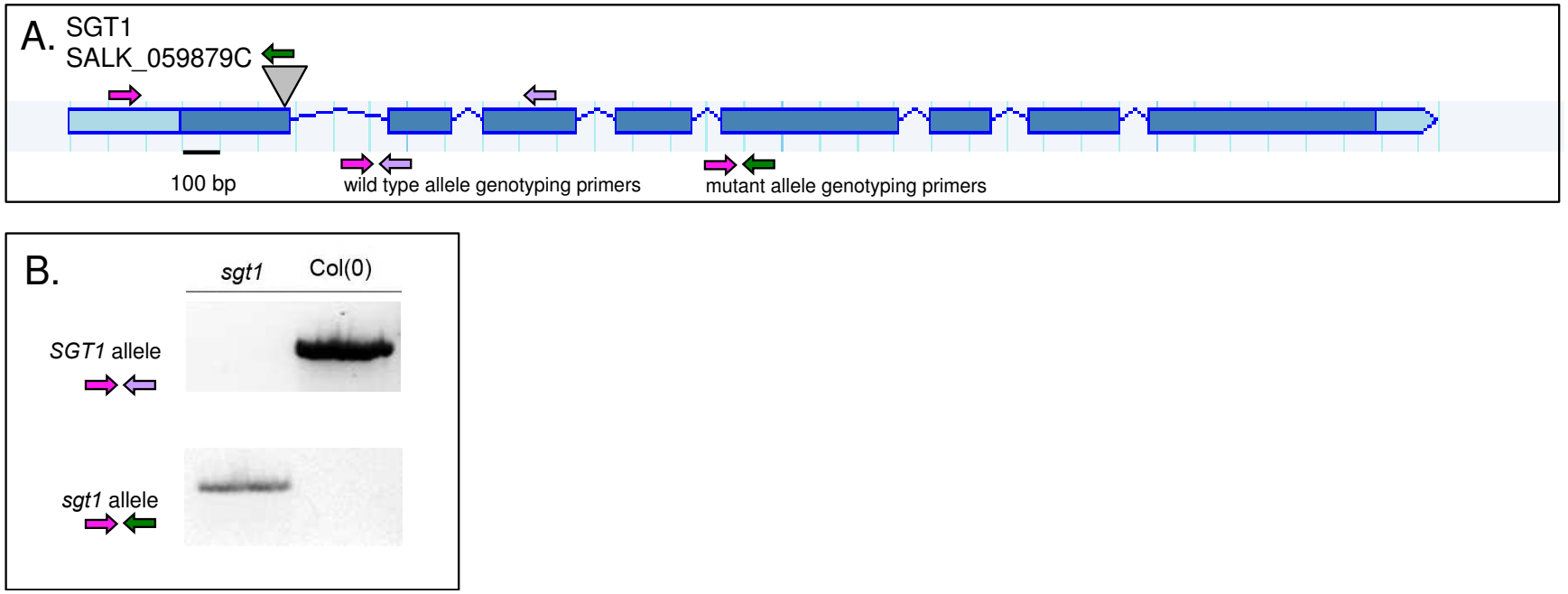

Figure 15. Genotyping of *sgt1*.
A. Gene model with locations of genotyping primers and T-DNA insert (locations of primers and T-DNA insert [grey triangle], shown above gene model, to scale). Gene model diagram from arabidopsis.org. Location of T-DNA insert determined by sequencing *sgt1* allele PCR product from B, below. B. Agarose gel electrophoresis results of genotyping PCR with wild-type (top row) and mutant (bottom row) allele-specific primers on mutant (left lane) and wild-type (right lane) plants.

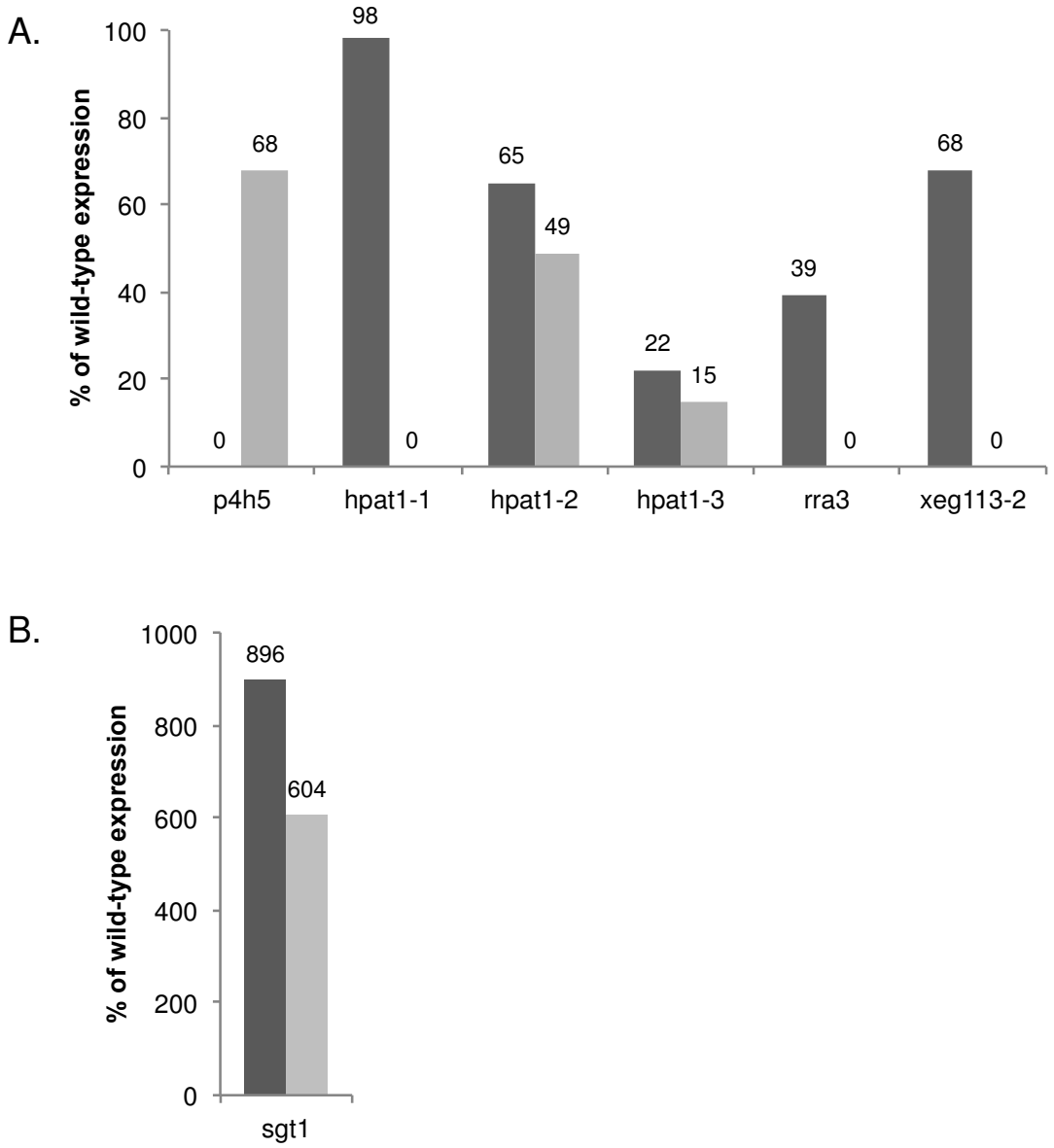

Figure 16. Mutant transcript abundance relative to wild-type.
Relative expression was calculated based on the ΔΔCt method, using amplified *ACT8* mRNA as the reference transcript. The Ct values for 5' primer set 1 (dark grey) and 3' primer set 2 (light grey) for each gene were normalized to the Ct value for Actin-8, and expression in the mutant relative to wild-type was calculated using the formula $2^{-\Delta\Delta Ct}$. A) % of wild-type expression in lines *p4h5*, *hpat1-1*, *hpat 1-2*, *hpat 1-3*, *rra3*, and *xeg113-2*. B) % of wild-type expression in *sgt1*.

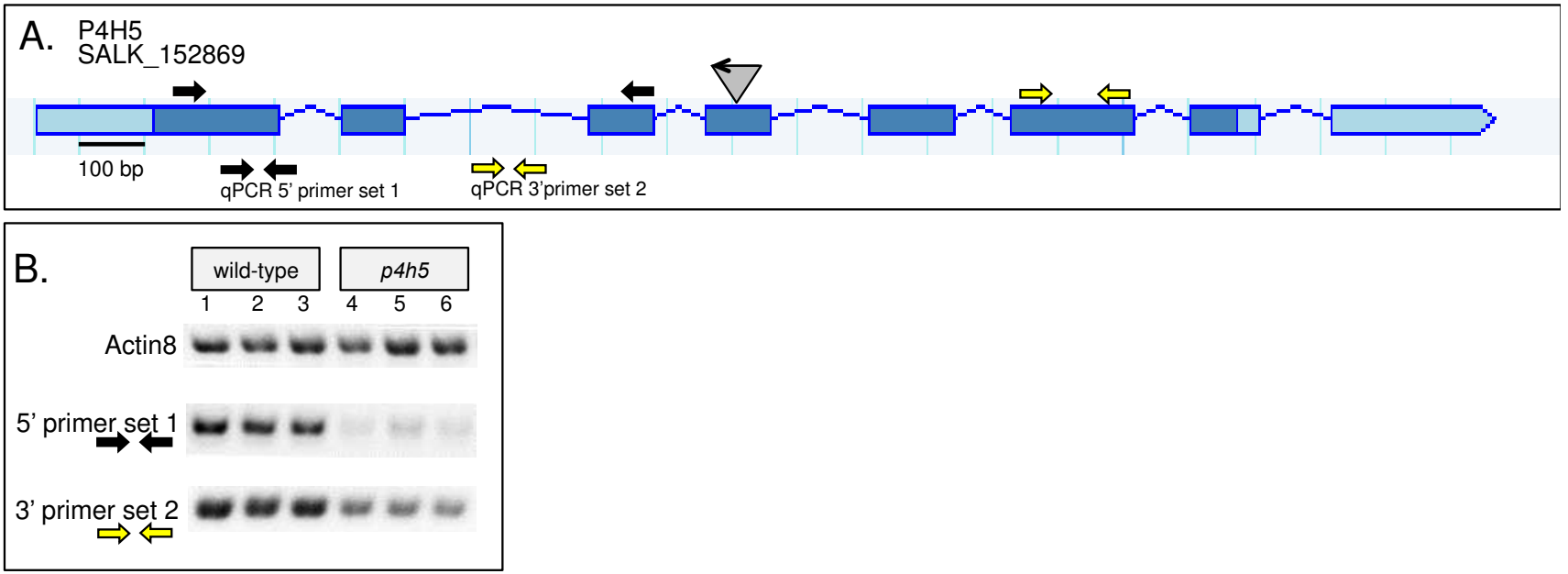

Figure 17. Expression analysis of *p4h5*.
A. Gene model with locations of qPCR primers and T-DNA insert (locations of primers and T-DNA insert [grey triangle, arrowhead points towards left border], shown above gene model, to scale). Primers were selected by NCBI Primer-BLAST to be specific only to the target template. B. Agarose gel electrophoresis results of qPCR on cDNA from each of three wild-type (lanes 1-3) and *p4h5* (lanes 4-6) plants, using reference gene (Actin 8) primers (top row), 5' primer set 1 (middle row), and 3' primer set 2 (bottom row).

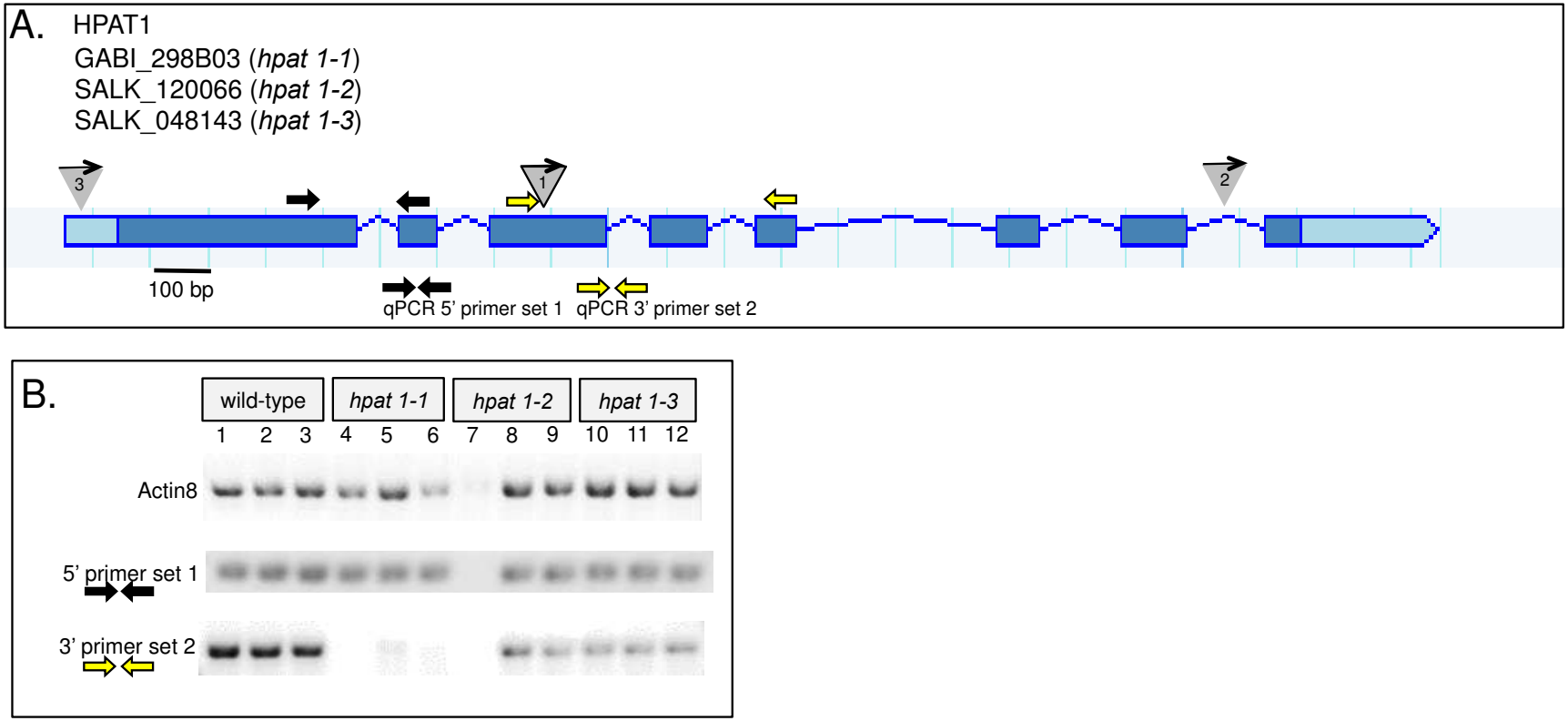

Figure 18. Expression analysis of *hpat1-1*, *hpat1-2*, and *hpat1-3*.
A. Gene model with locations of qPCR primers and T-DNA insert (locations of primers and T-DNA insert [grey triangle, arrowhead points towards left border], shown above gene model, to scale). Primers were selected by NCBI Primer-BLAST to be specific only to the target template. B. Agarose gel electrophoresis results of qPCR on cDNA from each of three wild-type (lanes 1-3), *hpat1-1* (lanes 4-6), *hpat1-2* (lanes 7-9), and *hpat1-3* (lanes 10-12) plants, using reference gene (Actin 8) primers (top row), 5' primer set 1 (middle row), and 3' primer set 2 (bottom row).

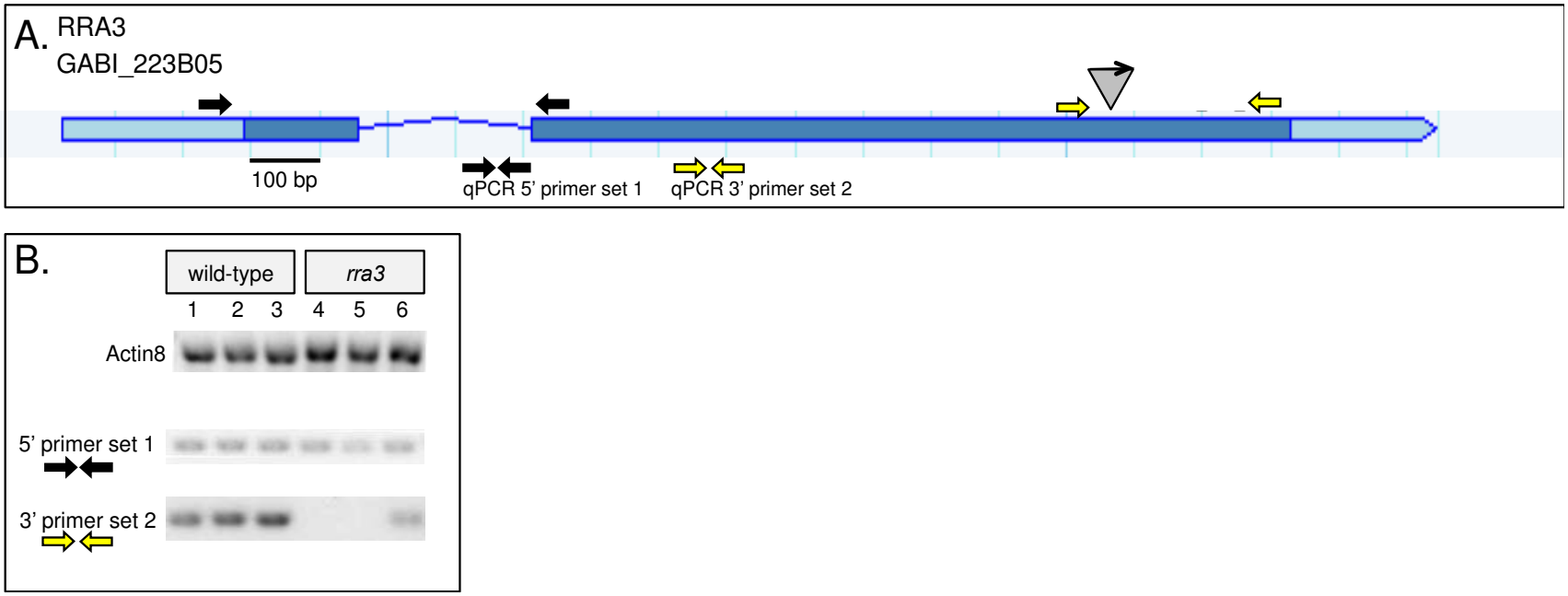

Figure 19. Expression analysis of *rra3*.
A. Gene model with locations of qPCR primers and T-DNA insert (locations of primers and T-DNA insert [grey triangle, arrowhead points towards left border], shown above gene model, to scale). Primers were selected by NCBI Primer-BLAST to be specific only to the target template. B. Agarose gel electrophoresis results of qPCR on cDNA from each of three wild-type (lanes 1-3) and *rra3* (lanes 4-6) plants, using reference gene (Actin 8) primers (top row), 5' primer set 1 (middle row), and 3' primer set 2 (bottom row).

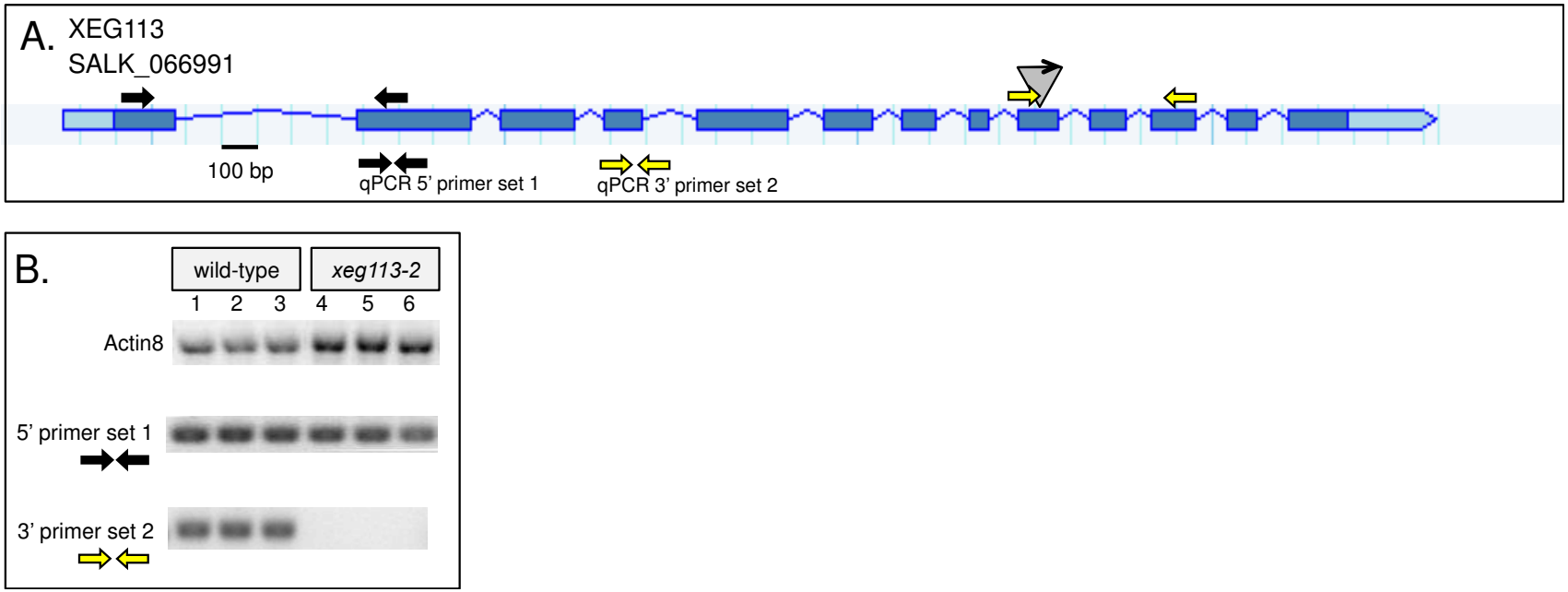

Figure 20. Expression analysis of *xeg113-2*.
A. Gene model with locations of qPCR primers and T-DNA insert (locations of primers and T-DNA insert [grey triangle, arrowhead points towards left border], shown above gene model, to scale). Primers were selected by NCBI Primer-BLAST to be specific only to the target template. B. Agarose gel electrophoresis results of qPCR on cDNA from each of three wild-type (lanes 1-3) and *xeg113-2* (lanes 4-6) plants, using reference gene (Actin 8) primers (top row), 5' primer set 1 (middle row), and 3' primer set 2 (bottom row).

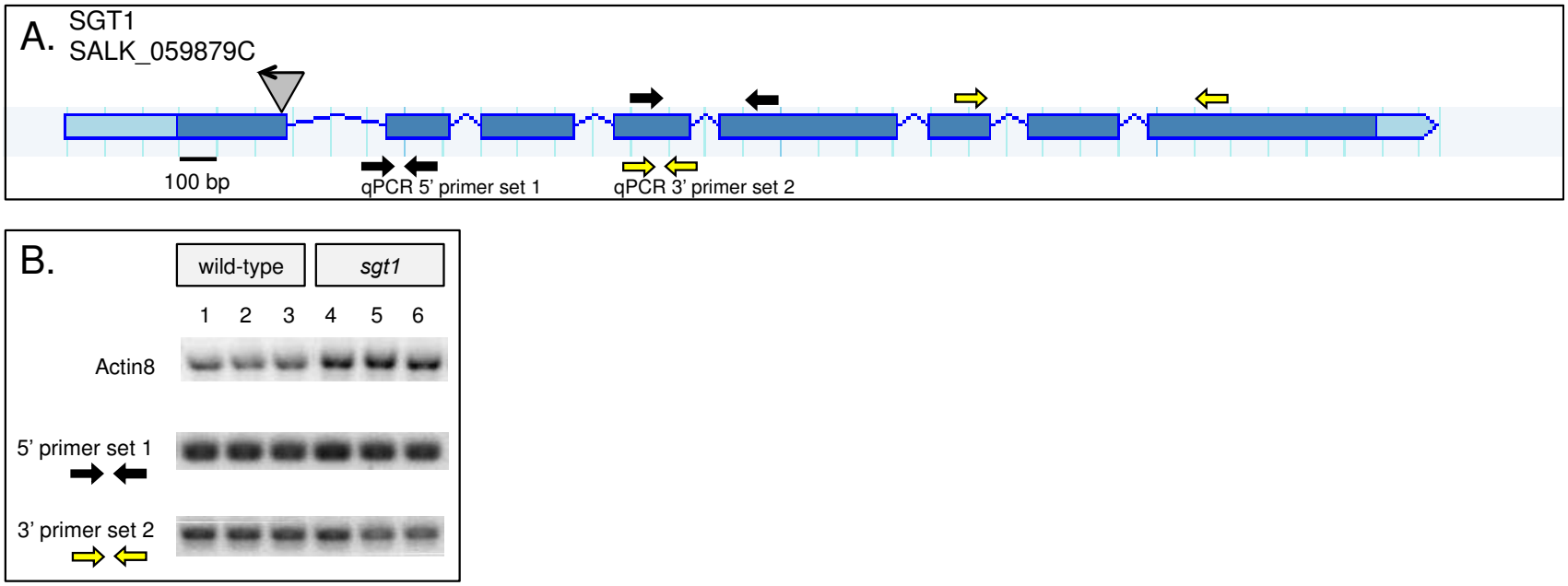

Figure 21. Expression analysis of *sgt1*.
A. Gene model with locations of qPCR primers and T-DNA insert (locations of primers and T-DNA insert [grey triangle, arrowhead points towards left border], shown above gene model, to scale). Primers were selected by NCBI Primer-BLAST to be specific only to the target template. B. Agarose gel electrophoresis results of qPCR on cDNA from each of three wild-type (lanes 1-3) and *sgt1* (lanes 4-6) plants, using reference gene (Actin 8) primers (top row), 5' primer set 1 (middle row), and 3' primer set 2 (bottom row).

Table 6. Summary of mutant Arabidopsis phenotypic data.
Data are averages ± standard error. Shaded boxes with bold values, $p < 0.05$ in comparison to wild-type (Student's T-test).

