CHARACTERIZATION OF TWO UDP GLYCOSYLTRANSFERASE GENES FROM HYBRID POPLAR

By

Ahmed Kenawy

M.Sc., The University of Alexandria, (Egypt), 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Forestry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December, 2016

Abstract

Glycosyltransferases (GTs) play important roles in plant growth and development. The biological functions of many GTs are unknown. In the present study, two putative GT genes (*PopGT1* and *PopGT2*) were cloned and their biological roles in growth and development of *Arabidopsis* and hybrid-poplar were investigated. *In silico*, *in vitro*, and *in vivo* methods were used to characterize the two encoded proteins. Phylogenetic analysis, enzyme activity assays, and transcript abundance were studied. In addition, plant growth and development, leaf morphology, stem anatomy, cell wall composition, biomechanical properties, soluble carbohydrate, and phenolic metabolite contents were determined.

The results indicated that PopGT1 showed high similarity to tobacco salicylic acid glycosyltransferase, and both PopGT1 and PopGT2 (annotated as AtUGT74F2) were clustered within phylogenetic group L of family-1 GTs (UGTs). *In vitro* characterization of the two recombinant proteins indicated that PopGT1 glycosylated several flavonoids, showed only trace activities towards cinnamic and indole butyric acid, and accepted UDP-glucose as a sugar donor. The optimum temperature and pH for *in vitro* PopGT1 activity was 35 ºC and pH 7.5, respectively. PopGT2 showed no enzymatic activity towards any substrates.

The two coding sequences (*PopGT1* and *PopGT2*) were cloned in the *pSM3* expression vector and over-expressed in *Arabidopsis* plants to investigate their *in vivo* functions. Phenotypically, plant height, stem diameter, rosette diameter, and stem number increased significantly in the transgenic plants. In addition, rosette morphology and root gravitropism were altered. Transgenic plants flowered earlier than the control plants. Chemically, cell wall compositions and phenolic metabolite contents changed significantly.

In parallel, transgenic trees showed changes in leaf morphology, stem diameter, phloem fibre arrangement, and early bud break. Wood density was reduced revealing a brittle-stem phenotype. Marginal increases in lignin and reductions in cellulose content were apparent. Salireposide content was reduced in the bark of transgenic trees. The results indicated that altering the expression of both genes in *Arabidopsis* and poplar affected plant growth and development, cell wall composition, phenolic metabolite profiles, and wood biomechanical properties. PopGT1 showed *in vivo* substrate specificity towards kaempferol and promiscuous *in vitro* enzyme activity. However, the substrate of PopGT2 remains unclear.

Preface

The research was conducted at the University of British Columbia, Faculty of Forestry, Department of Wood Sciences.

Ahmed Kenawy conducted the experimental work, data analysis, and wrote the thesis. Dr. Mansfield contributed by supervising the research program, developing the research questions, providing the research opportunities, preparing and editing the manuscript. Ji-Young Park provided scientific support during the experimental work.

Dr. Ellis helped in formulating the research questions, resolving research problems, providing some of the authentic chemicals, and revising the thesis.

Dr. El-Kassaby helped with formulating the research questions, experimental design, discussion of the results, and revising the thesis.

Dr. Hamelin helped in formulating the research questions, providing rust strain, helped in resolving research problems, discussing the results, and revising the thesis.

Dr. Ralph, the University of Wisconsin, Madison, USA helped in metabolites analysis by Nuclear Magnetic Resonance (NMR) identification of salireposide.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2Ca:	Double 35S cauliflower mosaic-virus constitutive promoter.
Aox:	Alcohol oxidase
BA:	Benzyl adenine
BLAST:	Basic local alignment search tool
CAPS:	3-(Cyclohexylamino)-1-propanesulfonic acid
CAZy:	Carbohydrates active enzymes database
cDNA:	DNA coding sequence
Col:	Columbia
CTAB:	Cetyltrimethyl ammonium bromide
C-terminus:	Carboxylic terminus
cTP:	Chloroplast transit peptide
DNA:	Deoxyribonucleic acid
DNase:	Deoxyribonuclease enzyme
GC:	Gas chromatography
GT:	Glycosyltransferase
HEPS:	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His:	Histidine
HPLC:	High-pressure liquid chromatography
IAA	Indole acetic acid
IBA:	Indole butyric acid
IPTG:	Isopropyl β-D-1-thiogalactopyranoside
IUBMB:	The International Union of Biochemistry and Molecular Biology
kb:	Kilobase
kDa:	Kilodalton
kg:	Kilograms
LB:	Luria broth
Len:	Length
Loc:	Localization
M:	Molar
MEO:	Modulus of Elasticity
MOPS:	3-(N-morpholino) propanesulfonic acid
MSA:	Methyl salicylic acid
MSL:	Michael Smith Laboratory
mTP:	Mitochondrial target peptide
NAA:	Naphthaline acetic acid
***nah*G:**	Salicylic acid carboxylase
NAPS:	Nucleic Acids Protein service unit
Ni^{+2}:	Nickle
NMR:	Nuclear magnetic resonance
N-terminus:	Amino terminal
º C:	Degrees celsius
PAL:	Phenyl alanine ammonia lyase
PAs:	Proanthocyanidins

PCR:	Polymerase Chain Reaction
PopGT:	Poplar glycosyltransferase
PSPG-box:	Plant secondary product glycosyltransferase
PtGT:	*Populus trichocarpa* glycosyltransferase
PVPP:	Polyvinylpolypyrrolidone
qRT-PCR:	quantitative real time PCR
RC:	Reliability class
RNA:	Ribonucleic acid
RNAi:	RNA interference
RNAi-PopGT:	RNAi suppressing lines of PopGT gene.
RNase:	Ribonuclease
rpm:	Revolutions per minute
SA:	Salicylic acid
SAG:	Salicylic acid glycosyltransferase
SGE:	Salicylic acid glucose ester
SAGT:	Salicylic acid glycosyltransferase
SDS:	Sodium dodecyl sulfate
SP:	Secretory pathway signal peptide
TAIR:	the *Arabidopsis* information resource
TAPS:	3-[[1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]amino]propane-1-sulfonic acid
TDZ:	Thidiazuron
TPlen:	Predicted pre-sequence length (for cleavage site prediction)
Tris:	2-Amino-2-hydroxymethyl-propane-1,3-diol
UBC:	The University of British Columbia
UDP:	Uridine diphosphate
UGT:	UDP glucose glycosyltransferase
WPM:	Woody plant medium
RACE:	Random amplification of cDNA ends
MS:	Murashige and Skoog
OE:	Over-expression

Acknowledgements

I would like to express my sincere thanks to my supervisor, Dr. Shawn D. Mansfield, for the research opportunity he offered me, and for his effort in helping to get this work done. Warm thanks to members of my supervisory committee for their immense contribution and support during the research and writing of the thesis. My thanks are extended to Dr. John Ralph, Dr. Patrick Martone, Dr. Rob Guy, and Dr. Lacey Samuels for allowing me to use their laboratory facilities. I would also like to express my appreciation to students and post-doctoral researchers in the Mansfield laboratory.

Special thanks to Dr. Ji Young Park, Dr. Faride Unda, Dr. Grant McNair, Dr. Mohamed Ismail, and Dr. Abdul-Azeem Zumrawi, for the unconditional scientific and spiritual help and support they offered. Many thanks go to Yassen Mottiar, Emily Murphy, Martha Kertesz, and Amanda Johnson for their help and support.

I would also like to extend my thanks to the UBC Access and Diversity Office, especially, Chandra Barber, for their support.

The contributions of the following funding institutions are gratefully acknowledged:

(a) The Egyptian Ministry of Higher Education and Scientific Research for financial support through scholarship (2007-2010);

(b) Members of the Egyptian Cultural and Educational Office Diplomatic Mission, Canada;

(c) Staff and Faculty of the City of Scientific Research and Technology Applications, Genetic Engineering, and Biotechnology Research Institute of Egypt;

(d) The Faculty of Forestry, University of British Columbia (UBC), Vancouver, for financial support through a partial tuition fee award;

(e) The British Columbia Muslim Association, Richmond Branch, for the special support.

To my parents, my daughters "Youmna and Mariam", my family and my friends

Chapter 1: Introduction

1.1 Importance of poplar trees

Poplar trees belong to the genus *Populus* and the willow family (*Salicaceae*). They are important forest trees both ecologically and economically. Currently, the genus *Populus* includes 30 different species worldwide, which are classified into six sections. Each section combines a group of species with similar ecological and morphological properties (Lexer et al., 2005; Eckenwalder, 1996). The six sections include:

Abaso: the Mexican poplar;
Aigeiros: the cottonwoods and black poplar;
Leucoides: the swamp poplars;
Populus: the white poplars and aspens;
Tacamahaca: the balsam poplars;
Turanga: the Afro-Asian poplars.

Hybridization is a common, naturally occurring phenomenon, either within the same section (intra-breeding), or between compatible species of different sections (inter-breeding). The resulting hybrid poplars usually demonstrate superior phenotypes over their parent species (DiFazio et al., 2011). For example, the interspecific F1 crosses between *P. deltoides* and either *P. trichocarpa* or *P. nigra* displayed hybrid vigor (heterosis) for shoot growth, making them ideal for agroforestry production (Dickmann and Keathley, 1996; Bradshaw et al., 2000; Pearce et al., 2004).

Poplars have extensive niche distributions and are spread widely over the boreal and temperate regions of the Northern Hemisphere. In North America, the range of poplar trees extends from

the northern tree line to the north of Mexico (DiFazio et al., 2011). Canada has approximately 40% of the world's natural poplars, with 28.3 million hectares of poplars existing in natural stands (Isebrands and Richardson, 2014).

Poplar is a pioneer species that grows rapidly, with extreme adaptability to harsh environments. It is a common colonizer of sites cleared by fire, wind, or logging. For example, *P. tremuloides* (quaking aspen) is highly intolerant of shade, as well as competition, and is able to establish pure stands in sites that have been disturbed (Dickmann et al., 2001).

Poplar species are dioecious, with distinct male and female sexes, although poplar trees displaying both sexes have been reported in some species (Cronk, 2005; Slavov et al., 2009). Poplar trees reproduce by different mechanisms, including sexual and asexual methods. Seeds are produced via sexual reproduction and can be carried by wind for long distances, which facilitates the geographic distribution of the species. Asexual reproduction in poplar via root sprouts (suckers) is common, distinguishing it from other tree species (Dickmann et al., 2001). The vegetative reproductive nature of *Populus* provides a fitness advantage that permits the species to spread across a variety of habitats. It also facilitates the propagation of genetically improved genotypes in a quick and effective manner (DiFazio et al., 2011; Rood et al., 2007; Smulders et al., 2008).

Given the unique characteristics of poplar trees, there has been an increasing interest in establishing high-yield, short-rotation forest plantations of poplars as a potential source of cellulose fibre, climate-change mitigation strategies, and for energy production via biofuels (Bradshaw et al., 2000; Taylor, 2002). In addition, it is commonly used in the production of

veneer, plywood, fibreboard, synthetic wood-cement, pallets, and paper products (Cooke and Rood, 2007; Isebrands and Richardson, 2014). The fact that poplar trees display wide distribution, rapid growth, fertile sexual reproduction, ease of cloning via vegetative reproduction, ease of transformation, relatively small genome size, and the availability of genetically defined plant materials has established it as a model tree species (Bradshaw et al., 2000; Taylor, 2002; Cooke and Rood, 2007). In addition, its genome sequence was completed and published in 2006 (Tuskan et al., 2006). In comparison with *Arabidopsis*, poplar displays characteristics that are not common in other non-woody plants, such as its dioecious nature, large size, seasonal dormancy, secondary xylem formation, and long life span (Bradshaw et al., 2000; Taylor, 2002).

1.2 Secondary cell wall formation and wood structure

As a woody plant, the most economically valuable part of a tree is the trunk. Wood formation is the final stage in a succession of biochemical processes, including initiation, elongation, thickening, lignification, and cell death. The deposition of the secondary cell wall components (cellulose, hemicellulose, and lignin) occurs during both the thickening and lignification steps. Wood tissue is a specialized structure that conducts water from the ground to the leaves through vessel elements, and provides the plant with mechanical support through its fibres. These two types of specialized cells (vessel elements and fibres) have lignified walls (Albersheim et al., 2010).

Cell walls are composed of different structural carbohydrate polymers (cellulose, hemicellulose, and pectin) and are reinforced by phenolic polymers called lignin (Mansfield, 2009). Cellulose

composes 50% of the secondary cell wall biomass, and consists of D-glucose monomers that are linked linearly via β (1→ 4) glycosidic bonds. Hydrogen bonding between cellulose chains forms a microfibril structure and permits the formation of crystalline domains (Albersheim et al., 2010). Furthermore, the arrangement of cellulose microfibrils along the longitudinal axis of the cell forms three secondary cell wall layers (S1, S2, and S3). In the S2 layer, the largest among the three layers, the arrangement of the microfibrils around the longitudinal axis of the cell largely determines the microfibril angle (MFA). MFA affects the physical properties of wood, such as stiffness, flexural properties, and shrinkage upon drying. It was shown that MFA negatively correlates with the modulus of elasticity (Bendtsen and Senft, 1986). Lignin is a phenolic polymer composed of the three monolignols (syringyl, guaiacyl, and *p*-coumaryl alcohols). It represents 18-35% of the secondary cell wall composition, and contributes to the cell wall strength by reinforcing the walls and improving strength. Hemicellulose represents 15-30% of the secondary cell wall structure, and is highly branched with variable sugars linked together via glyosidic bonds (Mansfield, 2009). The main hemicellulose in hardwood cell walls is xylan, which is composed of xylose units linked via β (1→4) bonds. The formation of the sugar-based polymers (cellulose, hemicellulose, and pectin) is the result of the integrated activity of several carbohydrate-active enzymes, including various glycosyltransferases (Geisler-Lee et al., 2006; Brown et al., 2007; Frankova and Fry, 2013).

1.3 Glycosylation

In plants, glycosylation is a process that is crucial to carbon metabolism, cellular energy conservation, cell integrity, and dynamic growth. It is one of the key enzymatic processes that

provides living cells with the requisite plasticity needed to adapt to changing environmental conditions (Vogt and Jones, 2000; Bowles et al., 2006; Malik and Black, 2012). Glycosylation also plays an important role in orchestrating the bioactivity, transportation, and metabolism of small metabolites in living cells. In plants, this process is often the last step in the biosynthesis of many natural products (Vogt and Jones, 2000). Glycosyltransferases (GTs) control many biological processes, such as cell wall synthesis, maintaining cellular homeostasis of plant hormones, localization of secondary metabolites, compartmentalization and detoxification of xenobiotics, such as 2,4,5-trichlorophenol (TCP) and the *Alternari brassicae* toxins (destruxin B) (Vogt and Jones, 2000; Lim and Bowles, 2004; Bowles et al., 2006; Offen et al., 2006). Glycosylation adds a sugar molecule to the hydroxyl (OH), carboxyl (-COOH), amino (-NH2), sulfhydryl (-SH), a acyl (C-C) group of the acceptor molecule, ultimately creating a plethora of unique secondary metabolites (Jones and Vogt, 2001; Gachon et al., 2005; Bowles et al., 2005; Wang and Hou, 2009).

1.4 Glycosyltransferases

Glycosyltransferases (GTs) are carbohydrate-modifying enzymes that exist in all living organisms and control many metabolic processes in living cells (Paquette et al., 2003; Liang et al., 2015). GTs belong to the EC 2.4.x.y class of enzymes responsible for transferring a sugar residue from a nucleoside diphosphate-activated sugar (NDP-sugar donor) to an acceptor molecule (Paquette et al., 2003). In 1984, the first cDNA sequence of a secondary-metabolite GT was isolated from maize (*bronze*-1), where it encodes a putative flavonol 3-O-GT (Fedoroff et al., 1984). Based on this sequence, many other GTs were subsequently cloned and characterized.

1.4.1 Classification of glycosyltransferases

The diverse functions of GT enzymes, and the unknown relationship between their primary sequences and their regio-specificity and regio-selectivity, have made their classification a challenging process (Osmani et al., 2009). Many systems have been suggested for the classification of GTs, but no single method has proven sufficient to assign an enzymatic function to a putative enzyme accurately without further *in vivo*/*in vitro* characterization (Bowles et al., 2005; Bowles et al., 2006; Osmani et al., 2009).

Glycosyltransferases have been classified based on the type of sugar donor into either the Leloir enzymes, where activated nucleotide sugars are the source of the glycosyl moiety, or into the non-Leloir enzymes, where non-nucleotide sugars are employed (Leloir and Cardini, 1957; Compain and Martin, 2001; Weijers et al., 2008). This method of classification did not take into consideration the stereochemical changes that occur in the anomeric carbon of the sugar and did not consider the specificity of the enzyme towards the acceptor molecule. A different approach to GT classification was later proposed based on improved knowledge of the enzymatic reaction mechanisms. In this classification scheme, GTs can be classified into either inverting or retaining enzymes, based on whether they retain or invert the stereochemical orientation of the anomeric carbon of the sugar (Sinnot, 1990; Kapitonov and Yu, 1999; Lairson et al., 2008). In addition, the International Union of Biochemistry and Molecular Biology (IUBMB) nomenclature classified GTs as members of the nucleoside diphosphate-glycosyltransferase (NDP-GTs: EC 2.4.x.y) class of enzymes (www.CAZy.com). However, neither the broad substrate specificity towards different phenolic substrates, nor the possible functions of the novel GT enzymes could be adequately predicted by the IUBMB system (Vogt and Jones, 2000). Therefore, these

classification methods are considered insufficient in classifying the UGT enzymes (Osmani et al., 2009).

The availability of the *Arabidopsis* genome sequence (The *Arabidopsis Genome Initiative*, 2000) and the development of improved bioinformatics approaches have enhanced the classification of GTs. The GT enzymes are now grouped into families based on a range of characteristics, including their 3D-fold structures (GT-A, GT-B, and the recently predicted GT-C), their catalytic reaction mechanisms (inverting or retaining), the existence of consensus sequences, and the degree of amino acid sequence identity (Campbell et al., 1997; Kapitonov and Yu, 1999; Keegstra et al., 2001; Lairson et al., 2008; Osmani et al., 2009).

According to the most recent update of the Carbohydrate-Active Enzyme database (http://www.cazy.com, 18/12/2016), GTs fall into 101 families, as well as a non-classified group of GT enzymes. Within each GT family, protein sequence identity is 40% or greater, and the protein 3-D fold-structure is expected to be the same for all members. However, functional prediction based on sequence similarities remains difficult, and the substrate-specificity of many GTs grouped within the same family is still unknown (Bowles et al., 2006). GT families have been classified into four large clans based on the 3D-folding type and the reaction mechanism: GT-A inverting, GT-A retaining, GT-B inverting, and GT-B retaining (Lairson et al., 2008). The distribution of the GT enzymes varies in number between organisms, and their classification is continually changing (Paquette et al., 2003; Osmani et al., 2009; Wang and Hou, 2009).

1.4.2 Glycosyltransferase diversity

Plants rely on a group of carbohydrate-active enzymes (CAZy) for the biosynthesis of different glycosylated photosynthetic-products, including disaccharides, polysaccharides, and cell wall polymers (Geisler-Lee et al., 2006). The two major enzyme categories in the CAZy database are the glycosyltransferases (GTs) and glycosylhydrolyses (GHs). The two groups of enzymes play an integral role in the biosynthesis of a large variety of glycosylated compounds, including glycolipids, glycoproteins, lignin precursors, and phenolic glycosides (Vogt and Jones, 2000; Keegstra and Raikhl, 2001). Among their diverse roles in plants, GTs are responsible for the biosynthesis of cell wall polymers (cellulose, hemicellulose, and pectin) (Lerouxel et al., 2006; Zhou et al., 2007; Perrin, 2008). Cellulose biosynthesis is considered the most important biological process on earth. Approximately 10^{11} tons of cellulose are produced each year (Mansfield, 2009).

Several GT families are integral to the formation of plant cell wall components, including GT1, GT2, GT4, GT5, GT8, GT29, GT 43, GT37, GT47, and GT48. For example, family GT2 combines cellulose synthase (CESA) and cellulose synthase-like (CSL-like) enzymes. These GTs form an enzyme complex (cellulose synthesis complex (CSC)) that is responsible for the biosynthesis of plant cell wall cellulose. The *At*CSLC4 enzyme has been shown to be involved in the synthesis of the xyloglucan backbone, while family GT8 contains galacturonosyltransferase (*GAUT*) and galacturonosyltransferase-like (*GATL*) genes, as well as pectin-synthesizing enzymes (Sterling et al., 2006). Four *Arabidopsis* proteins belonging to family GT43 (irregular xylem (IRX) 9, IRX14, IRX9-L, and IRX14-L), as well as two proteins (IRX10 and IRX10-L) belonging to the GT47 family, were reported to be involved in the synthases of xylan and

glucuronoxylan backbones, respectively (York and O'Neill, 2008). Family GT1 combines enzymes that glycosylate small phenolic metabolites, including plant hormones and phenylpropanoid metabolites (Ross et al., 2001). The members of this family contribute to the synthesis of the secondary cell wall in different manners (directly or indirectly). For example, UGT72E2 and UGT84A1 enzymes have been implicated in monolignol glycosylation. On the other hand, UGT84B1 (auxin GT) modulates hormone activity, indirectly affecting cell wall formation, while ectopic expression of flavonoid UGTs affect auxin transport mechanisms and thereby also indirectly affect cell wall formation (Jackson et al., 2001; Jackson et al., 2002; Jones et al., 2003; Lim et al., 2005; Bowles et al., 2006).

1.4.3 Glycosylation and homeostasis of secondary metabolites

Glycosylation of free (soluble) metabolites is believed to be an important step in maintaining a homeostatic level of plant metabolites and protecting plant cells from the toxic effect of high concentrations of free substances, enabling plants to store these metabolites in an inactive form until needed (Bowles et al., 2006). Adding a sugar moiety to a lipophilic phenolic molecule can modify its intrinsic chemical properties, functionality, and bioactivity (antioxidant activity), and plants have broadly adopted this mechanism to manipulate their secondary metabolites (Vogt and Jones, 2000). For example, glycosylation of cyanogenic and glucosinolate compounds stabilize them in intact plant cells. During microbial invasion or herbivore attack, the cleavage of the sugar moiety from the cyanogenic glycoside releases the unstable aglycone, which eventually breaks down and generates the toxic hydrogen cyanide (Bowles et al., 2006). Similarly, when the wall is damaged, stable glucosinolate glycosides hydrolyze, releasing bitter compounds (Fahey et

al., 2001). Solanidine glycosyltransferase reduces the toxicity of the steroidal alkaloid solanidine in *Solanaceae* plants by converting it to its tri-glycoside form (solanine) (Charles et al., 1997; Moehs et al., 1997). Glycosylation can also participate in maintaining the redox state of the cell by adjusting the homeostatic level of ascorbic acid. For example, UGT78A2 has been implicated in maintaining ascorbic acid homeostasis in *Arabidopsis* (von Saint Paul, 2010).

The role of GTs in determining the colour of plant pigments (anthocyanins) has been studied extensively. Morita et al., (2005) found that the loss of GT activity in Japanese morning glory flowers, which are naturally red or blue, resulted in a change in their colour to a purplish-gray or reddish-brown. These colour changes were correlated with a mutation in a GT gene, leading to the accumulation of anthocyanidin 3-O-glucoside in the mutant flowers instead of anthocyanidin 3-O-sophorosides (Morita et al., 2005; Bowles et al., 2006).

UGTs also have important roles in maintaining the stability and activity of plant defense-related metabolites. For example, phenolic metabolites (salicin, tremulacin, and salicortin) provide trees with a chemical defense against herbivores, affecting insect development and reproduction, as well as hydrolyzing inside the insect stomach, where they cause toxic effects (Ruuhola et al., 2001 a and b; Ruuhola and Julkunen-Tiitto, 2003; Mithofer and Boland, 2012). They can also be converted to signaling molecules that trigger plant defense mechanisms against pests, such as salicylic acid and its metabolites. However, the removal of the sugar molecules from the phenolic glycosides is also toxic to the plant cells, and induces stress responses, as well as programmed cell death (Dempsey et al., 1999; Dempsey et al., 2011). Interestingly, some phenolic metabolites have been found chemically active when they exist in glycoside forms. For example, sinapoylmalate (putative ultraviolet (UV) protectant) is synthesized from the

intermediate 1-O-sinapoylglucose ester (chemically active glycoside) in the leaves of *Brassicaceae* (Milkowski et al., 2000; Bowles et al., 2006). Another example is saponins (glycosylated terpenes) that have a role in antifungal resistance. They are characterized by the existence of a side chain of polysaccharide (up to 5 sugars) that is attached to the C-3 position, and is vital for their antifungal activity. The cleavage of the C3 polysaccharide side-chain by β galactosidase from *Gaeumannomyces graminis* resulted in the loss of the saponins antifungal activity (Sandrock and Van-Etten, 1998). Other glycosylation events also occur on different saponin carbons (C26 and C28) that influence the antifungal bioactivity of saponin (Bowles et al., 2006). For example, cleaving the C-26 glucose from the inactive forms of avenacosides A and B during microbial invasion results in their conversion to biologically active antifungal agents (Jones et al., 1999; Osbourn, 2003). A few GTs have been associated with this modification (Bowles et al., 2006). The solanum SaGT4A and SGT enzymes were found, *in vitro,* to glycosylate saponins. In addition, UGT73K1 and UGT71G1 were functionally characterized in *Medicago truncatula,* and UGT73K1 was found to glycosylate several saponin aglycones, such as hederagenin and soyasapogenols B and E. On the other hand, UGT71G1 showed specificity towards the triterpene medicagenic acid (Achnine et al., 2005; Bowles et al., 2006).

1.5 Family 1 glycosyltransferases

1.5.1 UGT family 1 nomenclature

Campbell et al., (1997) and subsequently Coutinho et al., (2003) proposed a novel scheme for classifying the UDP-GT enzymes (UGTs) into different families based on their protein sequence

identity. This method of classification has been recommended by the international nomenclature committee to create a systematic scheme for naming these enzymes (Mackenzie at al., 1997). Family 1 is a multigene superfamily and the enzymes belonging to it are clustered into separate families based on their sequence identity. According to this classification scheme, the UGT superfamily is divided into different families: animal UGT are assigned to families 1-50, yeast to families 51-70, plants to families 71-100, and bacterial enzymes to families 101-200 (Mackenzie et al., 1997). In this classification method, if two enzymes show at least 40% amino acid sequence identity, they are grouped into the same family. An Arabic number is then assigned to each cluster, identifying each family. The enzymes within each family are classified into subfamilies, and those that share 60% or more sequence identity are assigned a subfamily letter (Mackenzie et al., 1997; Ross et al., 2001; Osmani et al., 2009). Finally, within each subfamily a unique number is assigned to each UGT enzyme. For example; *UGT74F1* is a gene that codes an enzyme belonging to the UGT superfamily, family number 74, and subfamily F, and has been designated enzyme 1 (Figure 1).

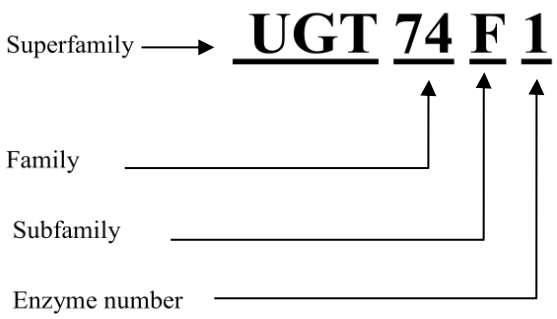


Figure 1: Enzyme nomenclature scheme of family 1 uridine diphosphate glycosyltransferases.

1.5.2 The roles of family 1 glycosyltransferase

Family 1 UGT enzymes employ a uridine di-phosphate sugar (UDP-sugar) as a donor molecule to add a glycosyl moiety to hydrophobic acceptor molecules (www.CAZy.org). Family 1 UGTs is one of the largest families of GTs, composed of 11,753 enzymes, distributed among many living organisms. For example, 195, 176, 6,797, and 4,561 enzymes of family-1 UGTs have been reported in viruses, archaea, bacteria, and eukaryotes, respectively, in addition to 24 unclassified UGTs. To date, only 368 have been functionally characterized and only 29 crystal structures have been resolved, including 19 structures belonging to bacteria and 9 belonging to eukaryotic organisms (http://www.cazy.org/GT1.html (18/12/2016)).

The plant UGT superfamily has been established based on the existence of the carboxylic-terminus consensus-sequence motif, which is a characteristic domain for plant secondary product glycosyltransferase (PSPG). This domain contains the conserved amino acids sequence [F/W]-x(2)-[Q/L]-x(2)-[L/I/V/M/Y/A]-[L/I/M/V]-x(4,6)-[L/V/G/A/C]-[L/V/F/Y/A/H/M]-[L/I/V/M/F]-[S/T/A/G/C/M]-[H/N/Q]-[S/T/A/G/C]-G-x(2)-[S/T/A/G]-x(3)-[S/T/A/G/L]-[L/I/V/M/F/A]-

x(4,5)-[P/Q/R]-[L/I/V/M/T/A]-x(3)-[P/A]-x(2,3)-[D/E/S]-[Q/E/H/N/R] (Hughes and Hughes, 1994), and binds the UDP moiety of the activated sugars (Hughes and Hughes, 1994; Lairson et al., 2008; Osmani et al., 2009; Wang and Hou, 2009). Analysis of the 3-D protein folding models of members of family 1 UGTs have shown that they are type-B enzymes, consisting of two β/α/β Rossmann-like domains (Lairson et al., 2008; Wang and Hou, 2009). In addition, it was also shown that family 1 GTs are inverting enzymes that change the stereo-chemical position of the anomeric carbon of the sugar donor, ultimately leading to their classification as a GT-B inverting enzymes (Lairson et al., 2008).

Several crystal structures of members of this family (eg., *Medicago truncatula* UGT71G1, *Arabidopsis thaliana* UGT72B1, *M. truncatula* UGT78G1, *M. truncatula* UGT85H2 and *Vitis vinifera* VvGT1) have been resolved (Shao et al., 2005; Offen et al., 2006; Brazier-Hicks et al., 2007 a and b; Li et al., 2007; Modolo et al., 2009). The conserved amino acid motif was found in the resolved structures, and its role in the regio-selectivity and regio-specificity of the enzyme has been examined (Shao et al., 2005; Offen et al., 2006; Brazier-Hicks et al., 2007 a and b; Li et al., 2007; Modolo et al., 2009). In addition, analysis of the crystal structures showed that there are conformational changes in the folding of the enzyme after the substrate is bound to the active site (Wang, 2009).

1.5.3 Sequence architecture of family 1 glycosyltransferases

When full-length protein sequences are compared, UGTs show low sequence similarity. This lack of sequence similarity has made it difficult to predict enzyme function (Paquette et al., 2003; Bowles et al., 2006; Osmani et al., 2009). However, the availability of several putative

UGT genes from different plant species has provided the opportunity to revise our understanding of the sugar binding mechanisms that occur at the PSPG-box consensus sequence at the C-terminus of the enzyme. Within this domain, certain amino acids were found to have special importance when interacting with nucleotide sugars (Osmani et al., 2009; Wang and Hou, 2009). For example, highly conserved amino acids were found in positions 1 (W), 4 (Q), 8 (L), 10 (H), 19–24 (HCGWNS), 27 (E), 39 (P), 43 (E/D) and 44 (Q), while residues W (22), D/E (43) and Q (44) form hydrogen bonds with the sugar component of the donor molecule (Osmani et al., 2009; Wang, 2009). The glutamine residue Q (44) is highly conserved across most UGTs (Wang, 2009; Wang and Hou, 2009) and is believed to be responsible for the sugar-donor specificity of UGTs. Glutamine Q (44) was found to be important for glucosyl moiety transfer activity, whereas substituting glutamine (Q) with histidine (H) results in galactosyl moiety transfer activity (Kubo et al., 2004; Ono et al., 2010; Caputi et al., 2012).

The amino terminus of UGT enzymes shows a lower level of sequence similarity than the carboxylic terminus of the enzyme. It has been suggested that the aglycone recognition site might occur in the amino terminus (Li et al., 2007), while certain amino acids found in the amino terminus have been found to be responsible for the interaction between the phenolic substrates and the enzyme (Wang and Hou, 2009).

1.5.4 Distribution of family 1 glycosyltransferase in plant genomes

As more genome sequences become available, information about the number and distribution of family 1 UGTs among eukaryotic organisms is being revealed. The number of these enzymes in different species reflects the degree of complexity of an organism. For example, in

Physcomitrella patens, a primitive moss with no developed vascular system, only 15 UGT sequences have been identified. On the other hand, the genome sequences of more complex organisms, such as *Arabidopsis* and poplar, contain 122 and 336 UGT sequences, respectively (Tuskan et al., 2006). Other plants showed variability in the number of UGT sequences found in their genomes, including:

- *Linum usitatissimim* with 137 putative family 1 UGT genes (Barvkar et al., 2012);
- *Oryza sativa* with 224 UGT sequences (Tanaka et al., 2008);
- *Vitis vinifera* with 210 UGTs genes (Jaillon et al., 2007);
- *Medicago truncatula* with 187 UGT genes (Modolo et al., 2007).

1.5.5 Phylogenetic and evolutionary studies on family 1 UGTs

Phylogenetic studies examining UGT enzymes in *Arabidopsis* have resulted in their classification into 14 different phylogenetic groups (A-N; as indicated in Table 1) (Li et al., 2001; Ross et al., 2001). Paquette et al., (2003) conducted a phylogenetic comparison between UGTs from several organisms (plants, animals, fungi, bacteria, and viruses) and showed that plant UGTs fall into three main phylogenetic groups (clades), with one dominant monophyletic clade. The appearance of this major group suggested that plant UGTs had diverged from those of the remaining groups of organisms after the divergence of plant, fungi, and animal kingdoms. The other two groups of plant UGTs (lipid and sterol GTs) form minor groups, and likely reflect a bacterial ancestry of chloroplast enzymes (Paquette et al., 2003).

The appearance of multiple pseudogenes, as well as chromosomal clustering of enzyme subfamilies is evidence for one or more recent duplication events (Paquette et al., 2003). It was

suggested that many gene duplication events occurred during the evolution of terrestrial plants, but most of these were lost (Caputi et al., 2012). An analysis of 1500 UGT gene sequences from six plants revealed 24 orthologous groups (OG1-OG24), and the results clearly showed that a significant expansion of UGTs occurred in a lineage-specific manner. The lineage-specific pattern of UGT expansion and the independent functional modification have led to an inaccurate prediction of an association between the structure and function of novel UGTs, which is probably responsible for the slow progress in identifying the functions of UGTs (Sakakibara and Hanada, 2011). The larger size of the GT1 family in poplar trees, compared to other herbaceous plants, reflects the whole genome duplication events that occurred during poplar evolution (Tuskan et al., 2006). It might also suggest that poplar employ a greater number of active secondary metabolite-related enzymes to accommodate the need for the longevity of life, structural complexity, and dormancy control (Geisler-Lee et al., 2006).

Table 1: Classification of 107 *Arabidopsis* UGTs into 14 phylogenetic groups (A-N) and some of the detected activity of their members as summarized by Ross et al., (2001).

Phylogenetic group	Family number	Activity
Group A	79B1, 79B2, 79B3, 79B4, 79B5, 79B6, 79B7, 79B8, 79B9, 79B10, 79B11, 91A1, 91B1, 91C1	Anthocyanidin 3-O-glycoside
Group B	89A2, 89B1, 89C1	NA
Group C	90A1, 90A2, 90A4	NA
Group D	73C1, 73C2, 73C3, 73C4, 73C5, 73C6, 73C7, 74B1, 74B2, 74B3, 74B4, 74B5, 74D1	Betanidin, flavonols esculetin, scopletin, solanidine
Group E	71B1, 71B2, 71B5, 71B6, 71B7, 71B8, 71C1, 71C2, 71C3, 71C4, 71C5, 71D1, 71D2, 72E1, 72E2, 72E3, 72B1, 72B2, 72EB3, 72D1, 72C1, 88A1	NA
Group F	78D1, 78D2, 78D3	Anthocyanidin, flavonols
Group G	85A1, 85A2, 85A3, 85A4, 85A5, 85A7	Hydroxymandelonitrile
Group H	76B1, 76C1, 76C2, 76C3, 76C4, 76C5, 76D1, 76E1, 76E2, 76E3, 76E4, 76E5, 76E6, 76E7, 76E9, 76E11, 76E12, 76F1, 76F2	NA
Group I	83A1	NA
Group J	87A1, 87A2	NA
Group K	86A1, 86A2	NA
Group L	74B1, 74C1,74D1,74E1, 74E2, 74F1, 74F2, 75B1, 75B2, 75C1, 75D1, 84A1, 84A2, 84A3, 84A4, 84B1, 84B2	Phenyl acetyl-thiohydroximate Limonoids *t*-cinnamic acid Benzoic acid Salicylic acid (SA)
Group M	92A1	NA
Group N	82A1	NA

NA = not available

This phylogenetic classification was helpful to identify plant UGTs functions. However, UGTs in plants tend to have low substrate-specificity and new functions are regularly discovered for

previously characterized enzymes (Lynch and Conery, 2000). For example, UGT74D1 was previously characterized as an auxin UGT, although a recent study indicated that this enzyme showed activity towards 2-oxindole-3-acetic acid (Tanaka et al., 2014). The *in vitro* activity of UGT74D towards jasmonic acid (JA) was also reported (Song, 2005). Several publications suggest that UGT74E2 and UGT84B1 display activity towards plant auxins (Jackson et al., 2001; Jackson, 2002), although they belong to different UGT families (Jackson et al., 2001; Tognetti et al., 2010). Similarly, UGT 73C6 and UGT78D1 are both flavonol 7-O-GTs that belong to different UGT phylogenetic groups (Jones et al., 2003). UGT75B1 and UGT75C1 belong to the same family (Ross et al., 2001), but UGT75B1 is primarily a *p*-aminobenzoate acyl-glucosyltransferase (Hong et al., 2001; Eudes et al., 2008), while UGT75C1 encodes an anthocyanin 5-O-glucosyltransferase (Tohge et al., 2005). Accordingly, the substrate specificity of UGT enzymes cannot be accurately predicted by using only the primary amino acid sequence information and phylogenetic classification. The integration of different analytical approaches is required to understand the regio-specificity and regio-selectivity of these enzymes towards different donor and acceptor substrates (Osmani et al., 2009).

1.6 Glycosylation of secondary metabolites in plants

Glycosylation is an important mechanism that has evolved to generate a wide range of secondary metabolites. These metabolites are crucial for plant growth, development, and survival in challenging environmental conditions (Jones and Vogt, 2001). Many classes of secondary metabolites have been isolated and characterized, including phenolic acids, terpenoids, tannins, alkaloids, thiohydroximates, steroids, cyanogens, glucosinolates, salicylates, and flavonoids

(Bowles et al., 2006; Zhong and Yue, 2005; Wang and Hou, 2009). Plant UGTs that are responsible for the production of several glycosylated secondary metabolites have been found to display regio-selectivity mechanisms rather than substrate specificity. However, the UGT regio-specificity and regio-selectivity mechanisms remain poorly understood (Jones and Vogt, 2001).

When plants face temporary or permanent environmental stresses, their fitness can be negatively affected. In order to overcome stresses, plants have evolved extraordinary biochemical plasticity that enable them to respond quickly and effectively to changing conditions, allowing them to alter their growth, development, metabolism, and physiology (Sultan, 2000; Offen et al., 2006). The ability of plants to overcome these conditions can be attributed to the synthesis and storage of numerous secondary metabolites (Gachon et al., 2005; Zhong and Yue, 2005). As plants evolved, the number, complexity, and dynamics of these metabolites increased, and the biosynthetic mechanisms involved in maintaining their homeostasis became more complex (Vogt and Jones, 2000; Gachon et al., 2005; Babst et al., 2014). Each group of plants developed a characteristic portfolio of secondary metabolites that has become an important taxonomic signature for each group (Vogt and Jones, 2000).

The products of the phenylpropanoid pathway are crucial for plant growth, development, and survival (Vogt, 2010). For example, lignin is an important polymer for cell wall formation providing the mechanical support for the cell, and is generally composed of three phenolic monolignols (Mottiar et al., 2016). These monolignols are phenylpropanoid pathway products, and their glycosylation is an important step in the storage and localization of these molecules (Vogt and Jones, 2000). For example, UGT72E2 and UGT72E3 have been shown to be phenylpropanoid GTs, using coniferyl and sinapyl alcohols as aglycones substrates (Lim et al.,

2005), while UGT73E1 and UGT72E2 were found to have activity toward the aldehyde forms of the aglycones (Lim et al., 2005; Bowles et al., 2005). Flavonoids, which are also products of the phenylpropanoid pathway, are found mostly in glycosylated forms in plants (Vogt and Jones, 2000; Lim et al., 2003; Bowles et al., 2006; Yin et al., 2012). This general class of metabolites includes; anthocyanins, flavonols, flavones, and isoflavones (Williams and Gryer, 2004; Bowles et al., 2006), all of which play important roles in plant protection, growth, and development, including fruit and flower colouring, protection against UV radiation, and pollen fertility. In addition, these metabolites modulate plant hormone transport, and affect plant morphology, stress resistance, and ecological signalling (Dixon et al., 2002; Taylor and Grotewold, 2005; Treutter, 2005; Peer and Murphy, 2007; Buer et al., 2010).

Small phenolic metabolites, such as salicylic acid (SA), have been associated with the control of signal transduction, hypersensitive responses and programmed cell-death mechanisms (Dempsey et al., 2011). The role of SA and its derivatives in these mechanisms has been studied under different biotic and abiotic stress conditions, since it is believed that SA and its glycosides are signalling molecules (Dempsey et al., 2011). However, the exact mechanism remains unclear. For example, methyl salicylic acid (MSA) was found to play an important role as an inter-/intra-plant signalling molecule, increasing the hypersensitive response towards both biotic and abiotic factors and initiating cellular cross talk (Shah et al., 1999; Song, 2005; Dempsey et al., 2011). It was found that UGT74F1 and UGT74F2 are active towards SA, and the ectopic expression of UGT74F2 in *Arabidopsis* compromised plant resistance towards the pathogenic bacteria *Pseudomonas syringae* (Song et al., 2008). UGT74F2 catalyzes glucose conjugation to anthranilate via an ester linkage (Quiel and Bender, 2003), as well as produces two glucose

conjugates of salicylic acid, salicylic acid glucoside (SAG) and salicylic acid glucose ester (SGE). In contrast, UGT74F1 only produces the SAG form with higher specificity to the 2-hydroxyl group of the benzene ring (Lim et al., 2002). Plants which harbour mutations in *ugt74f1* and *ugt74f2* genes were characterized and their soluble metabolite analysis showed altered metabolite profiles (Dean and Delaney, 2008). The mutant plants accumulated different SA, where *ugt74f1* mutant plants accumulated SGE, SAG, and 2,5-dihydroxybenzoic acid 5-O-β-D-glucose, while *ugt74f2* mutant plants accumulated SAG and 2,5-dihydroxybenzoic acid 2-O-β-D-glucose. The authors indicated the importance of the two genes in SA metabolism, where changes in the enzymatic activity of UGT74F1 or UGT74F2 had dramatic effects on SA metabolism in *Arabidopsis* (Dean and Delaney, 2008). The role of *Arabidopsis* UGT74F2 in nicotinate (NA) glycosylation was also studied (Li et al., 2015), and found that higher levels of free NA accumulated in *ugt74f2* mutant plants, while salicylic acid was not increased compared to the wild-type plants. In addition, the germination defect phenotype in WHAT PLANT that was detected in the *ugt74f2*-1 mutant was complemented by overexpression of *UGT74F2*. The authors suggest that UGT74F2 may function to detoxify plant cells from over-accumulation of NA during seed germination (Li et al., 2015).

The crystal structure of UGT74F2 was shown to complex with its substrates UDP and SA, and the residues that coordinate SA (H18, Y180, and M274) were identified. Moreover, it was found that the three residues are conserved between UGT74F1 and UGT74F2. Based on the *in silico* docking of SA to UGT74F1 and UGT74F2, the authors theorized that "amino acid substitutions around the binding site are responsible for the observed activity differences between these enzymes" (Thompson et al., 2016).

Salicylate glycosides have ecological importance as a component in plant herbivore interaction. Plants in the family *Salicaceae* are known for being rich in salicylate glycosides (up to 25% of the foliar weight), such as salicin and salicortin. They play an important role in the protection against biotic (herbivore) and abiotic (UV-B radiation) stress (Hwang and Lindroth, 1997; Ruuhola et al., 2003; Babst et al., 2010). However, the pathways and catalysts responsible for biosynthesis and glycosylation of this group of metabolites are not clear.

1.7 Control of plant hormone homeostasis via glycosyltransferases

The regulation of the levels of phytohormones is a vital process enabling rapid adaptation of plants to changing environments and maintaining adequate growth (Vogt and Jones, 2000; Jones and Vogt, 2001; Bowles et al., 2005; Piotrowska and Bajguz, 2011). Conjugation followed by enzymatic hydrolysis could provide a means of regulating the levels of active phytohormones in plant cells (Bowles et al., 2006). Glycosylation has evolved to control the levels and localization of active hormones and their activities in plant cells (Ostrowski and Jakubowska, 2013). Glycosylation of hormones can be either reversible, as in most hormone glycosides, or irreversible as in N-glucosides of cytokinins (Hou et al., 2004; Woo et al., 2007). Glycosides of plant hormones were initially considered inactive forms of plant hormones; however, some conjugates of IBA were shown to be very active in *in vitro* bioassays (Ljung et al., 2002). The formation and hydrolysis of auxin conjugates has been shown to be developmentally regulated, and varies significantly among plant tissues (Rampey et al., 2004).

Auxin glycosylation, such as indole acetic acid (IAA) glycosides, is widespread in plants. The first plant auxin glycosyltransferase identified was cloned from maize (IAGLU) and shown to

glycosylate IAA (Szerszen et al., 1994). Later, UGT84B1 was isolated from *Arabidopsis* and showed high activity *in vitro* towards IAA (Jackson et al., 2001). Over-expression of this gene in *Arabidopsis* results in altered rosette size, reduced height, loss of apical dominance, and a root impaired gravitropism phenotype. Metabolic profiling of these plants showed increased IAA-Glc as well as increased IAA content (Jackson et al., 2002; Tognetti et al., 2010).

Biochemical characterization of UGT84B1, UGT84B2, UGT75B1, and UGT75B2 enzymes indicated that UGT84B1 has high activity towards IAA and lower but significant activity towards IBA, cinnamic acid, and IPA. The remaining enzymes showed only trace activities towards plant IAA (Jackson et al., 2001). UGT74E2 was also found to be active with IBA (Jin et al., 2013), and ectopic expression of the gene in *Arabidopsis* showed that the enzyme maintains its catalytic activity towards IBA *in vivo*. Metabolic profiling of the transgenic plants showed increased levels of both IBA and its glucoside IBA-Glc, as well as altered IAA patterning. Plant growth was altered, including root branching, and rosette shape, and the plants displayed increased drought and salt stress intolerance (Tognetti et al., 2010)

Recently, UGT74D1 has been identified as a novel auxin GT. It displays *in vitro* activity towards IAA, IPA, IBA, and NAA auxins, converting them to the corresponding glucose esters (Jin et al., 2013). Tanka et al., (2014) demonstrated that UGT74D1 also shows activity towards 2-oxindole-3-acetic acid. Ectopic expression of this gene results in loss of root gravitropism in UGT74D1 over-expressing plants. Similar root phenotypes were observed when the UGT84B1 auxin GT was over-expressed in *Arabidopsis* plants (Jackson et al., 2002).

Glycosylation of cytokinins involves O-glucosylation, O-xylosylation, and N-glucosylation (Mok and Mok, 2001). Zeatin is the most common cytokinin in plants, and its glucosides, the transport and/or the storage form of this hormone. Zeatin O-glycosylating (ZOG) enzymes have been identified in several plant species, including maze (ZOG1 and ZOG2), and *Phaseolus vulgaris* (ZoX1) (Martin et al., 1999 a, b; Martin et al., 2001). These enzymes show substrate stereo-specificity by forming either *cis*- or *trans*-glycosides (Pineda et al., 2008). UGT76C1 and UGT76C2 from *Arabidopsis* were shown to recognize all cytokinins and form N- and O- zeatin glucosides. The *in vivo* function of UGT76C1 towards cytokinins was also confirmed in transgenic plants; however, these plants did not show altered morphology under normal growth conditions (Wang et al., 2013).

Abscisic acid glucose ester is the most abundant conjugate of the plant hormone abscisic acid (ABA), but several other glycosides of ABA have been identified in many plant species (Xu et al., 2002; Nambara and Marion-Poll, 2005). Six UDP-glucosyltransferases (UGT84B1, UGT75B1, UGT84B2, UGT71B6, UGT75B2, and UGT73B1) exhibit activity towards ABA. The highest catalytic activity was observed with UGT84B1, but only UGT71B6 showed *in vitro* enantio-selective glycosylation towards *cis*-ABA (Lim et al., 2005). The same protein was also shown *in vitro* to glycosylate a wide range of ABA analogues (Priest et al., 2005). Over-expressing this gene in *Arabidopsis* led to accumulation of high levels of ABA-Glc. Although the fully-grown transgenic plants did not show altered morphology, a post-germination growth phenotype, suggested that a low level of ABA deficiency had the resulted from over-expressing the gene (Priest et al., 2006).

Several brassinosteroid (BR) glycosides have also been identified in plants (Fujioka and Yokota, 2003; Bajguz, 2007). Two BR GTs (UGT73C5 and UGT73C6) were found in *Arabidopsis* (Poppenberger et al., 2005; Husar et al., 2011), and UGT73C5 was shown to catalyze O-glucosylation of both brassinolide (BR) and its biosynthetic precursor, castasterone. Interestingly, the same gene product has been shown to glycosylate a fungal deoxymenivalenol (DON) mycotoxin (Poppenberger et al., 2003).

Gibberellins (GAs) are tetracyclic diterpenoids that participate in all stages of plant development, including germination, hypocotyl elongation, phase transitions, root, leaf, stem, and fruit growth, greening of leaves, flowering, and the development of flowers and seeds (Ragni et al., 2011). The biosynthesis of GA conjugates has long been recognized in various gymnosperm and angiosperm plant species (Nadeau and Rapaport, 1974). The conjugation of GA appears to produce inactive metabolites, and this may be a key step in the process of further catabolism, compartmentalization steps within the cell, or long distance transport. On the other hand, the low activity of GA conjugates in bioassays suggests that these metabolites could actually be waste products that are deposited in cell vacuoles (Stoddart and Venis, 1980).

1.8 Phenolic glycosides and UGTs in poplar

Trees belonging to the family *Salicaceae* contain significant concentrations of phenolic glycosides, and these metabolites of the phenylpropanoid pathway can represent up to 35% of leaf dry weight (Tsai et al., 2006). The genes regulating the glycosylation of this diverse group of secondary metabolites have not been extensively studied in trees (Payyavula et al., 2009). Despite their important roles in the biotic and abiotic stresses (Dixon and Pavia, 1995; Dixon et

al., 2002; Ruuhulla et al., 2003; Major and Constabel, 2006; Yu et al., 2006; Babst et al., 2010), a negative correlation between the production of these metabolites and growth rate exists (Osier and Lindroth, 2001). However, the mechanisms underlying the relationship between non-structural phenylpropanoid metabolites and growth are not clear (Payyavula et al., 2009; Babst et al., 2010). In order to identify UGT genes that are active towards salicyl alcohol, aspen cell cultures were supplied with salicyl alcohol and the expression of UGTs gene was examined. Many poplar UGT genes (74) were found to be up-regulated, and 18 of these were further screened by Q-PCR. The results showed that GT-2 and GT-246 were the most abundant UGTs expressed in the cell cultures, displaying 2-3 fold increases in activity. The authors suggested that these two enzymes might therefore be involved in the glycosylation of salicyl alcohol (Payyavula et al., 2009).

In 2012, one putative poplar UGT gene (*PtGT1*) was cloned and expressed in tobacco plants (Wang et al., 2012a). When its amino acid sequence was aligned to the *Arabidopsis* UGT72E1-E3 homologs, PtGT1 showed highest similarity to UGT72E3 (75% similarity and 54% identity). Ectopic expression of the *PtGT1* gene in tobacco plants resulted in significant increases in klason lignin content, and the transgenic plants flowered earlier than the wild-type control (Wang et al., 2012 a). Interestingly, *in vitro* assays showed that the enzyme did not recognize coniferyl, sinapyl, or *p*-coumaryl alcohols as substrates, and the transgenic plants did not show increased levels of the corresponding glycosides. The authors therefore suggested that the observed impact on lignin accumulation is not a direct effect of the enzyme and further research is needed (Wang et al., 2012 a).

Two pathogen-induced poplar UGTs have been characterized *in vitro*. These two genes were previously *in silico* identified as putative pro-anthocyanidin (PA) GTs, and were grouped with flavonoid glycosylating UGTs based on phylogenetic placement (Veljanovski and Constabel, 2013). *In vitro* characterization of the two enzymes against flavonoid substrates showed that UGT78L1 was active towards quercetin and kaempferol. The second enzyme (UGT78M1) did not show activity towards any of the tested aglycones, and neither enzyme showed activity towards PA (Veljanovski and Constabel, 2013).

Coleman et al., (2007) overexpressed a uridine diphosphate glucose pyrophosphorylase (UGP) gene in hybrid poplar trees, in an effort to increase UDP-Glc levels and thereby improve tree growth. Interestingly, cellulose deposition was increased by 5%, but the concentration of salicylic acid 2-O-beta-D-glycoside (SAG) was also significantly increased. In addition, the transgenic trees showed a reduction in stem height and diameter, leaf area, leaf dry weight, elongated axial leaves, and shoots with smaller leaves. The authors suggested that SA might be acting as a sink for excess UDP-glucose and the resulting accumulation of SAG could be inducing stress responses that decreased growth (Coleman et al., 2007).

