Defining roles for *SHORT VEGETATIVE PHASE/AGAMOUS-LIKE 24-like* (*SAL*) genes in the activity-dormancy transitions of white spruce (*Picea glauca*) terminal buds

By

Amanda Gregoris

A thesis submitted in partial fulfillment of the requirements for the degree of

Masters of Science

in

Plant Biology

Department of Biological Sciences
University of Alberta

Abstract

An important part of the annual growth cycle of white spruce (*Picea glauca* [Moench.] Voss; *Pg*) trees is the transition from active growth to dormancy, which confers protection against the potentially destructive environmental elements of winter. Terminal bud formation and cessation of meristemic growth is a precursor to dormancy induction. Environmental cues, such as photoperiod, temperature, water stress and phytohormones influence the progression of bud development and growth cessation. In angiosperms, *SHORT VEGETATIVE PHASE* (*SVP*) genes have been implicated in the control of bud formation, growth cessation and dormancy induction. However, the roles of *SVP*-like genes in white spruce and other conifers have yet to be investigated in this context.

We identified a suite of white spruce genes with sequence similarity to *SVP* genes and explored whether these genes have a role in bud formation. To determine the white spruce genes that are mostly closely related to angiosperm *SVP* and *SVP*-like genes, we constructed a phylogenetic tree using nucleotide and deduced amino acid sequences from a range of land plants. This analysis showed that seven white spruce genes form a sister clade with both *SVP-like* sequences and the closely related *AGAMOUS-LIKE 24-like* (*AGL24-like*) sequences from angiosperm species. Based on this evolutionary relationship, we have called these white spruce genes *PgSVP/AGL24-like* (*PgSAL*). Transcriptional profiling revealed that the seven *PgSAL* genes plus the more distantly related GQ03118_H14 exhibited three major expression patterns, with five of the seven *PgSAL* genes showing declining expression at later time points. Based on transcriptional data, the genes that are most likely to be involved in regulating bud formation and/or growth cessation are *PgSAL1, PgSAL2*, *PgSAL3*, *PgSAL4*, and *PgSAL5*.

Based on these expression profiles, we selected two *PgSAL* genes for further functional characterization through identifying factors that regulate their expression. We targeted the promoter sequences of *PgSAL1* and *PgSAL5* to identify potential upstream regulators. *In silico* characterization revealed potential transcription factor binding sites in the *Pg*SAL1 and putative *Pg*SAL5 promoters that may be regulated by environmental cues associated with bud formation and growth cessation, such as low temperatures, light, water stress and hormones (abscisic acid, ethylene, cytokinin, gibberellins and auxin). DNA-protein interactions as determined by yeast one-hybrid revealed that the promoter of *PgSAL1* gene showed interactions consistent with a function in the bud formation pathways conserved with the angiosperm photoperiodic pathway. The putative *Pg*SAL5 promoter is regulated by factors that suggest a role outside of bud formation, based on the angiosperm model. Both the *Pg*SAL1 and putative *Pg*SAL5 promoters were regulated by transcription factors that participate in regulatory networks of low temperature, the abscisic acid response, plant defense and/or secondary growth. A subset of transcription factor binding sites suggest that *PgSAL1* and *PgSAL5* could be regulated by the defense pathway, which may indicate novel roles for these genes outside of the phase transition from active growth to dormancy.

We demonstrate that white spruce *SAL* genes are homologous to angiosperm *SVP* and *AGL24* genes, and propose that a subset of these genes have roles in the bud formation processes that precede winter dormancy based on expression patterns and associated upstream regulatory pathways, in addition to possible functions outside of bud formation.

Preface

The experimental design for white spruce seedlings was designed and grown by Ms. Carmen Gibbs-Allen and Dr. Eri Adams (Chapter 2) and Dr. Jill Hamilton (Chapter 3). The statistical analyses (Chapter 2) were carried out in consultation with Mr. Dean Koch and Ms. Melodie Kunegal-Lyon. I was responsible for conducting the molecular biology experiments and analyses described in this thesis.

Acknowledgements

I would like to thank Dr. Nathalie Isabel and Ms. Marie-Claude Gros-Louis (Natural Resources Canada – Canadian Forest Service, Québec) for providing the white spruce seedlings used for Chapters 2 and 3, and Ms. Carmen Gibbs-Allen and Dr. Eri Adams (Chapter 2) and Dr. Jill Hamilton (Chapter 3) for conducting the plant experiments used to generate experimental materials. The trees used in the Chapter 3 analysis were donated by Coast to Coast Reforestation. I am thankful to Dr. Walid El Kayal for his assistance with the gene expression studies in Chapter 2. I would like to thank Mr. André Gagné and Dr. John MacKay (formerly of Université Laval) for providing cDNAs used in Chapter 2 from the Arborea cDNA collection maintained at Université Laval. I would like to acknowledge Dr. Amy Brunner and Mr. Earl Petzold for training on the yeast one-hybrid assays, and Mr. Steven Rigoulot for suggestions on cDNA library construction in Chapter 3. I would like to thank Mr. Dean Koch and Ms. Mélodie Kunegel-Lion for assistance and recommendations with the multivariate analysis of variance (MANOVA) in Chapter 2. I would like to thank Dr. Rhiannon Peery for creating the alignments for the *SAL5* promoter and *SAL5* cDNA sequence against the white spruce genome assembly contigs in Chapter 3, which I used to create my final alignment figures. I am thankful to my committee member Dr. Enrico Scarpella for help with the promoter database search in Chapter 3, and his continuous advice throughout my research. I would most of all like to thank my supervisors Dr. Janice Cooke and Dr. Jocelyn Hall for their experimental advice and comments on my writing.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3-AT	3-amino-1,2,4-triazole
AC	*Actinidia chinensis*
AG	AGAMOUS
AGL	AGAMOUS-LIKE
AIC	Akaike Information Criterion
ANR1	ARABIDOPSIS NITRATE REGULATED 1
AP1	APETELA1
AP2	APETELA2
ASR	ABSCISIC ACID STRESS RIPENING
AT	*Arabidopsis thaliana*
ATHB9	ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 9
AUX	auxin
bp	base pairs
BIC	Bayesian Information Criterion
β-gal	β-galactosidase
BS	bootstrap
CPC	CAPRICE
cfu	colony forming units
CK	cytokinin
cm	centimeters
CO	CONSTANS
ct	cycle threshold
CT	*Citrus trifoliata*
CTAB	cetyltrimethylammonium bromide
DAM	DORMANCY ASSOCIATED MADS-BOX
DIW	deionized water
EDTA	ethylenediaminetetraacetic acid
EG	*Eucalyptus grandis*
EE	*Euphorbia esula*

EIL	ethylene insensitive-like
ERF	ethylene response factor
ETC	ENHANCER OF TRYPTYCHON AND CAPRICE
EVG	EVERGROWING
FLC	FLOWERING LOCUS C
FLM	FLOWERING LOCUS M
FLX	FLOWERING C EXPRESSOR
FRI	FRIGIDA
FT	FLOWERING LOCUS T
FTL1	FT/TERMINAL FLOWER1-like
FTL2	FT/TERMINAL FLOWER2-like
HV	*Hordeum vulgare*
His	histidine
HSD	honest significant difference
g	grams
GA	gibberellins
GTR	general time reversible
IAA	indole-3-acetic acid
JTT	Jones-Taylor-Thornton
L	litres
LB	Lennox broth
LD	long day
LFY	LEAFY
LiAc	lithium acetate
M	molal
mm	millimeters
MADS	MCM1, Antirrhinum, DEFICIENS, SRF
MCM1	MNICHROMOSOME MAINTENANCE1
MEF2	MYOCYTE ENHANCER FACTOR2A
MFT	MOTHER OF FT AND TFL1
MIKC	MADS, intervening, keratin, C-terminal

mL	milliliters
ML	maximum likelihood
MP	maximum parsimony
NBS-LRR	nucleotide-binding site/leucine-rich repeat
OD	optical density
O/N	overnight
PCR	polymerase chain reaction
PEBP	phosphatidylethanolamine-binding protein
PEG	polyethylene glycol
PG	*Picea glauca*
PIF3	POLYCHROME INTERACTING FACTOR 3
PhP	*Phsycometrella patens*
PHYA	PHYTOCHROME A
PA	*Prunus avium*
PG	*Picea glauca*
PM	*Prunus mume*
PP	*Prunus persica*
qRT-PCR	quantitative reverse transcriptase PCR
RAV1	RELATED TO ABI1/VP1
RELL	resampling estimated log-likelihoods
SAL	SHORT VEGETATIVE PHASE/AGAMOUS-LIKE 24-like
SAM	shoot apical meristem
SC	synthetic complete
SD	short day
SEP	SEPELLATA
SH	Shimodaira-Hasegawa
SHP	SHATTERPROOF
SOC1	SUPPRESSOR OF OVEREXPRESSOR OF CONSTANS 1
SRF	SERUM RESPONSE FACTOR
ST	*Solanum tuberosum*
SVP	SHORT VEGETATIVE PHASE

TBR	terminal branch rearrangement
TE	TRIS base EDTA
TF	transcription factor
TFBS	transcription factor binding sites
TIF5a	TRANSLATION INITIATION FACTOR 5a
Tm	melting temperature
Trp	tryptophan
µg	microgram
µl	microlitre
Ura	uracil
VV	*Vitis vinifera*
YNB	yeast nitrogen base
YPDA	yeast, peptone, dextrose, adenine sulfate
ZAP1	ZINC-DEPENDENT ACTIVATOR PROTEIN-1

1.0 Chapter 1: Introduction and Background

1.1 Importance of white spruce in Canadian forests

Canada has 348 million hectares of forest, which is 9% of the global forest cover, ranking Canada in third place, behind Russia and Brazil, in forested areas (Natural Resources Canada 2014a). The five species of spruce native to Canada, white spruce (*Picea glauca* (Moench) Voss), red spruce (*Picea rubens* Sarg.), black spruce (*Picea mariana* (Mill.) Britton, Sterns & Poggenb), Engelmann spruce (*Picea engelmannii* Parry ex Engelm.), Sitka spruce (*Picea sitchensis* (Bong.) Carrière), make up a large proportion of Canadian forests (Natural Resources Canada 2015a, Canadian Wildlife Federation 2017). A distinct feature of spruce trees that differentiate them from other conifers is their four-sided needles, with the exception of the two-sided needles of Sitka spruce.

White spruce trees and their hybrids are found in almost all forests within Canada except for northern regions of Nunavut and the Pacific Coast, and comprise 20% of Canada's forests (Government of Alberta 2006, Natural Resources Canada 2015, Canadian Wildlife Federation 2017). White spruce trees have a transcontinental distribution across Canada (Nienstaedt and Zasada 1990, Figure. 1.1). In Alberta, *Picea* make up approximately 45% of forested areas, with white spruce comprising 30% (Government of Alberta 2013). Natural hybrid zones of white x Engelmann spruce (*Picea glauca* (Moench) Voss x *Picea engelmannii* Parry ex Engelm.) occur where the distributions of these trees overlap in Alberta (Government of Alberta 2016). Other trees commonly found growing in forests containing white spruce include Sitka spruce, balsam poplar, aspen and birch (Government of Alberta 2003).

Spruce trees play an important role in contributing to the maintenance of forest ecology. Forests containing spruce trees provide habitats for many species of the *Cervidae* family, including deer, moose, caribou and elk, as well as sheep, goats and bears. In addition, many Indigenous communities (~70%) are located in Canadian forests (Natural Resources Canada 2014a). Continued attention is being placed towards conservation and regeneration of Canadian forests with caribou being declared a species at risk by the Canadian federal government (Government of Canada 2017). This is an effort to preserve caribou habitats, which have a preference for forests containing white spruce, among other tree species (Government of Canada 2017).

Canada's softwood lumber exports comprise mainly of spruce, pine and fir, and generate a GDP of $22 billion annually (Natural Resources Canada 2014b). Over 200,000 Canadians are employed by the forestry industry in Canada (Natural Resources Canada 2017). Spruce trees are harvested for use in solid wood and paper products, and grown commercially for the Christmas tree industry. Some of the products manufactured from white spruce include newsprint, construction materials, plywood, paddles, musical instruments and packing cases (Government of Alberta 2006, Government of Alberta 2003). Trees are generally harvested when trees have reached 80 to 120 years of age (Government of Alberta 2003).

1.2 Dormancy

1.2.1 Endodormancy, ecodormancy, paradormancy

Entrance into a dormant state is a key component of the perennial lifestyle in northern temperate climates. Dormancy aids to protect trees from the unfavourable conditions of winter so that they may go on to resume growth and thrive the following spring. There are multiple descriptions that have employed the word dormancy to describe the cessation of growth. In Lang

(1987) and Lang *et al.* (1987) three states of dormancy were described: (1) endodormancy, (2) paradormancy and (3) ecodormancy. Endodormancy is also described as innate or seasonal dormancy. Endodormancy is a state in which the cessation of a structure is imposed by the tissue itself, and regrowth of this structure will not occur even when placed under growth permissive conditions. Paradormancy is a state in which the inhibitory cues preventing regrowth at a structure is imposed by the plant, but this inhibitory signal originates from a different structure. Paradormancy is also referred to as correlative inhibition, and is commonly displayed through apical dominance, especially in conifers. Apical dominance is the circumstance that attribute the characteristic conical shape of conifers trees, in which auxin released from the apical bud inhibits the growth of lateral buds. Inhibitory effects of apical dominance can be removed by damage to the apical bud or decapitation, thereby removing the inhibitory auxin signal (Gocal *et al.* 1991) and allowing cytokinin originating from the roots to stimulate axillary bud growth (Bangerth 1994). Ecodormancy is when cessation of growth of a structure is external to the plant, and imposed by the environment. In ecodormancy growth permissive conditions such as warm temperatures, nutrient and water availability are absent and as a result growth of the structure does not occur.

Lang's (1987) definitions of dormancy are limited since they rely on the overall appearance of growth on a physiological level, and focus on the source of the dormancy imposing cues. This characterization is problematic because the dormancy status of a structure is based on whether the structure has the ability to resume growth, not if growth actually occurs (Rohde and Bhalerao 2007). Rohde and Bhalerao (2007) propose an alternative definition of dormancy which implicates the meristem as the main determinant of dormancy status. Rohde and Bhalerao (2007) define bud dormancy as the absence of growth in meristematic tissues even

when permissive growth conditions are present. This definition is similar to the description of endodormancy proposed by Lang (1987) but makes the clear distinction that once dormant, growth may still not occur as a result of environmental factors preventing resumption of growth, also referred to as ecodormancy. This description is also inclusive towards structures which may not resume growth as a result of inhibition being imposed by another plant structure. Here thereafter the term dormancy will refer to the simple and inclusive definition proposed by Rohde and Bhalerao (2007).

1.2.2 Dormancy depth

Despite improvements researchers have made in defining the state of dormancy, the current definition implies that dormancy is either present or absent in a structure. This definition does not take into account that dormancy is not a strict qualitative trait, and instead the ability of dormancy to inhibit regrowth exists on a continuum. Dormancy is now recognized as a quantitative trait, and the scale of dormancy establishment is referred to as "depth". This depth can be quantified either by the number of days of chilling or the temperature at the time of chilling and may be supplemented with photoperiod input (Worral and Mergen 1967**,** Sarvas 1974, Leinonen 1996). Dormancy depth can be influenced by the temperature during dormancy establishment. Intermediate temperatures can induce deeper dormancy in birch (Junttila *et al.* 2003), apple and pear (Jonkers 1979). Environmental factors influencing dormancy will be discussed further in Chapter 3.

Cooke *et al.* (2012) propose that the definition of dormancy should be expanded to incorporate the depth of dormancy of a particular structure like buds, similar to that of seed dormancy, using the terms deep, intermediate, or non-deep (Baskin and Baskin 2004, Graeber *et al. 2*012). This categorization of dormancy takes into account that dormancy is a quantitative

trait as opposed to a qualitative trait, integrating internal as well as external signals to modulate and regulate depth (Cooke *et al.* 2012).