	Wild-type	*p4h5*	*hpat1-1*	*rra3*	*xeg113-2*	*sgt1*
days to bolting	21 ± 1.3	20.7 ± 0.8	20.3 ± 0.7	20.8 ± 1.0	20.3 ± 0.6	22.5 ± 1.0
days to flowering	23.3 ± 1.5	23.9 ± 1.2	23.3 ± 0.9	24.1 ± 1.6	22.3 ± 1.0	25.9 ± 1.4
days to harvest	43.9 ± 2.2	43.5 ± 1.7	43.2 ± 1.5	45.0 ± 1.7	42.6 ± 1.6	47.2 ± 1.3
# of siliques at harvest	7.3 ± 1.9	5.0 ± 0.9	5.8 ± 0.9	7.8 ± 1.3	10.7 ± 2.3	6.8 ± 1.0
stem thickness (mm)	1.4 ± 0.06	**1.2 ± 0.04**	1.3 ± 0.05	1.4 ± 0.08	1.4 ± 0.06	1.4 ± 0.08
inflorescence length (mm)	362 ± 10.7	**319 ± 7.5**	**315 ± 13.5**	365 ± 18.3	377 ± 7.8	339 ± 15.8
cauline stem length (mm)	128 ± 5.4	**114 ± 3.7**	**114 ± 3.8**	143 ± 6.0	**107 ± 5.3**	126 ± 8.0
# of bolts	4 ± 0.4	4.4 ± 0.4	3.9 ± 0.3	4.2 ± 0.3	4.3 ± 0.3	4.6 ± 0.3
# of branches off main bolt	3.7 ± 0.2	**2.9 ± 0.2**	3.2 ± 0.1	3.5 ± 0.2	3.2 ± 0.2	3.9 ± 0.3
inflorescence wet weight (g)	2.54 ± 0.15	**1.91 ± 0.12**	**1.97 ± 0.14**	2.40 ± 0.21	2.40 ± 0.20	2.73 ± 0.22
cauline stem wet weight (g)	0.90 ± 0.05	**0.68 ± 0.04**	**0.68 ± 0.06**	0.91 ± 0.09	0.82 ± 0.07	0.94 ± 0.08
cauline stem dry weight (mg)	93 ± 7	**64 ± 5**	**64 ± 5**	87 ± 9	82 ± 8	90 ± 8
average glucose release (%/mg dry weight)	9.2 ± 0.04	9.4 ± 0.03	9.3 ± 0.02	9.0 ± 0.03	8.9 ± 0.02	9.2 ± 0.02
average hydroxyproline content (%/mg dry weight)	0.048 ± 0.001	0.050 ± 0.001	0.054 ± 0.001	0.044 ± 0.001	0.044 ± 0.001	0.046 ± 0.001
number of plants measured	9	10	11	11	12	11

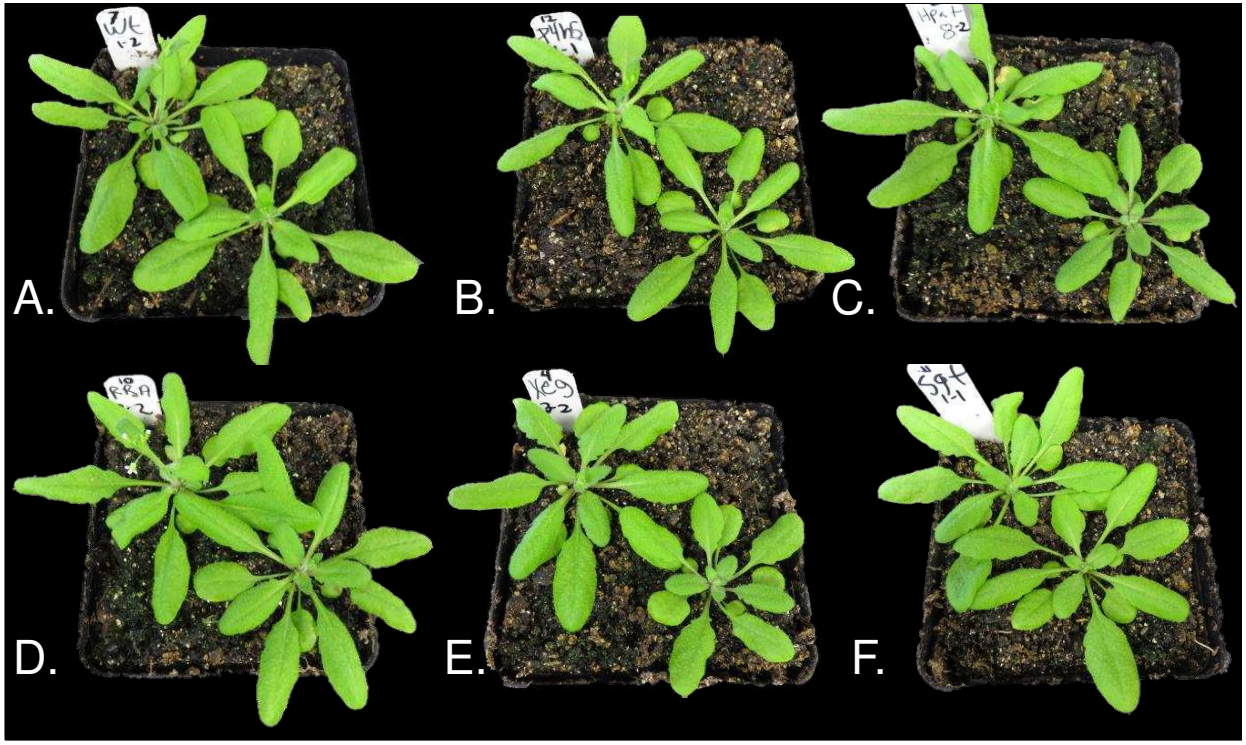

Figure 22. Five-week-old Arabidopsis plants.
A, Wild-type; B, *p4h5*; C, *hpat1*; D, *rra3*; E, *xeg113-2*; F, *sgt1*.

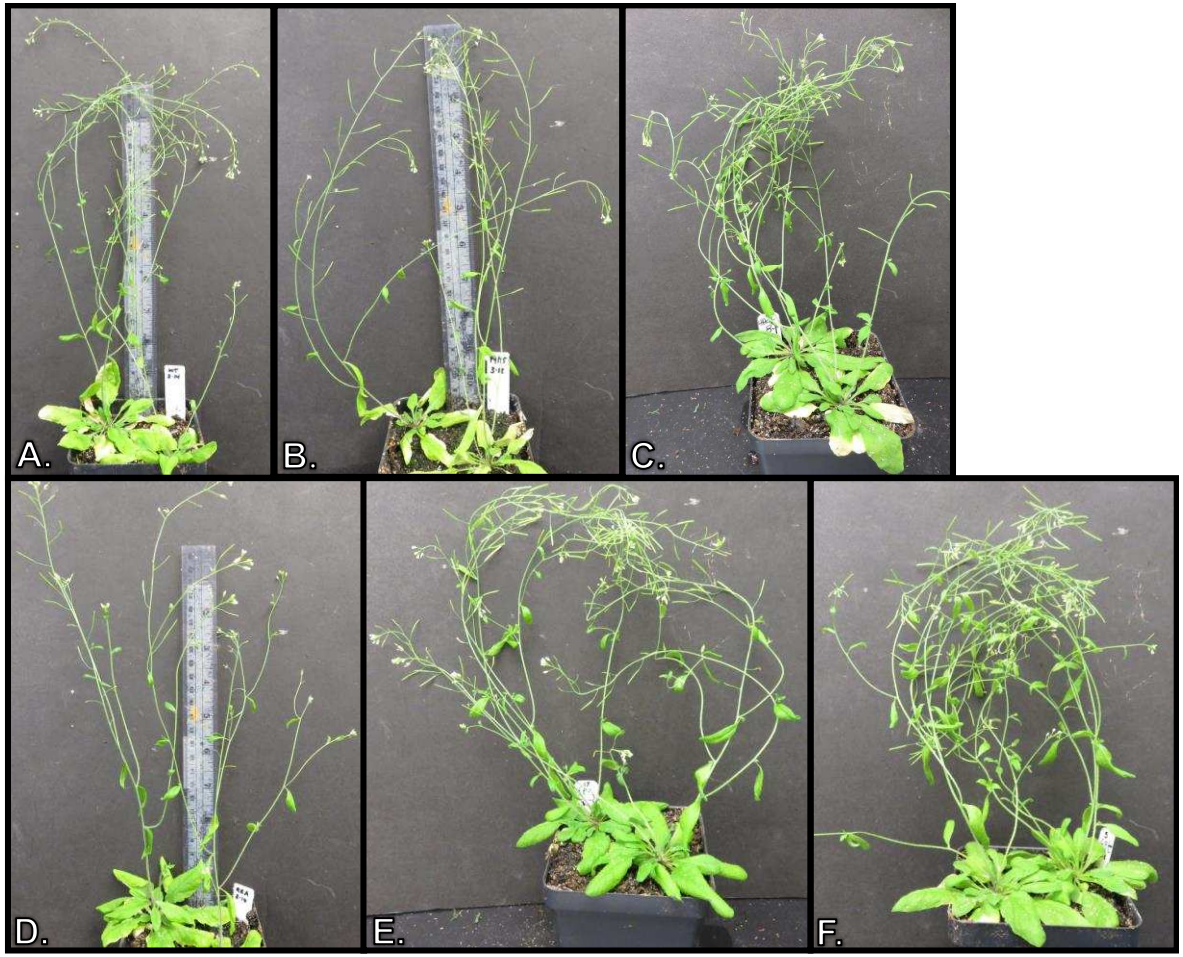

Figure 23. Seven-week-old Arabidopsis plants.
A, Wild-type; B, *p4h5*; C, *hpat1*; D, *rra3*; E, *xeg113-2*; F, *sgt1*.

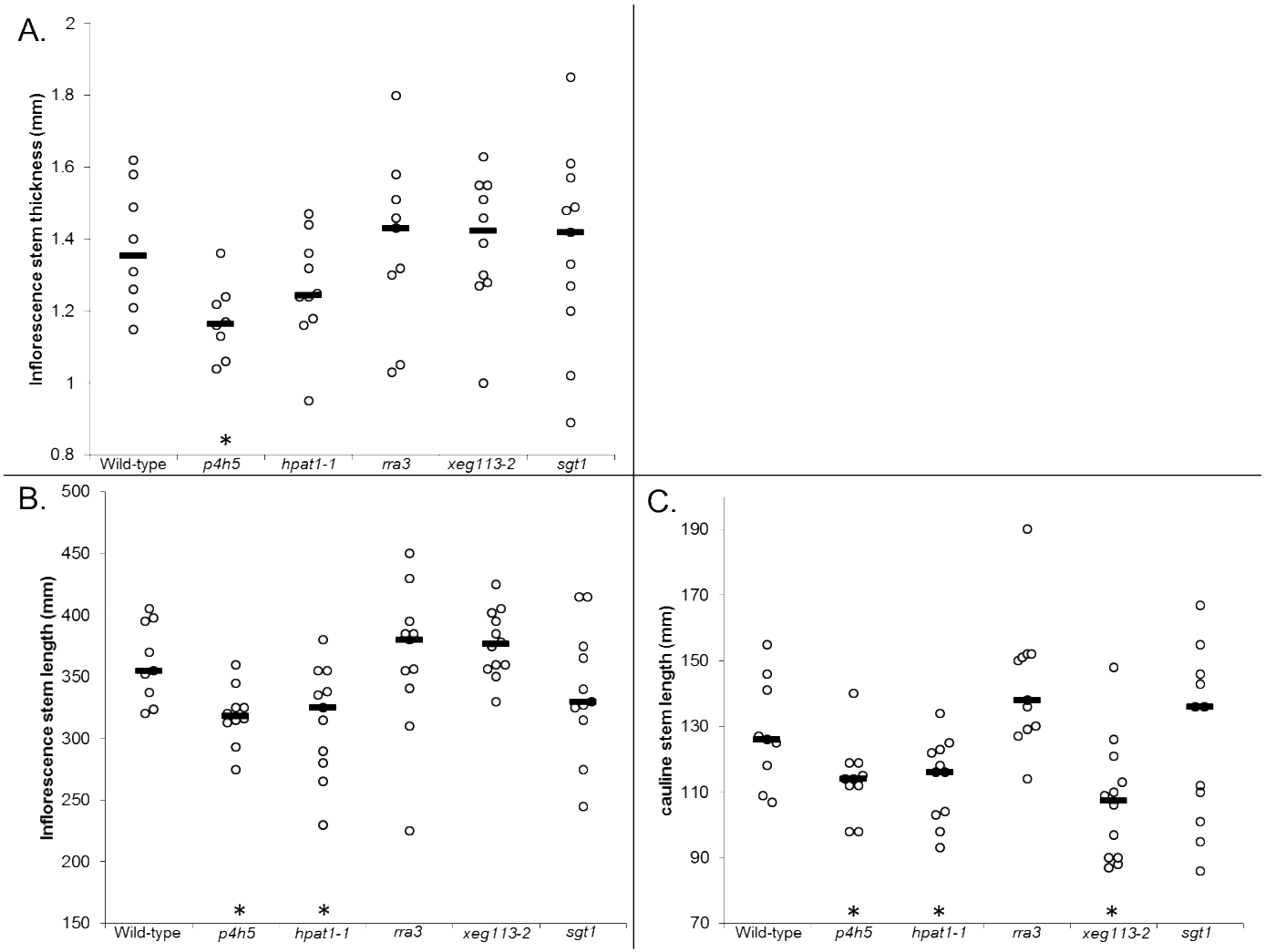

Figure 24. Stem phenotypes of wild-type and mutant Arabidopsis.
Each circle represents the measurement of a single plant; black bars, median value for that line; asterisks, lines with significant differences from wild-type ($p < 0.05$, Student's T-test). A, inflorescence stem thickness, measured at base of rosette. B, inflorescence stem length measured from base of rosette to apical meristem. C, cauline stem length, measured from base of rosette to first silique.

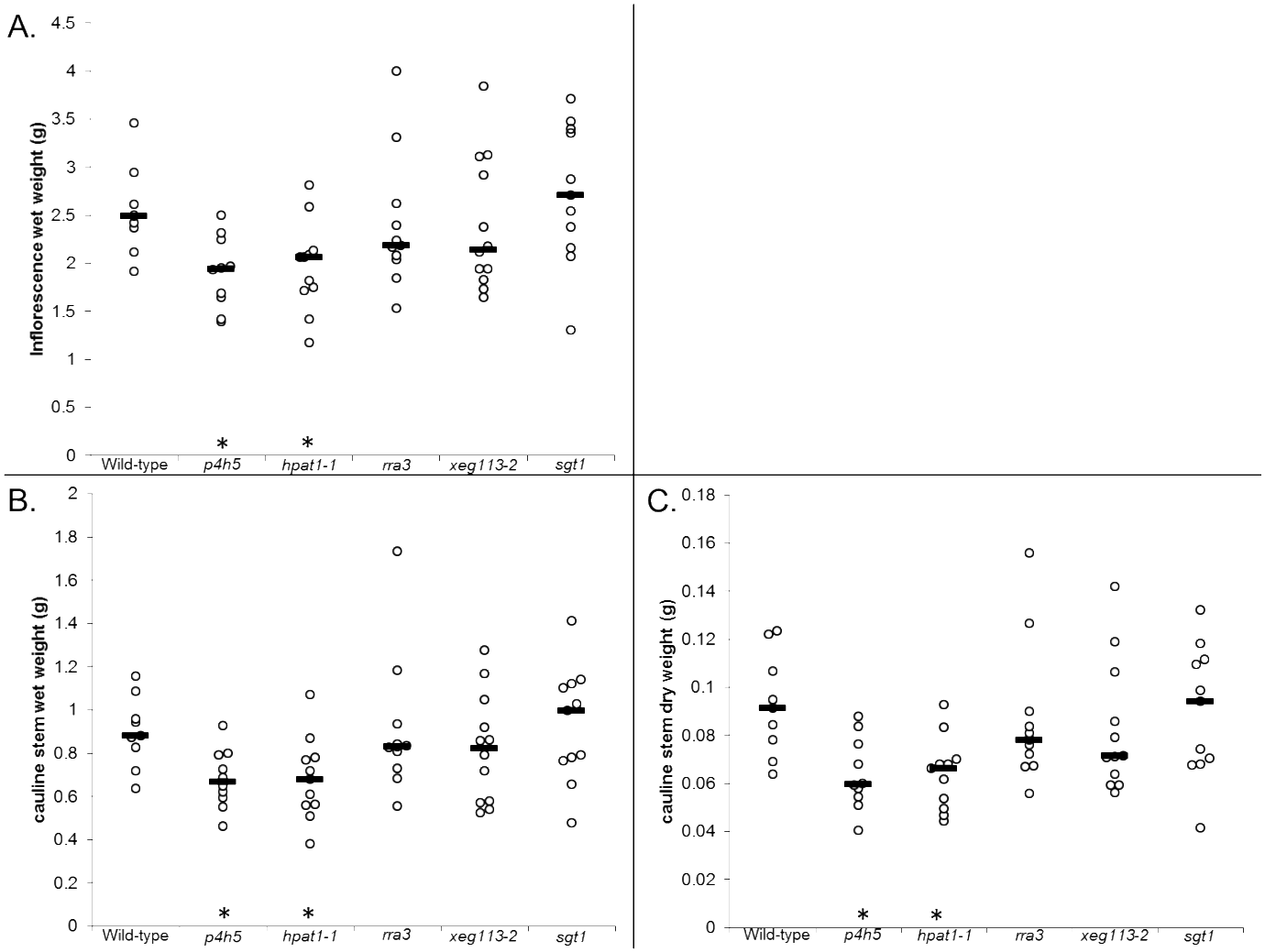

Figure 25. Biomass phenotypes of wild-type and mutant Arabidopsis.
Each circle represents the measurement of a single plant; black bars, median value for that line; asterisks, lines with significant differences from wild-type ($p < 0.05$, Student's T-test). A, inflorescence wet weight including leaves, flowers, siliques, and the entire inflorescence stem. B, wet weight of the cauline stem stripped of leaves. C, dry weight of the cauline stem measured in B.

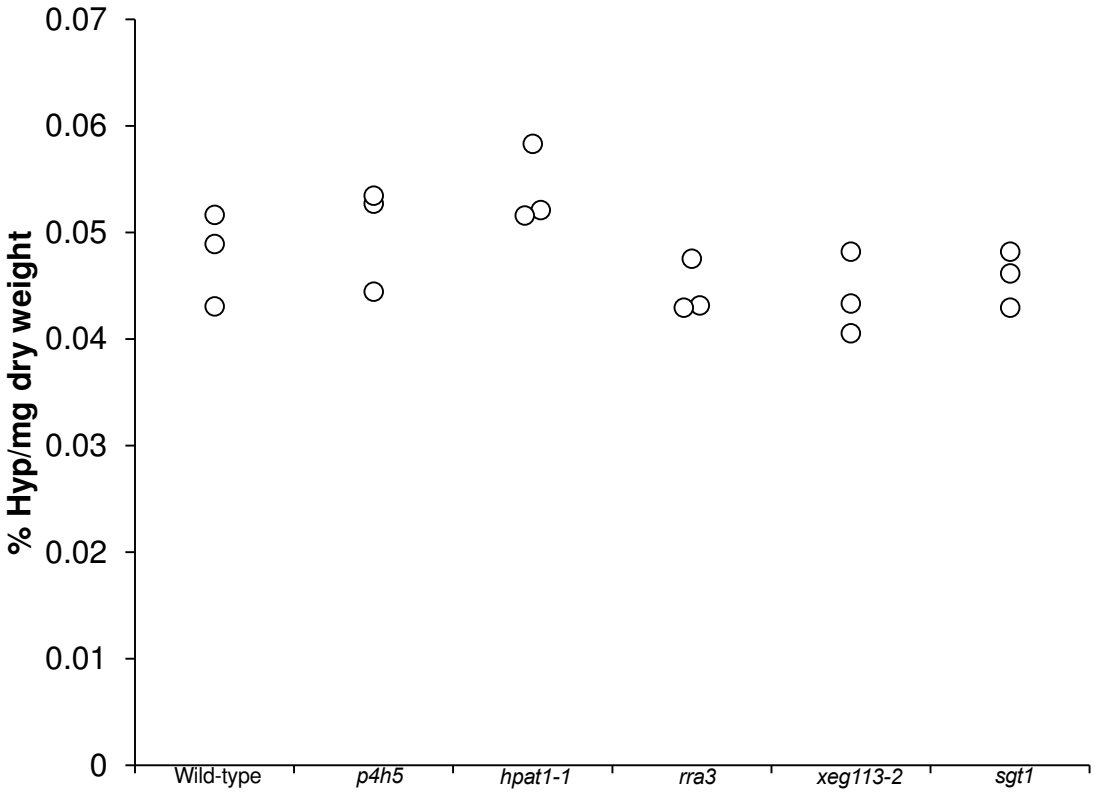

Figure 26. Hydroxyproline content of wild-type and mutant Arabidopsis cauline stems.
Each circle represents an average of three measurements of a single pool of three or four plants.

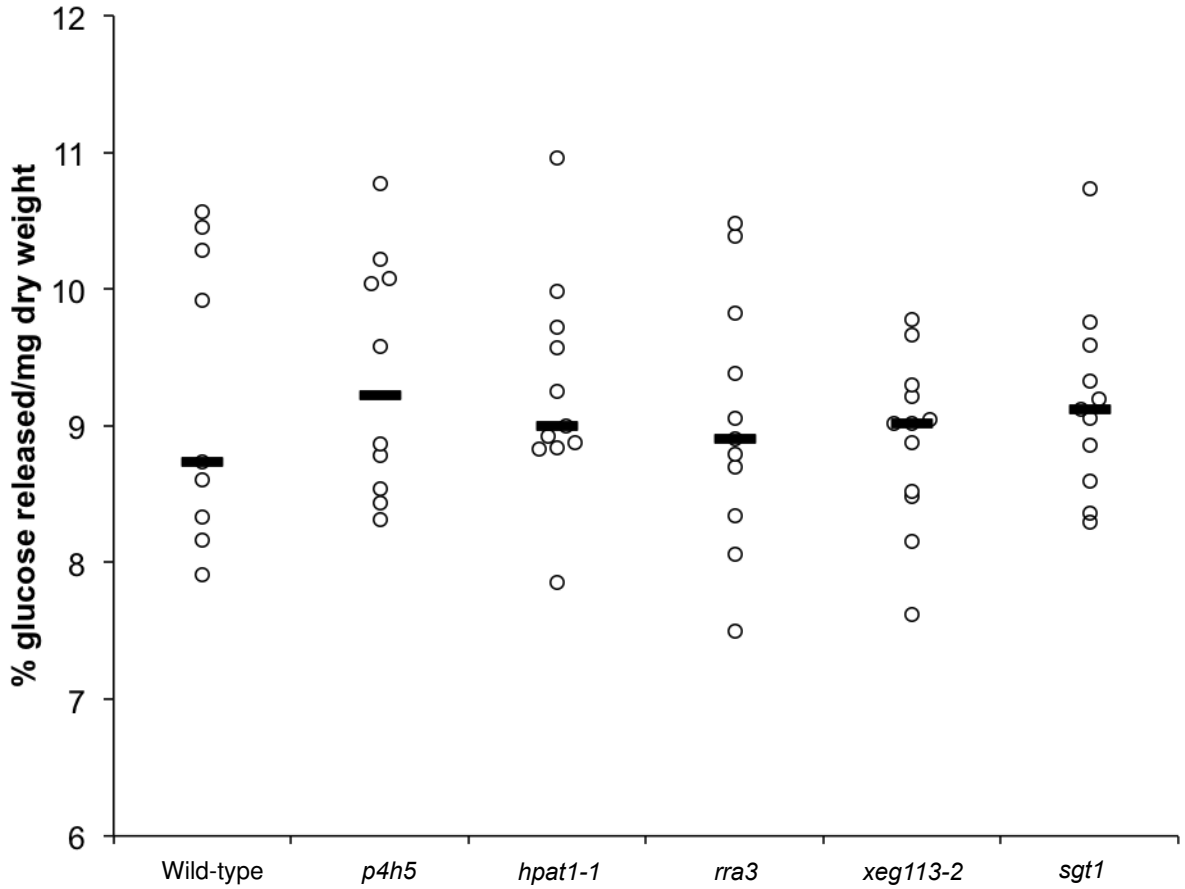

Figure 27. Glucose release from wild-type and mutant Arabidopsis cauline stems. Each circle represents an average of three measurements from a single plant. Black bars represent the median value for that line.

CHAPTER FOUR: FLUORESCENT REPORTER SYSTEM FOR PLANT CELL WALL PROTEINS IN ARABIDOPSIS

SYNOPSIS

In order to examine the association between extensin and the cell wall more closely, I created two synthetic gene constructs that encode either the complete extensin domain of *SlLRX1* (sec-tdT-long) or the C-terminal 20 amino acids of *SlLRX1* (sec-tdT-short). These extensin fragments are encoded fused to the fluorescent reporter protein tdTomato (tdT), and are directed to the cell wall with a secretory signal sequence. Transgenic Arabidopsis plants expressing these tdT-extensin constructs were analyzed for expression levels by fluorescence microscopy and Western blotting with anti-RFP antibody. The Western blot revealed likely post-translational modification of sec-tdT-long with hydroxylation and arabinosylation, as the strongest bands detected were much larger than predicted. Fluorescence in cauline stem tissue (the inflorescence stem from the base of the rosette to the first silique) of transgenic plants was observed after various treatments to denature or remove the fluorescent signal (including 1 M NaCl, 8 M LiCl, preparation of alchol-insoluble-residue [AIR], 1% SDS, CTEC2 cellulase digestion, and proteinase K digestion). In lines expressing sec-tdT-short, fluorescence was retained after 1 M NaCl and 1% SDS treatments, as well as AIR preparations. In lines expressing sec-tdT-long, fluorescence was additionally retained after 8 M LiCl treatment. These results indicate a strong interaction of tdT-extensin proteins with the cell wall, and in the case of sec-tdT-long, potentially covalent interactions. Developmental phenotypes were measured in several independently transformed lines expressing each tdT-extensin construct, but the differences observed were small, and confounded by the small sample size of the experiment. Hydroxyproline content was also measured in these same lines, and substantially more hydroxyproline was measured in lines

strongly expressing sec-tdT-long than in wild-type (long-2, 2.3 times more; long-7, 3 times more). Finally, glucose release from cauline stem biomass after pretreatment and cellulase digestion was measured, as overexpression and cross-linking of extensins may reduce biomass digestibility. No strong differences were observed, although, again, the small sample size confounds the analysis.

INTRODUCTION

In Chapter 3, I described my investigations into possible biofuel-related phenotypes in Arabidopsis mutants in the genes encoding enzymes that perform extensin post-translational modification, which should have reduced extensin crosslinking in their cell walls. Similarly, I wanted to investigate the biofuel-related phenotypes of transgenic Arabidopsis plants overexpressing extensin genes. In order to monitor extensin cross-linking directly, these plants also express, in tandem with the extensin gene, a fluorescent reporter protein tag (tdTomato) (Figure 28, Figure 29). Overexpression and increased cross-linking of extensin protein in plant cell walls could increase biomass recalcitrance and therefore decrease digestibility and glucose yield.

The extensin protein I chose for this experiment is part of an extensin chimera, SlLRX1[128]. This gene is expressed in tomato pollen, and encodes two functional domains (a leucine-rich-repeat domain and an extensin domain) separated by a flexible cysteine- and glycine-rich linker region. The homologous proteins in maize, mPEX1 and mPEX2, are tightly associated with the pollen tube cell wall[81,129]. Presumably the extensin domain is responsible for this association, which means that the extensin domain of leucine-rich-repeat extensin chimeras functions independently in a multi-domain protein. Therefore, substitution of the leucine-rich-

repeat domain of SlLRX1 with a fluorescent reporter protein should not affect the cross-linking abilities of the extensin domain.