In order to study the effect of SAG on tree growth and development, a bacterial SA hydroxylase gene (*nahG*) that converts SA into catechol was ectopically expressed in poplar trees. The *nahG* transgenic trees did not show any morphological phenotypic changes, but their metabolite profiles showed a decrease in quinic acid conjugates and increased levels of catechol glucoside, while exerting little effect on free SA and catechol levels (Morse et al., 2007). The author suggests that poplar trees might need to maintain normal levels of SA as an adaptive mechanism

(Morse et al., 2007). In contrast, tobacco plants, expressing *nah*G showed dramatic reduction in salicylic acid levels (Gaffney et al., 1993).

1.9 Objectives and hypothesis

Two poplar *PopGT1* and *PopGT2* coding sequences were isolated from hybrid poplar with high protein sequence identity to the *Arabidopsis* SAGT GTs (AtUGT74F1 and AtUGT74F2).

In order to characterize these two UGTs functionally, heterologous gene expression of the two enzymes in *E. coli* system was carried out and *in-vitro* substrate activity of the enzymes towards different classes of phenolic aglycones and activated sugars were assessed.

In order to study the effect of over-expression of the two genes on plant growth and development, the two coding sequences of *PopGT1* and *PopGT2* genes were cloned under the control of the cauliflower mosaic virus constitutive 35S promoter and transformed into *Arabidopsis* plants. The resultant transgenic plants were studied phenotypically, and their secondary metabolites profiles were analyzed.

To investigate the effect of down-regulation of the native glycosyltransferases (*PopGT1* and *PopGT2*) in hybrid-poplar, RNAi mediated suppression was employed using two independent constructs designed to reduce the gene expression of these two genes and their paralogues. The transgenic trees were assessed phenotypically, and the secondary metabolites profiles of the transgenic trees were studied.

1.9.1 Hypothesis

The two poplar glycosyltransferases (PopGT1 and PopGT2) are integral to maintaining the local tissue specific SA concentrations and maintaining normal plant growth, and development.

1.9.2 Aim of the study

The aim of this study was to isolate and functionally characterize putative poplar SAGT genes and their encoded proteins. In addition, the effects of mis-regulating SAGT gene expression on plant growth and development were to be assessed.

Specifically, the following studies were carried out:

1. Isolate and perform *in vitro* functional characterization of the poplar homologs of the *Arabidopsis* SAGT (*PopGT1* and *PopGT2*).
2. Examine the effect of ectopic expression of *PopGT1* and *PopGT2* in *Arabidopsis*.
3. Alter the expression of *PopGT1* and *PopGT2* in hybrid poplar to investigate the effects of these changes on plant growth.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals employed for buffers and media preparation were of analytical grade unless otherwise stated. Solvents were high-pressure liquid chromatography (HPLC) grade, and all molecular work was completed using molecular biology grade chemicals.

For protein assays and metabolite profiling, all the commercially available substrates were HPLC grade. The aglycone 2,5-dihydroxybenzyl benzoate was cleaved from its glycoside salireposide using almond β-glucosidase at the University of Wisconsin, Madison, USA (laboratory of Dr. J. Ralph).

2.1.2 Genotypes of hybrid poplar

The hybrid poplar genotype *Populus alba x grandidentata* (P39) was used to generate all transgenic poplar trees for the experiments described in this thesis.

2.1.3 Genotypes of *Arabidopsis thaliana*

- *Arabidopsis thaliana* ecotype Columbia (Col0)
- *Arabidopsis thaliana* ecotype Wassilewskija (Ws)

2.2 General plant methods

Poplar trees were propagated *in vitro* to obtain 10 trees of the same size for each line. Trees were then transferred to the University of British Columbia (Point Grey Campus, Vancouver, Canada) greenhouse facilities and planted in 7.56 litre plastic pots, containing perennial potting soil (50% peat, 25% fine bark and 25% pumice; pH 6.0). The pots were placed on flood-tables in a randomized design. The trees were grown under 16 hours of supplemental light (300 W m^{-2}), at 25 ºC and 20 ºC during the daytime and nighttime, respectively. All plants were watered daily with fertigated water containing macro and micronutrients (Appendix). Finally, the pots were randomly moved on the flood-tables periodically to avoid any positional effects.

2.2.1 *In vitro* propagation and maintenance of hybrid poplar trees

Hybrid poplar trees were propagated *in vitro* under aseptic conditions. Briefly, the stems were cut into short internodal segments, placed into GA7 Magenta jars containing standard woody plant medium (WPM) (Lloyd and McCown, 1980), and solidified with agar. All cultures were incubated for 3-4 weeks until new plantlets emerged. The newly emerged shoot tips were excised and transferred into new GA7 Magenta jars, containing solid WPM. The plantlets were allowed to grow for an additional 4-6 weeks. Each line was labeled and kept separately. The generated trees were then maintained by sub-culturing the shoot tips on WPM solid media every 4-6 weeks. All cultures were incubated at 22 ºC under 16 hours of daylight and an average photon flux of 40 µmol m^{-2} s^{-1}.

2.2.2 Phenotypic characterization of poplar trees

Trees were permitted to grow for six months (September - February), after which five trees per line were harvested for molecular and chemical analyses. Plant height, and stem diameter 20 cm above the root collar, were measured using a measuring tape and digital calipers, respectively, prior to harvest. Trees not harvested were maintained in the greenhouse, where they were allowed to grow, set buds, and senesce.

2.2.3 Generation of transgenic poplar trees

In order to study the effect of mis-regulating candidate gene expression on plant growth and cell wall structure, hybrid poplar (P39) was used to generate transgenic trees. Leaves were collected from 4 - 6 week-old tissue culture grown plants, and cut using a 7.0 mm cork-borer. *Agrobacterium tumefaceins* strain EHA105 (Hood et al., 1993), transformed with a chimeric binary construct, was inoculated in 5 mL liquid YEP medium containing rifampicin 50 mg/L and the appropriate selection antibiotics. The culture was incubated overnight in a shaker-incubator at 28 ºC and 200 rpm. The following day, 250 mL of liquid WPM medium, supplemented with 100 µM acetosyringone, was inoculated with the bacterial culture that was grown overnight. The new culture was then incubated at 28 ºC and 100 rpm. On the third day, the culture was diluted to an OD_{600} of 0.1 to 0.2, and then 10 mL of the bacterial culture was dispensed into 50 mL falcon tubes. To each tube, 25 leaf discs were added and co-incubated in a shaker-incubator for 1 hour at 28 ºC and 100 rpm. The leaf disks were then blotted dry on sterile filter paper and placed adaxial side down on petri dishes containing solid WPM medium supplemented with 0.1 µM of the following plant growth regulators: naphthalene acetic acid (NAA), benzyl adenine (BA), and

thidiazuron (TDZ). Twenty petri dishes were used for each of the transformations, and each petri dish contained between 20-25 leaf discs. All the plates were then incubated in the dark for three days. The leaf disks were then moved to new WPM plates supplemented with 50 mg/mL carbenicillin and cefotaxime, to eliminate the growth of the *Agrobacterium,* and incubated for two additional days in the tissue culture incubation room under subdued light. After two days, the explants were transferred to solid WPM media, containing kanamycin for selection of the *pHellsgate*12 vector.

After shoot emergence, one shoot per leaf disk was isolated and transferred to a new GA7 Magenta jar containing the same medium and supplemented with the appropriate antibiotics. The trees were sub-cultured twice on the selection medium. Vigorous shoots that established roots were chosen for further screening. The genotype of transgenic trees was confirmed by PCR screening of genomic DNA, using appropriate primer sets (Table 2). The PCR-positive plants were propagated *in vitro*, using solid WPM medium without antibiotics, until they were transferred to the greenhouse.

2.2.4 Stem sectioning and staining of hybrid poplar trees

Stems of 6-month old greenhouse-grown poplar trees were harvested 20-25 cm above the root collar, and then used to generate 40 µm cross-sections, using a Spencer AO860 hand sliding microtome. The stem sections were stained separately with either toluidine blue or 10% phloroglucinol/HCl histochemical stains, mounted onto glass slides, and visualized with a Leica DRM light microscope. Images were taken with a digital camera (Infinity 3) and visualized with

the associated Infinity Capture software. Three trees per line were sectioned and multiple sections per stem were visually inspected using different magnifications.

2.2.5 Growth conditions of *Arabidopsis* plants

Arabidopsis plants were grown in a growth chamber under a 16/8 hrs light/dark regime at 20 ºC during the day and 18 ºC at night. For light stress treatments, plants were allowed to grow under 24 hours light at 20 ºC.

2.2.6 Generation of transgenic *Arabidopsis* lines

The standard floral dip technique was used to generate transgenic *Arabidopsis* plants, as described by Clough and Bent (1998). An Erlenmeyer flask containing 50 mL of YEP medium containing the appropriate selection of antibiotics was inoculated with 5 mL of an overnight culture of *Agrobacterium tumefaciens* GV3101.

The culture was then incubated in a 28 ºC incubator shaker at 220 rpm for 16 hrs until the optical density of the culture reached an OD_{600} of 0.6-0.8. The culture was centrifuged at 5,000 rpm, re-suspended in 5% sucrose solution with 0.05% silwet and the optical density of the solution adjusted to 1.0 (OD_{600}). *Arabidopsis* plants with 2-5 opened flowers were then dipped into the *Agrobacterium* suspension solution and all plants were maintained overnight lying on their sides in a dark room. The following day, the plants were placed upright in the growth chamber and grown in isolation. The dipping event was repeated once more after 7 days.

2.2.7 Establishment of transgenic *Arabidopsis* lines

Plants that had been treated with *Agrobacterium* (T0 plants) were allowed to grow until seed maturation. Seeds were then harvested and placed in dry paper envelopes (T1 seeds). For plant screening, seeds were surface-sterilized using 70% ethanol for one minute followed by 20% commercial bleach solution for 15 minutes, and then rinsed with sterilized water. Seeds were then vernalized for three days at 4 ºC in the dark and plated on ½ MS medium containing 20-25 mg/L hygromycin for transgenic plant selection. Resistant plants that had healthy shoots and roots were transferred to 3-inch green injection molded pots (contain potting mixture), watered, and covered with transparent plastic sheets for 5 days to maintain a high level of humidity. The plastic sheets were then removed gradually over 3 days and the plants allowed to grow.

After approximately two weeks growth, one leaf from each plant was isolated and DNA extracted using the CTAB method. PCR screening for the transgenes was performed, and positive plants were identified and separated. Each such plant represents an individual insertion line (T1 plants). Plants were permitted to self-fertilize and the seeds were collected from each plant separately (T2 seeds). The T2 seeds from each line were then plated and plants (T2 plants) tested for gene expression using RT-PCR.

2.2.8 Phenotypic characterization of transgenic *Arabidopsis* plants

T2 plants of the high expressing lines were planted as previously described. For phenotypic characterization, 15 plants per line were planted and screened for the following phenotypes; rosette diameters, flowering time, stem height, stem diameter, and the number of stems per plant.

The phenotyping process was also performed for T3 generation plants. Plants were measured every week in order to estimate the increase in rosette diameter. For each plant, two readings per rosette were measured diagonally using a standard ruler. Stem height was measured following eight weeks of growth. Stem diameters were measured at maturity using digital calipers.

2.2.9 Stem sectioning and staining of *Arabidopsis* plants

Stem sections were cut by hand from 5-week old *Arabidopsis* plants using a razor blade. Thin sections were stained separately with toluidine blue, Maüle stain and phloroglucinol/HCl to investigate the effect of over-expressing the poplar genes on the anatomy of the plant inflorescence and tissue characteristics (interfascicular fibres, phloem tissue, xylem vessels, epidermal cells, and xylem fibres), as well as lignin localization patterns. Stem sections were examined visually using a Leica DMR light microscope and images were captured using a digital camera (Infinity 3).

2.3 Molecular biology methods

2.3.1 DNA extraction from plant tissues

A cetyltrimethyl ammonium bromide (CTAB) extraction method was used to isolate DNA from plant tissues. The tissues were harvested, flash frozen in liquid nitrogen and kept in a -80 ºC freezer until processing. Frozen tissues were ground to a fine powder in a pre-chilled mortar and pestle. One mL of freshly prepared CTAB extraction buffer (2% CTAB, 1.4 M NaCl, 20 mM

EDTA, 100 mM Tris HCl pH 8.0, 0.2% β-mercaptoethanol) was added to 100 mg of frozen tissue and mixed thoroughly by vortexing.

The samples were then incubated in a 70 ºC water bath for 15 minutes, cooled to room temperature, centrifuged for 10 minutes at 13,000 rpm, and the lysates transferred to new tubes. To the lysates, 700 µL phenol:chloroform:isoamyl alcohol (25:24:1) solution was added to each tube and immediately mixed by vortexing for 30 seconds. The samples were centrifuged at full speed and the aqueous phases transferred to new tubes. An equal volume of chloroform was added to each sample to extract the phenolic residues. Samples were then mixed by vortexing and centrifuged at maximum speed for 10 minutes, and the aqueous layers removed and transferred to new tubes. An equal volume of absolute isopropyl alcohol was added to each tube and the samples were incubated at room temperature for an additional 30 minutes.

All samples were then centrifuged for 15 minutes at full speed at 4 ºC. The alcohol was removed, the pellets washed with 70% ethanol and the tubes were centrifuged for 10 minutes at full speed at 4 ºC. The pellets were allowed to air dry and were then dissolved in sterile distilled water. The DNA concentration was measured using a NanoDrop spectrophotometer ND1000 (NanoDrop Technologies, USA). All DNA samples were stored at -20 ºC.

2.3.2 RNA extraction from poplar tissues

RNA was extracted from poplar tissues according to the method of Kolosova et al., (2004). Tissues were collected from actively growing trees, flash-frozen in liquid nitrogen and kept at -80 ºC until processing. For the RNA extraction process, plant tissues were ground to a fine

powder in a pre-chilled mortar and pestle using liquid nitrogen. Fifteen mL of freshly prepared RNA extraction buffer was added to 1.5 g tissue while grinding. The buffer contained 400 mM Tris-HCl (pH 8.5), 15% lauryl sulfate lithium salt, 320 mM lithium chloride, 10 mM EDTA disodium salt, 1% deoxycholate sodium salt, 1% tergitol NP-40 (w/v), 0.01 µM aurintricarboxylic acid, 1.5 mM dithiothreitol, 0.2 mM thiourea, 2% PVPP, and nuclease free H_2O. The tubes were then frozen in liquid nitrogen and kept in a -20 ºC freezer overnight.

The following day, the samples were allowed to thaw in a 37 ºC water bath with occasional mixing to ensure sample homogeneity. The samples were centrifuged at 4,000 rpm for 20 minutes at 4 ºC. The supernatants were then filtered through Kimwipes into 50 mL Falcon tubes. One thirtieth of the volume of 3.3 M sodium acetate (pH 6.1) and 1/10 volume of absolute ethanol were then added, and the samples incubated for 10 minutes on ice, and then centrifuged as previously described. The supernatants were transferred to new tubes and 1/10 volume of 3.3 M sodium acetate (pH 6.1) and an equal volume of chilled absolute isopropanol were added to each tube. All the samples were kept in a -80 ºC freezer for 30 minutes.

After incubation, the samples were allowed to thaw at room temperature and centrifuged for 15 minutes at 4,000 rpm at 4 ºC. The pellets were re-suspended in 2 mL TE buffer (20 mM Tris HCl, 1.0 mM EDTA, pH 8.0) and 2 mL of 5 M NaCl and 1 mL of 10% CTAB were added and mixed by vortexing. All tubes were incubated at 65 ºC for 5 minutes in a water bath. After incubation, an equal volume of chloroform:isoamyl alcohol was added to each tube, mixed thoroughly by vortexing, and centrifuged as previously described. This extraction was repeated once. One quarter volume of 8 M lithium chloride was added to the supernatants, and the tubes

were incubated overnight in a -20 ºC freezer. On the next day, the samples were centrifuged at 4,000 rpm for 30 minutes at 4 ºC.

The pellets were dissolved in 900 µL of TE buffer and the samples transferred to 2 mL micro-centrifuge tubes. One volume of chilled isopropanol and 100 µL of 3.3 M sodium acetate (pH 6.1) were added to the tubes and then incubated for 30 minutes at -80 ºC. The samples were centrifuged at 13,000 rpm for 30 minutes at 4 ºC. The pellets were washed with 70% ethanol and centrifuged again for 10 minutes. The pellets were re-suspended in 200 µL of autoclaved DEPC-treated water. DNase treatment was performed to eliminate any DNA contamination using the DNase I Digest kit (Ambion, USA). RNA was quantified using a NanoDrop ND 1000 spectrophotometer and 1 µg of RNA was used to generate cDNA using the SuperScript II RT kit (Invitrogen, Canada). All RNA samples were stored in a -80 ºC freezer.

2.3.3 RNA extraction from *Arabidopsis* tissues

RNA was isolated from *Arabidopsis* plant tissues using TRIzol reagent (Invitrogen, Canada) according to the manufacturer's instructions. Using a mortar and pestle, 100-150 mg tissue was ground to a fine powder in liquid nitrogen. One mL of the TRIzol reagent was added immediately to the frozen tissue, and the contents of each tube were incubated at room temperature for 5 minutes. To each sample, 0.2 mL chloroform was added, and the tubes were mixed by vortexing. All the tubes were then incubated for 2-3 minutes at room temperature, after which the tubes were centrifuged at 13,000 rpm in a bench-top centrifuge for 15 minutes at 4 ºC. The aqueous phases were transferred to new micro-centrifuge tubes and 0.5 mL of isopropanol was added to each tube and mixed. The tubes were incubated at room temperature for 10

minutes, and then centrifuged at 13,000 rpm for 10 minutes at 4 ºC in a bench-top centrifuge. The pellets were washed with 70% ethanol and centrifuged at 13,000 rpm for 5 minutes at 4 ºC. The pellets were left to air-dry, re-suspended in RNase-free water, and incubated at 60 ºC for 10 minutes to dissolve the pellets completely. The isolated RNA was treated with a *DNase*I Digest kit and quantified using a NanoDrop ND 1000 spectrophotometer. The purified RNA was stored in a -80 ºC freezer. One microgram of DNase-treated RNA was used for the synthesis of cDNA using the Superscript II RT kit. The resulting cDNA was stored in -20 ºC freezer.

2.3.4 Standard polymerase chain reaction (PCR)

PCR reactions were performed in a 20µL standard reaction mixture. The reaction mixture contained 0.2 U of Taq DNA polymerase (New England, BioLabs) or 0.2 U of Phusion High-Fidelity DNA polymerases (Thermo Fisher Scientific, Canada), 2 µL of 10X buffer, 200 mM of each dNTPs (New England, BioLabs), 20 pmol of each gene specific primers (Table 2), and milliQ water was added to bring the total volume to 20 µL. For each tube, 50-100 ng of the DNA-templet was added.

PCR amplification was done in a PTC 200 PCR thermal cycler (MJ Research, USA). The amplified products were then separated by electrophoresis using 1.5% agarose gel (agarose in Tris acetate EDTA buffer, pH 8.0). The separated DNA fragments were visualized in gel ethidium bromide staining technique. The fragment sizes were estimated using DNA molecular weight Marker, 1 kb DNA ladder (New England BioLabs, Canada). The gels were photographed using the Alpha Imager image-analysis system (Alpha Innotech, USA).

Colony PCR was performed in a similar fashion, except, the DNA template was exchanged with a 2-5 µl of 100x diluted bacterial cells suspension.

2.3.5 Gene isolation using the RACE-PCR technique

Random amplification of cDNA ends (RACE) PCR was used to clone the poplar GT genes from transgenic poplar trees (over-expressing UGP-pyrophosphorylase gene) that showed increased levels of SAG (Coleman et al., 2007). RNA was extracted and cDNA was synthesized using a SuperScript II RT kit. A RACE Clontech Kit (Clontech laboratories, USA) was used to generate cDNA fragments with adaptors attached to either the 5′- or the 3′-ends. This cDNA was the template for PCR reactions, where one of the primers was compatible with either the 5′- or the 3′-end adaptor and the second primer was a gene-specific primer. PCR products, which showed the expected product sizes, were cloned using a TA cloning kit and sequenced using M13 F and M13 R primers (Table 2). The DNA sequence obtained was then used as a query in a Blast database search to investigate its similarity to the SAGTs genes from other plants.

2.3.6 Purification of PCR product

PCR purification was performed using a QIAquick PCR purification kit (Qiagen, Canada) as per the manufacturer's instructions. The PCR product was mixed with a binding buffer and then applied to the membrane of the supplied column. The column was then centrifuged at maximum speed for 1 minute and washed with washing buffer. The PCR products were then eluted from the membrane, using 30 to 50 µL of elution buffer or milliQ water. The purified products were stored in a -20 ºC freezer.

2.3.7 Cloning of PCR product

Purified PCR product was cloned into the TOPO ZeroBlunt vector (Invitrogen, Canada) according to the manufacturer's instructions. One µL of the cloning vector was mixed with 50-150 ng PCR product, 1 µL of salt solution, and H_2O was then added to bring the volume to 6 µL. The reaction was left at room temperature for 15 minutes, and then 3 µL of the reaction mixture was used to transform competent *E. coli* cells (Top10). The sequence was then confirmed by sequencing, using M13 F and M13 R primers (Table 2).

2.3.8 DNA sequencing

In each sequencing reaction, 150 ng plasmid DNA was added to the Big-dye reaction mixture (3 µL) along with 1 µL 5 pmol/µL sequencing primer (Table 2) and water to a final volume of 10 µL. The sequencing reaction was performed in a PTC 200 PCR thermal cycler (MJ Research, USA) using the following PCR program; one denaturation step at 98 ºC for 10 minutes, 24 cycles of 98 ºC for 5 seconds, 55 ºC for 10 seconds, 72 ºC for 30 seconds, and a final extension step at 72 ºC for 7 minutes. The sequencing reaction products were purified and sequenced by the Nucleic Acids Protein service unit (NAPS), Michael Smith Laboratory (MSL), at UBC.

2.3.9 *In silico* analysis of the poplar UGT sequences

DNA sequence editing, translation into the corresponding translated amino acids, and alignment was performed using the BioEdit software (Hall, 1999). Multiple sequence alignments were completed using the online Clustal Omega sequence alignment software (Sievers et al., 2011)

(http://www.genome.jp/tools/clustalomega/). BLAST search for protein domain signature was completed online. Protein folding analysis of the translated amino acid sequences was performed using the Phyre-2 protein recognition online server (Kelley et al., 2015) (http://www.sbg.bio.ic.ac.uk/phyre2). The predicted fold-structures were viewed and cartoon representations were created using the ByMol software (www.ByMol.com).

In silico identification of the physico-chemical properties of the translated amino acid sequences was completed using the ExPASy-ProtParam software (http://web.expasy.org/protparam).

2.3.10 Construction of phylogenetic trees

In order to study the relatedness of the two poplar proteins to the members of the *Arabidopsis* family 1 UGTs, two phylogenetic trees were constructed, using the two poplar translated amino acid sequences, as well as the protein sequences of family 1 UGTs (obtained from the CAZY website). All sequences were aligned using multiple alignment software Clustal Omega (http://www.genome.jp/tools/clustalomega). The aligned sequences were used to generate 2 neighbouring join trees using Mega5 sequence analysis software (Tamura et al., 2011). The first tree was constructed to determine the phylogenetic group that positioned the isolated genes within family 1 UGTs. Meanwhile, the second tree indicated the relationship between the two poplar protein sequences and the members of the *Arabidopsis* phylogenetic group L (Ross et al., 2001).

2.4 Construction of expression vectors

2.4.1 *pSM-PopGT1* over-expression construct

The coding sequence of the *PopGT1* was amplified from the cDNA hybrid poplar using the RACE technique, as previously described (section 2.3.5). The amplified sequence was cloned into the TOPO ZeroBlunt vector. The plasmid was then transformed into competent *E. coli* Top10 cells and the gene sequence confirmed using M13F and M13R primers. Two primers containing *Bam*HI and *Kpn*I restriction sites were used to amplify the gene and clone it into the multiple cloning site of the *pSM*3 binary vector under the control of 2x35S constitutive promoter (Table 2). The resulting construct was transformed into competent *E. coli* Top10 cells, sequenced, and then used to transform *Agrobacterium tumefaciens* GV3101. *Arabidopsis* plant transformation was completed using the floral dip technique (Clough and Bent, 1998). The *Arabidopsis* lines over-expressing the *PopGT1* gene (OE-*PopGT1*) were screened and maintained.

2.4.2 *pSM-PopGT2* over-expression construct

The coding sequence of the *PopGT2* gene was cloned from the hybrid poplar P39 cDNA. The same procedure described in the generation of the *pSM-PopGT1* construct was employed to create the *pSM-PopGT2* over-expression construct. The construct was confirmed by sequencing and the binary construct was used to transform *Agrobacterium tumefaciens* GV3101 to generate transgenic *Arabidopsis* plants over-expressing the poplar gene (OE-*PopGT2*). The generated lines were screened and maintained.

2.4.3 *pHellsgate* RNAi-*PopGT* constructs

A 350 bp fragment (positions 70-420) of the *PopGT1* gene was amplified using gene-specific primers that were designed to include the attB sequences (Table 2). The amplicon was then cloned into the Topo ZeroBlunt vector and confirmed by sequencing. The fragment was inserted into the *pDONR*221 vector, and then recombined into the sense and antisense loci of the *pHellsgate*12 vector to generate the silencing construct *p*Hells-*PopGT1* using recombinase enzyme (Invitrogen, Canada). The same procedure was followed to generate the second silencing construct, *pHells-PopGT2*, also using a 350 bp (position 90-440) fragment. Both constructs were screened by restriction enzyme digestion to validate the insert size and intron orientation. Constructs with the correct insert size and orientation were then sequenced using two vector-specific primers (Table 2). The confirmed hybrid constructs were transformed into *Agrobacterium tumefaciens* EHA105 (Hood et al., 1993), which was subsequently used to transform hybrid poplar trees using the leaf-disc *Agrobacterium*-mediated transformation technique.

2.4.4 *pET* (+) bacterial expression constructs

The coding sequences of the two poplar genes were amplified using chimeric primers that contain the required restriction cleavage sites (Table 2), and the PCR products were cloned into the multiple cloning site of *pET30a* (+) and *pET28* vectors via the *Kpn*I and *Not*I restriction sites. The constructs were designed to express the proteins with N-terminus 6X-His tags. A stop codon was incorporated into the reverse primer to stop protein translation and obtain the protein only in the N-terminus. The constructs were sequenced to confirm that the coding sequences of the

genes were in frame with the promoter. The confirmed constructs were then used to transform chemically competent *E. coli* DE3 cells or a Rosetta DE3 expression strain.

2.4.5 *pGEX*-4T2 bacterial expression constructs

The coding sequences of the two poplar genes were amplified using chimeric primers containing the *Kpn*I and *Not*I restriction enzyme cleavage sites (Table 2). The PCR products were cloned into the multiple cloning site of the *pGEX*-4T2 vector via the *Kpn*I and *Not*I restriction sites. The constructs were designed to express the target proteins with an N-terminus GST tag. The constructs were sequenced to confirm the structure of the expression vector. The confirmed construct was used to transform the competent *E. coli* BL21 expression strain.

2.4.6 *pPICZB Pichia* expression constructs

The coding regions of the two poplar genes were again amplified using modified chimeric primers containing the *Kpn*I and *Not*I restriction enzyme cleavage sites (Table 2). The PCR products were cloned into the multiple cloning site of the *pPICZB* vector via the *Kpn*I and *Not*I restriction sites. The constructs were designed to express the proteins with a 6X-His tag at either the C or N terminus of the protein for the *pPICZB* vector, according to the manufacturer's protocol. The constructs were then transformed into *Pichia pastoris* GS115 strain via electroporation.

2.5 Transformations

2.5.1 Transformation of competent *E. coli* cells

Competent cells were removed from the -80 °C freezer and allowed to thaw slowly on ice. Three to five µL of the ligation mixture was added to each tube and the cells were mixed gently. The mixture was incubated on ice for 15-30 minutes. The cells were then exposed to heat-shock treatment at 42 ºC for 1 minute, and immediately returned to the ice to rest for 10 minutes. To each tube, 250 µL of SOC medium (Invitrogen, Canada) was added, and the tubes were incubated in a shaker-incubator at 37 ºC for 1.5 hours. On a petri dish containing solid LB medium supplemented with the appropriate antibiotics, 50-100 µL of the culture was plated and incubated at 37 ºC for 16 hours in a microbial incubator. Positive colonies were confirmed by colony PCR using gene-specific primers.

2.5.2 Transformation competent *A. tumefaciens* cells

Competent cells of the *Agrobacterium* strain were removed from the -80 °C freezer and allowed to thaw on ice slowly. One µg of plasmid construct was added to the cells and incubated on ice for 30 minutes, and the cells were then heat-shocked for 5 minutes in a 37 ºC water bath. To each tube, 1 mL YEP medium was added and the tubes were incubated at 28 ºC for 2-4 hours. Tubes were centrifuged, and the pellets were re-suspended in 100 µL of fresh YEP medium. The cells were then spread on YEP plates containing the appropriate antibiotics, and the plates were incubated in a 28 ºC microbial incubator for 2-3 days. Positive colonies were confirmed by colony PCR screening using gene specific primers. The positive strains were stored as 50% glycerol stocks in a -80 °C freezer.

2.6 Plasmid purification

Two to five mL of each of the bacterial liquid cultures was centrifuged at maximum speed for 5 minutes. The supernatants were discarded and the pellets were re-suspended in 250 µL of resuspension buffer. To each tube, 250 µL of lysis solution was added and the cells were mixed slowly for 2-5 minutes at room temperature followed by the addition of 350 µL of neutralization solution. All tubes were then centrifuged for 7 minutes at full speed. Seven hundred µL of the supernatant was then pipetted into a plasmid mini-preparation column (Qiagen, USA) and incubated for 1 minute at room temperature. The columns were centrifuged for 1 minute at full speed and 0.7 mL of wash solution was added to each column and allowed to set for 1 minute at room temperature. Columns were centrifuged for 1 minute at full speed and the flow-through was discarded. The spinning procedure was repeated one additional time. The DNA was eluted from the column using 30-50 µL of distilled water. Finally, the DNA concentration was determined using a NanoDrop ND 1000 spectrophotometer.

2.7 Primers design

Primers were designed using Primer3 software (Untergasser et al., 2012). The gene-coding sequences were amplified using forward primers that included the start codon and reverse primers that contained the stop codon. In order to clone the coding sequences into the multiple cloning sites of the binary vectors, gene-specific primers containing restriction enzyme recognition nucleotide sequences were used to add the restriction sites to both ends of the coding sequences. The chimeric PCR product was then inserted into the multiple cloning site of binary vectors, in frame with the promoter.

2.8 Real-time quantitative PCR (RT-PCR)

Plants originally screened using antibiotic resistance were subsequently screened using qRT-PCR to determine transcript abundance. Several dilutions of cDNA, which was synthesized from the extracted RNA, were used for the optimization of quantitative RT-PCR reactions. All primers employed are listed in Table 2. Three biological replicates and three experimental replicates were used for each experiment. For expression analysis, transcription-initiation factor 5α (TIF α5) and the actin genes were individually employed as reference housekeeping genes using gene specific primers (Table 2).

2.8.1 Data analysis

All data were analyzed using the CFX-Q PCR software and Microsoft Excel to calculate the relative gene expression using the difference between the gene of interest and the reference gene using the following equation: $(2^{-(\text{Cq sample - Cq reference})})$ (Unda et al., 2012).

Table 2: Primer sequences and their applications

Primer name	Primer Sequence	Modification	Usage
P7-F	5´-ATGGAGGAGTCATGGAAAGGGC-3´		Amplification of *PopGT2*
P7-R	5´-CTACCCCACCTACTCAACAATCT-3´		Amplification of *PopGT2*
660-F	5´-ATGGAGAGGGAACAAAAAACC-3´		Amplification of *PopGT1*
660-R	5´-TTAAGGTCTCTGATCAGGACTG-3´		Amplification of *PopGT1*
Pi-P7 His-F	5'-GTACCATAATAATGTCTCATCATCATCATCATCATGA GG AGTCATGGAAAGGGC-3´	6X-His tag-*Kpn*I	Expression *PopGT2* in *Pichia*
Pi-7R stop	5´-TGCGGCCGCTCAATTCTTGAATTTTAGAATCTTAGAT A CC-3´	Stop codon, *Not*I	Expression *PopGT2* in *Pichia*
Pi-660-F	5´-CGGTACCATAATAATGTCTCATCATCATCATCATCAT G AGAGG GAACAAAAAACCAG -3´	6X-His tag, *Kpn*I	Expression *PopGT1* in *Pichia*
Pi-660-R	5´-ATGCGGCCGCTTACGGACGCTGATCCGGGCA-3´	Stop codon, *Not*I	Expression *PopGT1* in *Pichia*
5´ *AOX1*	5´-GACTGGTTCCAATTGACAAGC-3´		Screening of *Pichia*
3´ *AOX1*	5´-GCAAATGGCATTCTGACATCC-3´		Screening of *Pichia*
P27-5	5´-GGGATGACGCACAATCC-3´		Sequencing *pHellsgate*
P27-3	5´-GAGCTACACATGCTCAGG-3´		Sequencing *pHellsgate*
820-F	5′-GCAAAAGATTTTAAGAGAGAAAGGGGGAG-3´		Amplification of *UGT74F2*
820-R	5´-CTGTATGTGAGTTACCTATTTGCTCTGAACCC-3´		Amplification of *UGT74F2*
At-Tub-F	5´-TTCTCGATGTTGTTCGTAAGGAAGC-3´		Q-RT PCR of *Arabidopsis*
At-Tub-R	5´-AGCTTTCGGAGGTCAGAGTTGAGTT-3´		Q-RT PCR of *Arabidopsis*
***pET*-P7F**	5´-CCGGATCCATGGAGGAGTCATGGAAAG-3´	*Bam*HI	Cloning of *PopGT2* in *pET* expression vector
***pET*-P7R**	5´-CCGCGGCCGCTCAATTTCTGAATTTTAG-3´	*Not*I	Cloning of *PopGT2* in *pET* expression vector
***pET*-660F**	5´-CCGGATCCATGGAGAGGGAACAAAAAACC-3´	*Bam*HI	Cloning of *PopGT2* in *pET* expression vector
***pET*-660 R**	5´-CCGCGGCCGCTTAAGGTCTCTGATCAGGACTCT-3´	*Not*I	Cloning of *PopGT2* in *pET* expression vector
Q-P7F	5´-GGATTGCCTCTGCTTGAGTC-3´		Q-RT PCR of poplar *PopGT2*
Q-P7R	5´-TATCCACTGCCTCAGCCTCT-3´		Q-RT PCR of poplar PopGT2

Primer name	Primer Sequence	Modification	Usage
Q-660F	5′-CTATGGTCTCAGCCTCTTCAAGC-3′		Q-RT PCR of poplar *PopGT1*
Q-660R	5′-ACCCATCTCTGTTGTCTCCTCTAC-3′		Q-RT PCR of poplar *PopGT1*
Q-TIF 5F	5′-GACGGTATTTTAGCTATGGAATTG-3′		Q-RT PCR of poplar
Q-TIF 5R	5′- CTGATAACACAAGTTCCCTGC-3′		Q-RT PCR of poplar
***pHells* P7-F**	gggg acaagtttgtacaaaaaagcaggctGCCACTCTTGTAACATCCAT	attB1	Cloning of *PopGT2* in *pHellsgate* 12
***pHells* P7-R**	ggggaccactttgtacaagaaagctgggtcctaCAGAGGCAATCCAGGAAT	attB2	Cloning of *PopGT2* in *pHellsgate* 12
***pHells* 660-F**	5’-ggggacaagtttgtacaaaaaagcaggctCACCTCCATGCATCAAGA CA-3′	attB1	Cloning of *PopGT1* in *pHellsgate*12
***pHells* 660-R**	ggggaccactttgtacaagaaagctgggtcctaCGAACCTGGACCATGAAC	attB2	Cloning of *PopGT1* in *pHellsgate*12
T7 promoter-F	5′-TAATACGACTCACTATAGGG-3′		Sequencing of protein expression vector
T7 promoter-R	5′-TAGTTATTGCTCAGCGGTGG-3′		Sequencing of protein expression vector
***pSM*-p7-F**	5′-CTGGATCCATGGAGGAGTCATGGAAAGGGCA-3′	*Bam*HI	Cloning of *PopGT2* in *pSM*3 vector
***pSM*-p7-R**	5′-CTGGTACCTCAATTCTTGAATTTTAGAATC-3′	*Kpn*I	Cloning of *PopGT2* in *pSM*3 vector
***pSM*-p660-F**	5′-CTGGATCCATGGAGAGGGAACAAAAAACCAGC-3′	*Bam*HI	Cloning of *PopGT1* in *pSM*3 vector
***pSM*-660-R**	5′-CTGGTACCTTAAGGTCTCTGATCAGGATCT-3′	*Kpn*I	Cloning of *PopGT1* in *pSM*3 vector
Kanamycin F	5′-ACGGGAAAAGGACATGATGC-3′		Kanamycin screening
Kanamycin R	5′-CGGACGCAGAAGGCAATGT-3′		Kanamycin screening

Capital letters represent the primer portion that was designed based on the gene sequence, and lower case letters represent the additional nucleotides required for gene cloning.

2.9 Recombinant protein expression analysis

2.9.1 Protein expression in *Pichia pastoris*

The candidate genes were cloned into a *pPICZB Pichia* expression vector with a 6X-His tag. One construct included the 6X-His tag at the C-terminal and another featured an N-terminal tag. The constructs were electroporated into the *Pichia pastoris* GS115 strain according to the recommended Invitrogen manual. The cells were screened on YPS medium containing up to 500 mg/L zeocin to select for colonies containing multiple copy-number insertions. Among the positive strains, 10 colonies were evaluated by PCR. The positive colonies were then tested for the mutant phenotype. All positive colonies were tested for small-scale protein expression under pH-controlled and non-controlled growth conditions. Protein expression was analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Two colonies were selected for further large-scale protein expression. Colonies were inoculated and methanol induction of protein expression was performed as described by the manufacturer's protocol. Protein purification was employed using Ni-NTA agarose beads, and the protein was analyzed by SDS-PAGE.

2.9.2 Protein expression in *E. coli*

To examine the function of the isolated genes, a heterologous gene expression approach was followed to produce recombinant protein using a prokaryotic system. An *E. coli* expression system was tested for the compatibility of producing soluble and functional enzymes. The coding regions of the genes were isolated and cloned in the multiple cloning site of the *pET*30 a (+), *pGEX* 4-T2, and *pET28* a (+) protein expression vectors using the appropriate restriction enzyme.

2.9.2.1 Induction protocol

An overnight culture was used to inoculate a 50 mL LB medium in a 250 mL flask. The culture was incubated at 37 ºC in a shaker incubator until the culture reached an OD_{600} of 0.4-0.8. The cultures were cooled to the required induction temperature, and then IPTG was added to the culture at a final concentration between 0.1-1.0 mM to initiate protein induction. The cultures were then incubated at the required temperature (16-37 ºC) for protein induction in a 225 rpm shaker incubator for 24-48 hours. The cultures were sampled every 6 hours to determine the best incubation time. During the incubation period, bacterial pellets were collected from 1 mL sample of the *E. coli* cultures by centrifugation at 13,000 rpm for 5 minutes. Each pellet was re-suspended in 250 µL BugBuster bacteria-lysis buffer and incubated at room temperature for 30 minutes. The tubes were then centrifuged and the supernatant was transferred to a fresh micro-centrifuge tube. Both the supernatant and the pellets were analyzed for the presence of recombinant protein using SDS-PAGE.

2.9.2.2 SDS polyacrylamide gel electrophoresis

SDS polyacrylamide gels (10% separating gel and 4% staking gels) were prepared using a pre-mixed solution of 37:1 acrylamide:bisacrylamide (BioRad, USA) in mini gel trays. Protein samples were mixed with 2X SDS loading buffer and heated for 5-10 minutes at 95 ºC. The samples were then loaded onto the SDS gel and run at 140 volts for approximately 1 hour. Gels were stained using Coomassie Blue stain and de-stained using a de-staining solution (methanol acetic acid water). The protein bands were inspected on a white light box and images were collected using an Alpha Imager Image Analysis system.

2.9.2.3 Recombinant protein purification

One litre of induced bacterial culture was pelleted by centrifugation. The cell membranes of the pelleted cells were disrupted in a cell lysis buffer (50 mM MOPS, 10 % glycerol, 300 mM NaCl, pH 7.2) containing a proteinase inhibitor cocktail by sonication for 3 minutes using short cycles of a 5 second burst followed by 10 seconds of resting on ice. The cell suspension was then centrifuged in a Thermo Scientific Sorvall Legand XTR ultra-centrifuge at 23,000 rpm for 30 minutes at 4 °C and the clear lysate was collected in a new tube.

Protein purification was performed using Ni^{+2}-NTA agarose beads (Qiagen, USA). The beads were washed and equilibrated with the binding buffer according to the manufacturer's recommended protocol. The supernatant was mixed with 1 mL of the equilibrated beads and incubated in a rocking shaker for 1-3 hours at 4 ºC. The mixture was then loaded on a BioRad column and allowed to elute by gravity. The Ni^{+2}-NTA protein retaining beads were washed using washing buffers containing 30 mM imidazole, while the protein was eluted using elution buffer containing 400 mM imidazole.

Protein desalting and buffer exchange were carried out using 30 kDa Vivaspin columns (VivaSpin, GE Healthcare, USA). The purified protein was kept in a buffer containing 50 mM MOPS, 15% glycerol, and 14 mM β-mercaptoethanol, pH 7.2. The enzyme was dispensed into microcentrifuge tubes, 50 µL in each tube, flash frozen in liquid nitrogen, and stored in a -80 ºC freezer.

2.9.3 Protein substrate specificity assay

The enzyme activity was tested towards different phenolic substrates. The assays were performed in a 200 µL reaction mixture containing 1.0 mM UDP-glucose, 1.0 mM phenolic substrate, 50 mM MOPS buffer pH 7.2, 14.0 mM β-mercaptoethanol, 0.8 mM $MgCl_2$ and de-ionized water to a final volume of 200 µL, with the recombinant enzyme added last. The reaction mixture was incubated for 0.5, 1, 2, 6, and 24 hours, and stopped by adding an equal volume of methanol. The samples were centrifuged, filtered using 0.45 µM syringe filters, and analyzed by reversed-phase HPLC.

2.9.3.1 Temperature profile analysis of the recombinant enzyme activity towards quercetin

The enzymatic activity of the recombinant enzyme at different incubation temperatures (20-70 °C) was measured. A reaction mixture was prepared in a master tube with all the enzyme assay components (1.0 mM UDP-glucose, 1.0 mM quercetin, 25 mM MOPS buffer pH 7.2, 0.8 mM $MgCl_2$, and deionized H_2O up to the final volume). The mixture was mixed gently to ensure homogeneity of the solution, and 500 µl aliquots of the master reaction mix were dispensed into separate tubes, then each tube was incubated separately in an incubator, previously adjusted to one of the target temperatures (25, 30, 35, 40, 45, 50, 55, 60, and 70 °C).

After mixing all the assay components, 100 µL was extracted and placed into a new tube containing an equal volume of absolute methanol, as the start-time control. After 30 minutes of incubation, 100 µL of the assay mixture was removed to a separate tube and the reaction stopped by adding an equal volume of absolute methanol, after which the tubes were vortexed

thoroughly. The reaction tubes were stored in a -20 °C freezer until samples were ready to be analyzed by reversed-phase HPLC. The same procedure was repeated for each incubation time (1, 2, 4, and 24 hours) and the samples were stored, as previously described.

2.9.3.2 pH profile of the recombinant enzyme activity towards quercetin

The activity of the recombinant enzyme towards quercetin as a substrate was tested in different pH-buffered reaction mixtures. Different buffers (Acetate, MOPS, HEPES, CAP, and TAP) were prepared and the pH adjusted for each buffer, within the buffering range. Reaction mixtures, containing the entire assay component except the buffer, were prepared and the solution was distributed into separate tubes. The buffers were then added separately to each tube to a final concentration of 25 mM. All the assay tubes were incubated at 37 °C for two hours. After two hours, all the enzymatic reactions were stopped by adding an equal volume of methanol followed by vortexing. The samples were stored at -20 °C until all samples were ready to be analyzed by reversed-phase HPLC.

2.9.3.3 Substrate specificity of the PopGT1 enzyme

The substrate specificity of the enzyme was tested on a wide variety of phenolic substrates that span several phenylpropanoid metabolites and plant hormones (quercetin, kaempferol, naringinin, catechine, taxifolin, hesperetin, phloretin, myricetin, hydroxyl cinnamic acid, benzoic acid, catechol, salicylic acid, salignin, gentisic acid, 4 hydroxy benzoic acid, 3 hydroxy benzoic acid, cyaniding, pelargonidin, cinnamic acid, sinapic acid, courmaric acid, ferulic acid, indole acetic acid, indole butyric acid, gibberellic acid, naphthalene acetic acid, 6-benzyl amino purine,

coumarin, 6-hydroxy 7-methoxy coumarin, 7-hydroxy 6-methoxy coumarin, sinapyl alcohol, cinnamyl alcohol, coumaryl alcohol, salicylaldehyde, , and 2,5-dihydroxybenzyl benzoate).

For each potential substrate, the assay was completed as described in section 2.13.4 under the optimum temperature and pH, and the enzymatic activity was assessed by the reversed-phase HPLC. Following the completion of the assay, positive reactions were acid hydrolyzed and the products analyzed again by reversed-phase HPLC, and the hydrolyzed products were again compared to the corresponding aglycone standards.

2.10 Chemical analysis methods

2.10.1 Chemical analysis of plants metabolites

Phenolic metabolites were extracted from 0.1 g ground leaf tissue with 1 mL of 80% methanol for 1 hour in an ice-cold sonicating water bath. Five µL of 10 mg/mL internal standard was added to the tissue during the extraction. The tubes were centrifuged for 5 minutes at maximum speed (13,000 rpm) and the extraction procedure repeated. The supernatants were combined and solvent evaporated in an Eppendorf vacuum concentrator model 5301 (Eppendorf, Germany). The pellets were re-suspended in 0.5 mL of 100 % methanol, filtered through a 0.45 µm syringe filter and phenolic metabolites were analyzed by reversed-phase HPLC.

2.10.1.1 Reversed-phase HPLC analysis of total phenolic metabolites

The methanolic extracts were analyzed using a Summit Analytical HPLC System (Dionex Corporation, USA), where the phenolic extracts were separated using a Symmetry® C18

reversed-phase column (4.6 x 250 mm). In order to detect and quantify the phenolic metabolites, a 10-20 µL sample was injected onto the column at 0.7 mL/min mobile phase and screened with a diode array detector using scanning range between 205-420 nm. The mobile phase was an increasing gradient of solvent A (acetonitrile:methanol (75:25) acidified with 0.1% trifluoroacetic acid) and solvent B (0.1% aqueous trifluoroacetic acid). The retention time and the UV spectra were compared among the phenolic compounds, and the data collection and peak area integration was done using Chromeleon software. Mass-spectral analysis was completed at the University of Wisconsin, Madison, USA (laboratory of Dr. J. Ralph).

2.10.1.2 Analysis of acid hydrolyzed phenolic metabolites

One-half mL of the plant methanol extract was mixed with an equal volume of 2 M HCl and heated at 90 ºC for 1 hour. The solution was then cooled to room temperature and 0.7 mL of ethylacetate:cyclohexane (1:1 (v/v)) added to phase-separate the reaction. The organic layer (upper-phase) was collected to a new Eppendorf tube and the same procedure was repeated twice. The ethylacetate was then evaporated using an Eppendrof Vacufuge 5301 apparatus. The pellet was dissolved in 0.5 mL absolute methanol and 10-20 µL of the methanol extract was analyzed by reversed-phase HPLC, as previously described.

2.10.1.3 Soluble sugars analysis

Frozen leaf tissues were ground in liquid nitrogen using a mortar and pestle. The ground tissues were lyophilized for 24 hours in a Labconco freeze dryer (Labconco, USA). Then, 50 mg of the dried tissue was weighed into a 15 mL centrifuge tube and 50 µL of galactitol (5 mg/mL) was

added as an internal standard. Four mL of methanol:chloroform:water solution (12:5:3) was added, and the samples were mixed by vortexing and left overnight at 4 ºC for extraction.

The next day, the samples were centrifuged at 6,000 rpm for 10 minutes and the supernatant was collected to a new tube. The pellet was washed twice with 4 mL of the extraction solution. The pellet was then dried and used later for starch analysis. Five mL of water was added to the pooled supernatant and mixed by vortexing, and then centrifuged for phase separation at 6,000 rpm for 5 minutes. The upper phase was transferred to a 20 mL scintillation vial, and 2 mL of the upper organic phase was then dried in an Eppendrof Vacufuge 5301 apparatus. The pellet was re-suspended in 1 mL deionized water and filtered through a 0.45 µm syringe filter into an HPLC vial. Soluble sugars were analyzed using a Dionex HPLC system (Dionex Corp., USA) fitted with a Hiplex CA column (Agilent) maintained at 70 ºC, and the soluble sugars were separated with a mobile phase of water with flow rate of 0.17 mL/min. Detection of the carbohydrates was achieved with an electrochemical detector Dionex ED-50.

2.11 Structural characteristics of poplar wood

2.11.1 Fibre quality analysis

Fibre cell length of woody stem tissue was measured on samples obtained approximately 25-30 cm above the root collar. The wood tissue was separated from the pith and digested with Franklin solution (1:1, 30% peroxide:glacial acetic acid) at 70 °C for 48 hours. The samples were then washed with water until a neutral pH was achieved. Fibre length was measured on a Fibre Quality Analyzer (FQA, OpTest Equipment, Canada) by suspending the washed fibres in 100-200 mL water to obtain a fibre count of 25-40 fibres per second. The average length was

estimated as the average of 5,000 readings. Fibre length was measured for five trees from each transgenic line and their corresponding controls.

2.11.2 Wood density

Debarked poplar stem segments, 1.0 cm in length, were precision cut to 1.68 mm-thick sections from bark to bark using a pneumatic saw. The sections were then extracted with hot acetone overnight in a Soxhlet apparatus. Samples were allowed to dry in an oven at 50 ºC, and then scanned by X-ray densitometry (Quintek Measurement System, USA). The density was measured on both sides of the pith and the average density calculated for each sample.

2.11.3 Determination of Young's Modulus of Elasticity (MOE)

Debarked poplar stems were cut from bark to bark into 1.68 mm-thick sections 3-4 cm in length in the longitudinal direction using a pneumatic saw. In order to have uniform sections, the two sides of the stem sections were split from the pith into two halves. Each half was sanded from both sides using a fine sand paper and the dimensions of the samples were measured using digital calipers. A three point bending test was then performed to determine Young's Modulus of Elasticity (MOE), using an Instron tensometer 5565 (Instron, USA) equipped with a custom-made three point bending jig to determine the stress-strain relationship of the wood samples. The distance between the two fixed blades was 1.4 cm and a constant force (15 ºN) was applied to break the wood samples. BlueHill 3 software (Instron, USA) was used to calculate the MOE, based on the slope of the straight part of the stress-strain curve.

2.12 Chemical composition of cell wall

2.12.1 Klason lignin analysis

The cell wall chemical composition of dried wood samples from 6-month old greenhouse grown trees was determined following a modified Klason protocol (Cullis et al., 2004). The stems were harvested and the bark separated from the wood tissue and left to dry at room temperature with ambient humidity. The pith was removed from the stem and the samples were ground in a Wiley mill to pass through a 40-mesh screen. Then samples were extracted overnight with hot acetone using a Soxhlet apparatus. Extractive-free wood powder (200 mg) was weighed into a test tube, 3 mL 72% sulphuric acid was added and the samples were incubated at room temperature for 2 hours. The acid was then diluted to 3% by adding 112 mL water. Samples were transferred into serum bottles that were sealed and autoclaved at 121 ºC for 60 minutes. Samples were filtered through medium-coarseness crucibles and the filtrate was used to determine the concentration of acid-soluble lignin and structural carbohydrates. Crucibles containing the retentate were washed with deionized water and oven-dried overnight at 105 ºC. The crucibles were kept in a desiccator until cool. Acid-insoluble lignin was determined by the weight difference between the weight of the dry, clean crucibles and the weight of the lignin-containing crucibles. Acid-soluble lignin was determined by spectrophotometry by measuring the UV absorption at 205 nm and calculated using the Beer-Lambert equation and an extinction coefficient of 105 (M^{-1} cm^{-1}).

2.12.2 Structural carbohydrates of poplar stems

Structural carbohydrates were analyzed by HPLC analysis of the aforementioned collected filtrate. One mL of the filtrate was weighed into an HPLC vial and 50 µL of 5 mg/mL fucose

added as an internal standard. Samples were filtered through a 0.45 µm syringe filter, and 20 µL injected onto a DX-600 anion exchange HPLC system (Dionex Corp, USA) fitted with a CarboPac PA1 column (Dionex-Thermofisher) and equipped with pulsed amperometric detector with a gold electrode. A flow rate of 1.0 mL/min was applied as a mobile phase and a post-column addition of 100 mM NaOH was used to facilitate the electrochemical detection of the carbohydrates by the detector. Three-point calibration curves using high, medium, and low standard-sugar mixtures were used to calculate the concentration of the sugars (glucose, mannose, xylose, galactose, arabinose, and rhamnose). Peaks were integrated using Chromeleon software and calculations were done using MS Excel software. The statistical significance between the wild-type control and the transgenic samples was assessed using Student's test (*t*-test).

2.12.3 Determination of alpha-cellulose

Alpha-cellulose was analyzed according to Yokoyama *et al.*, (2002). Two hundred mg extractive-free ground wood samples (40-mesh) were weighed in glass vials, then 3.5 mL (60 mL of glacial acetic acid, 1.3 g NaOH/L) and 1.5 mL of 20% sodium chlorite were added to each sample and the mixtures incubated at 50 °C with gentle shaking for 16 hours. The reactions were stopped by placing the vials on ice and the reaction was repeated a second time, transferred to a pre-weighed coarse crucible, washed twice with 50 mL of 1% glacial acetic acid, and rinsed once with 10 mL acetone.