1.2.3 Dormancy establishment

Dormancy can be induced by environmental or endogenous factors, with the importance and strength of each factor tending to be species specific (Singh *et al.* 2017, Hänninen and Tanino 2011). Following bud burst in the spring, preformed needle primordia and stem units contained within the bud will elongate. Preformed growth occurs in the previous growth season, whereas neoformed growth occurs in the same growth season. Indeterminate growth refers to a plant, e.g. poplar, which produces neoformed stem units and elongates internodes in the same growing season (Kozlowski and Pallardy 19977, Rohde *et al.* 2000). While determinate growth (e.g. white spruce), refers to the majority or all of the season's current growth to be predetermined by the number of preformed stem units from the previous growing season (Kozlowski and Pallardy 19977, Rohde *et al.* 2000). After the summer equinox, the days begin to shorten, which is perceived by plants as short days (SDs). SDs are recognized by the plant when the period of light falls below the critical day length (Taiz and Zeiger 2010). Perception of these changing photoperiods are believed to be perceived in the needles and leaves (Eagles and Wareing 1964, Wareing 1970, Singh *et al.* 2017). SDs have been shown to trigger bud set and cessation of growth in some species, such as *Populus* species including hybrid aspen (*Populus tremula* L. *x Populus tremuloides Michx*; Olsen *et al.* 1997b), bay willow (*Salix pentandra* L.; Junttila 1980), and downy birch (*Betula pubescens* Ehrh.; Junttila 1980). SD signals developmental changes in buds, causing a subset of primordia to differentiate into bud scales instead of needles or leaves (Okuba 2000). Terminal bud set is a prerequisite to dormancy

induction, and is followed by cessation of cell division at meristems (Rohde and Bhalerao 2007). Evidence has shown that the cell-to-cell communication networks, plasmodesmata, become blocked with callose prevents signaling molecules, such as transcription factors and hormones, from reaching the shoot apical meristem (SAM; Rinne and van der Schoot 1998, Rinne *et al.* 2005, Levy *et al.* 2007, Rinne *et al.* 2011). This model proposes that bud dormancy is the result of the symplasm of the bud becoming physically isolated from the rest of the plant. The establishment and removal of these plasmodesmata plugs are associated with dormancy establishment and release in birch and poplar (Jian *et al.* 1997, Rinne and van der Schoot 1998, Rinne *et al.* 2011). It is believed that plasmodesmata plugs must be removed in order for the bud to regain communication with the rest of the plant and to subsequently receive the necessary signals to resume growth, which is associated with dormancy release (Rinne *et al.* 2011). However, at this time, no genetic or molecular biology has confirmed this hypothesis on bud dormancy (Singh *et al.* 2017).

In white spruce, SDs are not necessary for the formation of terminal buds (El Kayal *et al.* 2011, Hamilton *et al.* 2016). However, SDs have been found to accelerate the development of terminal buds and suppress the same-season expansion of needle primordia in partially formed buds (El Kayal *et al.* 2011, Hamilton *et al.* 2016), otherwise known as second flush or lammas growth (Figure 2). Younger white spruce trees are more susceptible to growth cessation and terminal bud set under SDs, however this trait declines as the tree matures and likely becomes regulated by endogenous signals (Cooke *et al.* 2012, Singh *et al.* 2017).

Photoperiod is not a ubiquitous stimulus for the induction of bud set and growth cessation, as some species' growth cycle is driven by temperature. Species from the Rosaceae family, including apple (*Malus pumila* Mill.) and pear (*Pyrus communis* L.), use low

temperatures as an indication to commence the processes associated with bud set and growth cessation (Heide and Prestrud 2005). Unlike day length, temperature can have great variation year-to-year. Although dormancy is induced by photoperiod in most species, some species use both temperature and photoperiod cues in regulating their annual growth cycles. For instance, high day temperatures and low night temperatures can serve to replace the photoperiod requirement for dormancy induction in Norway spruce, bay willow and hybrid aspen (Heide 1974, Junttila 1980, Mølmann *et al.* 2005). Fall temperatures can also affect growth cessation, rate of dormancy acquisition and depth of dormancy in poplar (Kalcsits *et al.* 2009, Tanino *et al.* 2010). Low temperatures are not a requirement for bud formation in white spruce (El Kayal *et al.* 2010), however low temperatures delay bud formation and do not prevent second flush in trees grown under long day (LD) conditions (Hamilton *et al.* 2016). The delay in bud formation caused by low temperatures in LD condition is in agreement with evidence in poplar that suggests temperature alters the tree's responsiveness to photoperiod (Rohde *et al.* 2011).

1.3 Molecular regulation of bud formation and dormancy acquisition

1.3.1 MADS-box genes

Since many genes that have been implicated in bud formation belong to the MADS-box gene family, here I have included an overview of the structure of these genes. The MADS-box gene family is a family of transcription factors (TFs) that has roles in development and differentiation in plant, fungi and animal species. The designation "MADS" is derived from the earliest described members of this family: the "M" stands for *MINICHROMOSOME MAINTENANCE1* (*MCM1*) in yeast (*Saccharomyces cerevisiae* Meyen Ex. Hansen) (Passmore *et al.* 1988), the "A" stands for *AGAMOUS*, discovered in *Arabidopsis thaliana* (L.) Heynh

(hereafter referred to as *Arabidopsis*, Yanofsky *et al.* 1990), the "D" stands for or *DEFICIENS* from snapdragon (*Antirrhinum majus* L.) (Sommer *et al.* 1990, Schwarz-Sommer *et al.* 1992) and the "S" stands for *SERUM RESPONSE FACTOR* (*SRF*) from humans (*Homo sapiens* L.) (Norman *et al.* 1988). Members in this family are categorized as Type I and Type II *MADS-box* genes based on conserved domains, and are believed to have undergone a duplication event preceding the divergence between plant and animals (Alvarez-Buylla *et al.* 2000). Type I *MADS-box* genes contain the *SRF-like* domain, whereas Type II genes encode a *MYOCYTE ENHANCER FACTOR2-like* (*MEF2-like*) domain and are exclusively found in plants (De Bodt *et al.* 2003, Alvarez Buylla *et al.* 2000).

Type II *MADS-box* genes are also referred to as MIKC genes due to the characteristic four domain structure: "M" (MADS), "I" (intervening), "K" (keratin-like) and "C" (C-terminal) (Theissen *et al.* 1996). The "M" and "K" domains are well conserved and participate in DNA binding and protein-protein interactions, respectively (Davies *et al.* 1996, Fan *et al.* 1997). The "I" domain is comparatively less conserved, and is believed to contribute to dimerization specificity (Parěnicová *et al.* 2003). The "C" domain is the most divergent domain, and is involved in multimeric protein complex formation, as well as transcriptional activation (Egea-Cortines *et al.* 1999, Honma and Goto 2001). Type II *MADS-box* genes are further subdivided into MIKC* and MIKC^{c}. The additional "^{c}" in MIKC^{c} refers to "classic", since MIKC^{c} genes possess the classic MIKC-type domains found in this family (Becker and Theissen 2003). The asterisk, "*", in MIKC* denotes that this group of *MADS-box* genes deviate from the classic MIKC domains via an elongated "I" domain, in addition to the divergence in the "K" domain (Becker and Theissen 2003). There is also documentation of a less well conserved "N" domain, which proceeds the "M" domain, but it is only found in minority species (Henschel *et al.*

2002). MIKC proteins bind to promoter regions as homo-dimers, hetero-dimers, or higher order protein complexes to regulate transcription (Egea-Corines *et al.* 1999, Honma and Goto 2001).

1.3.2 The photoperiodic flowering pathway

Considerable research has been put forth to characterize the function of MIKC genes in *Arabidopsis* and other angiosperm species. This work has demonstrated MIKC genes have roles in floral organ and meristem identity determination (Kaufmann *et al.* 2005, Gramzow and Theissen 2010), in addition to the regulation of flowering time, also referred to as the phase transition from vegetative to reproductive growth. The pivotal research by Böhlenius *et al.* (2006) showed that a poplar flowering gene, *FLOWERING LOCUS T* (*FT*), also regulates the activity-dormancy transition, demonstrating that the transition to dormancy is regulated by a signaling network analogous to the photoperiodic pathway that regulates the transition from vegetative to reproductive growth. Given the importance surrounding the regulation of these processes, many researchers use the flowering pathway as a guide to understand the transition from vegetative growth to dormancy. Both flowering and dormancy represent transitions in developmental stages, which may employ a similar pathway. Endogenous and environmental signals are integrated at the SAM to regulate this phase transition in *Arabidopsis* (Hartmann *et al.* 2000, Tao *et al.* 2012), including light, temperature and endogenous signals (reviewed in Amasino 2010). Several transcription factors within and outside the MADS-box family participate in the photoperiodic transition to flowering which include, but are not limited to the following gene: *CONSTANS* (*CO*), *FT*, *APETELA1* (*AP1*), *LEAFY* (*LFY*), and the *MADS-box* genes *FLOWERING LOCUS C* (*FLC*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), *AGAMOUS-LIKE 24* (*AGL24*) and *SHORT VEGETATIVE STAGE* (*SVP)* (reviewed in Amasino 2010 and Irish 2010).

Figure 1.3 is a simplified overview of the photoperiodic flowering pathway in *Arabidopsis*. Under LDs in *Arabidopsis*, the expression of *CO* peaks at the end of the day allowing the CO protein to remain stable to transcribe *FT* (Suarez-López *et al.* 2001). In the absence of the LDs the transcription of *CO* peaks in the dark and the protein is degraded, thereby it is unable to transcribe *FT* (Suárez-López *et al.* 2001). *FT* is transcribed in the leaf and is believed to translocate to the SAM to induce the expression of *AP1* and *SOC1*, which leads to the meristem transitioning from vegetative to reproductive growth (Abe *et al.* 2005, Yoo *et al.* 2005, Nakamura *et al.* 2013). SOC1 and AGL24 upregulate each other's expression (Michaels *et al.* 2003). SOC1 is proposed to complex with AGL24 to transcribe *LFY* (Lee *et al.* 2008), and LFY goes on to upregulate *AP1* to induce flowering (Liu *et al.* 2009). SVP directly binds to the *SOC1* promoter to inhibit the transcription of *SOC1* in the SAM and leaf (Li *et al.* 2008), thereby inhibiting the transition to flowering. SOC1 is believed to play a part in the inhibition of the transcription of *SVP* through its ability to bind an intron within the *SVP* gene (Immink *et al.* 2012). When the plant is exposed to ambient temperatures (16°C) *SVP* complexes with FLC to repress the transcription of *FT* in the leaf, and block the transition to flowering until a chilling requirement is met (Lee *et al.* 2007, Li *et al.* 2008). The mechanism behind the increase of *AGL24* expression in response to vernalization is not clear, although it is believed this regulation occurs independent of FLC (Michaels *et al.* 2003).

1.3.3 Conceptual model of bud formation and dormancy acquisition in angiosperms and conifers

Many genes have been implicated in the regulation of dormancy induction based on key players in the controlling the flowering pathway which are highlighted here. The most widely

accepted theory of flowering is the external coincidence model (Bünning 1936), of which I will give a very brief overview. The external coincidence model assumes the plant has an internal circadian oscillation of gene expression that is reset daily based on photoperiod, which must coincide with an external cue, such as light, to bring about flowering. In the LD plant *Arabidopsis*, PHYTOCHROME A (PHYA) perceives light and prevents degradation of the CO protein during LDs (Valverde *et al.* 2004, Langercrantz 2009). CO will now remain stable for an extended time during periods of light to promote transcription of the flowering inducer, *FT* (Suárez-López *et al.* 2001). PHYA may further function to regulate growth cessation as demonstrated in *Arabidopsis*. In *phya* mutants, the transcript levels of genes associated with flowering, *CO* and *FT*, are reduced as a consequence of the absence of a functional PHYA (Yanovsky and Kay 2002). A pivotal study determined that overexpression of the oat (*Avena sativa* L.) *PHYA* gene in hybrid aspen (*Populus tremula x tremuloides*) renders the tree unresponsive to changes in day length when maintained at constant temperatures (Olsen *et al.* 1997b). This result supported the proposition that phytochromes, particularly *PHYA,* are involved in sensing and signaling SD induced growth cessation, bud set, cold acclimation and induction of dormancy. It has also been confirmed that orthologs of *Arabidopsis FT* and *CO* did not display a reduced transcription in the hybrid aspen overexpressing oat *PHYA* (Böhlenius *et al.* 2006). This lack of reduction suggests the *Arabidopsis* regulatory mechanism of PHYA over *FT* and *CO* is similar in hybrid aspen, and regulation of *FT* and *CO* are important for SD induced growth cessation and bud set.

A key flowering time regulator *FT* promotes the transition to flowering in monocot and eudicot species (Pin and Nilsson 2012). It has been demonstrated that *FT* orthologs can have roles outside of the flowering pathway. In experiments conducted in *Populus tremula x*

tremuloides, a *FT* ortholog from *Populus trichocarpa* Torr. & A. Gray has been shown to be involved in this early response to SD induced growth cessation and bud set genes (Böhlenius *et al.* 2006). *FT* orthologs also participate in bud set and growth cessation in Norway spruce (Gyllenstrand *et al.* 2007, Karlgren *et al.* 2011), as well as growth termination in tomato (*Solanum lycopersicum* L.; Lifschitz *et al.* 2006) and tuberization in potato (*Solanum tuberosum* L.; Navarro *et al.* 2011). Research in white spruce has identified that genes with similarity to angiosperm genes that participate in the initiation of floral buds, such as *MOTHER OF FT AND TFL1* (*MFT*) and *AP1*, are also differentially expressed during white spruce bud formation (El Kayal *et al.* 2011). Karlgren *et al.* (2011) found that *FT* genes sister to *FT/TERMINAL FLOWER1* (*TFL1*) were implicated in bud formation and growth cessation in Norway spruce, and these orthologs were able to affect flowering time and one also altered flower morphology in wildtype *Arabidopsis* plants. This accumulation of research demonstrates that the transition from vegetative growth to reproductive growth in conifers shares similarities with the molecular pathway involved in the transition between active growth to dormancy in angiosperms. However, due to the evolutionary distance between angiosperms and conifers it is quite possible that are divergent functions of the orthologous genes involved in these pathways. Even though conifers and flowering plants shared a common ancestor approximately 310 million years ago (Schneider *et al.* 2004), it is possible there is conserved regulatory mechanism associated with bud formation and/or phase transitions. Furthermore, it is possible for these genes to function outside of the traditional roles that have been functionally characterized in *Arabidopsis*.

In addition to transcription factors, phytohormones play a role in the developmental processes leading to dormancy, although further evidence is necessary to establish direct roles in dormancy establishment. Ethylene and abscisic acid (ABA) may function cooperatively in bud

formation (Rohde *et al.* 2002, Ruttink *et al.* 2007). One of the roles of ABA is preventing the growth of shoots (Davies 2010). However, there is evidence that ABA participates in bud development and maturation (Rohde *et al.* 2002, Ruttink *et al.* 2007). In hybrid aspen, the rate of bud maturation is slowed in the presence of decreased ABA sensitivity (Petterle *et al*. 2011). Two weeks of SD treatment upregulated genes involved in the transcription in ethylene biosynthesis and signaling in *Populus tremula L.* x *Populus alba L.* (Ruttink *et al.* 2007). Ethylene insensitive in birch (*ETHYLENE RESPONSE1* [*etr1*] Ruonala *et al.* 2006) and *ABSCISIC ACID-INSENSITIVE3* (*ABI3*) in poplar result in similar altered bud morphology (Rohde *et al.* 2002, Ruttink *et al.* 2007). However, the expression of *ABI3* is not affected by ABA levels in poplar, and therefore the link between ABA and bud maturation is unclear (Maurya and Bhalerao 2017). Fewer studies have been conducted on the effect of indole-3-acetic acid/auxin (IAA/AUX) and cytokinins in SD-induced growth cessation and apical bud formation.

The family of phytohormone that has been most well studied in SD-induced growth cessation are the gibberellins (GA). GA play a key role in cell division and elongation in plants, and it is believed decreasing GA levels contribute to growth cessation. *Arabidopsis SVP* inhibits transcription of a key GA biosynthesis gene (Andrés *et al.* 2014), which prevents the transcription of key flowering genes, thereby delaying the transition to flowering (Blázquez *et al.* 1998, Moon *et al.* 2003). In *Salix*, phytochromes recognize the photoperiod shift to SD, which diminishes the GA and IAA/AUX content (Olsen *et al.* 1995a, b, Olsen *et al.* 1997a). In addition to decreased amounts of GA, continued exposure to SD also causes the tree to become insensitive to GA in *Salix* (Juntilla and Jensen 1988). In controlled growth chamber conditions photoperiod-induced transcriptional changes can be detected as early as two weeks following the switch from LD to SD conditions in *Populus*, and bud formation is seen as early as three weeks

of SD treatment (Ruttink *et al.* 2007). Work performed in hybrid aspen suggests PHYA may control GA levels during SD induction. Overexpression of oat *PHYA* resulted in no reductions in GA content nor decreased activity of GA 20-oxidase, a rate limiting enzyme in GA biosynthesis (Olsen *et al.* 1997a, Mølmann *et al.* 2003). Further evidence of GA's role in growth cessation is demonstrated with a delay in growth cessation in the presence of overexpression of *GA 20-oxidase* in hybrid aspen (Eriksson and Moritz 2002). However, most evidence of GA's role in growth cessation has been demonstrated in the bud and this mechanism is yet to be supported in dormancy at the cambium (Druart *et al.* 2007).