SlLRX1 is also an interesting example of an extensin protein, because its tyrosine content is unlike that of other extensins. Most extensins contain many instances of the YXY motif, which has been implicated in extensin cross-linking through several *in vitro* experiments. Extensin peroxidases have been identified which form isodityrosine and di-isodityrosine from tyrosines in YXY motifs specifically, in the presence of hydrogen peroxide; this reaction *in vitro* can occur both intra- and inter-molecularly[13,52,130]. Evidence for *in vivo* cross-links is scarce, due to the difficulties in analyzing cross-links in insoluble extensins from cell walls without breaking them, but intermolecular isodityrosine cross-links have been identified in soybean suspension-cultured cells[17]. It has been tacitly assumed that tyrosine must be present in the form of the YXY motif for these cross-links to be assembled. In *SlLRX1*, however, only two tyrosines are encoded in the entire extensin domain, and they are separated by 10 amino acids, which include a single SP_4 motif. One tyrosine directly precedes the stop codon, while the other is surrounded by non-phenolic amino acids, which are unlikely to participate in peroxidase-mediated cross-linking reactions. Prior analysis of SlLRX1 homologs mPEX1 and mPEX2 suggests that these proteins may be covalently bound to the cell wall, as treatment with reagents that disrupt non-covalent interactions (such as LiCl and SDS) did not reduce the SlLRX1 protein content of the pollen tube cell wall[129]. Therefore, tyrosine in any context may permit cross-linking, or perhaps some other reaction is responsible for the strength of the association of SlLRX1 with the cell wall.

To test the sequence requirements for cell wall association, I designed two tdT-fusion protein constructs (Figure 28, Figure 29). One, called sec-tdT-short, encodes the C-terminal 20 amino acids of *SlLRX1*, which include a single SP_4 repeat as well as both tyrosines. The other,

called sec-tdT-long, encodes the 105 amino acids of the cysteine- and glycine-rich linker region as well as the entire 290 amino acids of the extensin domain of *SlLRX1*. The fluorescent protein tag permits analysis of cells expressing the constructs with fluorescence microscopy, Western blotting, or more conventional immunohistochemistry. The reporter protein construct can also be distinguished from a cell's native extensins by virtue of the fluorescent reporter protein tag.

The tag I chose for this experiment, tdTomato, has several remarkable features. First, like all fluorescent reporter proteins, it self-assembles from a linear polypeptide chain into a complex mature structure of beta-barrels that shield the chromophore, so no co-factors are required to observe fluorescence, and fluorescence can readily be observed *in vivo*. It is tolerant of being expressed as part of a fusion protein, making it suitable as a tag[131,132]. It is also quite robust and will maintain fluorescence in a variety of normally denaturing conditions. For example, as opposed to the more commonly used green fluorescent protein (GFP) or enhanced GFP, which are pH-sensitive and will not reliably fluoresce under acidic conditions such as those present in the cell wall, tdTomato fluorescence is not affected by pH values from 5 – 12[133]. It can therefore be imaged directly in cell walls without any pH adjustment step.

Overexpression of extensin proteins and of artificial extensin domains has been reported previously[34,124,134,135]. Most experiments were performed in *Nicotiana tabacum* plants, in which no gross phenotypic changes resulting from either extensin under- or over-expression were detected[134,135]. Arabidopsis *EXTENSIN-1* (*EXT-1*) was overexpressed in stably transformed Arabidopsis plants under the control of the Cauliflower Mosaic Virus (CaMV) 35S promoter[124]. A detailed analysis of macroscopic developmental phenotypes was performed, as described below, but effects at the level of the cell wall were not analyzed. Overexpression of synthetic extensin domains, which were composed of many repeats of "hallmark" extensin motifs tagged

with eGFP, was investigated both in suspension cultured BY-2 tobacco cells and in stably transformed Arabidopsis plants. The eGFP tag of the synthetic protein was used to observe any microscopic phenotypes associated with synthetic extensin protein expression, as well as to purify and analyze post-translational modifications of the synthetic protein, as described below. However, macroscopic phenotypes were not measured.

In this study, I observed Arabidopsis plants stably transformed with the tdT constructs I created (control tdT constructs cyto-tdT and sec-tdT, and experimental tdT-extensin constructs sec-tdT-short and sec-tdT-long). I examined cellular phenotypes by means of the fluorescent reporter protein tag, such as levels of transgene expression and association of the fluorescent reporter protein construct with the cell wall after various treatments. I hypothesized that sec-tdT-long would have a very tight association with the cell wall due to the presence of the complete extensin domain, while the association of sec-tdT-short would depend on whether a single SP_4 motif or the associated two tyrosines were sufficient for cell wall attachment. Gross phenotypes relevant for biofuels production were also measured, including characteristics such as stem thickness and length, biomass yield, hydroxyproline content, and glucose yield after pretreatment and digestion. I hypothesized that plants overexpressing sec-tdT-short and sec-tdT-long would have increased extensin cross-linking in their cell walls, as shown by increased hydroxyproline content, leading to smaller and less digestible plants.

MATERIALS AND METHODS

Plasmid construction

The binary vector pART27[136] encoding tdTomato[131,132] and secreted tdTomato (using the secretory signal sequence from tomato polygalacturonase[137,138]), under the control of the CaMV 35S promoter, were kind gifts of Dr. Jocelyn Rose, Cornell University[139] (named pART-tdT and

pART-sec-tdT). The entire reporter protein construct was excised from pART-tdT and pART-sec-tdT by digesting with *Not*I. The resulting 3573 and 3651 bp fragments were gel-purified with the QIAquick Gel Extraction Kit (28704, QIAGEN; used for all gel purifications described below) and ligated using T4 DNA Ligase (M0202, New England Biolabs; used for all ligations described below) into dephosphorylated pBlueScript KS+ that had also been digested with *Not*I and gel-purified to make pBS-tdT and pBS-sec-tdT. The plasmids were confirmed by sequencing with primers 35S-FP (5'-CCTTCGCAAGACCCTTCCTC-3') and OCS-RP (5'-CGTGCACAACAGAATTGAAAGC-3'). All sequencing was performed by the Proteomics and Metabolomics Facility at CSU on an ABI 3130xL Genetic Analyzer.

The short and long extensin domain sequences from *SlPEX1* (NCBI accession AF159296[128]) were codon-optimized by GenScript for expression in *Escherichia coli* and yeast‡ (Figure 29, Figure 30), synthesized, and cloned by GenScript into pUC57 to make pUC57-less and pUC57-high, respectively. The long extensin domain was then removed from pUC57-high and subcloned into pUC57-less by digestion of both pUC57-high and pUC57-less with *Pst*I and *Nsi*I followed by gel purification of the 1119 bp band from pUC57-high and the 4698 bp band from pUC57-less, and ligation of the two gel-purified fragments to make pUC57-full length. The plasmid was confirmed by sequencing with the primers tdT-seq-1900-FP (5'-CCCGTTCAATTGCCTGGT-3') and OCS-RP. The plasmid pBS-sec-tdT-short was made by digesting pUC57-less and pBS-sec-tdT with *Nde*I and *SgrA*I, followed by gel purification of the 1142 bp band from pUC57-less and the 5545 bp band from pBS-sec-tdT, and ligation of the two gel-purified fragments. The plasmid pBS-sec-tdT-long was made by digestion of pUC57-full

‡ Codon optimization was performed in order to express protein for *in vitro* studies; codon-optimized protein was predicted to also function in plants.

length and pBS-sec-tdT with *Nde*I and *SgrA*I, followed by gel purification of the 2243 bp band from pUC57-full length and the 5545 bp band from pBS-sec-tdT, and ligation of the two gel-purified fragments. pBS-sec-tdT-short plasmids were confirmed by sequencing with the 35S-FP and OCS-RP primers. pBS-sec-tdT-long plasmids were confirmed by sequencing with the 35S-FP, tdT-seq-1900-FP, and OCS-RP primers. All pBS plasmids were also confirmed by digestion with *Not*I and examination of the resulting insert and backbone fragment lengths by agarose gel electrophoresis.

Binary vectors for plant transformation were made by gel purifying the *Not*I insert fragments from the pBS-sec-tdT-short and pBS-sec-tdT-long plasmids and ligating them with pART27-sec-tdT backbone that had been digested with *Not*I, gel purified, and dephosphorylated, to obtain pART27-sec-tdT-short and pART27-sec-tdT-long. These two plasmids were confirmed by sequencing with the 35S-FP, tdT-seq-1900-FP, and OCS-RP primers.

A bacterial expression vector for tdTomato expression was made by amplifying the tdTomato sequence from pUC57-less using the following reaction mixture: 0.4 µL of Platinum Pfx polymerase (11708013, Life Technologies), 5 µL of 10x Pfx buffer, 0.3 mM DNTPs, 1 mM $MgSO_4$, 0.3 mM primers EcoRV (5'-ACGTGGACACTAAACTCGATATCACCTCCCACAACG-3') and PstI (5'-CTGCTGCTGCAGCTGCTGCTTACTTGTAC-3'), which adds an in-frame stop codon (underlined) to the end of the tdTomato coding sequence, 50 ng of pUC57-less, and H_2O to 50 µL. The template was amplified in a BioRad C1000 Thermal Cycler with the following program: 94 ºC for 3 minutes, followed by 34 cycles of 94 ºC for 15 seconds (denaturation), 60 ºC for 30 seconds (annealing), and 68 ºC for 60 seconds (extension), then a final extension at 68 ºC for 4 minutes, and holding at 4 ºC. The 844 bp PCR product was purified with the QIAquick PCR

Purification kit (28104, QIAGEN) and digested with EcoR*V* and Pst*I*, then gel purified and religated into the 3852 bp backbone of pUC57-less that had also been digested with EcoR*V* and Pst*I* and gel purified, to make a plasmid encoding the tdTomato coding sequence without any promoter (pUC57-tdT). This 1450 bp sequence was digested from pUC57-tdT with EcoR*I* and Pst*I*, gel purified, and ligated into pET28a that had been similarly digested and gel-purified to make pET28a-tdT. The plasmid was confirmed by sequencing with the primers T7 (5'-TAATACGACTCACTATAGGG-3') and T7-term (5'-GCTAGTTATTGCTCAGCGG-3') to ensure the start codon of tdTomato was in frame with the coding sequence for the His-tags included in the pET28 backbone.

Transient transformation

Transient expression assays to assess the subcellular localization of cyto-tdT, sec-tdT, sec-tdT-short, and sec-tdT-long in white onion epidermal cells were performed with particle bombardment. DNA for bombardment was prepared by transforming *E. coli* strain dH5α with each pBS-tdT plasmid (pBS-tdT, pBS-sec-tdT, pBS-sec-tdT-short, and pBS-sec-tdT-long) by electroporation, culturing positive colonies, and extracting plasmid DNA using the QIAquick Spin Midiprep kit (12143, QIAgen). 2 μg of plasmid DNA from each of the four plasmids were adsorbed onto separate aliquots of W10 tungsten particles, according to the manufacturer's instructions (Bio-Rad). The abaxial epidermis of a white onion bulb leaf was removed by hand and placed on a half-strength MS plate (2.16 g/L Murashige and Skoog [MS] salts [M524, *Phyto*Technology Laboratories], 1% sucrose, 0.5% Agargel [A3301, Sigma], pH 5.7), with the inner side oriented upwards. For each bombardment, 1.2 mg of DNA-coated tungsten particles were placed on the macrocarrier. Onion epidermal cells were bombarded using the PDS-1000/He Biolistic Particle Delivery System (Bio-Rad) at 1100 psi under a vacuum of 20 kPa at a distance

of 9 cm from the stopping screen. Following bombardment, the onion peels incubated in the same petri dish for 63 h at room temperature in the dark before observation. Transformed onion peels were imaged before and after plasmolysis in 1 M NaCl using a Leica 5500 microscope (Leica Microsystems) running IPLab version 4 software (BD Biosciences) with a C4742-95 camera (Hamamatsu Photonics). Fluorescence was observed with a TRITC filter cube (570 nm long pass; Leica Microsystems). Fluorescence images were taken using a 250 ms exposure, with white and black points set to their respective maxima, to be able to compare fluorescence between cells.

Arabidopsis transformation

Agrobacterium tumefaciens, strain LBA4404, was transformed with each of the four pART27-tdT plasmids (pART27-tdT, pART27-sec-tdT, pART27-sec-tdT-short, pART27-sec-tdT-long) using the standard freeze-thaw method[140]. Correct transformation, including maintenance of intact border sequences, was confirmed by extracting the plasmid from transformed Agrobacterium, inserting the plasmid into *E. coli* strain dH5α, extracting the plasmid from dH5α with the QIAprep Spin Miniprep kit (27104, QIAgen), and sequencing with primers 35S-FP, tdT-seq-1900-FP, OCS-RP, OCS-FP (5'-GATAGAGCGCCACAATAACAAAC-3'), 35S-RP (5'-GACCAGAGTGTCGTGCTCCA-3'), and 332-RP (5'-CAGCAGGATGCTTAACGTATG-3'). Wild-type Arabidopsis plants (ecotype Columbia-0, Col-0) were transformed with each transformed Agrobacterium line by the simple floral dip method[141]. Transgenic plants were selected on half-strength MS plates containing 50 μg/mL of kanamycin. Selected plants were transferred to soil and allowed to self-pollinate, and several independent lines for each transgene were bred to homozygosity over several generations, determined by all progeny (at least 100 plants/line) demonstrating tdTomato

fluorescence and resistance to kanamycin. Subsequent experiments were performed on T3 generation plants.

Purification of tdTomato protein

E. coli strain BL21 (DE3) was transformed with the pET28a plasmid for inducible expression in bacteria containing the coding sequence for tdTomato (pET28a-tdT). A 500 mL LB culture containing 50 μg/mL kanamycin was grown to mid-log phase (OD_{600} = 0.6). Expression of tdTomato was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG, FERR0391, Fisher) to 1 mM final concentration. Cells were harvested four hours after the start of induction by centrifugation at 6000 g for 6 minutes. The cell pellet was washed by resuspension in STE buffer (100 mM NaCl, 10 mM Tris pH 8.0, 1 mM EDTA) followed by centrifugation at 6000 g for 6 minutes. The cell pellet was frozen at -80 ºC, then thawed on ice and resuspended in lysis buffer (100 mM NaCl, 20 mM Tris pH 8.0, 20 mM βME, 5 mM imidazole, 1 mM PMSF, protease inhibitor mix). After resuspension, Thesit (88315, Sigma) was added to a final concentration of 1%. Cells were lysed by the addition of 5 mg lysozyme, followed by incubation at room temperature until the solution became viscous. The salt concentration was then adjusted to 0.5 M NaCl and the lysate was clarified by centrifugation at 27,000 g for 15 minutes. Clarified lysate was applied to a TALON Metal Affinity Resin column (Clontech) pre-equilibrated with wash buffer (500 mM NaCl, 20 mM Tris pH 8.0, 20 mM βME, 5 mM imidazole, 0.1% Thesit) at a flow rate of 60 mL/hour at 4 ºC. The A_{280} of the flow-through was monitored with a UV flow cell, and the column was washed until the absorbance returned to baseline. Bound protein was then eluted by washing with elution buffer (500 mM NaCl, 200 mM imidazole, 20 mM Tris pH 8.0, 20 mM βME, 0.01% Thesit) and the eluate was monitored by eye as well as A_{280} for the presence of tdTomato. Purified tdTomato was concentrated by dialysis

overnight at 4 ºC against sucrose, then buffer (50 mM NaCl, 20 mM Tris pH 8.0, 1 mM DTT), then 50% glycerol with 1 mM DTT, and stored at -20 ºC.

Genotyping

DNA was extracted from young leaves with the Shorty prep method[120]. Leaves were frozen at -80 ºC, then ground with a micropestle for 15 seconds over dry ice. 500 µL of Shorty buffer (0.2 M Tris/HCl, pH 9.0, 0.4 M LiCl [a substitution for the 0.25 M NaCl of the original protocol], 25 mM EDTA, 1% SDS) was added and the tissue was ground for another 15 seconds. Samples were centrifuged for 5 minutes at 21,130 g. 350 µL of supernatant was mixed with 350 µL of isopropanol and centrifuged for 10 minutes at 21,130 g. The supernatant was poured off and the pellet was dried at room temperature for 30 minutes. Samples were resuspended in 200 µL of TE buffer (10 mM Tris, pH 8.0, 0.1 mM EDTA).

PCR was performed on extracted DNA as well as the appropriate pBS-tdT plasmid using a reaction mixture of 1 µL of DNA, or 1 ng of the appropriate plasmid, 0.25 µL of Phusion High-Fidelity DNA Polymerase (M0530S, NEB), 5 µL Phusion High-Fidelity buffer, 0.4 mM DNTPs, and 0.2 mM 35S-FP and OCS-RP in 25 µL total, in a BioRad C1000 Thermal Cycler with the following program: 98 ºC for 1 minute, then 30 cycles of denaturation at 98 ºC for 10 seconds, annealing at 56 ºC for 15 seconds, and elongation for l minute at 72 ºC, followed by 5 minutes at 72 ºC and holding at 4 ºC. 10 µL of each PCR product were analyzed on a 1% agarose gel.

Western blots

Crude protein was extracted from a young leaf of one wild-type plant and one plant from each transgenic line by grinding each leaf for 15 seconds in 2x SDS sample buffer (4% SDS, 125 mM Tris pH 6.8, 20% glycerol, 0.01% bromophenol blue, 50 mM dithiothreitol [DTT]), using 10 µL of sample buffer per milligram of tissue. Samples were then boiled for at least five minutes in

a water bath and centrifuged for 5 minutes at 21,130 g. The supernatant was removed and used in Western blotting, below.

Wild-type Arabidopsis protein for blocking the primary anti-RFP antibody was extracted by grinding whole Arabidopsis plants in liquid nitrogen, incubating in 1 volume of 100 mM Tris pH 7.5, 1 mM PMSF, and 2% SDS, and centrifuging for 10 minutes at 21,130 g. The supernatant was transferred to a glass tube and heated in a water bath for 10 minutes at 95 ^{o}C. After cooling, 5 volumes of 50 mM Tris pH 7.5, 150 mM NaCl, and 2% Triton X-100 were added. 1 µL of anti-RFP antibody (A00682, GenScript) was blocked for 1 hour at room temperature in 1 mL wild-type Arabidopsis blocking extract and 9 mL TBST.

10 µL of crude protein extract from each wild-type and transgenic plant, as well as 10 µL of a 1:9 dilution of purified tdTomato in water, were loaded on a 12% polyacrylamide gel and electrophoresed for 1 hour at 200 V, then transferred onto PVDF. The PVDF was blocked in 5% non-fat milk in TBST (10 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.05% Tween-20) for 1 hour at 4 ^{o}C and washed four times for 15 minutes in TBST.

Blots were treated overnight at 4 ^{o}C with blocked anti-RFP antibody. Blots were washed four times for 15 minutes in TBST at room temperature, then treated for two hours at room temperature with a 1:2,500 dilution of goat anti-rabbit alkaline-phosphatase conjugated secondary antibody (A3687, Sigma), diluted in TBST. Blots were then washed four times for 15 minutes in TBST at room temperature and developed in 10 mL alkaline phosphatase buffer (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 3 mM $MgCl_2$) containing 80 µL NBT (35 mg/mL in 70% DMSO; N6876, Sigma) and 30 µL BCIP (50 mg/mL in 100% DMSO; 8503, Sigma).

Leaf fluorescence

Young leaves from the same plants used for protein extraction for the Western blot were imaged with a TRITC filter cube on an Olympus SZX12 microscope using an Insight Color Mosaic camera (Model 11.2, SPOT Imaging Solutions), running SpotBasic 5.1 software. Exposures were 100 ms for all fluorescence images, with the white and black points fixed at the maximum values to ensure fluorescence levels could be compared between images.

Sub-cellular localization of fluorescent reporter protein constructs in Arabidopsis

Leaves from confirmed, brightly fluorescent transgenic Arabidopsis lines tdT-5, sec-5, short-5, and long-7 were plasmolyzed in 1 M NaCl. Trichomes were imaged as described for onion epidermal peels, above.

Cross-linking analysis

Approximately 100 µg of freeze-dried, ground cauline stems from transgenic lines with the strongest tdTomato expression (tdT-5, sec-5, short-11, short-12, long-2, and long-7) were suspended in 500 µL of one of the following solutions, made in Pipes (piperazine-N,N'-bis-[2-ethanesulfonic acid])-sucrose buffer (50 mM Pipes, pH 6.8, 10% sucrose, 1 mM EGTA, 0.5 mM MgC12): 1 M NaCI, 8 M LiCI, 1% SDS, or Pipes-sucrose buffer only. A fifth treatment consisted of 200 µg/mL proteinase K in 50 mM Tris pH 8.0, 0.5% SDS, and 2 mM $CaCl_2$. Samples rotated on a rocker for 1 hour at room temperature. A sixth treatment consisted of digestion with 5 µL of Cellic CTEC2 in 30 mM sodium citrate, pH 4.5, in a 50 ^{o}C rotisserie oven for 2 hours. After treatment, all samples were washed three times in Pipes-sucrose buffer. A seventh treatment consisted of a crude cell wall preparation by the alcohol-insoluble-residue (AIR) method, made by incubating the biomass in 95% ethanol at 50 ^{o}C for 30 minutes or until all green color was removed, followed by 15 minutes in acetone at room temperature, then 15

minutes in methanol at room temperature[142]. The biomass was dried at room temperature and resuspended in Pipes-sucrose buffer.

For imaging the biomass, dry samples were dry-mounted and immediately imaged. Treated samples were mounted in Pipes-sucrose buffer. Imaging proceeded as described above for onion epidermal peel and Arabidopsis leaf sub-cellular localization. All fluorescence images were taken with 100 ms exposure, with the white and black points fixed at the maximum values to ensure fluorescence levels could be compared between images.

Phenotyping

Homozygous transgenic and wild-type (Col-0) plants were grown in soil (Fafard 4P mix) in 4-inch pots in a temperature-controlled growth chamber (Percival Scientific) under long-day conditions (16 h light/8 h dark) at 22 ºC. Two flats were grown, each of which contained one pot of wild-type, tdT-2, tdT-5, tdT-7, sec-2, sec-4, sec-5, short-5, short-11, short-12, long-2, long-7, long-8, and long-9. Each pot contained two plants.

Plants were photographed 5 weeks after germination. A number of parameters were measured, based on the growth stages defined by Boyes et al (2001), that could have relevance to extensin function in inflorescence stems[122]. These parameters included stem diameter at the inflorescence base, inflorescence stem length, cauline stem length, number of stems, number of branches off the main stem, wet weight of the entire inflorescence and wet weight of the cauline stem only.

Inflorescence stems were harvested when at least two siliques had shattered, between 43 and 61 days after germination. The cauline stems were flash-frozen in liquid nitrogen and freeze-dried for three days in a Labconco FreeZone[1] at -50 ºC. The stems were re-weighed to determine

dry weight and ground in a TissueLyser II (QIAgen) twice for 1 minute at 30 oscillations/second each time. Ground samples were stored at -20 ºC.

Hydroxyproline assay

Three 150 mg pools of freeze-dried, ground cauline stem biomass from 2-4 plants were made for each line. Three 50 mg samples from each pool were weighed into 2 mL Sarstedt screwtop tubes and 1.4 mL of 6 M HCl was added to each tube. Samples were hydrolyzed at 100 ºC for 18 hours. The supernatant was assayed for hydroxyproline according to a modification of the methods of Kivrikko and Liesmaa[68] as follows: 500 µL were transferred to a new tube and pH-adjusted by the addition of 12 M NaOH to pH 3.0 (± 0.1). A higher pH led to development of a dark brown color in the samples that interfered with subsequent colorimetric analysis. The samples were centrifuged at 21,130 g for 1 minute to clarify the supernatant. 125 µL of sample supernatant was then mixed with 250 µL of 50 mM sodium hypobromide and incubated at room temperature for 5 minutes. The oxidation reaction was stopped by the addition of 125 µL of 6 M HCl, followed by 250 µL of para-dimethylaminobenzaldehyde (DMAB, 5% in n-propanol) to all tubes. Tubes were sealed, mixed by hand, and incubated at 70 ºC for 15 minutes. After cooling, the absorbance at 560 nm of each sample was measured in triplicate (200 µL each) in a 96-well plate, using a BioTek Synergy HT plate reader. Data were analyzed in Microsoft Excel 2011. Hydroxyproline standards with the same salt concentrations (4.2 M NaCl) and pH (3) as the samples were used to construct a standard curve.

Digestibility assay

Glucose release (digestibility) was analyzed following an adaptation of the protocol of Santoro *et al*[70]. Triplicate aliquots of freeze-dried, ground biomass for each plant of 5 mg ± 0.3 mg were weighed into Sarstedt 2 mL screwtop tubes. 700 µL of 6.25 mM NaOH was added and

the tubes were vortexed and placed in a 90 ºC heat block for 3 hours. At the end of the pretreatment, the samples were placed on ice, then centrifuged at 21,130 g for 30 seconds, and 50 µL of cell wall digestion mix was added, consisting of Cellic CTEC2 (70 mg/g biomass, Novozymes), Cellic HTEC2 (2.5 mg/g biomass, Novozymes), 0.45 M sodium citrate, pH 4.5 (final concentration 30 mM), and 0.1% NaN_3 (final concentration 0.007%). Samples were incubated in a 50 ºC rotisserie oven with end-over-end rotation for 20 hours.

The tubes were centrifuged at 21,130 g for 1 minute. 650 µL of supernatant was removed to a new tube, diluted 1:10 in water, and measured for glucose content in a GOPOD assay (Megazyme). Triplicate 20 µL aliquots of sample, glucose standard, or an enzyme-only mixture were mixed with 200 µL of GOPOD reagent and incubated at 40 ºC for 45 minutes in a 96-well plate. The absorbance at 510 nm was read in a BioTek Synergy HT plate reader. Data were analyzed in Microsoft Excel 2011. Glucose standards were used to construct a standard curve. The absorbance from the enzyme-only mixture was subtracted from the absorbance of the samples to correct for glucose present in Cellic CTEC2 and HTEC2 enzyme preparations.

RESULTS

Construct confirmation with restriction enzyme digestion and sequencing

Sequences encoding the four tdTomato constructs (cyto-tdT, sec-tdT, sec-tdT-short, and sec-tdT-long; Figure 28A) were successfully transferred into the pBS vector for transient transformation of onion cells, confirmed by sequencing as well as by restriction enzyme digestion of the vector with *Not*I to remove the cassette (cleaving before the 35S promoter and after the OCS transcription terminator), followed by agarose gel electrophoresis (Figure 31). The bands on the gel reflect the expected sizes for each construct, particularly the ~80 bp difference between cyto-tdT and sec-tdT, the ~80 bp difference between sec-tdT and sec-tdT-short, and the

~1100 bp difference between sec-tdT-short and sec-tdT-long. Correct transformation of Agrobacterium strain LBA4404 with the pART27-fluorescent reporter protein plasmids was confirmed by sequencing (data not shown).

Sub-cellular localization of tdTomato in transiently transformed cells

As a first test of the reporter protein constructs, I transiently transformed onion epidermal cells to express each of the four tdTomato constructs described above. After transformation, cells transformed with each of the construct looked similar; no differences in sub-cellular localization could be seen (Figure 32). After plasmolysis with 1 M NaCl, in every case tdTomato was localized in the appropriate sub-cellular compartment: cyto-tdT appeared exclusively in the cytoplasm, while sec-tdT, sec-tdT-short, and sec-tdT-long all appeared both in the cytoplasm and cell wall (Figure 33). I then washed the onion peels with various solutions, including 1 M NaCl and 10% SDS, but fluorescence did not noticeably decrease in any of the transformed cells, including those transformed with cyto-tdT (data not shown).