The crucibles were dried in a 50 °C oven overnight and then weighed. The α-cellulose fraction was isolated from 100 mg holocellulose by reacting the substrate with 17.5% NaOH for 30

minutes. The solution was diluted to 8.75% NaOH and left for an additional 30 minutes. The remaining retentates were filtered through pre-weighed coarse crucibles and washed with deionized water. The retentate inside the crucible was then soaked in 1 M acetic acid for 5 minutes and washed with deionized water. The crucibles were left to dry at 50 °C overnight and weighed the next day to calculate cellulose content.

2.12.4 Monolignols analysis

The lignin monomeric composition (syringyl:guaiacyl; S:G) was determined using the thioacidolysis procedure according to Robinson and Mansfield (2009). Ten mg of oven-dried extractive-free wood was weighed into a screw-cap vial, to which tetracosane was added as an internal standard (2 mL of 25 mg/L in dichloromethane). Following derivatization, the monomer ratio was determined by gas chromatography (GC) on a Polaris-Q ion trap system equipped with an AS2000 autosampler, a 30 m, 0.25 mm diameter J&W DB-5 column. The injector temperature was maintained at 250 ºC, the initial oven temperature was 130 ºC, and the detector chamber was kept at 270 ºC. The oven was maintained at 130 ºC for 3 minutes after injecting 2 µL of the sample, ramped up at 3 ºC/minute to 260 ºC and then held for 5 minutes. The area of the syringyl:guaiacyl peaks were integrated using the internal standard peak area and the S:G ratio was calculated using Microsoft Excel. The results were analyzed for statistical significance using *t*-test at α level = 0.05.

2.13 Phloem tissues chemical analysis

Klason lignin and monolignol analyses were carried out as described in sections 2.12.1, and 2.12.4, respectively.

2.13.1 HPLC analysis

Phloem tissue was separated and frozen in liquid nitrogen. The tissue was ground to a fine powder and metabolites were extracted by 80% acidified methanol using the previously described extraction method (section 2.10.1). The extracts were dried using an Eppendrof Vacufuge 5301. The dry pellets were dissolved in 0.5 mL of absolute methanol, and the metabolites analyzed by reversed-phase HPLC. LC-mass spectra and NMR analyses of the significantly different metabolites was conducted at the University of Wisconsin, Madison, USA (laboratory of Dr. J. Ralph).

Chapter 3: Results

3.1 Isolation of poplar glycosyltransferases genes

Initially, the *P. trichocarpa* genome sequence (JGI v.1.1) was queried using two *Arabidopsis* genes, *UGT74F1* (At2g43840) and *UGT74F2* (At2g43820), and the tobacco SAGT (cDNA sequences AF190634.1*)* to identify putative poplar homologs of genes known to display SAGT activity. The basic local alignment search tool (BLAST) did not identify poplar gene sequences with high similarity to the queried cDNAs. The translated amino acid sequences of these genes were then used to BLAST the GenBank database, and one poplar UGT gene with approximately 53% identity was identified. This gene was previously annotated as *P. deltoides* x *trichocarpa* UDP-GT, similar to At2g43820. Its homolog in *P. trichocarpa* genome was the POPTR_0007s00870, which is located on linkage group VII of the poplar genome sequence (JGI v.1.1). Two primers were designed based on the sequence of the *P. trichocarpa* POPTR_0007s00870 gene and used to isolate the *PopGT2* from the hybrid poplar cDNA.

In order to isolate *PopGT1*, nucleotide sequences of SAGTs from *Arabidopsis*, tobacco, and the isolated poplar *PopGT2* sequence were used to design degenerate primers. Subsequently, RACE-PCR was used with these primers to amplify expressed putative homologs of the *AtUGT74F1* and *AtUGT74F2* from the hybrid poplar cDNA. The degenerate primers amplified a 550 bp cDNA fragment and BLAST search located the fragment on scaffold 182 of the *P. trichocarpa* genome-sequence (JGI v.1.1) (POPTR_0001s39980). The full length of the *PopGT1* was then amplified from the hybrid poplar using the two primers, which were designed based on the gene coding sequence of POPTR_0001s39980. I subsequently named these genes as *PopGT1* and *PopGT2,* based on the first 3 letters of the word *Populus* (Pop), the abbreviation for

glycosyltransferase (GT), and an Arabic number representing the closest *Arabidopsis* homologs (*AtUGT74F1* and *AtUGT74F2*). The putative homologs of the *Arabidopsis* SAGT genes were then targeted by RNAi-suppression constructs to down-regulate the expression of these homologs in the hybrid poplar. Sequence identity was calculated from the translated sequence using Clustal Omega and a sequence identity matrix was generated (Table 3).

Table 3: Protein sequence identity matrix calculated from the poplar (PopGT1and PopGT2), tobacco, and *Arabidopsis* (AtUGT74F1, AtUGT74F2, and AtUGT74E2) SAGT translated amino acids. Values represent the identity percentage between each pair of the aforementioned glycosyltransferases.

Sequence	1	2	3	4	5	6
1: AtUGT74F1	100.00					
2: AtUGT74F2	76.84	100.00				
3: PopGT2	50.13	52.24	100.00			
4: Tobacco SAGT	48.99	49.89	51.43	100.00		
5: PopGT1	45.95	47.75	50.11	54.07	100.00	
6: AtUGT74E2	46.15	46.15	47.88	49.44	54.57	100.00

UGT74F1 (At2g43840); UGT74F2 (At2g43820); UGT74E2 (At1G05680); Tobacco SAGT (AF190634); PopGT1 (POPTR_0001s39980); PopGT2 (POPTR_0007s00870).

In silico analysis of the PopGT1 and PopGT2 translated amino acid sequences was carried out using the ExPASy-ProtParam online protein analysis server. The calculated physicochemical parameters indicated that PopGT1 is 471 amino acids (aa) in length, while PopGT2 is 457 aa in length. The predicted protein size of PopGT1 and PopGT2 was 53.00 and 51.20 kDa, respectively, while the calculated isoelectric points were 4.98 for PopGT1 and 5.00 for PopGT2 (Table 4). The calculated characteristics of both PopGT1 and PopGT2 are comparable to the general characteristics of family 1 UGTs.

Table 4: Physico-chemical protein characteristics predicted from the translated amino acid sequences of PopGT1, PopGT2, using the ExPASy-ProtParam protein analysis server. The values for the *Arabidopsis* (UGT74F1, UGT74F2, and UGT74E2) protein properties were obtained from the TAIR database.

Protein name	No. of amino acids	Calculated M.W. kDa	Calculated I.P.
PopGT1	471	53.0	4.98
PopGT2	457	51.2	5.00
AtUGT74F1	449	50.3	5.46
AtUGT74F2	449	50.77	5.00
AtUGT74E2	453	51.05	5.67

UGT74F1 (At2g43840); UGT74F2 (At2g43820); UGT74E2 (At1g05680); Tobacco SAGT (AF190634); PopGT1 (POPTR_0001s39980); PopGT2 (POPTR_0007s00870). M.W.: Molecular weight in kilo Dalton (kDa); I.P.: Iso-electric point.

Family 1 UGT enzymes are hypothesized to maintain the cellular homeostatic level of toxic phenolic metabolites and their functional activities are believed to occur in the cytosol (Bowles et al., 2005). In order to investigate the subcellular protein localization pattern of the poplar enzymes (PopGT1 and PopGT2), the translated amino acid sequences were interrogated for the presence of signal peptides using the TargetP v.1.1 server (http://www.cbs.dtu.dk/services/TargetP). The results indicated that neither protein contains a signal peptide, nor a localization pattern. In addition, no pre-sequence cleavage sites were detected in either protein sequence (Table 5). These results suggest that these proteins are most likely cytosolic enzymes.

Table 5: PopGT1 and PopGT2 signal peptides and localization analyses using the TargetP v.1.1 online server.

Protein name	cTP	mTP	SP	Other	Loc	RC	TPlen
PopGT1	0.125	0.076	0.198	0.456	_	4	-
PopGT2	0.072	0.233	0.042	0.793	_	3	-

cTP: chloroplast transit peptide; mTP: mitochondrial target peptide; SP: secretory pathway signal peptide; RC: reliability class, Loc: localization, and Tplen: predicted pre-sequence length (for cleavage site prediction).

UGTs use activated sugar as a donor molecule to glycosylate a phenolic compound. The carboxylic termini of the UGT protein sequences contain the conserved amino acid sequence PSPG-box that consists of 44 amino acids (Hughes and Hughes, 1995).

Protein domain analysis for both PopGT1 and PopGT2 translated amino acid sequences was carried out using the Prosit protein domain analysis online server (http://prosite.expasy.org/). The UDP-glycosyltransferase signature was detected in both protein sequences. Sequence alignment of the two poplar protein sequences with the *Arabidopsis* and tobacco homologs identified the presence of this conserved domain at the carboxylic terminus of the analyzed sequences (Figure 2). This result suggests that the PopGT1 and PopGT2 proteins are putative members of family 1 UGTs.

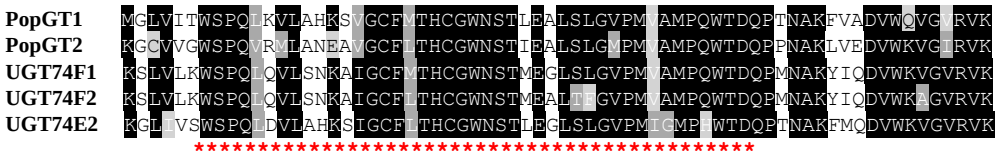


Figure 2: Protein domain analysis of the poplar (PopGT1 and PopGT2) and *Arabidopsis* (AtUGT74F1, AtUGT74F2, and AtUGT74E2) UGTs, showing the presence of the conserved UDP-glucose binding domain (ps00027). Asterisks represent the conserved PSPG 44-amino-acid sequence, black asterisk depict the Glutamine (Q44), which is specific for UDP-glucose as a donor. 100% identical amino acids are shaded in black and gray-coloured shaded amino acids are similar. Shaded graph was generated using shad box online server.

To determine the predicted 3D-protein fold-structure of the translated amino acid sequences of PopGT1 and PopGT2, the Phyre2 server was used. The protein tertiary-structure graphics for both PopGT1 and PopGT2 were generated using the PyMOL software (DeLano, 2002) (available at: www.pymol.org). The analysis showed typical GT-B 3D protein fold-structures for both protein sequences (Figures 3a and 3b), which is a characteristic fold in family 1 UGTs

(Lairson et al., 2008). The two proteins were aligned to each other using the PyMOL software, and shown to fold similarly, with alpha helices and beta sheets in near identical alignment (Figure 3c). This result again suggests that both sequences belong to family 1 UGTs.

a) b) c)

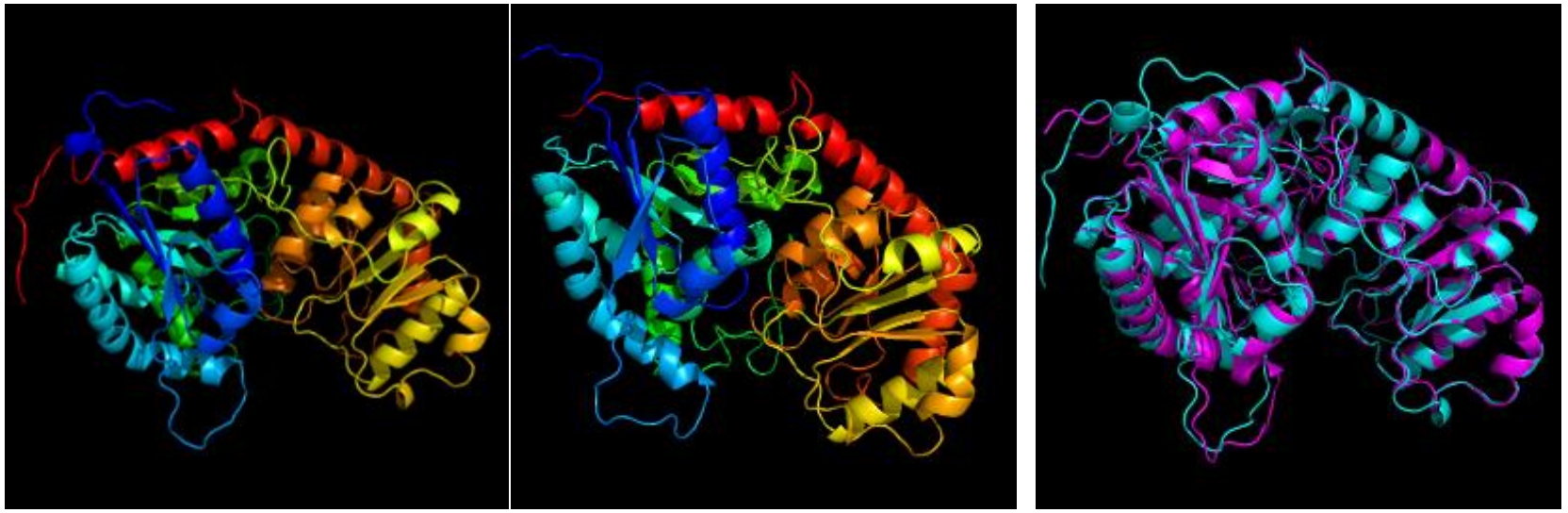


Figure 3: Protein 3D fold-structures of (a) PopGT1 and (b) PopGT2, depicting a GT-B folding pattern. 3D alignment of the poplar PopGT1 (blue) to the PopGT2 (pink) proteins is shown in panel (c). The folding-structures were generated using the Phyre2 protein modelling software. The 3D protein alignment and graphics were created using PyMOL software (www.pymol.org).

Phylogenetic analysis

Phylogenetic analysis of protein sequences has been used to classify UGTs into different phylogenetic groups (Ross et al., 2001). In some cases, the results agreed with the substrate-recognition classification of UGTs (Li et al., 2001; Ross et al., 2001). However, some members of the family 1 UGT proteins do not show the same pattern, raising questions about the efficiency of using sequence similarity and phylogenetic approaches in the classification of UGTs (Bowles et al., 2005; Osmani et al., 2009). To investigate the relatedness of the isolated poplar enzymes (PopGT1 and PopGT2), and their phylogenetic relationship to *Arabidopsis* UGTs, I aligned the translated amino acid sequences of both poplar proteins to the 40 *Arabidopsis* UGTs, representing 40 GT families and subfamilies of the family 1 UGTs in the CAZy online database (http://www.cazy.org). The phylogenetic tree (Figure 4) indicated that the two poplar proteins (red labels) were clustered with UGT families belonging to the *Arabidopsis* phylogenetic group L (clades UGT74, UGT75, and UGT84), which is labelled with a brown line (Figure 4).

To gain more insight into the relatedness of PopGT1 and PopGT2 to the other members of the *Arabidopsis* UGT phylogenetic group L, all UGT sequences belonging to this group were downloaded from the CAZy database and aligned with the two poplar translated protein sequences (PopGT1 and PopGT2), as well as the tobacco SAGT, using the Clustal Omega online server. A second Neighbour-Joining (NJ) phylogenetic tree (Figure 5) was generated to relate the two poplar enzymes to their closest homologs. The phylogenetic analysis indicated that PopGT1 and PopGT2 amino acid sequences clustered together in clade UGT74 of the phylogenetic group L with a 100% confidence level. Within this cluster, PopGT2 was coherently clustered with *Arabidopsis* UGT74F1 and UGT74F2 in one sub-group (88% confidence level). In contrast, PopGT1 clustered with the

tobacco SAGT and UGT74E2 in a second sub-group with relatively low bootstrap values (32), making the separation between the two sub-groups uncertain. In this case, I assumed that PopGT1 and PopGT2 belong to the same phylogenetic clade L within family UGT74 (yellow box in Figure 5). This clade is known to contain UGTs (AtUGT74F1, AtUGT74F2, and the *N. tabacum* SAGT) that glycosylate different phenolic substrates (benzoic acid, cinnamic acid, salicylic acid, anthralinic acid and flavonols), as well as auxin (AtUGT74D1, AtUGT74E1, and AtUGT74E2) and glucosinolate (AtUGT74C1 and AtUGT74B1). These results, in combination with the previously published results, suggested that PopGT1 and PopGT2 may be active towards SA as well as other phenolic substrates. In addition, the results suggest that these enzymes might be important in maintaining normal plant growth and development, as well as cell wall lignification. To explore this further, I examined RNAi hybrid poplar lines and *Arabidopsis* transgenic plants expressing *PopGT1* and *PopGT2* for growth phenotypes, cell wall structure, wood quality, and chemical composition. In addition, I attempted to characterize recombinant proteins encoded by the target poplar genes *in vitro*.

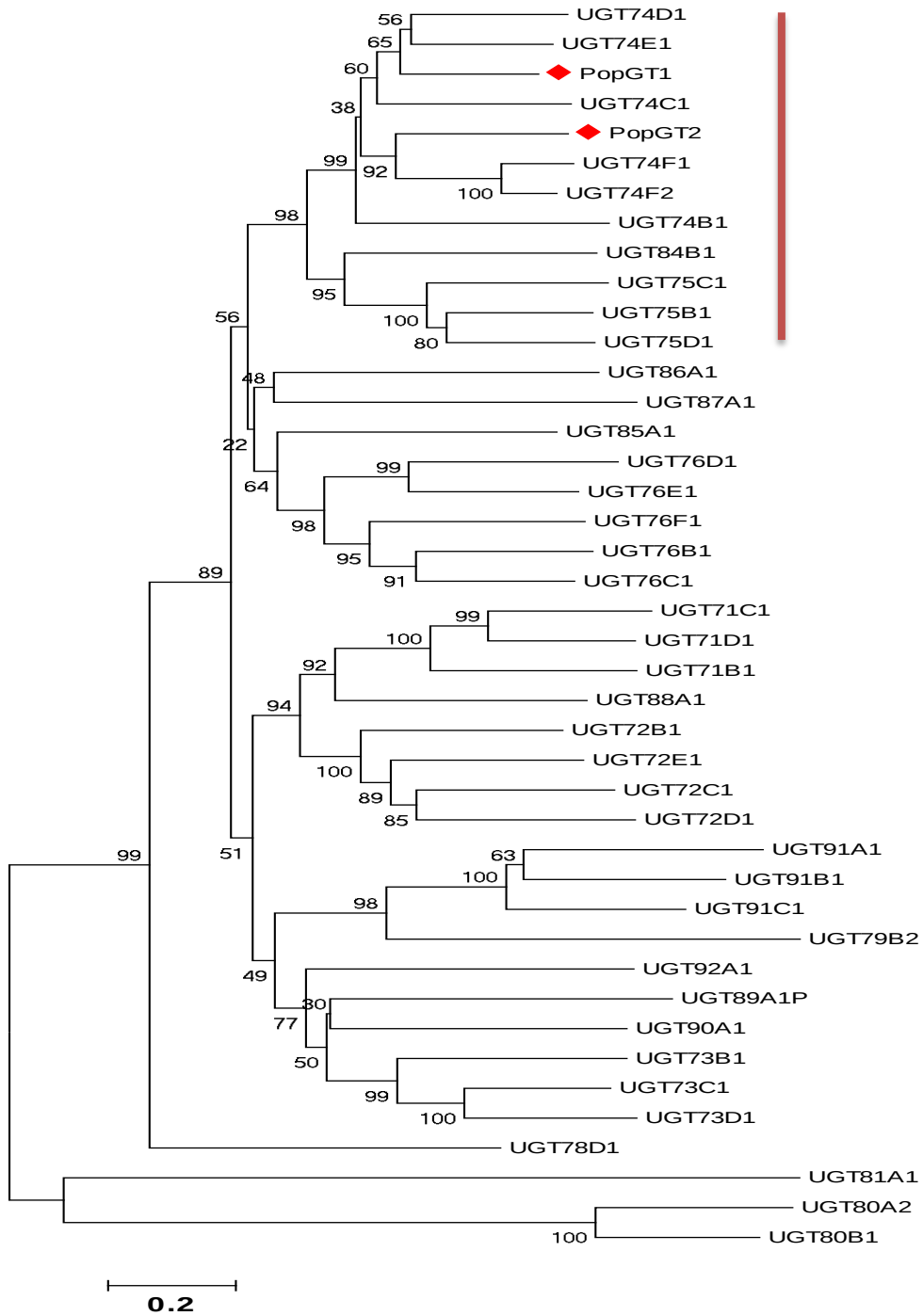


Figure 4: Neighbour-Joining (NJ) phylogenetic tree generated for 42 UGT amino acid sequences, including two poplar protein sequences, PopGT1 and PopGT2 in red diamond, and the 40 *Arabidopsis* UGT protein sequences, representing the members of family 1 UGTs. Sequences were downloaded from the CAZy database (www.cazy.org), and were aligned using the Clustal Omega server. The tree was generated using MEGA 5.2 software. The percentages of bootstrap (1000 replicates) are shown next to the branches and members of phylogenetic group L are labeled with the brown line.

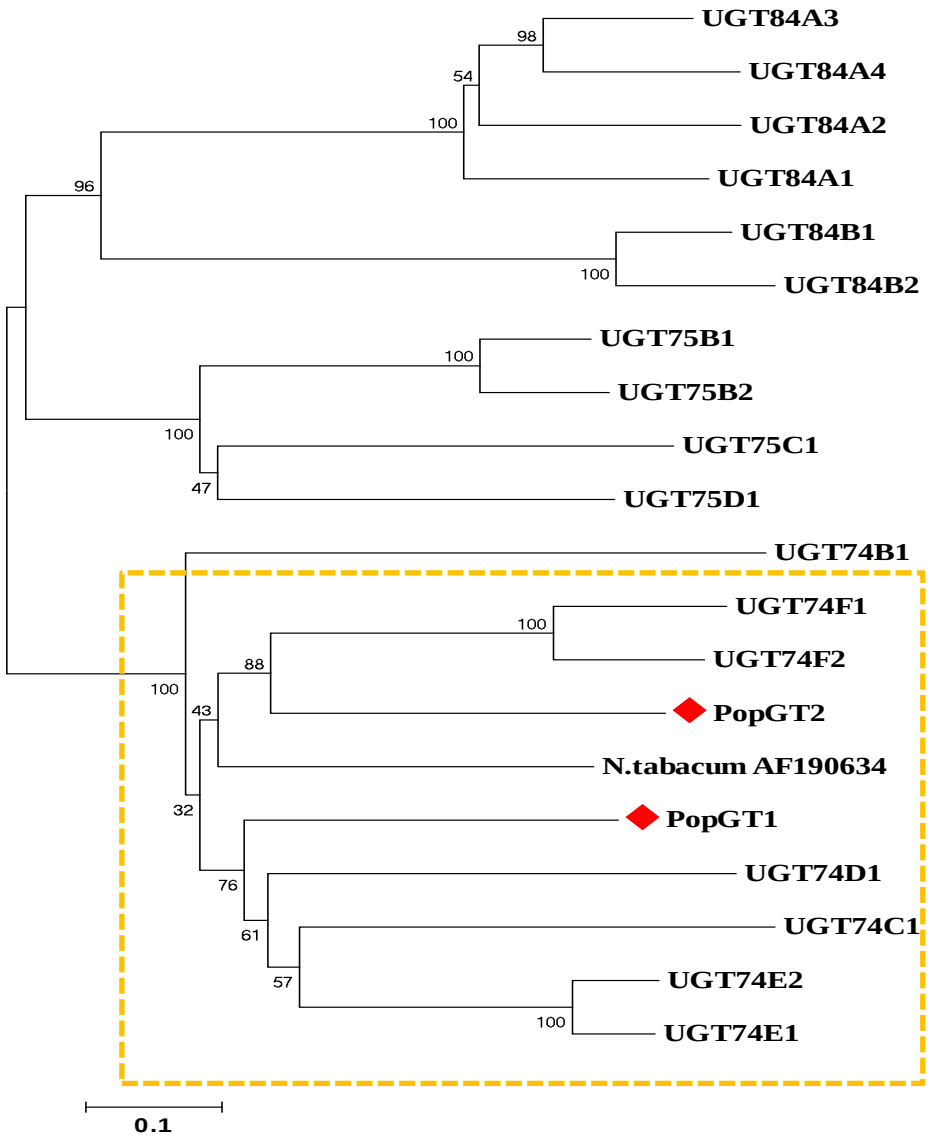


Figure 5: Neighbour-Joining (NJ) phylogenetic tree generated from amino acid sequences representing 20 *A. thaliana* group L UGTs, two poplar protein sequences (PopGT1 and PopGT2, labeled in red diamond), and the tobacco SAGT. Clade 74 is identified by a yellow box. The bootstrap values (1000 replicates) are shown next to the branches as a percentage. The protein sequences were downloaded from the CAZy database (www.cazy.org), and the SAGT sequence was obtained from the GenBank database. Protein sequence alignment was carried out using the Clustal Omega server and phylogenetic analysis was conducted using the MEGA5 software.

3.2 Protein expression and function characterization

3.2.1 Protein expression in *Pichia pastoris* system

In an attempt to reveal the substrate specificity of the isolated poplar enzymes (PopGT1 and PopGT2), I attempted to produce recombinant proteins to examine their activity against different phenolic substrates in two protein expression systems (*Pichia pastoris* and *E. coli*). The *Pichia* protein expression system was initially chosen, given its ability to express eukaryotic proteins with the post-translational modifications required for activity (Canam et al., 2008; Unda et al., 2012). I cloned the Open Reading Frames (ORFs) of the two poplar genes (*PopGT1* and *PopGT2*) into the multiple cloning site (MCS) of the *pPICZB* vector under the control of the *AOX* methanol-inducible promoter. To facilitate affinity chromatography purification of the expressed protein (using Ni^{+2}), the constructs were designed to have a 6X-histidine tag fused either to the C- or the N-terminus of the recombinant proteins. The constructs were then transformed in *P. pastoris* strain Gs115. Protein expression was induced using methanol and protein production was assessed in the soluble and insoluble fractions after 4 days growth at 28 ºC. SDS-PAGE analysis showed that no differentially expressed proteins were visually detected in the protein profiles of the induced *AOX*::*PopGT1* or *AOX*::*PopGT2* yeast strains when compared to either non-induced or the empty-vector strains (Figure 6).

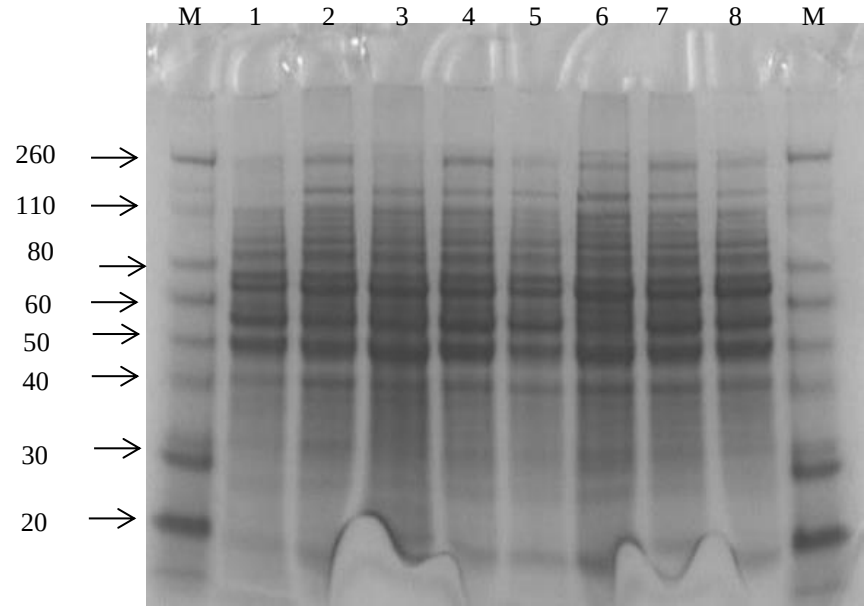


Figure 6: SDS-PAGE gel electrophoresis of total protein extracts of *Pichia* strains: lane M, Protein molecular weight marker; lane 1, non-induced G115 strain; lane 2, empty vector strain; lanes 3-5, AOX::*PopGT1* induced strains; and lanes 6-8, AOX::*PopGT2* induced strains. White arrow indicates the predicted position of the recombinant protein approximate band size between 52-53 kDa.

As recombinant proteins were not visually identified in the total protein extracts, I used Ni^{+2}-affinity chromatography in attempt to purify trace amounts of the recombinant proteins that potentially could not be detected on the SDS-PAGE gels. However, no recombinant proteins were obtained post purification. In order to confirm these results, western blot analysis was carried out on all induced and non-induced protein fractions, employing an anti-His antibody, but the blot did not show any specific binding between the antibody and the *Pichia* protein samples.

3.2.2 Protein expression in bacterial host system

Since my trials to express the two poplar genes in *Pichia* were unsuccessful, recombinant protein expression was then performed using the bacterial expression systems *pET30*(a), *pGEX4T2*, and

pET28(a) under several induction conditions (Table 6). Under all conditions tested, the *pGEX4T2* BL21 *E. coli* expression system did not produce any detectable soluble PopGT1 or PopGT2 proteins. However, both proteins were found in the centrifugation pellets that include all the insoluble cellular components.

The two poplar genes were subsequently expressed in both the *pET30* +(a) *E. coli* BL21 (DE3) and *pET28* +(a) *E. coli* Rosetta (DE3) systems, expression systems were created to enhance the expression of eukaryotic proteins in functional fold-structures and soluble forms in bacterial cells. SDS-PAGE analysis of the soluble and insoluble protein fractions indicated that no soluble PopGT proteins were retrieved when the induction temperature was between 16-37 ºC and IPTG concentration between 0.5-5.0 mM. Instead, the recombinant proteins were located in insoluble pellets. However, by reducing the IPTG concentration to 0.1 mM and conducting the induction at 16 ºC it was possible to obtain some production of soluble recombinant PopGT1 and PopGT2 (Table 6). SDS-PAGE analysis showed that both the PopGT1 and PopGT2 proteins were in the expected size range (52 and 55 kDa; Figure 7 and 8).

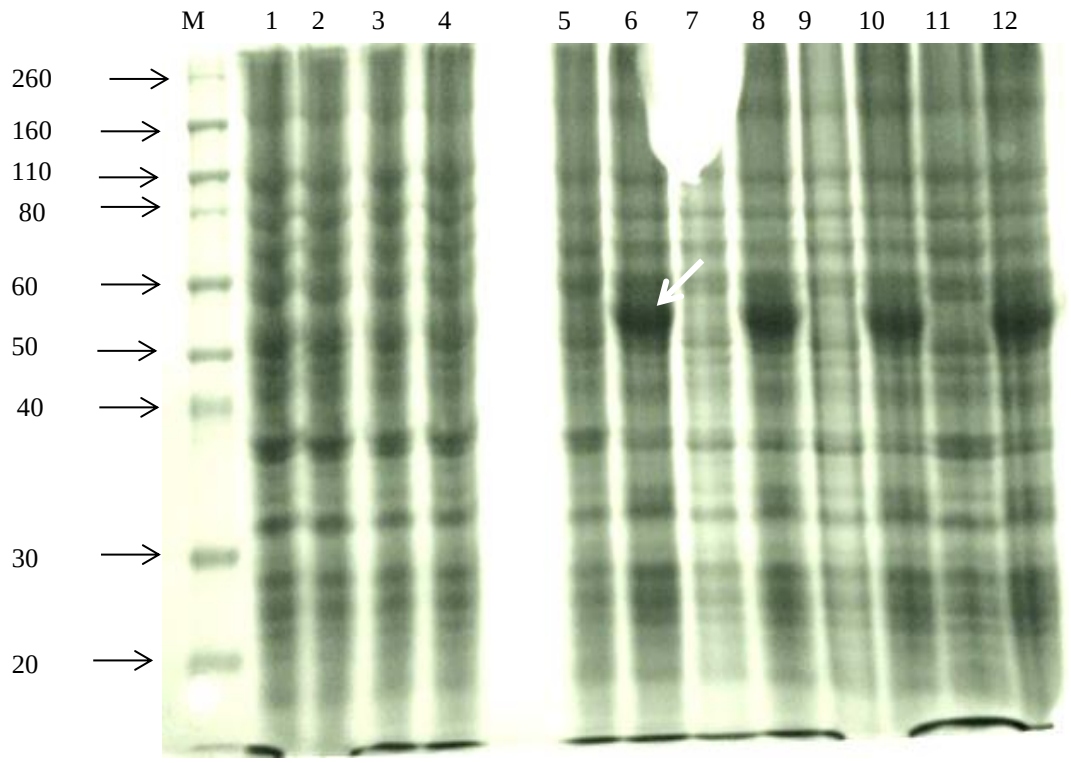


Figure 7: Recombinant protein expression in *pET* expression system. Lane M: molecular weight marker; lanes 1-4: soluble fraction of induced proteins; lanes 5-12: total cellular protein; lanes 5, 7, 9, 11: non-induced strains; lanes 6 and 8: PopGT1 induced; lanes 10 and 12: PopGT2 induced strains. The white arrow labels the induced protein in the insoluble fraction. The expected band size is between 52-55 kDa.

Table 6: **IPTG concentrations and induction temperatures that were used to optimization PopGT1 and PopGT2 protein expression in the *pET* protein expression systems.**

Temp. / **IPTG conc.**	**37 ºC**	**28 ºC**	**16 ºC**
5mM	Insoluble*	Insoluble	Insoluble
2mM	Insoluble*	Insoluble	Insoluble
1mM	Insoluble*	Insoluble	Insoluble
0.5 mM	Insoluble	Insoluble	Insoluble
0.1mM	Insoluble	Insoluble	**Soluble**

* Company's recommended induction condition.

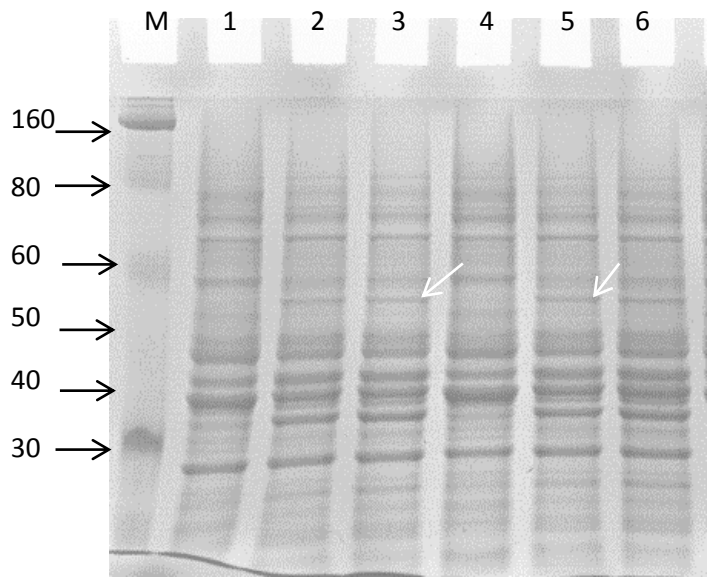


Figure 8: Recombinant protein expression in the *pET* expression system. Lane M: protein molecular weight marker; lane 1: non-induced PopGT1; lanes 2 and 3: induced soluble fraction of PopGT1; lane 4: non-induced PopGT2 strain; lanes 5 and 6: induced PopGT2. Black arrows indicate the molecular weight size and white arrows indicate the soluble protein band of both proteins. Expected molecular weight between 52 and 55 kDa.

Under all induction conditions, the majority of the recombinant proteins were detected in the pellet fraction of the bacterial cells.

3.3 Purification of the PopGT1 and PopGT2 recombinant enzymes

In order to purify the recombinant PopGT1 and PopGT2 for further biochemical characterization, Ni^{+2} affinity chromatography was employed. The clear lysates that contain the soluble fractions of the recombinant proteins were passed through Ni-NTA agarose beads either by gravity or using automated FPLC system (Figures 9 and 10). The two purification attempts could not separate the proteins of interest from native bacterial proteins, even though different gradients of imidazole were

employed in the washing and elution steps. The low production yields and the apparent instability of the recombinant proteins appeared to hinder any further purification process. Therefore, I tested both the partially purified enzymes and the total bacterial soluble lysates for enzymatic activity.

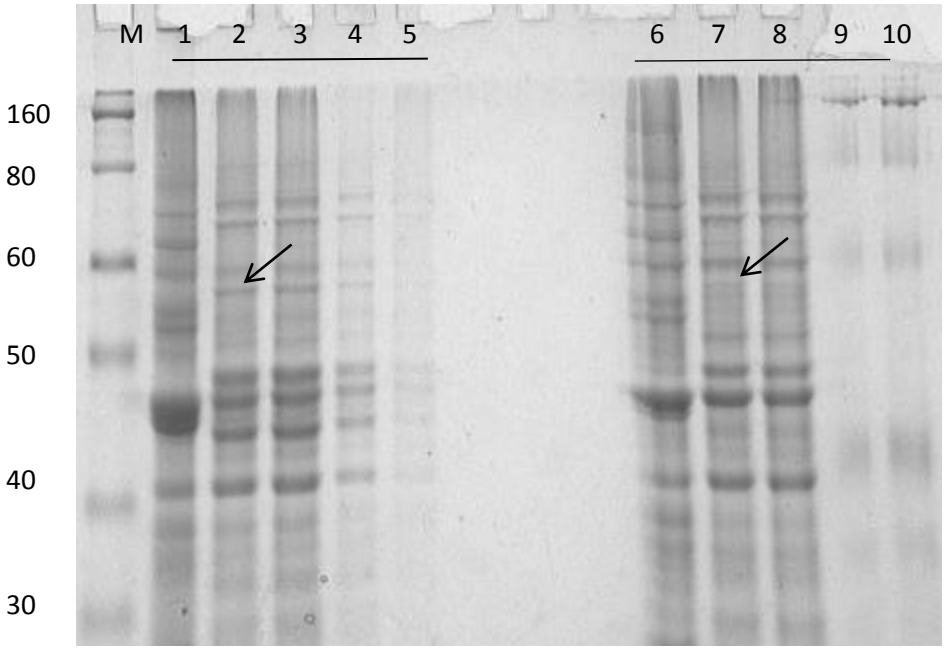


Figure 9: Protein purification of PopGT1 (1-5) and PopGT2 (6-10) using a Ni-NTA agarose beads column. Lane M: protein molecular weight marker; lanes 1 and 6: non-induced strain; lanes 2 and 7: bacterial clear lysates; lanes 3 and 8: column flow through; lanes 4, 5, 9, and 10: elution fractions with 40 and 250 mM imidazole.

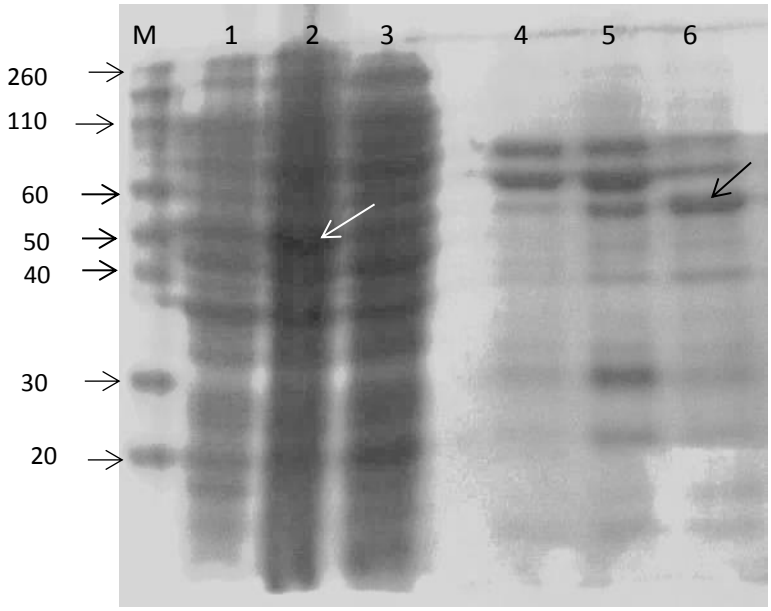


Figure 10: **FPLC purification of concentrated PopGT1 protein collected from 2-litre bacterial culture (lane 2). Lane 1: non-induced culture; lane 2 induced culture; lane 3: washing step, using 25 mM Imidazole; lane 4, 5, and 6: elution fractions. White and black arrows identify the protein bands in the washing and elution fractions, respectively. Band size (52-55 kDa) was compared to pre-stained molecular weight marker (M).**

3.4 Characterization of PopGT1 and PopGT2 enzyme functions

3.4.1 *In vitro* substrate specificity of the enzyme

As shown by *in silico* and phylogenetic analyses, both of the amino acid sequences of PopGT1 and PopGT2 were clustered with members of family 1 UGTs that belong to phylogenetic group L and clade UGT74 of the *Arabidopsis* UGT enzymes. Enzymes belonging to this clade have previously been shown to possess glycosylation activity against different classes of phenolic substrates and plant hormones. Therefore, I screened a wide array of phenolic substrates that are members of the major classes of the phenylpropanoids, including phenolic acids, monolignols, flavonoids, anthocyanin,

coumarins, and salicylates. In addition, activity against plant hormones and phenolic amino acids was tested.

The partially purified PopGT1 and PopGT2 did not show any activity towards the aforementioned substrates under the experimental conditions employed. When the total bacterial clear lysate containing PopGT1 was used in the biochemical assays, glycosylation activity towards several phenolic substrates, including quercetin, kaempferol, taxifolin, naringenin, phloretin, biochanin A, and hesperetin was detected compared to the boiled enzyme and empty vector controls (Figure 11 and 12, and Table 7). In addition, PopGT1 showed trace activity towards IBA and cinnamic acid. Unexpectedly, no PopGT1 activity was detected towards benzoic acid or salicylic acid or against the phenolic glycoside salireposide aglycone (2,5-dihydroxybenzyl benzoate). Instead, the 2,5-dihydroxybenzyl benzoate substrate was hydrolyzed during incubation.

Similarly, the biochemical activity of PopGT2 was investigated by incubating the total bacterial clear lysate containing the recombinant enzyme with the same substrates tested with PopGT1. However, PopGT2 did not appear to catalyze any biochemical reaction with the substrates tested.

PopGT1 and PopGT2 activities were then tested for sugar donor specificity. The enzymatic assays were repeated, using UDP-galactose instead of UDP-glucose and incubated for 2 hours and over-night. HPLC analysis showed that no glycosylation products were detected in any of the assays using all the aforementioned tested substrates. These results indicate that the PopGT1 enzyme is a UDP-glucose-specific utilizing enzyme that is active towards a wide range of phenolic substrates. In contrast, the substrate specificity of PopGT2 remains unknown.

Table 7: Phenolic substrates that were employed to evaluate *in vitro* enzymatic activities of PopGT1 and PopGT2 enzymes. The activity of PopGT1 is presented as detected or not-detected (nd) based on the appearance of a glycoside product. The percentage activity of PopGT1was calculated relative to the activity towards quercetin. No glycosylation activity was detected for PopGT2.

Substrate	Glycoside	% Relative activity	Substrate	Glycoside
Quercetin	Detected	100	Salicylic acid	nd
Kaempferol	Detected	84	Saligenin	nd
Naringenin	Detected	75	Salicylaldehyde	nd
Hesperetin	Detected	60	Salicin	nd
Phloretin	Detected	19	2,5 dihydroxybenzyl alcohol	nd
Taxifolin	Detected	19	2,5-dihydroxybenzoic acid	nd
Biochanin A	Detected	16	4-hydroxybenzoic acid	nd
IBA	Detected	06	3-hydroxybenzoic acid	nd
Cinnamic acid	Detected	04	Catechol	nd
Myricetin	Nd	-	Coumarin	nd
NAA	Nd	-	6-hydroxy 7-methoxy coumarin	nd
IAA	Nd	-	7-hydroxy 6-methoxy coumarin	nd
Benzyl adenine	Nd	-	Benzoic acid	nd
Phenyl alanine	Nd	-	Ferulic acid	nd
Tyrosine	Nd	-	Coumaric acid	nd
2,5-dihydroxybenzyl benzoate	nd/hydrolyzed	-	Sinapic acid	nd
Catechin	Nd	-	Cinnamyl alcohol	nd
Cyanidin	Nd	-	Coumaryl alcohol	nd
Pelargonidin	Nd	-	Sinapyl alcohol	nd

Indole butyric acid (IBA), Indole acetic acid (IAA), Naphthaline acitic acid (NAA), Salicylic acid (SA), % relative activity = (peak area of the product/peak area of quercetin glycoside peak area) x 100.

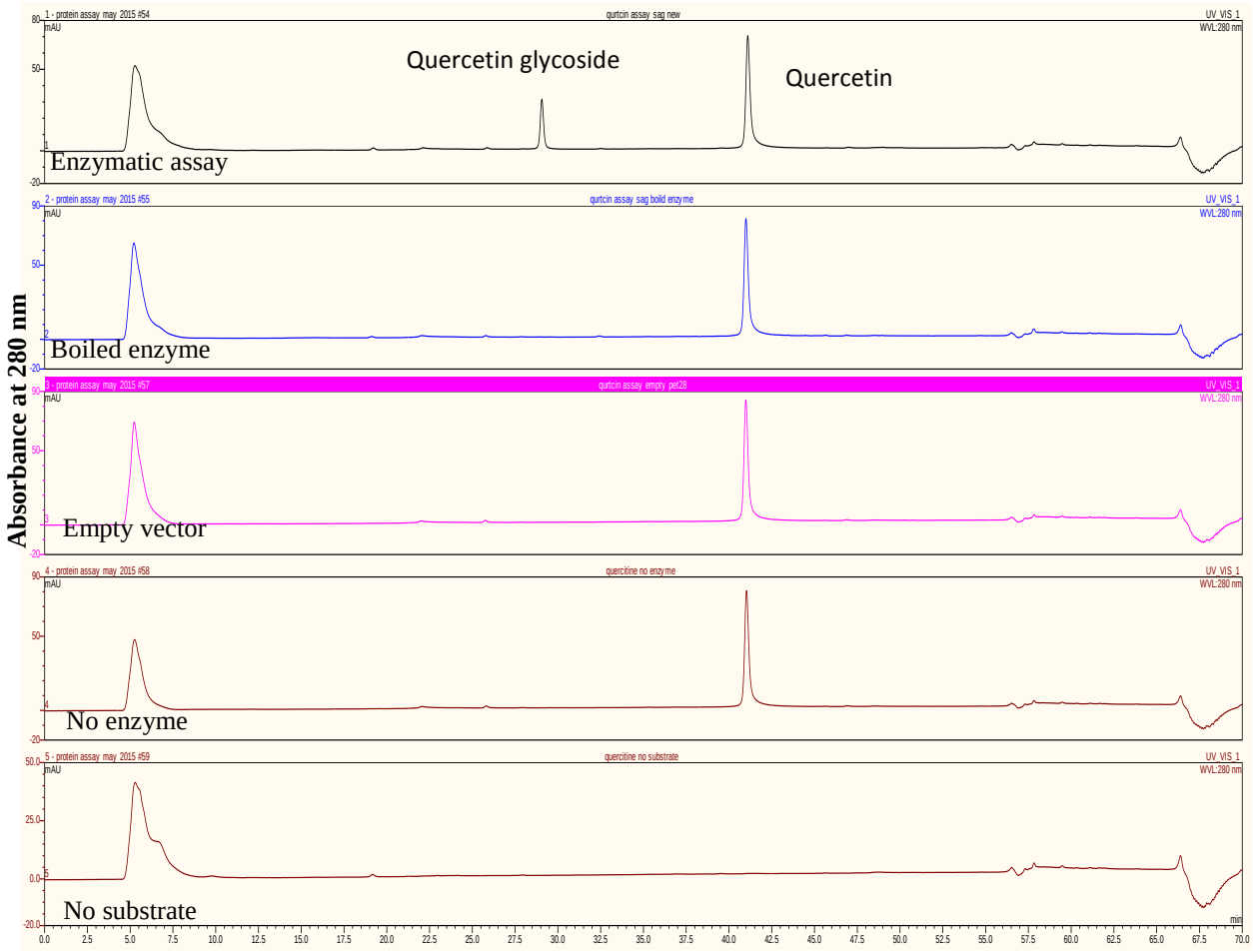


Figure 11: **HPLC chromatograms of PopGT1 enzymatic assay towards quercetin at room temperature compared to different experimental control reactions, including: boiled bacterial clear-laysate, empty vector bacterial clear-laysate, no clear-laysate added, and no substrate added. Only one product was observed when quercetin was incubated with the bacterial lysate expressing PopGT1 enzyme.**

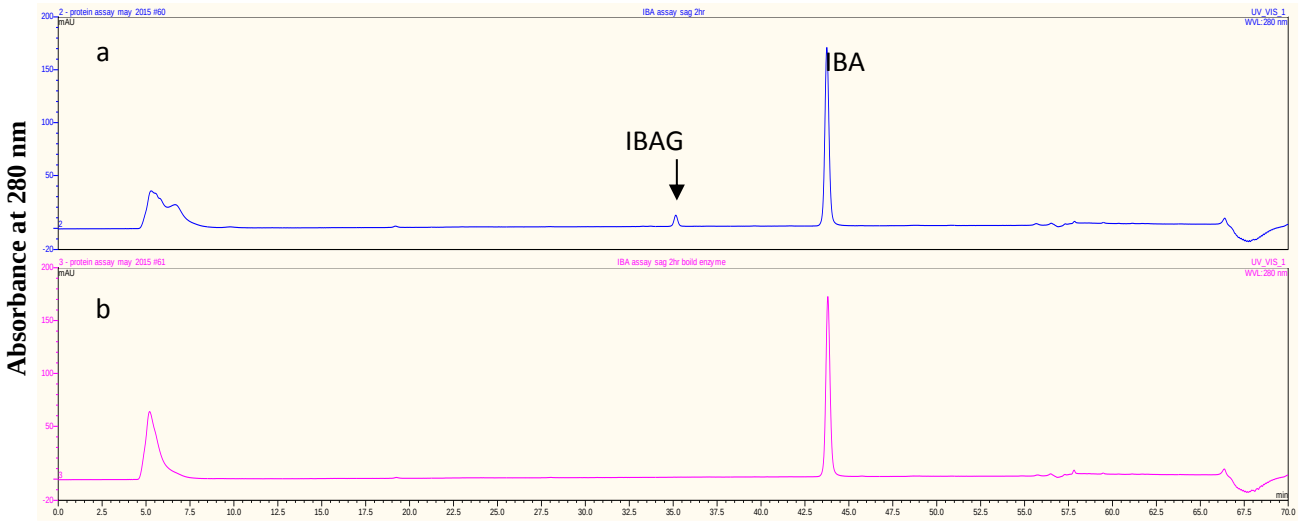


Figure 12: PopGT1 enzyme showed trace activity towards indole butyric acid (IBA) and produced IBA-glycoside (black arrow). HPLC trace of the enzymatic assay (a) compared to the boiled enzyme control assay (b).

3.4.2 Temperature and pH profiles of the PopGT1 enzyme

Temperature optimization was carried out by testing the activity of PopGT1 towards quercetin at different temperatures (20-70 ºC). The glycosylation product was analyzed using HPLC and the peak area of the quercetin glycoside was compared at each temperature tested. The results of the biochemical assays suggested that the optimum temperature was 35 ºC (Figure 13).

PopGT1 was also tested over a range of pHs (4.0-10) towards quercetin to identify the optimal pH for the enzyme. The PopGT1 pH profile showed a bell-shaped profile with the optimum activity at pH 7.5 (Figure 14).

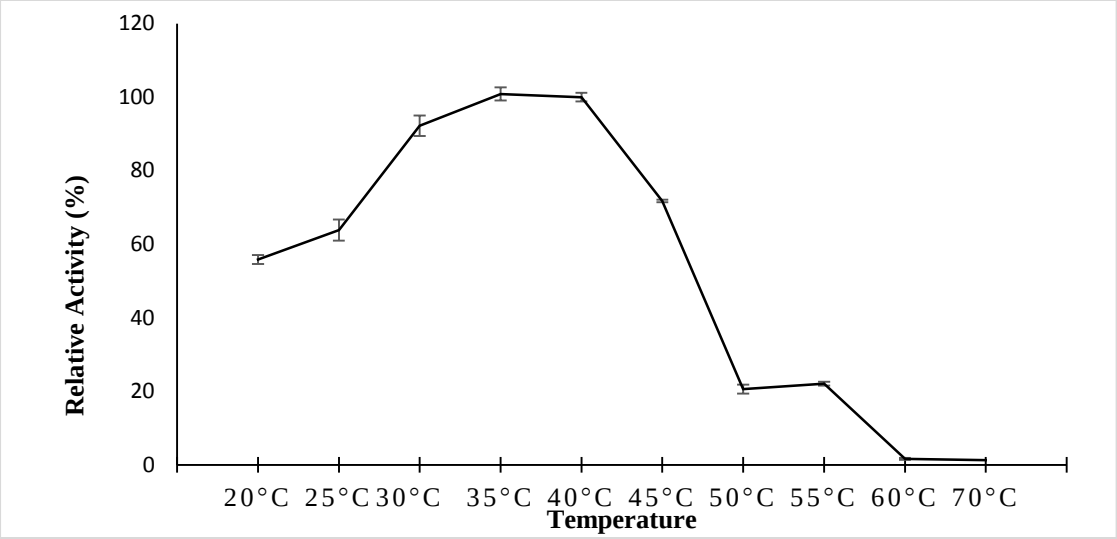


Figure 13: Relative enzymatic activity (%) of PopGT1 towards quercetin at different temperatures (20-70 ºC) using reversed-phase HPLC to detect the formation of a glycoside. Relative activity (%) was calculated compared to the average peak area at 35 °C. Error bars present the standard error for 3 replicates.