1.3.4 SHORT VEGETATIVE PHASE/AGAMOUS-LIKE 24-like genes

SVP is a member of the MADS-box gene family, and has been extensively examined in *Arabidopsis* in relation to flowering (Blázquez *et al.* 1998, Hartmann *et al.* 2000, Moon *et al.* 2003, Gregis *et al.* 2006, Lee *et al.* 2007, Li *et al.* 2008, Li *et al.* 2008, 2006, Gregis *et al.* 2009, Liu *et al.* 2009, Andrés *et al.* 2014). *Arabidopsis* possesses one *SVP* gene and another sequence with high similarity to *SVP*, *AGAMOUS-LIKE 24* (*AGL24*). *SVP* is a negative regulator of flowering, demonstrated by the knock-out phenotype that displays an early flowering phenotype, while overexpression induces the formation of leaf-like sepals and flowers later than wildtype (Hartmann *et al.* 2000, Masiero *et al.* 2004). Despite the high sequence similarity and close evolutionary history between *AGL24* and *SVP* (Parěnicová *et al.* 2003), *AGL24* plays an antagonistic function by promoting the transition from vegetative to reproductive phase (Michaels *et al.* 2003). *AGL24* loss of function mutants flower later, while overexpression results in early flowering (Michaels *et al.* 2003). During vegetative growth *SVP* is expressed in leaves and the SAM to maintain the vegetative state (Hartmann *et al.* 2000), while *AGL24* is

primarily expressed in the infloresence meristem and promotes the development of the floral meristem (Michaels *et al.* 2003, Yu *et al.* 2004).

Arabidopsis AGL24 and *SVP* have roles that extend beyond regulating the timing of flowering. AGL24 and SVP participate in the regulation of *AGAMOUS* (*AG*) in a transcription factor complex to affect normal flower development (Gregis *et al.* 2006). Overexpression of *AGL24* and *SVP* independently cause similar floral abnormalities, such as the development of structures resembling leaves where one would expect petals and sepals (Michaels *et al.* 2003, Masiero *et al.* 2004). Barley (*Hordeum vulgare* L.) *SVP*-like genes are also believed to control meristem identity as demonstrated by the floral reversion phenotypes observed in mutant barley and *Arabidopsis* (Trevaskis *et al.* 2007). At early stages of floral development SVP and AGL24 both inhibit transcription of *SOC1*, a *MADS-box* gene which promotes the floral transition (Gregis *et al.* 2006). SVP and AGL24 also hetero-dimerize with AP1 to repress expression of floral meristem identity genes (Gregis *et al.* 2006, Gregis *et al.* 2009, Liu *et al.* 2009). *Arabidopsis* mutant and phenotyping experiments have demonstrated that *SVP* is epistatic to *AGL24* in the flowering pathway (Gregis *et al.* 2006).

DORMANCY ASSOCIATED MADS-BOX (*DAM*) genes are a group of *SVP*-like genes that have been associated with roles in bud formation, flowering and/or dormancy acquisition (Bielenberg *et al.* 2004, Li *et al.* 2009, Jiménez *et al.* 2009, Yamane *et al.* 2011). *DAM* genes have been identified in peach (*Prunus Persica* (L.) Batsch, *Pp*; Bielenberg *et al.* 2004, Jiménez *et al.* 2009), Japanese apricot (*Prunus mume* (Siebold) Siebold & Zucc.; Saski *et al.* 2011), leafy spurge (*Euphorbia esula* L.; Horvath *et al.* 2010), raspberry (*Rubus idaeus* L.; Mazzitelli *et al.* 2007), potato (*Solanum tuberosum* L.; Campbell *et al.* 2008), trifoliate orange (*Poncirus trifoliata* (L.) Raf.; Li *et al.* 2010), kiwifruit (*Actinidia deliciosa* (A. Chev.) C.F. Liang & A.R.

Ferguson; Wu *et al.* 2011), apple (*Malus domestica* Borkh.; Mimida *et al.* 2015) and Asian pear (*Pyrus pyrifolia* (Burm. F.) Nakai; Liu *et al.* 2012). Initial evidence that the *DAM* genes play a role in dormancy arose from the naturally occurring peach *EVERGROWING* (*EVG*) mutant (Rodriguez *et al.* 1994). The *EVG* mutant does not form terminal vegetative buds in response to dormancy-inducing conditions such as shortened photoperiod and low temperatures, does not cease growth at terminal meristem, and does not enter an endodormant state (Rodriguez *et al.* 1994). This phenotype is attributed to the deletion of six tandemly arranged *Pp DAM* genes (Bielenberg *et al.* 2004) which demonstrate seasonal expression patterns (Jiménez *et al.* 2009). Based on expression profiling, these genes are hypothesized to have non-redundant roles in growth cessation and/or terminal bud formation, and may have undergone sub- or neofunctionalization (Jiménez *et al.* 2009). It should be noted that *EVG* also has a reduced level of cold hardiness in comparison to wildtype peach trees (Rodriguez *et al.* 1994, Arora *et al.* 1996), however it has not yet been investigated if the *PpDAM* genes have direct or indirect roles in this pathway.

Research from angiosperm *SVP*- and *DAM*-like genes across a range of angiosperms provide strong evidence that that these genes have roles in cessation of growth at meristem and terminal bud formation, and possibly other functional roles as well. Horvath (2009) proposes that dormancy may be partially regulated by *DAM* genes regulating *FT* homologs, considering the recent evolutionary divergence between *DAM* and *SVP* genes. Based on these studies and the observation that a sequence showing similarity to *SVP* was differentially regulated during white spruce bud formation (El Kayal *et al.* 2011), we chose to investigate the role of *SVP*-like genes in white spruce terminal bud formation.

1.4 The current study

While *SVP*-like genes have been well characterized in angiosperms species, prior to this research, little if anything was known about *SVP*-like genes, their function and regulation in conifers. The long-term goal of this research is to determine if *MADS-box* genes related to *SVP* regulate bud formation and/or transition to dormancy in white spruce.

The following are the specific objectives of my thesis research:

(1) identify the white spruce genes most closely related to functionally characterized *SVP* genes of angiosperm species, and determine their evolutionary relationship;

(2) establish if the expression profiles of candidate white spruce *SVP*-like genes correlate with the developmental events of bud formation; and

(3) discover upstream regulators of white spruce *SVP*-like genes using yeast one-hybrid and *in silico* promoter motif identification.

Through addressing these objectives, I tested the following hypotheses: (1) white spruce *SV*P-like genes share a common ancestor with angiosperm *SVP*-like genes; (2) white spruc*e SV*P-like genes are involved in bud formation and possibly dormancy establishment; and (3) white spruce *SV*P-like genes are regulated by transcription factors which have also been found to regulate bud formation or dormancy acquisition in other species.

This thesis is composed of four chapters. Chapter 1 contains an overview of the photoperiodic flowering pathway and a summary of background material related to the molecular and developmental processes involved in growth cessation and bud formation. Chapter 2 presents a phylogenetic analysis of the white spruce genes related to *SVP,* and hypothesized functions of these genes based on qRT-PCR transcript profiling data obtained from developing white spruce buds. Chapter 3 is an investigation of the promoters for two of these white spruce

genes, identifying transcription factors and other regulatory molecules that may regulate their expression. Chapter 4 presents a synthesis of these results and conclusions, and proposes future directions.

Chapter 1 Figures

Figure 1.1 Distribution of white spruce trees across Canada (Natural Resource Canada, Canadian Forest Services, 2015b). This figure is a copy of an official work that is published by the Government of Canada and the reproduction has not been produced in affiliation with, or with the endorsement of the Government of Canada.

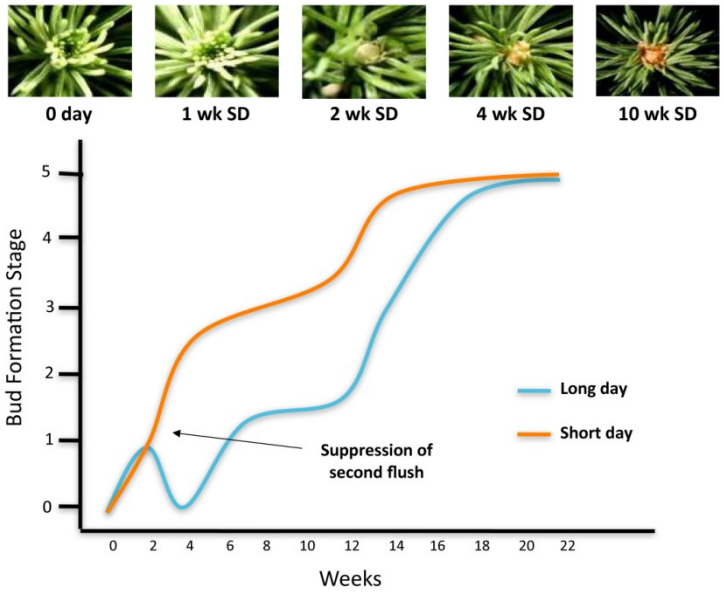


Figure 1.2 Summary of phenotypic stages of bud formation across long day and short day conditions from Hamilton *et al.* (2016). Under long day conditions stage of bud formation shifts back to an average of stage 0 at four weeks because of the occurrence of second flush. This observation is not seen in the short day conditions because second flush is repressed. Modified from Hamilton *et al.* (2016).

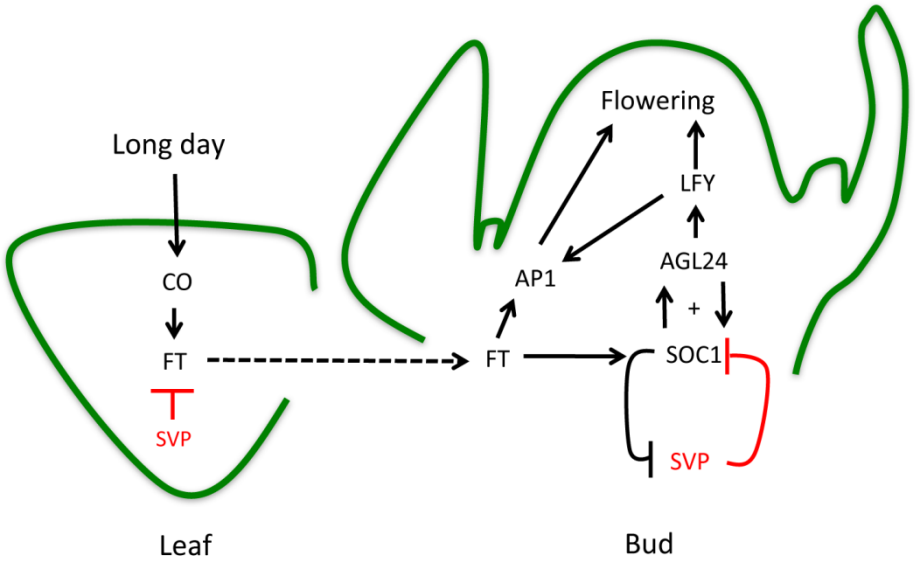


Figure 1.3. A simplified summary of a subset of transcription factors involved in the flowering pathway in *Arabidopsis* under long photoperiods (modified from Amasino 2010 and Andrés *et al.* 2014). Under long days in the leaf the CONSTANS (CO) protein is stabilized in the light to induce transcription of *FLOWERING LOCUS T* (*FT*). FT is translocated to the shoot apical meristem to upregulate *APETELA1* (*AP1*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*). SOC1 and SHORT VEGETATIVE PHASE (SVP) act to inhibit each other's expression. SOC1 and AGAMOUS-like 24 (AGL24) upregulate one another, and dimerize to increase transcription of *LEAFY* (*LFY*). LFY transcribes *AP1*, and LFY and AP1 will go on to induce transcription of downstream genes involved in inducing flowering.

2.0 Chapter 2: Roles for *SHORT VEGETATIVE PHASE/AGAMOUS-LIKE 24-like* genes in distinct phases of white spruce apical bud formation

2.1 Introduction

Successful timing of the transition from active growth to dormancy is critical to the survival of perennial species in Northern temperate forests. Endodormancy (hereafter referred to as dormancy) is the cessation of growth in meristematic tissue, in which growth will not resume even under permissive conditions (Rohde and Bhalerao 2007). Processes leading to dormancy acquisition are interconnected, since bud formation (Ruttink *et al.* 2007) and growth cessation (Weiser 1970, Kalcsits *et al.* 2009) are important for cold acclimation, and apical bud formation is a prerequisite for dormancy induction (Rohde and Bhalerao 2007). Vegetative bud formation is the process by which bud scales form to protect and enclose the shoot apical meristem, and leaf primordia and subtending stem units for the next growing season are created. While photoperiod is the primary environmental cue to induce bud initiation in many tree species (Ingvarsson *et al.* 2006, Luquez *et al.* 2007, Cooke *et al.* 2012, Ding and Nilsson 2016), we have demonstrated that white spruce (*Picea glauca* (Moench) Voss) is able to complete bud formation in the absence of dormancy inducing cues such as shortening photoperiod and low temperatures, although short days (SDs) accelerate bud formation by suppressing the occurrence of lammas growth (El Kayal *et al.* 2011, Hamilton *et al.* 2016).

To prevent damage to the shoot apical meristem (SAM), perennials integrate endogenous and environmental signals to promote correct timing of bud formation during the autumnal transition from active growth. The SAM contains densely packed cells and maintains the population of undifferentiated cells, some of which go on to differentiate into leaf or

reproductive primordia (Rohde *et al.* 2000). As with the activity-dormancy transition, the SAM also integrates various endogenous and environmental cues to regulate the transition from vegetation to reproductive growth (Hartmann *et al.* 2000, Tao *et al.* 2012). An accumulating body of research suggests that regulatory components of the network signaling the transition from vegetative to reproductive growth and the transition from active growth to dormancy are evolutionarily conserved (reviewed in Petterle *et al.* 2013 and Singh *et al.* 2017). For example, the phosphatidylethanolamine-binding protein (PEBP) family member *FLOWERING LOCUS T* (*FT*) acts as a floral promoter (reviewed in Pin and Nilsson 2012). *FT* orthologs have been shown to regulate bud set and growth cessation in angiosperm perennial species such as poplar (Böhlenius *et al.* 2006, Hsu *et al.* 2011), while a related set of PEBP genes named *FT/TERMINAL FLOWER1-like (FTL1)* have been implicated in bud formation and growth cessation in the conifer species Norway spruce (*Picea abies* (L.) H. Karst., Gyllenstrand *et al.* 2007, Karlgren *et al.* 2011, Klintenäs e*t al.* 2012). Following the seminal finding that *FT* orthologs regulate bud set and growth cessation, other orthologs of regulators downstream of the *CO/FT* module have been identified that govern aspects of the activity-dormancy transition in apical buds, including *FLOWERING LOCUS D-like* 1 (*FDL1*, Tylewicz *et al.* 2015), *Like APETELA1* (*LAP1*, Azeez *et al.* 2014), and *AINTEGUMENTALIKE1* (*AIL1*, Karlberg *et al.* 2011). Taken together, these studies suggest that these and other putative orthologs in the regulatory network that control time to flowering may function as regulators of bud formation in conifers such as white spruce.

Several of the aforementioned genes belong to the MADS-box family of transcription factors (Tao *et al.* 2012, Hartmann *et al.* 2000). Within the large and diverse MADS-box family is a subgroup of genes referred to as MIKC-type genes based on their four conserved domains.

The MIKC gene *SHORT VEGETATIVE PHASE* (*SVP)* is widely known as an important negative regulator of flowering, and the *Arabidopsis SVP* knockout mutant *svp-41* causes an early flowering phenotype (Hartmann *et al.* 2000). *SVP*-like genes, also called *DORMANCY-ASSOCIATED MADS-BOX* (*DAM*) genes, have been implicated in regulation of bud formation in peach (*Prunus persica* (L.) Batsch, Jiménez *et al.* 2009, Yamane et al. 2011), as well the acquisition and/or release of dormancy in peach (Jiménez *et al.* 2009), raspberry (*Rubus idaeus* L.; Mazzitelli *et al.* 2007), Japanese apricot (*Prunus mume* (Siebold) Siebold & Zucc.; Sasaki *et al.* 2011), leafy spurge (*Euphorbia esula* L.; Horvath *et al.* 2010) and kiwifruit (*Actinidia deliciosa* (A. Chev.) C.F. Liang & A.R. Ferguson; Wu *et al.* 2011). Despite the high sequence similarity between *AGAMOUS-LIKE 24* (*AGL24*) and *SVP*, these genes have opposing role with *AGL24* being a positive regulator of flowering (Parěnicová *et al.* 2003, Michaels *et al.* 2003). To our knowledge there have been no studies that have looked into the role of *AGL24*-like genes in the activity-dormancy transition. The phase change between vegetative and reproductive growth at the SAM is regulated by genes that include *AGL24* and *SVP* (Becker and Theissen 2003).