Expression in stably transformed Arabidopsis plants

In order to create a greater population of cells expressing the four tdTomato constructs, I stably transformed Arabidopsis with each construct using Agrobacterium. After breeding at least three independently transformed lines for each construct to be homozygous, I confirmed the transgene insertion by PCR (Figure 34). As expected, bands from plants transformed with cyto-tdT were about 80 bp smaller than those from plants transformed with sec-tdT, which were themselves about 80 bp smaller than those from plants transformed with sec-tdT-short; bands from plants transformed with sec-tdT-long were about 1100 bp larger than those from plants transformed with sec-tdT-short. Three lines were confirmed to be transformed with the cyto-tdT construct, named tdT-2, tdT-5, and tdT-7. Three lines were confirmed to be transformed with the

sec-tdT construct, named sec-2, sec-4, and sec-5. Three lines were confirmed to be transformed with the sec-tdT-short construct, named short-5, short-11, and short-12. Four lines were confirmed to be transformed with the sec-tdT-long construct, named long-2, long-7, long-8, and long-9.

I analyzed the level of tdTomato expression in the lines confirmed with PCR using both an anti-RFP Western blot on protein extracted from leaves and fluorescence microscopy on leaves from the same plants (Figure 35). The lines transformed with each construct exhibited a range of expression, reflecting the transcriptional activity in the region where the transgene inserted. The intensity of the anti-RFP band on the Western blot generally correlated to the fluorescence detected in the leaves, where tdTomato from lines with low fluorescence (sec-4, short-5, long-8, and long-9) was weakly detected on the blot, while tdTomato from lines with strong fluorescence (tdT-5, sec-5, short-11, short-12, long-2, and long-7) was strongly detected on the blot. A few lines had strong bands on the blot but weak fluorescence (tdT-2, tdT-7, sec-2); no lines had strong fluorescence but weak bands on the blot. The reduced fluorescence compared to high levels of tdTomato in tdT-2, tdT-7, and sec-2 may reflect other factors involved in tdTomato fluorescence.

The size of the bands detected on the Western blot was the same for purified tdTomato as for the cyto-tdT and sec-tdT lines, as expected. Puzzlingly, the predicted size for tdTomato is 54 kDa (assuming a molecular weight of 110 daltons per amino acid), but the apparent size from the Western blot is greater than 75 kDa. SDS does not affect tdTomato fluorescence (as seen in biomass imaging, described below); perhaps tdTomato is insufficiently denatured by SDS sample buffer to run true to size. The sec-tdT-long protein is predicted to be 94 kDa, or 40 kDa larger than tdTomato. In the long-2, long-7, and long-8 lanes, proteins of various sizes, all larger

than tdTomato, were detected. The strongest band in all cases was much more than 40 kDa larger than tdTomato, suggesting post-translational modification of the sec-tdT-long reporter protein had occurred to increase its molecular mass.

Sub-cellular localization of tdTomato in Arabidopsis

Leaves of representative transgenic lines (tdT-5, sec-5, short-5, and long-7) were imaged after plasmolysis to confirm appropriate sub-cellular localization of tdTomato fluorescence. Plasmolysis clearly separated the cytoplasm from the cell wall, which could be most readily imaged in trichomes (Figure 36). In tdT-5, fluorescence was observed only in plasmolyzed cytoplasm, while in sec-5, short-5, and long-7, fluorescence was observed in both the cytoplasm and cell wall. As with transiently transformed onion cells, the secretory sequence of the sec-tdT, sec-tdT-short, and sec-tdT-long constructs was both necessary and sufficient to direct tdTomato to the cell wall.

Effect of different treatments on tdTomato fluorescence in muro

Since plants transformed with sec-tdT, sec-tdT-short, and sec-tdT-long all exhibited fluorescence in cell walls, I wanted to know whether the presence of the extensin domain in sec-tdT-short or sec-tdT-long could be correlated with levels of tdTomato fluorescence in cell walls after treatments intended to remove cell-wall-associated proteins. I analyzed freeze-dried, ground, cauline stem biomass from a wild-type plant as well as a plant from each of the most fluorescent lines, namely, tdT-5, sec-5, short-11, short-12, long-2, and long-7. Biomass was examined with fluorescence microscopy either in a dry, untreated state, or after preparation of the alcohol-insoluble residue (AIR) or treatment with 1 M NaCl, 8 M LiCl, CTEC2 cellulase, 1% SDS, 200 µg/mL proteinase K, or PIPES-sucrose buffer only (Figure 37, Figure 38,

Table 7). No background fluorescence was detected in wild-type biomass under any treatment conditions. Dry biomass was highly fluorescent for all transgenic lines. Treatment in buffer removed fluorescence from tdT-5, and fluorescence was slightly reduced in the other transgenic lines, indicating removal of cytoplasmic reporter protein. Treatment with 1 M NaCl, which should remove ionically bound proteins (such as perhaps sec-tdT), only decreased fluorescence in tdT-5. Fluorescence in all other lines was comparable to that seen after buffer-only treatment. Treatment with 8 M LiCl, a chaotropic agent capable of disrupting the outer hydroxyproline-rich glycoprotein cell wall matrix of *Chlamydomonas reinhardii*[143], dramatically reduced fluorescence in all transgenic lines except long-2 and long-7. No fluorescence was observed in tdT-5, while slight fluorescence in some biomass particles could be observed in sec-5, short-11, and short-12. Fluorescence in long-2 and long-7 was still strong, comparable to that seen after buffer-only or 1 M NaCl treatment. An ethanol-washed crude cell-wall preparation (AIR) dramatically reduced fluorescence in all transgenic lines. In tdT-5, sec-5, and short-12, essentially no fluorescence was observed. However, in short-11, long-2, and long-7, fluorescence could still be seen, although at much reduced levels compared to buffer-only treatment. Treatment with 1% SDS reduced fluorescence in all transgenic lines except long-2 and long-7, with tdT-5 and sec-5 exhibiting barely observable fluorescence and short-11 and short-12 exhibiting noticeably reduced fluorescence compared to buffer-only treatment (Figure 38). Treatment with either proteinase K or CTEC2 reduced fluorescence in all the transgenic lines, including long-2 and long-7. Slight fluorescence could still be detected in thicker cell wall chunks of all lines except tdT-5 after proteinase K digestion, likely because tdTomato was more protected in the more intact pieces of inflorescence stem.

Expression of tdT or tdT-extensin constructs does not affect developmental phenotypes

In order to investigate whether overexpression of an extensin domain has any effect on plant growth and development, I grew each of the transgenic lines described above (tdT-2, tdT-5, tdT-7; sec-2, sec-4, sec-5; short-5, short-11, short-12; long-2, long-7, long-8, long-9) together in a growth chamber, along with wild-type plants. I measured a suite of developmental characteristics in which increased extensin cross-linking might produce a phenotype, based on the growth stages of Arabidopsis described by Boyes *et al.*[122] (Table 8, Figure 39 – Figure 45).

No plants displayed any major growth phenotypes, implying that any effects from extensin overexpression would be subtle (Figure 39). There were a few small differences, including several characteristics that were statistically significantly different from wild-type in a few cyto-tdT and sec-tdT lines ($p < 0.05$, Student's T-test). These characteristics also differed in one sec-tdT-short and one sec-tdT-long line, although not always in the expected direction. Stems were thicker in tdT-5, sec-2, and (as expected based on previous studies[124]) long-9 (wt, 1.2 ± 0.04 mm; tdT-5, 1.6 ± 0.04 mm; sec-2, 1.5 ± 0.03 mm; long-9, 1.6 ± 0.06 mm) (Figure 40). The inflorescence stem (but not the basal, cauline stem, which was the only part of the stem used in subsequent experiments) was unexpectedly longer in short-12 (wt, 324 ± 12 mm; short-12, 393 ± 2 mm) (Figure 41, Figure 42). Sec-2, sec-4, and long-9 also had significantly more biomass than wild-type (*e.g.*, dry weight of wt cauline stems, 47 ± 4 mg; sec-2, 130 ± 11 mg; sec-4, 97 ± 6 mg; long-9, 112 ± 3 mg) (Figure 43 - Figure 45).

Expression of sec-tdT-long increases hydroxyproline content

I measured the amount of insoluble hydroxyproline in inflorescence stems of transgenic and wild-type lines in order to determine whether the extensin domain of sec-tdT-short or sec-tdT-long increased the abundance of extensins bound in cell walls (Figure 46). As expected, the

hydroxyproline content in stems was not significantly different from wild-type in any cyto-tdT lines ($p > 0.2$, Student's T-test), sec-tdT lines ($p > 0.8$), or sec-tdT-short lines ($p > 0.16$). I measured an average of 0.047% Hyp/mg dry weight in wild-type plants, and a comparable amount in the cyto-tdT (an average for all three lines of 0.048%), sec-tdT (an average for all three lines of 0.047%), and sec-tdT-short lines (an average for all three lines of 0.043%). The stem hydroxyproline content I measured in wild-type plants is comparable to what has been reported in tomato stems (0.05%)[123] as well as other Arabidopsis ecotypes (0.06% in ecotype Landsberg *erecta* and 0.03% in Wassilewskija[124]). The two sec-tdT-long lines with the lowest tdTomato fluorescence and expression (long-8 and long-9) also had comparable levels of hydroxyproline in stems as wild-type (an average of 0.054% in long-8 and 0.041% in long-9, $p > 0.12$). However, the two sec-tdT-long lines with high tdTomato fluorescence and expression, as well as strong association of fluorescence with cauline stem cell walls, had much higher levels of hydroxyproline than wild-type, with approximately 2.5 times more in long-2 (an average of 0.11%, $p = 0.0001$) and 3 times more in long-7 (0.14% [n=1; the significance of this difference could not be evaluated]).

Expression of tdT or tdT-extensin constructs does not affect biomass digestibility

Since long-2 and long-7 had increased abundance of hydroxyproline-rich cell-wall-associated extensins compared to wild-type, I anticipated that inflorescence stems of plants from these lines would be less digestible than wild-type, while plants from the cyto-tdT, sec-tdT and perhaps sec-tdT-short lines should have comparable digestibility. I measured glucose release from biomass of 2 – 4 plants from each line after pretreatment and cell wall digestion in triplicate (Figure 47). As expected, no cyto-tdT line had significant differences in glucose released compared to wild-type ($p > 0.3$, Student's T-test). Unexpectedly, glucose release from several

lines transformed with the other constructs was significantly different from wild-type. Sec-5 and short-5 both had less glucose released than wild-type (wt, 9.1 ± 0.1%; sec-5, 8.0 ± 0.2%; short-5, 7.9 ± 0.3%; p = 0.04 in both cases), while long-7 and long-8 had more (long-7, 10.2 ± 0.02%, p = 0.03; long-8, 10.6 ± 0.2, p = 0.047). Long-2, which is the other line that had more hydroxyproline, had 8.5 ± 0.3% glucose released, an insignificant difference compared to wild-type. The average glucose release from all the lines analyzed was as follows (± standard error): wt, 9.1 ± 0.1%; cyto-tdT, 9.6 ± 0.4%; sec-tdT, 8.9 ± 0.1%; sec-tdT-short, 8.5 ± 0.6%; sec-tdT-long, 9.7 ± 0.2%.

DISCUSSION

Expression and modification of tdT and tdT-extensin proteins

The two extensin domains, "short" and "long", were chosen in order to evaluate the relative contributions of tyrosine and SP_n sequences to extensin function in the cell wall. The short extensin domain contains a single SP_4 motif and two tyrosines, which are separated by 10 amino acids rather than being found in the classic YXY motif involved in isodityrosine formation and extensin cross-linking (Figure 28, Figure 29). The long extensin domain encodes a version of *SlPEX1* lacking the leucine-rich repeat domain, and includes 37 SP_n repeats and the same two tyrosines as found in the short domain.

Transgenic Arabidopsis plants were made to assess the functionality of these extensin domains indirectly, by observing the tagging reporter protein tdTomato with fluorescence microscopy. A range of tdTomato expression, from barely detectable to overwhelmingly bright, was established in several independently transformed lines expressing each construct. These lines were evaluated qualitatively in terms of the relative fluorescence signal under identical fluorescence microscopy conditions, as well as relative detection of tdTomato in an anti-RFP

Western blot (Figure 35). Lines with the strongest fluorescence also had the strongest signal on the blot.

Earlier investigations by Estevez *et al.* of transgenic Arabidopsis lines using the CaMV 35S promoter to overexpress GFP-tagged synthetic extensin domains revealed post-translational modification, including hydroxylation and glycosylation, of the synthetic extensin[34]. These modifications were first identified by the unexpectedly large size of the synthetic protein detected on an anti-GFP Western blot, and confirmed by affinity purification of the synthetic proteins followed by measurement of molar protein and carbohydrate content to quantify hydroxyproline conversion and arabinosylation. Furthermore, GFP-tagged synthetic extensins were reported to vary, depending on the organ, in degree of post-translational modification (perhaps due to local control of the activity of extensin post-translational modification enzymes) and sub-cellular localization (proteins were apparently not secreted in roots, and poorly secreted in other organs[§]). In my studies, lines expressing detectable levels of sec-tdT-long (long-2, long-7, and long-8) showed similarly unexpectedly large sizes of protein on the Western blot, with bands much larger than 94 kDa, which is the predicted size based on the encoded amino acid sequence of the unmodified protein (Figure 35A). Performing a similar analysis as Estevez *et al.* (affinity purification using the tdTomato tag, followed by protein and carbohydrate analysis of the purified protein) may also reveal post-translational modification of the single SP_4 motif of sec-tdT-short, as well as clarifying the extent of post-translational modification of sec-tdT-long. I only examined aerial organs for this study (leaves and stems) due to their greater relevance for

[§] GFP is unstable under acidic conditions, although raising the pH can result in recovery of fluorescence as new GFP is synthesized. Since the cell wall is acidic, GFP fluorescence in the cell wall is not detectable without pH adjustment (Fleming, unpublished results, Scott [1999][144] and Genovesi [2008][145]). In Estevez, 2006, cells were plasmolyzed in mannitol without any pH adjustment, so secreted GFP may not have been detectable.

biofuel applications, but closer examination of sec-tdT-short and sec-tdT-long in all organs may expose variation in sub-cellular localization and post-translational modification, similar to the findings of Estevez *et al.*

Relative strength of cell wall association of tdTomato

Since sec-tdT-short and sec-tdT-long were secreted and, in the case of sec-tdT-long, post-translationally modified, it seemed possible that the fluorescent reporter protein constructs would show the same strong, and possibly covalent, cell-wall association as detected for the SlPEX1 homologs mPEX1 and mPEX2. I performed a qualitative imaging experiment, in which the presence of fluorescent reporter protein constructs in cauline stem biomass from transgenic lines with the strongest reporter protein expression (tdT-5, sec-5, short-11, short-12, long-2, and long-7) was evaluated by fluorescence microscopy after treatment in various conditions designed to extract non-covalently-linked proteins (NaCl, LiCl, AIR, SDS, proteinase K, and CTEC2) (Figure 37, Figure 38, Table 7). All treatments removed observable fluorescence from tdT-5, and all treatments except 1 M NaCl drastically reduced or eliminated fluorescence from sec-5. However, fluorescence was maintained after certain treatments in lines short-11 (1 M NaCl, AIR, 1% SDS), short-12 (1 M NaCl and 1% SDS), long-2, and long-7 (1 M NaCl, 8 M LiCl, AIR, and 1% SDS for both lines). Treatment with proteinase K or CTEC2 was able to drastically reduce or eliminate fluorescence from all lines, indicating that retention of fluorescence in the sec-tdT-short and sec-tdT-long lines after the other treatments was not due to inaccessibility of the tdTomato domain. Treatment with CTEC2, a cellulase preparation, was not expected to affect fluorescence directly. However, loss of fluorescence after CTEC2 treatment correlates with early results of Derek Lamport, where cellulase treatment released 70 – 75% of cell-wall-associated hydroxyproline in tomato and tobacco cell cultures[94]. The CTEC2 preparation may have protease

activity, or its known cellulase activity may have disrupted the overall cell wall structure and released tdTomato from the cell wall. The comparatively strong fluorescence observed in long-2 and long-7 after treatment with LiCl (which is capable of dissociating the non-covalently-linked HRGP cell walls of *C. reinhardii*) and 1% SDS strongly suggests that sec-tdT-long is covalently linked to the cell wall in these lines. Interestingly, even though sec-tdT-short fluorescence was mostly removed by 8 M LiCl treatment, observable but reduced fluorescence was retained after 1% SDS treatment (Figure 38). The short extensin domain may be covalently bound in the cell wall, but in such a way that LiCl can affect tdTomato fluorescence. Quantitative and non-fluorescence based experiments, such as detection of tdTomato in the supernatants resulting from each treatment, will help clarify the nature of the association of sec-tdT-short and sec-tdT-long with the cell wall.

Plant growth phenotypes related to tdTomato expression

Since my data suggest that sec-tdT-long is covalently cross-linked in the cell wall, and sec-tdT-short has a strong association with the cell wall, I expected to see similar phenotypes in transgenic Arabidopsis expressing sec-tdT-short and sec-tdT-long as were found in Arabidopsis overexpressing *EXTENSIN-1* with the CaMV 35S promoter[124]. The differences found in stem height in that study were small but significant (a reduction of 5-8% in inflorescence stem length, n = 11-15), and were directly correlated with the level of overexpression of *EXT1*, in that lines with greater hydroxyproline content had a greater reduction in stem length. Presumably, overproduction of EXT1 leads to more highly cross-linked cell walls and reduced cell expansion, producing smaller cells and therefore shorter plants. I was unable to detect major plant growth phenotypes in transgenic lines overexpressing tdT or tdT-extensin constructs compared to wild-type plants (the correlation with hydroxyproline content is discussed below) (Table 8). Some

differences appeared statistically significant, particularly in the amount of biomass produced, which was greater in several transgenic lines compared to wild-type (sec-2, sec-4, and long-9). However, as I expected to see either no difference (in cyto-tdT and sec-tdT lines) or less biomass (in sec-tdT-short or sec-tdT-long), these findings are difficult to explain.

Looking at all the data I collected on wild-type plants for this and other experiments (see Chapter 3, Table 6), it is clear that there is a wide range of values for many of the traits I measured in wild-type plants alone. This experiment comparing plant growth of transgenic and wild-type Arabidopsis had small samples sizes (n = 2 – 4 for each independently transformed line), making it very difficult to determine whether the small effects are truly significant. The experiments with *EXT1*, in contrast, had n of 11 – 15, making small effects easier to detect. That the different lines expressed the transgenic constructs at different levels also adds another variable to the experiment. It is difficult to obtain transgenic lines with identical expression characteristics given the random insertion of the transgene into the genome, but perhaps for detailed phenotypic analyses some level of normalization for the expression of tdTomato should be performed. However, even when comparing lines with similarly high levels of tdTomato expression (tdT-5, sec-5, short-11, short-12, long-2 and long-7), no parameters stand out as being noticeably different from wild-type.

In contrast, the differences detected in the hydroxyproline content of long-2 and long-7 were quite large (2.3 and 2.9 times greater than wild-type, respectively)[**] (Figure 46). This compares well to the increase measured in hydroxyproline in *AtEXT1* overexpression lines (1.5 –

[**] The only developmental phenotype with a similarly large difference between a transgenic line and wild-type was cauline stem dry weight. However, as noted above, cauline stem dry weights seem to have a wide range of wild-type values, so this difference may not be significant.

4 times greater)[124]. Hydroxyproline content was less variable than the developmental characteristics I measured, as the levels of hydroxyproline measured in wild-type plants and the other transgenic lines were extremely consistent, and compared well with the values obtained from my previous experiment (0.047%% Hyp/mg in this experiment *versus* 0.048%) as well as those measured by other investigators (0.06% in ecotype Landsberg *erecta* and 0.03% in Wassilewskija[124]). The differences seen in hydroxyproline content are therefore large enough that the small sample size is not problematic, at least in the case of long-2. As long-7 had 2.9 times more hydroxyproline compared to wild-type but was only measured once, the experiment needs to be replicated to validate these findings. Given the high levels of sec-tdT-long expression in long-2 and long-7, the evidence of post-translational modification of the sec-tdT-long construct from the Western blot, and the retention of fluorescence in long-2 and long-7 cell walls after various treatments, it is highly likely that the increased hydroxyproline content detected in long-2 and long-7 is due to overexpression, post-translational modification, and cross-linking of sec-tdT-long. It is not remarkable that no differences in hydroxyproline content were measured between wild-type and any sec-tdT-short lines, as the predicted hydroxyproline content of sec-tdT-short is so low (one SP_4 motif) compared to sec-tdT-long (37 SP_n motifs).

Surprisingly, given the greatly increased hydroxyproline content of long-2 and long-7, glucose release was not dramatically affected in any sec-tdT-long line (Figure 47). As with the plant growth phenotypes, the effect sizes were quite small (for example, a difference of 1.4% between wild-type and long-8, which had the most glucose released). Replication will be necessary to be confident that any differences in glucose release are due to expression of sec-tdT-long. Since no line had considerable differences in glucose release from the cauline stem, even though long-2 and long-7 have substantially increased levels of extensin in their cauline

stem cell walls that is probably cross-linked, it is unlikely that extensin overexpression and increased cross-linking are highly relevant in terms of biofuel production, at least in the Arabidopsis model system. Then again, there may be a maximum amount of recalcitrance that can be attributed to extensin cross-linking, and native extensin expression reaches this maximum. In this case, overexpression of extensins cannot further increase recalcitrance. However, further experiments, such as treating transgenic biomass with a protease prior to digestion with cellulase (similar to the experiments described in Chapter 2; a significant increase in digestibility would be predicted in sec-tdT-long lines compared to wild-type), may reveal more about the role of extensins in biomass recalcitrance. Plants over-expressing tdT-extensin fusion proteins can also be forced to increase extensin cross-linking by stimulating hydrogen peroxide production through wounding or infection of the plants, since extensin cross-linking depends on hydrogen peroxide[29,55]. Increased cross-linking of the fusion proteins could make the digestibility phenotypes more pronounced.

In this study, leaves were examined because they were easy to acquire and their removal would not affect the plant, while stems were examined because they are an analog for woody biomass used in biofuel production. However, roots may be an organ of greater interest with respect to extensin-related phenotypes (particularly cell-wall-associated fluorescence, hydroxyproline content, and glucose release), because in Arabidopsis, both extensins and the enzymes for extensin post-translational modification are most highly expressed in roots[23,33]. In addition, monitoring the fluorescent tdTomato tag in transgenic plants crossed with mutant lines (as described in Chapter 3) having reduced extensin post-translational modification could also reveal the importance of each modification for secretion and cell wall association.

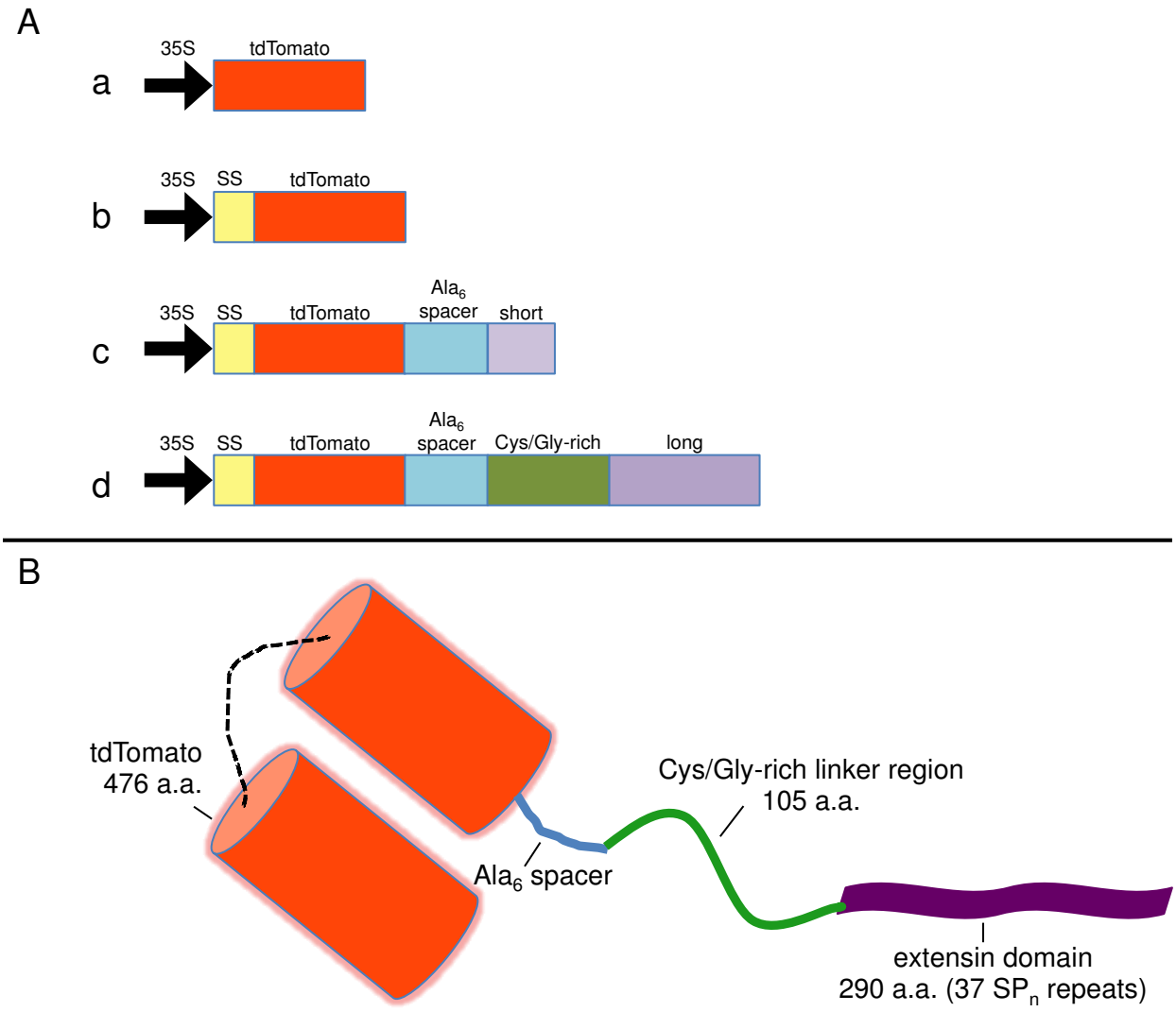

Figure 28. Schematic diagrams of the tdT and tdT-extensin constructs expressed under the control of the 35S promoter in Arabidopsis plants.
A. Diagrams (not to scale) of each of the four constructs: a) cyto-tdT, b) sec-tdT, c) sec-tdT-short, d) sec-tdT-long. SS, tomato polygalacturonase signal sequence. Ala-spacer, 6 alanines separating tdTomato from extensin domain. Short, C-terminal 20 amino acids of *SlPEX1*. Cys/Gly-rich, sequence from *SlPEX1* intervening between native LRR and EXT domains. Long, C-terminal 290 amino acids of *SlPEX1*.
B. Cartoon of the mature sec-tdT-long reporter protein construct

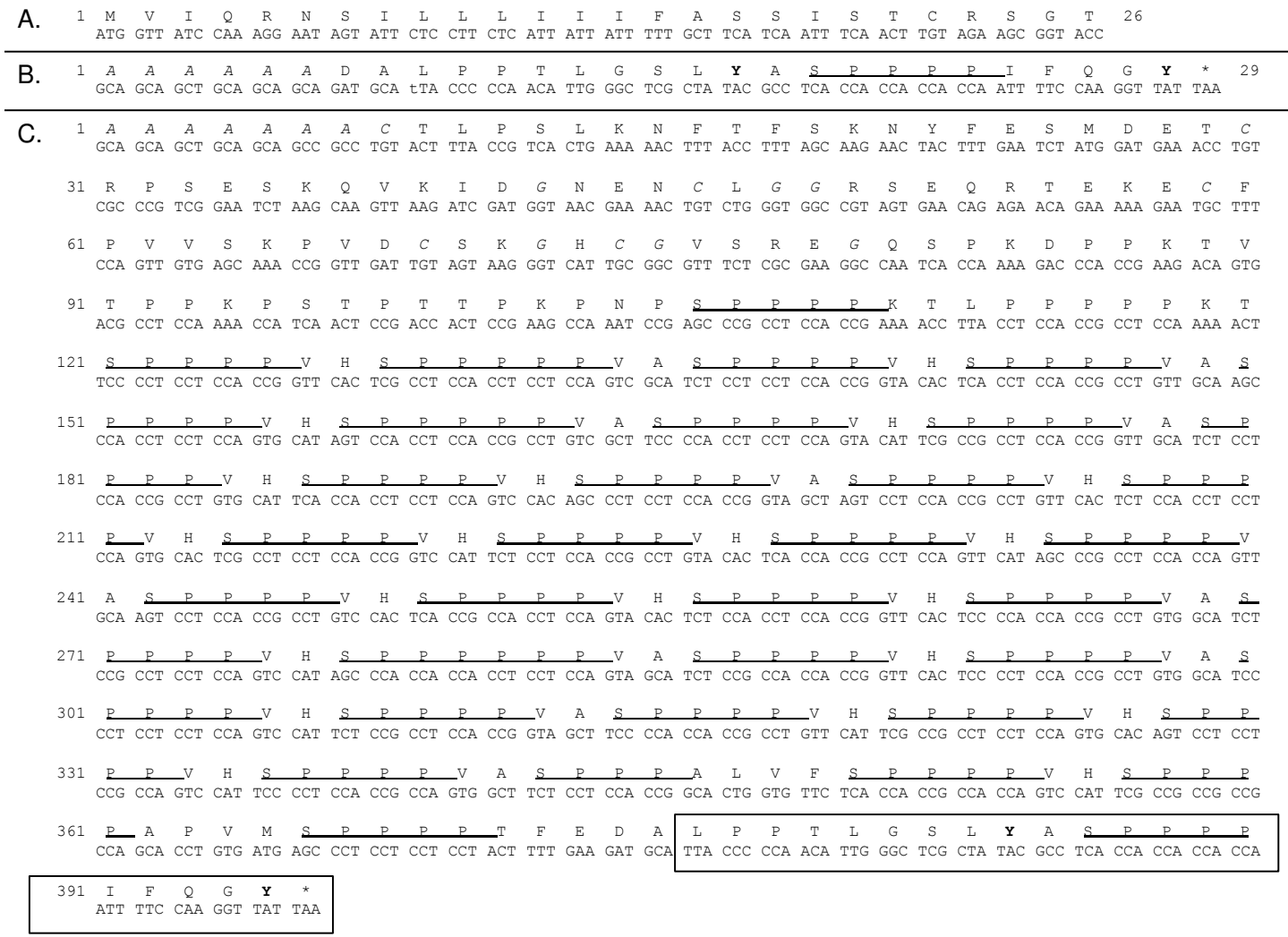

Figure 29. Coding sequence and proposed translation of crucial gene fragments.
A) secretory sequence from tomato polygalacturonase; B) the short extensin domain; C) the long extensin domain used in the tdT-extensin constructs. The 6-alanine spacer and cysteine and glycine residues of the cys/gly-rich linker region are italicized. SPn motifs are underlined. Tyrosines are in bold. Box indicates sequence of short extensin domain within the context of the long extensin domain.