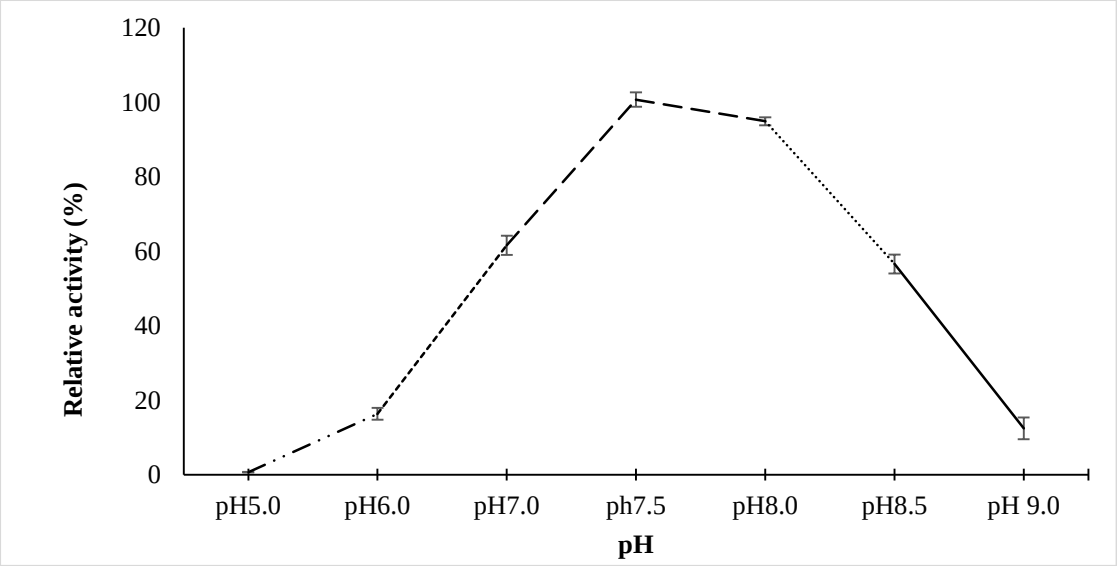


Figure 14: Relative enzymatic activity (%) of PopGT1 towards quercetin at different pHs (4-9) using reversed-phase HPLC to detect the formation of a glycoside. Relative activity (%) was calculated compared to the average peak area at pH 7.5. Assays were performed at 35 °C. Error bars present the standard error for 3 replicates. Different line patterns depict different buffers employed for the assays.

3.5 Over-expression of the two poplar *PopGT1* and *PopGT2* genes in *A. thaliana*

3.5.1 Generation of transgenic lines

In order to examine the biological function of the two UGT (*PopGT1* and *PopGT2*) genes, both were independently cloned into the multiple cloning site of the *pSM3* vector under the control of constitutive double 35S promoter, generating two constructs (*pSM-PopGT1* and *pSM-PopGT2*). These constructs were used to generate transgenic *Arabidopsis* plants, using *Agrobacterium*-mediated transformation. Fifteen hygromycin-positive plants were then screened from each construct, using qRT-PCR. Six over-expressing lines, showing different levels of transcript abundance, were selected from each construct. No gene expression was detected in the wild-type plants, as expected. Gene expression analysis indicated that lines OE-*PopGT1*-1, OE-*PopGT1*-3, OE-*PopGT1*-4 and OE-*PopGT1*-6 were the highest among the OE-*PopGT1* transgenic lines. In contrast, lines OE-*PopGT2*-1, OE-*PopGT2*-4, OE-*PopGT2*-5, and OE-*PopGT2*-6 showed the highest gene expression for OE-*PopGT2* plants (Figure 15 a and b).

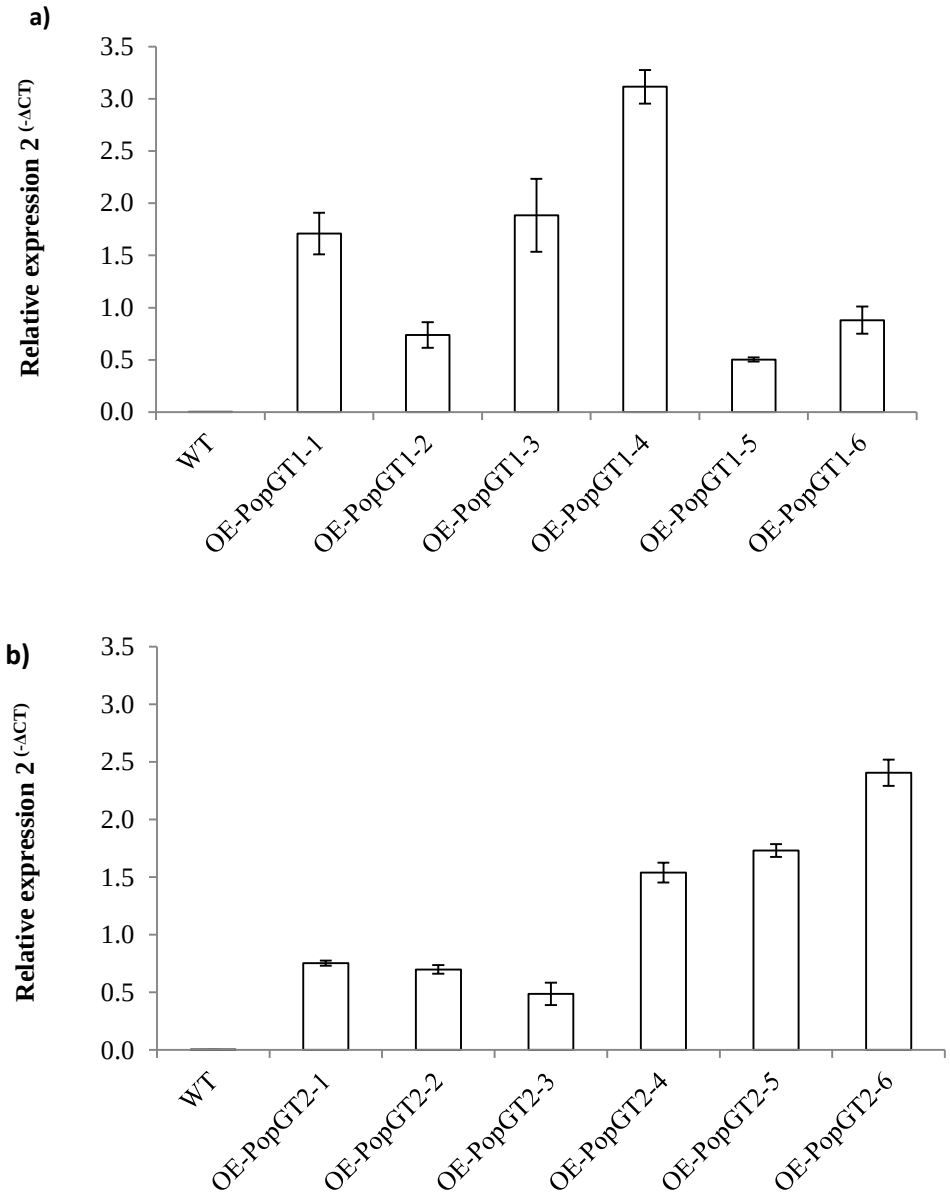


Figure 15: Transcript abundance of (a) *PopGT1* and (b) *PopGT2* in transgenic *Arabidopsis* lines (OE-*PopGT1* and OE-*PopGT2)* compared to wild type (WT) using tubulin as a reference gene. Error bars represent standard deviation (n = 3).

3.5.2 Phenotypic characterization of OE-*PopGT1* and OE-*PopGT2* transgenic lines

3.5.2.1 Plant growth phenotype

In order to study the effect of over-expressing the poplar UGTs (*PopGT1* and *PopGT2*) on plant growth, all transgenic lines were vernalized and planted in growth chambers under controlled conditions. Stem height, stem diameter, number of stems per plant, rosette diameter, rosette shape, and flowering time were recorded. All transgenic lines displayed significant increases in plant height and rosette diameter (Figures 16 and 17). Under normal growth conditions (16/8 hrs light regime), the maximum stem heights were 46.53 and 49.21 cm for OE-*PopGT1-3* and OE-*PopGT2*-6 transgenic lines, respectively, compared to 35.92 cm for Columbia (0), and 38.5 cm for Ws (0) wild-type controls (Tables 8 and 9).

In all the OE-*PopGT1* over-expressing lines, the increase in plant height over the controls ranged between 11-30%, except line OE-*PopGT1*-4, which showed significant reductions in all measured growth parameters. In addition, the increase in the average stem height of line OE-*PopGT1*-5 (39.92 cm) was not statistically significantly different from the wild-type controls. Meanwhile, lines OE-*PopGT2*-2, 4, 5, and 6 showed significant increases in stem height (Table 8).

Under continuous light (24 hours), OE-*PopGT1* and OE-*PopGT2*, showed significant increases in stem height (Figures 18) and rosette diameter compared to wild-type control. OE-*PopGT1* transgenic lines showed increases in stem height, ranging between 14-22% of the wild-type control, while, OE-*PopGT2* transgenic lines showed 5-24% increases.

Both transgenic lines, OE-*PopGT1* and OE-*PopGT2,* showed larger stem diameters when grown under a normal light regime. In all OE-*PopGT1* transgenic lines a significant increase in stem diameter was observed; the minimum and maximum increases in stem diameters was 4 and 19% in lines OE-*PopGT1*-2, and OE-*PopGT1*-3, respectively. However, the increases in the stem diameter of lines OE-*PopGT1*-2 and 5 were not statistically significant compared to wild-type control. Line OE-*PopGT1*-4 showed a different pattern than the rest of the transgenic lines, as the average stem diameter was reduced significantly (28.5%) compared to the control plants. In addition, non-significant increases in stem diameter were observed in all transgenic lines, when the plants were grown under 24 hours light (Figure 18). These increases ranged between 2-8% of the wild-type diameter. In contrast, lines OE-*PopGT1*-2 and 4 showed a non-significant reduction in stem diameter (Table 8).

Transgenic lines, over-expressing the *PopGT2* poplar gene, showed growth patterns similar to the *PopGT1* over-expressing plants. Lines OE-*PopGT2*-2, -4, -5, and -6 showed significant increases in stem diameter, when they were grown under a 16/8 hrs (light/dark) regime. The increases in stem diameter ranged between 7-19% of the average diameter of the wild-type controls (1.09 mm). Transgenic lines showed approximately an 8% increase in stem diameter, when grown under 24 hours light. However, these increases were not statistically different from wild-type control plants (Table 9).

Rosette diameters showed significant increases in all transgenic lines, however, the variability in their diameter was apparent compared to wild-type controls. These variabilities were also associated with morphological changes in rosette shape (Figure 19).

Transgenic OE-*PopGT1 Arabidopsis* lines showed significant changes in the average rosette diameters (lines OE-*PopGT1*-1, 3, 4, and 6), including increases in stem diameter that ranged between 8-24%. Lines OE-*PopGT1*-2 and 5 did not display significant morphology changes compared to wild-type controls. When the plants were grown under 24 hours light, 5-12% increase in rosette diameters was apparent. Again, line OE-*PopGT1*-4 showed a reduction in the rosette diameter (16%) compared to the wild-type controls.

Similarly, all OE-*PopGT2* transgenic lines showed increasing trends in rosette diameters compared to wild-type controls. Lines OE-*PopGT2*-1, 2 and 3 showed non-significant increases ranging between 4-8% of the wild-type rosette diameter. Meanwhile, lines OE-*PopGT2*-4, 5 and 6 showed significant increases in rosette diameters that ranged between 9.5-21% of the wild-type, when plants were grown under 16/8 hrs (light/dark) regime. When plants were grown under 24-hour light, stem diameter increased in all transgenic lines, except line OE-*PopGT2*-3. The increase in rosette diameters ranged between 5.6-20% of the wild-type control (Table 9). In addition, as a result of over-expressing the poplar genes in *Arabidopsis*, all transgenic lines showed an increased average number of stems per plant compared to wild-type controls. The OE-*PopGT1* transgenic lines, growing under 16/8 hrs light regime generated 3.65 stems per plant compared to wild-type controls (1.25 stems per plant). The maximum number of stems per plant was found on line OE-*PopGT1*-3 (3.65), whereas, the minimum number of stems was observed in line OE-*PopGT1*-6 (2.18). All the increases were significantly different compared to wild-type controls. When plants were grown under a 24-hour light regime, the number of stems per plant also increased on average compared to the control plants. Line OE-*PopGT1*-3 generated the highest number of stems (3.17) compared to 1.25 stems per plant in the wild-type controls.

Similarly, the *PopGT2* over-expressers showed similar phenotypes. The maximum and minimum number of stems per plant was 3.50 and 1.94, respectively, compared to the wild-type control (1.30) when plants were grown under normal (16/8 hrs) light regime. In contrast, under 24-hour light regime, the maximum number of stems per plant was 3.16 in line OE-*PopGT2*-3. Together, these results indicate that *Arabidopsis* growth patterns were significantly affected in response to over-expressing the poplar genes (*PopGT1* and *PopGT2*).

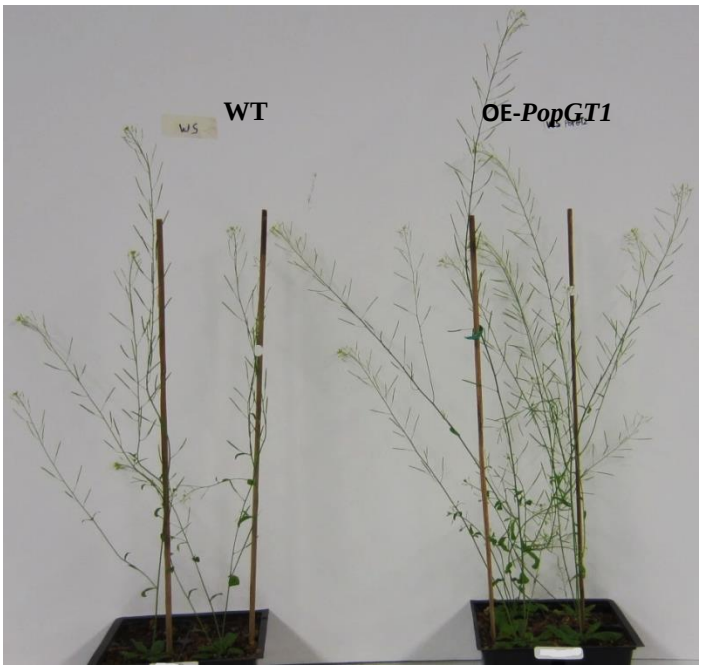


Figure 16: **Representative stem height of transgenic *Arabidopsis* lines over-expressing the poplar *PopGT1* gene under normal growth conditions compared to wild-type control plants.**

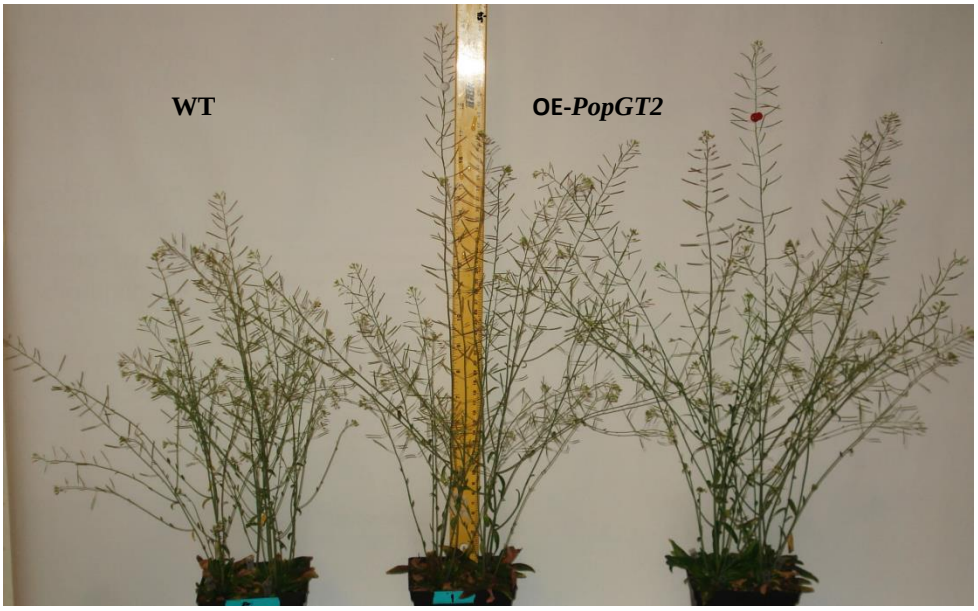


Figure 17: Representative stem height of transgenic *Arabidopsis* lines over-expressing poplar *PopGT2* gene, under normal growth conditions compared to wild-type control plants.

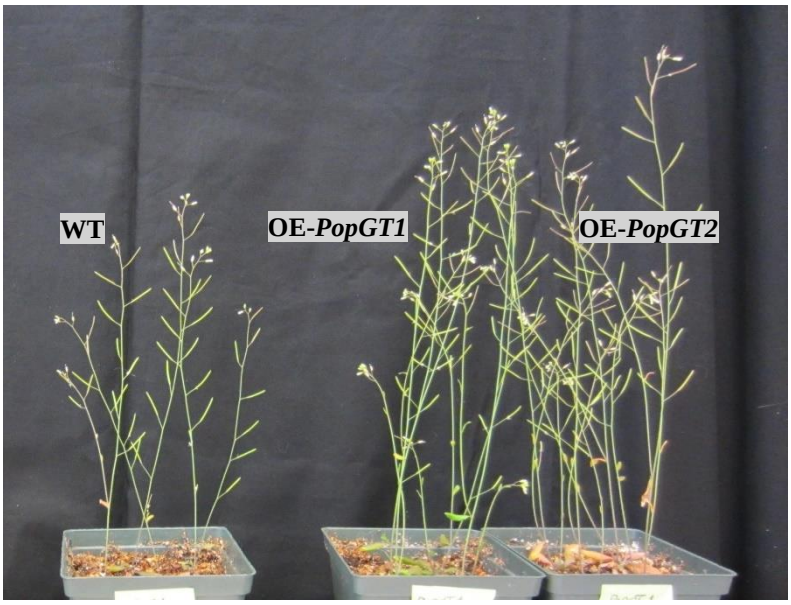


Figure 18: **Increased stem height phenotype of OE-*PopGT1* and OE-*PopGT2* transgenic *Arabidopsis* lines grown under 24 hours of continuous light.**

Table 8: Growth characteristics of transgenic OE-*PopGT1* compared to wild-type *Arabidopsis* plants grown under different light regimes (16/8 hours light/dark or 24-hour continuous light). Stem height (cm), stem diameter (mm), rosette diameter (mm), and number of stems per plant were recorded. Numbers in bold depict the significantly different means compared to wild-type controls, using *t*-test at α = 0.05. Numbers in brackets represent standard error of means (n = 10).

Genotype	16/8 hours (light/dark) grown plants				24-hour (light) grown plants			
	Stem height (cm)	**Stem diameter (mm)**	**Rosette diameter (mm)**	**Average number of stems**	**Stem height (cm)**	**Stem diameter (mm)**	**Rosette diameter (mm)**	**Average number of stems**
WT (Ws)	35.92 (0.36)	1.05 (0.02)	85.77 (0.21)	1.25 (0.10)	18.50 (0.32)	0.87 (0.03)	65.62 (0.75)	1.25 (0.25)
OE-*PopGT1*-1	**45.26 (1.07)**	**1.22 (0.04)**	**97.25 (0.64)**	**3.03 (0.15)**	**23.86 (0.45)**	0.94 (0.03)	**71.65 (0.58)**	**2.25 (0.22)**
OE-*PopGT1*-2	**42.44 (0.97)**	1.09 (0.03)	87.75 (0.83)	**2.25 (0.21)**	19.33 (0.53)	0.84 (0.03)	69.21 (0.73)	**1.65 (0.34)**
OE-*PopGT1*-3	**46.53 (0.62)**	**1.25 (0.02)**	**95.75 (0.40)**	**3.65 (0.22)**	**22.63 (0.66)**	0.89 (0.02)	**74.45 (0.67)**	**3.17 (0.25)**
OE-*PopGT1*-4	**23.41 (1.52)**	**0.65 (0.07)**	**65.12 (1.89)**	**3.67 (0.12)**	16.38 (0.35)	0.75 (0.01)	**54.40 (0.84)**	**2.37 (0.42)**
OE-*PopGT1*-5	39.91 (1.10)	1.07 (0.09)	86.33 (0.98)	**2.54 (0.23)**	**22.50 (0.67)**	0.90 (0.04)	71.77 (1.59)	**2.55 (0.25)**
OE-*PopGT1*-6	**43.93 (0.55)**	**1.12 (0.06)**	**93.67 (1.33)**	**2.18 (0.13)**	**22.95 (0.43)**	0.92 (0.02)	**73.15 (0.45)**	**1.85 (0.47)**

Table 9: Growth characteristics of transgenic OE-*PopGT2* compared to wild-type *Arabidopsis* plants grown under different light regimes (16/8 hours light/dark or 24-hour continuous light). Stem height (cm), stem diameter (mm), rosette diameter (mm), and number of stems per plant were recorded. Numbers in bold depict the significantly different means compared to wild-type controls, using t-test at α = 0.05. Numbers in brackets represent standard error of means (n = 10).

Genotype	16/8 hours (light /dark) grown plants				24-hour (light) grown plants			
	Stem height (cm)	**Stem diameter (mm)**	**Rosette diameter (mm)**	**Average number of stems**	**Stem height (cm)**	**Stem diameter (mm)**	**Rosette diameter (mm)**	**Average no of stems**
WT (col)	38.50 (0.50)	1.09 (0.02)	88.67 (0.62)	1.30 (0.23)	21.57 (0.99)	0.75 (0.05)	63.61 (0.55)	1.50 (0.25)
OE-*PopGT2*-1	**43.33 (0.48)**	1.17 (0.02)	101.20 (2.12)	**1.94 (0.23)**	**22.91 (0.31)**	0.81 (0.06)	**71.12 (0.52)**	**2.66 (0.12)**
OE-*PopGT2*-2	**45.14 (0.75)**	**1.20 (0.01)**	94.14 (1.31)	**2.50 (0.31)**	**23.34 (0.57)**	0.72 (0.05)	**68.56 (0.56)**	**2.50 (0.15)**
OE-*PopGT2*-3	41.76 (0.87)	1.07 (0.01)	92.57 (0.89)	**2.45 (0.28)**	**23.62 (0.89)**	0.80 (0.04)	69.17 (1.02)	**3.16 (0.30)**
OE-*PopGT2*-4	**48.88 (0.83)**	**1.23 (0.02)**	**104.38 (1.13)**	**3.10 (0.44)**	**26.75 (0.47)**	**0.83 (0.07)**	**69.84 (0.65)**	**3.00 (0.48)**
OE-*PopGT2*-5	**46.71 (0.79)**	**1.27 (0.02)**	**105.14 (0.78)**	**3.40 (0.19)**	**25.43 (0.54)**	0. 81 (0.08)	**73.23 (0.37)**	**2.57 (0.22)**
OE-*PopGT2*-6	**49.21 (1.04)**	**1.30 (0.01)**	**110.17 (1.67)**	**3.50 (0.31)**	**25.11 (0.36)**	0.84 (0.04)	**76.79 (0.36)**	**3.00 (0.33)**

Under normal growth conditions, both OE-*PopGT1* and OE-*PopGT2* over-expression lines showed varied rosette morphology. In the OE-*PopGT1* plants approximately 10% of the transgenic lines showed irregular rosette morphology. These plants showed asymmetric rosette growth, irregular leaf expansion, stunted growth, hyponastic leaf growth, dense leaves of variable sizes, and flowering on stunted inflorescences (Figure 19 a-e).

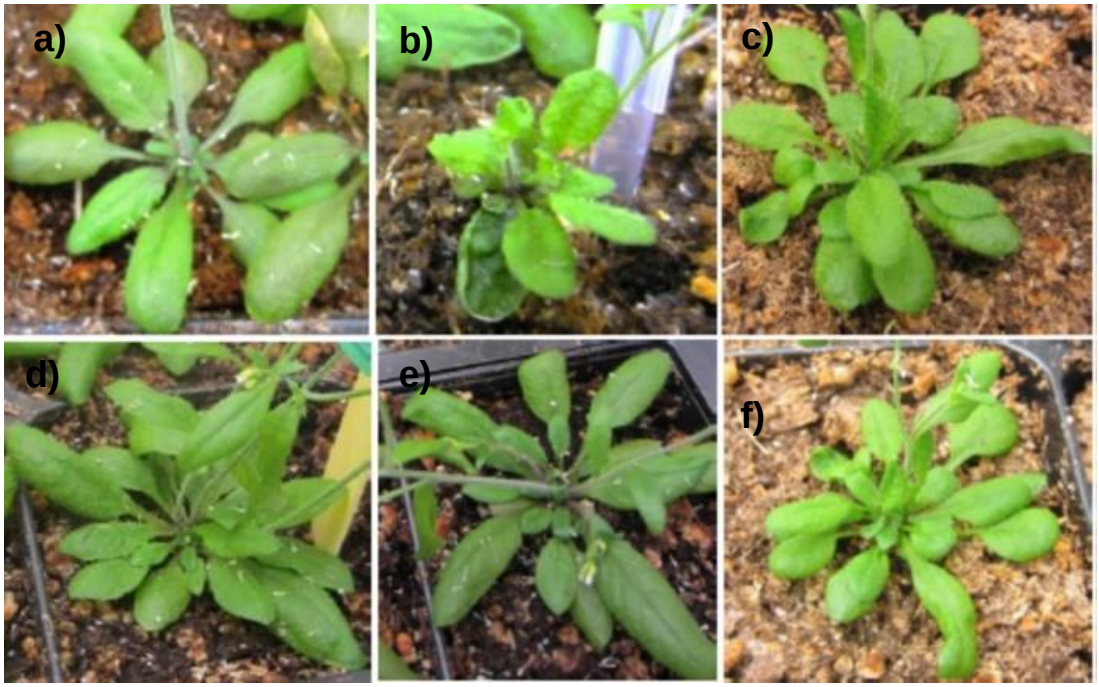


Figure 19: Irregular rosette morphology of five-week old transgenic OE-*PopGT1* and OE-*PopGT2* *Arabidopsis* plants compared to (a) wild-type control, (b and c) OE-*PopGT1*, (d, e, and f), (b) OE-*PopGT2* Stunted growth, (c, d, and f) condensed leaves hyponastic, (b, c and f) leaf-growth, (e) irregular flowering on stunted inflorescences, (b, c, d, and f) asymmetric rosette growth, (d and e) multiple stems per plant.

3.5.2.2 Plant flowering phenotype

Flowering time was monitored and recorded daily after bolting (two weeks post-planting) in a growth chamber. Both transgenic lines (OE-*PopGT1* and OE-*PopGT2*) flowered 4 days earlier

than wild-type control (Figure 20). The number of plants that flowered was scored each day, and the percentage of flowering plants was then calculated. In general, the flowering rates of the OE-*PopGT1* transgenic lines were faster than both the wild-type control plants and the OE-*PopGT2*. At 24 days post planting, 80-100% of the OE-*PopGT1* transgenic lines were flowering, while 60-80% of the OE-*PopGT2* were flowering (Figure 21 a and b).

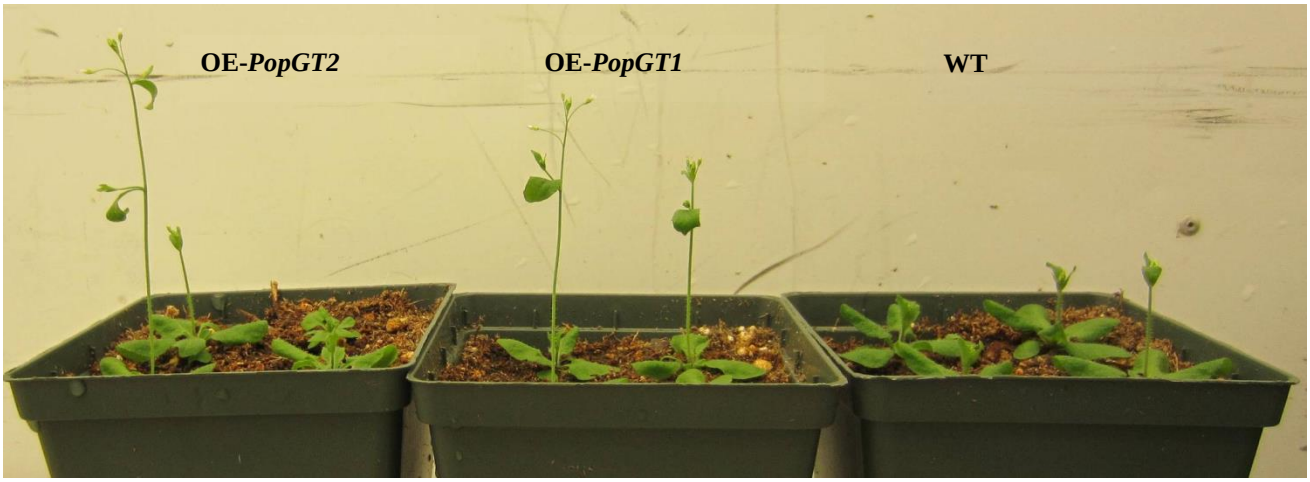


Figure 20: Early bolting and flowering of transgenic *Arabidopsis* plants compared to wild-type control.

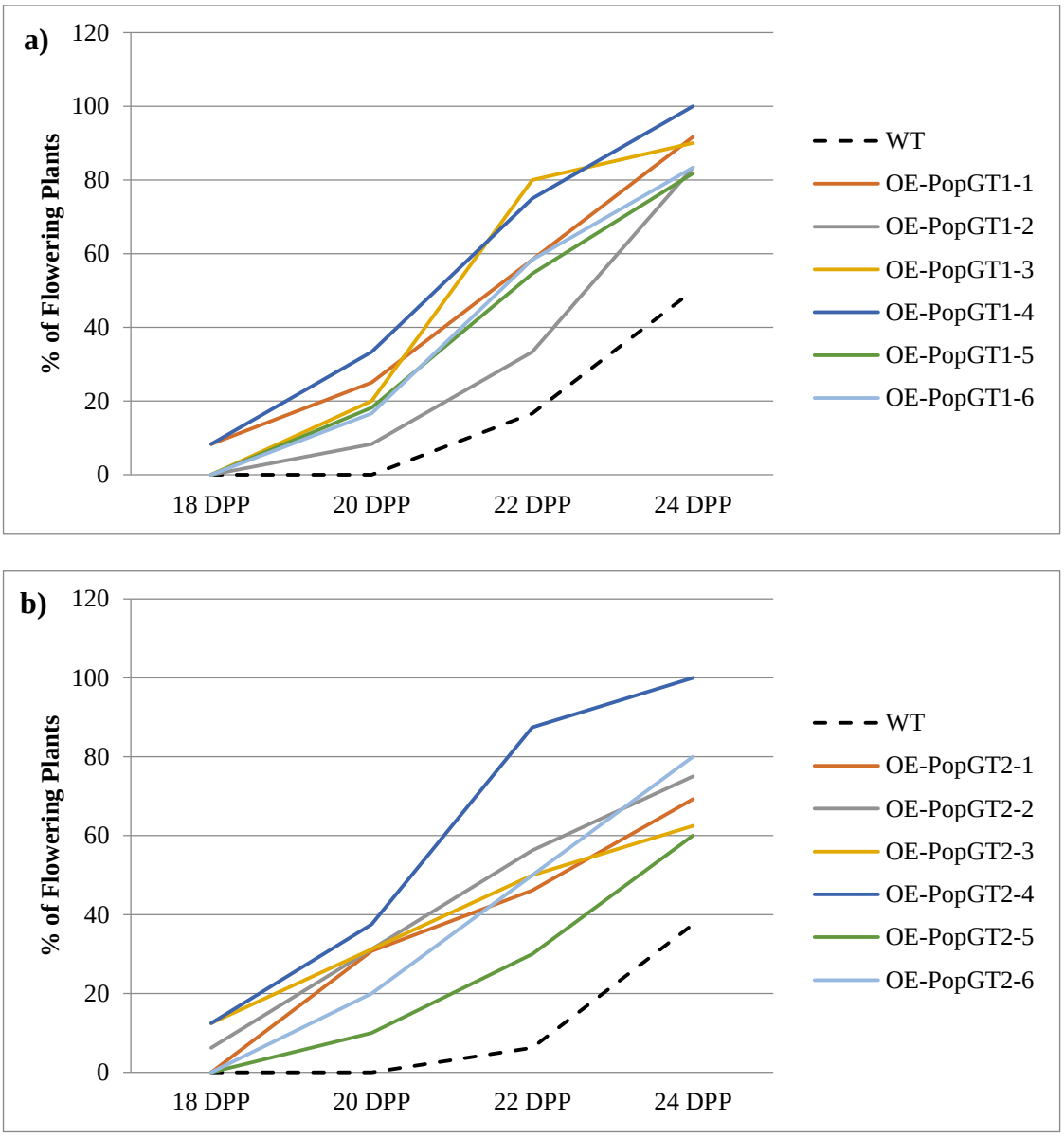


Figure 21: **Percentage of flowering plants in (a) OE-*PopGT1* and (b) OE-*PopGT2* transgenic *Arabidopsis* lines. Each line represents the percentages of flowering plants daily between 18-24 days post planting (DPP). Wild-type controls are represented by the dashed lines (n = 12).**

3.5.3 Stem cross sectioning and microscopy

The effect of introducing the two poplar genes (*PopGT1* and *PopGT2*) on the stem anatomy of *Arabidopsis* plants was investigated, using histochemical and microscopic observations.

Comparison of stem cross-sections of both transgenic lines (OE-*PopGT1* and OE-*PopGT2*) with those of wild-type *Arabidopsis* main inflorescences showed no noticeable changes in the tissue anatomy. These results were consistent for plants grown either under a 16/8 hrs (light/dark) or 24-hour light regimes (Figures 22 and 23).

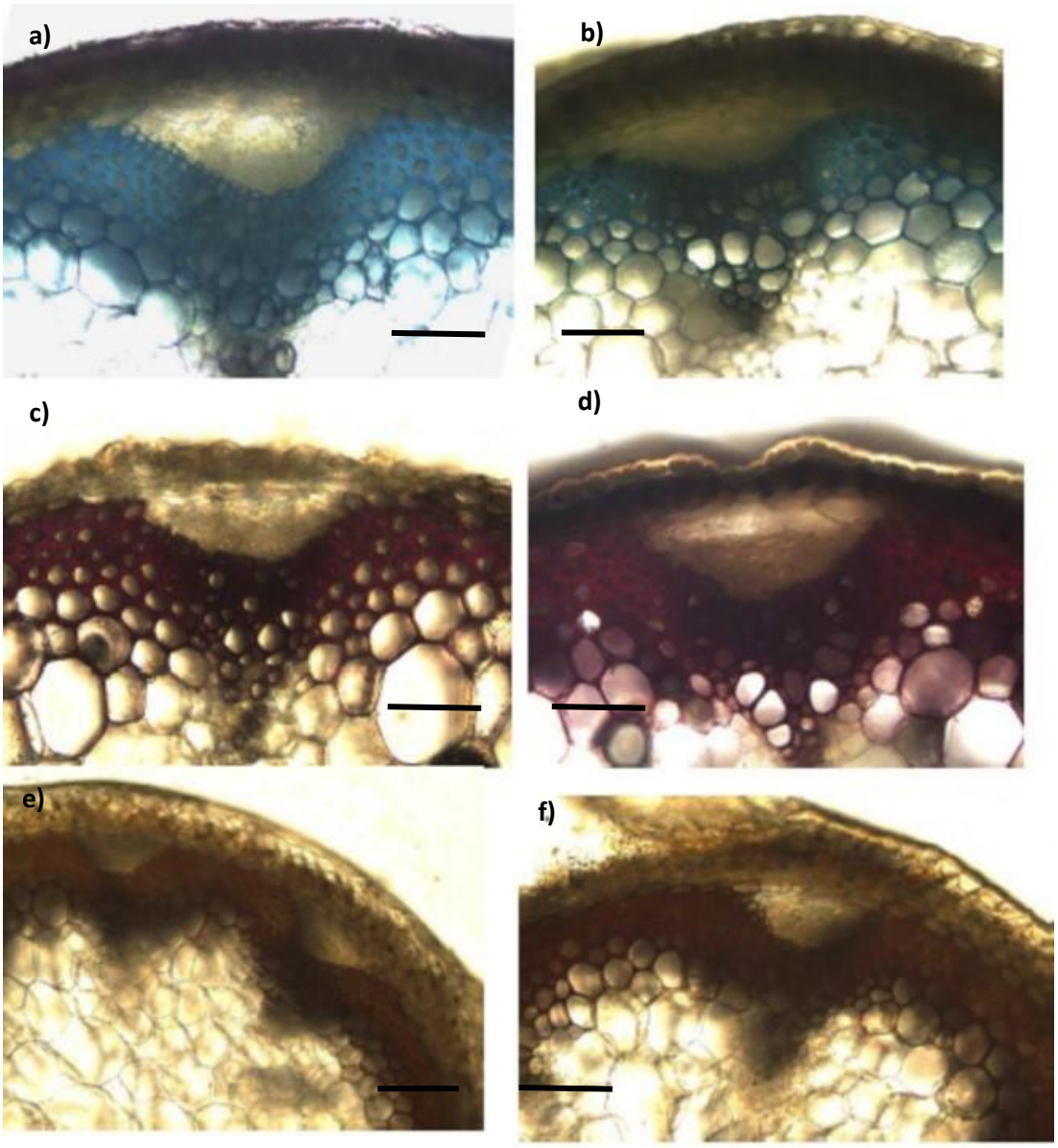


Figure 22: **Stem cross-section of wild type (a, c, and e) and OE-*PopGT1* (b, d, and f) inflorescences. Cross-sections stained with toluidine blue (a, b), phloroglucinol HCl (c, d), and Maüle stain (e, f) are shown. Scale bar = 50 µM.**

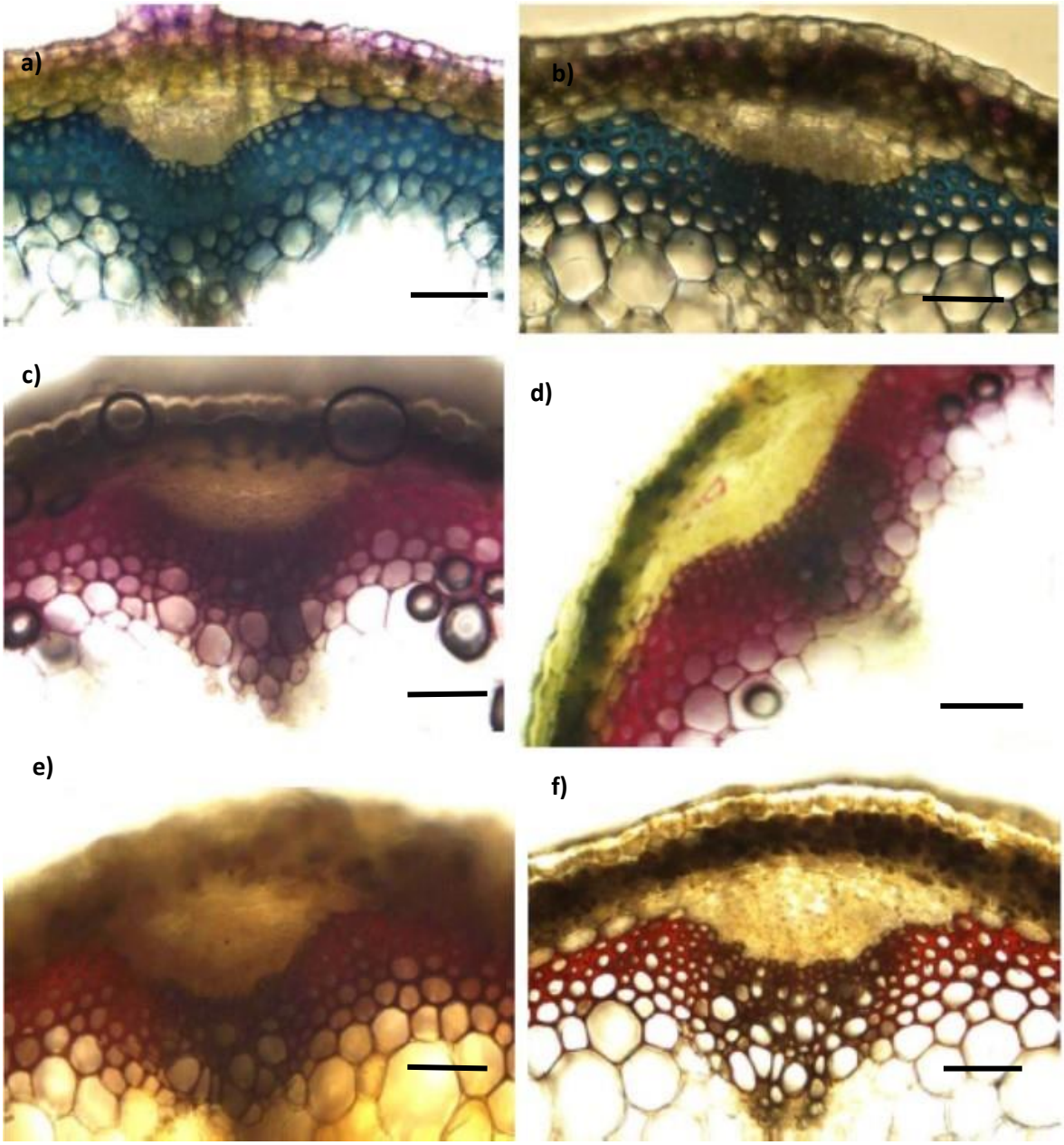


Figure 23: Stem cross-section of wild type (a, c, and e) and OE-*PopGT2* (b, d, and f) inflorescences. Cross-sections stained with toluidine blue (a, b), phloroglucinol HCl (c, d), and Maüle stain (e, f) are shown. Scale bar = 50 µM.

3.5.4 Root phenotype

Both OE-*PopGT1* and OE-*PopGT2* transgenic *Arabidopsis* lines showed a spiral-root phenotype (Figure 24 a). The transgenic seeds, when germinated on flat agar plates with or without the selection antibiotic (hygromycin), showed the same root phenotype. In some transgenic lines (1, 3, 4 and 6 of OE-*PopGT1* and lines 1, 3, 4, 5 and 6 of OE-*PopGT2*) the newly emerging roots initially grew straight, then began to grow as spiral roots (Figure 24 a). However, when the same seeds were grown on upright plates, some seedlings displayed arrested root growth, as their roots started to loop without any sense of gravity. In contrast, other roots initially grew normally and showed similar gravitropic curvature on the tilted plates, but then started to loop, creating knot-like structures. Interestingly, the roots continued to grow normally again after forming a loop (Figure 24 b). This growth phenotype was not observed in any of the wild-type plants, which showed a normal gravitropic response. The transgenic roots showed random growth patterns when they started to loop around themselves. In addition, more root hairs were observed on the transgenic lines (Figure 24 b and c). Microscopic investigation of the roots at the looping points showed that root epidermal cell-size was asymmetric on either side of the looping root, compared to the wild-type controls (arrows in Figure 24 c). It was apparent that the transgenic lines with spiral roots showed reduced growth parameters (stem height and diameter), more branching, more rosette deformation, and more stems per plant than those exhibiting a weaker spiral root phenotype, increasing the degree of variability in the measured traits.

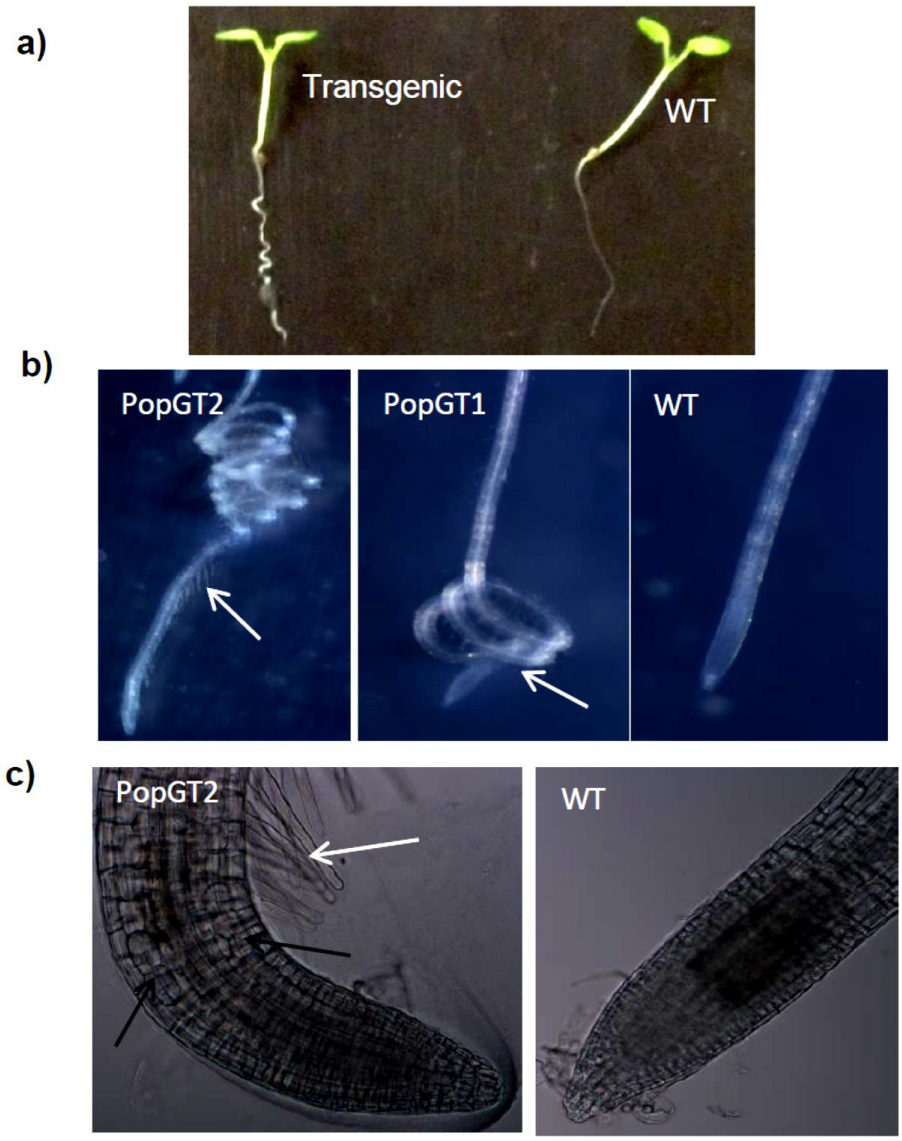


Figure 24: Spiral root phenotype of transgenic *Arabidopsis* plants compared to wild-type control, Spiral roots of *Arabidopsis* OE-*PopGT1* and OE-*PopGT2* transgenic plants (a, b). Black arrows depict enlarged cells on both sides of the bending points (c), while white arrows indicate grown more hairy roots on the transgenic plants (b and c).

3.5.5 Cell wall chemical composition

Chemical analysis carried out on the main inflorescences of six *Arabidopsis* OE-*PopGT1* and six OE-*PopGT2* transgenic lines revealed that the levels of structural sugars changed in both OE-

PopGT1 and OE-*PopGT2* lines in response to ectopic expression of the poplar genes in *Arabidopsis* plants. In OE-*PopGT1*plants, arabinose, galactose, and mannose were reduced compared to the wild-type control, and these reductions were statistically significant in the transgenic lines, OE-*PopGT1*- 1, 2, 3 and 4. Lines OE-*PopGT1*-5 and OE-*PopGT1*-6 followed the same trend, but the changes were not statistically significant. In addition, glucose and xylose levels in three OE-*PopGT1* transgenic lines (OE-*PopGT1*-1, 3, and 4) showed significant increases, while the remaining lines showed increasing trends compared to the wild-type control. Rhamnose levels varied in all transgenic lines, and only OE-*PopGT1*-4 showed a significant reduction in its content compared to wild-type controls (Table 10).

Total lignin content also showed increasing trends in the transgenic lines. The changes were statistically significant in lines OE-*PopGT1*-1, 3, and 4, while the total lignin content in the three remaining lines (OE-*PopGT1*-2, 5, and 6) was increased, but not significantly. The altered total lignin content was due to increases in acid-soluble lignin (Table 10).

In the OE-*PopGT2* transgenic lines, the structural sugar content was also variable. Three lines showed significant reductions in arabinose, rhamnose, and galactose content (OE-*PopGT2*-4, 5, and 6). In addition, all lines showed a significant reduction in mannose content, except line OE-*PopGT2*-3 that showed a higher mannose content than the wild-type plants, however, the increase in mannose content in line OE-*PopGT2*-3 was not significantly different from the wild-type control. Glucose showed an increasing trend in all transgenic lines, except line OE-*PopGT2*-3. The increase in glucose content was significantly different from the wild-type controls in three lines, namely OE-*PopGT2*-4, 5, and 6. The OE-*PopGT2*-2 transgenic line showed a 0.36%

increase in cell wall glucose content, but this increase was not statistically significant from the wild-type controls (Table 11).

Total lignin content increased in all transgenic lines, but these increases were not statistically different from the wild-type controls, except in line OE-*PopGT2*-6. This line showed a significant increase in the total lignin content, as well as the highest acid-soluble lignin content compared to the wild-type plants. All the lines showed a non-significant change in acid-insoluble lignin content, as well as an increased amount of acid-soluble lignin. These increases were significantly different in OE-*PopGT2*-2, 4, 5 and 6 transgenic lines, compared to the control plants. As described above, the amount of hemicellulose-derived sugars was variable in the transgenic plants, compared to wild-type controls. The general trend was a reduction in the total amount of sugars.

Table 10: Cell wall chemical composition of transgenic *Arabidopsis* lines (OE-*PopGT1*) overexpressing the poplar *PopGT1* and wild-type plants, represented as a percentage of the weight of pooled extract-free primary inflorescence tissues. Bold numbers represent the statistically different averages (%), using *t*-test at α = 0.05 (n = 4). Standard error values are shown in brackets.

	Structural sugars (%)						Lignin (%)		
	A	**R**	**Ga**	**G**	**X**	**M**	**Total lignin**	**Acid non-soluble**	**Acid soluble**
Wild type	0.84 (0.02)	0.59 (0.03)	1.49 (0.02)	27.13 (0.44)	10.94 (0.28)	1.65 (0.02)	20.75 (0.20)	18.88 (0.14)	1.97 (0.14)
OE-*PopGT1*-1	**0.72 (0.01)**	0.54 (0.05)	**1.41 (0.02)**	**28.12 (0.13)**	**11.70 (0.15)**	**1.47 (0.02)**	**21.35 (0.11)**	18.76 (0.19)	**2.59 (0.05)**
OE-*PopGT1*-2	**0.77 (0.01)**	0.64 (0.04)	**1.40 (0.02)**	27.49 (0.32)	10.52 (0.19)	**1.52 (0.03)**	20.82 (0.25)	18.82 (0.12)	2.00 (0.19)
OE-*PopGT1*-3	**0.74 (0.02)**	0.55 (0.01)	**1.39 (0.02)**	**28.60 (0.18)**	**11.57 (0.06)**	**1.58 (0.03)**	**21.75 (0.11)**	19.05 (0.18)	**2.71 (0.12)**
OE-*PopGT1*-4	**0.62 (0.04)**	**0.51 (0.02)**	**1.35 (0.03)**	**28.32 (0.33)**	**12.05 (0.21)**	**1.45 (0.03)**	**21.57 (0.16)**	18.71 (0.20)	**2.86 (0.14)**
OE-*PopGT1*-5	0.82 (0.03)	0.56 (0.03)	1.45 (0.02)	27.46 (0.20)	11.06 (0.13)	1.67 (0.03)	21.08 (0.23)	18.93 (0.26)	2.15 (0.16)
OE-*PopGT1*-6	0.80 (0.01)	0.57 (0.03)	1.47 (0.02)	27.16 (0.29)	11.03 (0.26)	1.68 (0.05)	21.10 (0.32)	19.06 (0.14)	2.04 (0.13)

Arabinose (A), Rhamnose (R), Galactose (Ga), Glucose (G), Xylose (X), Mannose (M).

Table 11: Cell wall chemical composition of transgenic *Arabidopsis* lines (OE-*PopGT2*), overexpressing the poplar *PopGT2* and wild-type plants, represented as a percentage of the weight of pooled extract-free primary inflorescence tissues. Bold numbers represent the statistically different means, using *t*-test at α = 0.05 (n = 4). Standard error values are shown in brackets.

	Structural sugars (%)						Lignin		
	A	**R**	**Ga**	**G**	**X**	**M**	**Total lignin**	**Acid non-soluble**	**Acid soluble**
Wild type	0.74 (0.03)	0.49 (0.01)	1.27 (0.03)	23.13 (0.31)	10.64 (0.08)	1.29 (0.03)	20.40 (0.29)	18.29 (0.27)	2.11 (0.04)
OE-*PopGT2*-1	0.65 (0.04)	0.47 (0.02)	1.20 (0.02)	23.32 (0.19)	10.80 (0.17)	**1.17 (0.01)**	20.53 (0.20)	18.44 (0.55)	2.09 (0.09)
OE-*PopGT2*-2	0.72 (0.21)	0.45 (0.01)	1.16 (0.02)	23.49 (0.26)	11.20 (0.19)	**1.12** (0.02)	20.70 (0.17)	18.20 (0.69)	**2.50 (0.02)**
OE-*PopGT2*-3	0.74 (0.01)	0.47 (0.02)	1.15 (0.01)	23.06 (0.08)	10.07 (0.06)	1.25 (0.04)	20.72 (0.44)	18.45 (0.32)	2.27 (0.10)
OE-*PopGT2*-4	**0.62 (0.02)**	**0.41** (0.01)	**1.05** (0.03)	**24.20 (0.13)**	10.05 (0.17)	**0.95 (0.03)**	20.92 (0.26)	18.06 (0.15)	**2.86 (0.04)**
OE-*PopGT2*-5	**0.63 (0.01)**	**0.44 (0.01)**	**1.08 (0.02)**	**23.86 (0.26)**	10.76 (0.13)	**1.05 (0.03)**	21.15 (0.23)	18.21 (0.79)	**2.94 (0.06)**
OE-*PopGT2*-6	**0.57** (0.01)	**0.39** (0.01)	**1.01 (0.02)**	**24.32 (0.14)**	11.03 (0.06)	**1.08 (0.01)**	**21.29 (0.18)**	18.14 (0.26)	**3.15 (0.04)**

Arabinose (A), Rhamnose (R), Galactose (Ga), Glucose (G), Xylose (X), Mannose (M)

3.5.6 Analysis of phenolic metabolites

HPLC profiling of the methanolic extracts of the transgenic *Arabidopsis* leaves showed significant changes in the phenolic profiles, in both total and acid-hydrolyzed metabolites. However, SA was not detected among the phenolic metabolites (Figure 25). Phenolic metabolites showing UV spectra typical of flavonoids were detected in the total methanol extract of equivalently developed *Arabidopsis* leaves. However, none of these metabolites could be identified based on comparison to a wide range of authentic standards.