Previously, we conducted a microarray transcriptomic analysis that identified genes with sequence similarity to *Arabidopsis*. *CO/FT* network regulators of flowering, including *SVP*, that are differentially expressed during white spruce bud formation (El Kayal *et al.* 2011). Based on this finding, in this study, we explored the hypothesis that a lineage of white spruce genes with sequence similarities to angiosperm SVP genes may play a role in regulating developmental events associated with the activity-dormancy transition in apical buds of white spruce. We first identified candidate genes to study by generating phylogenies of a broad sampling of MIKC genes across multiple species. We also investigated patterns of expression during bud

development using quantitative reverse transcriptase PCR (qRT-PCR), and used the resulting transcript profiles to speculate on roles of these genes.

2.2 Materials and Methods

2.2.1 Plant material

White spruce seedlings originating from Québec provenances obtained from the Canadian Forestry Service (Québec, Canada) were used to generate materials for qRT-PCR experiments. Seedlings represented the same population used in El Kayal *et al.* (2011), Galindo-Gonzalez *et al.* (2012) and Galindo-Gonzalez *et al.* (2015). Trees in their second growth cycle were grown under long day conditions (LD; 16 h days/8 h nights) at 20°C with 50% relative humidity for approximately eight weeks of active growth. At Day 0, half of the plants were switched to short day conditions (SD; 8 h days/16 h nights) at 20°C with 50% relative humidity. A complete randomized design was used for the experiment, with plant materials within each photoperiod condition arranged within randomized blocks. Apical shoot tips/developing buds from four to five plants were harvested from the leader at five time points (Day 0, 7, 14, 28 and 70) following transfer to SD or LD conditions. Following harvest, tissues were immediately frozen in liquid nitrogen and stored at -80°C.

2.2.2 Phylogenetic analyses

Nucleotide and amino acid sequences of 88 MIKC sequences (32 from white spruce, three from *Physcometrella patens*, 28 from *Arabidopsis*, 25 from range of angiosperm species) sampled from 14 different species were obtained from GenBank and GenPept, respectively (Table S1). *Arabidopsis* sequences were used as a backbone to resolve major topologies, and additional characterized *SVP/AGL24* (*SA*) genes from various angiosperm species were added to

diversify the *SA* clade. White spruce sequences were identified by submitting *Arabidopsis* MIKC coding sequences, from Pařenicová *et al.* (2003), to BLASTx of the GCAT database (Rigault *et al.* 2011). This white spruce expressed gene resource represents 27,720 unique, mostly full-length cDNA sequences, developed from sequencing of 42 different libraries (Rigault *et al.* 2011). The top ten white spruce hits from each query were pared down to a non-redundant list of representative unigenes after constructing a tentative Neighbour Joining phylogenetic tree. Sequences were deemed redundant is they had a >95% similarity. If the contig comprised multiple ESTs (sequences representing physical cDNA clones), the longest clone that had been sequenced from both the 5' and 3' ends were used. If these were not available, clones that had been sequenced from the 5' end was used. Three *MADS-box* genes from the moss species *Physcomitrella patens*, a representative of an early diverging lineage of land plants, were selected as outgroups. Nucleotide sequences were derived from the open reading frame (ORF) of the cDNA sequences using NCBI's ORF Finder (ncbi.nlm.nih.gov/orffinder/). White spruce amino acid sequences were not available on GenPept and were predicted by translating the cDNA ORF into amino acids.

Amino acid alignments (Figure 1) were generated in MAFFT v7 (Katoh and Standley 2013), using amino acid partition by L-INS-i (single domain alignment) parameters. Nucleotide sequences were then forced to appropriate codon triplet to their respective amino acid sequence in Mesquite v2.75 (Maddison and Maddison 2011).

Phylogenetic relationships of the amino acid and nucleotide partitions were inferred using maximum parsimony (MP) and maximum likelihood (ML). MP searches were conducted in PAUP* 4.0b10 (Swofford 2002) with the following parameters: 300 random addition replicates, terminal branch rearrangement (TBR), 50 trees held in the construction of the initial starting tree,

1000 bootstrap (BS), 1000 trees with a length greater than or equal to 1 held during 1000 times BS.

We conducted both unweighted (e.g., Fitch 1971) and weighted searches under parameters above to test if the topology of the major clades would be altered by a greater importance assigned to the more highly conserved gene regions. Weighted searches incorporated variable weighting schemes according to MIKC domain conservation: (1) the "M" (MADS-box) domain, which recognizes and binds to the MADS-box domain on downstream target genes and facilitates dimerization, (2) the "I" (intervening) domain that specifies the formation of DNA dimers (Theissen *et al.* 1996); (3) the "K" (keratin-like) domain participates in protein-protein interactions and is well conserved (Kaufman *et al.* 2005), (4) the "C" (C-terminal) domain that has roles in transcriptional activation and higher order complex formation (Kaufmann *et al.* 2005, Cseke and Podila 2004); (5)"N" (N-terminal) domain precedes the "M" domain, however it is only found in a minority of genes. As the "M" and "K" domain are highly conserved, and the "I" and "C" domains are less well conserved across land plants (Davies *et al.* 1996, Fan *et al.* 1997, Parěnicová *et al.* 2003, Egea-Cortines *et al.* 1999), we ran weighted analyses following a weighting scheme. Domains were weighted according to the defined "N", "M", "I", "K", "C" regions outline in Henschel *et al.* (2002): "N" domain weight of 0.5, "M" domain weight of 3, "I" domain weight of 2, "K" domain weight of 3, "C" domain weight of 1.

Maximum likelihood (ML) analyses were conducted with nucleotide and amino acid partitions using GARLI 2.0 (molecularevolution.org/software/phylogenetics/garli/, Zwicki 2006, Bazinet and Cummings 2008, Sukumaran and Holder 2010). Models of molecular evolution for the nucleotide and amino acid data were determined using the AIC (Akaike Information Criterion) and BIC (Baysian Information Criterion) as implemented in jModelTest2 (Darriba *et*

al. 2012, Guidon and Gascuel 2003) and ProtTest (Abascal *et al.* 2005), respectively. The GTR+I+Γ (general time reversible + invariable + gamma) substitution model was selected for nucleotide data. The JTT+I+Γ (Jones-Taylor-Thornton + invariable + gamma) model was selected for amino acid data. All tree searches were conducted with estimated state frequencies, proportion of invariant sites was estimated, 4 rate categories, 100 times bootstrap**.**

Alternative topologies of constraint trees were tested against the original unconstrained ML tree in PAUP* using the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999). Likelihood settings in PAUP* were adjusted to meet the optimality parameters of the original GARLI analysis of the unconstrained tree that correspond to the GTR+I+Γ model. Likelihood scores were estimated in PAUP* using the Roger-Swofford approximation method (Rogers and Swofford 1998) branch-length optimization with the one-dimensional Newton-Raphson with pass limit= 20 and delta = 1e-06. The SH test with the following parameters 1000 RELL (Resampling Estimated Log-Likelihoods) bootstrap one-tailed test, assuming $p<0.05$ was significant.

2.2.3 qRT-PCR

RNA extractions were performed as described by Pavy *et al.* (2008). Quantity and quality was assessed with an Infinite® 200 NanoQuant (Tecan Group Ltd., Männedorf, Switzerland) and gel electrophoresis, as well as 2100 Bioanalyzer (Agilent, Mississauga, ON, Canada) for a subset of samples. Primer design was carried out using Primer Express® v3.0 (Applied Biosystems, Carlsbad, CA, USA; Table 2.2). cDNA synthesis and qRT-PCR using a SYBR Green assay was carried out according to El Kayal *et al.* (2011). Three to four biological replicates and two technical replicates were used for each time point. Reactions were performed using an Applied Biosystems® 7500 Fast Real-Time PCR System (Applied Biosystems, Foster

City, CA, USA) or an Applied Biosystems® Quant Studio™ 6 Flex Real Time PCR System (Applied Biosystems, Carlsbad, CA, USA). Standard curves were used to quantify transcript abundance of the reference gene *TRANSLATION INITIATION FACTOR5A* (*TIF5*A, GQ00410_I10, GenBank DR448953). Due to pipetting error *TIF5A* values from the qRT-PCR plate run for GQ03707_I04 was substituted for the *TIF5A* values from another plate, after being normalized to the calibrators present on the plates.

2.2.4 Statistical analyses

Statistical analyses to detect significant differences of transcript abundance was carried out in RStudio v3.4.1 (R Core Team 2017), the FDR (false discovery rate) test for the MANOVA, the Levene test for homogeneity of variance using the "car" package v2.1-5 (Fox and Weisberg 2011), and the "lsmeans" package (Lenth 2016). A split-plot two-way ANOVA was used for analysis of the reference gene expression. Transcript quantities log transformed to fulfill normality and heterogeneity of variance assumptions, and a MANOVA (multivariate analysis of variance) was run. Shapiro-Wilk test for normality and histograms were used to assess normality. We were unable to acquire a p value > 0.05 for homogeneity of variance photoperiod for GQ03118_H14 ($p = 0.385$). A FDR test (alpha = 0.05) was used to determine significant differences.

2.3 Results

2.3.1 A clade of white spruce genes is sister to the angiosperm clade containing *Arabidopsis thaliana SVP* and *AGL24*

A total of 32 MIKC-like cDNA sequences were identified in the white spruce expressed gene catalogue representing 27,720 unique cDNA sequences derived from 42 different libraries

(Rigault *et al.* 2011). We found all four MIKC domains were present and conserved in our alignment (Figure 2.1). We identified six major clades consistent with gene families (Figure 2.2, the following ML and MP BS support are listed): *ARABIDOPSIS NITRATE REGULATED 1* (*ANR1*; 95, 76), *FLOWERING LOCUS C* (*FLC*; 100, 100), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*; 98, 69), *SEPELLATA* (SEP; 75, 54), *SHATTERPROOF* (*SHP*; 66, 62) and *SVP/AGAMOUS-LIKE 24* (*SA*; 85, 99). The topology of the six major clades were consistent across phylogenetic analyses (Figure 2.2, Supplemental Figure S2.1, S2.2, S2.3 and S2.4). For each of these major clades, white spruce genes form sister clades to the angiosperm clades. Consistent with this pattern, seven white spruce sequences form a sister clade (ML and MP BS support: 100, 100) to the *SA* clade (Figure 2.1). An eighth gene, *Pg*GQ03118_H14, was not a part of the *PgSAL* clade, but resolved as ancestral to the *PgSAL* clade in the nucleotide ML before BS (Supplemental Figure S2.6), nucleotide weighted MP (Supplemental Figure S2.3) and the nucleotide MP (Supplemental Figure S2.2) analyses.

To confirm the robustness of the sister relationship between *SA* and *PgSAL* genes, we performed an SH test (Shimodaira and Hasegawa 1999). Six alternative topologies were created to constrain the seven white spruce genes that are sister to the *SA* clade, the *SOC1* clade (Supplemental Figure S2.11), the *SEP* clade (Supplemental Figure S2.9), the *SHP* clade (Supplemental Figure S2.10), *ANR1* (Supplemental Figure S2.7), *FLC* (Supplemental Figure S2.8) and the *AGL15* clade (*AGAMOUS-LIKE15*, *AtAGL15*, *AtAGL18*, *Pg*GQ03118_H14; Supplemental Figure S2.6). The SH test indicated that the ML score of the constraint trees were significantly different from the ML score of the unconstrained tree (Table 2.1), thus the topology in Figure 2.2 is the best explanation of the data.

2.3.2 White spruce *SAL* genes show distinct transcript abundance profiles during terminal bud development

Based on their phylogenetic relationship with the *SA* clade and presence of MIKC motifs (Figure 2.1), the seven white spruce sequences sister to the *SA* clade were named *PgSA-like1* (*PgSAL1*, GQ03605_C12), *PgSA-like2* (*PgSAL2*, GQ03707_I04), *PgSA-like3* (*PgSAL3*, GQ02822_N14), *PgSA-like4* (*PgSAL4*, GQ03702_K12), *PgSA-like5* (*PgSAL5*, GQ03806_I20), *PgSA-like6* (*PgSAL6*, GQ04010_J13) and *PgSA-like7* (*PgSAL7*, GQ03232_K15). Since these seven *PgSAL* sequences and the *SA* angiosperm sequences are inferred to have had a common ancestor, we hypothesized that at least some of the *PgSAL* genes might play roles during the activity-dormancy transition in white spruce. As the first exploration of this hypothesis, qRT-PCR transcript profiling was carried out over the course of apical bud development under both short days and long days for each of these seven genes plus the more distantly related GQ03118_H14 (Figure 2.3). Bud formation followed the same developmental progression as reported in El Kayal *et al.* (2011). Developing buds were first visible between 7 and 14 days. At 70 days, SD buds had completed development but were not dormant, while LD buds were still under development. While lammas growth (second flush) can occur in white spruce under LD conditions (El Kayal *et al.* 2011, Hamilton *et al.* 2016), buds were not sampled from any trees that showed indications of lammas growth.

All seven *PgSAL* genes and *Pg*GQ03118_H14 displayed significant differences in transcript abundance across time during the course of SD or LD bud formation (Figure 2.3). A split-plot two-way ANOVA using Ct's demonstrated *TIF5A* transcript abundance was significantly different across "photoperiod" nested in "day" ($p = 0.015$), with a significant difference in short-day Day 0 versus Day 70 ($p = 0.048$). As a result, we will not compare

directly between photoperiods, and make conservative statements about any significant differences for short Day 0 versus Day 70 for our genes of interest. Relative transcript abundance and transcript profiles over bud development differed markedly between genes. Five of the eight genes – *PgSAL1* (GQ03605_C12), *PgSAL2* (GQ03707_I04), *PgSAL3* (GQ02822_N14), *PgSAL4* (GQ03702_K12), and *PgSAL5* (GQ03806_I20) - exhibited significantly greater transcript abundance during the first two weeks of bud formation than at later time points. *PgSAL1* and *PgSAL5* expression is reduced in SD at Day 28, whereas LD expression did not decline until Day 70. *PgSAL3* expression began to decline in LD at Day 28, and was followed by a further decline in expression at Day 70 in both SD and LD conditions (Table 2.3, Figure 2.3). *PgSAL6* (GQ04010_J13) exhibited significantly greater transcript abundance only at later time points, with an increase seen in SD at Day 28 and Day 70, and in LD at Day 70 (Table 2.3, Figure 2.3). *PgSAL7* (GQ03232_K15) expression significantly decreased at SD Day 28, whereas no significant change in expression was found in the LD treatment (Table 2.3, Figure 2.3). GQ03118_H14 expression fluctuated during bud formation. In SD, GQ03118_H14 declined at Day 14 and reached a maximum at Day 70. Peak expression of GQ03118_H14 during SD contrasted with the significant increase in expression during mid-phase development observed in LD, at Day 14 (Table 2.3, Figure 2.3).

Given the known roles of *AtSVP* and *AtAGL24* in the photoperiodic flowering pathway, we further tested whether photoperiod affected transcript profiles for any of these genes during bud development. *PgSAL3* showed a significant difference in response to overall SD and LD photoperiod treatments (Table 2.3, Figure 2.3). *PgSAL3* SD expression declined sooner in SD versus LD, with SD expression beginning to decrease at Day 28.

We observed that the four genes that share the most closely related evolutionary relationship, *SAL1*, *SAL2*, *SAL4* and *SAL5*, also have a similar expression profile across bud formation in SD and LD conditions (Figure 2.1, 2.2). These *SAL* genes have a greater transcript abundance at earlier time points in bud development, in comparison to later later time points in bud development.