A. gaattcATGGTTTCCAAGGGTGAGGAGGTTATCAAAGAGTTCATGAGATTCAAGGTTAGGATGGAAGGTTCCATGAACGGTCACGAGTTCGAGATCGAGGGCGAGGG
TGAAGGTAGACCCTACGAGGGCTCCCAAACCGCAAAGCTCAAAGTGACTAAGGGTGGTCCTTTGCCCTTCGCTTGGGACATCTTGTCCCCCCAATTCATGTATGGCT
CTAAGGCATACGTTAAGCATCCTGCTGACATCCCCGATTACAAAAAGTTGTCCTTCCCAGAGGGTTTCAAGTGGGAAAGGGTCATGAACTTCGAGGATGGAGGTCTT
GTGACTGTGACCCAAGATTCTAGTTTGCAGGACGGCACTTTGATCTACAAGGTGAAGATGAGAGGCACAAACTTTCCTCCCGATGGTCCAGTCATGCAAAAGAAAAC
TATGGGTTGGGAAGCCTCCACTGAGAGGCTTTACCCAAGAGACGGCGTTCTTAAGGGTGAAATCCACCAAGCTCTCAAACTTAAGGATGGAGGCCACTACTTGGTGG
AGTTCAAGACCATCTACATGGCTAAGAAGCCCGTGCAACTCCCCGGCTATTACTACGTGGACACTAAACTCGATATCACCTCCCACAACGAGGACTACACCATCGTT
GAACAATATGAGAGGTCTGAGGGTCGCCATCACCTTTTCTTGGGTCATGGTACTGGAAGCACCGGTAGTGGCAGCTCTGGCACCGCTTCATCCGAGGATAATAACAT
GGCTGTGATCAAGGAGTTTATGCGCTTCAAAGTCCGTATGGAGGGCTCAATGAATGGCCACGAGTTCGAGATCGAAGGAGAGGGTGAGGGCCGCCCATATGAGGGCA
CTCAGACAGCTAAGTTGAAAGTCACCAAGGGTGGACCACTTCCTTTCGCTTGGGATATTCTCTCACCACAGTTTATGTACGGTTCCAAGGCTTACGTGAAACACCCA
GCCGATATTCCAGATTATAAGAAGTTGTCTTTCCCAGAAGGATTTAAGTGGGAGCGCGTTATGAACTTCGAGGACGGTGGTTTGGTTACAGTCACCCAAGACTCCTC
CCTTCAAGATGGTACGCTTATCTACAAGGTCAAAATGCGTGGAACCAATTTCCCACCAGACGGCCCAGTTATGCAGAAGAAGACTATGGGCTGGGAGGCTTCAACAG
AGCGCTTGTATCCCCGCGATGGAGTGTTGAAGGGCGAGATTCACCAGGCATTGAAGTTGAAGGACGGTGGACATTACCTCGTGGAGTTTAAGACCATCTACATGGCC
AAGAAACCCGTTCAATTGCCTGGTTATTACTACGTTGATACCAAGTTGGACATTACCTCCCACAACGAGGATTACACCATTGTCGAACAGTACGAGCGTTCCGAGGG
CCGCCACCACCTCTTCCTCTACGGTATGGACGAGTTGTACAAGGCAGCAGCTGCAgcagcagatgcat**ttacccccaacattgggctcgctatacgcctcaccaccac**
caccaattttccaaggttattaactcgagtgctttaatgagatatgcgagacgcctatgatcgcatgatatttgctttcaattctgttgtgcacgttgtaaaaaacc
tgagcatgtgtagctcagatccttaccgccggtttcggttcattctaatgaatatatcacccgttactatcgtatttttatgaataatattctccgttcaatttact
gattgtaccctactacttatatgtacaatattaaaatgaaaacaatatattgtgctgaataggtttatagcgacatctatgatagagcgccacaataacaaacaatt
gcgttttattattacaaatccaattttaaaaaaagcggcagaaccggtcaaacctaaaagactgattacataaatcttattcaaatttcaaaaggccccaggggcta
gtatctacgacacaccgagcggcgaactaataacgttcactgaagggaactccggttcccCGCCGGCG

B. CTGCAGCAGCCGCCTGTACTTTACCGTCACTGAAAAACTTTACCTTTAGCAAGAACTACTTTGAATCTATGGATGAAACCTGTCGCCCGTCGGAATCTAAGCAAGTTAA
GATCGATGGTAACGAAAACTGTCTGGGTGGCCGTAGTGAACAGAGAACAGAAAAAGAATGCTTTCCAGTTGTGAGCAAACCGGTTGATTGTAGTAAGGGTCATTGCGGC
GTTTCTCGCGAAGGCCAATCACCAAAAGACCCACCGAAGACAGTGACGCCTCCAAAACCATCAACTCCGACCACTCCGAAGCCAAATCCGAGCCCGCCTCCACCGAAAA
CCTTACCTCCACCGCCTCCAAAAACTTCCCCTCCTCCACCGGTTCACTCGCCTCCACCTCCTCCAGTCGCATCTCCTCCTCCACCGGTACACTCACCTCCACCGCCTGT
TGCAAGCCCACCTCCTCCAGTGCATAGTCCACCTCCACCGCCTGTCGCTTCCCCACCTCCTCCAGTACATTCGCCGCCTCCACCGGTTGCATCTCCTCCACCGCCTGTG
CATTCACCACCTCCTCCAGTCCACAGCCCTCCTCCACCGGTAGCTAGTCCTCCACCGCCTGTTCACTCTCCACCTCCTCCAGTGCACTCGCCTCCTCCACCGGTCCATT
CTCCTCCACCGCCTGTACACTCACCACCGCCTCCAGTTCATAGCCCGCCTCCACCAGTTGCAAGTCCTCCACCGCCTGTCCACTCACCGCCACCTCCAGTACACTCTCC
ACCTCCACCGGTTCACTCCCCACCACCGCCTGTGGCATCTCCGCCTCCTCCAGTCCATAGCCCACCACCACCTCCTCCAGTAGCATCTCCGCCACCACCGGTTCACTCC
CCTCCACCGCCTGTGGCATCCCCTCCTCCTCCAGTCCATTCTCCGCCTCCACCGGTAGCTTCCCCACCACCGCCTGTTCATTCGCCGCCTCCTCCAGTGCACAGTCCTC
CTCCGCCAGTCCATTCCCCTCCACCGCCAGTGGCTTCTCCTCCACCGGCACTGGTGTTCTCACCACCGCCACCAGTCCATTCGCCGCCGCCGCCAGCACCTGTGATGAG
CCCTCCTCCTCCTACTTTTGAAGATGCAT

Figure 30. Sequences ordered from GenScript
A) pUC57-less; B) pUC57-high, shown 5'-3'.
In A), bases encoding tdTomato are uppercase. Bases encoding the short anchor sequence are in bold. Single underline, recognition sites for *Pst*I and *Nsi*I; double underline, recognition sites for *Nde*I and *SgrA*I.

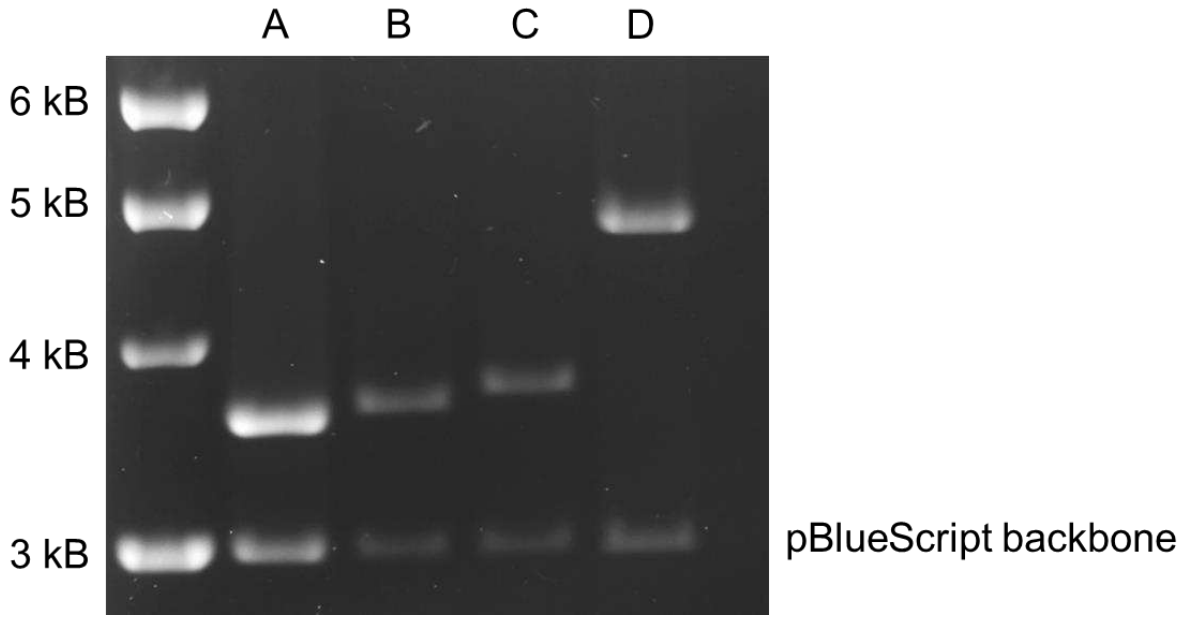

Figure 31. Results of agarose gel electrophoresis after restriction enzyme digestion of positive clones of the four pBS-tdT plasmids
Each construct in the pBS backbone (3 kb) was digested with *Not*I and separated by gel electrophoresis. A) pBS-cyto-tdT; B) pBS-sec-tdT; C) pBS-sec-tdT-short; D) pBS-sec-tdT-long.

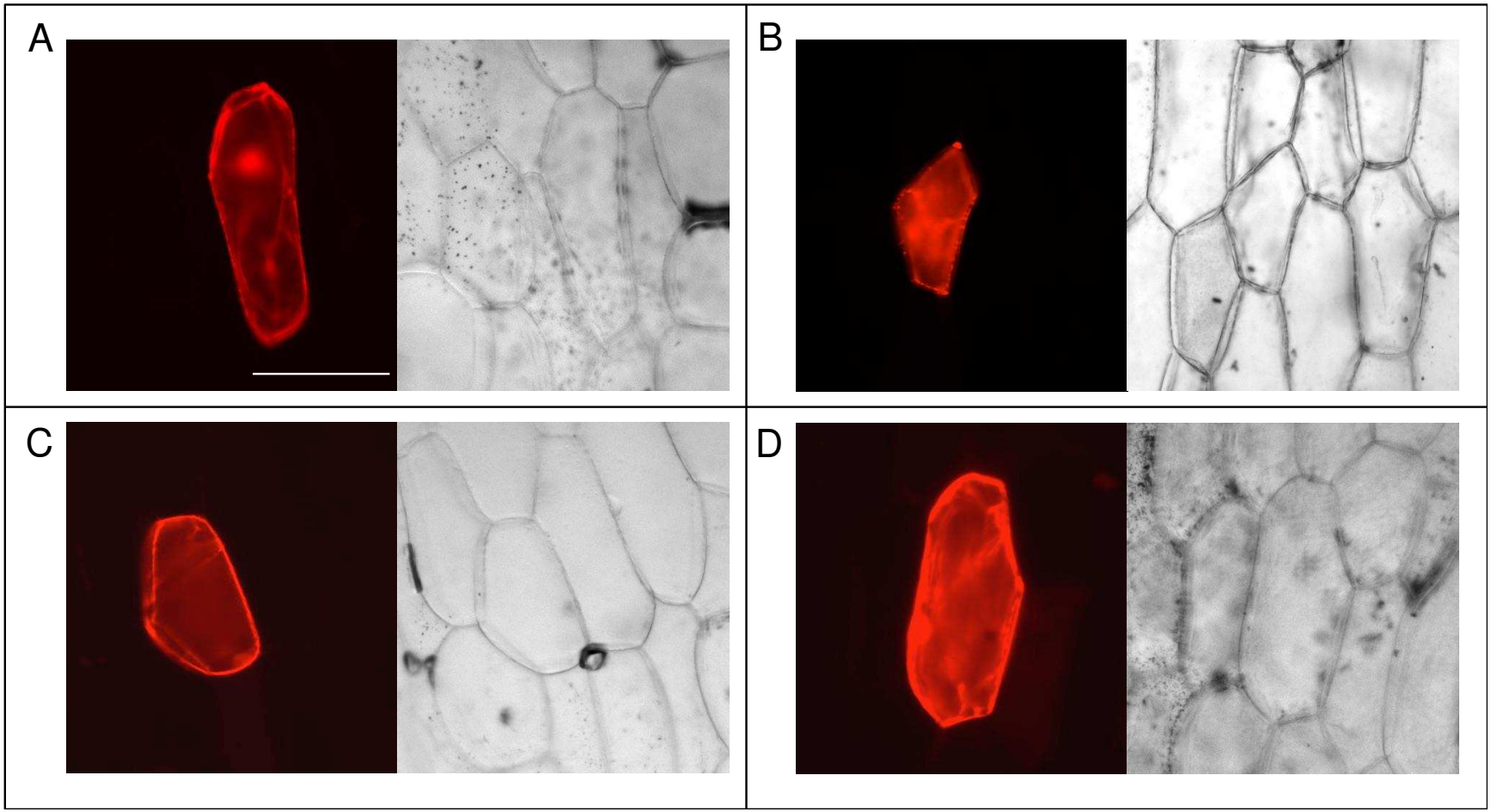

Figure 32. Fluorescent signal from tdTomato in transiently transformed onion epidermal cells.
Onion epidermis was transformed by bombardment with tungsten particles coated with A) pBS-cyto-tdT; B) pBS-sec-tdT; C) pBS-sec-tdT-short; D) pBS-sec-tdT-long and imaged after transformed cells recovered in the dark at room temperature for 63 hours. Exposures for all fluorescent images were 250 ms. Camera white and black points were set at their respective maxima. Scale bar, 100 μm

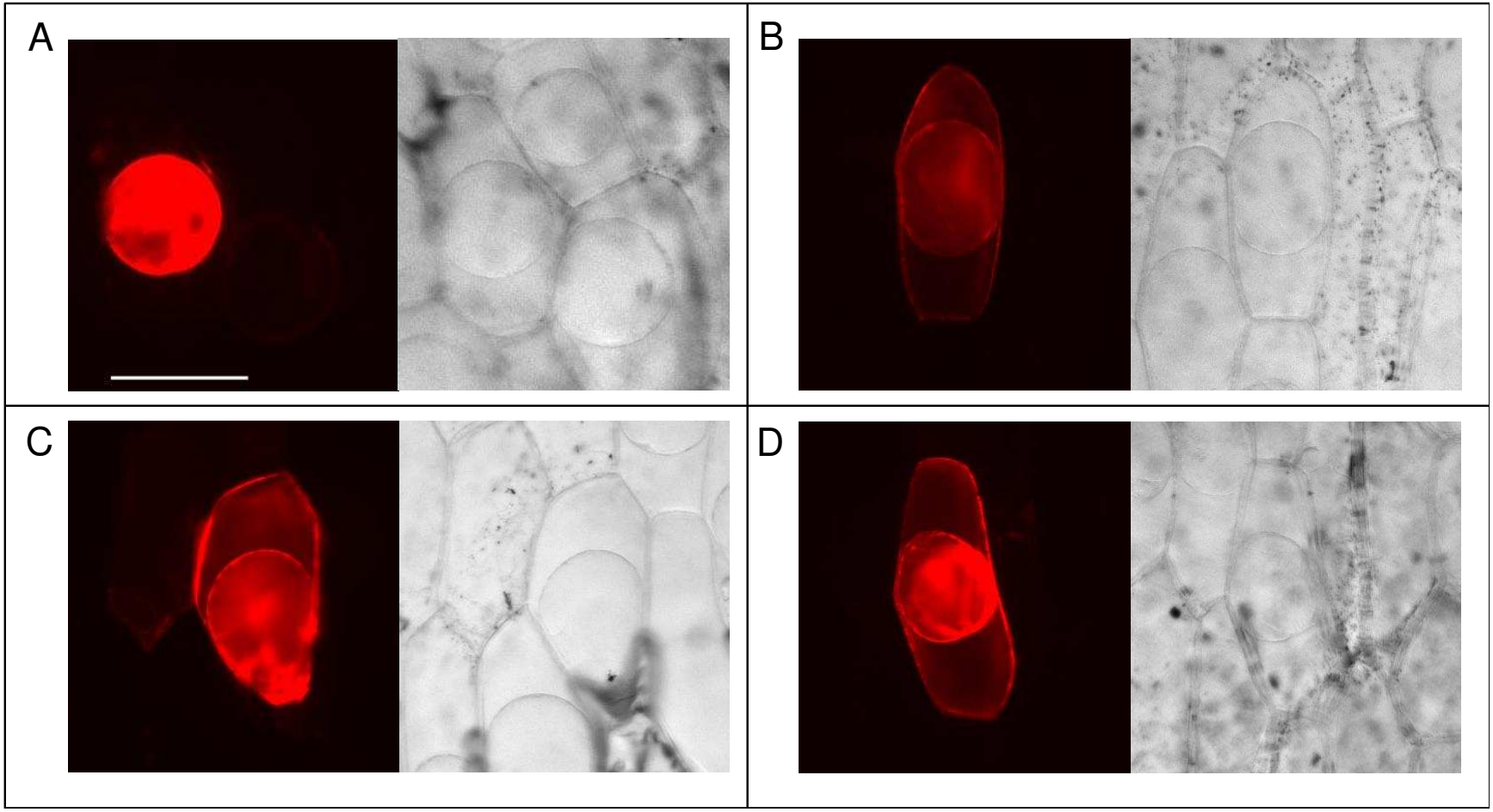

Figure 33. Subcellular localization of tdTomato in transiently transformed onion epidermal cells.
Onion epidermis was transformed by bombardment with tungsten particles coated with A) pBS-cyto-tdT; B) pBS-sec-tdT; C) pBS-sec-tdT-short; D) pBS-sec-tdT-long and imaged after transformed cells first recovered in the dark at room temperature for 63 hours to permit sufficient secretion of fluorescent protein and then were plasmolyzed in 1 M NaCl. Exposures for all fluorescent images were 250 ms. Camera white and black points were set at their respective maxima. Scale bar, 100 μm.

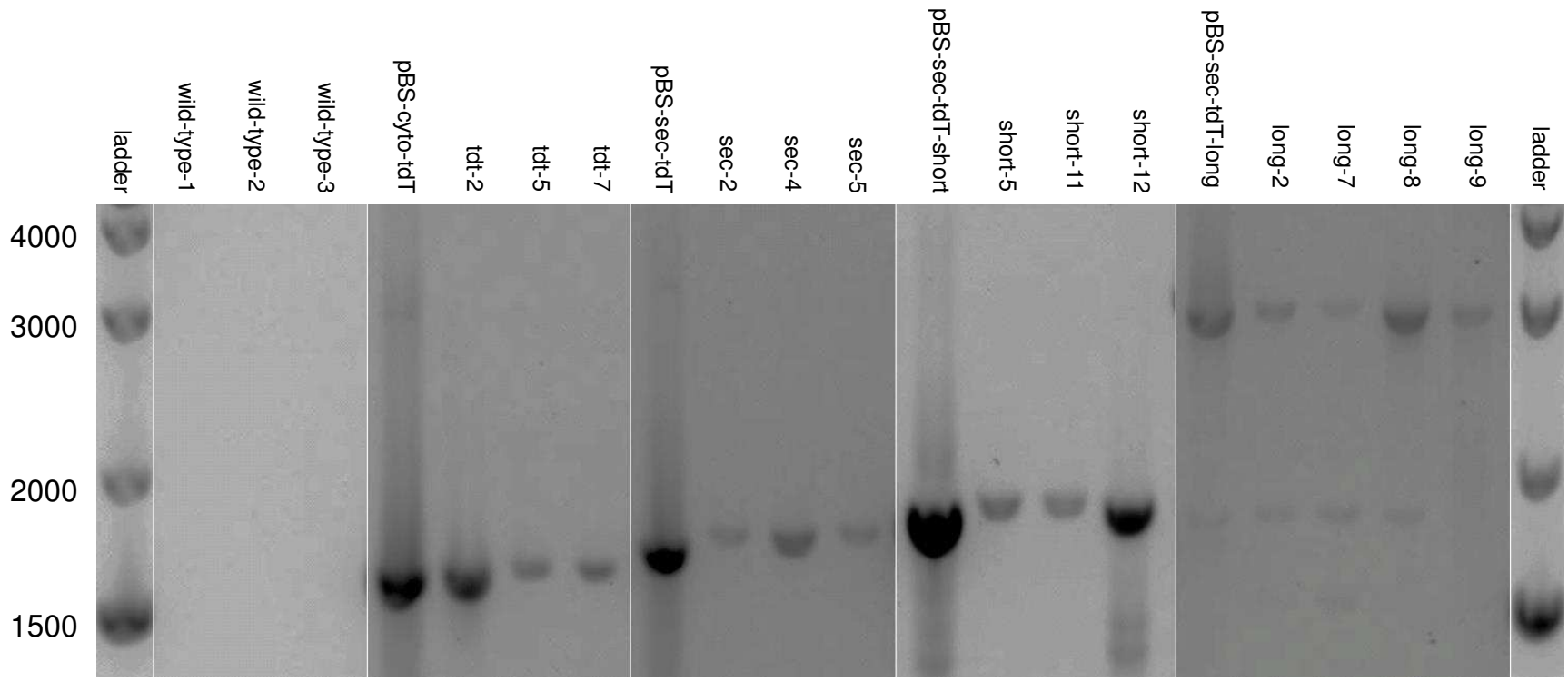

Figure 34. Genotyping of transgenic Arabidopsis plants transformed with each of the four constructs using primers 35S and OCS. PCR products were separated on a 1% agarose gel

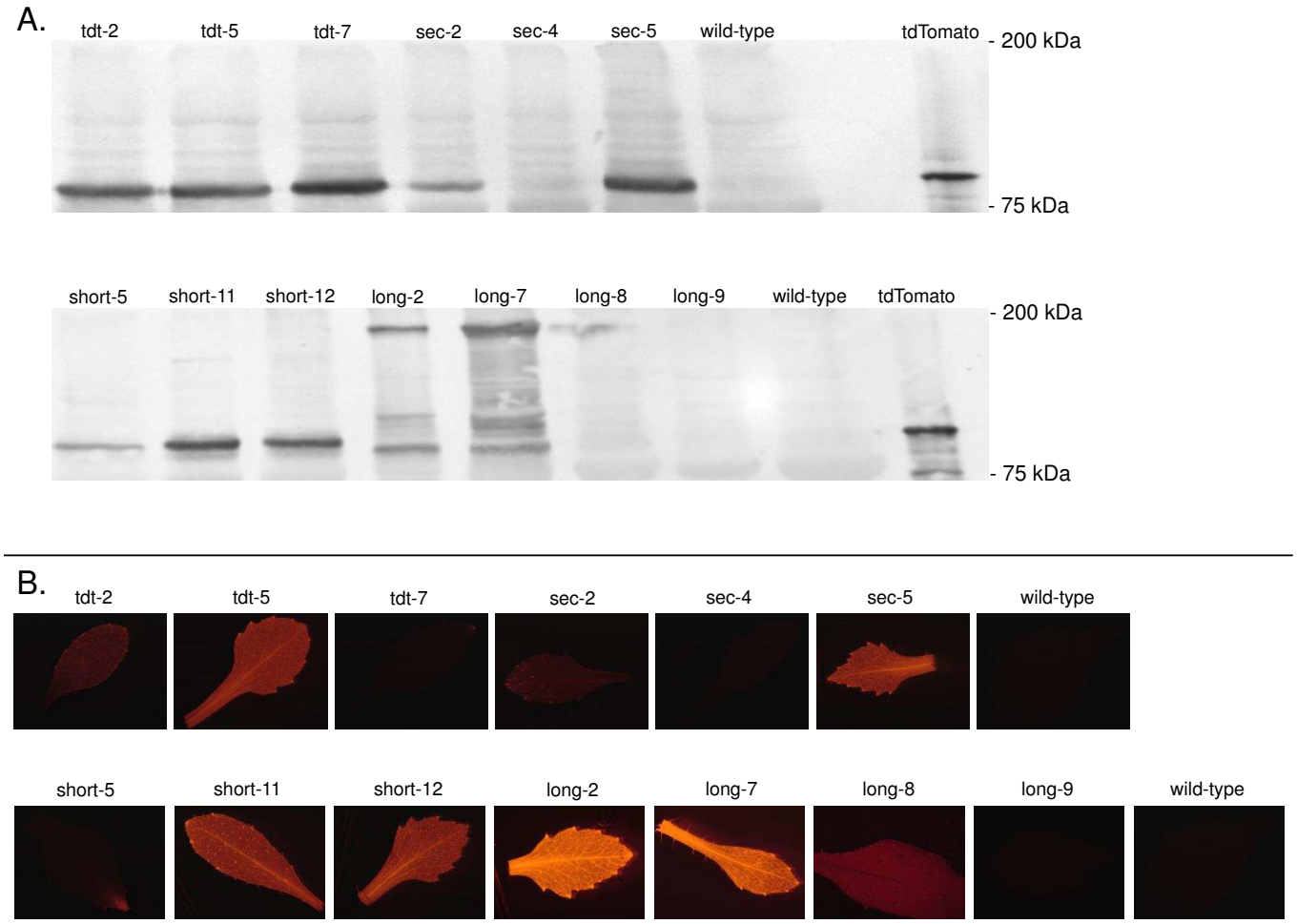

Figure 35. Expression of tdTomato in homozygous Arabidopsis lines

In A), 10 L of crude protein extract or a 1:9 dilution of purified tdTomato in water were loaded on a 12% polyacrylamide gel and electrophoresed at 200 V for 1 hour. After transfer to PVDF, tdTomato bands were visualized by blotting with rabbit anti-RFP that had been pre-incubated in wild-type Arabidopsis protein extract, followed by development with alkaline phosphatase.