In parallel, acid hydrolysis was conducted on the methanolic extracts (containing the total metabolites of the *Arabidopsis* leaves), which again showed clear differences in metabolite concentrations between the transgenic and wild-type plants. Some of these metabolites showed UV spectra similar to the flavonoid-specific spectra. Among them, kaempferol was identified and its concentration was significantly increased in the transgenic lines compared to the wild-type controls (Figure 26).

The HCl-hydrolyzed metabolites showed different patterns in both OE-*PopGT1* and OE-*PopGT2* transgenic lines. In OE-*PopGT1* lines, acid hydrolyzed metabolites were variable in all transgenic lines compared to wild-type controls. The only identified compound, kaempferol, showed an increased normalized peak area in all transgenic lines (normalized using the internal standard peak area). The significant increases ranged between 1.2-2.7 times more than the wild-type controls. Peak a1 showed increasing trend in peak area, which ranged between 1.4-2.6 times the average peak area of the same peak in the wild-type controls. This compound showed a high variability in lines 1 and 6, and showed significant increases compared to the wild-type controls. Peak a2 showed significant increases in lines OE-*PopGT1*-1, 3, 4, and 6. The maximum increase was 2.3 times that of the

average peak area of wild-type plants (line OE-*PopGT1*-4). The increases in peaks a3 and a6 were not statistically significant from the wild-type plants (Table 12 and Figure 27).

In the OE-*PopGT2* lines, the normalized peak area (relative to the internal standard) significantly higher for peaks p1, p2, p5, and p6. Peak p5 was identified as kaempferol, by comparison to the authentic compound (Figure 28). The increase in the kaempferol normalized peak area in the transgenic lines ranged between 1.1-4.8 times compared to the wild-type plants. The change in the kaempferol concentrations was significantly higher in lines OE-*PopGT2*-1, 3, 4, 5, and 6, with the maximum concentration found in line OE-*PopGT2*-6 (Table 13). Meanwhile, OE-*PopGT2*-2 showed no significant difference compared to wild-type control at $\alpha = 0.05$. In addition, the unknown peaks (p1, p2, p3, and p6) showed variable trends in the transgenic lines compared to the wild-type controls. Peak p1 showed significant increases in all the OE-*PopGT2* transgenic lines. These increases ranged between 1.3- and 1.9-fold compared to wild-type plants. Peak p2 showed the same trend and the increase in its peak area ranged between 1.6 and 2.6 times the wild-type plants. Peak p3 did not show significant changes in the transgenic lines compared to the wild-type plants. Peak p6 showed increasing trends in the transgenic lines compared to the wild-type controls (on average). However, the increases were not statistically significant from the wild-type controls, except in line OE-*PopGT2*-4, which showed twice the wild-type plants content.

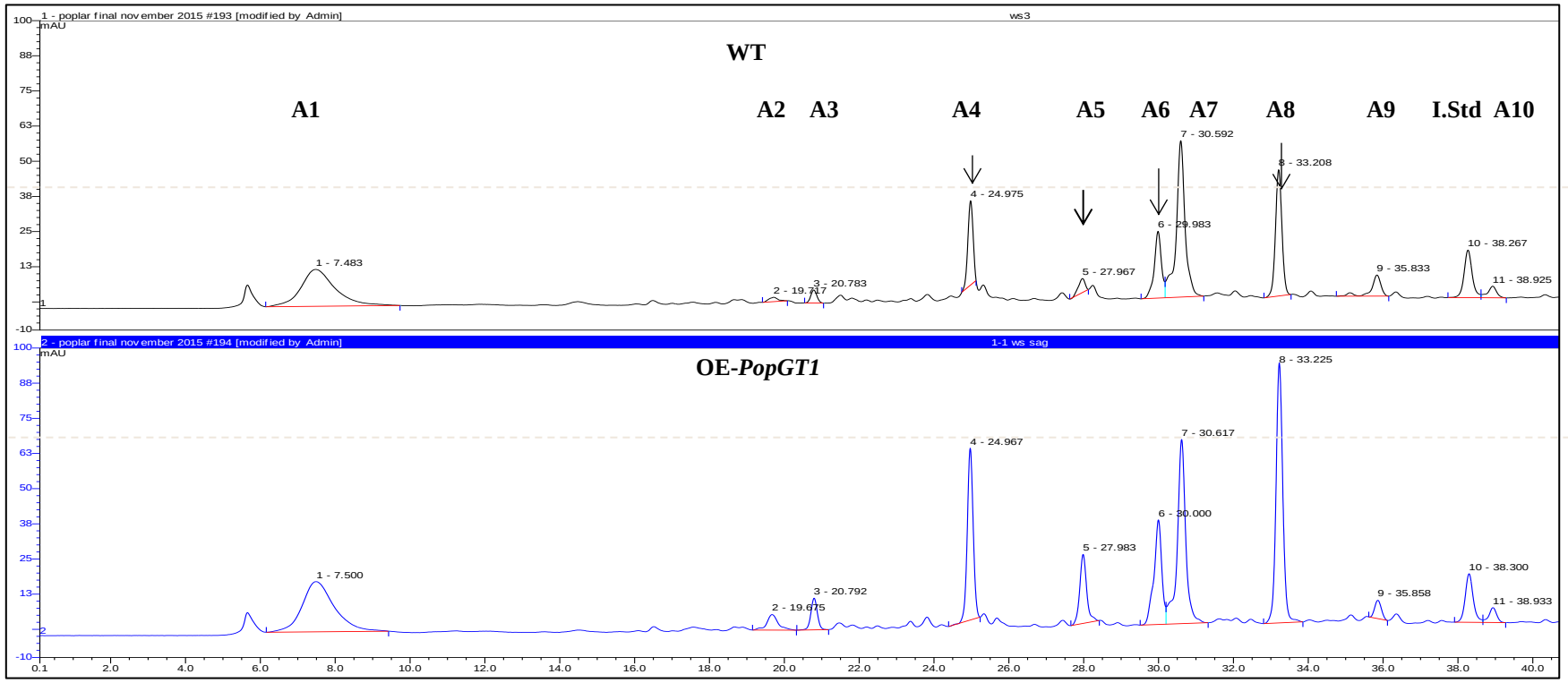


Figure 25: Representative HPLC trace for the total-metabolite profiles of both wild-type *Arabidopsis* and OE-*PopGT1* transgenic lines. Peaks are labelled (A1 to A10) and the 3,4,5-trimethoxycinnamic acid internal standard (I.Std). Arrows depict peaks that showed altered concentrations in transgenic plants compared to wild-type plants.

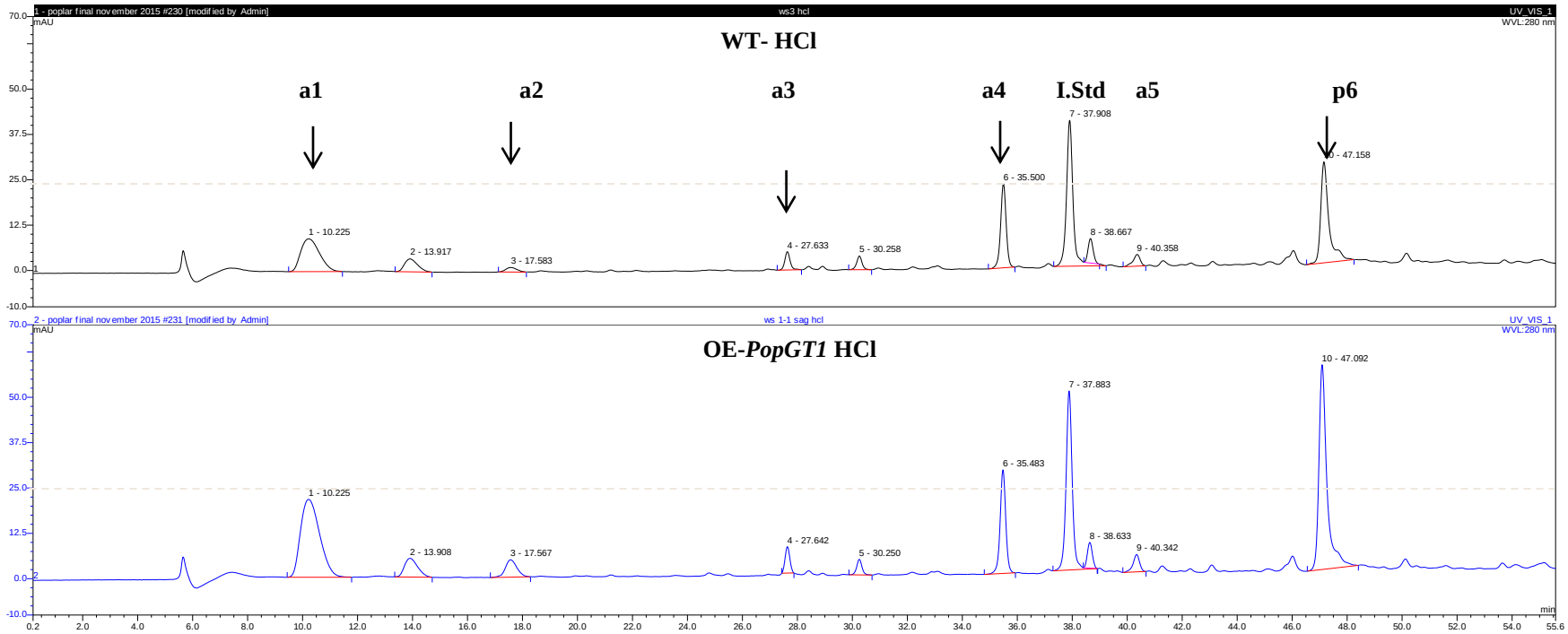


Figure 26: Representative HPLC trace for the acid-hydrolyzed metabolite profiles of both wild-type *Arabidopsis* and OE-*PopGT1* transgenic lines. The major peaks were labelled (a1 to a6) and the 3,4,5-trimethoxycinnamic acid internal standard (I.Std). Arrows depict metabolites that showed altered concentrations in transgenic plants compared to wild-type plants.

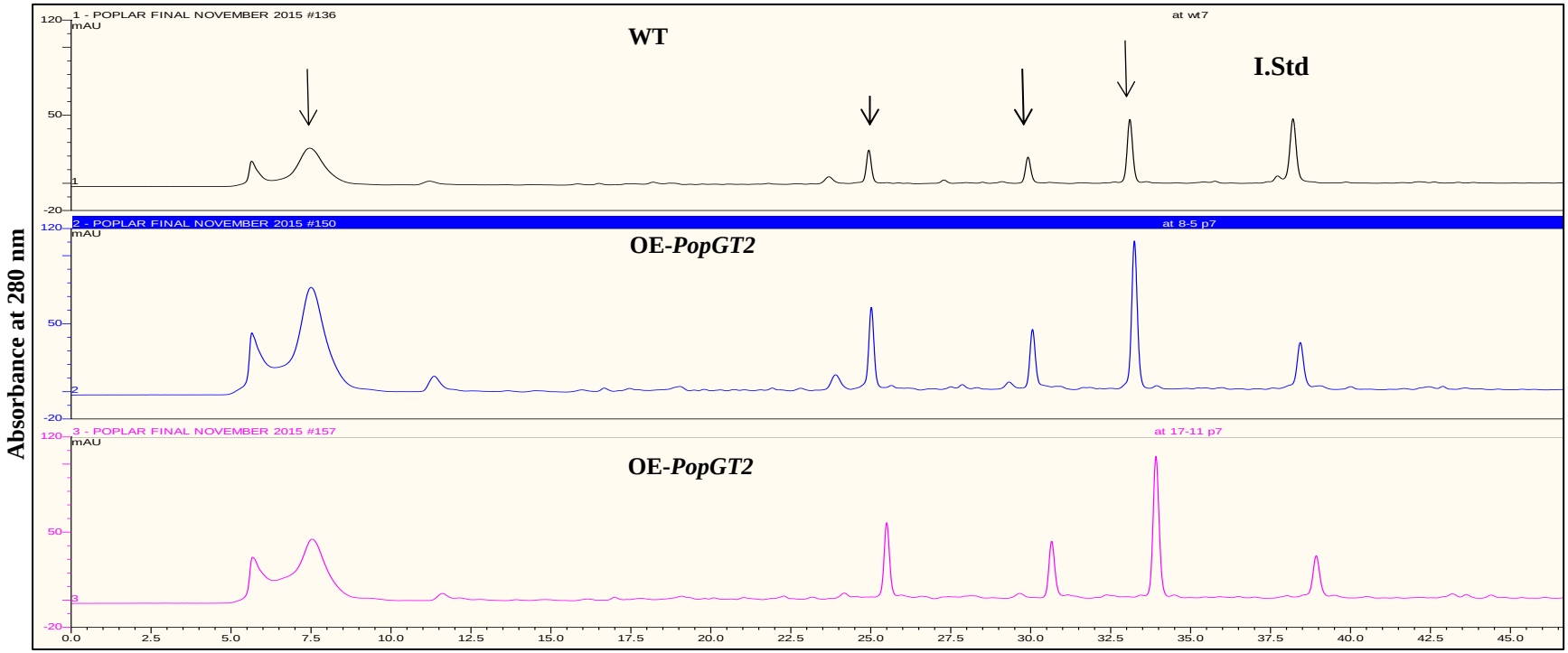


Figure 27: Representative HPLC trace for the total-metabolite profiles of both wild-type *Arabidopsis* and OE-*PopGT2* transgenic lines. Arrows depict peaks that showed altered concentrations in transgenic plants compared to the wild-type plants and the 3,4,5-trimethoxycinnamic acid internal standard was labelled (I.Std).

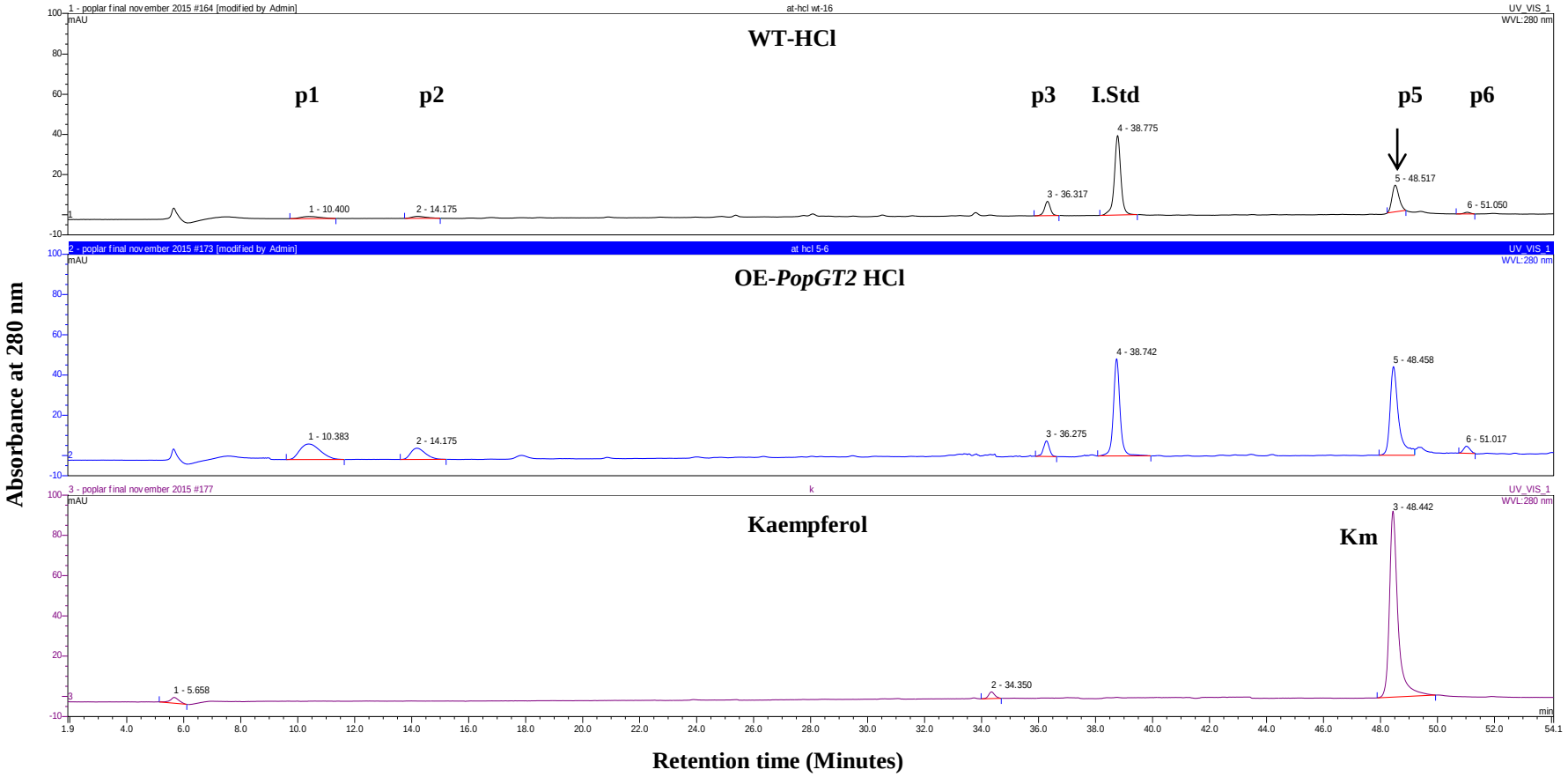


Figure 28: Representative HPLC trace for the acid-hydrolyzed-metabolite profiles of both wild-type *Arabidopsis* and OE-*PopGT2* transgenic lines. The peaks were labelled (p1 to p6) and the 3,4,5-trimethoxycinnamic acid internal standard (I.Std). Arrow depicts the kaempferol (Km) peak that showed altered concentrations in transgenic plants compared to the wild-type plants.

Table 12: Normalized peak area of the acid-hydrolyzed methanol extracts of *Arabidopsis* OE-*PopGT1* freeze-dried leaf tissues. Peaks (a1-a6) are the major peaks detected, where p5 was identified as kaempferol. Bold numbers depict the statistically significant mean, using t-test at α = 0.05. Standard error are shown in brackets (n = 4).

Genotype	Peak a1	Peak a2	Peak a3	Kaempferol a5	Peak a6
WT (Ws)	3.99 (0.51)	1.30 (0.12)	6.43 (0.37)	6.82 (0.49)	0.43 (0.03)
OE-*PopGT1*-1	**10.41 (2.76)**	**2.14 (0.68)**	5.55 (0.05)	**13.23 (0.90)**	0.61 (0.13)
OE-*PopGT1*-2	**5.44 (1.48)**	1.72 (0.16)	7.34 (1.51)	7.37 (0.54)	0.47 (0.02)
OE-*PopGT1*-3	**7.88 (0.35)**	**2.55 (0.65)**	6.27 (0.48)	**11.19 (1.19)**	0.52 (0.04)
OE-*PopGT1*-4	**8.10 (0.13)**	**3.06 (0.16)**	7.72 (1.11)	**18.04 (0.31)**	0.77 (0.13)
OE-*PopGT1*-5	**5.80 (1.99)**	2.46 (0.71)	6.34 (0.76)	**8.45 (0.52)**	0.55 (0.29)
OE-*PopGT1*-6	**7.39 (3.78)**	**3.31 (0.66)**	5.60 (0.02)	**9.26 (0.16)**	0.46 (0.18)

Table 13: Normalized peak area of the acid-hydrolyzed methanol extracts of *Arabidopsis* OE-*PopGT2* freeze-dried leaf tissues. Peaks (p1-p6) are the major peaks detected, where p5 is kaempferol. Bold numbers depict the statistically significant mean, using t-test at α = 0.05. Standard error are shown in brackets (n = 4).

Genotype	Peak p1	Peak p2	Peak p3	Kaempferol p5	Peak p6
WT (Col)	1.16 (0.02)	2.06 (0.44)	2.49 (0.21)	3.89 (0.85)	0.70 (0.07)
OE-*PopGT2*-1	**2.43** (0.05)	**4.40** (1.15)	2.28 (0.14)	**5.58** (0.29)	0.90 (0.04)
OE-*PopGT2*-2	1.42 (0.03)	**4.19** (0.87)	2.57 (0.09)	4.38 (0.38)	0.94 (0.05)
OE-*PopGT2*-3	**2.04** (0.02)	**3.34** (0.73)	1.81 (0.11)	**6.53 (0.59)**	1.15 (0.03)
OE-*PopGT2*-4	**2.28** (0.08)	**4.07** (0.68)	2.14 (0.24)	**14.31** (0.43)	**2.01 (0.05)**
OE-*PopGT2*-5	**2.08** (1.04)	**4.07** (0.33)	2.28 (0.22)	**12.41** (0.77)	1.05 (0.07)
OE-*PopGT2*-6	**2.74** (0.07)	**5.38** (1.08)	2.01 (0.17)	**15.95** (0.35)	0.91 (0.03)

3.6 Down-regulation of *PopGT1* and *PopGT2* in hybrid poplar trees

To investigate the effect of down-regulation of the two putative poplar UGT genes (*PopGT1* and *PopGT2*) on growth and development, two *pHellsgate12*-RNAi constructs were generated and used to suppress the expression of *PopGT1* and *PopGT2*. The first construct, designed to down-

regulate the *PopGT1*, resulted in the generation of RNAi-suppression lines (RNAi-*PopGT1* A1-A9), while the second construct, designed to down-regulate the *PopGT2*, also resulted in 9 transgenic lines (RNAi-*PopGT2* lines B1-B9).

The RNAi-transgenic poplar lines clearly showed reductions in the expression of the target genes relative to wild-type trees (Figures 29 a and b).

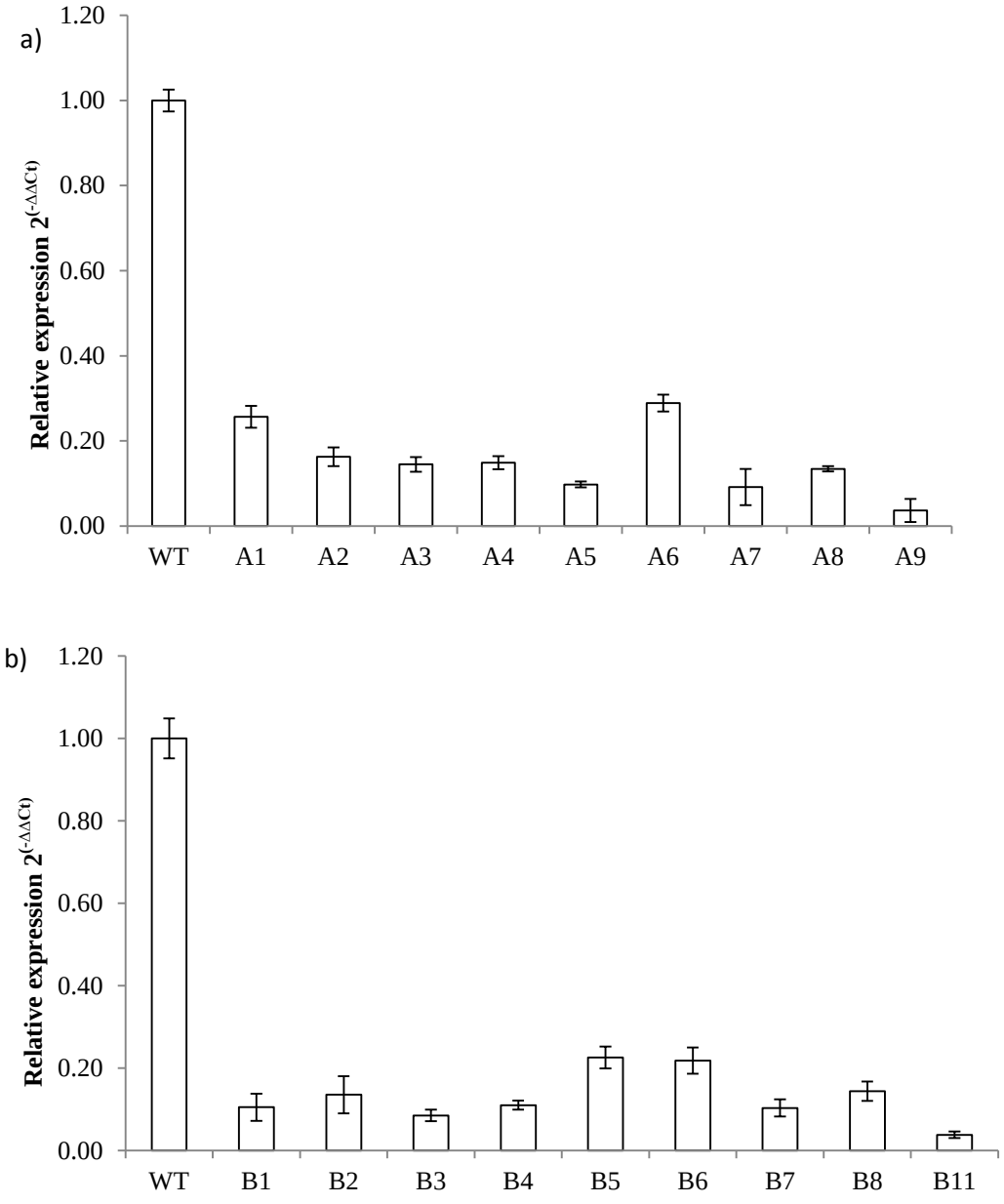


Figure 29: Gene expression of (a) RNAi-PopGT1 (A1-A9) and b) RNAi-PopGT2 (B1-B8 and B11) hybrid poplar transgenic lines relative to the gene expression of *PopGT1* and *PopGT2* genes in wild-type leaves. Error bars represent the standard deviation of 3 biological replicates.

3.6.1 Plant growth phenotype

3.6.1.1 Stem height and diameter

The effects of RNAi-suppression of *PopGT1* and *PopGT2* on the growth of six-month-old greenhouse-grown trees showed that tree height was not affected by RNAi-suppression of *PopGT1* and *PopGT2*. However, all transgenic lines showed significant reduction in stem diameter when compared to wild-type trees, except line RNAi-*PopGT1* A6 and RNAi-*PopGT2* line B8 (Table 14).

Table 14: Average plant height and diameter of RNAi-*PopGT1* and RNAi-*PopGT2* six-month-old transgenic hybrid poplar lines (n = 10). Values in brackets represent the standard deviation and statistically different values are in bold.

RNAi-*PopGT1*			RNAi-*PopGT2*		
Genotype	**Height (cm)**	**Diameter (mm)**	**Genotype**	**Height (cm)**	**Diameter (mm)**
WT	251.33 (9.06)	14.71 (0.56)	**WT**	248.33 (7.06)	14.65 (0.32)
A1	241.50 (7.04)	**13.20** (0.32)	**B1**	238.80 (12.54)	**12.29** (0.43)
A2	249.40 (4.85)	**12.46** (0.46)	**B2**	239.75 (5.25)	**12.69** (0.55)
A3	241.40 (5.67)	**12.95 (**0.82)	**B3**	260.00 (11.79)	**13.07** (0.62)
A4	239.80 (9.49)	**11.79** (0.63)	**B4**	246.40 (5.04)	**13.20** (0.49)
A5	240.17 (13.04)	**11.80** (0.78)	**B5**	246.80 (6.27)	**12.66** (0.87)
A6	246.00 (3.58)	14.27 (1.17)	**B6**	245.20 (5.38)	**12.63** (0.51)
A7	247.67 (6.96)	**12.48** (0.34)	**B7**	244.00 (6.36)	**13.43** (0.41)
A8	252.20 (5.00)	**13.18** (0.48)	**B8**	251.00 (3.78)	14.26 (1.59)
A9	247.00 (4.70)	13.06 (0.61)	**B11**	247.40 (4.16)	11.72 (0.36)

3.6.1.2 Leaf morphology

During the second growing season, it was apparent that the transgenic trees leaf morphology changed, consistently displaying smaller leaves in both RNAi-*PopGT1* and RNAi-*PopGT2* lines (Figure 30). In addition, the leaf edges were less serrated (fewer teeth) compared to the wild-type leaves (Figure 30). Transgenic lines (RNAi-*PopGT2* lines A2, A4, A5, A8, A9, and RNAi-

PopGT1 B2, B5, B7, B11) also did not drop their stipules, whereas the other lines did. However, the stipules of lines A3, A7, B1, and B4 were temporarily retained, displaying some growth and then abscised (Figure 31).

The transgenic trees also showed a propensity for increased branching compared to wild-type trees. Sylleptic branches were denser on the transgenic lines and their leaves were much smaller than on the wild-type branches. In addition, the sylleptic branches generally showed reduced apical dominance, as their lateral buds appeared to be growing continuously (Figure 32).

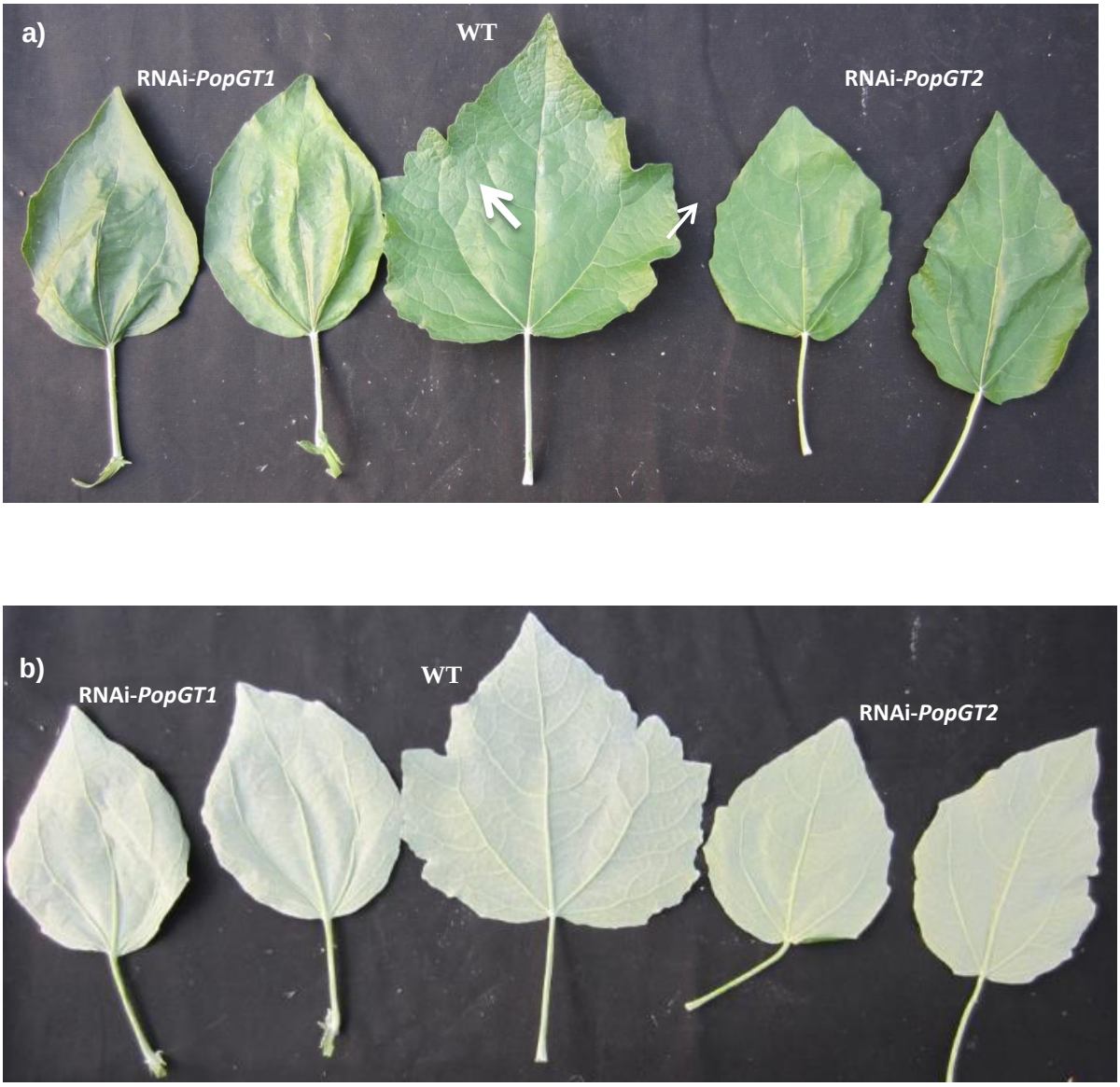


Figure 30: Leaf morphological changes of (a) adaxial RNAi-*PopGT1* and RNAi-*PopGT2* leaves and (b) abaxial transgenic poplar trees in the second growing season compared to the wild-type trees. Arrows indicate the wild-type lobes that were reduced in the transgenic leaves.

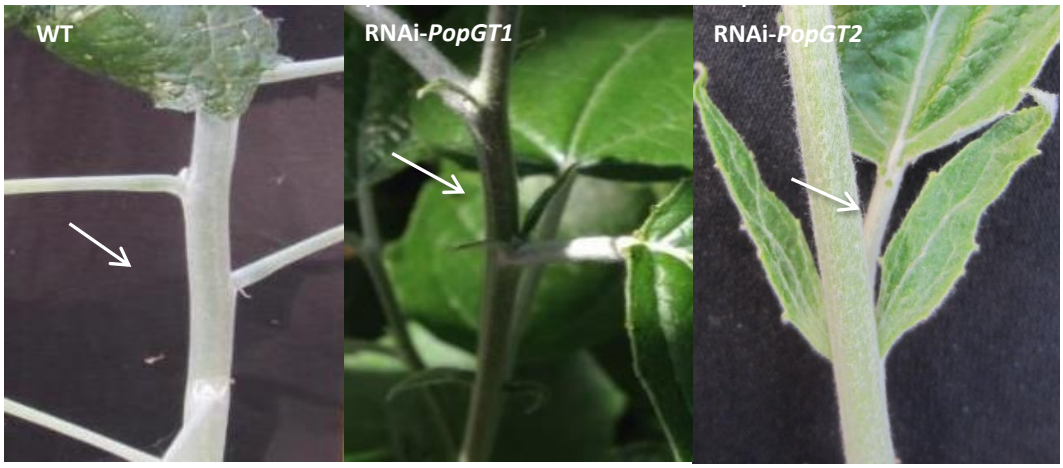


Figure 31: Stipule growth on transgenic poplar RNAi-*PopGT1* and RNAi-*PopGT2* poplar lines compared to the wild-type abscised stipules. Arrows indicate the stipule positions.

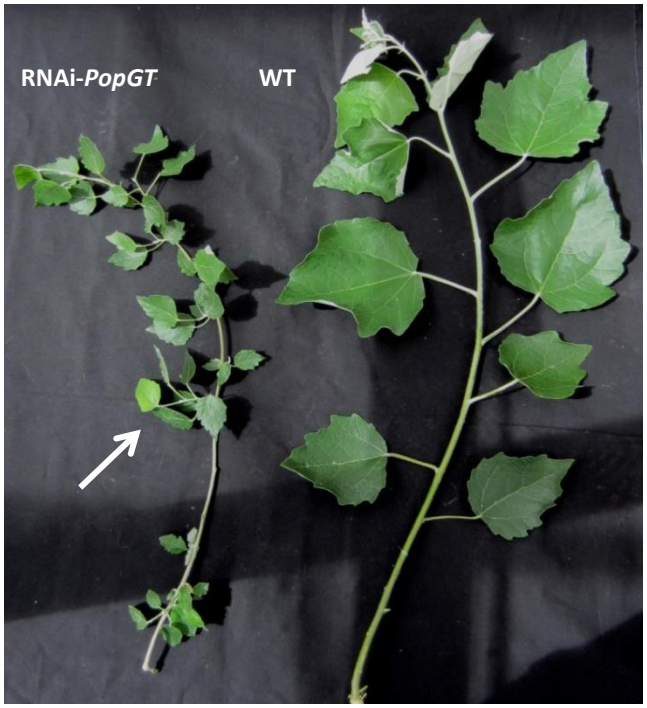


Figure 32: Sylleptic branches of three-year-old RNAi-mediated suppression transgenic trees (on the left) compared to wild-type trees (on the right). White arrows indicate the side branching and the small leaves.

3.6.2 Bud break

Trees were allowed to senesce, and the timing of bud break was recorded for both transgenic and wild-type trees for two growing seasons (2014 and 2015). The results clearly and consistently showed that the trees broke bud 2 weeks earlier (on average) in the transgenics compared to wild-type trees (Figures 33 a, b, and c). The first bud break in 2014 was in the last week of February, and in 2015, the bud break was in the first week of March (Figures 34 a, b).

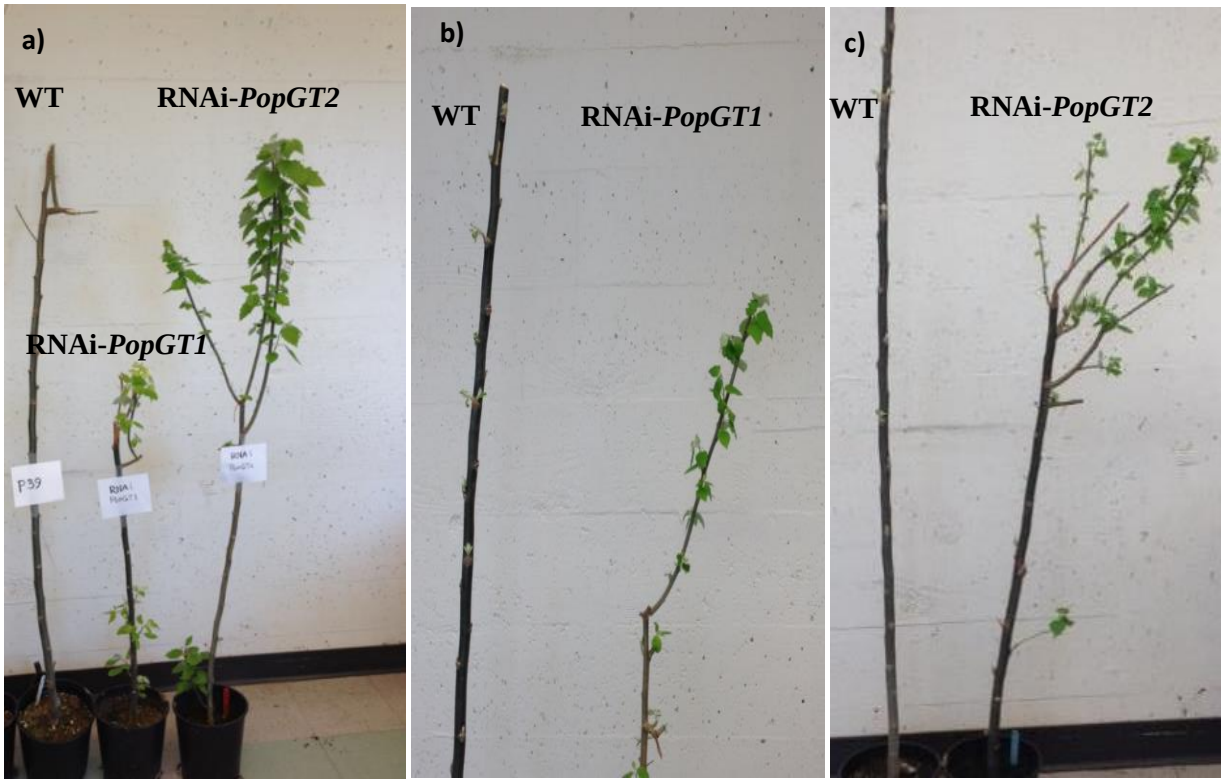


Figure 33: Early bud break of RNAi-*PopGT1* and RNAi-*PopGT2* transgenic lines compared to wild-type trees in (a) March 2014, and (b, c) March 2015.

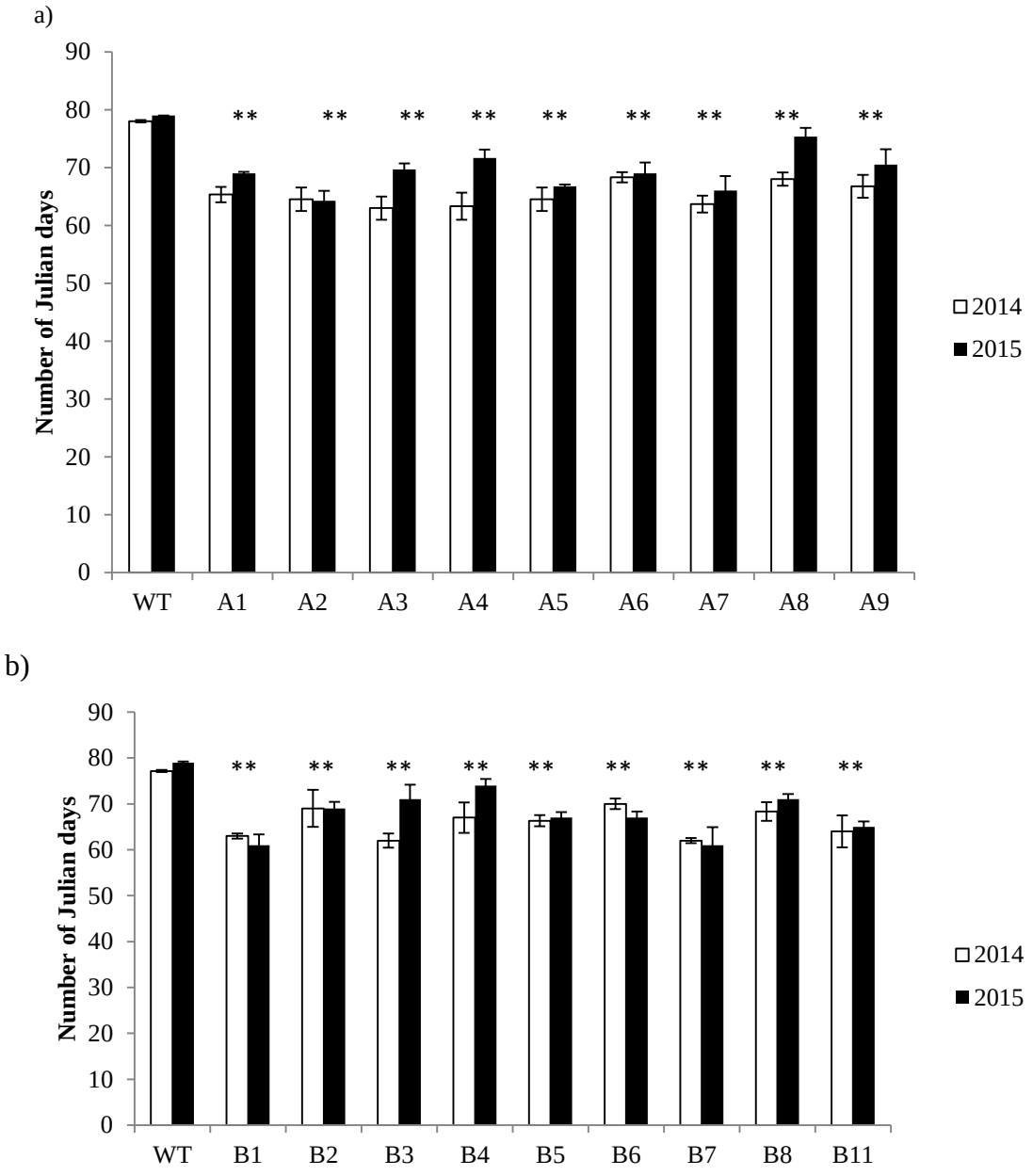


Figure 34: Timing of bud break in Julian days of (a) RNAi-*PopGT1*, and (b) RNAi-*PopGT2* poplar transgenic lines and wild-type controls in 2014 (white columns) and 2015 (black columns). Columns represent the average number of Julian days before bud break (January 1st is day 1, March 1st is day 60). Error bars represent standard error (n = 4). All transgenic trees broke their buds earlier than the wild-type controls. All differences were statistically significant at $\alpha = 0.05$ (asterisks depict the significant values).

3.6.3 Stem cross sectioning and microscopy

Stem cross-sectioning and histochemical staining showed unique structural differences between the transgenic and wild-type poplar trees in both xylem and phloem tissues. There were clear differences in vessel distribution patterns in transgenic lines, as the vessels appeared more clustered in the transgenic lines compared to the wild-type trees (Figure 35). No differences were apparent in the intensity of phloroglucinol/HCl staining of the xylem tissues. Interestingly, when poplar stems of all transgenic lines and wild-type trees where sectioned 10 cm above the root collar, it was found that the arrangement of the phloem fibre bundles was altered in all transgenic lines, where they formed 1 or 2 layers of fibre bundles compared to the 3 to 4 layers common in the wild-type trees. The phloem fibre bundles of all transgenic lines were more aggregated and were larger compared to wild-type trees (Figures 36 and 37). Histochemical staining of the phloem fibres also showed different cell wall deposition patterns (Figure 36), notably, the presence of an apparent G-layer. In addition, in some of the transgenic lines, some lignified cells were formed near the cortex in bundle-like structures. These tissues were only observed in the RNAi-*PopGT1* lines (A1, A2, A4, and A9), and RNAi-*PopGT2* lines (B1, B4, B3, and B11). These tissues showed different cell wall patterns than the normal phloem fibre cell walls. They were severely deformed and the lumen of the cells was hollow, lacking the aforementioned G-layer. Finally, the bark layer was thinner in the transgenic lines compared to the wild-type trees (Figure 37).

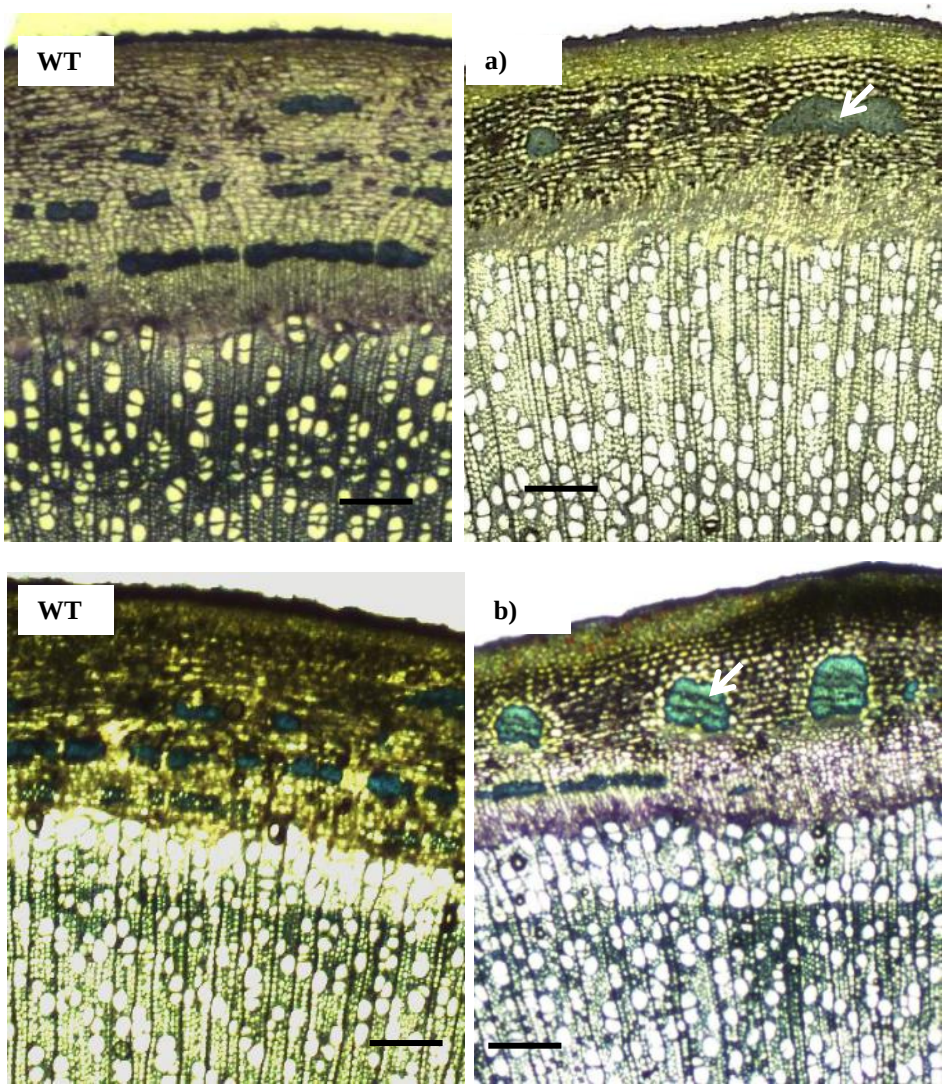


Figure 35: Stem cross sections of (a) RNAi-*PopGT1* and (b) RNAi-*PopGT2* transgenic lines compared to wild-type trees, showing abnormal phloem fibre bundles (white arrows). Scale bars are 200 µm.

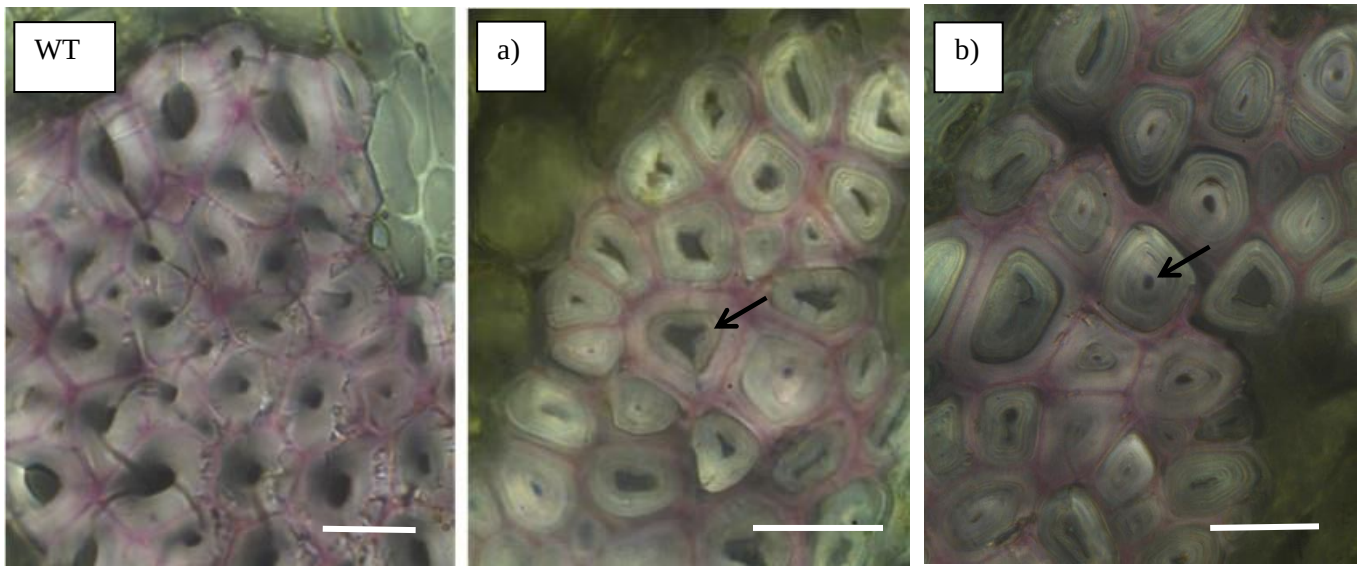


Figure 36: Phloem fibre of wild-type trees compared to (a) RNAi-*PopGT1* transgenic poplar and (b) RNAi-*PopGT2*. Arrows depict the G-layer structure in the fibre cell walls. Scale bars = 25µm.

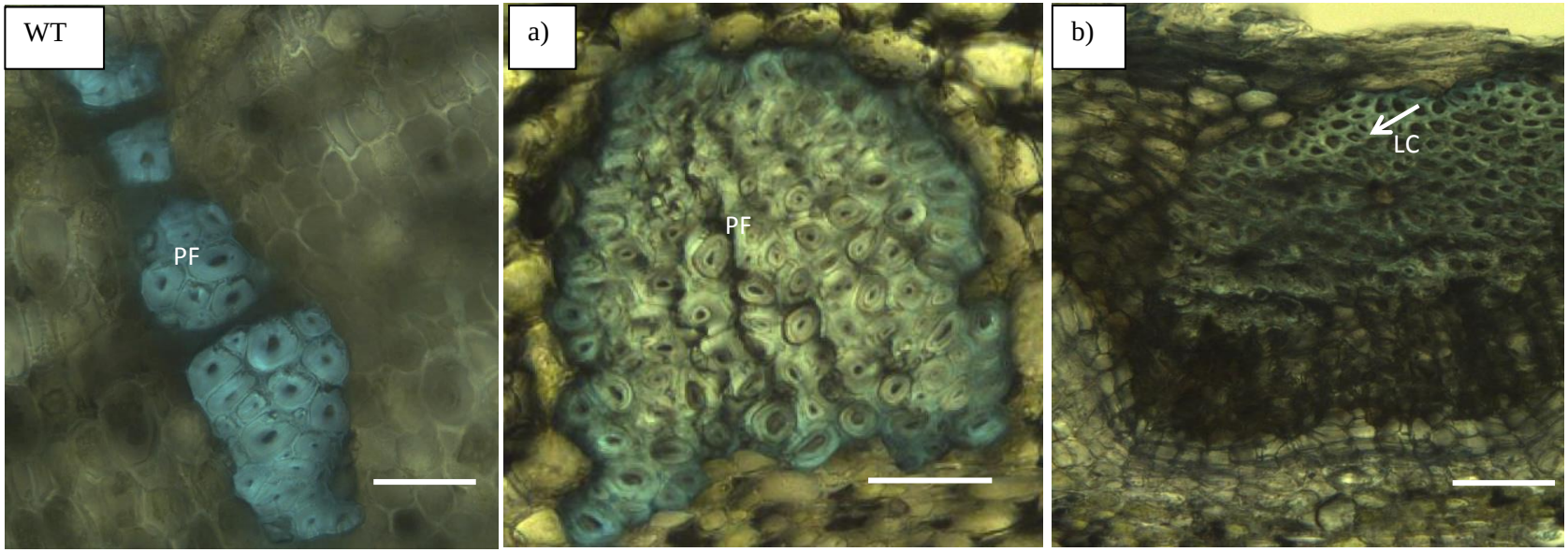


Figure 37: Example of altered phloem fibre (PF) morphology in (a) RNAi-*PopGT1* and (b) RNAi-*PopGT2* transgenic trees compared to wild-type trees. Lignified fibres near the cortex are depicted by white arrows. Scale bar = 50 µm.