2.4 Discussion

2.4.1 A clade of white spruce *SAL* genes is sister to functionally characterized angiosperm *SA* genes

Studies over the last decade have identified orthologs of photoperiodic flowering pathway genes, including multiple MIKC *MADS-box* genes such as *SVP* that regulate events during the activity-dormancy transition in perennial species (reviewed in Singh *et al.* 2017). However, most studies from which the current conceptual model of this regulatory network have been carried out in angiosperms. El Kayal *et al.* (2011) found that genes showing sequence similarity to MIKC *MADS-box* floral regulators, such as *SVP*, *MOTHER OF FT AND TFL1* (*MFT*) and *APETELA2* (*AP2*), are also differentially expressed during white spruce bud formation, suggesting that orthologs of *MADS-box* genes of the photoperiodic flowering pathway may function in regulation of bud formation in conifers. Thus, we addressed the hypothesis that white spruce genes with sequence similarity to *SVP* and *SVP*'s closest relative in *Arabidopsis*, *AGL24*, are involved in regulating processes associated with bud set in the coniferous species, white spruce.

As the first step, we demonstrated using multiple phylogenetic methods that seven MIKC white spruce sequences, denoted *PgSAL1* to *PgSAL7*, form a sister clade to the angiosperm clade

containing *SVP* from *Arabidopsis* and other species (Figure 2.2). *AtAGL24*, which is closely related to *AtSVP* but exhibits contrasting function (Hartmann *et al.* 2000, Yu *et al.* 2002), also fell within the angiosperm *SA* clade, as previously shown (Pařenicová *et al.* 2003). Other phylogenetic analyses carried out on *SA* genes from the angiosperm perennials – such as the *DORMANCY-ASSOCIATED MADS-BOX* (*DAM*) genes – identified *SA* orthologs within the same clade (Jiménez *et al.* 2009, Yamane *et al.* 2011, Mazzitelli *et al.* 2007, Sasaki *et al.* 2011, Horvath *et al.* 2010, Wu *et al.* 2011). Since angiosperms and gymnosperms are widely agreed to be sister clades (Qiu *et al.* 2010, Soltis *et al.* 2011, Wickett *et al.* 2014), conifer genes resolving as a sister clade with their most closely related angiosperm genes is consistent with their evolutionary history. The relationship of conifer genes resolving as sister to their angiosperm homologs is consistent with the topology of the other major clades in our phylogenetic trees. The sister relationship between the conifer *PgSAL* and angiosperm *SA* genes is reflective of the relationship between conifer and angiosperm genes described for other gene families. A sister relationship has been reported, for example, between conifer *FTL1-like* genes that are implicated in bud formation and angiosperm *FT* and *TERMINAL FLOWER1* (*TFL*1) genes implicated in time of flowering by Karlgren *et al.* (2011) and Klintenäs *et al.* (2012), as well as for conifer and angiosperm MYBs implicated in regulation of secondary metabolism pathways (Bedon *et al.* 2010). We provided additional evidence of the robustness of our topology by performing a constraint analysis using an SH test. This test indicated that the topology of *PgSAL* and angiosperm *SA* as sister clades is a significantly better explanation of the data than alternative topologies. An eighth sequence, *Pg*GQ03118_H14 resolves near the *PgSAL* clade, but cannot conclusively labeled a member of a gene-specific clade in the nucleotide MP unweighted and weighted trees. We hypothesize that *Pg*GQ03118_H14, which showed weak association with

the *AGL15* clade, may have some conserved functions with the *AGL15* clade. Other weighting schemes were also tested ("N" = 0.5, "M" = 4, "I" = 2, "K" = 3, "C" = 1), but were not found to significantly differ from the topology of the original weighted tree (data not shown). It is possible that *Pg*GQ03118_H14 may have developmental roles at the meristem. *Arabidopsis AGL15* and *AGL18* are expressed in the embryo and developing endosperm (Lehti-Shiu *et al.* 2005), and more recent evidence suggests *AGL15* may suppress *FT* expression (Fernandez *et al.* 2014). It is important to acknowledge that the accuracy of the SH test is dependent on the number of trees included in the analysis (Buckley *et al.* 2001).

The number of *SAL* genes differs between species. White spruce appears to have at least seven *SAL* genes based on an extensive expressed gene catalogue (Rigault *et al.* 2011), although we cannot discount that additional genes may be identified as the white spruce draft genome sequence matures to a reference quality genome assembly (Birol *et al.* 2013). *Arabidopsis* has two *SA* genes (*SVP* and *AGL24*; Yu *et al.* 2002), while peach has a minimum of six (*DAM1-6*; Li *et al.* 2009, Jiménez *et al.* 2009), kiwi has a minimum of four (*SVP1-4*; Wu *et al.* 2011) and Japanese apricot has a minimum of six (*DAM1*-6; Sasaki *et al.* 2011). Within the small subset of species that we considered, perennials appear to possess a greater number of *SAL* genes than annuals. As has been proposed for other conifer gene families (e.g. Bedon *et al.* 2010), the expansion of the *SVP* subfamily in perennials, and the maintenance of these duplicated genes, may reflect functional redundancy and/or regulation of additional processes associated with the perennial lifestyle by signaling networks analogous to the photoperiodic signaling network. The seven *PgSAL* genes plus the *AGL15-like Pg*GQ03118_H14 showed both distinct and overlapping expression profiles over the course of bud formation under LD and SD, as was found for DAM genes in peach (Li *et al.* 2009), supporting the notion that *PgSAL* genes perform both redundant

and non-redundant roles in regulating gene expression during early, mid, and late bud formation in white spruce. Perennials may require more genes in order to tightly regulate processes of vegetative and reproductive bud formation, initiation of bud set, bud burst, dormancy initiation and dormancy release. Annuals, such as *Arabidopsis,* simply need to regulate processes involved in reproductive bud formation. Therefore, it is reasonable to expect perennials would have a larger, more diversified group of *SVP*-like genes in order to tightly regulate these processes.

2.4.2 A subset of *PgSALs* may share conserved role in bud formation and/or growth cessation

Expression of *PgSAL1, PgSAL2*, *PgSAL3*, *PgSAL4*, and *PgSAL5* were significantly higher during the first two week of bud formation than at later stages of bud formation, suggesting that these transcription factors are positive regulators of bud development processes. Given the function of *SVP* and *AGL24* genes in angiosperm flowering time and development, we believe that *PgSAL* genes may also participate in cone development (Mouradov *et al.* 1998, Sundström *et al.* 1999). If the putative roles of *PgSAL1-5* hold true, they would contrast with the repressive role of *AtSVP* in floral transition (Hartmann *et al.*, 2000), and make their function more similar to that of *AtAGL24* (Michaels *et al.* 2003) in promoting flowering. Interestingly, this predicted function is analogous to the picture emerging for the Norway spruce *FT/TFL1-like* gene *FTL2* – a regulator of bud formation and growth cessation (Karlgren *et al.* 2011) – which has a biochemical function more similar to the flowering repressor *TFL1* than to the flowering activator *FT* (Klintenäs *et al.* 2012).

MADS-box genes are widely known for functioning as dimers and quaternary complexes (Riechmann 1996, Egea-Cortines *et al.* 1999, Honma and Goto 2001). It may also be possible

that the function of *SAL* genes is dependent on expression of their hetero-dimer partner. This possibility would add another layer of fine regulation for a process as complex as bud formation, which is reliant on multiple environmental cues as well endogenous signals.

The *PgSAL* genes did not show strong transcript abundance responses to photoperiod, a pattern consistent with our previous findings that SD is not required to initiate bud formation in white spruce (Hamilton *et al.* 2016). SD accelerates the completion of bud formation while suppressing lammas growth (Hamilton *et al.* 2016), consistent with our postulated roles for these genes in regulating early- and mid-stage bud formation processes.

Taken together, our results suggest that *PgSAL1, PgSAL2*, *PgSAL3*, *PgSAL4*, and *PgSAL5* are the most likely candidates to function as transcription factors in regulating bud formation and/or growth cessation, analogous to roles postulated for *SAL* genes in woody angiosperm species such as peach (Jiménez *et al.* 2009, Yamane *et al.* 2011), raspberry (Mazzitelli *et al.* 2007), Japanese apricot (Sasaki *et al.* 2011), leafy spurge (Horvath *et al.* 2010) and kiwifruit (Wu *et al.* 2011). Of these five genes, *PgSAL1* and *PgSAL4* are the most closely related genes based on phylogenetic analyses, and show similar patterns of expression. On the other hand, *PgSAL5* and *PgSAL6* are also closely related, but show opposite patterns of expression. The most divergent expression pattern is observed in *Pg*GQ03118_H14, which is not a *bona fide SAL* gene. These findings demonstrate that topology is not strong predictor of gene expression.

A limitation in our qRT-PCR analysis is that the amplicons were not subjected to sequencing to confirm target identity. Sequencing the amplicon would have confirmed that the desired amplicon has been transcribed, and that the desired amplicon was the only reaction product.

2.4.3 A *PgSAL* may have acquired novel functions

Further evidence for non-redundant functions comes from the distinct expression profiles exhibited by *PgSAL6*. *PgSAL6* showed significant upregulation only at later stages of bud formation, leading us to speculate that this transcription factor regulates processes associated with completion of bud formation and possibly transition to dormancy. At this time point, SD trees have largely completed bud formation and are transitioning to dormancy, while LD trees are still undergoing active bud development (El Kayal *et al.* 2010, Hamilton *et al.* 2016). Future experiments should focus on functional characterization of this *PgSAL*.

2.5 Conclusion

In this study, we have shown that conifer *SAL* genes likely share a common ancestor with angiosperm *SA* and *SAL* genes. Gene expression profiling suggests that the *PgSAL* genes may have acquired diverse regulatory roles during the course of bud formation. *PgSAL1, PgSAL2*, *PgSAL3*, *PgSAL4*, and *PgSAL5* exhibited gene expression that are consistent with overlapping but perhaps non-redundant *SA* roles in regulating early and/or mid stages of bud formation in white spruce. *PgSAL6* may regulate processes associated with later stages of bud formation and possibly dormancy transition, and thus may participate in a different signaling network. Further functional characterization of these *PgSAL* genes is warranted, given that these MIKC *MADS-box* genes potentially play novel roles that have yet to be described in angiosperms.

Given the well-documented role for angiosperm *SAs* in the seasonal response network regulating the transition to flowering, and evidence for the involvement of angiosperm *SA* genes in regulating bud formation, we hypothesize that *PgSAL1, PgSAL2*, *PgSAL3*, *PgSAL4*, and/or *PgSAL5* function as part of a conifer signaling network that shares an evolutionary history with the angiosperm *CO/FT* signaling network regulating bud formation (Singh *et al.* 2017) and

flowering transition (Andrés and Coupland 2012). Similar to *SVP* and *AGL24* in *Arabidopsis*, *PgSAL1, PgSAL2*, *PgSAL3*, *PgSAL4*, and *PgSAL5* may also have roles outside of timing of bud formation (Gregis *et al.* 2006, Gregis *et al.* 2009, Liu *et al.* 2009). At the same time, this study and others (e.g. Gyllenstrand *et al.* 2007, Karlgren *et al.* 2011, Klintenäs *et al.* 2012, Karlgren *et al.* 2013) suggest that the long period of evolutionary divergence between these taxonomic groups has also given rise to substantive differences between angiosperm and conifer activity-dormancy signaling networks. Consequently, care must be taken when applying the angiosperm model of signaling networks regulating bud formation, growth cessation and dormancy entrance to conifer species.

Chapter 2 Figures

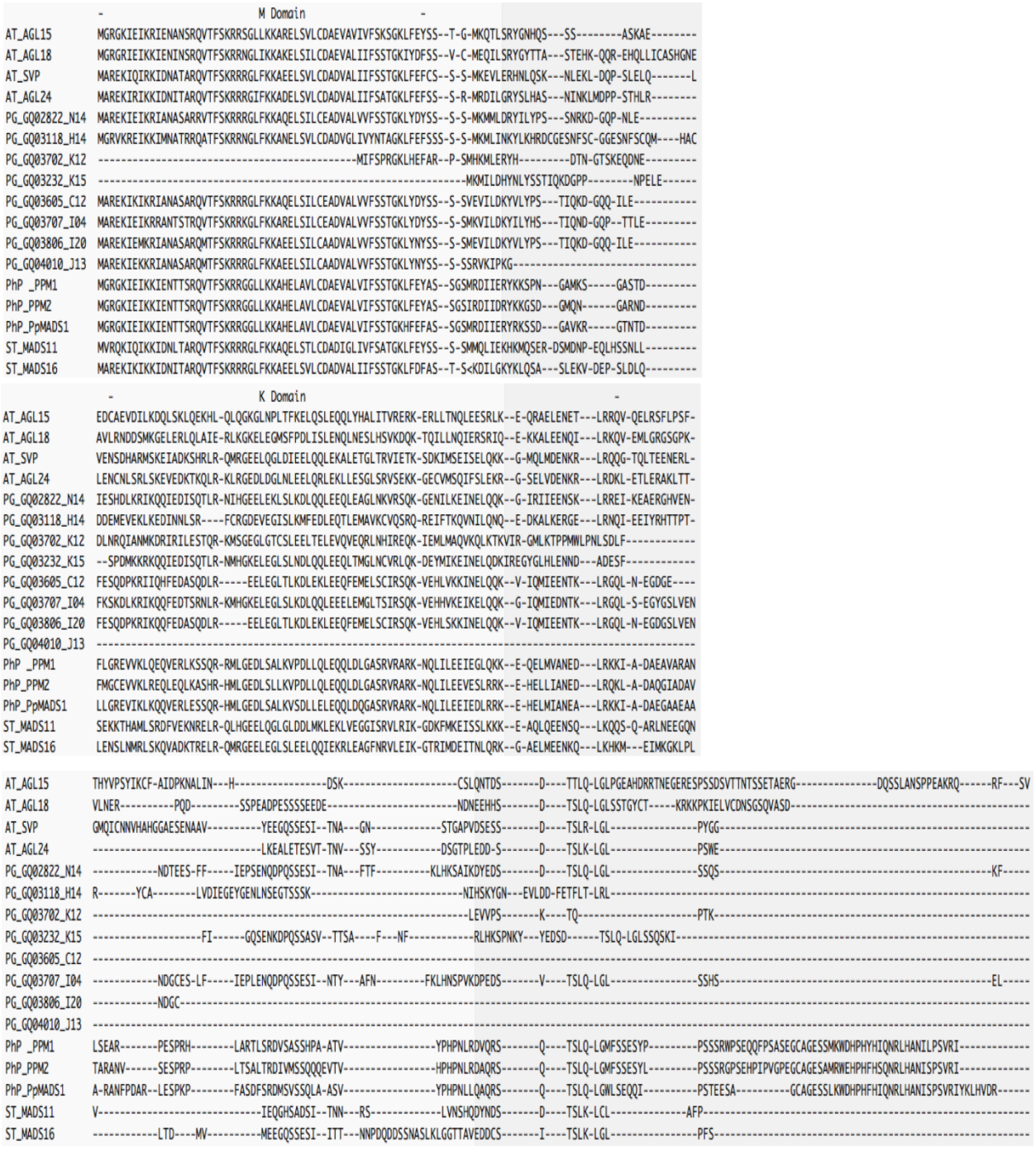


Figure 2.1 Amino acid alignment of MIKC sequences, with the individual "M" and "K" domains labeled. A subset of taxa was included for purpose of demonstrating conservation of the MIKC domains. The defined domains are based on *Physcomitrella patens* (PhP) from

Henschel *et al.* (2002). Here we also display seven *Picea glauca* (PG) *SAL* genes with their unique identifier from the GCAT assembly (Rigault et al. 2011) that is included in each sequence's NCBI flat file, *Arabidopsis thaliana* (AT) *SA* genes, and *Solanum tuberosum* (ST) genes. *AtAGL15*, *AtAGL16* and *Pg*GQ03118_H15 were also included based on their placement as closely related to the *PgSAL* clade based on the ML prior to BS.