In B), leaves from the same plants used in the Western blot were imaged directly by fluorescence microscopy. All images used an exposure of 100 ms; the white and black points of the camera were set to their respective maxima.

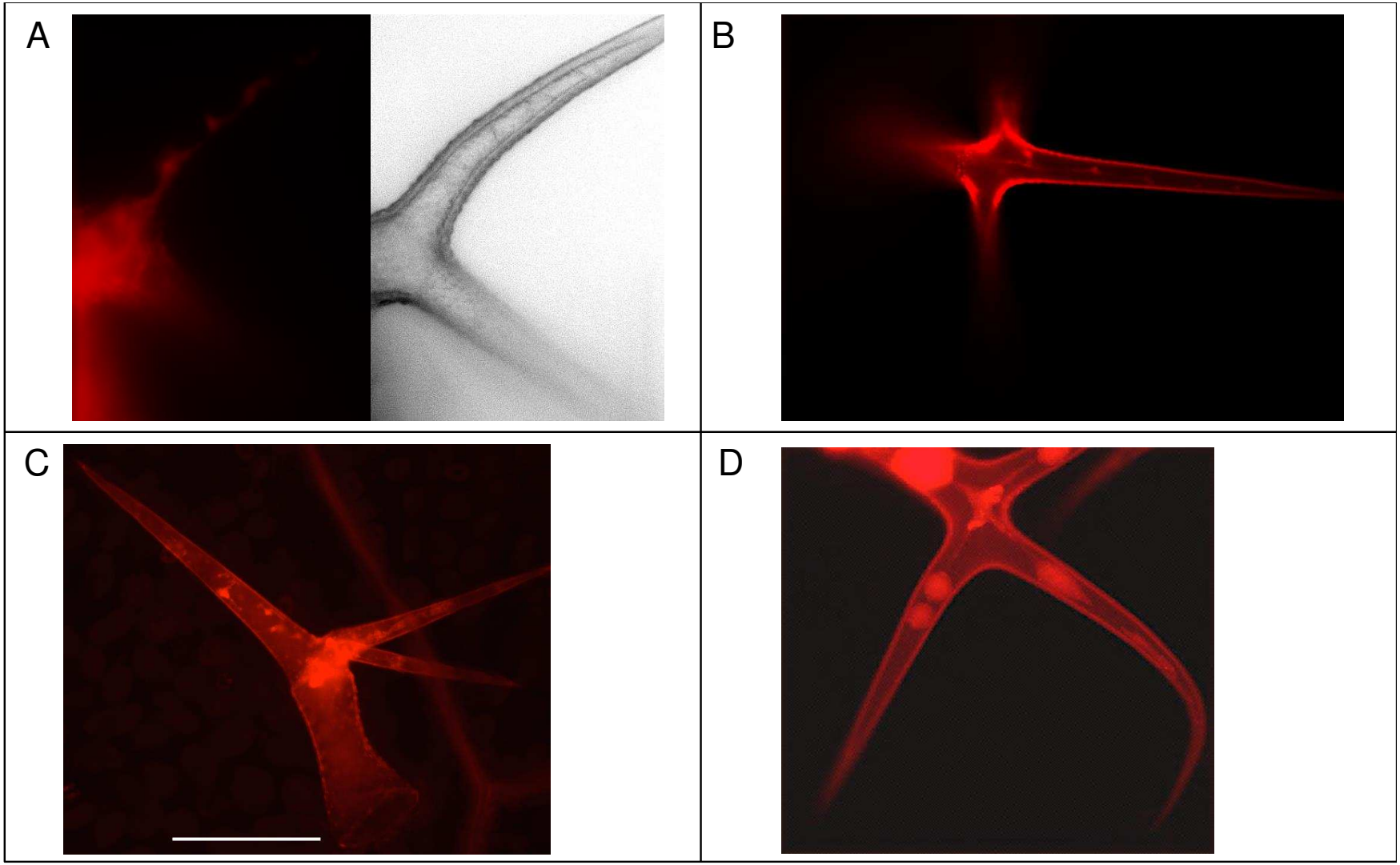

Figure 36. Subcellular localization of tdTomato in stably transformed Arabidopsis.
Leaves of lines A) tdT-5; B) sec-5; C) short-5; D) long-7 were plasmolyzed in 1 M NaCl and trichomes were imaged. Fluorescence images used a 250 ms exposure. White and black points for the camera were set at their respective maxima. Scale bar, 100 μm.

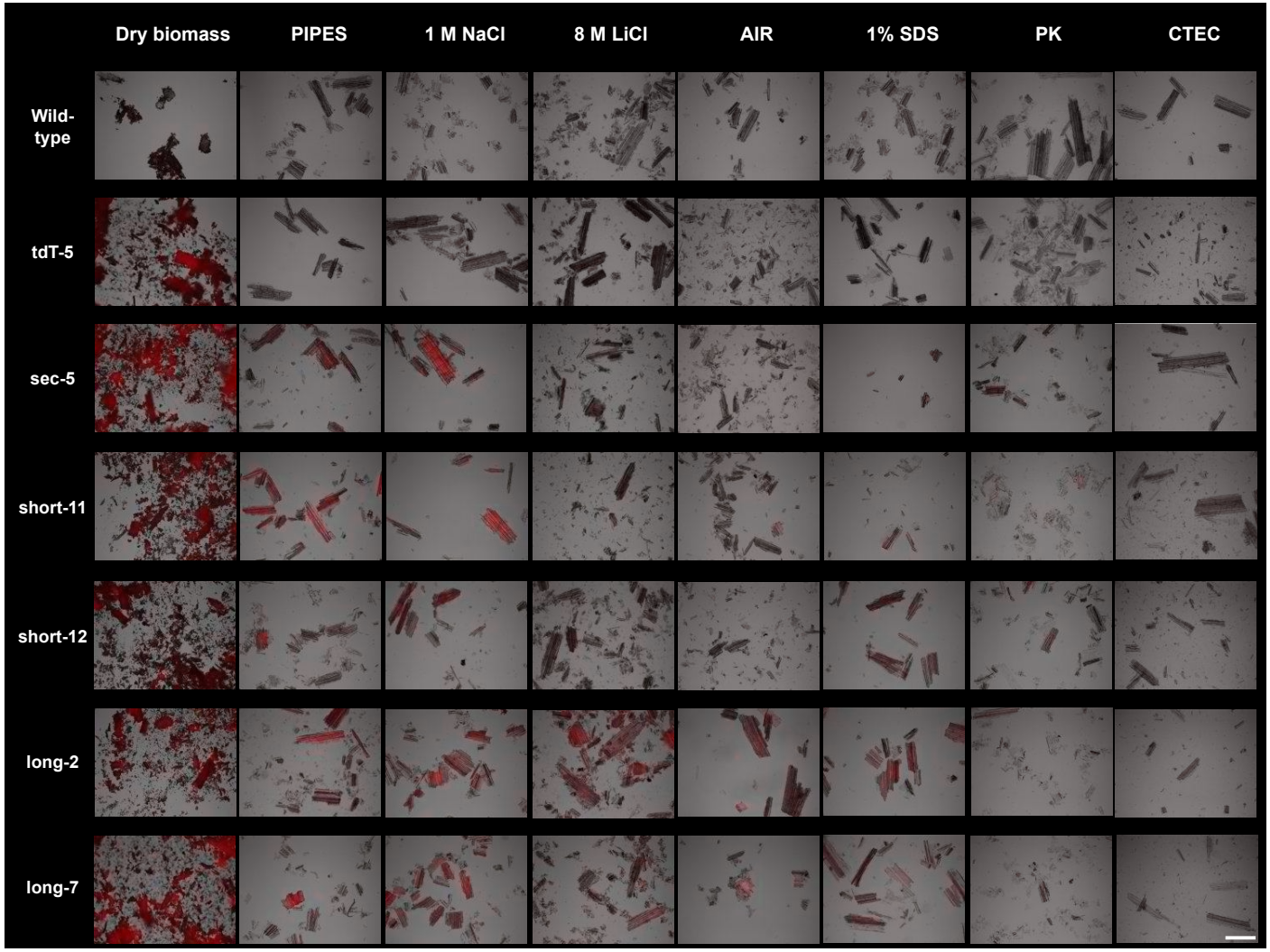

Figure 37. Fluorescence in transgenic and wild-type Arabidopsis cauline stem biomass after treatment with various solutions. All fluorescence images used a 100 ms exposure, the same objective lens (10x), and the same camera settings. Scale bar, 250 μm (in lower right image). Figure was made by overlaying the bright-field and fluorescent images, and setting the transparency of the fluorescent image to 50%.

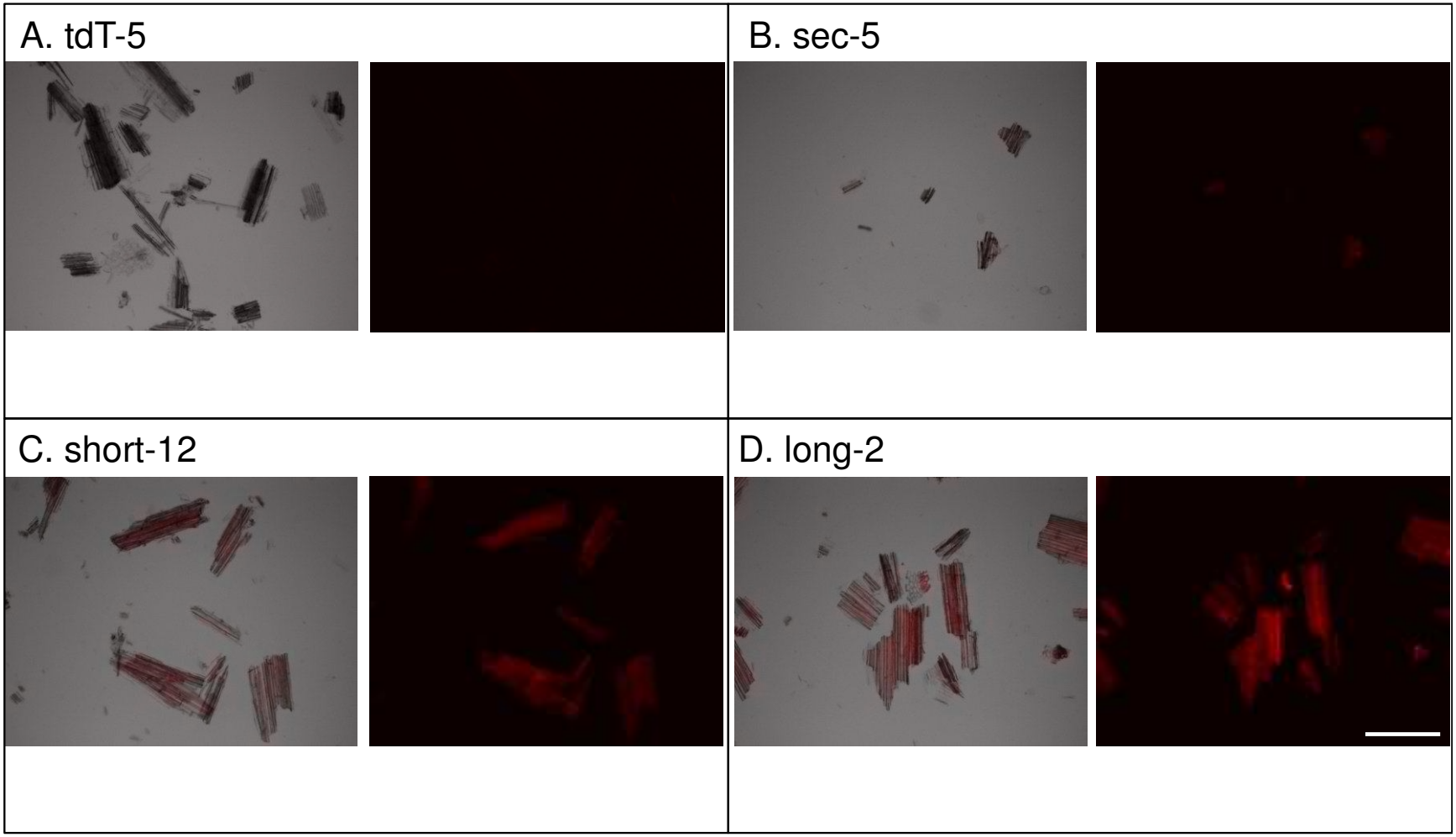

Figure 38. Enlargement of images from Figure 37, 1% SDS.
For each genotype, left, overlay of 50% transparent fluorescent image and bright-field image; right, fluorescent image. Scale bar, 250 μm (in lower right image).

Table 7. Summary of results from fluorescence dissociation experiments.
Each + indicates more fluorescence observed, on average. -, no fluorescence observed. +/-, fluorescence observed only in isolated chunks of biomass.

	Dry biomass	PIPES	1 M NaCl	8 M LiCl	AIR	1% SDS	PK	CTEC
Wild-type	-	-	-	-	-	-	-	-
tdT-5	+++	-	-	-	-	-	-	-
sec-5	+++	++	++	+/-	-	-	+/-	-
short-11	+++	++	++	+/-	+	+	+/-	-
short-12	+++	++	++	+/-	-	+	+/-	-
long-2	+++	++	++	++	+	++	+/-	-
long-7	+++	++	++	++	+	++	+/-	-

Table 8. Summary of transgenic Arabidopsis phenotypic data.
Data shown are averages ± standard error. Shaded cells, $p < 0.05$ (Student's T-test) in comparison to wild-type.

	Wild-type	tdT-2	tdT-5	tdT-7	sec-2	sec-4	sec-5	short-5	short-11	short-12	long-2	long-7	long-8	long-9
days to bolting	25.8 ± 1.3	25.7 ± 1.5	23.0 ± 0.9	25.0 ± 1.3	23 ± 0.6	21.5 ± 0.7	21.8 ± 0.7	22.7 ± 0.7	20.8 ± 0.6	19.5 ± 0.3	23.3 ± 0.7	25.0 ± 0	19.7 ± 0.3	25.0 ± 0.5
days to flowering	28.8 ± 1.9	31.3 ± 2.2	26.7 ± 0.9	29.7 ± 2.3	26.0 ± 0.7	24.5 ± 1.1	24.0 ± 0.8	26.2 ± 0.9	22.0 ± 0.7	21.8 ± 0.6	26.5 ± 0.9	30.0 ± 0	22.0 ± 0.6	27.3 ± 0.9
days to harvest	48.3 ± 0.5	53.0 ± 0.9	51.0 ± 5.0	51.3 ± 2.0	49.3 ± 1.0	47.3 ± 2.8	49.0 ± 0.7	48.3 ± 1.9	43.0 ± 1.8	44.0 ± 3.3	50.0 ± 1.9	48.0 ± 0	46.3 ± 2.0	49.0 ± 4.3
# of siliques at harvest	5.3 ± 1.7	5.3 ± 1.5	21.3 ± 0.6	10.7 ± 1.2	***12.5 ± 0.8***	15.5 ± 1.6	6.8 ± 1.4	10.6 ± 0.3	8.5 ± 1.2	18.3 ± 1.4	13.0 ± 0.5	5.0 ± 0.5	12.3 ± 1.3	17.0 ± 0
stem thickness (mm)	1.2 ± 0.04	1.4 ± 0.10	***1.6 ± 0.04***	1.4 ± 0.07	***1.5 ± 0.03***	1.4 ± 0.03	1.2 ± 0.02	1.4 ± 0.02	1.0 ± 0.05	1.2 ± 0.03	1.3 ± 0.03	1.1 ± 0.03	1.1 ± 0.07	***1.6 ± 0.06***
inflorescence length (mm)	324 ± 12	352 ± 10	361 ± 6	370 ± 12	368 ± 9	343 ± 5	424 ± 17	357 ± 3	278 ± 8	***393 ± 2***	341 ± 11	330 ± 10	329 ± 4	388 ± 13
cauline stem length (mm)	112 ± 4	128 ± 7	1201 ± 2	153 ± 18	123 ± 7	117 ± 4	112 ± 5	125 ± 2	95 ± 3	114 ± 4	124 ± 6	114 ± 4	91 ± 3	147 ± 6
# of bolts	3.5 ± 0.4	3.0 ± 0.3	3.3 ± 0.6	4.7 ± 0.2	4.5 ± 0.3	4.8 ± 0.2	3.5 ± 0.5	4.0 ± 0.4	2.8 ± 0.1	3.5 ± 0.3	4.5 ± 0.1	3.5 ± 0.3	4.0 ± 0.3	4.3 ± 0.2
# of branches off main bolt	3.3 ± 0.1	4.0 ± 0.3	***4.5 ± 0.1***	5.0 ± 0.5	3.8 ± 0.2	3.0 ± 0.2	3.0 ± 0	4.0 ± 0.1	3.0 ± 0.2	3.5 ± 0.1	3.8 ± 0.1	3.0 ± 0	2.7 ± 0.2	4.7 ± 0.3
inflorescence wet weight (g)	1.78 ± 0.17	2.72 ± 0.25	3.33 ± 0.39	2.82 ± 0.28	***3.66 ± 0.06***	***3.11 ± 0.17***	2.57 ± 0.18	2.61 ± 0.25	1.09 ± 0.06	2.63 ± 0.21	2.17 ± 0.12	1.39 ± 0.01	1.88 ± 0.09	***2.94 ± 0.05***
cauline stem wet weight (g)	0.62 ± 0.07	0.93 ± 0.12	1.08 ± 0.11	0.88 ± 0.09	***1.13 ± 0.06***	1.00 ± 0.06	0.85 ± 0.08	0.89 ± 0.08	0.31 ± 0.01	0.80 ± 0.08	0.79 ± 0.06	0.48 ± 0.01	0.58 ± 0.03	1.01 ± 0.02
cauline stem dry weight (mg)	47 ± 4	100 ± 14	130 ± 16	87 ± 9	***130 ± 11***	***97 ± 6***	89 ± 8	85 ± 6	32 ± 2	87 ± 8	68 ± 6	44 ± 0	57 ± 2	***112 ± 3***
average glucose release (%/mg dry weight)	9.1 ± 0.1	8.9 ± 0.1	9.4 ± 0.1	7.3 ± 0.1	8.9 ± 0.1	9.7 ± 0.1	***8.0 ± 0.2***	***7.9 ± 0.3***	8.8 ± 0.1	9.5 ± 0.3	8.5 ± 0.3	***10.2 ± 0.02***	***10.6 ± 0.2***	9.9 ± 1.1
average Hyp content (%/mg dry weight)	0.047	0.052	0.041	0.051	0.047	0.047	0.048	0.045	0.044	0.041	***0.109***	0.140	0.054	0.041
# of plants measured	3	3	4	3	4	4	4	4	4	4	3	2	3	2

Figure 39. Five-week-old Arabidopsis plants in four-inch diameter pots. Numbers in white boxes to the right of each plant indicate the line number.

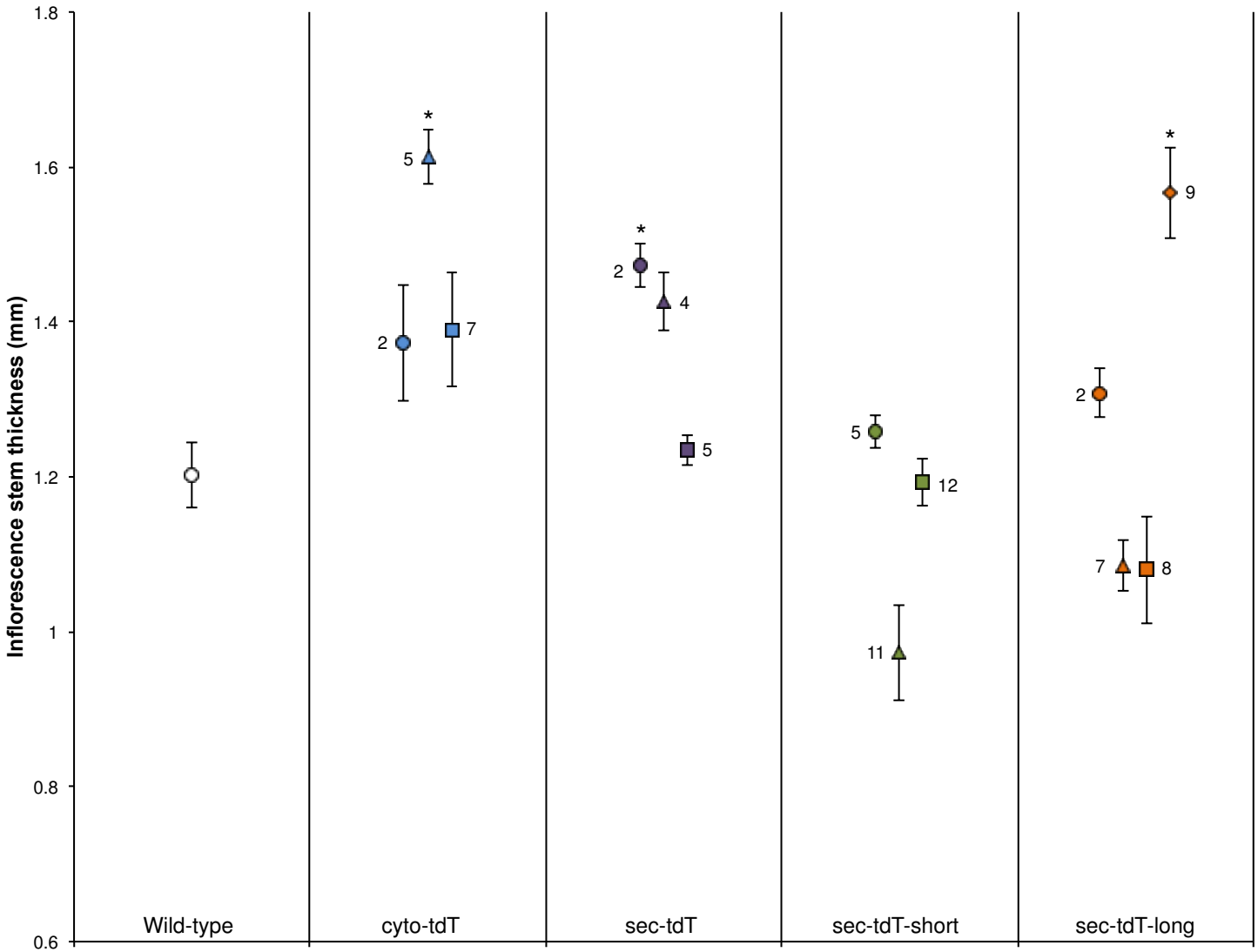

Figure 40. Average inflorescence stem thickness in independently transformed Arabidopsis lines expressing each transgene. Error bars indicate standard error. *, $p < 0.05$ in comparison to wild-type (Student's T-test).

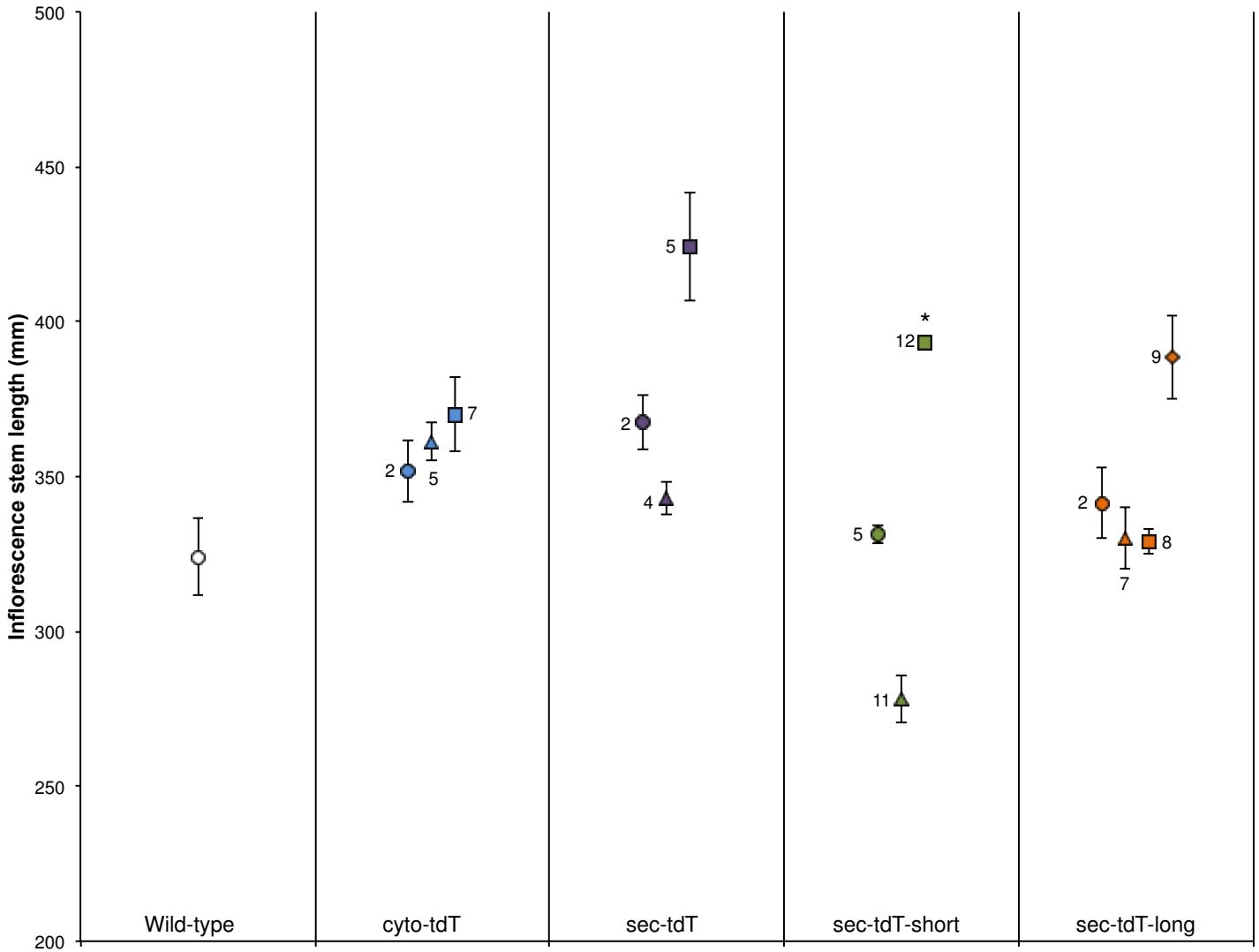

Figure 41. Average inflorescence stem length in independently transformed Arabidopsis lines expressing each transgene. Error bars indicate standard error. *, $p < 0.05$in comparison to wild-type (Student's T-test).