3.6.4 Soluble carbohydrates analysis

Quantification of the total soluble carbohydrates in the mature leaves of six-month-old greenhouse-grown RNAi-*PopGT1* and RNAi-*PopGT2* transgenic and wild-type trees showed a reduction ranging between 3-31% in the total soluble sugars in both RNAi-suppression lines. All lines, except A6 and A8, showed a significant reduction in the total soluble carbohydrate contents, while only B1, B5, and B7 showed significant differences compared to wild-type controls (Figure 38).

In all RNAi-*PopGT1* transgenic lines, the sucrose concentration was significantly reduced (13-34% of the wild-type sucrose content), while glucose concentrations were variable. Fructose concentrations showed an increasing trend in all transgenic lines, but the increases were only significant in lines A2, A4, A6, and A8 (Figure 38 a)

In RNAi-*PopGT2* trees, sucrose concentration was also reduced (16-32%) in lines B1, B3, B4, B6, B7, and B8 compared to wild-type controls (Figure 38 b). The glucose content again showed variable concentrations compared to wild-type controls. Similarly, the RNAi-*PopGT2* transgenic lines showed variable averages of fructose concentrations. Transgenic lines B1, B3, B5, B7, and B11 showed a non-significant reduction in fructose content, while lines B2, B4, B6, and B8 showed a significant increase in fructose compared to wild-type controls.

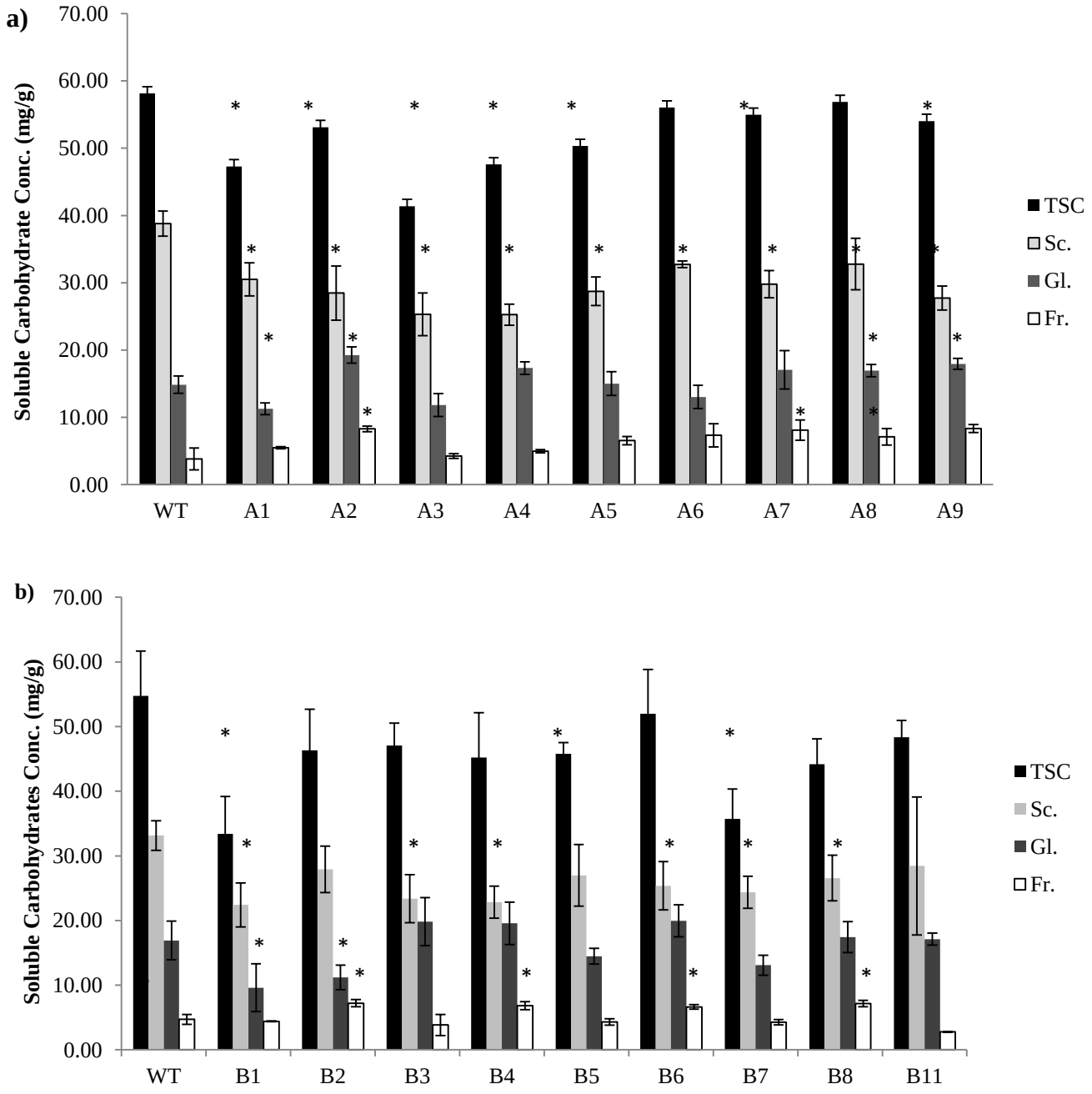


Figure 38: Average concentration of total soluble carbohydrates (TSC), sucrose (Sc), fructose (Fr), and glucose (Gl) in (a) RNAi-*PopGT1* transgenic lines, and (b) RNAi-*PopGT2*. Error bars represent standard error of means and asterisks depict the statistically significant values at $\alpha = 0.05$, using *t*-test (n = 5).

3.6.5 Wood properties

Since the transgenic trees showed a brittle-stem phenotype, wood properties were examined in the transgenic (RNAi-*PopGT1* and RNAi-*PopGT2*) lines, compared to wild-type trees. The traits studied included fibre quality (via FQA), microfibre angle (MFA), wood density by X-ray densitometry (bark-to-bark stem sections), and wood stiffness (Young's Modulus of Elasticity).

Initially, two lines from each suppression construct were screened to determine if any traits had been markedly influenced by down regulation of the two poplar UGT genes. Wood density and wood stiffness showed significant decrease when compared to wild-type samples, while MFA was unchanged (Table 1A, Appendix). Wood density ranged between 9-16% lower in all transgenic lines than the wild-type values (Figures 39 a and b). These reductions in wood density appeared to be associated with an increase in the wood fibre diameter, since fibre quality analysis showed significant increases in fibre diameter in all transgenic lines. The increase in fibre diameter was 8.5-18.7% and 8.5-20.7% in RNAi-*PopGT1* and RNAi-*PopGT2* lines, respectively (Tables 15 a and b), relative to wild-type trees.

Modifying the expression level of either *PopGT1* or *PopGT2* did not affect fibre length, as no significant differences were apparent between the transgenic lines and wild-type trees. Although a non-significant reduction was observed in the transgenic lines compared to the wild-type fibres. Vessel length and width did not change significantly between transgenic and wild-type trees (Tables 15 a and b).

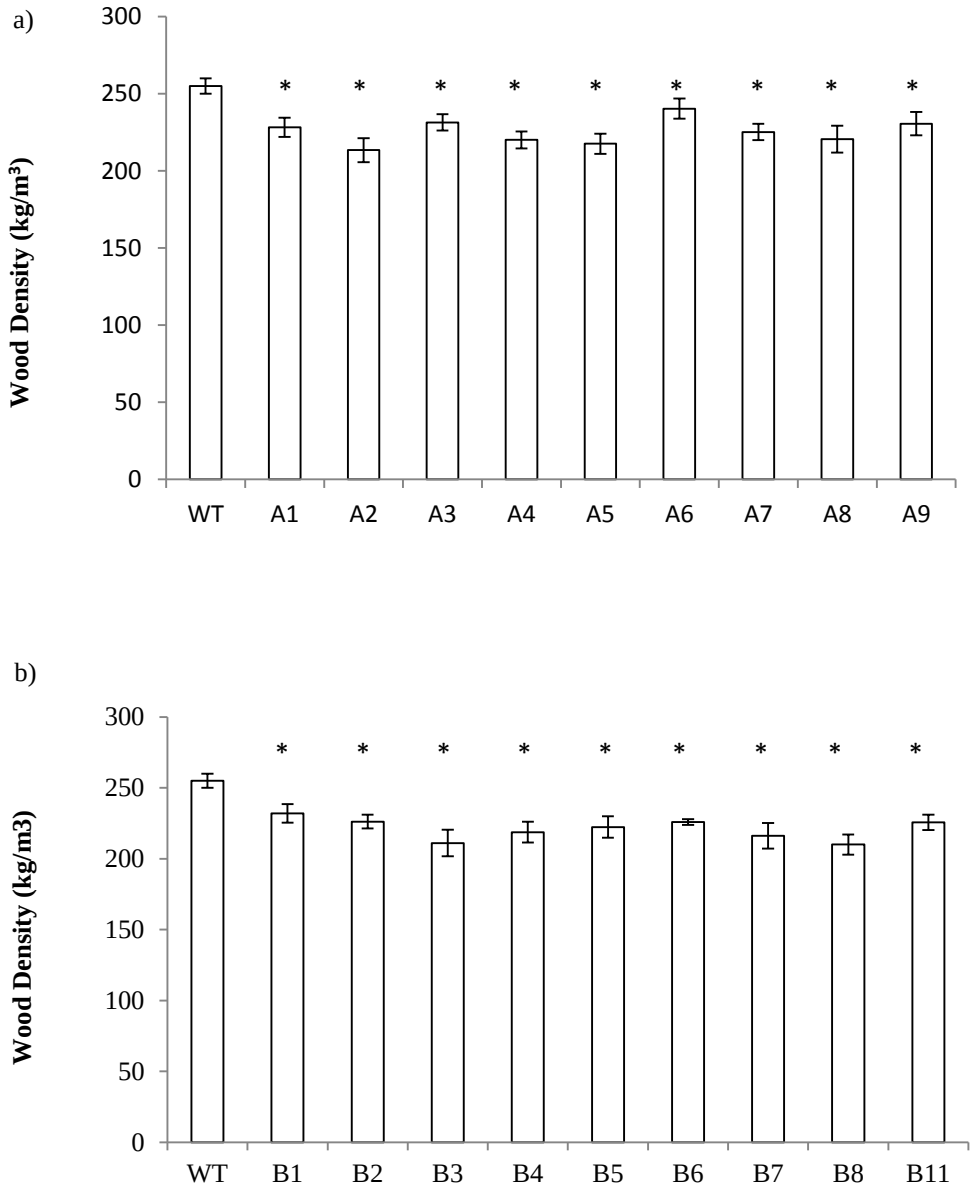


Figure 39: Wood density (kg/m^3) of (a) RNAi-*PopGT1* and (b) RNAi-*PopGT2* transgenic lines compared to wild-type trees. Error bars represent standard error (n = 5). Asterisks represent the significantly different means, using *t*-test at $\alpha = 0.05$.

Table 15: Fibre quality analysis of six-month-old greenhouse-grown transgenic poplar stems of (a) RNAi-*PopGT1* and (b) RNAi-*PopGT2* compared to the wild-type trees. Fibre length (mm), fibre width (µm), vessel length (mm), and vessel width (µm) were measured. Bold numbers depict the significantly different means using t-test at α = 0.05. Standard deviation values are represented in brackets (n = 5).

a)

	FL (mm)	FW (µm)	VL (mm)	VW (µm)
WT	0.48 (0.01)	23.07 (0.32)	0.63 (0.04)	106.13 (4.57)
A1	0.48 (0.01)	**26.45 (0.25)**	0.61 (0.03)	106.75 (1.94)
A2	0.47 (0.01)	**26.25 (0.11)**	0.56 (0.03)	104.30 (0.99)
A3	0.47 (0.01)	**27.85 (0.35)**	0.64 (0.01)	110.35 (0.53)
A4	0.48 (0.01)	**26.00 (0.18)**	0.61 (0.01)	105.95 (2.30)
A5	0.50 (0.01)	**27.35 (0.45**)	0.58 (0.02)	105.25 (2.46)
A6	0.48 (0.01)	24.02 (0.90**)**	0.67 (0.01)	106.00 (1.58)
A7	0.46 (0.01)	**27.13 (0.53)**	0.60 (0.02)	104.60 (2.98)
A8	0.47 (0.01)	**26.80 (0.46)**	0.59 (0.03)	106.00 (2.37)
A9	**0.44 (**0.02**)**	**25.37 (0.42)**	0.56 (0.01)	99.77 (2.83)

b)

	FL (mm)	FW (µm)	VL(mm)	VW (µm)
WT	0.48 (0.01)	23.07 (0.32)	0.63 (0.05)	106.13 (5.28)
B1	0.46 (0.01)	**26.60 (0.38)**	0.61 (0.01)	107.63(2.69)
B2	0.48 (0.01)	**27.43 (0.27)**	0.61 (0.03)	105.60 (1.22)
B3	0.48 (0.01)	**27.07 (0.22)**	0.61 (0.01)	106.13 (0.73)
B4	0.47 (0.01)	**27.05 (0.37)**	0.63 (0.01)	111.20 (1.55)
B5	0.47 (0.01)	**26.75 (0.20)**	0.64 (0.02)	105.50 (2.37)
B6	0.44 (0.01)	**27.25 (0.45)**	0.60 (0.01)	110.25 (2.58)
B7	**0.45 (0.01**)	**25.30 (0.37)**	0.62 (0.01)	**101.80 (1.55)**
B8	0.49 (0.01)	**25.60 (0.65)**	0.61 (0.02)	104.97 (2.23)
B11	0.47 (0.01)	**25.93 (0.55)**	0.59 (0.01)	105.27 (1.58)

In order to compare the changes in wood stiffness in both RNAi-*PopGT1* and RNAi-*PopGT2* transgenic lines, Young's Modulus of Elasticity (MOE) was measured for all genotypes, employing the three-point bending test. On average, the modulus values for the transgenic trees were significantly lower than the corresponding wild-type trees (Figures 40 a and b). The reduction in MOE and wood-density values appeared to have similar trends, since RNAi-

PopGT1 lines A2, A4, and A5 showed the lowest values in both wood density and MOE (Figure 40 a), while RNAi-*PopGT2* suppression lines B1, B6, and B11 also showed the lowest values in both traits (Figure 40 b).

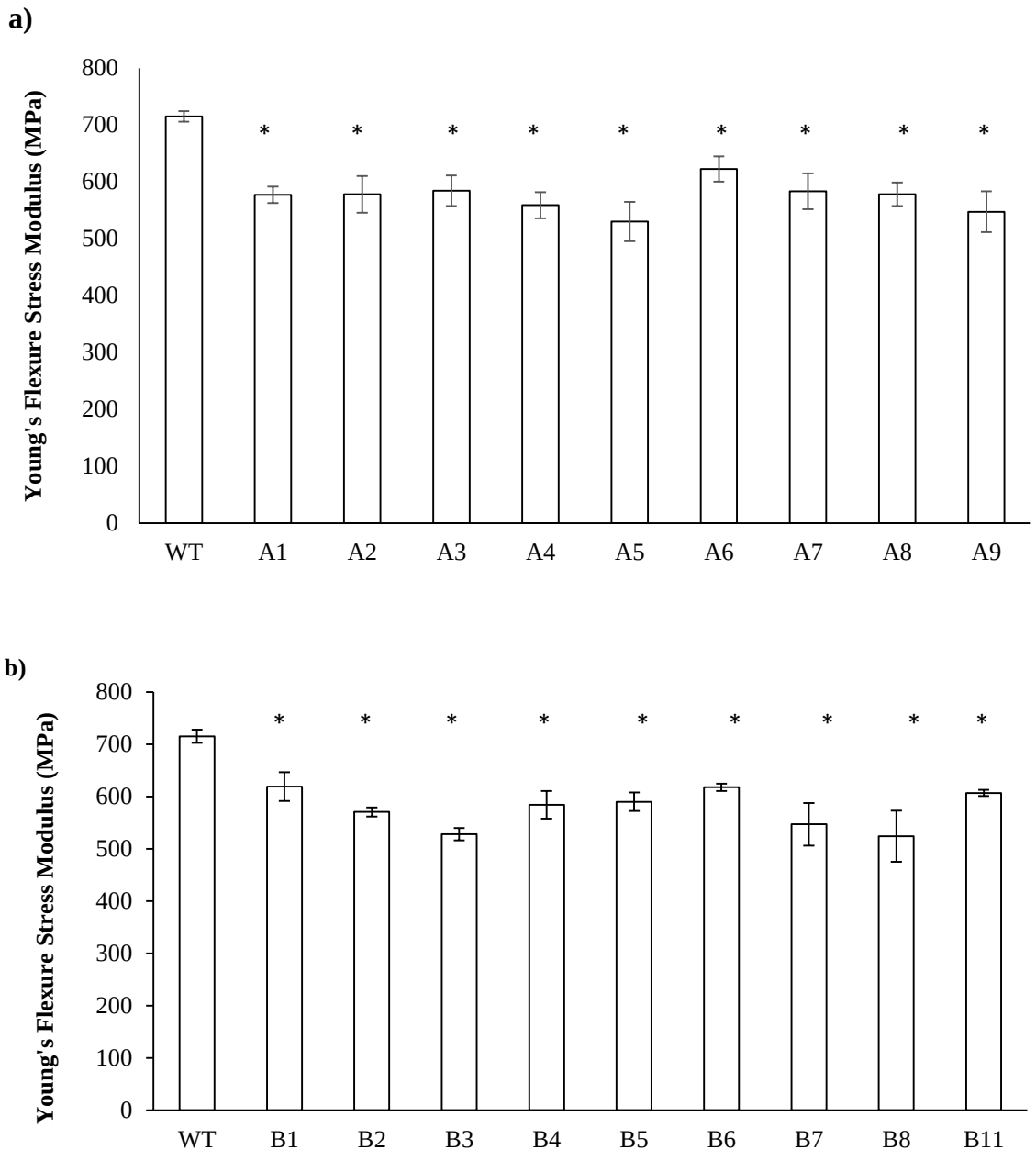


Figure 40: Young's modulus of Elasticity (a) RNAi-*PopGT1* and (b) RNAi-*PopGT2* lines compared to wild-type trees. Columns represent the average value of the modulus of three internodes tested per tree (n = 4), with a minimum of 6 stress tests per plant. Asterisks represent the significant values at α = 0.05. Error bars represent the standard error values.

3.6.6 Cell wall chemical analysis

Cell wall composition analysis was carried out for wild-type, RNAi-*PopGT1*, and RNAi-*PopGT2* trees. A significant increase in Klason lignin content was observed in several transgenic lines compared to wild-type controls (Tables 16 a and b). Total lignin content in the wild type was on average 21.5%, while the lignin content in the transgenic lines ranged between 22.7 and 24%, representing a 1.3 to 2.5% increase in cell wall lignin accumulation. In contrast, significant reductions (5-6%) in structural glucose were observed between wild-type trees and the transgenic lines. Hemicellulose-derived sugars, on the other hand, did not appear to change significantly (Tables 16 a and b).

Table 16: **Cell wall chemical composition of (a) RNAi-*PopGT1* and (b) RNAi-*PopGT2* transgenic lines compared to wild-type trees. Bold values represent the lines with significant differences, using *t*-test at α = 0.05 and the standard deviation values are shown in brackets (n = 4).**

a)

Genotype	% Glucose	% Hemicellulose-derived sugars	% Lignin
WT	50.70 (1.08)	19.08 (1.14)	21.56 (0.74)
A1	**45.92 (1.67)**	20.57 (1.26)	**23.12 (0.34)**
A2	**46.52 (1.08)**	19.32 (1.06)	**22.95 (0.54)**
A3	**45.71 (1.58)**	20.36 (0.87)	**23.56 (0.26)**
A4	**45.97 (1.09)**	19.10 (1.19)	23.09 (1.16)
A5	**44.64 (0.85)**	19.65 (0.52)	**24.31 (1.31)**
A6	48.87 (2.28)	19.18 (0.82)	23.02 (3.20)
A7	**44.74 (0.59)**	19.20 (0.21)	**24.06 (1.06)**
A8	46.72 (2.05)	18.67 (1.95)	21.60 (1.50)
A9	**45.49 (1.19)**	18.49 (0.90)	**23.43 (0.39)**

b)

Genotype	% Glucose	% Hemicellulose-derived sugars	% Lignin
WT	50.70 (1.08)	19.08 (1.14)	21.56 (0.74)
B1	**47.45 (1.09)**	18.72 (1.05)	**23.03 (1.02)**
B2	**48.51 (1.02)**	19.04 (1.22)	22.71 (1.17)
B3	**46.75 (0.85)**	19.23 (0.79)	**23.32 (0.86)**
B4	**47.41(1.28)**	18.53 (1.20)	**24.14** (0.55)
B5	**46.24 (1.29)**	19.68 (0.95)	**23.51 (0.75)**
B6	48.96 (1.09)	19.76 (1.10)	22.05 (0.42)
B7	**44.91 (2.39)**	19.62 (0.78)	**23.17** (0.60)
B8	**45.65 (1.39)**	18.68 (0.88)	**24.25 (1.17)**
B11	**46.76 (0.58)**	19.32 (0.62)	**23.57** (0.64)

In order to test if the reduction in the glucose content was a result of reduced cellulose content in the cell walls, α-cellulose content was measured directly in two representative lines from each

RNAi-suppression line. The results confirm that the transgenic lines showed a significant reduction in α-cellulose content compared to the wild-type trees. The reduction in the α-cellulose content of RNAi-*PopGT2* trees was greater than the reduction seen in RNAi-*PopGT1* trees (on average) when the two constructs were compared to each other. The α-cellulose content of the RNAi-*PopGT1* lines A5, and A9 was 28.7 and 29.2%, respectively, while the levels in the RNAi-*PopGT2* lines B4 and B11 were 25.4, and 25.8%, respectively (Tables 17 a and b). The results suggest that the reduction in structural glucose is the result of impaired cellulose accumulation in the cell walls of the transgenic lines.

Table 17: Average (%) cell wall α-cellulose and the S/G monomers ratio of two representative (a) RNAi-*PopGT1* transgenic poplar lines (A5 and A9) and (b) RNAi-*PopGT2* (B4 and B11). Bold numbers represent statistically significant differences between transgenic line and wild-type trees, using *t*-test at α = 0.05 (n = 3). Standard error values are shown in brackets.

a)

Genotype	(%) Alpha Cellulose	S/G
WT	34.01 (0.80)	2.62 (0.07)
A5	**28.75** (0.80)	2.67 (0.18)
A9	**29.23 (2.19)**	2.70 (0.05)

b)

Genotype	(%) Alpha Cellulose	S/G
WT	34.01 (0.80)	2.62 (0.07)
B4	**25.42 (0.90)**	2.58 (0.09)
B11	**25.89 (1.13)**	2.61 (0.07)

3.6.7 Bark tissue analysis of transgenic poplar trees

3.6.7.1 Cell wall analysis of bark tissue

The chemistry (lignin and structural-carbohydrate content) of the phloem tissue in the transgenic trees was assessed in addition to the evaluation of their xylem cell walls. A significant reduction

in both cell wall hemicellulose-derived carbohydrates and total lignin content was apparent in many lines of both RNAi-*PopGT1* and RNAi-*PopGT2* transgenic trees. In the *PopGT1* RNAi-suppression lines, all lines showed a significant (15-20%) reduction in hemicellulose content. All lines had reduced mannose and rhamnose, except line 7, where the reduction in rhamnose content was not significant. Lines A1-A4 also showed a significant reduction in xylose content, and the remaining lines (A5-A9) contained less xylose than the wild-type controls, but the reduction was not significant. In comparison to the wild-type trees, galactose content appeared to increase slightly but was not significantly different from the wild-type controls. In addition, lines A1-A6 showed a significant increase in glucose content, while lines A7-A9 showed reductions, but not significantly. Several lines (A1, A3, and A5-A9) showed significant reductions in total lignin content, with values ranged between 6-21% of the wild-type levels. The remaining lines (A2 and A4) contained less lignin than the wild-type samples (6% and 15% of the wild-type content), but were not statistically significant (Tables 18 a and b).

In the RNAi-*PopGT2* transgenic lines, similar trends in both structural sugar and lignin content were observed. The transgenic lines showed a significant reduction in hemicellulose-derived sugars, with 5-18% reduction, except for galactose, which increased in the transgenic lines. Line RNAi-*PopGT2* B3 showed a non-significant reduction (5.3%) in hemicellulose content compared to wild-type controls. All transgenic lines showed a significant reduction in mannose content (25-45%) compared to wild-type controls. In addition, the galactose content increased significantly (8-55%) compared to wild-type, except with lines RNAi-*PopGT2* B3 and B7, which showed non-significant increases (8% and 16%, respectively). In all transgenic lines, rhamnose, xylose and arabinose showed a reduction in the structural sugar contents. The reduction in xylose was not significantly different in transgenic lines when compared to wild-type trees. Four lines

showed a significant reduction in arabinose (RNAi-*PopGT2* B1, B2, B4, and B6). Rhamnose content was significantly reduced in RNAi-*PopGT2* B1, B2, and B6, by 5-19% compared to wild-type trees. Lignin content was significantly reduced in all transgenic lines, except lines RNAi-*PopGT1* A2 and A5, which showed a reducing trend (5 and 7%, respectively) (Tables 18 a and b). No significant differences in lignin monomer ratios were apparent between the transgenic lines and the wild-type trees (Table 18 b).

Table 18: Structural cell wall carbohydrates, lignin content (%), and syringyl/guaiacyl (S/G) ratio of 6-month-old bark tissues of (a) RNAi-*PopGT1* and (b) RNAi-*PopGT2* transgenic lines compared to wild-type poplar trees. Bold values represent the statistically significant values using *t*-test at $\alpha = 0.05$. Standard deviation is presented in brackets (n = 3).

a)	Structural Carbohydrates %						Lignin	Monolignols
	Arabinose	**Rhamnose**	**Galactose**	**Glucose**	**Xylose**	**Mannose**	**Total lignin**	**S/G**
WT	3.88 (0.58)	1.01 (0.05)	2.80 (0.15)	23.26 (0.64)	5.90 (0.26)	2.03 (0.09)	18.55 (1.38)	1.97 (0.18)
A1	2.81 (0.22)	**0.70 (**0.10)	2.88 (0.40)	**27.52 (**0.86)	**4.91 (**0.67)	**1.27 (**0.11)	**15.09 (0.69)**	1.78 (0.40)
A2	3.17 (0.42)	**0.68 (**0.09)	2.86 (0.29)	**25.28 (**0.56)	**4.41 (**0.32)	**1.22 (**0.06)	17.50 (2.61)	1.86 (0.22)
A3	3.03 (0.51)	**0.68 (0.13)**	2.97 (0.28)	**26.10 (1.72)**	**4.05 (**1.19)	**1.16 (**0.11)	**15.32 (1.11)**	1.65 (0.63)
A4	**2.83 (0.48)**	**0.80 (0.10)**	**2.98 (0.10)**	**25.18 (**1.18)	**4.56 (**1.71)	**1.20 (**0.24)	15.66 (1.65)	1.71 (0.29)
A5	2.81 (0.59)	**0.88 (0.02)**	2.77 (0.41)	**25.99 (**1.91)	5.15 (1.73)	**1.11 (**0.28)	**14.61 (1.43)**	1.88 (0.50)
A6	3.16 (0.21)	**0.95 (0.04)**	2.97 (0.44)	**24.95 (**2.26)	4.05 (1.63)	**1.17 (**0.55)	**14.94 (1.75)**	1.85 (0.41)
A7	3.40 (0.49)	0.90 (0.06)	3.17 (0.32)	24.00 (2.19)	3.99 (1.11)	**1.38 (0.35)**	**16.09 (0.75)**	1.72 (0.59)
A8	3.16 (0.54)	**0.85 (0.07)**	3.24 (0.35)	25.62 (2.21)	4.47 (1.06)	**1.54 (**0.34)	**15.47 (1.25)**	1.76 (0.37)
A9	3.19 (0.61)	**0.86 (0.11)**	3.07 (0.30)	24.83 (1.65)	4.52 (0.07)	**1.56 (**0.12)	**14.75 (2.16)**	1.91 (0.16)

b)	Structural Carbohydrates %						Lignin	Monolignols
	Arabinose	**Rhamnose**	**Galactose**	**Glucose**	**Xylose**	**Mannose**	**Total lignin**	**S/G**
WT	3.76 (0.56)	1.00 (0.07)	2.84 (0.21)	23.37 (0.75)	5.47 (1.06)	2.55 (0.34)	18.48 (1.41)	1.97 (0.18)
B1	**2.53 (0.48)**	**0.84 (0.04)**	**3.26 (0.13)**	**28.52 (1.25)**	4.71 (0.50)	**1.73 (0.11)**	**15.07 (1.91)**	2.19 (0.22)
B2	**2.26 (0.25)**	**0.85 (0.04)**	**3.30 (0.19**)	**29.14 (1.34)**	5.19 (0.73)	**1.74 (0.09)**	**15.26 (1.05)**	2.20 (0.16)
B3	3.56 (0.41)	0.97 (0.06)	3.07 (0.51)	24.65 (1.36)	5.29 (0.42)	2.03 (0.21)	17.20 (2.14)	1.97 (0.55)
B4	**3.05** (0.13)	0.99 (0.01)	**3.29 (0.10)**	**25.64 (0.65)**	5.11 (0.46)	**1.94 (0.06)**	**16.33 (0.95)**	2.22 (0.16)
B5	3.31 (0.55)	0.95 (0.05)	**3.18 (0.15)**	**26.53 (1.02**)	4.70 (0.61)	1.93 (0.33)	16.86 (2.40**)**	**2.48 (0.08)**
B6	**2.92 (0.68)**	**0.91 (0.02)**	**3.42 (0.31**)	**27.03 (0.85**)	4.70 (0.34)	**1.83 (0.25)**	**15.92 (1.71)**	2.03 (0.54)
B7	3.05 (0.35**)**	0.99 (0.08)	3.29 (0.42)	**25.44 (0.66)**	5.11 (0.58)	**1.94 (0.15**)	**16.33 (1.08)**	1.86 (0.63)
B8	3.56 (0.24)	0.91 (0.03)	**4.43 (0.50)**	24.52 (2.55)	3.86 (1.61)	**2.06 (0.10)**	**16.25 (1.14)**	2.12 (0.88)
B11	3.37 (0.59)	0.95 (0.10)	**3.22 (0.13)**	**26.88 (0.43)**	4.29 (0.75)	**1.93 (0.21)**	17.62 (2.77)	2.07 (0.12)

3.6.8 HPLC analysis of phenolic metabolites in transgenic-poplar trees

3.6.8.1 Mature leaf metabolites

HPLC analysis was carried out on the phenolic extracts of paired mature leaves of two transgenic RNAi-suppression genotypes (RNAi-*PopGT1* and RNAi-*PopGT2*) and the wild-type controls, using either total phenolic extracts or acid hydrolyzed metabolites. The results showed unique metabolic profiles for both transgenic genotypes compared to the wild-type controls (Figures 41-44). Overall, there was reduction in the relative abundance of many phenolic metabolites in both RNAi-*PopGT1* and RNAi-*PopGT2* transgenic lines (Figures 41 and 42). Statistical analysis of the normalized peak areas (Tables 3A and 4A, Appendix) indicated that the levels of several compounds were significantly different.

RNAi-*PopGT1* transgenic lines showed significant changes in peaks with the following retention times: 23.35, 24.69, 25.78, 27.14, 27.85, 28.13, 28.47, 30.62, 31.40, 31.71, 32.73, 33.96, 36.68, 37.15, 39.84, and 40.43 minutes. However, all other phenolic metabolites showed variability (Figure 41 and Table 3A, Appendix).

RNAi-*PopGT2* lines showed similar trends in the associated metabolic profiles. The peaks appearing at the following retention times: 16.46, 17.71, 26.08, 28.9, 30.82, 31.08, 31.46, 32.99, 33.51, 34.43, 34.91, 35.52, 36.16, 36.32, 36.68, 37.13, 39.43, 39.79, 46.78, 47.3, 53.12, and 57 minutes, were significantly changed compared to the wild-type trees. However, some other metabolites showed variable and inconsistent peak areas (Figure 42 and Table 4A, Appendix).

In both RNAi-suppression lines, RNAi-*PopGT1* and RNAi-*PopGT2*, SA (RT = 31.81) was detected and its concentration showed a significant reduction in five (RNAi-*PopGT1*: A2, A3, A4, A5 and A9) of six lines tested. An unidentified peak (RT = 13.5) also showed a significant reduction in four transgenic lines (RNAi-*PopGT1*: A2, A3, A4, and A5) among the six lines analysed (Figures 43 and 44). The unidentified peak (17.45) showed increasing trend in the transgenic RNAi-*PopGT1* lines compared to the wild-type controls, while the apparent reduction in the peak area of another unidentified peak (RT = 26.72) was not significant. Five transgenic lines (RNAi-*PopGT1* A2, A3, A4, A5, and A9) showed a significant reduction in the peak area of an unidentified peak at RT = 46.5-47.25 minutes.

RNAi-*PopGT2* lines showed a non-significant reduction in the normalized peak area of SA compared to the wild-type controls, except in line RNAi-*PopGT2* B11, which showed a significant reduction. In addition, the unidentified peak (RT = 13.5) showed significant reduction in three of six transgenic lines (RNAi-*PopGT2* B2, B4, and B5) examined, while the remaining lines showed non-significant reductions. Two transgenic lines (RNAi-*PopGT2* B2 and B11) showed significant increases in the average peak areas of the peak appearing at 17.27 minutes (RT = 17.27), with reducing trends in the remaining lines compared to the wild-type controls. The chromatographic peak appearing at RT = 26.56 minutes consistently showed non-significant reduction in the transgenic lines (RNAi-*PopGT2*).

Together, the data indicated that there was some inconsistency and variability in the metabolic profile of the trees tested. This variability is reflected in the statistical power, and therefore it is difficult to draw firm conclusions to describe the substrate specificity of the enzyme in trees. Total metabolite analysis showed no detectable amount of free SA in the trees analyzed.

However, there was a significant reduction in the conjugated SA concentration in some of the transgenic lines compared to the wild-type control after acid hydrolysis treatment (Figures 43 and 44). It is unclear if this reduction was the result of reducing SA glucoside conjugates or because of a reduction in other salicylate metabolites.

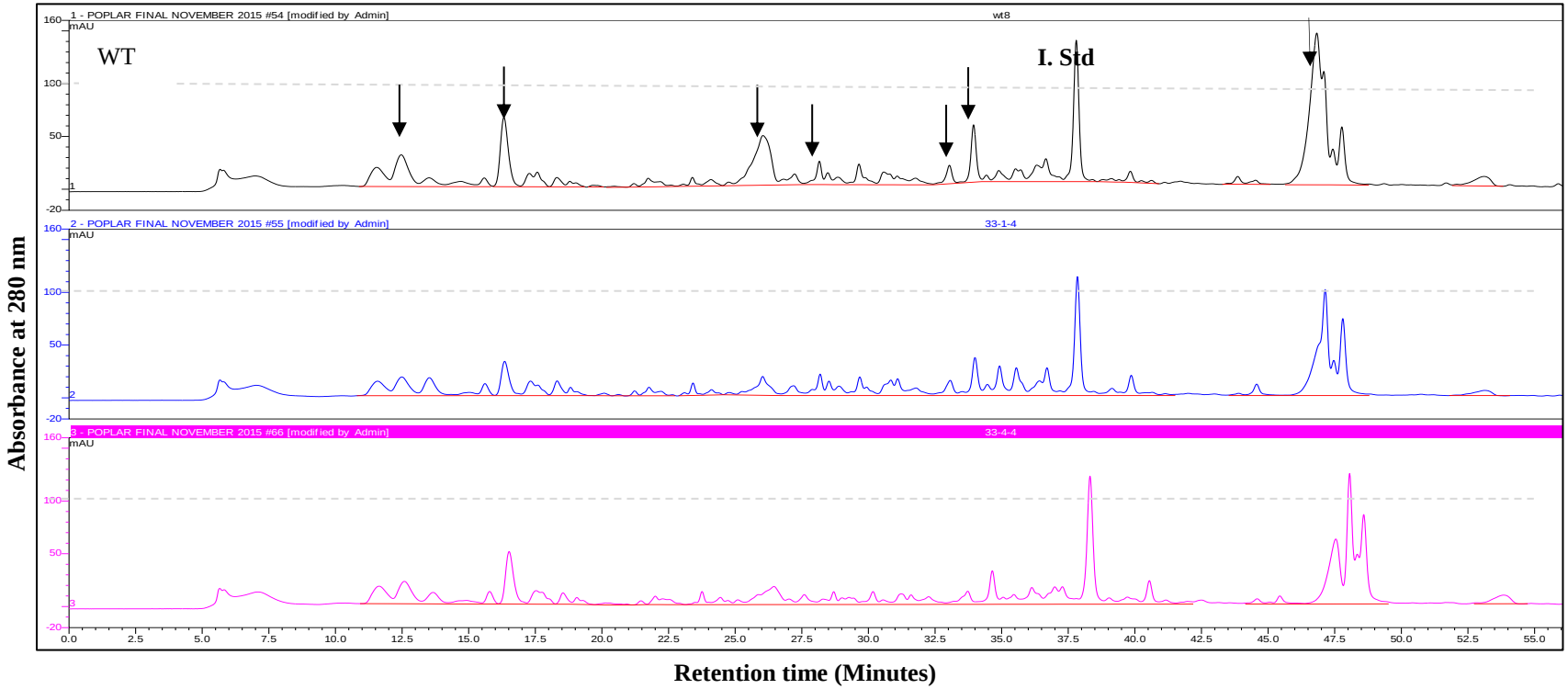


Figure 41: HPLC traces of total phenolic metabolites extracted from RNAi-*PopGT1* hybrid poplar mature leaves, using 80% methanol. Arrows indicate some of the unidentified compounds that showed reduced concentrations in the transgenic trees compared to the wild-type controls. Peaks represent absorbance at 280 nm. I. Std is the internal standard.

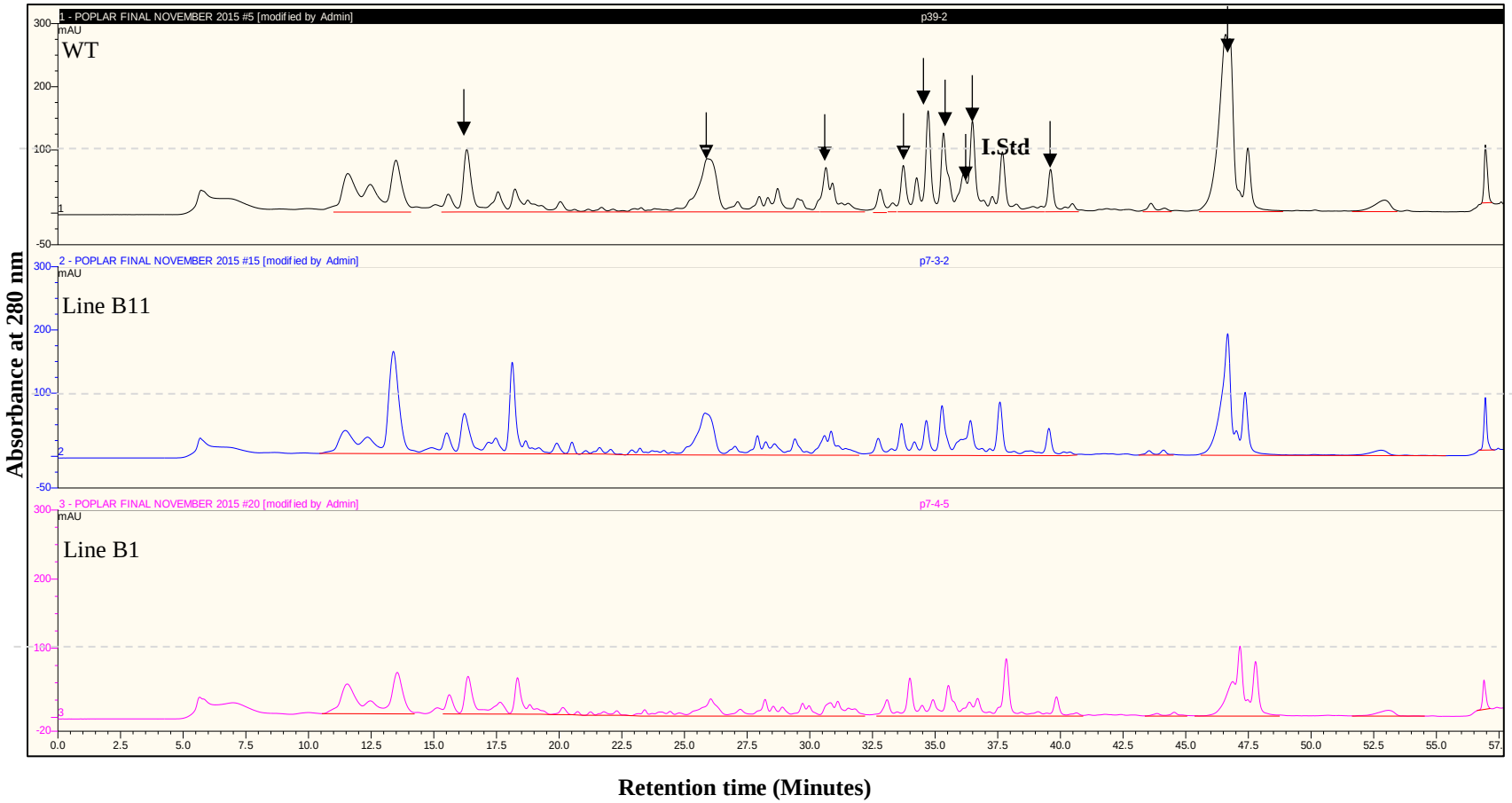


Figure 42: **HPLC traces of total phenolic metabolites extracted from RNAi-*PopGT2* hybrid poplar mature leaves using 80% methanol. Arrows indicate some compounds that showed reduced concentrations in the transgenic lines compared to the wild-type controls. Peaks represent absorbance at 280 nm. I. Std is the internal standard.**

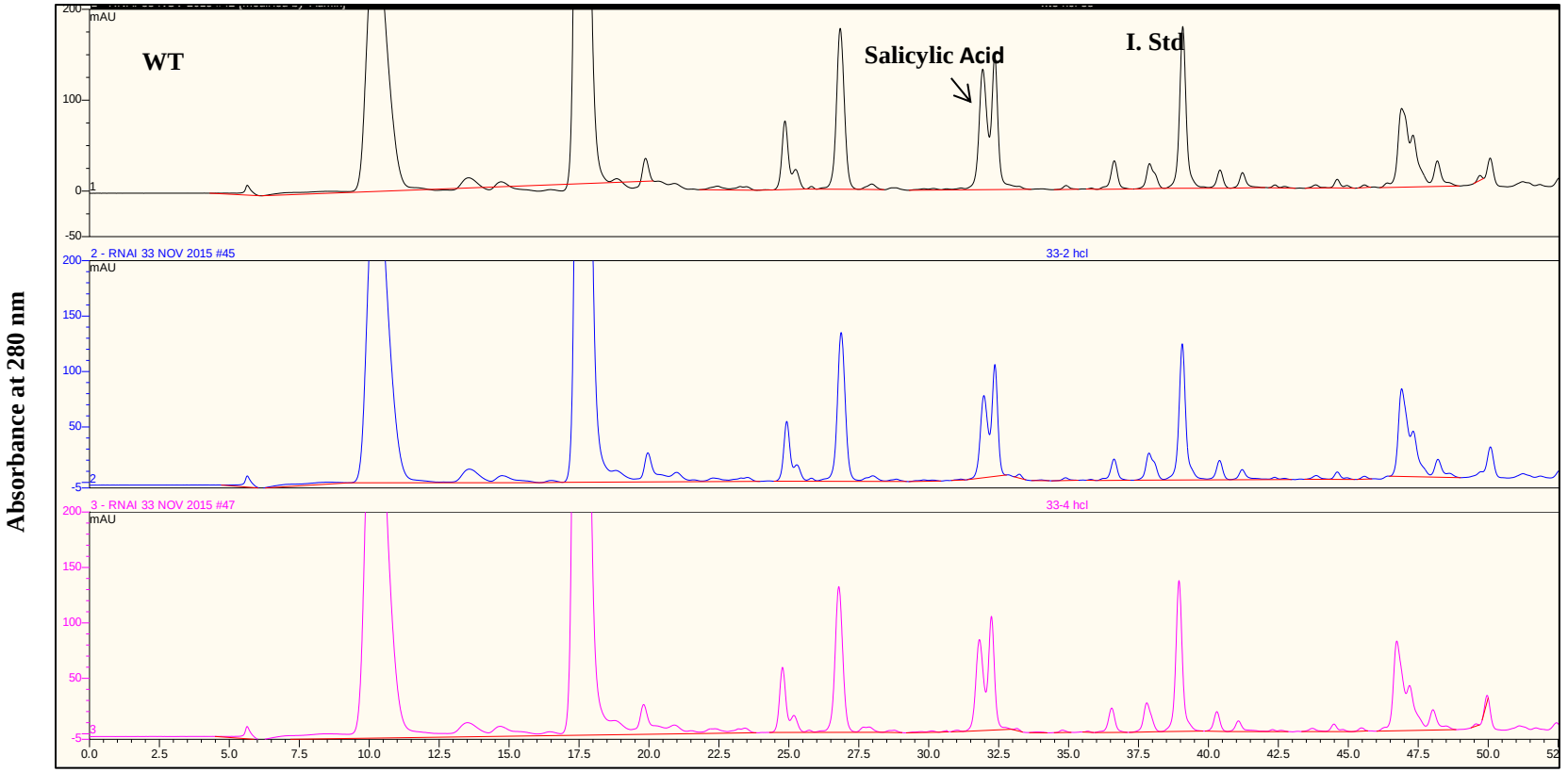


Figure 43: HPLC traces of the acid-hydrolyzed phenolic metabolites of the RNAi-*PopGT1* hybrid poplar mature leaves, using 80% methanol. Arrows indicate some compounds that showed reduced concentrations in the transgenic trees compared to the wild-type controls. Peaks represent absorbance at 280 nm. I. Std is the internal standard.

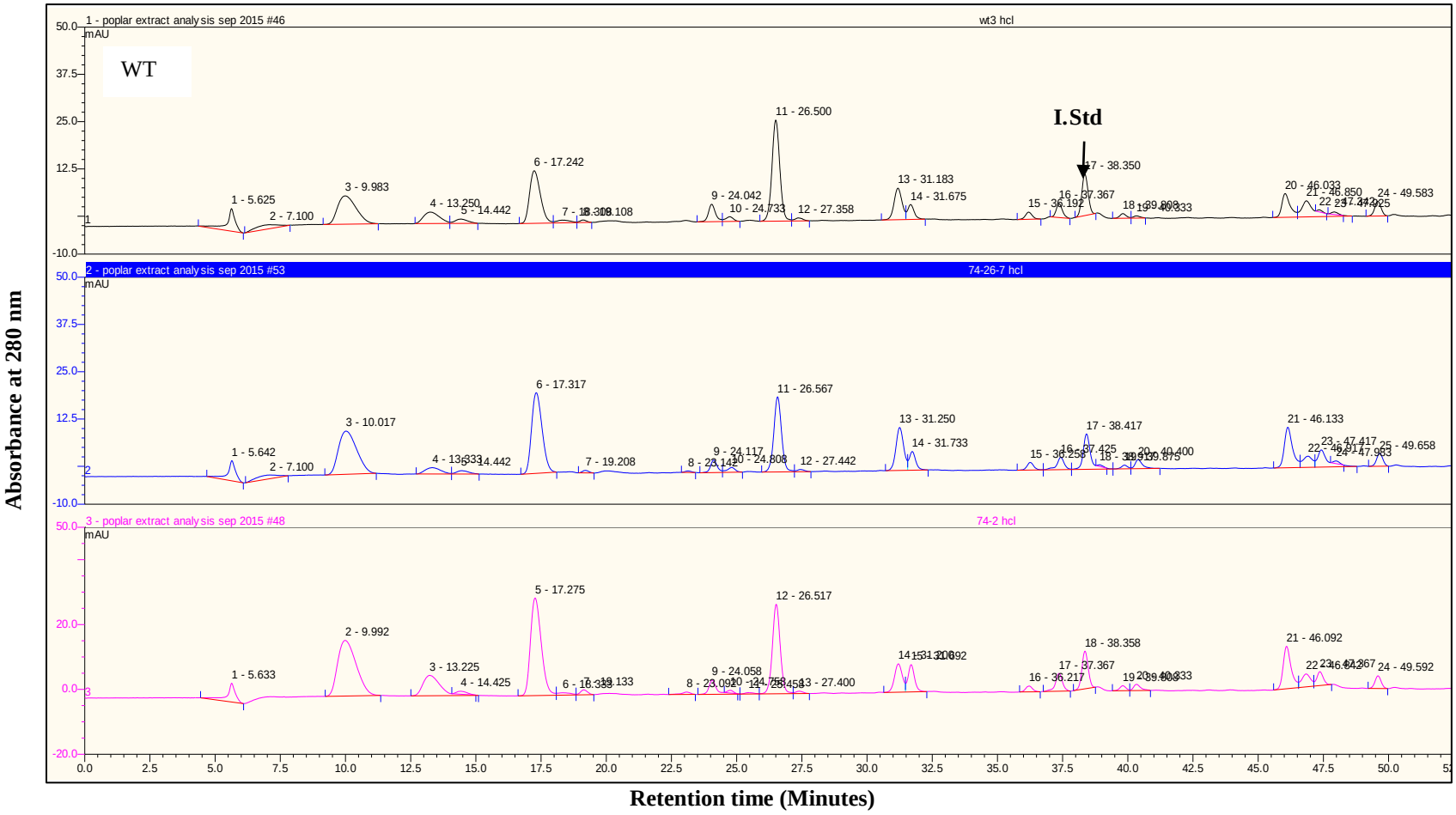


Figure 44: HPLC traces of the acid-hydrolyzed phenolic metabolites of the RNAi-*PopGT2* hybrid poplar mature leaves, using 80% methanol. Arrows indicating some peaks with reduced concentrations in the transgenic trees compared to the wild-type controls. Peaks represent absorbance at 280 nm. I. Std is the internal standard.

3.7 Bark tissues

Lyophilized bark tissues were analyzed via reversed-phase HPLC and the separated metabolite profiles were compared between transgenic lines and wild-type trees. Chromatographic peaks at RT = 20.9, 23.68, 28.82, and 31.69 minutes showed increased peak area in transgenic trees compared to wild type, while, peaks at RT = 23.8 and 28.54 minutes were reduced in the transgenic lines (Figure 45). The chromatographic peak at RT = 28.54 minutes was further analyzed, as this compound was found to be a glycoside that was hydrolyzable by almond glucosidase. The other peak was not hydrolyzable when treated with almond β-glucosidase. The molecular mass of the glucoside was 406 g/mol and NMR analysis of the corresponding fraction indicated that this glycoside is salireposide (Figures 45 and 46 and Table A2, Appendix). NMR analysis of one of the reduced compounds in both *PopGT1* and *PopGT2* RNAi-suppressed trees was identified as salireposide, which is an ester glucoside of benzoic acid and 2,5-dihydroxylbenzyl alcohol (Figure 46).