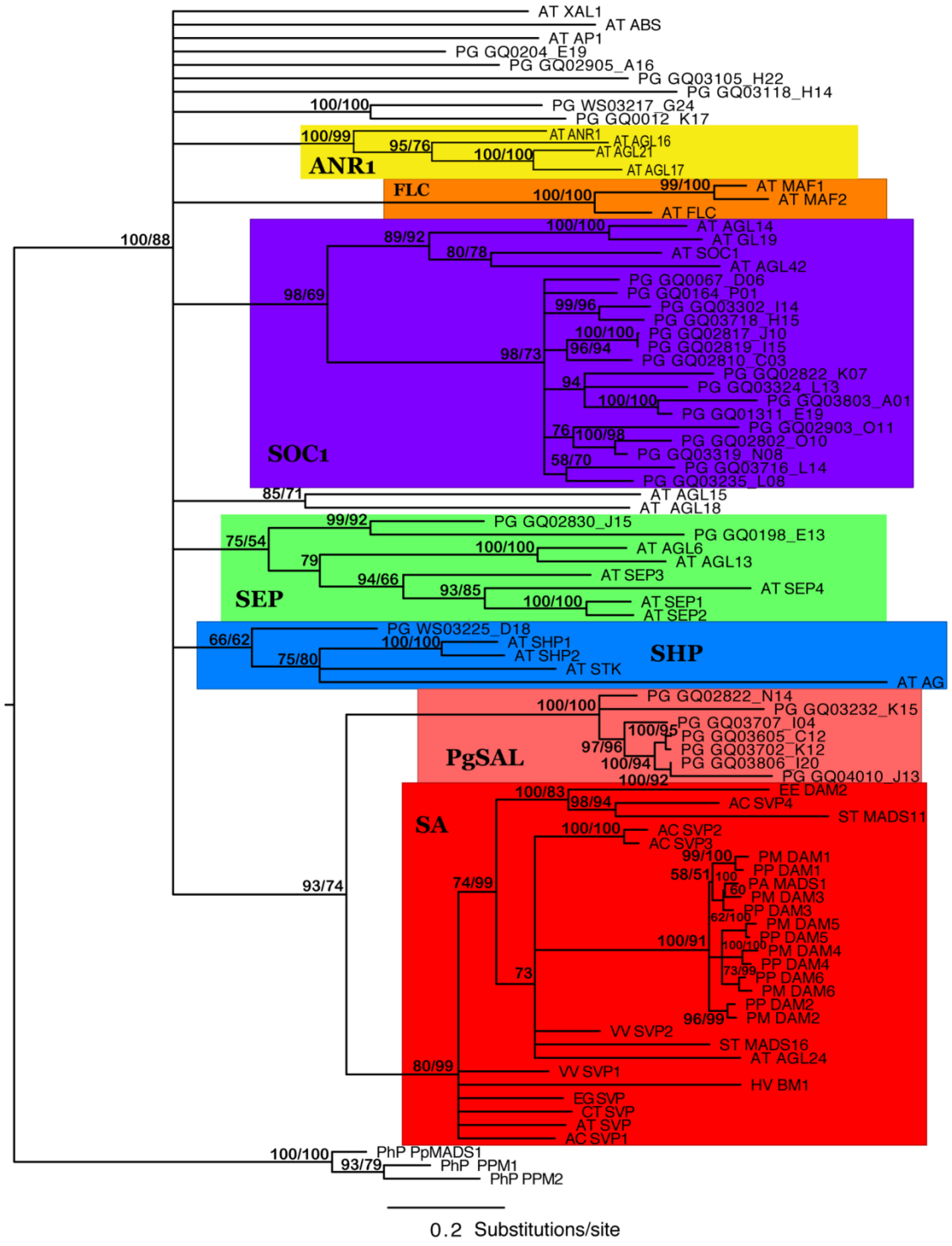


Figure 2.2 Maximum likelihood (ML) tree constructed from MIKC nucleotide partition (-ln = 42000.21361). Branches with less than 50% bootstrap support have been collapsed. Values

above nodes represent bootstrap values (maximum likelihood/maximum parsimony). Clade names, based on gene function, are indicated in boxes. The following abbreviations accompanying gene names refers to species of origin: AC = *Actinidia chinensis*, AT = *Arabidopsis thaliana*, CT = *Citrus trifoliata*, EG = *Eucalyptus grandis,* EE = *Euphorbia esula,* HV = *Hordeum vulgare*, PhP = *Physcomitrella patens*, PG = *Picea glauca*, PA = *Prunus avium,* PM = *Prunus mume,* PP = *Prunus persica,* ST = *Solanum tuberosum*, VV = *Vitis vinifera*. The white spruce genes in the *SAL* clade were later named *SAL1* (GQ03605_C12), *SAL2* (GQ03707_I04), *SAL3* (GQ02822_N14), *SAL4* (GQ03702_K12), *SAL5* (GQ03806_I20), *SAL6* (GQ04010_J13) and *SAL7* (GQ03232_K15).

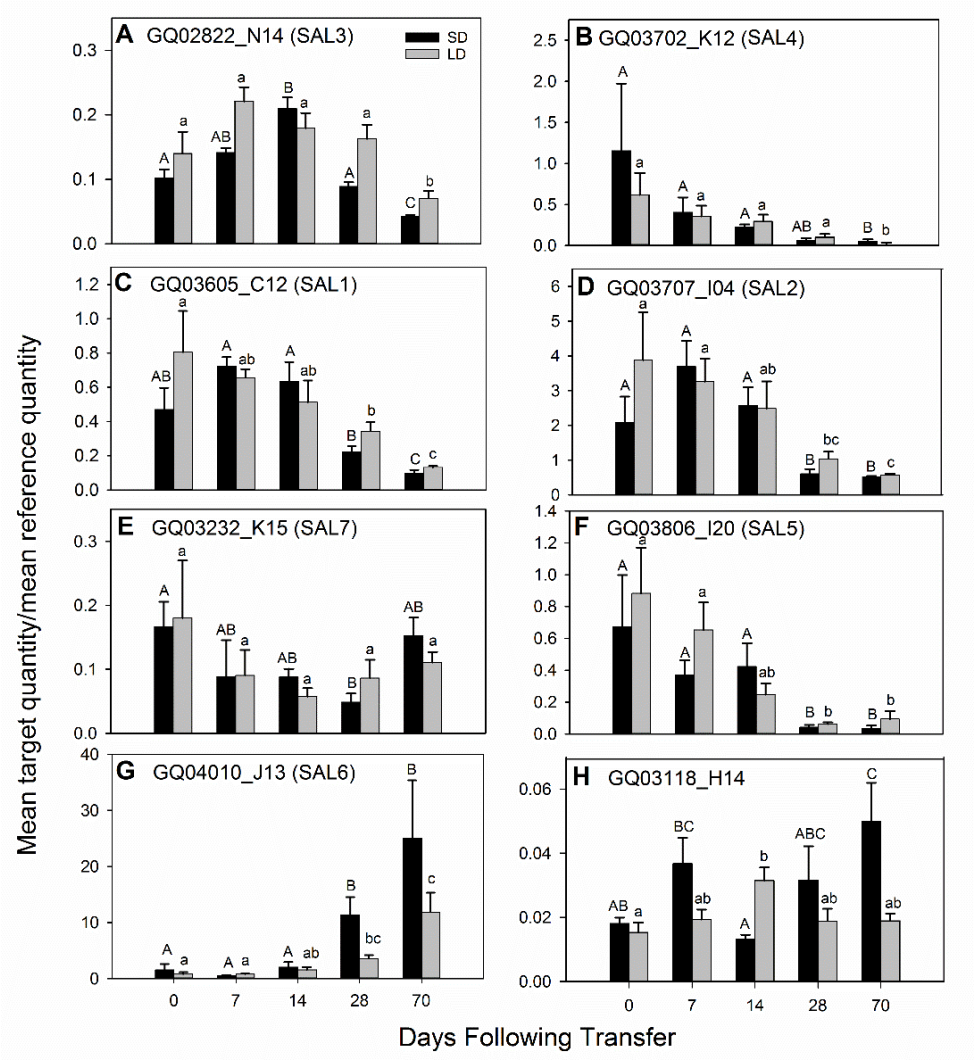


Figure 2.3 Transcript abundance profiles. Expression data corresponds to eight white spruce MIKC genes quantified in terminal shoot apices undergoing bud development under either SD or LD conditions. Transcript abundance was quantified by qRT-PCR using a standard curve method. *TIF5A* was used as a reference. Standard error bars represent three to four biological replicates. Letters above bars represent FDR grouping as determined by a MANOVA. Upper case letters represent significant differences ($p<0.05$) across time points within short days, and lower-case letters represent significant differences across time points within long days. Statistical comparisons are not made between photoperiod within days.

Chapter 2 Tables

Table 2.1 Shimodaira-Hasegawa (SH) test of alternative topologies. SH test was performed on constrained maximum likelihood (ML) trees. Log-likelihood score of the original ML tree is significantly greater than the alternative constraint trees (see Supplementary Figures S6-11)

Tree	Log-likelihood score (-ln)	p-value
Unconstrained	4200.21	-
SEP constraint	42619.76	<0.001
SHP constraint	42499.24	<0.001
SOC1 constraint	42143.22	0.002
FLC constraint	42228.67	<0.001
ANR1 constraint	42145.91	0.004
AGL15 constraint	42086.43	0.031

Table 2.2 Gene specific primers used for qRT-PCR analysis. Primers were designed with Primer Express® v3.0.

Gene	Primer Name	Primer Sequence (5' to 3')
GQ02822_N14	GQ02822_N14 FW	CAGATGTAGCCCTCGTCGTTTT
	GQ02822_N14 RV	ATGCTGGAGCTCGAGTAGTCGTA
GQ03702_K12	GQ03702_K12 FW	CGGGAGCTATCGATTCTATGTGA
	GQ03702_K12 RV	TAGTCGTACAGCTTCCCAGTTGAA
GQ03605_C12	GQ03605_C12 FW	GGCCCGCGAGAAAATAAAAA
	GQ03605_C12 RV	CCTGCGCCTCTTCGAGAAC
GQ03707_I04	GQ03707_I04 FW	CACAAGACTGCCATATCCTTCACT
	GQ03707_I04 RV	GGGAATACAAATGATAGAGGACAATACA
GQ03232_K15	GQ03232_K15 FW	CGCTTTCGAAGTACGGTGTTG
	GQ03232_K15 RV	GGCCTGTGGAGAATAACCCTAA
GQ03806_I20	GQ03806_I20 FW	ACCCCCCGTCATCTGAATCTAT
	GQ03806_I20 RV	TAGCTGCAAGGAAGTAACATAATCATC
GQ04010_J13	GQ04010_J13 FW	TTTGTCGTTTGATTTTAGGGTTCTC
	GQ04010_J13 RV	CCGAAGGCCTACACCAAGATT
GQ03118_H14	GQ03118_H14 FW	GGAGGGTAGGCTTTGCTTTGT
	GQ03118_H14 RV	TGCCAATTCCCCACAGACA
TRANSLATION INITIATION FACTOR5A (*TIF5A*)	GQ00410_I10 FW	TCGGCGGTGGCAGAGT
	GQ00410_I10 RV	TCCCCACAACTACGAAATCTCA

Table 2.3 p-values from a multivariate analysis of variance (MANOVA) performed on quantitative reverse transcriptase PCR (qRT-PCR) values of white spruce terminal buds. MANOVA's were performed across photoperiod and across time (i.e. day) nested in each photoperiod. Pillai test was used to calculate approximate F-value for the overall MANOVA, and sum of squares was used to calculate the F-value for the ANOVAs applied to the individual genes. Shapiro-Wilk test for normality and histograms were used to assess normality, and the Levene test was used to assess for homogeneity of variance. Statistics were conducted in RStudio, with an alpha value of 0.05.

	p-value	
MANOVA across all genes	Photoperiod	Photperiod/Day
Degrees of freedom	1	8
Pillai	0.585	2.958
Approx. F-value	4.404	2.347
p-value	0.002	<0.001
Gene	Photoperiod	Photperiod/Day
GQ02822_N14/*SAL3*		
Degrees of freedom	1	8
Sum of squares	3.078	30.471
F-value	13.205	16.354
p-value	<0.001	<0.001
GQ03702_K12/*SAL4*		
Degrees of freedom	1	8
Sum of squares	0.297	24.075
F-value	0.571	5.791
p-value	0.455	<0.001
GQ03605_C12/*SAL1*		
Degrees of freedom	1	8
Sum of squares	0.030	1.351
F-value	1.655	18.726
p-value	0.207	<0.001

GQ03707_I04/*SAL2*		
Degrees of freedom	1	8
Sum of squares	0.201	29.203
F-value	0.554	10.073
p-value	0.462	<0.001
GQ03232_K15/*SAL7*		
Degrees of freedom	1	8
Sum of squares	1.311	11.612
F-value	1.494	1.654
p-value	0.231	0.149
GQ03806_I20/*SAL5*		
Degrees of freedom	1	8
Sum of squares	0.703	24.861
F-value	1.457	6.442
p-value	0.236	<0.001
GQ04010_J13/*SAL6*		
Degrees of freedom	1	8
Sum of squares	0.428	31.280
F-value	1.475	13.466
p-value	0.2334	<0.001
GQ03118_*H14*		
Degrees of freedom	1	8
Sum of squares	1.417	20.982
F-value	2.438	4.512
p-value	0.128	<0.001

Chapter 2 Supplementary Data

Figure S2.1 Maximum parsimony tree constructed from nucleotide partition.

Figure S2.2 Weighted maximum parsimony tree. MP tree constructed from nucleotide partition in PAUP*.

Figure S2.3 Maximum parsimony tree from amino acid partition. MP tree constructed from amino acid partition in PAUP*.

Figure S2.4 Maximum likelihood tree from amino acid partition.

Figure S2.5 Maximum likelihood tree without bootstrap.

Figure S2.6 AGL15 maximum likelihood constraint.

Figure S2.7 ANR1 maximum likelihood constraint.

Figure S2.8 FLC maximum likelihood constraint.

Figure S2.9 SEP maximum likelihood constraint.

Figure S2.10 SHP maximum likelihood constraint.

Figure S2.11 SOC1 maximum likelihood constraint.

Table S2.1 List of nucleotide and amino acid sequences of 88 MIKC sequences from 14 different species used for phylogenetic trees.

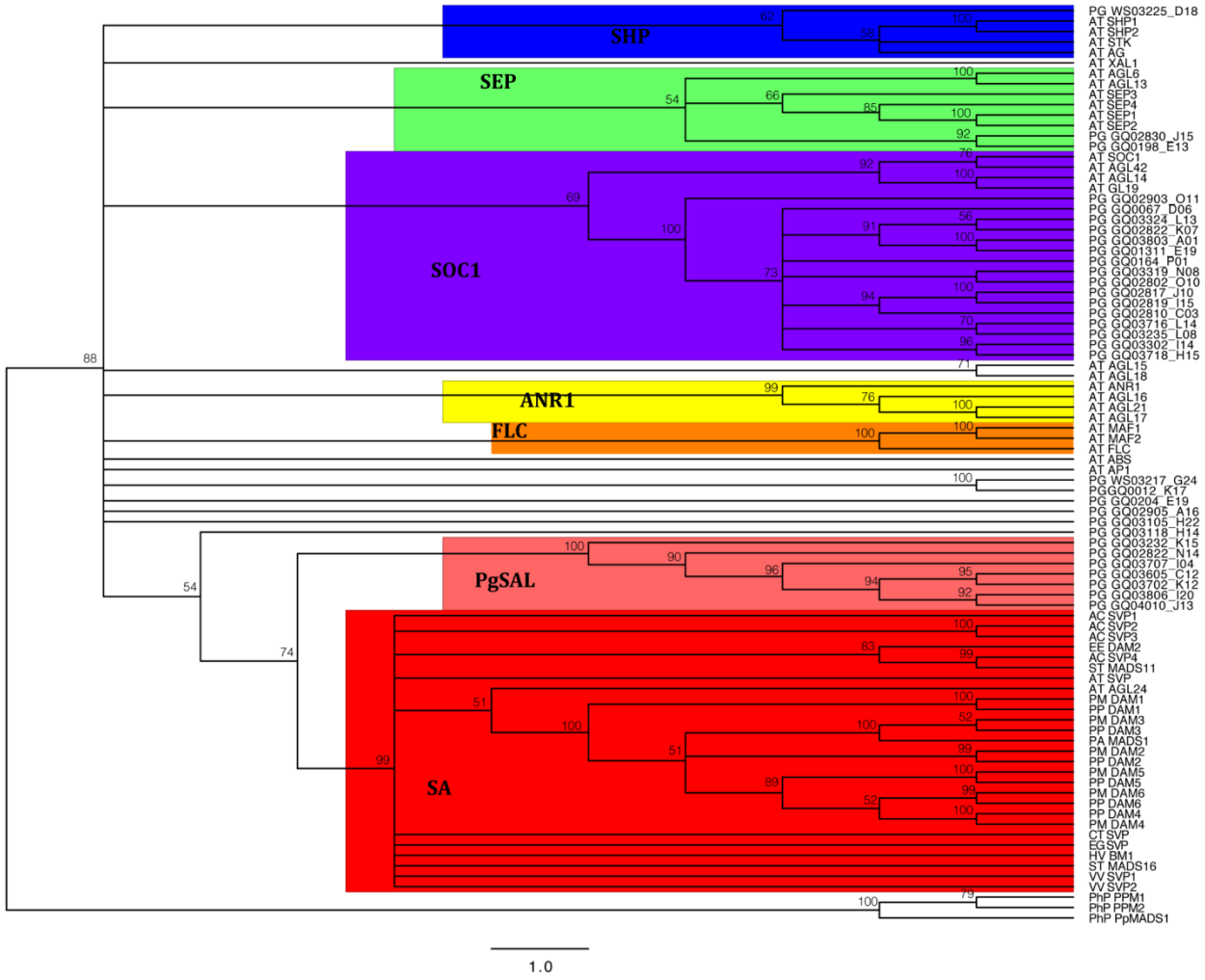


Figure S2.1 Maximum parsimony tree constructed from nucleotide partition. Branches with less than 50% BS support have been collapsed. Values above nodes represent BS values. Values above nodes represent bootstrap values. Clade names, based on gene function, are indicated in boxes. The following abbreviations accompanying gene names refers to species of origin: AC = *Actinidia chinensis*, AT = *Arabidopsis thaliana*, CT = *Citrus trifoliata*, EG = *Eucalyptus grandis,* EE = *Euphorbia esula,* HV = *Hordeum vulgare*, PhP = *Physcomitrella patens*, PG = *Picea glauca*, PA = *Prunus avium,* PM = *Prunus mume,* PP = *Prunus persica,* ST = *Solanum tuberosum*, VV = *Vitis vinifera*.