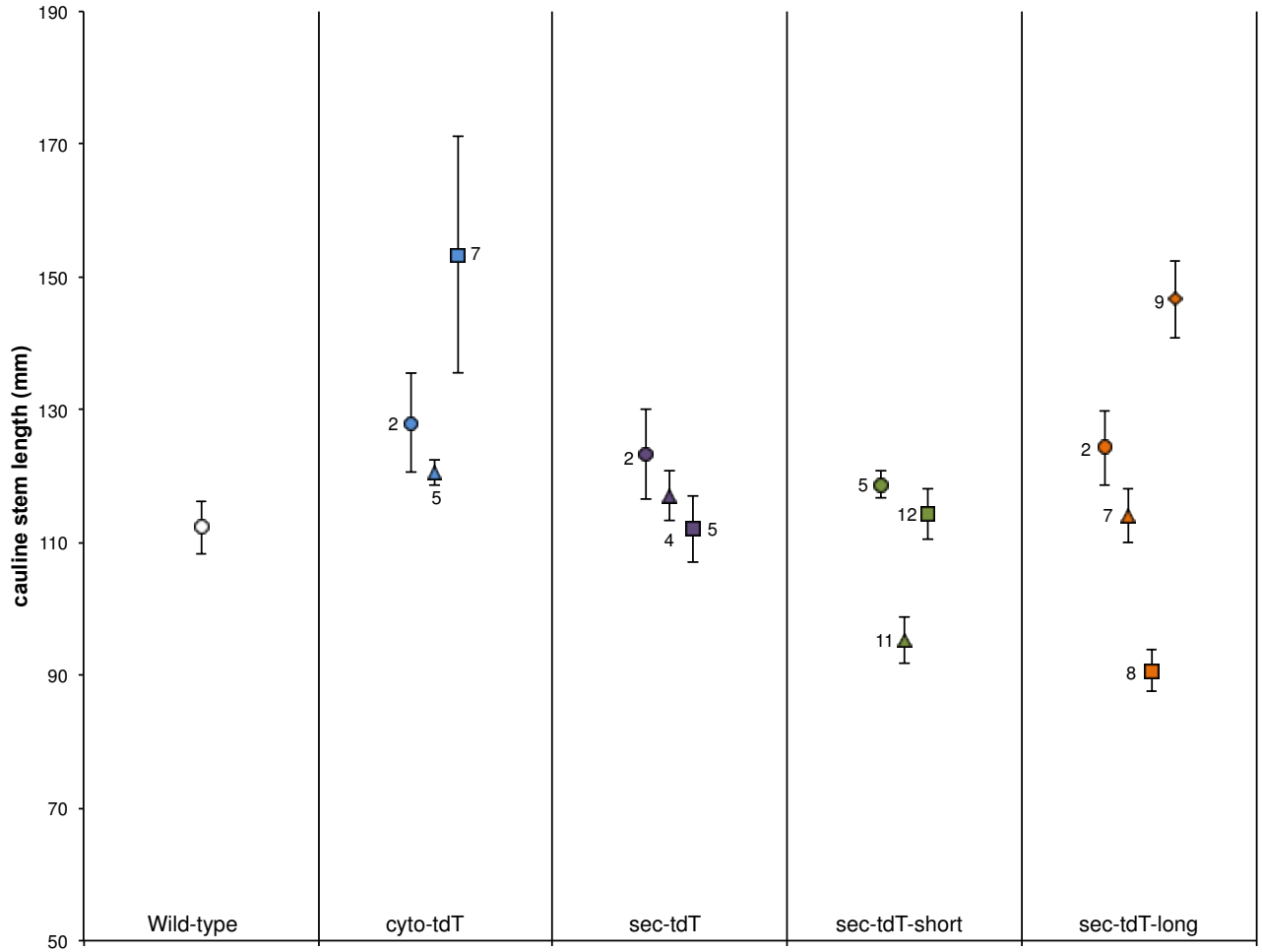

Figure 42. Average cauline stem length in independently transformed Arabidopsis lines expressing each transgene. Error bars indicate standard error.

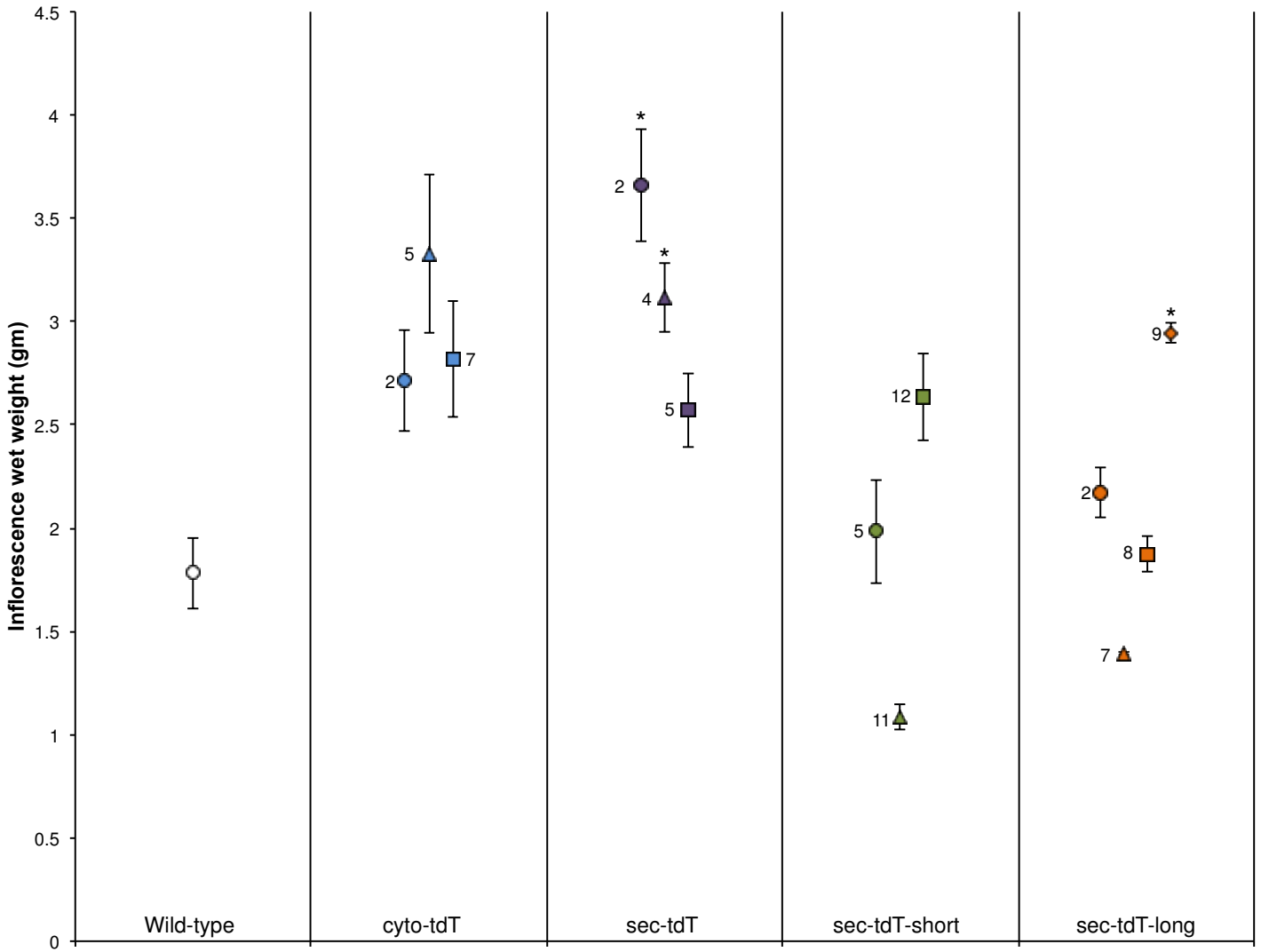

Figure 43. Average inflorescence wet weight in independently transformed Arabidopsis lines expressing each transgene. Error bars indicate standard error. *, $p < 0.05$ in comparison to wild-type (Student's T-test).

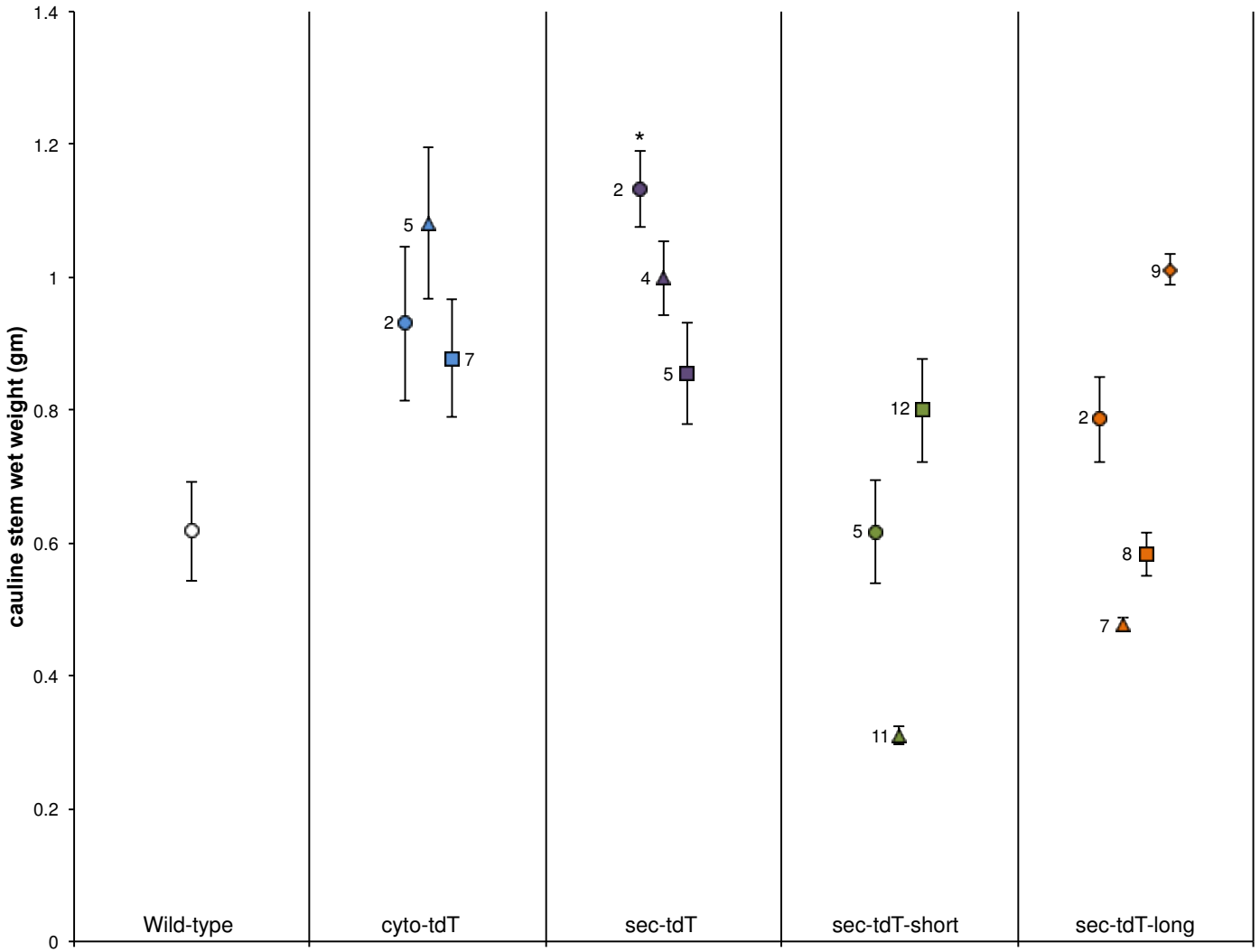

Figure 44. Average cauline stem wet weight in independently transformed Arabidopsis lines expressing each transgene. Error bars indicate standard error. *, $p < 0.05$ in comparison to wild-type (Student's T-test).

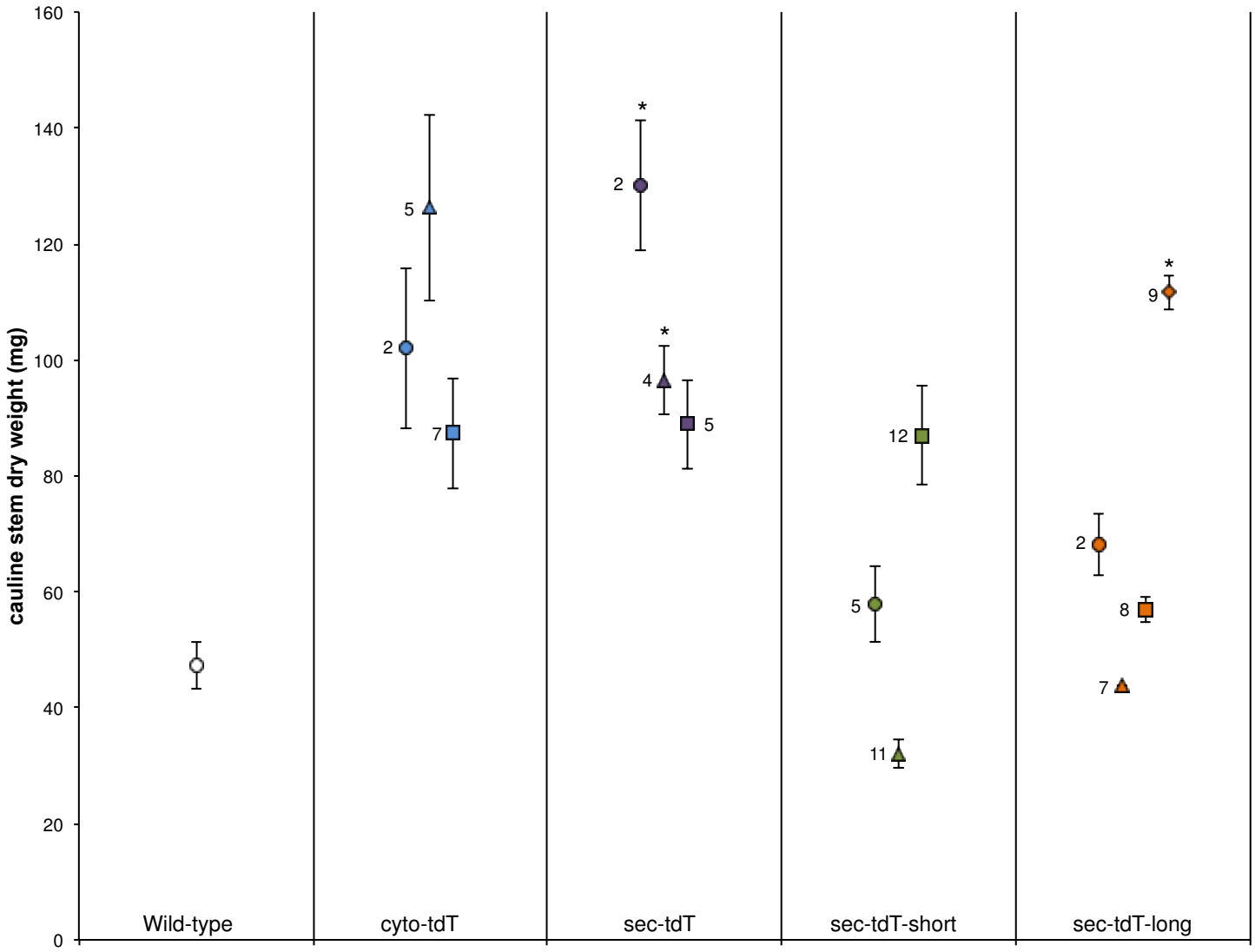

Figure 45. Average cauline stem dry weight in independently transformed Arabidopsis lines expressing each transgene. Error bars indicate standard error. *, $p < 0.05$ in comparison to wild-type (Student's T-test).

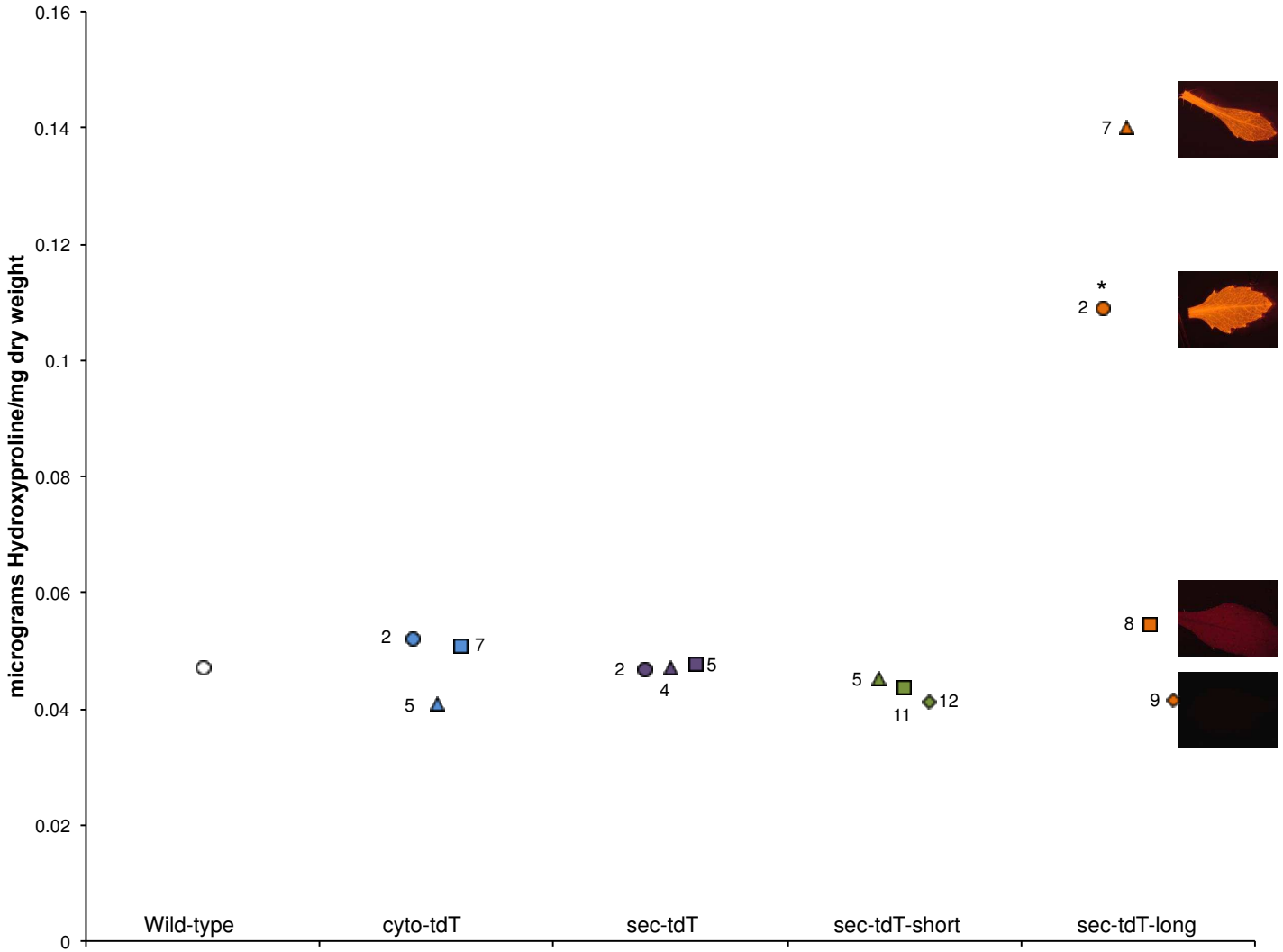

Figure 46. Average hydroxyproline content of independently transformed Arabidopsis lines expressing each transgene.
Images of fluorescent leaves for each sec-tdT-long line included for comparison of hydroxyproline content with relative fluorescence.
*, p = 0.0001 in comparison to wild-type (Student's T-test).

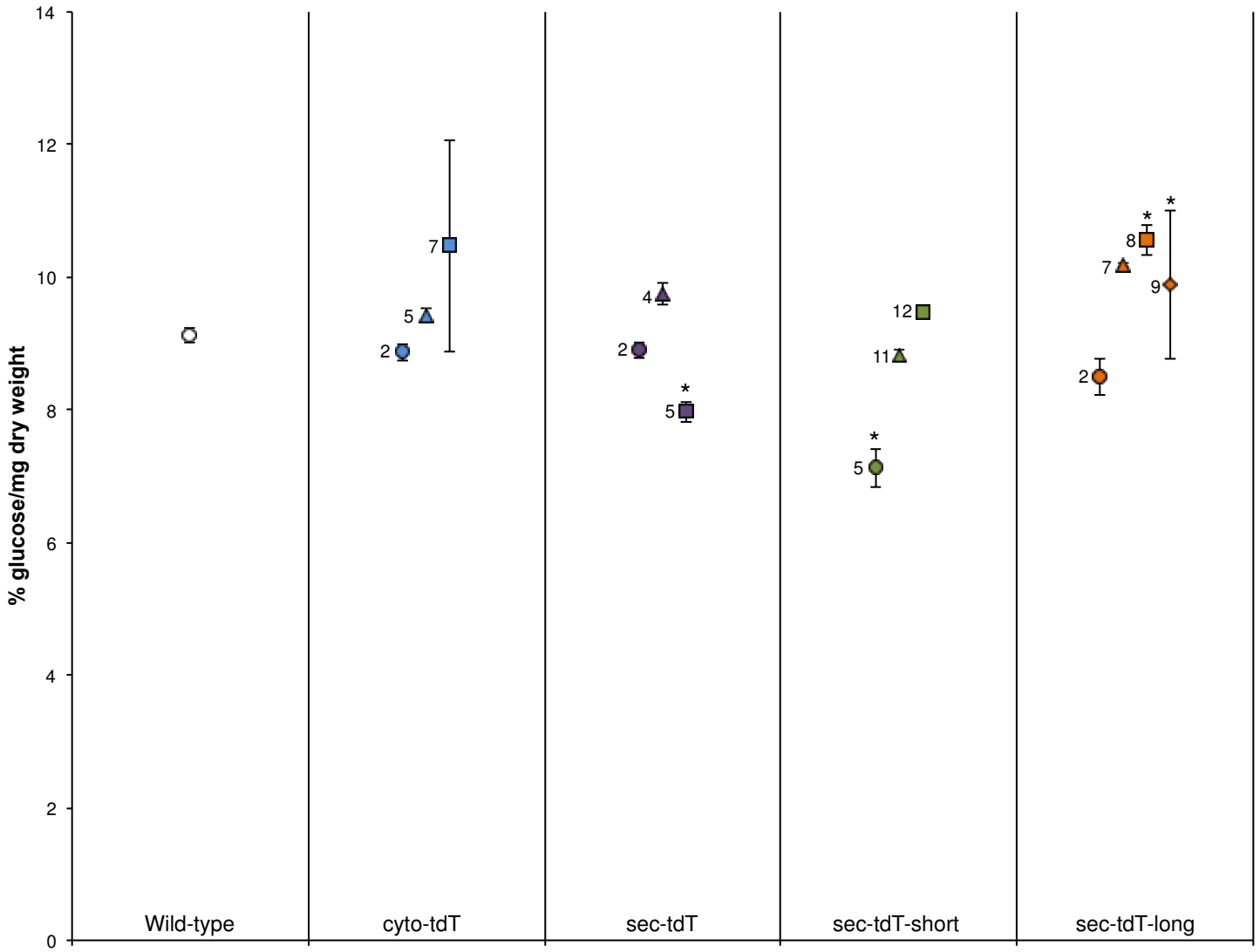

Figure 47. Average % glucose released per mg dry weight in independently transformed Arabidopsis lines expressing each transgene. Error bars indicate standard error. *, $p < 0.05$ in comparison to wild-type (Student's T-test).

CHAPTER FIVE: SUMMARY DISCUSSION

In the research performed for this dissertation, I found support for the hypothesis that extensin cross-linking increases biomass recalcitrance in some cases. In Chapter 2, I tested whether extensins could be removed from pretreated poplar biomass by protease treatment, and whether extensin removal could be associated with subsequent glucose release after cellulase digestion. I found that Fermgen protease treatment was able to remove extensins from liquid hot water pretreated poplar, but not from dilute acid or alkaline peroxide pretreated poplar. Glucose release was also increased after Fermgen treatment of liquid hot water pretreated poplar, but not dilute acid or alkaline peroxide pretreated poplar. These results indicate that, depending on the pretreatment, extensins may be a meaningful target for improving glucose yields from biomass. A residual pool of extremely tightly bound extensins in dilute acid and alkaline peroxide pretreated poplar resists protease degradation; further work to characterize methods of removing these strongly cross-linked extensins may increase glucose yields in poplar after several different types of pretreatments.

In Chapter 3, I tested whether mutations in the genes encoding the enzymes that perform extensin post-translational modifications affect cauline stem cell wall extensin content or digestibility in Arabidopsis. Previously, root hair growth defects were described in these mutant lines. I expected post-translational modification of all extensins to be reduced by the mutations, causing alterations in extensin cross-linking in cell walls and increased plant growth due to increased cell elongation. I found either no effect on plant growth phenotypes or reduced plant growth. I also found no effect on cauline stem hydroxyproline content or digestibility. Since no change in cauline stem hydroxyproline content was observed, no conclusions can be drawn from

these experiments regarding the interplay of extensins and biomass digestibility. The known defective root hair growth phenotypes as well as the high expression levels in roots of extensins and their modifying enzymes suggest that examination of roots rather than stems may prove more illuminating of extensin function in future studies.

In Chapter 4, I tested whether increased extensin abundance in cell walls caused increased biomass recalcitrance in transgenic Arabidopsis lines overexpressing tdT-extensin fusion proteins. I found that overexpression of the complete *SlLRX1* extensin domain caused robust association of tdTomato with the cell wall, indicating that the fusion protein is likely covalently bound in the cell wall like the native extensin proteins. These lines also contained 2 - 3 times more hydroxyproline in their cauline stems than wild-type plants. However, no change in biomass digestibility was observed. Again, using this system to examine roots, with their more active extensin expression and post-translational modification, may reveal more about extensin function. It may be that there is a maximum amount of recalcitrance that can be attributed to extensin cross-linking, and native extensin expression reaches this maximum. In this case, overexpression of extensins cannot further increase recalcitrance.