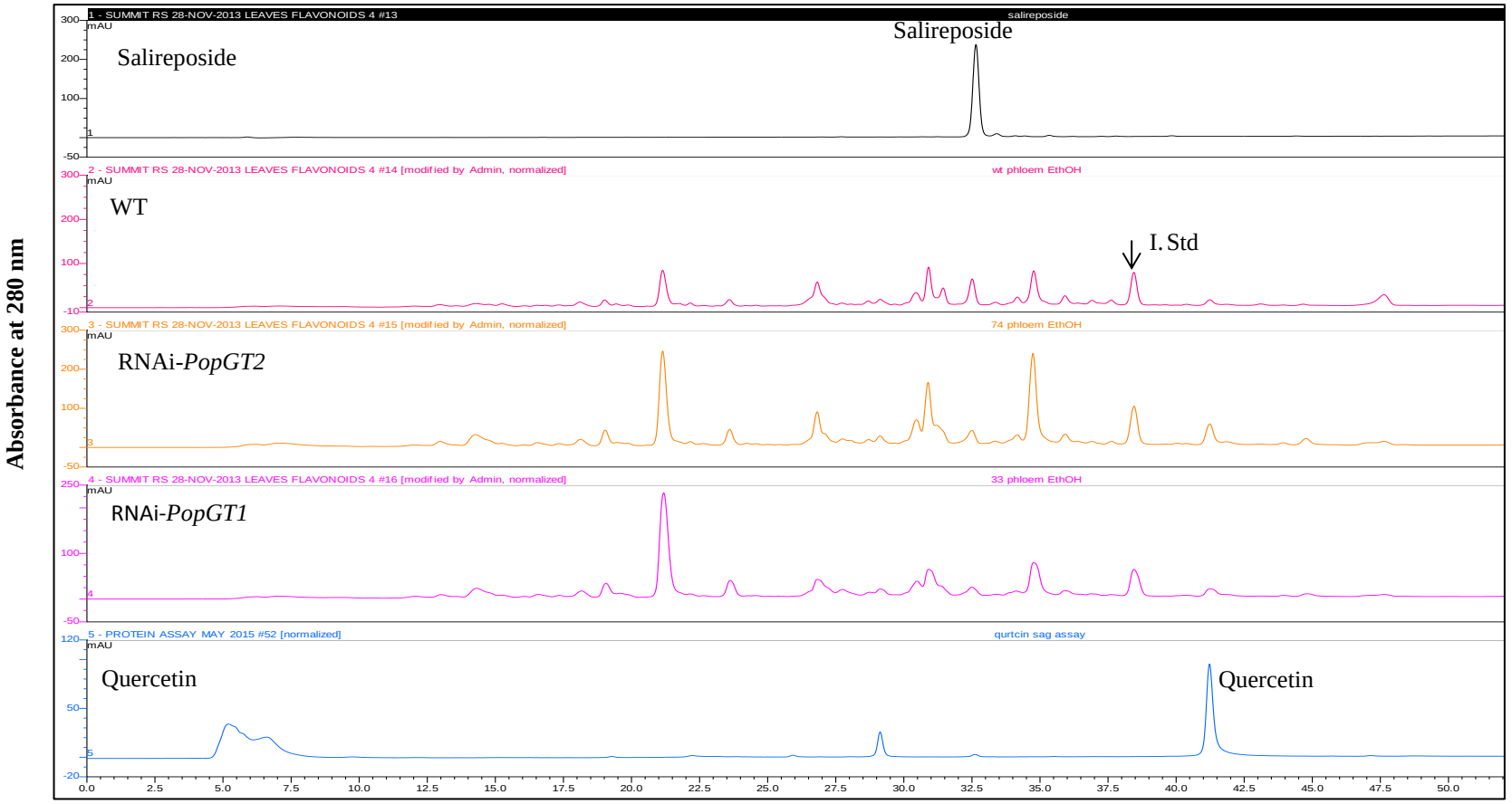
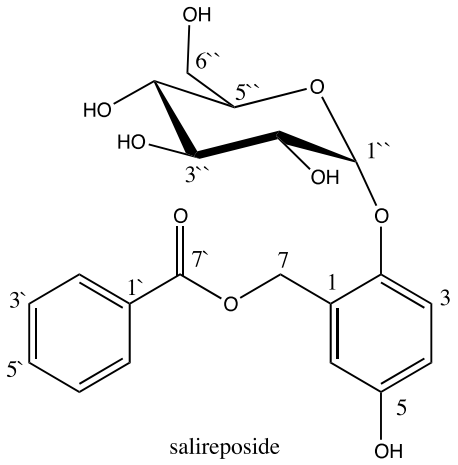


Figure 45: Phloem metabolites analysis of transgenic poplar trees compared to wild-type control. Salireposide and quercetin standards are shown. Peaks represent absorbance at 280 nm. I. Std. is the internal standard.



Chemical Formula: $C_{20}H_{22}O_9$
Exact Mass: 406.1264

Figure 46: Chemical structure of salireposide.

Chapter 4: Discussion

Phenolic metabolites in plants and glycosyltransferases

Phenolic metabolites are an important component of plant secondary metabolism that contribute to many aspects of plant growth and development, as well as reproduction, disease and stress resistance, and adaptation to changing environmental conditions (Jones and Vogt, 2001). In their non-conjugated forms, phenolic metabolites are largely toxic to plant cells (Bowles et al., 2006). Accordingly, plants maintain homeostatic levels of phenolic metabolites using different modification mechanisms, such as methylation, acylation, and glycosylation. These modifications change the physical and chemical properties of the metabolites, providing a means to control their biological activity, cellular localization and translocation to various tissues (Vogt and Jones, 2000; Zhong and Yue, 2005). In addition, the modifications generate a diverse array of novel secondary metabolites that ensure metabolic plasticity.

Several enzymes have evolved in plant cells to help generate these modified secondary metabolites from a limited number of available precursors, including glycosyltransferases (Paquette et al., 2003; Bowles et al., 2005; Vogt, 2010). Glycosylation, where a glycosyltransferase transfers a sugar moiety to an aglycones substrate, is one of the most important mechanisms that plants have evolved to regulate the homeostatic level of the phenolic secondary metabolites and xenobiotics (Vogt and Jones, 2000).

Plants have different families of glycosyltransferases, and each family contains GTs that are specialized in certain carbohydrate modification mechanisms (Geisler-Lee et al., 2006). Superfamily 1 of glycosyltransferases (UGTs) contains enzymes that are able to glycosylate

different types of small secondary metabolites, xenobiotics, and plant hormones (Bowles et al., 2006). UGTs were originally believed to provide a modification mechanism for activation, deactivation, and storage of lipophilic secondary metabolites, but new evidence suggests that UGTs are also active in hormonal and transcriptional regulation (Jones and Vogt, 2001; Bowles et al., 2005; Wang et al., 2012b).

Altering the gene expression of phenylpropanoid-specific UGTs can impact morphological and physiological phenotypes of the resultant plants. Examples of some phenotype changes that have been observed include cell wall composition, stem height, rosette shape, leaf morphology, root structure, altered life cycle, and different stress responses. Surprisingly, many of these observed phenotypic changes have never been explained biologically (Lim et al., 2001; Jackson et al., 2002; Woodward and Bartel, 2005; Bowles et al., 2006; Lanot et al., 2006; Wang et al., 2012a; Yin et al., 2014; Tanaka et al., 2014). Moreover, hundreds of putative glycosyltransferase genes have been identified within the increasing number of available genome sequences, but their catalytic functions and biological roles in plants remain essentially unknown (Tuskan et al., 2006; Osmani et al., 2009; Caputi et al., 2012).

In the present study, *in silico*, *in vitro,* and *in vivo* methods were employed to elucidate the substrate specificity of PopGT1 and PopGT2 enzymes and to understand their potential roles in plant growth and development.

4.1 Gene isolation and characterization

Two putative poplar UGTs (PopGT1 and PopGT2) that have high amino acid sequence similarity were isolated and characterized. Although PopGT2 had already been annotated as a

putative SAGT, based on its sequence similarity to the *Arabidopsis* gene At2g43820, *PopGT1* was an unknown poplar UGT gene. PopGT1 showed high sequence similarity to members of group L of family 1 UGTs (UGT74F1, UGT74F2, UGT74E2, and the *N. tabacum* salicylic acid glycosyltransferases) (Ross et al., 2001). However, sequence similarity between different members of family 1 UGTs has proven to be an inaccurate indicator for the exact enzyme function of UGTs (Bowles et al., 2006; Osmani et al., 2009), although it is still helpful in drawing a broad classification for an enzyme's specificity.

4.1.1 *In silico* analysis of the protein sequences of PopGT1 and PopGT2

Protein domain analysis

In silico signal peptide analysis of the deduced amino acid sequences of both PopGT1 and PopGT2 revealed no evidence for a signal peptide, suggesting that the two poplar enzymes are most likely cytosolic. This finding is in agreement with the results of Ross et al., (2001), who demonstrated that, of the 107 UGTs in *Arabidopsis*, none of those in family 1 have a signal peptide motif (Paquette et al., 2003; Lim and Bowles, 2004). This absence is in contrast to other membrane-bound GT families, such as the cellulose-synthase or callose-synthase GTs, which possess single or multiple transmembrane domains, and belong to different GT families (Gibeaut, 2000).

Further *in silico* characterization of both PopGT1 and PopGT2 showed the presence of a conserved UDP-binding domain PSPG-box, which is a characteristic domain of plant UGTs (Hughes and Hughes, 1994). Previous studies have shown that the presence of glutamine (Q) at residue number 44 in the PSPG box is important for glucosyl moiety transfer activity, while

substituting Q44 with a histidine (H) results in galactosyl moiety transfer activity (Kubo et al., 2004; Osmani et al., 2009; Wang and Hou, 2009; Ono et al., 2010). In the present study, sequence analysis of the PopGT1 and PopGT2 PSGP-box showed the presence of a glutamine (Q) residue at position 44 of both PopGT1 and PopGT2 domains. Accordingly, this finding was consistent with the biochemical results subsequently obtained with recombinant proteins, which showed that one of the isolated UGTs (PopGT1) employs UDP-glucose as a substrate.

Veljanovski and Constabel (2013) earlier found that poplar UGT78M1 harbours a histidine residue at position 44 of the PSPG-box domain and shows sugar-substrate specificity for UDP-galactose. Correspondingly, UGT78L1, which possesses a glutamine residue at the last position of the PSPG box, showed preference for UDP-glucose as a sugar donor.

Phylogenetic analysis

Phylogenetic analysis showed that the two poplar enzymes were clustered with members of group L UGTs (Figure 4), a group that includes enzymes that show catalytic activity towards a wide variety of phenolic substrates, including phenylpropanoids and plant hormones (Lim et al., 2003; Bowles et al., 2005; Lim et al., 2005; Li et al., 2001). When a second phylogenetic analysis was performed, using the two poplar genes and all members of group L, this clearly showed that the two poplar enzymes are closely related to the SAGTs and to auxin UGT74E2 that have been functionally characterized in other studies (Li et al., 2001; Lanot et al., 2006). These various *in silico* analyses suggested that PopGT1 and PopGT2 might be active towards SA and/or plant auxins.

4.2 Biochemical characterization of recombinant PopGT1 and PopGT2

Based on the *in silico* analysis it was anticipated that both PopGT1 and PopGT2 enzymes might show catalytic activity towards small phenolic substrates, such as phenylpropanoids and hydroxy benzoates. Surprisingly, in this study, no *in vitro* enzymatic activity was detected by PopGT2 towards any of the model substrates that were tested, although PopGT1 showed low activity towards cinnamic acid as well as the plant hormone IBA. In addition, PopGT1 showed broad activity towards different flavonoids, indicating promiscuous enzymatic activity. These biochemical results are not in agreement with the predictions arising from the phylogenetic analysis, indicating once again that while phylogenetic analysis based on sequence information might help in drawing a general idea about enzyme function, it does not represent an accurate method for protein function prediction (Vogt and Jones, 2000; Lim et al., 2001; Bowles et al., 2006). The observed broad catalytic activity found for PopGT1 is consistent with the wide regio-selectivity typically displayed by the UGT enzymes within group L (Bowles at al., 2006; Cartwright et al., 2008). For example, Lim et al., (2002) demonstrated that two *Arabidopsis* SAGT enzymes (UGT74F1 and UGT74F2), although sharing 90% sequence similarity and 78% sequence identity, possess different regio-specificity towards SA (UGT74F1 produces SAG, while UGT74F2 produces both SAG and SA glucose ester). In another study, PtUGT78L1 and PtUGT78M1, belonging to phylogenetic group F, were reported to be likely proanthocyanidin GTs based on co-expression data (Ross et al., 2001; Veljanovski and Constabel, 2013). However, Veljanovski and Constabel (2013) reported that in biochemical assays PtUGT78L1 showed substrate activity towards two flavonols and one cyanidin, while no *in vitro* activity was detected for PtUGT78M1 towards any of the tested biological substrates. Instead, the enzyme was able to glycosylate the model substrate 2,4,6-trichlorophenol. The authors suggested that despite the

high sequence similarity between the two enzymes, PtUGT78M1 might be active towards substrates that were not tested in their study (Veljanovski and Constabel, 2013). In contrast, two enzymes isolated from *Dorotheanthus bellidiformis*, UGT73A5 and UGT71F2, both glycosylated betanidin as a substrate although they shared only 19% sequence identity (Vogt, 2002).

In the present study, despite the broad substrate specificity of PopGT1, it displayed no activity towards the two model substrates (scopoletin and isoscopoletin). No *in vitro* enzymatic activity was observed for PopGT2 towards any of the tested phenolic compounds, including the two model substrates. The possibility remains that this result may reflect loss of enzymatic activity of PopGT2 during expression and purification. Most of the expressed PopGT1 and PopGT2 proteins accumulated in the insoluble fraction of the bacterial cell, which may have resulted in protein mis-folding and the loss of the catalytic activity of PopGT2. In general, protein mis-folding is a common problem when expressing eukaryotic proteins in *E. coli* systems, which lack the post-translational modification machinery required for proper protein folding (Baneyx and Mujacic, 2004). In addition, the enzymatic activity assay used in the present study might be missing an essential component for PopGT2 enzymatic activity. Alternatively, the enzyme might have been functionally active towards different phenolic and/or sugar donor substrates that were not tested in this study.

4.3 *In vivo* characterization of PopGT1 and PopGT2

4.3.1 *In vivo* activity of PopGT1 and PopGT2 in *Arabidopsis*

In the present study, the transgenic lines obtained were assessed phenotypically and biochemically to investigate the *in vivo* substrate activity, as well as the biological roles of both

PopGT1 and PopGT2 in plant growth and development. The results obtained from the metabolite analysis showed that OE-*PopGT1* transgenic plants accumulated high levels of the flavonol kaempferol. This is in agreement with the results obtained in the *in vitro* assay of PopGT1 substrate specificity, as kaempferol was indeed glycosylated by PopGT1. Interestingly, OE-*PopGT2* plants showed similar growth and cell wall phenotypes, even though no *in vitro* enzyme activity could be detected.

Plants generally do not accumulate free forms of flavonoids; instead, these compounds usually exist in their conjugated (multiply glycosylated) forms (Vogt and Jones 2000; Jones et al., 2003; Vogt, 2010). As an example of the remarkable diversity of these compounds, three hundred different quercetin glycosides have been reported in plants (Harborne and Baxter, 1999).

In vivo substrate recognition

Although the transgenic *Arabidopsis* leaves accumulated significant amounts of kaempferol conjugates (Tables 12 and 13), the PopGT1 enzyme showed higher relative *in vitro* activity towards quercetin. This difference between the detected *in vitro* and the *in vivo* activities towards kaempferol might occur because significant quantities of free quercetin may not be present in the leaves of the transgenic plants. My observation is in agreement with the results of Jones et al., (2003), who reported that kaempferol glycosides are more abundant in *Arabidopsis* leaves than are other flavonols. Other studies reported that in the absence of UV light, *Arabidopsis* leaf tissues accumulated kaempferol glycosides, whereas quercetin glycosides were found to accumulate in the flowers (Veit and Pauli, 1999; Bloor and Ibrahams, 2002; Jones et al., 2003; Modolo et al., 2007). This is consistent with the apparent absence of quercetin from the total and

the acid hydrolyzed metabolites of OE-*PopGT1* and OE-*PopGT2* transgenic *Arabidopsis* leaves. Similar results were obtained by Woo et al., (2007), when they over-expressed the *PsUGT1* (a gene that is known to glycosylate) in *Arabidopsis* plants. They observed the accumulation of kaempferol, which was the only flavonoid detected in the plant tissue, even though the PsUGT1 enzyme showed broad *in vitro* substrate specificity towards other flavonoids (Woo et al., 2007).

The total phenolic profiles of the OE-*PopGT* transgenic *Arabidopsis* plants showed several changes, which suggests that the introduced enzymes might not be specific to a single substrate. The apparent broad substrate recognition might reflect wide regio-selective activity of the enzymes. Several studies have reported family 1 UGTs have evolved wide regio-selectivity towards plant secondary metabolites and xenobiotics (Lim et al., 2004; Bowles et al., 2005; Wies et al., 2006; Isayenkova et al., 2006; Modolo et al., 2007; Cartwright et al., 2008; Veljanovski and Constabel., 2013; Yin et al., 2014). Therefore, the observed changes in the metabolic profile of the transgenic plants in response to over-expressing the *PopGT1* and *PopGT2* genes in the present study, might be a consequence of altered homeostatic level of the naturally existing secondary metabolites.

PopGT1 and PopGT2 did not show similar behavior as UGT74E2

PopGT1 showed close relatedness to the SA (UGT74F1 and UGT74F2) UGTs and the auxin (UGT74E2) UGT. However, neither PopGT1 nor PopGT2 showed *in vitro* activity towards SA, and only PopGT1 showed trace activity towards IBA. Tongetti et al., (2010) over-expressed UGT74E2, which showed activity towards several plant auxins, in *Arabidopsis*. The OE-*UGT74E1 Arabidopsis* plants showed typical auxin deficiency phenotypes, as well as increased

accumulation of both the free and conjugated forms of IAA and IBA. Furthermore, delayed bolting and flowering phenotypes were observed (Tongetti et al., 2010). Similar phenotypes have been reported in other studies where auxin-UGTs were over-expressed (Jackson et al., 2002; Tanaka et al., 2014). However, the phenotypic changes I observed when over-expressing the *PopGT1* and *PopGT2* genes are not consistent with the phenotypic changes observed in the UGT74E2 over-expressing *Arabidopsis* plants, as the plants showed early flowering, increased number of inflorescences per rosette, and increased plant height.

Growth and development phenotypic changes in transgenic *Arabidopsis*

OE-*PopGT1* and OE-*PopGT2* transgenic lines showed general morphological changes in the transgenic lines. It is well known that plant hormones control cell division and elongation as well as development (Bjorklund et al., 2007; Moubayidin et al., 2009). Jackson et al., (2002) showed that over-expressing *UGT84B1* resulted in auxin deficiency phenotypes, which included short stems, branching plants, increased hypocotyl elongation, and loss of root gravitropic response. However, those plants also showed a high auxin content (Jackson et al., 2002). Similarly, Tanaka et al., (2014) over-expressed the *UGT74D2* gene in *Arabidopsis,* they found similar auxin deficiency phenotypes. The direct effects of over-expression of *PopGT1* and *PopGT2* on plant hormones are not clear, and further studies are required to elucidate such impacts, if they exist.

In contrast, over-expression of *Stevia* GT (SrUGT74G1) led to an increase in plant height, rosette diameter, seed yield, pollen viability, increased trichrome branching, and accumulation of catechin (Guleria and Yadav, 2014). Moreover, over-expression of *PsUGT1* in *Arabidopsis* led to altered leaf morphology, root development, auxin response and shortened life cycle of

transgenic plants. These effects were suggested to be indirect results of altering the homeostatic levels of auxin and its conjugates, consistent with a regulatory role for flavonoids in maintaining plant auxin homeostasis (Woo et al., 2007). Some of these phenotypes are similar to the phenotype observed in the *PopGT1* and *PopGT2* over-expressing plants, but the substrate specificity of these poplar UGTs is different from that of UGTs studied earlier.

The *PopGT1* and *PopGT2* over-expressing plants showed differences in the levels of flavonoids, which are known to inhibit polar auxin transport (Peer and Murphy, 2007). Thus, glycosylation of endogenous flavonoids may have an effect on polar auxin transport, and thus plant growth and development (Yin et al., 2014). Alternatively, flavonoids (aglycone forms or glycosylated compounds) have been shown to interfere with many proteins including transcription factors, the auxin transporter proteins, kinases and phosphorylases (Hichri et al., 2011; Peer et al., 2011; Yin et al., 2012; Onkokesung et al., 2014; Yin et al., 2014; Wang et al., 2016). For example, inhibition of flavonoid synthesis in the transparent testa4 (*tt4*) mutant plants resulted in increased auxin transport (Peer et al., 2004; Buer and Muday, 2004; Buer et al., 2013). It has been reported that the effect of flavonoids on auxin transport may be mediated through binding to the PIN and PINOID proteins. Therefore, when flavonoids bind PINOID, they alter PIN transporter localization and as such alter polar auxin transport (Santelia et al., 2008; Geisler et al., 2014; Yin et al., 2014). Mutations in the PIN1 protein resulted in an auxin deficiency phenotype, including altered leaf morphology, increased branching, and loss of root gravitropism response. Similar phenotypes were observed in *Arabidopsis* plants over-expressing the auxin *UGT84B1* gene (Jackson et al., 2002). Thus, the poplar genes misregulated in this study may be indirectly impacting auxin transport in transgenic *Arabidopsis*, which may be the cause for the altered plant development.

The OE-*PopGT1* and OE-*PopGT2* transgenic *Arabidopsis* plants demonstrated early bolting and flowering, traits that are also closely linked with plant hormones status (Bjorklund et al., 2007; Moubayidin at al., 2009). Similar results were observed by Wang et al., (2012) who demonstrated that transgenic tobacco expressing *PtGT1* had an early flowering phenotype (Wang et al., 2012a). Woo et al., (2007) also showed that the expression of *UGT85A7* is required for plant growth and development, and *UGT85A7* over-expressing transgenic plants had a shorter life cycle, indicating a cell cycle regulation effect of the enzyme. Furthermore, the *ugt87a2* mutant showed a delayed flowering phenotype, while over-expression of *UGT87A2* in the mutant background recovered the phenotype (Wang et al., 2012b). The authors went on to show that UGT87A2 is a regulator of flowering time, which is achieved through its effect on the flowering repressor FLOWERING LOCUS C (FLC). However, the enzyme catalytic function and intracellular target(s) remains unknown (Wang et al., 2012b).

In the present study, OE-*PopGT1* and OE-*PopGT2* transgenic plants showed increased branching (Tables 7 and 8) and loss of root gravitropism phenotypes (Figure 24), similar to the phenotypes observed in studies by Jackson et al., (2001), Lim et al., (2001) and Tanaka et al., (2014). However, in contrast to these published studies, the stem height and rosette diameter of the OE-*PopGT1* and OE-*PopGT2* plants were increased. Recently, it was reported that when a poplar UGT gene (*PtGT1*) is over-expressed in tobacco plants, the transgenic plants show altered morphology, as well as cell wall structure alterations, including reduction in klason lignin and early flowering phenotype (Wang et al., 2012 a). In addition, the recombinant protein lacked *in vitro* activity, when tested against a suite of possible substrates, including monolignol phenolic acids and alcohols (Wang et al., 2012 a).

Expansion of UGTs in poplar

The secondary metabolite complexity of poplar has likely evolved from the genome duplication, which resulted in more UGTs and/or enzymes that have wide substrate specificities (Babst et al., 2010). This implies that functional evolution may have occurred, where paralogous enzymes are recruited to perform different roles in trees (Caputi et al., 2012; Sakakibara and Hanad, 2011). Accordingly, the two coding sequences that were isolated from poplar trees (*PopGT1* and *PopGT2*) might have evolved to perform different functions than that in *Arabidopsis.* Consequently, the observed change in the phenotypes of the transgenic *Arabidopsis* plants in the present study may be the result of non-specific glycosylation of physiologically important substrates by the poplar enzymes. This non-specific glycosylation could in turn lead to mis-localization and/or the deactivation/activation of their substrates in several plant compartments. The *in vitro* substrate specificity of PopGT2 was not detected, and more characterization experiments are required to confirm this hypothesis.

Caputi et al., (2012) studied the expansion of group D UGTs in different plant species, and demonstrated that the increased numbers of these genes might be the result of gene duplication via unequal crossover, producing gene clusters on the chromosomes and paralogs of the same gene. In time, enzyme functional divergence may have occurred and the proteins could have acquired new functions (Caputi et al., 2012). These such processes secure a diverse array of enzymes capable of producing a variety of secondary metabolites, providing a means to supply a plant with the requirements for growth and development, as well as increasing the chemical plasticity and fitness (Vogt and Jones, 2000; Li et al., 2001; Gachon et al., 2005; Caputi et al., 2012). Sakakibara and Hanad (2011) analyzed family 1 UGTs and suggested that evolutionary

processes, such as convergent and repeated evolutionary events, might be an explanation for the independent development of enzymes that recognize the same substrates in different plants, and the poor correlation between substrate recognition and sequence similarity might have occurred as UGTs ability to recognize substrate evolved after expansion in new species that allows the UGT enzymes to recognize new substrates (Sakakibara and Hanad, 2011). This suggestion might explain the results obtained in the present study, in which the sequence similarity and phylogenetic characterization of the PopGT1 and PopGT2 enzymes have classified them as SA UGTs, while the *in vitro* substrate activity demonstrated a different catalytic activity for PopGT1.

4.3.2 Down-regulation of *PopGT1* and *PopGT2* in hybrid poplar

The impact of down-regulation of *PopGT1* and *PopGT2* genes resulted in changes in tree morphology (secondary growth), and in biochemical and biomechanical properties. Plant secondary growth is associated with the activity of the cambial meristem, and stem diameters have been shown to correlate with IAA concentration, but not with ABA (Pearce et al., 2004). Similarly, Tuominen et al., (1997) reported that IAA and gibberellins control primary and secondary growth developmental regulation, where IAA can modify GA levels. Early studies by Ceulemans and Isebrands (1996) indicated that biochemical characterization of young trees may provide an early screen for identifying superior growth phenotypes in poplar. For instance, it has been demonstrated that enhanced hormone concentration in young trees is positively correlated with rapid growth of the trees (Pharis et al., 1991). Therefore, it is possible that the results obtained in this study through the down-regulation of *PopGT1* and *PopGT2* are a consequence of

disturbing hormonal homeostasis, although hormone analysis would be required to test this hypothesis.

The transgenic trees generated showed a brittle stem phenotype. Stem cross-sectioning and microscopic examination indicated that there were some alterations in wood structure, including vessel distribution patterns. It has been recognized that auxins and gibberellins are responsible for controlling vessel formation during early development (Aloni, 1979; Aloni, 2013), and the ratio between auxin and gibberellin is an important factor in cell differentiation (Aloni, 2013). However, no biochemical data is available from the current study that would directly address this possibility.

In addition to changes in vessel distribution, the phloem fibres showed structural changes in the fibre-bundle distribution, as well as fibre aggregation patterns (Figures 36). These results suggest an essential role for the two poplar UGTs in both xylem and phloem development. The mechanism behind the observed phloem and xylem phenotype alteration is unknown. According to the canalization theory, early movement of auxin in the cambium regulates vessel formation (Sachs, 1980). Auxin polar transport was found to be essential for the functional continuity of the vessel, since inhibition of PIN and PID polar auxin transport generates discontinuous vessels. The rate of auxin production was also found to influence the width of the generated vessels (Yin et al., 2014).

Biomechanical analysis of both RNAi-*PopGT1* and RNAi-*PopGT2* transgenic trees in the present study indicated that wood density and Young's Modulus of Elasticity were each reduced significantly compared to the wild-type trees. Wood density has been associated with wood

mechanical properties, insofar as wood with higher density is generally stiffer and more resistant to breakage than wood with lower density (Chave et al., 2009).

Fibres Quality Analysis (FQA) revealed insignificant changes in fibre length in the transgenic trees, but fibre width increased significantly. However, this was not observed in microscopic views of stem cross-sections, which showed that the cell wall thickness was apparently unaltered. It seems likely, therefore, that the increase in fibre and vessel lumen diameter would account for the reduced wood density.

Previously, it was shown that increased concentration of auxins results in formation of narrower vessels, while low concentration of auxins results in wider vessels (Aloni et al., 2013). The effect of auxin concentration on vessel width was ascribed to the rate of vessel formation, since low auxin concentration provides more time for the vessel to expand, while high levels of auxin accelerate the vessel development process (Aloni, 2013; Woodward and Bartel, 2005). This may not entirely have a direct effect of auxin, since gibberellin has also been shown to affect fibre length (Aloni, 1979), fibre wall thickness, and fibre lumen diameter, all properties, which affect wood density (Fujiwara et al., 1991; Jacobsen et al., 2007; Martínez-Cabrera et al., 2009).

Leaf morphology and bud break phenotypes were also altered. In the transgenic trees, an early bud break phenotype was scored in two consecutive growing seasons, showing significant differences in the timing at which trees break their bud dormancy. Down-regulation of *PopGT1* and *PopGT2* appears to indirectly affect genes that impact plant growth and development, likely by altering the pools of specific secondary metabolites in emerging tissues, which may then influence the hormonal status of these plants. In a previous study, over-expressing a maize auxin glycosyltransferase in poplar trees resulted in a dramatic effect on tree growth and development

(Salyaev et al., 2006). The transgenic trees showed improved growth and an accelerated bud and branch development phenotype, which was associated with increased levels of free and glucose-conjugated auxin.

The effects of UGTs that glycosylate phenylpropanoid metabolites during plant growth and development are obvious (Bowles et al., 2006), but the underlying mechanism is not clear. Research has shown some effects of altering gene expression on cell wall structures; however, these effects have not been explained clearly (Bowles et al., 2006; Wang et al., 2012a). Nevertheless, these changes might be secondary responses of the plant as it works to control the homeostatic level of other secondary products.

The results of the present study identified salireposide as one of the compounds where the concentration in the bark of transgenic trees was reduced in response to down-regulation of *PopGT1* and *PopGT2*. Poplar trees accumulate high concentrations of phenolic glycosides, including salicin and salicylate glycosides in their leaves, but no UGTs with *in vivo* catalytic activities towards this group of metabolites has yet have been identified, except UGT71L1 which has been reported to have activity towards salicin and salicortin in the roots of poplar (Tsai et al., 2006; Tsai et al., 2011). The reduction in salireposide accumulation observed here suggests that PopGT1 and PopGT2 do play a role in salicin based phenolic glycoside biosynthesis in poplar trees.

Cell wall chemical analysis of the transgenic lines showed a marginal but significant increase in lignin content and reductions in cellulose content (determined as a structural glucose and supported by alpha cellulose analysis). In contrast, bark tissue showed a significant reduction in lignin content. The marginal changes in cell wall chemistry do not provide strong evidence for a

direct contribution of the two enzymes to cell wall biosynthesis or glycosylation of monolignols. However, it does suggest that altering flavonoid glycosylation can impact growth and development, including cell wall formation.

Plants control rising levels of flavonoids through glycosylation, as well as feedback inhibition mechanisms (Yin et al., 2012). The spatial and temporal distribution of both the enzymes and their target substrate availability is an important factor that might affect plant metabolite homeostasis. A study by Wang et al., (2012) investigating the impacts of over-expressing *PtGT1* in tobacco plants showed that lignin deposition was increased, and an early flowering phenotype was observed, even though the recombinant enzyme did not show the expected *in vitro* specificity towards monolignols (Wang et al., 2012 a).

HPLC analysis and spectroscopic analysis of the UGT-suppressed trees showed that several soluble metabolites varied in concentration, and that flavonoids appeared to be impacted substantially. Flavonoid biosynthesis is linked to phenylpropanoid metabolism. For example, it was found that silencing hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase led to a strong reduction in plant growth and lignin deposition. This was associated with an increase in flavonoid glycosides that accumulated in the transgenic plants, as did acylated anthocyanins (Besseau et al., 2007). Besseau et al., (2007) also reported that stem cross-sections showed altered tracheary elements phenotype in the transgenic plants.

Based on the data generated in the present study, PopGT1 and PopGT2 are most likely two new flavonol glycosyltransferases that have not been previously characterized.

Altering *PopGT1* and *PopGT2* gene expression in both *Arabidopsis* and poplar caused noticeable phenotypic and biochemical changes in both species. These phenotypic changes suggest an important role for these genes in plant growth and development, as well as plant physiology. However, further studies are needed to clarify the mechanisms behind their action. The identification of the salicin-based phenolic glycoside, salireposide, as a potential *in vivo* substrate of the PopGT1 and PopGT2 is a new finding that reflects possible regio-switching events that could occur during the evolution of poplar glycosyltransferases, which altered their regio-selectivity. Further in-depth studies are required to clarify the mechanism behind these findings.

Chapter 5: Conclusion

5.1 Conclusion

In this thesis, two putative poplar UGTs coding sequences (*PopGT1* and *PopGT2*) were isolated and characterized, using *in silico*, biochemical, and molecular biology techniques. The isolated sequences showed similarities to the previously annotated *Arabidopsis* SAGT and *N. tabacum Nt*SAGT (Lee and Raskin, 1999). An *in silico* analysis suggested that the two proteins are most likely cytosolic proteins possessing a carboxylic terminus conserved amino acid domain (PSPG-box), which is a characteristic domain for plant UGTs (Hughes and Hughes, 1994). Sequence analysis of the 44 amino acid residues in this domain predicted that the two proteins are putative UDP-glucose GTs. Biochemical analysis using recombinant protein assays confirmed that UDP-glucose is indeed a substrate for *PopGT1*. Phylogenetic analysis was performed to investigate the relatedness of the two proteins to other members of family 1 UGTs, and revealed that the two proteins clustered with members of phylogenetic group L, which are generally active towards phenylpropanoids and plant hormones (Ross et al., 2001; Li et al., 2001). The coding sequence of the two genes was cloned separately into the protein expression vector *pET28a* and recombinant proteins were generated in *E. coli*. Enzyme activity assays were carried out to determine the substrate specificity towards different phenolic substrates and two UDP-sugars (UDP-glucose and UDP-galactose). The results showed that only PopGT1 was active *in vitro* and no activity for PopGT2 towards any of the substrates tested were detected. PopGT1 showed activity towards nine different substrates, mostly flavonoids, and trace activity towards cinnamic acid and IBA.

Protein functional characterization was then pursued in the model plant *Arabidopsis*. The coding sequence of the two genes was cloned into the *pSM3* vector under the control of the double 35S

constitutive promoter. Several transgenic lines were characterized phenotypically and biochemically to evaluate the effect of over-expressing this gene on plant growth and development. All transgenic lines showed increased plant height, rosette diameter, and number of stems per plant. Loss of root gravitropic response, as well as early bolting and flowering phenotypes were also observed. Histochemical analysis showed no difference in the inflorescence anatomy and vascular system, as well as cell wall development. High performance liquid chromatographic (HPLC) analysis of total and acid-hydrolyzed metabolites indicated that the transgenic lines accumulated a flavonol (kaempferol) in the leaf tissues of transgenic plants in greater quantities than that of wild-type control. These findings support the *in vitro* enzyme characterization results.

RNAi-mediated suppression constructs were designed using *pHellsgate12* vector under the control of the 35S constitutive promoter, transformed into poplar using *Agrobacterium*. Phenotypic characterization of the transgenic trees, which were allowed to grow in the green house, indicated that stem height was not affected, but stem diameters were reduced significantly. The stems also had a brittle phenotype and significant reductions in wood elasticity were apparent as determined by Young's Modulus of Elasticity. In addition, significant reductions in wood density were observed. Histochemical staining and microscopic examination showed different distribution patterns of phloem fibres and phloem fibre bundles, in addition to reduced bark thickness. Changes in xylem tissues were inconsistent. Fibre quality analysis showed that fibre length was not affected in the transgenic trees, while fibre width increased significantly. Cell wall composition analysis showed a slight, but significant increase in lignin content, while significant reductions in glucose content were also apparent. These results were confirmed using alpha cellulose analysis of sub-samples.

Interestingly, morphological changes were obvious in the subsequent growing seasons: including retained leaf stipules, increased branching, as well as smaller and deformed leaves.

HPLC analysis of the total soluble metabolites and acid hydrolyzed metabolites indicated altered phenolic profile in the transgenic trees compared to wild-type controls. The UV-spectra generally matched common signatures of flavonoid UV-spectra. The phenotypic changes in the RNAi-mediated suppression poplar trees were attributed to the altered phenylpropanoid metabolic profile, and these changes may have a knock-on effect causing changes in the hormonal balance in trees.

Although the present study indicates that the two-isolated genes PopGT1 and PopGT2 are most likely flavonoid glycosyltransferases that can impact plant growth and development, further study is need to better understand the possible biological and physiological roles of these enzymes in plant development.

5.2 Future work

Based on the phenotypic changes that were observed in the transgenic plants generated in this thesis, several follow-up experiments are suggested:

(i) An analysis of the status of plant hormones in transgenic plants.

(ii) In order to avoid the *in vivo* functional redundancy of the two poplar genes, RNAi-suppression that targets both *PopGT1* and *PopGT2* simultaneously is required.

(iii) RNAseq of transgenic trees to study the global changes in gene regulation, and more specifically examine the correlation between the transcription abundance of UGT

genes belonging to clade 74 (which are normally low) and the changes in secondary metabolite contents, leaf morphology, and plant hormonal levels.

(iv) Elucidate the *in vitro* function of PopGT2 in different eukaryotic protein expression systems, such as insect cell-culture system to overcome the limitations of prokaryotic expression systems.

(v) Screen more salicylate aglycones in the enzymatic assay of PopGT1 and PopGT2 to provide a better understanding of the role of poplar GTs in the formation of high levels of phenolic glucosides in trees.

References

Achnine, L., Huhman, D. V., Farag, M. A., Sumner, L. W., Blount, J. W., and Dixon, R. A. (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. Plant Journal, 41: 875-87.

Albersheim, P., Darvill, A., Roberts, K., Sederoff R. et al., (2010). Plant Cell Walls, Garland Science, Taylor and Francis Group, LLC, New York.

Aloni, R. (1979). Role of auxin and gibberellin in differentiation of primary phloem fibres. Plant Physiology, 63: 609-614.

Aloni, R. (2013). Role of hormones in controlling vascular differentiation and the mechanism of lateral root initiation. Planta, 238: 819-830.

Babst, B. A., Chen, H. Y., Wang, H. Q., Payyavula, R, S., Thomas T. P, Harding, S. A., and Tsai, C. J. (2014). Stress-responsive hydroxycinnamate glycosyltransferase modulates phenylpropanoid metabolism in *Populus*. Journal of Experimental Botany, 65: 4191-4200.

Babst, B. A., Harding, S. A., and Tsai, C. J. (2010). Biosynthesis of phenolic glycosides from phenylpropanoid and benzenoid precursors in *Populus*. Journal of Chemical Ecology, 36: 286-297.

Bajguz, A. (2007). Metabolism of brassinosteroids in plants. Plant Physiology and Biochemistry, 45: 95-107.

Baneyx, F., and Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. Nature Biotechnology, 22: 1399 -1408.

Barvkar, V. T., Pardeshi, V. C., Kale, S. M., Kadoo, N. Y. et al., (2012). Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in *Linum usitatissimum* identified genes with varied expression patterns. BMC Genomics, 13:175.

Bendtsen, B. A., and Senft, J. (1986). Mechanical and anatomical properties in individual growth rings of plantation-grown eastern cottonwood and loblolly pine. Wood and Fibre Science, 18: 23-38.

Besseau, S., Hoffman, L., Geoffory P., Lapierre, C., Pollet, B., Legrand, M. (2007). Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affect auxin transport and plant growth. The Plant Cell, 19: 148-162.

Bjorklund, S., Antti, H., Uddestrand, I., Moritz, T., and Sundberg, B. (2007). Cross-talk between gibberellin and auxin in development of *Populus* wood: Gibberellin stimulates polar auxin transport and has a common transcriptome with auxin. Plant Journal, 52: 499–511.

Bloor, S. J., and Abrahams, S. (2002). The structure of the major anthocyanin in *Arabidopsis thaliana*. Phytochemistry, 59: 343-346.

Bowles, D., Isayenkova, J., Lim, E-K., and Poppenberger, B. (2005). Glycosyltransferases: managers of small molecules. Current Opinions in Plant Biology, 8: 254-263.

Bowles, D., Lim, E-K., Poppenberger, B., and Vaistij F. E. (2006). Glycosyltransferases of lipophilic small molecules. Annual Review of Plant Biology, 57: 567-597.

Bradshaw, H. D., Ceulemans, R., Davis, J., and Stettler, R. (2000). Emerging Model Systems in Plant Biology: Poplar (*Populus*) as A Model Forest Tree. Journal of Plant Growth Regulators, 19: 306-313.

Brazier-Hicks, M., Edwards, L. A., and Edwards, R. (2007a). Selection of plants for roles in phytoremediation: the importance of glucosylation. Plant Biotechnology Journal, 5: 627-635.

Brazier-Hicks, M., Offen, W. A., Gershater, M. C., Revett, T. J., et al., (2007b). Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. Proceeding of Natural Academy of Science, USA, 104: 20238-20243.

Brown, D. M., Goubet, F., Wong, V. W., Goodacre, R., et al., (2007). Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. Plant Journal of Cell Molecular Biology, 52: 1154-1168.

Buer, C. S., and Muday, G. K. (2004). The transparent testa4 mutation prevents flavonoid synthesis and alters auxin transport and the response of *Arabidopsis* roots to gravity and light. The Plant Cell, 16: 1191-1205.

Buer, C. S., Imin, N., and Djordjevic, M. A. (2010). Flavonoids: new roles for old molecules. Journal of Integrated Plant Biology, 52:98-111.

Buer, C. S., Kordbacheh, F., Truong, T. T., Hocart, C. H., and Djordjevic, M. A. (2013). Alteration of flavonoid accumulation patterns in transparent testa mutants disturbs auxin transport, gravity responses, and imparts long-term effects on root and shoot architecture. Planta, 238: 171-189.

Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochemistry Journal, 326: 929-939.

Canam, T., Unda, F., and Mansfield, S. D. (2008). Heterologous expression and functional characterization of two hybrid poplar cell wall invertases. Planta, 228: 1011.

Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., and Martens, S. (2012). A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. Plant Journal of Cell Molecular Biology, 69: 1030-1042.

Cartwright, A. M., Lim, E-K., Kleanthous, C., and Bowles, D. J. (2008). A kinetic analysis of regiospecific glucosylation by two glycosyltransferases of *Arabidopsis thaliana* domain swapping to introduce new activities. The Journal of Biological Chemistry, 283: 15724-15731.

Ceulemans, R., and Isebrands J.G. (1996). Carbon acquisition and allocation. In Biology of *Populus* and its Implications for Management and Conservation. In Stettler R.F., Bradshaw H.D., Heilman Jr., P.E. and Hinckley T.M. (Eds). NRC Research Press, Ottawa, pp 355-399.

Charles, P. M., Paul, V. A., Mendel, F., and William, R. B. (1997). Cloning and expression of solanidine UDP-glucose glucosyltransferase from potato. Plant Journal, 11: 227-236.

Chave, J., Coomes, D., Jansen, S., Lewis, S. L., Swenson, N. G., Zanne, A. E. (2009). Towards a worldwide wood economics spectrum. Ecology Letters, 12: 351-366.

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

Coleman, H. D., Canam, T., Kang, K. Y., Ellis, D. D., et al., (2007). Overexpression of UDP-glucose pyrophosphorylase in hybrid poplar affects carbon allocation. Journal of Experimental Botany, 58: 4257-4268.

Compain, P. P., and Martin, O. R. (2001). Carbohydrate mimetics-based glycosyltransferase inhibitors. Bioorganic and Medicinal Chemistry, 9: 3077-3092.

Cooke, J. E. K., and Rood, S. B. (2007). Trees of the people: The growing science of poplars in Canada and worldwide. Canadian Journal of Botany-Revue Canadienne de Botanique, 85: 1103-1110.

Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. Journal of Molecular Biology, 328: 307-317.

Cronk, Q. C. B. (2005). Plant eco-devo: the potential of poplar as a model organism. New Phytologists, 166: 39-48.

Cullis, I. F., Saddler, J. N., and Mansfield, S. D. (2004). Effect of initial moisture content and chip size on the bioconversion efficiency of softwood lignocellulosics. Biotechnology and Bioengineering, 85: 413-421.

Dean, J.V., and Delaney, S. P. (2008). Metabolism of salicylic acid in wild-type, *ugt74f1* and *ugt74f2* glucosyltransferase mutants of *Arabidopsis thaliana*. Physiology Plantarum, 132: 417-425.

DeLano, W. L. (2002) PyMOL molecular graphics system. Available at: http://www.pymol.org.

Dempsey, D. A., Shah, J., and Klessig, D. F. (1999). Salicylic Acid and Disease Resistance in Plants. Critical Reviews in Plant Science, 18: 1360-1385.

Dempsey, D. A., Vlot, A. C., Wildermuth, M. C., and Klessig, D. F. (2011). Salicylic acid biosynthesis and metabolism. The *Arabidopsis* book/American society of plant biologists, 9, e0156. http://doi.org/10.1199/tab.0156

Dickmann, D. I., and Keathley, D.A. (1996). Linking physiology, molecular genetics, and the *Populus* ideotype. In: Stettler, R.F., Bradshaw, H.D., Jr., Heilman, P.E., and Hinckley, T.M., editors. Biology of *Populus* and its implications for management and conservation. NRC Research Press, Ottawa, pp. 491-514.

Dickmann, D. I., Isebrands, J. G., Eckenwalder, J. E. and Richardson, J. (editors) (2001). Poplar culture in North America. National Research Council of Canada, Research Press, Ottawa, ON. 397.

DiFazio, S. P., Slavov, G. T., and Joshi, C. P. (2011). *Populus*: A premier pioneer system for plant genomics. Genetics, genomics and breeding of poplar, Science Publishers, 1-28.

Dixon, R. A., Achnine, L., Kota, P., Liu, C. J., Reddy, M. S. S., and Wang, L. (2002). The phenylpropanoid pathway and plant defence: a genomics perspective. Molecular Plant Pathology, 3: 371-390.

Dixon, R. A., and Paiva, N. L. (1995). Stress-Induced Phenylpropanoid Metabolism. Plant Cell, 7: 1085-1097.

Eckenwalder, J. E. (1996). Systematics and evolution of *Populus*. In: Stettler, R. F., Bradshaw, H. D., Jr., Heilman, P. E., and Hinckley, T. M., editors. Biology of *Populus* and its implications for management and conservation. NRC Research Press, Ottawa, pp. 7-32.

Eudes, A., Bozzo, G. G., Waller, J. C., Naponelli, V., et al., (2008). Metabolism of the folate precursor p-aminobenzoate in plants: glucose ester formation and vacuolar storage. Journal of Biological Chemistry, 283: 15451-15459.

Fahey, J. W., Zalcmann, A. T., and Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochemistry, 56: 5-51.

Fedoroff, N., Furtekt, D., and Nelson, O. (1984). Cloning of the bronze locus in maize by a simple and generalizable procedure using the transposable controlling element Activator (Ac). Proceeding of National Academy of Science, USA, 81: 3825-3829.

Frankova, L., and Fry, S. C. (2013). Biochemistry and physiological roles of enzymes that 'cut and paste' plant cell wall polysaccharides. Journal of Experimental Botany, 64: 3519–3550.

Fujioka, S., and Yokota, T. (2003). Biosynthesis and metabolism of brassinosteroids. Annual Review of Plant Biology, 54: 137-164.

Fujiwara, S., Sameshima, K., Kuroda, K., Takamura, N. (1991). Anatomy and properties of Japanese hardwoods. I. Variation of fibre dimensions and tissue proportions and their relation to basic density. IAWA Bulletin, NS, 12: 419-424.

Gachon, C. M., Langlois-Meurinne, M., and Saindrenan, P. (2005). Plant secondary metabolism glycosyltransferases: the emerging functional analysis. Trends in Plant Science, 10: 542-549.

Gaffney, T., Friedrich, L., Vernooij, B., Negrotto, D., et al., (1993). Requirement of salicylic Acid for the induction of systemic acquired resistance. Science, 61: 754-756.

Geisler, M., Wang, B. and Zhu, J. (2014), Auxin transport during root gravitropism: transporters and techniques. Plant Biology Journal, 16: 50-57.

Geisler-Lee, J., Geisler, M., Coutinho, P. M., Segerman, B., et al., (2006). Poplar carbohydrate-active enzymes. Gene identification and expression analyses. Plant Physiology, 140: 946-962.

Gibeaut, D. M. (2000). Nucleotide sugars and glycosyltransferases for synthesis of cell wall matrix polysaccharides. Plant Physiology and Biochemistry, 38: 69-80.

Guleria, P., and Yadav, S. K. (2014). Overexpression of a glycosyltransferase gene SrUGT74G1 from Stevia improved growth and yield of transgenic *Arabidopsis* by catechin accumulation. Molecular Biology Reports, 41: 1741–1752.

Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series, 41: 95-98.

Harborne, J. B., and Baxter H. (1999) The Handbook of Natural Flavonoids, vol. 1, Wiley and Sons, Chichester, west Sussex, England.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. Journal of Experimental Botany, 62: 2465-2483.

Hong, Z., Zhang, Z., Olson, J. M., and Verma, D. P. (2001). A novel UDP-glucose transferase is part of the callose synthase complex and interacts with phragmoplastin at the forming cell plate. Plant Cell, 13: 769-779.

Hood, E. E., Gelvin, S. B., Melchers, L. S., and Hoekema, A. (1993). New Agrobacterium helper plasmids for gene transfer to plants. Transgenic Research, 2: 208-218.

Hou, B., Lim, E-K., Higgins, G. S., and Bowles, D. J. (2004). N-glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. Journal of Biological Chemistry, 279: 47822-47832.

Hughes, J., and Hughes, M. A. (1994). Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (*Manihot esculenta* Crantz) cotyledons. DNA Sequence, 5: 41-49.

Husar, S., Berthiller, F., Fujioka. S., Rozhon, W., et al., (2011). Overexpression of the UGT73C6 alters brassinosteroid glucoside formation in *Arabidopsis thaliana*. BMC Plant Biology, 11: 51.

Hwang, S. Y., and Lindroth, R. L. (1997). Clonal variation in foliar chemistry of aspen: Effects on gypsy moth and forest tent caterpillar. Oecologia, 111: 99-108.

Isayenkova, J., Wray, V., Nimtz, M., Strack, D., and Vogt, T. (2006). Cloning and functional characterisation of two regioselective flavonoid glucosyltransferases from *Beta vulgaris*. Phytochemistry, 67: 1598-612.

Isebrands, J. G., and Richardson, J. (2014). Poplars and Willows: Trees for Society and Environment; CABI: Oxford, UK.

Jackson, R. G., Lim, E-K., Li, Y., Kowalczyk, M., et al., (2001). Identification and biochemical characterization of an *Arabidopsis* indole-3-acetic acid glucosyltransferase. Journal of Biological Chemistry, 276: 4350-4356.

Jackson, R.G., Kowalczyk, M., Li, Y., Higgins, G., and Ross, J., et al., (2002) Over-expression of an *Arabidopsis* gene encoding a glucosyltransferase of indole-3-acetic acid: phenotypic characterisation of transgenic lines. Plant Journal, 32: 573–83.

Jaillon, O., Aury, J. M., Noel, B., Policriti, A., et al., (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature, 449: 463-467.

Jin, S. H., Ma, X. M., Han. P., Wang, B., Sun, Y.G., Zhang G-Z., et al., (2013). UGT74D1 Is a Novel Auxin Glycosyltransferase from *Arabidopsis thaliana*. PLoS ONE, 8: 10.1371

Jones, P. R., Moller, B. L., Hoj, P. B. (1999). The UDP-glucose:*p*-hydroxymandelonitrile-O-glucosyltransferase that catalyzes the last step in synthesis of the cyanogenic glucoside

dhurrin in *Sorghum bicolor*. Isolation, cloning, heterologous expression, and substrate specificity. Journal of Biological Chemistry, 274: 35483–91.

Jones, P., and Vogt, T. (2001). Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. Planta, 213: 164-174.

Jones, P., Messner, B., Nakajima, J., Schaffner, A. R., and Saito, K. (2003). UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*. Journal of Biological Chemistry, 278: 43910-43918.

Kapitonov, D., and Yu, R. K. (1999). Conserved domains of glycosyltransferases. Glycobiology, 9: 961-678.

Keegstra, K., and Raikhel, N. (2001). Plant glycosyltransferases. Current Opinion in Plant Biology, 4: 219-224.

Kelley, L.A., Mezulis, S., Yates C.M., Wass, M. N., and Sternberg, M. E. (2015).The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols, 10: 845–858. [doi:10.1038/nprot.2015.053].

Kubo, A., Arai, Y., Nagashima, S., and Yoshikawa, T. (2004). Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Archive Biochemistry and Biophyics, 429: 198-203.

Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. Annual Review of Biochemistry, 77: 521-555.

Lanot, A., Hodge, D., Jackson, R. G., George, G. L., Elias, L., Lim, E-K., Vaistij, F. E., and Bowles D. J. (2006). The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in *Arabidopsis thaliana*. The Plant Journal, 48: 286-295.

Lee, H. I., and Raskin, I. (1999). Purification, cloning, and expression of a pathogen inducible UDP-glucose:Salicylic acid glucosyltransferase from tobacco. Journal of Biological Chemistry, 274: 36637-36642.

Leloir, L. F., and Cardini, C. E. (1957). Biosynthesis of glycogen from uridine diphosphate glucose. Journal of American Chemistry Society, 79: 6340-6341.

Lerouxel, O., Cavalier, D. M., Liepman, A. H., and Keegstra, K. (2006). Biosynthesis of plant cell wall polysaccharides - a complex process. Current Opinion in Plant Biology, 9: 621-630.

Lexer, C., Fay, M. F., Joseph, J. A., Nica, M. S., and Heinze, B. (2005). Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. Molecular Ecology, 14: 1045-1057.

Li, L., Modolo, L. V., Luis, L., Escamilla-Trevino, L. L., Achnine, L., Dixon, R. A., and Wang, X. (2007). Crystal Structure of *Medicago truncatula* UGT85H2 -Insights into the Structural Basis of a Multifunctional (Iso)flavonoid Glycosyltransferase. Journal of Molecular Biology, 370: 951-963.

Li, Y., Baldauf, S., Lim, E. K., and Bowles, D. J. (2001). Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. Journal of Biological Chemistry, 276: 4338–4343.

Li, W., Zhang, F., Chang, Y., Zhao, Eric Schranz, T. M., and Wang, G. (2015). Nicotinate O-glucosylation is an evolutionarily metabolic trait important for seed germination under stress conditions in *Arabidopsis thaliana*. Plant Cell, 27: 1907-1924.

Liang, D. M., Liu, J. H., Wu, H., Wang, B. B., et al., (2015). Glycosyltransferases: mechanisms and applications in natural product development. Chemical Society Review, 44: 8350-8374.

Lim E-K., Ashford, D. A., Hou, B., Jackson R. G., Bowles, D. J. (2004). *Arabidopsis* glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. Biotechnology and Bioenergy, 87:623–31

Lim, E-K., and Bowles, D. J. (2004). A class of plant glycosyltransferases involved in cellular homeostasis. EMBO Journal, 23: 2915-2922.