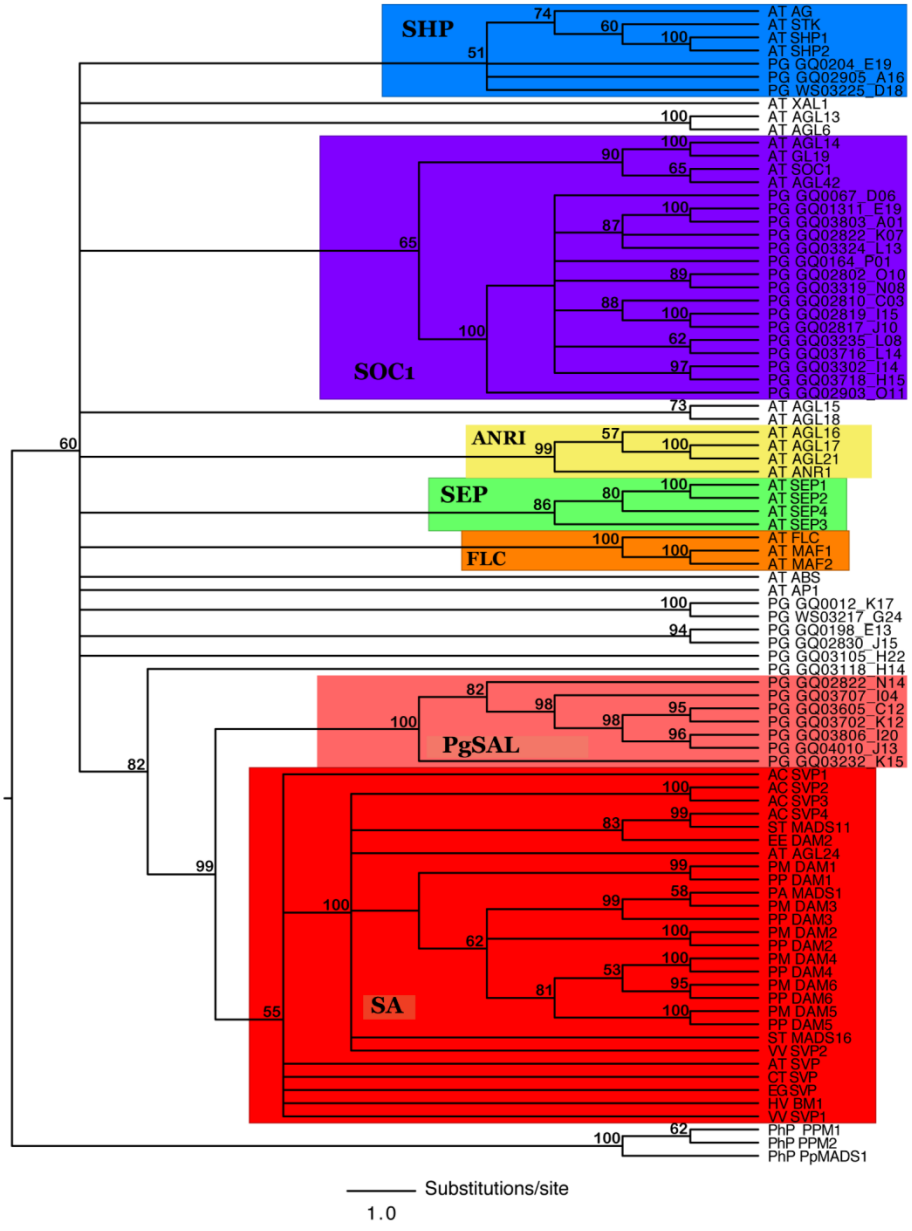


Figure S2.2 Weighted maximum parsimony tree. MP tree constructed from nucleotide partition in PAUP*. Tree search was conducted with 300 random addition replicates, TBR, 50 trees held in the construction of the initial starting tree, 1000 BS, 1000 nchuck with a chuckscore of greater than or equal to 1, majority rule 50% consensus tree. Weighting imposed across domains: "N" = 0.5, "M" = 3, "I" = 2, "K" = 3, "C" = 1. Branches with less than 50% BS support have been collapsed. Values above nodes represent bootstrap values. Clade names, based on gene

function, are indicated in boxes. The following abbreviations accompanying gene names refers to species of origin: AC = *Actinidia chinensis*, AT = *Arabidopsis thaliana*, CT = *Citrus trifoliata*, EG = *Eucalyptus grandis,* EE = *Euphorbia esula,* HV = *Hordeum vulgare*, PhP = *Physcomitrella patens*, PG = *Picea glauca*, PA = *Prunus avium,* PM = *Prunus mume,* PP = *Prunus persica,* ST = *Solanum tuberosum*, VV = *Vitis vinifera*.

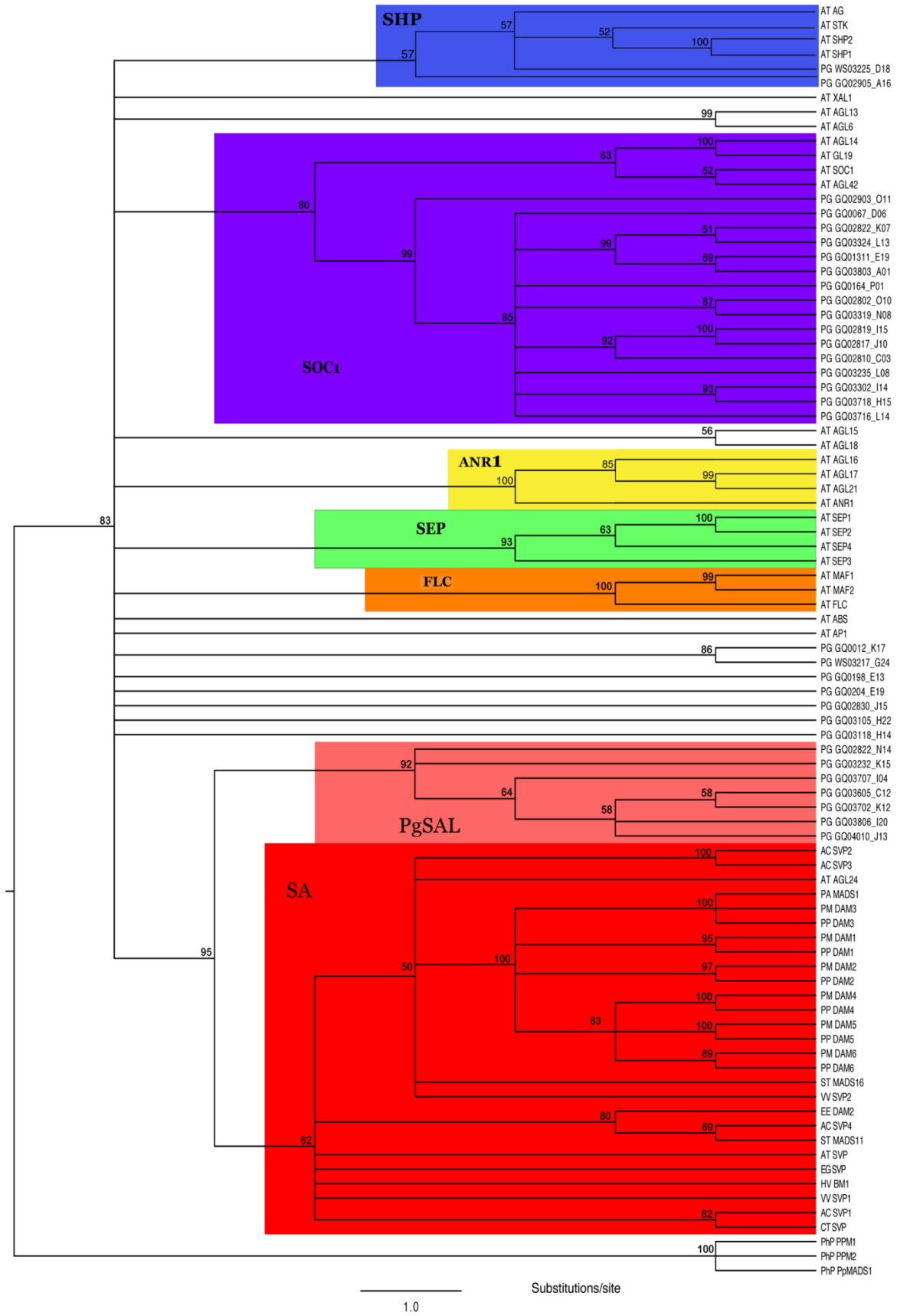

SHP
SOC1
ANR1
SEP
FLC
PgSAL
SA
AT AG
AT STK
AT SHP2
AT SHP1
PG WS03225_D18
PG GQ02905_A16
AT XAL1
AT AGL13
AT AGL6
AT AGL14
AT GL19
AT SOC1
AT AGL42
PG GQ02903_O11
PG GQ0067_D06
PG GQ02822_K07
PG GQ03324_L13
PG GQ01311_E19
PG GQ03803_A01
PG GQ0164_P01
PG GQ02802_O10
PG GQ03319_N08
PG GQ02819_I15
PG GQ02817_J10
PG GQ02810_C03
PG GQ03235_L08
PG GQ03302_I14
PG GQ03718_H15
PG GQ03716_L14
AT AGL15
AT AGL18
AT AGL16
AT AGL17
AT AGL21
AT ANR1
AT SEP1
AT SEP2
AT SEP4
AT SEP3
AT MAF1
AT MAF2
AT FLC
AT ABS
AT AP1
PG GQ0012_K17
PG WS03217_G24
PG GQ0198_E13
PG GQ0204_E19
PG GQ02830_J15
PG GQ03105_H22
PG GQ03118_H14
PG GQ02822_N14
PG GQ03232_K15
PG GQ03707_I04
PG GQ03605_C12
PG GQ03702_K12
PG GQ03806_I20
PG GQ04010_J13
AC SVP2
AC SVP3
AT AGL24
PA MADS1
PM DAM3
PP DAM3
PM DAM1
PP DAM1
PM DAM2
PP DAM2
PM DAM4
PP DAM4
PM DAM5
PP DAM5
PM DAM6
PP DAM6
ST MADS16
VV SVP2
EE DAM2
AC SVP4
ST MADS11
AT SVP
EG SVP
HV BM1
VV SVP1
AC SVP1
CT SVP
PhP PPM1
PhP PPM2
PhP PpMADS1
Substitutions/site
1.0

Figure S2.3 Maximum parsimony tree from amino acid partition. MP tree constructed from amino acid partition in PAUP*. Tree search was conducted with 300 random addition replicates, TBR, 50 trees held in the construction of the initial starting tree, 1000 bootstrap, 1000 nchuck with a chuckscore of greater than or equal to 1, majority rule 50% consensus tree. Branches with less than 50% BS support have been collapsed. Values above nodes represent bootstrap values. Clade names, based on gene function, are indicated in boxes. The following abbreviations accompanying gene names refers to species of origin: AC = *Actinidia chinensis*, AT = *Arabidopsis thaliana*, CT = *Citrus trifoliata*, EG = *Eucalyptus grandis,* EE = *Euphorbia esula,* HV = *Hordeum vulgare*, PhP = *Physcomitrella patens*, PG = *Picea glauca*, PA = *Prunus avium,* PM = *Prunus mume,* PP = *Prunus persica,* ST = *Solanum tuberosum*, VV = *Vitis vinifera*.

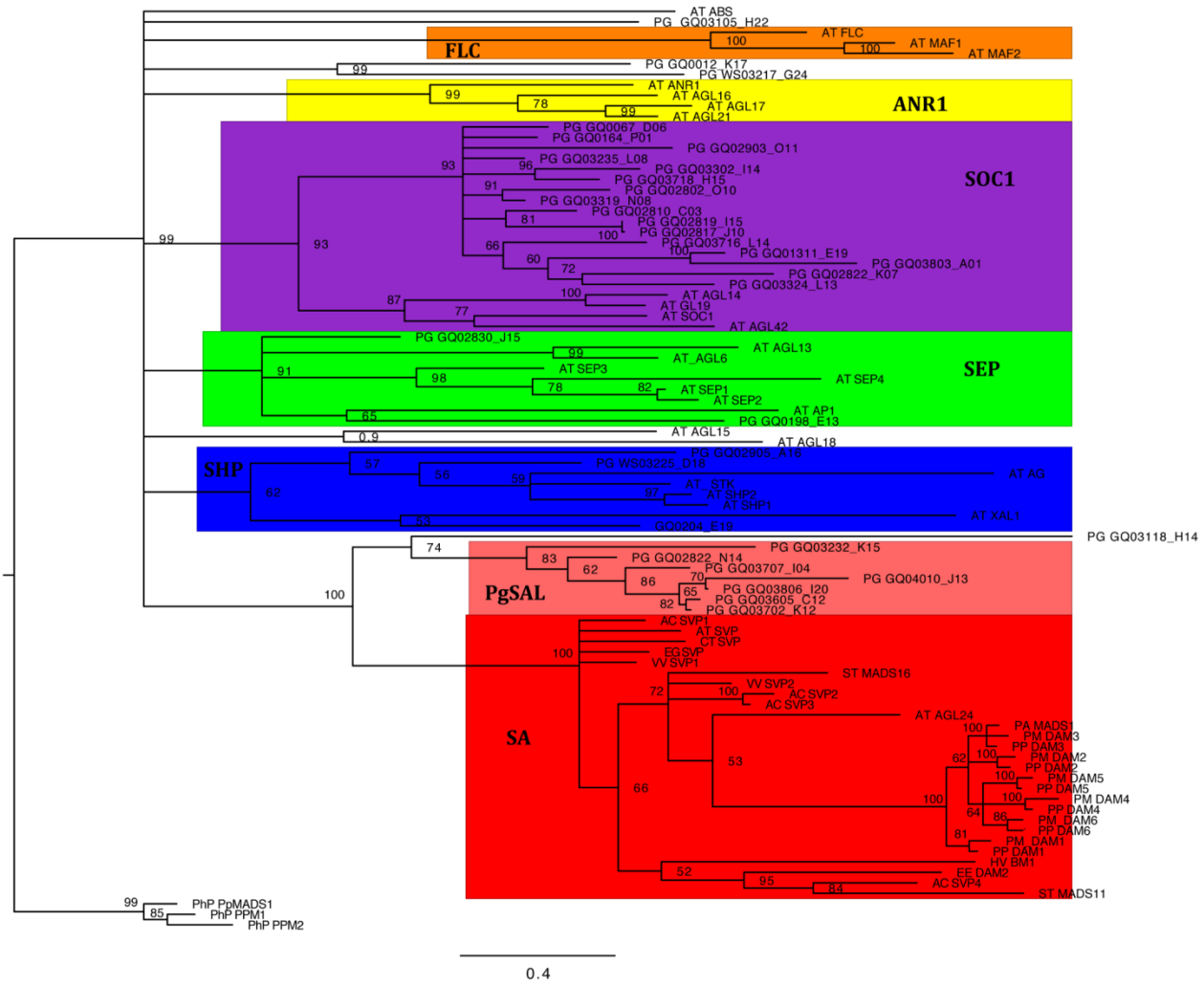


Figure S2.4 Maximum likelihood tree from amino acid partition. ML tree constructed from nucleotide data with GARLI 2.0 under the JTT+I+G model. Tree search was conducted with estimated state frequencies, proportion of invariant sites was estimated, 4 rate categories, 1000 times BS. Branches with less than 50% BS support have been collapsed. Values above nodes represent bootstrap values. Clade names, based on gene function, are indicated in boxes. The following abbreviations accompanying gene names refers to species of origin: AC = *Actinidia chinensis*, AT = *Arabidopsis thaliana*, CT = *Citrus trifoliata*, EG = *Eucalyptus grandis,* EE = *Euphorbia esula,* HV = *Hordeum vulgare*, PhP = *Physcomitrella patens*, PG = *Picea glauca*, PA

= *Prunus avium,* PM = *Prunus mume,* PP = *Prunus persica,* ST = *Solanum tuberosum*, VV = *Vitis vinifera*.