These first investigations do not negate the hypothesis that extensins are involved in biomass recalcitrance. The transgenic lines described in Chapter 4 will prove especially useful for further work to demonstrate the nature of extensin cross-links more precisely, as well as to indicate other methods that may effectively remove extensins from cell walls. Addressing extensin cross-linking may be key to decreasing biomass recalcitrance, helping to move biofuels from the realm of laboratories to that of day-to-day life.

REFERENCES

1. Wyman, C. E. Biomass ethanol: technical progress, opportunities, and commercial challenges. *Annu. Rev. Energy Environ.* **24,** 189–226 (1999).

2. Béguin, P. & Aubert, J. P. The biological degradation of cellulose. *FEMS Microbiol. Rev.* **13,** 25–58 (1994).

3. U. S. DOE. Breaking the biological barriers to cellulosic ethanol: a joint research agenda. *U.S. Department of Energy Office of Science and Office of Energy Efficiency and Renewable Energy* (2006), DOE/SC-0095. At www.doegenomestolife.org/biofuels/

4. Kumar, P., Barrett, D. M., Delwiche, M. J. & Stroeve, P. Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. *Ind. Eng. Chem. Res.* **48,** 3713–3729 (2009).

5. Jeoh, T. *et al.* Cellulase digestibility of pretreated biomass is limited by cellulose accessibility. *Biotechnol. Bioeng.* **98,** 112–22 (2007).

6. Mosier, N., Hendrickson, R., Ho, N., Sedlak, M. & Ladisch, M. Optimization of pH controlled liquid hot water pretreatment of corn stover. *Bioresour. Technol.* **96,** 1986–1993 (2005).

7. Selig, M., Vinzant, T., Himmel, M. & Decker, S. The effect of lignin removal by alkaline peroxide pretreatment on the susceptibility of corn stover to purified cellulolytic and xylanolytic enzymes. *Appl. Biochem. Biotechnol.* **155,** 397–406 (2009).

8. Olson, A. C. Proteins and plant cell walls. Proline to hydroxyproline in tobacco suspension cultures. *Plant Physiol.* **39,** 543–50 (1964).

9. Cooper, J. B. & Varner, J. E. Cross-linking of soluble extensin in isolated cell walls. *Plant Physiol.* **76,** 414–417 (1984).

10. Cooper, J. B., Chen, J. A., van Holst, G.-J. & Varner, J. E. Hydroxyproline-rich glycoproteins of plant cell walls. *TIBS* **12,** 24–27 (1987).

11. Tierney, M. L. & Varner, J. E. The extensins. *Plant Physiol.* **84,** 1–2 (1987).

12. Lamport, D. T. A., Kieliszewski, M. J., Chen, Y. & Cannon, M. C. Role of the extensin superfamily in primary cell wall architecture. *Plant Physiol.* **156,** 11–19 (2011).

13. Epstein, L. & Lamport, D. T. A. An intramolecular linkage involving isodityrosine in extensin. *Phytochemistry* **23,** 1241–1246 (1984).

14. Stafstrom, J. P. & Staehelin, L. A. The role of carbohydrate in maintaining extensin in an extended conformation. *Plant Physiol.* **81,** 242–246 (1986).

15. Fry, S. C. Isodityrosine, a new cross-linking amino acid from plant cell-wall glycoprotein. *Biochem. J.* **204,** 449–455 (1982).

16. Brady, J. D., Sadler, I. H. & Fry, S. C. Di-isodityrosine, a novel tetrameric derivative of tyrosine in plant cell wall proteins: a new potential cross-link. *Biochem. J.* **315,** 323–327 (1996).

17. Cegelski, L. *et al.* Plant cell-wall cross-links by REDOR NMR spectroscopy. *J. Am. Chem. Soc.* **132,** 16052–16057 (2010).

18. Qi, X., Behrens, B. X., West, P. R. & Mort, A. J. Solubilization and partial characterization of extensin fragments from cell walls of cotton suspension cultures. Evidence for a covalent cross-link between extensin and pectin. *Plant Physiol.* **108,** 1691–701 (1995).

19. Nuñez, A., Fishman, M. L., Fortis, L. L., Cooke, P. H. & Hotchkiss, A. T. Identification of extensin protein associated with sugar beet pectin. *J. Agric. Food Chem.* **57,** 10951–10958 (2009).

20. Lange, B. M., Lapierre, C. & Sandermann, H. Elicitor-induced spruce stress lignin (structural similarity to early developmental lignins). *Plant Physiol.* **108,** 1277–1287 (1995).

21. Bao, W., O'Malley, D. M. & Sederoff, R. R. Wood contains a cell-wall structural protein. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 6604–6608 (1992).

22. Moore, T. S. An extracellular macromolecular complex from the surface of soybean suspension cultures. *Plant Physiol.* **51,** 529–536 (1973).

23. Velasquez, S. M. *et al.* O-glycosylated cell wall proteins are essential in root hair growth. *Science.* **332,** 1401–1403 (2011).

24. Cannon, M. C. *et al.* Self-assembly of the plant cell wall requires an extensin scaffold. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 2226–31 (2008).

25. Ito, M., Kodama, H., Komamine, A. & Watanabe, A. Expression of extensin genes is dependent on the stage of the cell cycle and cell proliferation in suspension-cultured Catharanthus roseus cells. *Plant Mol. Biol.* **36,** 343–351 (1998).

26. Keller, B. & Lamb, C. J. Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation. *Genes Dev.* **3,** 1639–1646 (1989).

27. Sadava, D. & Chrispeels, M. J. Hydroxyproline-rich cell wall protein (extensin): role in the cessation of elongation in excised pea epicotyls. *Dev. Biol.* **30,** 49–55 (1973).

28. Merkouropoulos, G. & Shirsat, A. H. The unusual Arabidopsis extensin gene atExt1 is expressed throughout plant development and is induced by a variety of biotic and abiotic stresses. *Planta* **217,** 356–366 (2003).

29. Bradley, D. J., Kjellbom, P. & Lamb, C. J. Elicitor- and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: A novel, rapid defense response. *Cell* **70,** 21–30 (1992).

30. Jackson, P. A. *et al.* Rapid deposition of extensin during the elicitation of grapevine callus cultures is specifically catalyzed by a 40-kilodalton peroxidase. *Plant Physiol.* **127,** 1065–1076 (2001).

31. Tiré, C. *et al.* Extensin gene expression is induced by mechanical stimuli leading to local cell wall strengthening in Nicotiana plumbaginifolia. *Planta* **195,** 175–181 (1994).

32. Shirsat, A. H., Bell, A., Spence, J. & Harris, J. N. The Brassica napus extA extensin gene is expressed in regions of the plant subject to tensile stresses. *Planta* **199,** 618–624 (1996).

33. Showalter, A. M., Keppler, B., Lichtenberg, J., Gu, D. & Welch, L. R. A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol.* **153,** 485–513 (2010).

34. Estévez, J. M., Kieliszewski, M. J., Khitrov, N. & Somerville, C. Characterization of synthetic hydroxyproline-rich proteoglycans with arabinogalactan protein and extensin motifs in Arabidopsis. *Plant Physiol.* **142,** 458–70 (2006).

35. Mort, A. J. & Lamport, D. T. A. Anhydrous hydrogen fluoride deglycosylates glycoproteins. **309,** 289–309 (1977).

36. Lamport, D. T. A. in *The Biochemistry of Plants, Vol. 3* (ed. Stumpf, P.) 501–541 (1980).

37. Heath, M. & Northcote, D. H. Glycoprotein of the wall of sycamore tissue-culture cells. *Biochem. J.* **125,** 953–961 (1971).

38. Pokora, A. R. & Johnson, M. A. Protease catalyzed treatments of lignocellulose materials. US Patent 5374555 (1994).

39. Zhu, J. Y. & Pan, X. J. Woody biomass pretreatment for cellulosic ethanol production: Technology and energy consumption evaluation. *Bioresour. Technol.* **101,** 4992–5002 (2010).

40. Carpita, N. & McCann, M. C. in *Biochemistry and Molecular Biology of Plants* (eds. Buchanan, B., Gruissem, W. & Jones, R.) 52–108 (American Society of Plant Physiologists, 2000).

41. Liang, H. *et al.* Improved sugar release from lignocellulosic material by introducing a tyrosine-rich cell wall peptide gene in poplar. *CLEAN - Soil, Air, Water* **36,** 662–668 (2008).

42. Himmel, M. E. *Biomass recalcitrance: deconstructing the plant cell wall for bioenergy*. (Blackwell Publishing Ltd., 2009). doi:10.1002/9781444305418

43. Chaturvedi, V. & Verma, P. An overview of key pretreatment processes employed for bioconversion of lignocellulosic biomass into biofuels and value added products. *3 Biotech* **3,** 415–431 (2013).

44. Yang, B. & Wyman, C. E. Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuels, Bioprod. Biorefining* **2,** 26–40 (2008).

45. Showalter, A. Structure and function of plant cell wall proteins. *Plant Cell* **5,** 9–23 (1993).

46. Lamport, D. T. A. The protein component of primary cell walls. *Adv. Bot. Res.* **2,** 151–218 (1965).

47. Kieliszewski, M. J., Lamport, D. T. A., Tan, L. & Cannon, M. C. Hydroxyproline-rich glycoproteins: form and function. *Annu. Plant Rev.* **41,** 321–342 (2011).

48. Lamport, D. T. A. Hydroxyproline-O-glycosidic linkage of the plant cell wall glycoprotein extensin. *Nature* **216,** 1322–1324 (1967).

49. Lamport, D. T. A. & Miller, D. H. Hydroxyproline arabinosides in the plant kingdom. *Plant Physiol.* **48,** 454–6 (1971).

50. Lamport, D. T. A., Katona, L. & Roerig, S. Galactosylserine in extensin. *Biochem. J.* **133,** 125–132 (1973).

51. Van Holst, G.-J. & Varner, J. E. Reinforced polyproline II conformation in a hydroxyproline-rich cell wall glycoprotein from carrot root. *Plant Physiol.* **74,** 247–251 (1984).

52. Brady, J. D. & Fry, S. C. Formation of di-isodityrosine and loss of isodityrosine in the cell walls of tomato cell-suspension cultures treated with fungal elicitors or H_2O_2. *Plant Physiol.* **115,** 87–92 (1997).

53. Keller, B. Structural cell wall proteins. *Plant Physiol.* **101,** 1127–1130 (1993).

54. Ribeiro, J. M. *et al.* The contribution of extensin network formation to rapid, hydrogen peroxide-mediated increases in grapevine callus wall resistance to fungal lytic enzymes. *J. Exp. Bot.* **57,** 2025–35 (2006).

55. Merkouropoulos, G., Barnett, D. C. & Shirsat, A. H. The Arabidopsis extensin gene is developmentally regulated, is induced by wounding, methyl jasmonate, abscisic and salicylic acid, and codes for a protein with unusual motifs. *Planta* **208,** 212–219 (1999).

56. Elliott, K. A. & Shirsat, A. H. Promoter regions of the extA extensin gene from Brassica napus control activation in response to wounding and tensile stress. *Plant Mol. Biol.* **37,** 675–687 (1998).

57. Guo, L. *et al.* Differential retention and expansion of the ancestral genes associated with the paleopolyploidies in modern rosid plants, as revealed by analysis of the extensins super-gene family. *BMC Genomics* **15,** 612 (2014).

58. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8,** 785–6 (2011).

59. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39,** W29–37 (2011).

60. Sjödin, A., Street, N. R., Sandberg, G., Gustafsson, P. & Jansson, S. The Populus Genome Integrative Explorer (PopGenIE): A new resource for exploring the Populus genome. *New Phytol.* **182,** 1013–1025 (2009).

61. Wilkins, O., Nahal, H., Foong, J., Provart, N. J. & Campbell, M. M. Expansion and diversification of the Populus R2R3-MYB family of transcription factors. *Plant Physiol.* **149,** 981–93 (2009).

62. Yang, X. *et al.* The F-box gene family is expanded in herbaceous annual plants relative to woody perennial plants. *Plant Physiol.* **148,** 1189–200 (2008).

63. Hefer, C. A., Mizrachi, E., Myburg, A. A., Douglas, C. J. & Mansfield, S. D. Comparative interrogation of the developing xylem transcriptomes of two wood-forming species: Populus trichocarpa and Eucalyptus grandis. *New Phytol.* (2015). doi:10.1111/nph.13277

64. Cassab, G. I. & Varner, J. E. Tissue printing on nitrocellulose paper: a new method for immunolocalization of proteins, localization of enzyme activities and anatomical analysis. *Cell Biol. Int. Rep.* **13,** 147–152 (1989).

65. Smallwood, M. *et al.* Localization of cell wall proteins in relation to the developmental anatomy of the carrot root apex. *Plant J.* **5,** 237–246 (1994).

66. Smallwood, M., Martin, H. & Knox, J. P. An epitope of rice threonine- and hydroxyproline-rich glycoprotein is common to cell wall and hydrophobic plasma-membrane glycoproteins. *Planta* **196,** 510–522 (1995).

67. Selig, M. J. *et al.* Lignocellulose recalcitrance screening by integrated high-throughput hydrothermal pretreatment and enzymatic saccharification. *Ind. Biotechnol.* **6,** 104–112 (2010).

68. Kivirikko, K. I. & Liesmaa, M. A. Colorimetric method for determination of hydroxyproline in tissue hydrolysates. *Scand. J. Clin. Lab. Invest.* **11,** 128–133 (1959).

69. Kumar, R. & Wyman, C. E. Does change in accessibility with conversion depend on both the substrate and pretreatment technology? *Bioresour. Technol.* **100,** 4193–202 (2009).

70. Santoro, N. *et al.* A high-throughput platform for screening milligram quantities of plant biomass for lignocellulose digestibility. *BioEnergy Res.* **3,** 93–102 (2010).

71. Chen, J. & Varner, J. E. An extracellular matrix protein in plants: characterization of a genomic clone for carrot extensin. *Eur. Mol. Biol. Organ. J.* **4,** 2145–2151 (1985).

72. Chen, J. & Varner, J. E. Isolation and characterization of cDNA clones for carrot extensin and a proline-rich 33-kDa protein. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 4399–4403 (1985).

73. Yoshiba, Y. *et al.* Characterization of four extensin genes in Arabidopsis thaliana by differential gene expression under stress and non-stress conditions. *DNA Res.* **8,** 115–122 (2001).

74. Stafstrom, J. P. & Staehelin, L. A. Antibody localization of extensin in cell walls of carrot storage roots. *Planta* **174,** 321–332 (1988).

75. Rioux, D. & Baayen, R. P. A suberized perimedullary reaction zone in Populus balsamifera novel for compartmentalization in trees. *Trees* **11,** 389–403 (1997).

76. Teichmann, T. *et al.* GH3::GUS reflects cell-specific developmental patterns and stress-induced changes in wood anatomy in the poplar stem. *Tree Physiol.* **28,** 1305–1315 (2008).

77. Saha, P. *et al.* Self-rescue of an EXTENSIN mutant reveals alternative gene expression programs and candidate proteins for new cell wall assembly in Arabidopsis. *Plant J.* **75,** 104–116 (2013).

78. Smith, J. J., Muldoon, E. P., Willard, J. J. & Lamport, D. T. A. Tomato extensin precursors P1 and P2 are highly periodic structures. *Phytochemistry* **25,** 1021–1030 (1986).

79. Zhou, J., Rumeau, D. & Showalter, A. M. Isolation and characterization of two wound-regulated tomato extensin genes. *Plant Mol. Biol.* **20,** 5–17 (1992).

80. Valentin, R. *et al.* Elaboration of extensin-pectin thin film model of primary plant cell wall. *Langmuir* **26,** 9891–9898 (2010).

81. Rubinstein, A. L., Broadwater, A. H., Lowrey, K. B. & Bedinger, P. A. Pex1, a pollen-specific gene with an extensin-like domain. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 3086–3090 (1995).

82. Nakhamchik, A. *et al.* A comprehensive expression analysis of the Arabidopsis proline-rich extensin-like receptor kinase gene family using bioinformatic and experimental approaches. *Plant Cell Physiol.* **45,** 1875–81 (2004).

83. Lind, J. L., Bacic, A., Clarke, A. E. & Anderson, M. A. A style-specific hydroxyproline-rich glycoprotein with properties of both extensins and arabinogalactan proteins. *Plant J.* **6,** 491–502 (1994).

84. Cassab, G. I. Plant cell wall proteins. *Annu Rev Plant Physiol Plant Mol Biol* **49,** 281–309 (1998).

85. Kieliszewski, M. J. The latest hype on Hyp- O -glycosylation codes. *Phytochemistry* **57,** 319–323 (2001).

86. Shpak, E., Barbar, E., Leykam, J. F. & Kieliszewski, M. J. Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in Nicotiana tabacum. *J. Biol. Chem.* **276,** 11272–11278 (2001).

87. Shimizu, M. *et al.* Experimental determination of proline hydroxylation and hydroxyproline arabinogalactosylation motifs in secretory proteins. *Plant J.* **42,** 877–89 (2005).

88. O'Neill, M. A. & Selvendran, R. R. Glycoproteins from the cell wall of Phaseolus coccineus. *Biochem. J.* **187,** 53-63 (1980).

89. Smith, J. J., Muldoon, E. P. & Lamport, D. T. A. Isolation of extensin precursors by direct elution of intact tomato cell suspension cultures. *Phytochemistry* **23,** 1233–1239 (1984).

90. Kieliszewski, M., De Zacks, R., Leykam, J. F. & Lamport, D. T. A. A repetitive proline-rich protein from the gymnosperm douglas fir is a hydroxyproline-rich glycoprotein. *Plant Physiol.* **98,** 919–926 (1992).

91. Dill, I., Salnikow, J. & Kraepelin, G. Hydroxyproline-rich protein material in wood and lignin of Fagus sylvatica. *Appl. Environ. Microbiol.* **48,** 1259–1261 (1984).

92. Westermark, U., Hardell, H.-L. & Iversen, T. The content of protein and pectin in the lignified middle lamella/primary wall from spruce fibers. *Holzforschung* **40,** 65–68 (1986).

93. Whitmore, F. W. Lignin-protein complex in cell walls of Pinus elliottii: Amino acid constituents. *Phytochemistry* **21,** 315–318 (1982).

94. Lamport, D. T. A. The isolation and partial characterization of hydroxyproline-rich glycopeptides obtained by enzymic degradation of primary cell walls. *Biochemistry* **8,** 1155–63 (1969).

95. Pope, D. G. Relationships between hydroxyproline-containing proteins secreted into the cell wall and medium by suspension-cultured Acer pseudoplatanus cells. *Plant Physiol.* **59,** 894–900 (1977).

96. Fong, C., Kieliszewski, M. J., De Zacks, R., Leykam, J. F. & Lamport, D. T. A. A gymnosperm extensin contains the serine-tetrahydroxyproline motif. *Plant Physiol.* **99,** 548–552 (1992).

97. Cowling, E. B. & Merrill, W. Nitrogen in wood and its role in wood deterioration. *Can. J. Bot.* **44,** 1539–1554 (1966).

98. Tanger, P. *et al.* Cell wall composition and bioenergy potential of rice straw tissues are influenced by environment, tissue type, and genotype. *BioEnergy Res.* **8,** 1165-1182 (2015).

99. Sung, W. L., Wood, M. & Huang, F. Enzymatic preparation of plant fibers. US Patent 8603802 B2 (2011).

100. Luo, X. L., Zhu, J. Y., Gleisner, R. & Zhan, H. Y. Effects of wet-pressing-induced fiber hornification on enzymatic saccharification of lignocelluloses. *Cellulose* **18,** 1055–1062 (2011).

101. Esteghlalian, A. R., Bilodeau, M., Mansfield, S. D. & Saddler, J. N. Do enzymatic hydrolyzability and Simons' stain reflect the changes in the accessibility of lignocellulosic substrates to cellulase enzymes? *Biotechnol. Prog.* **17,** 1049–1054 (2001).

102. Dow, J. M., Davies, H. A. & Daniels, M. J. A metalloprotease from Xanthomonas campestris that specifically degrades proline/hydroxyproline-rich glycoproteins of the plant extracellular matrix. *Mol. plantmicrobe Interact. MPMI* **11,** 1085–1093 (1998).

103. Feng, T. *et al.* Characterization of an extensin-modifying metalloprotease: N-terminal processing and substrate cleavage pattern of Pectobacterium carotovorum Prt1. *Appl. Microbiol. Biotechnol.* **98,** 10077–89 (2014).

104. Carlile, A. J. *et al.* Characterization of SNP1, a cell wall-degrading trypsin, produced during infection by Stagonospora nodorum. *MPMI.* **13,** 538–50 (2000).

105. Strabala, T. J. & MacMillan, C. P. The Arabidopsis wood model - The case for the inflorescence stem. *Plant Sci.* **210,** 193–205 (2013).

106. Heckman, J. W., Terhune, B. T. & Lamport, D. T. A. Characterization of native and modified extensin monomers and oligomers by electron microscopy and gel filtration. *Plant Physiol.* **86,** 848–856 (1988).

107. Velasquez, S. M. *et al.* Low sugar is not always good: Impact of specific O-glycan defects on tip growth in Arabidopsis. *Plant Physiol.* **168,** 808–813 (2015).

108. Hall, Q. & Cannon, M. C. The cell wall hydroxyproline-rich glycoprotein RSH is essential for normal embryo development in Arabidopsis. *Plant Cell* **14,** 1161–72 (2002).

109. Ringli, C. The hydroxyproline-rich glycoprotein domain of the Arabidopsis LRX1 requires Tyr for function but not for insolubilization in the cell wall. *Plant J.* **63,** 662–669 (2010).

110. Baumberger, N., Ringli, C. & Keller, B. The chimeric leucine-rich repeat/extensin cell wall protein LRX1 is required for root hair morphogenesis in Arabidopsis thaliana. *Genes Dev.* **15,** 1128–1139 (2001).

111. Vlad, F. *et al.* Arabidopsis prolyl 4-hydroxylases are differentially expressed in response to hypoxia, anoxia and mechanical wounding. *Physiol. Plant.* **130,** 471–483 (2007).

112. Velasquez, S. M. *et al.* Complex regulation of prolyl-4-hydroxylases impacts root hair expansion. *Mol. Plant* **8,** 734–746 (2014).

113. Ogawa-Ohnishi, M., Matsushita, W. & Matsubayashi, Y. Identification of three hydroxyproline O-arabinosyltransferases in Arabidopsis thaliana. *Nat. Chem. Biol.* **9,** 726–30 (2013).

114. Egelund, J. *et al.* Molecular characterization of two Arabidopsis thaliana glycosyltransferase mutants, rra1 and rra2, which have a reduced residual arabinose content in a polymer tightly associated with the cellulosic wall residue. *Plant Mol. Biol.* **64,** 439–51 (2007).

115. Gille, S., Hänsel, U., Ziemann, M. & Pauly, M. Identification of plant cell wall mutants by means of a forward chemical genetic approach using hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 14699–704 (2009).

116. Saito, F. *et al.* Identification of novel peptidyl serine α-galactosyltransferase gene family in plants. *J. Biol. Chem.* **289,** 20405–20420 (2014).

117. Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* **301,** 653–7 (2003).

118. Scholl, R., May, S. & Ware, D. Seed and molecular resources for Arabidopsis. *Plant Physiol.* **124,** 1477–1480 (2000).

119. Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P. & Weisshaar, B. GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. *Nucleic Acids Res.* **40,** D1211–1215 (2012).

120. Edwards, K., Johnstone, C. & Thompson, C. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.* **19,** 1349 (1991).

121. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25,** 402–408 (2001).

122. Boyes, D. C. *et al.* Growth stage–based phenotypic analysis of Arabidopsis: A model for high throughput functional genomics in plants. *Plant Cell* **13,** 1499–1510 (2001).

123. Fragkostefanakis, S., Sedeek, K. E. M., Raad, M., Zaki, M. S. & Kalaitzis, P. Virus induced gene silencing of three putative prolyl 4-hydroxylases enhances plant growth in tomato (Solanum lycopersicum). *Plant Mol. Biol.* **85,** 459–71 (2014).

124. Roberts, K. & Shirsat, A. H. Increased extensin levels in Arabidopsis affect inflorescence stem thickening and height. *J. Exp. Bot.* **57,** 537–545 (2006).

125. De Cnodder, T., Vissenberg, K., Van Der Straeten, D. & Verbelen, J.-P. Regulation of cell length in the Arabidopsis thaliana root by the ethylene precursor 1-aminocyclopropane- 1-carboxylic acid: a matter of apoplastic reactions. *New Phytol.* **168,** 541–50 (2005).

126. Cooper, J. B., Heuser, J. E. & Varner, J. E. 3,4-Dehydroproline inhibits cell wall assembly and cell division in tobacco protoplasts. *Plant Physiol.* **104,** 747–752 (1994).

127. Smith, M. A. Characterization of carrot cell wall protein. *Plant Physiol.* **68,** 956–963 (1981).

128. Stratford, S. *et al.* A leucine-rich repeat region is conserved in pollen extensin-like (Pex) proteins in monocots and dicots. *Plant Mol. Biol.* **46,** 43–56 (2001).

129. Rubinstein, A. L., Marquez, J., Suarez-Cervera, M. & Bedinger, P. A. Extensin-like glycoproteins in the maize pollen tube wall. *Plant Cell* **7,** 2211–2225 (1995).

130. Held, M. A. *et al.* Di-isodityrosine is the intermolecular cross-link of isodityrosine-rich extensin analogs cross-linked in vitro. *J. Biol. Chem.* **279,** 55474–82 (2004).

131. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **22,** 1567–72 (2004).

132. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 7877–82 (2002).

133. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 11984–9 (2000).

134. Memelink, J. *et al.* Structure and regulation of tobacco extensin. *Plant J.* **4,** 1011–1022 (1993).

135. Jamet, E., Guzzardi, P. & Salva, I. What do transgenic plants tell us about the regulation and function of cell-wall structural proteins like extensins? *Russ. J. Plant Physiol.* **47,** (2000).

136. Gleave, A. P. A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Mol. Biol.* **20,** 1203–7 (1992).

137. Bird, C. R. *et al.* The tomato polygalacturonase gene and ripening-specific expression in transgenic plants. *Plant Mol. Biol.* **11,** 651–62 (1988).

138. Rose, R. E. *et al.* The nucleotide sequence of the 5' flanking region of a tomato polygalacturonase gene. *Nucleic Acids Res.* **16,** 7191 (1988).

139. Kelley, B. S. *et al.* A secreted effector protein (SNE1) from Phytophthora infestans is a broadly acting suppressor of programmed cell death. *Plant J.* **62,** 357–366 (2010).

140. An, G., Ebert, P. R., Mitra, A. & Ha, S. B. Binary vectors. *Plant Mol. Biol. Man.* **A3,** 1–19 (1988).

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

142. Pettolino, F. A., Walsh, C., Fincher, G. B. & Bacic, A. Determining the polysaccharide composition of plant cell walls. *Nat. Protoc.* **7,** 1590–607 (2012).

143. Hills, G. J., Phillips, J. M., Gay, M. R. & Roberts, K. Self-assembly of a plant cell wall in vitro. *J. Mol. Biol.* **96,** 431–441 (1975).

144. Scott, A., Wyatt, S., Tsou, P.-L., Robertson, D. & Stromgren Allen, N. Model system for plant cell biology: GFP imaging in living onion epidermal cells. *Biotechniques* **26,** 1125–1132 (1999).

145. Genovesi, V. *et al.* ZmXTH1, a new xyloglucan endotransglucosylase/hydrolase in maize, affects cell wall structure and composition in Arabidopsis thaliana. *J. Exp. Bot.* **59,** 875–889 (2008).