Lim, E-K., Baldauf, S., Li. Y., Elias, L., Worrall, D., et al., (2003). Evolution of substrate recognition across a multigene family of glycosyltransferases in *Arabidopsis*. Glycobiology, 13:139–45

Lim, E-K., Doucet, C. J., Li, Y., Elias, L., Worrall, D. (2002). The activity of *Arabidopsis* glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. Journal of Biological Chemistry, 277: 586-592.

Lim, E-K., Jackson, R. G., and Bowles, D. J. (2005). Identification and characterisation of *Arabidopsis* glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde. FEBS Letters, 579: 2802–2806.

Lim, E-K., Li, Y., Parr, A., Jackson, R., et al., (2001). Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in *Arabidopsis*. Biological Chemistry, 276: 4344-4349.

Ljung, K., Hull, A. K., Kowalczyk, M., Marchant, A., Celenza, J., Cohen, J. D., and Sandberg, G. (2002). Biosynthesis, conjugation, catabolism and homeostasis of indole-3-acetic acid in *Arabidopsis thaliana*. Plant Molecular Biology, 505: 309–332.

Lloyd, G., and McCown, B. (1980). Commercially-feasible micropropagation of mountain laurel, Kalmia latifolia, by use of shoot-tip culture. Combined Proceedings, International Plant Propagators' Society, 30: 421-427.

Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. Science, 290: 1151-1155.

Mackenzie, P. I., Owens, I. S., Burchell, B., Bock, K. W., et al., (1997). The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. Pharmacogenetics, 7: 255-269.

Major, I. T., and Constabel, C. P. (2006). Molecular analysis of poplar defense against herbivory: comparison of wound- and insect elicitor-induced gene expression. New Phytologists, 172: 617-635.

Malik, V., and Black, G. W. (2012). Structural, functional and mutagenesis studies of UDP-lycosyltransferases. Advanced Protein Chemical and Structural Biology, 87: 87-115.

Mansfield, S. D. (2009). Solutions for dissolution-engineering cell walls for deconstruction. Current Opinion in Biotechnology, 20: 286-294.

Martin, R. C., Mok, M. C., and Mok, D. W. S. (1999a). A gene encoding the cytokinin enzyme zeatin O-xylosyltransferase of *Phaseolus vulgaris*. Plant Physiology, 120: 553-557.

Martin, R. C., Mok, M. C., and Mok, D. W. S. (1999b). Isolation of a cytokinin gene, ZOG1, encoding zeatin O-glucosyltransferase of *Phaseolus lunatus*. Proceeding of National Academy of Science, USA, 96: 284-289.

Martin, R. C., Mok, M. C., Habben, J. E., and Mok, D. W. S. (2001). A cytokinin gene from maize encoding an O-glucosyltransferase specific to cis-zeatin. Proceeding of National Academy of Science, USA, 98: 5922-5926.

Milkowski, C., Baumert, A., and Strack, D. (2000). Cloning and heterologous expression of a rape cDNA encoding UDP-glucose:sinapate glucosyltransferase. Planta, 211: 883–886.

Mithofer, A., and Boland, W. (2012). Plant defense against herbivores: chemical aspects. Annual Review of Plant Biology, 63: 431-450.

Modolo L. V., Blount, J. W., Achnine, L., Naoumkina, M. A., Wang, X., and Dixon, R. A. (2007). A functional genomics approach to (iso)flavonoid glycosylation in the model legume *Medicago truncatula*. Plant Molecular Biology, 64: 499-518.

Modolo, L. V., Li, L., Pan, H., Blount, J. W., et al., (2009). Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso) flavonoids. Journal of Molecular Biology, 392: 1292-1302.

Moehs, C. P., Allen, P. V., Friedman, M., and Belknap, W. R. (1997). Cloning and expression of solanidine UDP-glucose glucosyl-transferase from potato. Plant Journal, 11: 227–236.

Mok, D. W. S., and Mok, M. C. (2001). Cytokinin metabolism and action. Annual Review of Plant Physiol and Plant Molecular Biology, 52: 89-118

Morita, Y., Hoshino, A., Kikuchi, Y., and Okuhara, H. (2005). Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose:anthocyanidin 3-O-glucoside-2″-O-glucosyltransferase, due to 4-bp insertions in the gene . Plant Journal, 42: 353-363.

Morse, A. M., Tschaplinski, T. J., Dervinis, C., Pijut, P. M., et al., (2007). Salicylate and catechol levels are maintained in nahG transgenic poplar. Phytochemistry, 68: 2043-2052.

Mottiar, Y., Vanholme, R., Boerjan, W., Ralph, J., and Mansfield S. D. (2016) Designer lignins: harnessing the plasticity of lignification. Current Opinion in Biotechnology, 37:190-200.

Moubayidin, L., Mambro, R. D., Sabatini, S. (2009). Cytokinin-auxin crosstalk. Trends in Plant Science, 14: 557-562.

Nadeau, R., and Rappaport, L. (1974). An amphoteric conjugate of [h]gibberellin a (1) from barley aleurone layers. Plant Physiology, 54: 809-812.

Nambara, E., Marion-Poll, A. (2005). Abscisic acid biosynthesis and catabolism. Annual Review of Plant Biology, 56:165-185.

Offen, W., Martinez-Fleites, C., Yang, M., Lim, E-K., et al., (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO Journal, 25: 1396-1405.

Ono, E., Homma, Y., Horikawa, M., Kunikane-Doi, S., et al., (2010). Functional differentiation of the glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (*Vitis vinifera*). Plant Cell, 22: 2856-2871.

Osbourn A. E. (2003) Saponins in cereals. Phytochemistry, 62:1-4.

Osier, T. L., and Lindroth, R. L. (2001). Effects of genotype, nutrient availability, and defoliation on aspen phytochemistry and insect performance. Journal of Chemical Ecology, 27: 1289-1313.

Osmani, S. A., Bak, S., and Moller, B. L. (2009). Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry, 70: 325-347.

Ostrowski, M., and Jakubowska, A. (2013). GH3 expression and IAA-amide synthetase activity in pea (*Pisum sativum* L.) seedlings are regulated by light, plant hormones and auxinic herbicides. Journal of Plant Physiology, 170: 361-368.

Paquette, S., Moller, B. L., and Bak, S. (2003). On the origin of family 1 plant glycosyltransferases. Phytochemistry, 62: 399-413.

Payyavula, R. S., Babst, B. A., Nelsen, M. P., Harding, S. A, and Tsai, C. J. (2009). Glycosylation-mediated phenylpropanoid partitioning in *Populus tremuloides* cell cultures. BMC Plant Biology, 9:151. [doi: 10.1186/1471-2229-9-151]

Pearce, D.W., Rood, S. B. and WU, R. (2004). Phytohormones and shoot growth in a three-generation hybrid poplar family. Tree Physiology, 24: 217-224.

Peer, W. A., and Murphy, A. S. (2007). Flavonoids and auxin transport: modulators or regulators? Trends in Plant Science, 12: 556–563.

Peer, W. A., Bandyopadhyay, A., Blakeslee, J., Makam, N., et al., (2004). Variation in expression and protein localization of the PIN family of auxin efflux facilitator proteins in flavonoid mutants with altered auxin transport in *Arabidopsis thaliana*. Plant Cell, 16: 1898-1911.

Peer. W. A., Blakeslee, J. J., Yang, H., and Murphy, A.S. (2011). Seven things we know about Auxin transport. Molecular Plant, 4: 487-505.

Perrin, R. M. (2008). Glycosyltransferases in Plant Cell Wall Synthesis. In: eLS. John Wiley & Sons Ltd, Chichester. http://www.els.net [doi: 10.1002/9780470015902.a0020102]

Pharis, R.P., Yeh F.C., and Dancik, B.P. (1991). Superior growth potentialin trees: what is its basis, and can it be tested for at an early age?. Canadian Journal of Forest Research, 21: 368-374.

Pineda, A. R., Brugie, N., Vankova R., Malbeck, J. M., et al., (2008). Over-expression of a Zeatin O-glucosylation gene in maize leads to growth retardation and tassel seed formation. Journal of Experimental Botany, 59: 2673-2686.

Piotrowska, A., and Bajguz, A. (2011). Conjugates of abscisic acid, brassinosteroids, ethylene, gibberellins, and jasmonates. Phytochemistry, 72: 2097-2112.

Poppenberger, B., Berthiller, F., Lucyshyn, D., Sieberer, T., et al., (2003). Detoxification of the Fusarium mycotoxin deoxynivalenol by a UDP-glucosyltransferase from *Arabidopsis thaliana*. Journal Biological Chemistry, 278: 47905-47914.

Poppenberger, B., Fujioka, S., Soeno, K., George, G. L., et al., (2005). The UGT73C5 of *Arabidopsis thaliana* glucosylates brassinosteroids. Proceeding of National Academy of Science, USA, 102: 15253-15258.

Priest, D. M., Ambrose, S. J., Vaistij, F. E., Elias, L., Higgins, G. S., Ross, A. R., Abrams, S. R., and Bowles, D. J. (2006). Use of the glucosyltransferase UGT71B6 to disturb abscisic acid homeostasis in *Arabidopsis thaliana*. Plant Journal, 46: 492-502.

Priest, D. M., Jackson, R. G., Ashford, D. A., Abrams, S. R., and Bowles, D. J. (2005). The use of abscisic acid analogues to analyse the substrate selectivity of UGT71B6, a UDP-glycosyltransferase of *Arabidopsis thaliana*. FEBS Letters, 579: 4454-4458.

Quiel, J. A., and Bender, J. (2003). Glucose conjugation of anthranilate by the *Arabidopsis* UGT74F2 glucosyltransferase is required for tryptophan mutant blue fluorescence. Journal of Biological Chemistry, 278: 6275-6281.

Ragni, L., Nieminen, K., David Pacheco-Villalobos, D., Sibout C., Schwechheimer, R., and Hardtkea, C. S. (2011). Mobile gibberellin directly stimulates *Arabidopsis* hypocotyl xylem expansion. The Plant Cell, 23: 1322–1336.

Rampey, R. A., LeClere, S., Kowalczyk, M., Ljung, K., et al., (2004). A family of auxin-conjugate hydrolases that contributes to free indole-3-acetic acid levels during *Arabidopsis* germination. Plant Physiology, 135: 978-988.

Rood, S. B., Goater, L. A., Mahoney, J. M., Pearce, C. M., et al., (2007). Flood, fire, and ice: disturbance ecology of riparian cottonwoods. Canadian Journal of Botany, 85: 1019-1032.

Ross, J., Li, Y., Lim, EK., Bowles, D. J. (2001). Higher plant glycosyltransferases. Genome Biology, 2: 3004-3011.

Ruuhola, T. M., Sipura, M., Nousiainen, O., and Tahvanainen, J. (2001a). Systemic induction of salicylates in *Salix myrsinifolia*. Annals Botany, 88: 483-497.

Ruuhola, T., Julkunen-Tiitto, R. (2003). Trade-off between synthesis of salicylates and growth of micropropagated *Salix pentandra*. Journal of Chemical Ecology, 29: 1565-1588.

Ruuhola, T., Tikkanen, O. P., and Tahvanainen, J. (2001b). Differences in host use efficiency of larvae of a generalist moth, *Operophtera brumata* (*Lepidoptera*: *Geometridae*) on three chemically divergent Salk species. Journail of Chemical Ecology, 27: 1595-1615.

Salyaev, R., Rekoslavskaya, N., Chepinoga, A., Mapelli, S., and Pacovsky, R., (2006). Transgenic poplar with enhanced growth by indroduction of ugt and acb genes. New Forests, 32: 211-229.

Sakakibara,Y. K., and Hanada, K. (2011). An evolutionary view of functional diversity in family 1 glycosyltransferases. The Plant Journal, 66: 182-193.

Sandrock, R. W., Van-Etten, H. D. (1998). Fungal sensitivity to and enzymatic degradation of the phytoanticipin alpha-tomatine. Phytopathology, 88: 137-43.

Santelia, D., Henrichs, S., Vincenzetti, V., Sauer, M., Bigler, L., et al., (2008). Flavonoids redirect PIN-mediated polar auxin fluxes during root gravitropic responses. Journal of Biological Chemistry, 283: 31218-31226.

Shah, J., Kachroo, P., and Klessig, D. F. (1999). The *Arabidopsis ssi1* mutation restores pathogenesis-related gene expression in npr1 plants and renders defensin gene expression salicylic acid dependent. Plant Cell, 11: 191-206.

Shao, H., He, X., Achnine, L., Blount, J. W., et al., (2005). Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. Plant Cell, 17: 3141-3154.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., et al., (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular System Biology, 7: 539.

Sinnot, M.L. (1990). Catalytic mechanisms of enzymatic glycosyltransfer. Chemical Review, 90: 1171-1202.

Slavov, G. T., Leonardi, S., Burczyk, J., Adams, W. T., et al., (2009). Pollen flow in two ecologically contrasting populations of *Populus trichocarpa*. Molecular Ecology, 18: 357-373.

Smulders, M. J. M., Cottrell, J. E., Lefèvre, F., Van-Der Schoot. J., et al., (2008). Structure of the genetic diversity in black poplar (*Populus nigra* L.) populations across European river systems: Consequences for conservation and restoration. Frontiers Ecological Management, 255: 1388-1399.

Song, J. T. (2005). Biochemical characterization of an *Arabidopsis* glucosyltransferase with high activity toward jasmonic acid. Journal of Plant Biology, 48: 422–428.

Song, J. T., Koo, Y. J., Seo, H. S., Kim, M. C., Do Choi, Y., & Kim, J. H. (2008). Overexpression of AtSGT1, an *Arabidopsis* salicylic acid glucosyltransferase, leads to increased susceptibility to *Pseudomonas syringae*. Phytochemistry, 69: 1128-1134.

Sterling, J. D., Atmodjo, M.A., Inwood, S.E., Kumar E., Kolli, V. S., et al., (2006). Functional identification of an *Arabidopsis* pectin biosynthetic homogalacturonan galacturonosyltransferase. Proceeding of National Academy of Science, USA, 103: 5236-5524.

Stoddart, J. L., and Venis, M. A. (1980). Molecular and Subcellular Aspects of Hormone Action. In Hormonal regulation of development. I, Molecular Aspects of Plant Hormones. Encyclopedia of Plant Physiology, New series, 9: 445-510.

Sultan, S. E. (2000) Phenotypic plasticity for plant development, function and life history. Trends in Plant Science, 5: 537-542.

Szerszen, J. B., Szczyglowski, K., and Bandurski, R. S. (1994). *iaglu*, a gene from Zea mays involved in conjugation of growth hormone indole-3-acetic acid. Science, 265: 1699-1701.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Molecular Biology and Evolution, 28: 2731-2739.

Tanaka, K., Hayashi, K., Natsume, M., Kamiya, Y., Sakakibara, H., et al., (2014). UGT74D1 catalyzes the glucosylation of 2-oxindole-3-acetic acid in the auxin metabolic pathway in *Arabidopsis*. Plant Cell Physiology, 55: 218-228.

Tanaka, T., Antonio, B. A., Kikuchi, S., Matsumoto, T., et al., (2008). The Rice Annotation Project Database (RAP-DB): 2008 update. Nucleic Acids Research, 36: D1028-D1033.

Taylor, G. (2002). *Populus*: *Arabidopsis* for forestry. Do we need a model tree?. Annals Botany, 90: 681-689.

Taylor, L. P., and Grotewold, E. (2005). Flavonoids as developmental regulators. Current Opinions in Plant Biology, 8: 317-323.

The *Arabidopsis* Initiative: Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. (2000). Nature, 408:796-815.

Thompson, A. M. G., Iancu, C. V., Dean, J., and Choe J. (2016). Structural basis of distinct salicylic acid glucosylation in Arabidopsis thaliana by two homologous enzymes: implications for plant stress response. FASEB Journal, 30: no. 1 Supplement 1142.3.

Tognetti, V, B., van-Aken, O., Morreel, K., Vandenbroucke, K., et al., (2010). Perturbation of indole-3-butyric acid homeostasis by the UDP-glucosyltransferase UGT74E2 modulates *Arabidopsis* architecture and water stress tolerance. Plant Cell, 22: 2660-2679.

Tohge, T., Nishiyama, Y., Hirai, M. Y., Yano, M., et al., (2005). Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant Journal, 42: 218-235.

Treutter, D. (2005). Significance of flavonoids in plant resistance and enhancement of their biosynthesis. Plant Biology, 7: 581-591.

Tsai, C. J., Harding, S. A., Tschaplinski, T. J., Lindroth, R. L., and Yuan, Y. (2006). Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus* . New Phytologist, 172: 47-62.

Tsai, C. J., Guo, W., Babst B., Nyamdari B., Yuan, Y., Payyavula, R., et al., (2011). Salicylate metabolism in *Populus*. . BMC Proceedings, 2011: 5 (Suppl 7):I9.

Tuominen, H., Puech, L., Fink S., and Sundberg, B. (1997). A radial concentration gradient of indole-3-acetic acid is related to secondary xylem development in hybrid aspen. Plant Physiology, 115: 577-585.

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., et al., (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science, 313: 1596-1604.

Unda, F., Canam, T., Preston, L., and Mansfield, S. D. (2012). Isolation and characterization of galactinol synthases from hybrid poplar. Journal of Experimental Botany, 63: 2059-2069.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., et al., (2012). Primer3 - new capabilities and interfaces. Nucleic Acids Research, 40: e115.

Veit, M., and Pauli, G. F. (1999). Major Flavonoids from *Arabidopsis thaliana* Leaves. Journal of Natural Products, 62: 1301–1303.

Veljanovski, V., and Constabel, C. P. (2013). Molecular cloning and biochemical characterization of two UDP-glycosyltransferases from poplar. Phytochemistry, 91: 148-157.

Vogt, T. (2002). Substrate specificity and sequence analysis define a polyphyletic origin of betanidin 5- and 6-O-glucosyltransferase from *Dorotheanthus bellidiformis*. Planta, 214: 492-495.

Vogt, T. (2010). Phenylpropanoid biosynthesis. Molecular Plant, 3: 2-20.

Vogt, T., and Jones, P. (2000). Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends in Plant Science, 5: 380-386.

von-Saint Paul, V. (2010). Stress inducible glycosyltransferases in *Arabidopsis thaliana* and their impact on plant metabolism and defense mechanisms. PhD thesis. Faculty of pharmacy, University of Munich.

Wang B., Jin, S. H., Hu, H. Q., Sun, Y. G., Wang, Y. et al., (2012b). UGT87A2, an *Arabidopsis* glycosyltransferase, regulates flowering time via FLOWERING LOCUS C. New Phytologists, 194: 666-675.

Wang Y. W., Wang, W. C., Jin, S. H., Wang, J., Wang, B., and Hou, B. K. (2012a). Over-expression of a putative poplar glycosyltransferase gene, PtGT1, in tobacco increases lignin content and causes early flowering. Journal of Experimental Botany, 63: 2799-2808.

Wang, F. B., Zhu, H., Chen, D. H., Li, Z. J., Peng, R. H., and Yao, Q. H. (2016). A grape bHLH transcription factor gene, VvbHLH1, increases the accumulation of flavonoids and enhances salt and drought tolerance in transgenic *Arabidopsis thaliana*. Plant Cell Tissue and Organ [doi:10.1007/s11240-016-0953-1].

Wang, J., and Hou, B. (2009). Glycosyltransferases: key players involved in the modification of plant secondary metabolites. Frontiers of Biology in China, 4: 39-46.

Wang, J., Ma X. M., Kojima, M., Sakakibara, H., and Hou, B. K. (2013). Glucosyltransferase UGT76C1 finely modulates cytokinin responses via cytokinin N-glucosylation in *Arabidopsis thaliana*. Plant Physiology and Biochemistry, 65: 9-16.

Wang, X. (2009). Structure, mechanism and engineering of plant natural product glycosyltransferases. FEBS Letters, 583: 3303–3309.

Weijers, C. A. G. M., Franssen, M. C. R., and Visser, G. M. (2008). Glycosyltransferase-catalyzed synthesis of bioactive oligosaccharides. Biotechnology Advances, 26: 436-456.

Williams, C. A., and Grayer, R. J. (2004). Anthocyanins and other flavonoids. Natural Products Report, 21: 539-573.

Woo, H. H., Jeong, B. R., Hirsch, A. M., and Hawes, M. C. (2007). Characterization of *Arabidopsis AtUGT85A* and *AtGUS* gene families and their expression in rapidly dividing tissues. Genomics, 90: 143-153.

Woodward, A. W., and Bartel, B. (2005). Auxin: regulation, action, and interaction. Annals Botany, 95: 707-735.

Xu, Z. J., Nakajima, M., Suzuki, Y., and Yamaguchi, I. (2002). Cloning and characterization of the abscisic acid-specific glucosyltransferase gene from adzuki bean seedlings. Plant Physiology, 129: 1285-1295.

Yin, R., Han, K., Heller, W., Albert, A., Dobrev, P. I., et al., (2014). Kaempferol 3-O-rhamnoside-7-O-rhamnoside is an endogenous flavonol inhibitor of polar auxin transport in *Arabidopsis* shoots. New Phytologists, 201: 466-475.

Yin, R., Messner, B., Faus-Kessler, T., Hoffmann, T., Schwab, W., et al., (2012). Feedback inhibition of the general phenylpropanoid and flavonol biosynthetic pathways upon a compromised flavonol-3-O-glycosylation. Journal of Experimental Botany, 63: 2465-2478.

York, W. S., O'Neill, M. A. (2008). Biochemical control of xylan biosynthesis - which end is up?. Current Opinions in Plant Biology, 11: 258-265.

Zhong, J. J., Yue, C.J. (2005). Plant cells: secondary metabolite heterogeneity and its manipulation. Advances in Biochemical Engineering and Biotechnology, 100: 53-88.

Zhou, G. K., Zhong, R., Himmelsbach, D. S., Mc Phail, B. T., and Ye, Z. H. (2007). Molecular characterization of PoGT8D and PoGT43B, two secondary wall-associated glycosyltransferases in poplar. Plant and Cell Physiology, 48: 689-699.

Appendix

Nutrient solution used for watering poplar trees

Poplar trees were grown in greenhouse on flooding tables, and trees were watered with nutrient solution containing macro and micro nutrients with the following molecular composition: E.C. 1.6, NO3 4.9 mM/l, Cl 0.3 mM/l, S <0.2 mM/l, HCO3 3.2 mM/l, P 2.76 mM/l, NH4 8.0 mM/l, K 4.1 mM/l, Na 0.3 mM/l, Ca 0.3 mM/l, Mg <0.2 mM/l, Si <0.10 mM/l, Fe 17.1 µM/l, Mn 7.7 µM/l, Zn 8.2 µM/l, B 33 µM/l, Cu 8.4 µM/l, Mo <0.2 µM/l. (pH 7.6).

Table 1A: Percent increase in the fibre lumen diameter, wood microfibre angle (MFA), wood density (kg/m^3), and Youngs Modulus Elasticity for two transgenic lines (a) RNAi-*PopGT1* and (b) RNAi-*PopGT2* compared to wild-type trees. Standard error values are represented in brackets, bold values depict the significantly different means, using *t*-test at $\alpha = 0.05$ (n = 5).

a)	**% Increase in fibre lumen diameter**	**MFA (^{0})**	**Wood density (kg/m^3)**
WT		16.3 (1.20)	255.06 (4.93)
A5	21.8 (1.85)	15.5 (0.81)	**217.66 (6.61)**
A9	25.3 (3.20)	15.1 (0.97)	**230.63 (7.54)**
b)			
	% Increase in fibre lumen diameter	**MFA (^{0})**	**Wood density (kg/m^3)**
WT		16.3 (1.20)	255.06 (4.93)
B5	17.5 (2.50)	15.6 (0.90)	**222.47 (4.93)**
B11	19.2 (1.76)	14.9 (1.55)	**225.73 (6.88)**

Table 2A: NMR profile of salireposide

	C	CH	CH2	CH3	H	HMBC	COSY
1		130.5			2H, 8.07, d, 7.7Hz	130.5, 134.2, 167.8	7.49
2		134.2			1H, 7.61,t, 7.93 Hz	130.5	8.07, 7.49
3		129.5			2H, 7.49, t, 7.4 Hz	129.5, 131.3	7.61
4		119.1			1H, 7.11, d, 8.7 Hz	128.6, 149.8, 153.9	6.72
5		116.0			1H, 6.85, d, 2.3 Hz	149.8	6.72
6		116.4			1H, 6.72, dd, 2.3, 8.7 Hz	149.8	6.85, 7.11
7			63.0		1H, 5.51, d, 12.8 Hz	116.0, 128.6, 149.8, 167.8	5.42
8			63.0		1H, 5.42, d, 12.8 Hz	116.0, 128.6, 149.8, 167.8	5.51
9		104.2			1H, 4.77, d, 7.2 Hz	149.8	3.43
10			62.4		1H, 3.87, d, 12 Hz		3.68
11			62.4		1H, 3.68, d, 12 Hz	77.8	3.87
12		74.8			1H, 3.44, m		
13		77.8			1H, 3.43, m		
14		71.2			1H, 3.36, m		
15		78.0			1H, 3.36, m		
16	167.8						
17	131.3						
18	128.6						
19	149.8						
20	153.9						

Table 3A: **Normalized peak area (mAU min) of the total phenolic metabolites (non-hydrolyzed) of six RNAi-*PopGT1* and wild-type control poplar trees. Bold numbers indicate the significant differences in metabolite normalized peak areas between the transgenic lines and the wild-type controls, using *t*-test at α = 0.05, Numbers in brackets represent the standard error of means (n = 4). RT is the retention time value for each compound**

RT	WT	A1	A2	A3	A4	A5	A9
11.48	22.49 (6.43)	16.61 (9.76)	31.78 (8.79)	26.25 (8.98)	11.89 (5.89)	26.20 (6.67)	22.49 (6.55)
12.41	20.80 (4.87)	19.33 (2.87)	18.25 (5.00)	26.36 (2.63)	14.87 (2.74)	22.93 (2.94)	26.24 (1.37)
13.45	14.22 (4.32)	11.84 (3.39)	16.74 (4.85)	15.73 (6.14)	8.87 (2.46)	9.91 (1.04)	8.34 (0.44)
14.98	4.95 (0.63)	10.29 (5.43)	4.76 (0.24)	5.11 (0.60)	7.91 (1.88)	5.08 (0.94)	10.51 (3.97)
15.54	6.77 (1.22)	6.52 (1.18)	6.92 (0.88)	7.89 (1.21)	5.60 (0.68)	6.21 (0.34)	5.89 (0.95)
16.28	28.53 (4.75)	20.17 (4.64)	25.08 (1.26)	23.18 (3.05)	19.63 (5.47)	21.08 (3.15)	20.39 (0.28)
17.25	7.91 (0.79)	5.79 (1.96)	8.71 (0.81)	6.52 (1.44)	5.93 (1.41)	7.35 (0.51)	9.27 (0.37)
17.53	6.30 (0.41)	5.31 (1.41)	7.02 (0.69)	7.35 (1.39)	4.44 (0.81)	4.92 (0.34)	6.89 (0.13)
18.23	8.54 (2.47)	5.32 (1.83)	11.35 (5.54)	8.41 (3.44)	3.42 (0.89)	6.64 (1.24)	4.99 (0.61)
18.75	3.08 (0.60)	2.22 (0.85)	3.67 (0.65)	3.61 (0.96)	1.71 (0.42)	2.63 (0.66)	2.95 (0.15)
18.99	2.40 (0.48)	1.71 (0.24)	2.28 (0.41)	2.30 (0.61)	1.66 (0.26)	1.64 (0.48)	1.91 (0.23)
20.00	1.92 (0.69)	0.81 (0.38)	2.07 (0.36)	1.94 (0.76)	1.12 (0.41)	1.52 (0.46)	0.84 (0.07)
20.59	0.57 (0.11)	0.32 (0.08)	0.77 (0.56)	0.98 (0.48)	0.54 (0.12)	0.69 (0.08)	0.65 (0.01)
21.15	1.42 (0.17)	1.16 (0.27)	1.97 (0.27)	1.75 (0.41)	0.85 (0.40)	1.69 (0.14)	1.98 (0.09)
21.56	0.84 (0.12)	0.64 (0.16)	1.85 (1.03)	0.79 (0.26)	1.59 (0.98)	1.63 (0.94)	0.86 (0.12)
21.69	3.37 (0.26)	2.21 (0.66)	3.25 (0.60)	3.18 (0.99)	2.34 (0.21)	2.72 (0.64)	3.99 (0.35)
22.16	2.76 (0.09)	2.21 (0.37)	3.13 (0.68)	2.19 (0.58)	2.18 (0.29)	3.80 (0.13)	4.58 (0.53)
22.56	0.44 (0.06)	0.30 (0.06)	0.65 (0.16)	1.04 (0.55)	0.64 (0.27)	0.52 (0.08)	0.90 (0.29)
23.02	1.04 (0.14)	0.86 (0.29)	1.48 (0.19)	0.97 (0.23)	0.87 (0.27)	0.92 (0.32)	1.40 (0.12)
23.35	3.08 (0.29)	2.88 (0.75)	4.73 (**0.39**)	2.48 (0.96)	3.32 (0.41)	2.78 (0.91)	4.31 (0.26)
23.59	0.55 (0.08)	0.41 (0.09)	0.91 (0.15)	1.53 (1.06)	0.68 (0.18)	1.63 (0.91)	0.77 (0.12)
23.93	0.97 (0.15)	0.61 (0.1)	1.13 (0.11)	0.64 (0.19)	0.68 (0.09)	0.94 (0.12)	0.97 (0.07)
24.04	2.30 (0.25)	1.76 (0.30)	3.16 (0.09)	2.42 (0.76)	2.52 (0.36)	2.86 (0.28)	2.80 (0.28)
24.30	1.00 (0.22)	0.64 (0.22)	1.90 (0.30)	1.08 (0.46)	1.14 (0.17)	1.43 (0.20)	1.44 (0.11)
24.69	1.57 (0.10)	1.14 (0.23)	2.32 (**0.18**)	1.71 (0.59)	1.91 (0.26)	2.26 (0.16)	2.07 (0.20)
25.21	2.38 (0.42)	1.35 (0.17)	2.75 (0.65)	2.77 (0.97)	1.74 (0.21)	2.16 (0.10)	3.14 (0.16)

RT	WT	A1	A2	A3	A4	A5	A9
25.78	6.21 (0.35)	**3.60 (0.72)**	**3.02 (0.60)**	**4.57 (0.38)**	**2.58 (0.76)**	**3.14 (0.37)**	**4.26 (0.22)**
25.97	27.52 (10.76)	12.27 (5.23)	16.52 (2.14)	14.85 (3.62)	9.76 (4.06)	15.42 (1.95)	15.07 (1.71)
27.14	5.97 (1.28)	5.31 (1.37)	6.44 (1.22)	7.23 (1.30)	7.53 (0.95)	**9.18 (0.81)**	8.18 (0.96)
27.85	1.63 (0.14)	1.77 (0.53)	3.11 (0.69)	3.01 (0.47)	2.20 (0.58)	**3.30 (0.21)**	4.15 (0.94)
28.13	6.85 (0.58)	5.40 (1.69)	5.96 (2.02)	6.32 (0.63)	**3.04 (0.64)**	5.84 (0.67)	3.98 (0.48)
28.47	4.73 (0.32)	3.95 (1.20)	3.86 (0.21)	5.45 (0.31)	**2.63 (0.20)**	4.17 (0.67)	4.04 (0.40)
28.81	4.45 (0.39)	4.15 (1.38)	3.44 (1.15)	3.33 (1.42)	3.45 (0.52)	4.79 (0.21)	4.14 (0.54)
29.43	0.82 (0.14)	0.79 (0.44)	2.23 (0.47)	1.68 (0.29)	1.69 (0.94)	1.24 (0.16)	0.97 (0.29)
29.63	6.52 (0.32)	4.47 (1.50)	4.72 (4.06)	5.17 (2.38)	3.19 (1.11)	6.34 (0.38)	5.29 (0.63)
29.87	2.23 (0.26)	2.39 (0.80)	4.39 (0.84)	3.89 (0.71)	2.27 (0.90)	2.87 (0.20)	2.37 (0.48)
30.05	1.01 (0.18)	0.89 (0.50)	2.47 (0.74)	1.41 (0.73)	1.31 (0.68)	1.01 (0.17)	0.63 (0.09)
30.62	4.27 (0.70)	4.31 (1.17)	**7.96 (0.77)**	5.11 (0.52)	2.77 (0.47)	3.92 (0.29)	3.55 (0.40)
30.81	4.36 (0.82)	4.75 (1.26)	5.67 (1.80)	4.42 (0.39)	2.16 (0.43)	4.16 (0.47)	4.18 (0.08)
31.08	4.52 (0.71)	4.65 (1.86)	3.04 (0.92)	6.34 (2.17)	2.03 (0.65)	6.26 (0.42	4.31 (1.18)
31.40	2.38 (0.26)	1.43 (0.60)	**4.16 (0.34)**	2.18 (0.36)	1.29 (0.40)	1.99 (0.08)	2.24 (0.28)
31.71	3.51 (0.32)	3.24 (1.09)	**0.96 (0.35)**	3.73 (0.72)	3.19 (1.08)	4.26 (0.75)	5.30 (0.23)
31.98	1.10 (0.14)	1.05 (0.39)	1.03 (0.56)	1.67 (0.31)	1.42 (0.48)	1.24 (0.27)	0.79 (0.34)
32.73	0.56 (0.14)	0.70 (0.32)	2.21 (0.83)	1.12 (0.43)	1.40 (0.50)	1.77 (0.12)	1.11 (0.53)
33.03	7.60 (0.29)	7.04 (1.75)	6.00 (0.90)	8.76 (1.06)	5.91 (1.18)	8.04 (0.48)	5.87 (0.86)
33.48	1.03 (0.31)	1.34 (0.32)	1.56 (0.81)	8.27 (5.63)	1.42 (0.14)	2.19 (0.17)	1.30 (0.66)
33.96	18.61 (1.13)	16.94 (3.03)	15.74 (1.07)	12.82 (5.60)	11.24 (2.71)	15.00 (0.35)	**13.35 (0.48)**
34.43	3.60 (0.72)	4.28 (1.83)	3.56 (0.64)	6.77 (1.37)	1.92 (0.55)	3.73 (0.51)	2.59 (0.68)
34.89	11.76 (2.82)	14.08 (5.72)	8.75 (1.61)	9.81 (4.22)	4.30 (0.57)	8.05 (0.94)	7.93 (0.70)
35.53	8.94 (2.73)	10.68 (3.79)	11.45 (3.55)	13.92 (5.10)	5.12 (1.89)	9.70 (0.98)	7.94 (1.43)
35.68	4.10 (0.27)	5.25 (1.44)	3.71 (0.42)	4.25 (1.25)	3.11 (0.81)	7.14 (3.59)	3.03 (0.46)
36.53	8.78 (0.52)	7.66 (2.27)	6.75 (2.46)	10.39 (2.20)	5.16 (3.04)	9.13 (0.46)	7.57 (1.42)
36.68	12.13 (0.93)	12.30 (4.00)	8.65 (1.37)	10.35 (1.82)	4.37 (1.39)	23.30 (15.44)	**7.93 (0.54)**
37.15	1.81 (0.11)	1.93 (0.37)	3.87 (1.79)	2.53 (0.93)	2.25 (0.80)	2.82 (0.78)	**1.04 (0.14)**
38.42	1.36 (0.33)	2.18 (0.29)	1.68 (0.75)	20.16 (17.01)	1.91 (0.93)	2.33 (0.29)	1.37 (0.22)
39.10	3.43 (0.63)	3.64 (0.61)	3.46 (1.08)	3.77 (1.34)	2.96 (1.29)	6.27 (2.29)	4.01 (0.57)
39.42	0.92 (0.05)	1.30 (0.24)	1.21 (0.21)	1.80 (0.20)	1.53 (0.39)	2.04 (0.97)	2.95 (1.46)
39.84	4.37 (1.29)	**6.84 (0.75)**	9.13 (1.71)	9.33 (4.61)	4.32 (2.47)	5.08 (1.51)	2.22 (0.45)
40.43	0.83 (0.02)	**2.99 (0.68)**	**2.06 (0.33)**	**2.66 (0.46)**	1.86 (0.30)	**4.28 (2.21)**	0.86 (0.54)

RT	WT	A1	A2	A3	A4	A5	A9
40.63	0.78 (0.16)	2.37 (0.49)	1.52 (0.52)	0.67 (0.24)	1.54 (0.54)	2.41 (0.58)	0.78 (0.24)
43.88	2.21(0.66)	4.30 (1.61)	2.40 (1.00)	2.38 (0.48)	1.56 (0.77)	2.95 (0.96)	2.27 (0.57)
44.54	3.85 (0.72)	4.92 (1.37)	4.28 (0.64)	3.99 (0.18)	3.22 (0.71)	3.20 (1.21)	5.45 (0.88)
46.92	63.99 (14.91)	45.08 (7.97)	47.15 (13.76)	41.37 (10.47)	37.33 (8.40)	40.09 (3.79)	55.11 (1.80)
47.12	57.87 (9.91)	50.17 (20.26)	59.31 (6.76)	58.77 (8.00)	43.14 (6.77)	46.68 (3.75)	56.14 (11.07)
47.43	14.69 (1.89)	23.72 (6.89)	13.41 (1.78)	28.81 (11.69)	11.87 (1.29)	13.65 (0.20)	13.55 (1.52)
47.78	42.18 (8.72)	33.24 (14.39)	44.76 (6.93)	31.40 (11.15)	33.60 (5.50)	38.20 (2.46)	46.85 (9.31)
53.11	9.44 (1.16)	15.24 (5.35)	10.55 (2.43)	20.05 (9.71)	12.18 (3.43)	8.33 (0.30)	9.96 (0.75)
57.00	21.78 (1.91)	14.90 (5.97)	20.68 (2.12)	16.74 (5.36)	12.78 (3.53)	20.54 (0.37)	23.48 (4.60)

Table 4A: **Normalized peak area (mAU) of the total phenolic metabolites (non-hydrolyzed) of six RNAi-*PopGT2* and wild-type control poplar trees. Bold numbers indicate the significant differences in metabolite peak areas between transgenic lines and wild-type controls, using *t*-test at $\alpha = 0.05$. Numbers in brackets represent the standard error of means, (n = 4). RT is the retention time value for each compound**

RT	WT	B1	B2	B3	B4	B5	B11
11.68	7.33 (2.47)	6.83 (1.72)	6.55 (1.31)	5.22 (1.68)	5.69 (2.45)	7.52 (2.58)	6.36 (2.28)
12.52	19.44 (5.82)	18.02 (1.43)	13.98 (3.50)	14.21 (2.74)	8.82 (4.43)	13.14 (4.15)	19.88 (6.51)
13.63	63.30 (11.27)	48.48 (7.59)	59.99 (8.67)	52.74 (22.16)	41.00 (4.44)	58.30 (24.21)	35.69 (8.39)
15.75	10.08 (1.24)	9.04 (1.93)	10.59 (1.62)	6.09 (3.14)	9.95 (1.42)	6.78 (2.04)	6.67 (2.19)
16.46	38.49 (3.10)	25.61 (1.03)	29.03 (1.69)	15.92 (2.58)	24.83 (2.21)	20.28 (4.20)	13.24 (2.52)
17.45	8.44 (0.49)	8.09 (2.42)	7.08 (0.46)	18.16 (6.47)	5.46 (1.69)	14.29 (9.64)	20.72 (8.11)
17.71	9.71 (1.31)	7.70 (0.84)	8.04 (1.18)	7.09 (0.56)	7.62 (1.11)	5.54 (0.58)	5.73 (0.19)
18.37	25.24 (10.20)	31.05 (5.20)	34.02 (6.91)	17.42 (10.05)	22.83 (3.51)	17.43 (10.03)	9.11 (7.63)
18.88	4.64 (0.81)	3.36 (0.50)	4.31 (0.21)	18.61 (12.98)	4.04 (0.34)	14.02 (10.94)	17.33 (7.68)
19.15	2.51 (0.93)	1.63 (0.36)	1.81 (0.03)	3.05 (1.03)	2.10 (0.13)	2.13(0.51)	3.73 (1.11)
19.47	2.75 (0.54)	2.65 (0.85)	3.68 (1.02)	2.48 (0.23)	2.07 (0.44)	2.92 (1.21)	2.57 (0.22)
20.19	5.28 (0.82)	3.54 (1.23)	5.97 (1.04)	3.76 (1.26)	3.69 (0.30)	4.19 (1.95)	6.25 (3.51)
20.78	2.84 (1.44)	5.07 (2.04)	3.04 (0.25)	3.23 (1.56)	1.62 (0.32)	2.64 (0.78)	1.90 (0.87)
21.29	1.40 (0.24)	1.24 (0.12)	1.44 (0.11	2.27 (0.99)	1.26 (0.15)	1.38 (0.52)	1.34 (0.59)
21.85	2.88 (0.71)	2.81 (0.15)	3.30 (0.33)	2.22 (0.58)	2.53 (0.32)	1.79 (0.19)	2.10 (0.41)
22.33	1.80 (0.55)	2.43 (0.43)	2.01 (0.15)	2.60 (0.46)	1.66 (0.16)	1.34 (0.57)	2.92 (0.87)
23.17	0.73 (0.37)	1.37 (0.42)	1.03 (0.16)	1.48 (0.35)	1.17 (0.08)	1.00 (0.08)	1.08 (0.50)
23.43	1.59 (0.30)	3.69 (0.94)	2.61 (0.32)	2.63 (0.39)	2.46 (0.19)	2.22 (0.21)	1.52 (0.55)
24.14	2.48 (0.40)	2.52 (0.79)	2.75 (0.73)	2.79 (0.51)	3.52 (0.22)	2.73 (0.42)	2.59 (0.49)
24.46	2.43 (0.89)	1.43 (0.34)	1.40 (0.30)	2.00 (0.66)	1.68 (0.29)	1.11 (0.26)	3.06 (0.50)
24.87	1.09 (0.84)	0.81 (0.23)	0.85 (0.23)	1.09 (0.45)	1.25 (0.29)	0.96 (0.38)	1.64 (0.44)
25.48	3.80 (1.88)	4.57 (1.18)	3.34 (0.97)	1.97 (0.22)	3.66 (0.24)	4.01 (1.03)	2.89 (1.25)
26.08	35.85 (2.39)	25.80 (8.32)	30.18 (1.79)	14.32 (2.91)	17.76 (2.54)	27.08 (9.94)	19.81 (5.11)
27.06	1.12 (0.54)	2.58 (0.95)	1.92 (0.41)	1.21 (0.14)	0.82 (0.14)	0.58 (0.26)	1.63 (0.27)
27.28	4.20 (0.68)	3.20 (0.83)	3.40 (0.82)	3.32 (0.69)	3.43 (0.73)	3.22 (0.58)	4.13 (1.21)
28.15	7.17 (1.43)	8.89 (0.62)	9.26 (0.49)	6.26 (1.34)	6.07 (1.32)	6.88 (1.92)	7.00 (1.32)
28.49	4.63 (1.20)	5.22 (0.10)	3.97 (0.80)	5.59 (2.38)	3.22 (0.41)	3.56 (0.78)	4.37 (1.49)

RT	WT	B1	B2	B3	B4	B5	B11
28.90	8.19 (1.78)	6.16 (0.88)	4.60 (0.58)	4.30 (0.84)	3.73 (0.48)	3.86 (1.14)	5.24 (1.11)
29.67	5.05 (1.22)	6.51 (0.90)	7.27 (0.79)	6.31 (0.65)	4.22 (0.70)	5.73 (1.50)	6.57 (1.73)
29.77	2.93 (1.00)	6.39 (3.08)	2.73 (0.36)	4.70 (2.68)	2.67 (0.78)	2.34 (0.49)	2.54 (1.06)
30.03	0.95 (0.38)	1.12 (0.29)	1.21 (0.31)	1.21 (0.38)	0.83 (0.13)	0.67 (0.15)	1.17 (0.42)
30.53	2.67 (0.47)	6.13 (2.61)	2.47 (0.62)	1.99 (0.08)	1.55 (0.48)	2.24 (1.01)	5.19 (1.65)
30.82	12.51 (4.17)	9.57 (1.46)	6.84 (2.34)	6.06 (1.94)	5.14 (0.39)	5.54 (1.38)	8.74 (3.42)
31.08	12.49 (0.44)	8.02 (0.39)	6.44 (1.25)	6.26 (1.98)	6.08 (0.48)	6.74 (2.27)	5.44 (2.38)
31.46	3.96 (0.49)	6.31 (2.99)	3.39 (1.21)	2.81 (1.74)	3.42 (1.13)	2.88 (1.90)	2.54 (1.74)
31.62	4.97 (0.89)	3.93 (0.65)	3.68 (0.56)	3.94 (0.86)	2.90 (0.45)	2.99 (0.94)	2.65 (1.09)
32.99	11.68 (1.22)	11.20 (0.80)	8.66 (0.77)	7.32 (1.30)	6.34 (0.76)	8.51 (2.51)	9.03 (2.36)
33.51	4.88 (0.68)	2.39 (0.32)	1.73 (0.30)	2.34 (0.82)	2.15 (0.11)	1.97 (0.54)	2.46 (0.90)
33.92	17.15 (2.67)	22.25 (2.30)	15.12 (2.92)	13.30 (0.90)	11.54 (2.08)	14.49 (3.89)	17.14 (5.90)
34.43	12.35 (1.50)	8.03 (1.10)	5.48 (0.87)	4.82 (0.94)	4.28 (0.21)	4.36 (1.23)	5.93 (2.03)
34.91	26.40 (9.05)	25.86 (4.08)	13.76 (2.48)	11.62 (2.51)	8.22 (0.44)	9.02 (2.57)	18.19 (8.08)
35.52	43.59 (1.12)	24.01 (2.15)	19.79 (1.80)	18.27 (2.70)	15.88 (2.28)	19.24 (3.02)	18.58 (3.04)
36.16	5.87 (0.17)	3.15 (0.20)	2.83 (0.64)	2.66 (0.83)	2.97 (0.48)	3.85 (1.38)	3.28 (1.25)
36.32	16.32 (1.35)	7.66 (0.86)	6.73 (0.38)	6.44 (1.48)	6.07 (1.26)	6.55 (1.23)	8.57 (3.01)
36.68	28.48 (6.74)	21.77 (4.85)	13.98 (5.18)	11.68 (2.78)	8.62 (1.10)	9.54 (2.30)	15.66 (6.16)
37.13	5.70 (0.37)	2.99 (0.41)	2.47 (0.23)	2.24 (0.57)	1.75 (0.06)	2.29 (0.82)	2.63 (0.99)
37.46	4.78 (0.98)	3.27 (0.37)	2.62 (0.29)	2.17 (0.43)	2.64 (0.58)	2.59 (0.48)	3.15 (0.94)
38.43	3.57 (0.82)	2.76 (0.76)	1.84 (1.41)	1.63 (0.99)	1.78 (0.78)	1.74 (0.64)	2.35 (0.78)
39.06	3.35 (0.36)	3.47 (0.60)	3.01 (0.38)	3.11 (0.93)	2.90 (0.21)	2.56 (0.57)	3.72 (1.27)
39.43	2.16 (0.32)	1.31 (0.53)	0.86 (0.20)	0.89 (0.15)	0.96 (0.19)	1.10 (0.41)	1.52 (0.64)
39.79	20.08 (0.78)	11.71 (1.21)	9.77 (1.00)	9.06 (1.53)	7.34 (0.78)	8.74 (2.31)	11.97 (3.70)
40.38	1.85 (0.35)	1.02 (0.15)	0.99 (0.20)	1.14 (0.33)	0.96 (0.09)	0.81 (0.24)	1.40 (0.17)
40.68	2.52 (0.71)	1.16 (0.31)	1.15 (0.48)	1.09 (0.22)	1.23 (0.28)	1.18 (0.48)	1.54 (0.68)
43.80	3.67 (1.01)	2.35 (0.27)	1.24 (0.30)	1.64 (0.32)	1.45 (0.24)	1.84 (1.16)	3.00 (1.48)
44.35	2.13 (0.35)	2.12 (0.11)	2.61 (0.53)	2.59 (0.25)	2.77 (0.27)	1.92 (0.60)	3.32 (0.52)
46.78	68.81 (22.36)	60.80 (17.48)	10.59 (4.10)	19.65 (0.85)	22.87 (5.07)	31.45 (2.86)	35.03 (4.75)
46.96	68.40 (11.01)	49.14 (8.59)	40.35 (6.06)	55.95 (15.05)	38.82 (7.61)	41.41 (16.14)	52.49 (12.25)
47.30	9.89 (0.65)	7.98 (0.40)	16.93 (6.87)	11.57 (1.37)	12.39 (5.36)	7.23 (0.61)	10.37 (2.78)
47.65	30.82 (5.67)	22.82 (3.64)	22.04 (6.84)	32.30 (0.96)	29.89 (3.55)	29.63 (5.38)	35.05 (8.26)
53.12	16.66 (1.49)	4.55 (0.88)	20.11 (12.18)	8.46 (1.87)	9.67 (2.72)	8.79 (1.84)	8.45 (1.75)
57.00	18.70 (1.69)	12.85 (1.60)	14.97 (1.70)	14.95 (1.38)	13.22 (3.82)	12.74 (0.87)	16.68 (3.88)

Table 5A: **Acid-hydrolyzed metabolites of RNAi-*PopGT1* and wild-type control plants that showed significant changes in normalized peak area compared to wild-type control. Bold values indicate the significant differences compared to the wild-type controls, using *t*-test at α = 0.05. Numbers in brackets represent the standard error of means (n = 4). RT is the retention time value for each compound**

RT	WT	A1	A2	A3	A4	A5	A9
13.5	18.60 (3.27)	14.54 (4.01)	8.11 (2.32)	7.68 (1.84)	5.46 (1.08)	9.11 (2.84)	9.76 (5.02)
14.62	9.45 (0.86)	9.41 (0.80)	7.57 (0.61)	6.88 (0.43)	5.47 (1.32)	8.20 (1.85)	5.84 (1.15)
17.45	126.68 (17.56)	128.08 (18.02)	146.89 (4.28)	146.23 (14.28)	149.89 (10.32)	167.97 (21.72)	159.60 (25.94)
19.78	9.89 (4.33)	5.51 (1.07)	4.84 (0.73)	4.87 (2.04)	3.87 (0.70)	5.20 (2.27)	3.04 (1.27)
24.76	21.13 (8.66)	18.19 (1.16)	16.91 (4.73)	18.87 (2.39)	13.41 (2.19)	19.32 (5.06)	21.91 (5.18)
25.13	8.90 (2.06)	5.75 (0.64)	2.89 (0.73)	2.29 (0.33)	2.70 (0.63)	2.83 (0.63)	3.22 (1.27)
26.72	61.13 (7.16)	55.46 (3.95)	44.31 (4.30)	54.99 (3.06)	46.64 (4.11)	51.78 (5.14)	53.31 (7.44)
31.81	47.73 (5.06)	41.58 (10.43)	28.16 (2.91)	20.63 (6.03)	12.75 (2.29)	23.65 (6.48)	21.95 (5.49)
32.21	10.59 (1.79)	11.92 (2.13)	8.64 (1.33)	8.89 (2.89)	6.45 (0.26)	10.21 (4.66)	6.25 (2.29)
37.75	52.25 (2.67)	50.45 (1.83)	49.64 (3.46)	48.69 (1.96)	46.64 (4.71)	53.84 (2.71)	57.25 (5.99)
38.96	27.22 (4.15)	22.65 (0.95)	17.67 (4.59)	19.16 (0.63)	15.55 (1.70)	21.29 (3.70)	24.12 (4.94)
40.29	1.37 (0.62)	1.72 (0.21)	1.20 (0.22)	1.66 (0.13)	1.37 (0.15)	1.57 (0.39)	1.64 (0.39)
46.87	8.73 (0.57)	7.62 (1.30)	9.93 (3.07)	12.11 (1.75)	10.46 (1.22)	10.54 (0.32)	12.43 (2.30)
47.25	22.39 (1.75)	20.19 (3.70)	10.03 (0.54)	14.66 (1.21)	12.40 (1.33)	12.22 (1.09)	12.56 (0.94)

Table 6A: **Acid-hydrolyzed metabolites of RNAi-*PopGT2* and wild-type control plants that showed significant changes in normalized peak area compared to wild-type control. Bold values indicate a significant difference when compared to wild-type controls using *t*-test at α = 0.05. Numbers in brackets represent the standard error of means (n = 4). RT is the retention time value for each compound.**

RT	WT	B1	B2	B3	B4	B5	B11
13.5	8.60 (1.27)	7.54 (4.01)	4.11 (1.32)	8.68 (1.84)	3.46 (1.08)	4.11 (2.84)	9.76 (2.02)
17.27	50.68 (3.56)	58.08 (8.02)	68.89 (14.28)	57.23 (14.28)	62.89 (9.32)	61.97 (2.72)	75.60 (5.94)
24.04	2.13 (0.77)	1.83 (1.16)	1.91 (0.73)	1.89 (0.39)	2.01 (0.19)	1.93 (0.60)	2.19 (0.58)
26.56	50.13 (3.16)	39.46 (3.95)	38.00 (4.30)	43.99 (1.06)	36.64 (4.11)	41.78 (5.14)	35.31 (6.34)
31.25	18.73 (2.06)	15.58 (3.43)	16.16 (3.91)	13.63 (2.03)	16.05 (1.29)	14.65 (1.48)	12.95 (2.12)
31.7	10.59 (1.79)	11.92 (2.13)	8.64 (1.33)	8.89 (2.89)	12.45 (1.26)	10.21 (4.66)	8.25 (2.29)
40.29	1.22 (0.15)	1.25 (0.95)	1.37 (0.59)	2.06 (0.59)	1.35 (0.71)	1.29 (0.50)	1.12 (0.41)
46.13	12.73 (0.57)	15.72 (1.30)	12.93 (1.07)	15.17 (1.75)	16.76 (1.22)	18.41 (0.32)	13.13 (2.30)