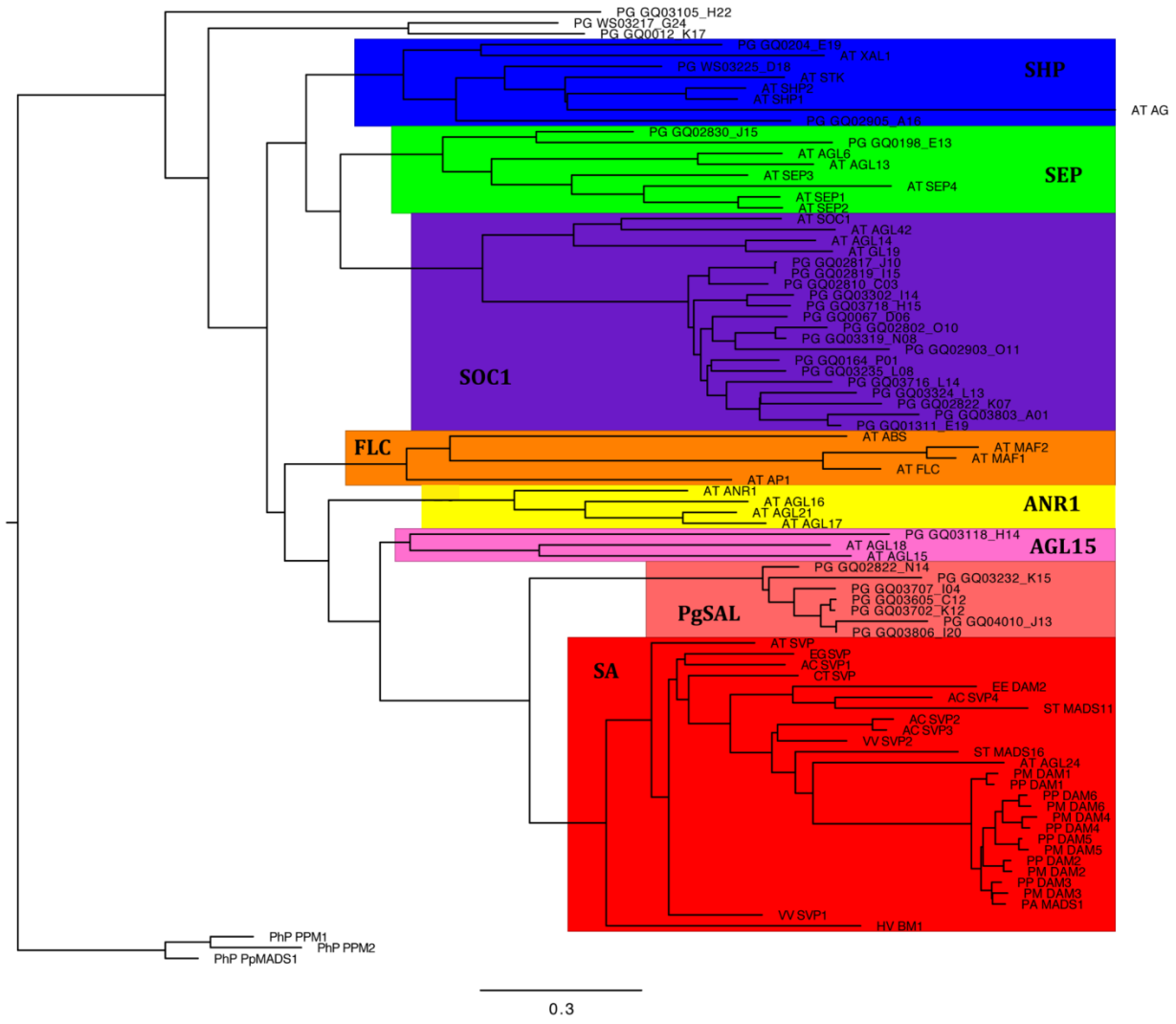


Figure S2.5 Maximum likelihood tree without bootstrap. The best ML tree constructed from nucleotide data (-ln = 42000.21361) with GARLI 2.0. Tree search was conducted with GTR+I+Γ substitution model, estimated state frequencies, proportion of invariant sites was estimated, 4 rate categories. Clade names, based on gene function, are indicated in boxes. The following abbreviations accompanying gene names refers to species of origin: AC = *Actinidia chinensis*, AT = *Arabidopsis thaliana*, CT = *Citrus trifoliata*, EG = *Eucalyptus grandis,* EE = *Euphorbia esula,* HV = *Hordeum vulgare*, PhP = *Physcomitrella patens*, PG = *Picea glauca*, PA

= *Prunus avium,* PM = *Prunus mume,* PP = *Prunus persica,* ST = *Solanum tuberosum*, VV = *Vitis vinifera*.

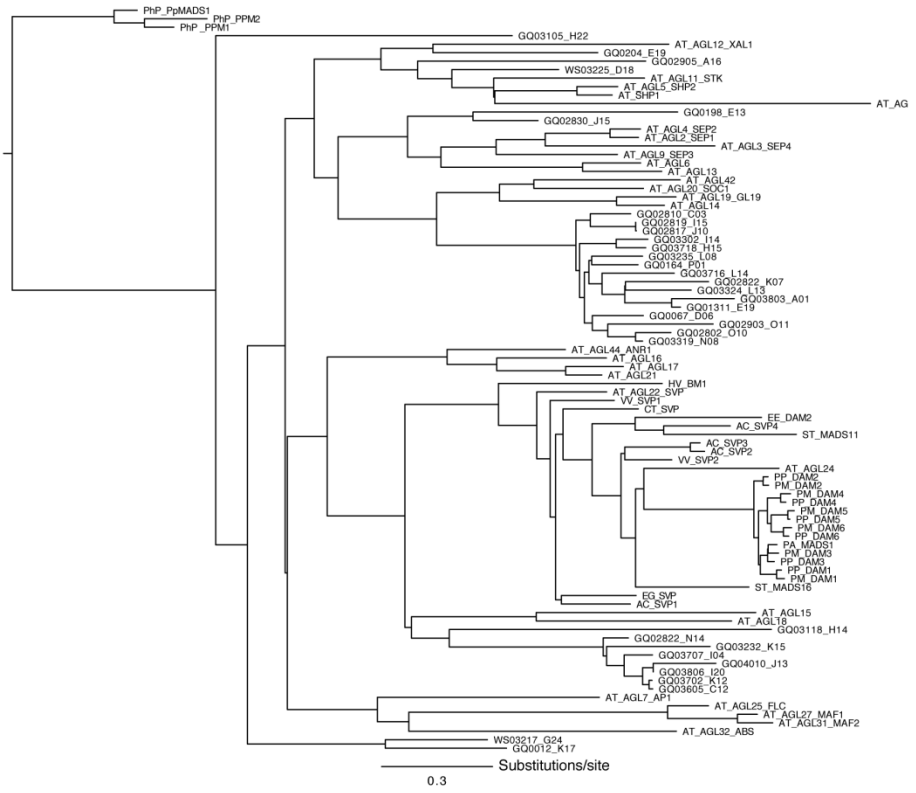


Figure S2.6 *AGAMOUS-LIKE 15* (*AGL15*) maximum likelihood constraint. The best ML tree constructed from nucleotide data (-ln = 42086.43568) with GARLI 2.0. Spruce *SAL* clade constrained with the *AGL15* clade. Tree search was conducted with GTR+I+Γ substitution model, proportion of invariant sites was estimated, 4 rate categories.

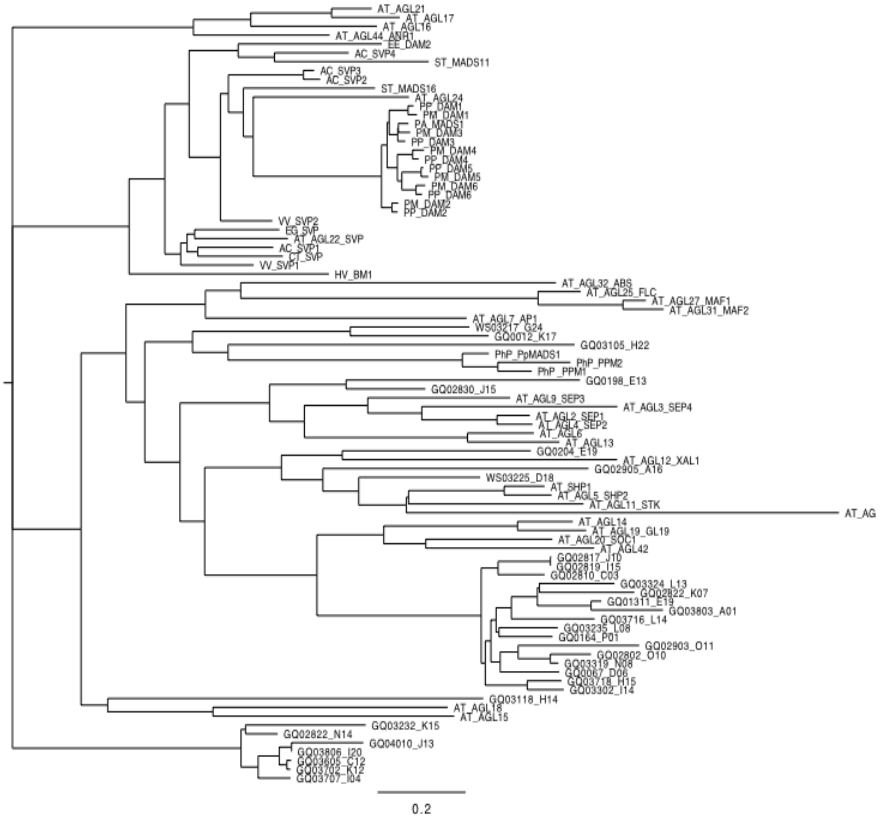


Figure S2.7 *ARABIDOPSIS NITRATE REGULATED 1* (ANR1) maximum likelihood constraint. The best ML tree constructed from nucleotide data (-ln = -42024.59) with GARLI 2.0. Spruce *SAL* clade constrained with the *ANR1* clade. Tree search was conducted with GTR+I+Γ substitution model, proportion of invariant sites was estimated, 4 rate categories.

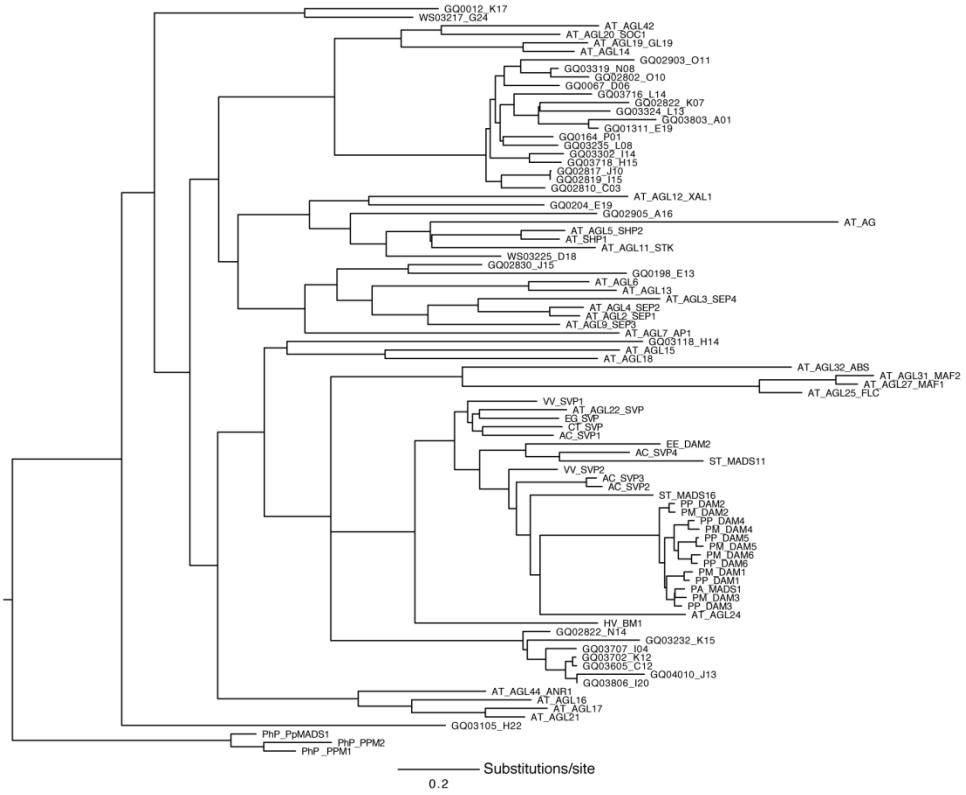


Figure S2.8 *FLOWERING LOCUS C* (*FLC*) maximum likelihood constraint. The best ML tree constructed from nucleotide data (-ln = 42228.67448) with GARLI 2.0. Spruce *SAL* clade constrained with the *FLC* clade. Tree search was conducted with GTR+I+Γ substitution model, estimated state frequencies, proportion of invariant sites was estimated, 4 rate categories.

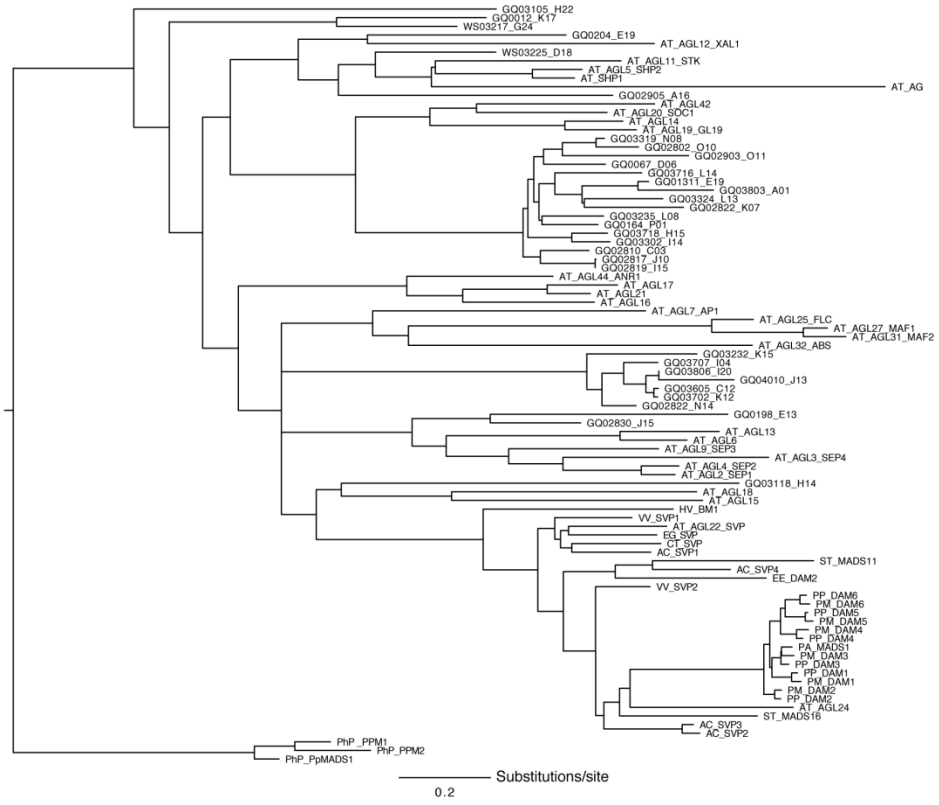


Figure S2.9 *SEPELLATA* (*SEP*) maximum likelihood constraint. The best ML tree constructed from nucleotide data (-ln = 42619.76438) with GARLI 2.0. Spruce *SAL* clade constrained with the *SEP* clade. Tree search was conducted with GTR+I+Γ substitution model, estimated state frequencies, proportion of invariant sites was estimated, 4 rate categories.

Figure S2.10 *SHATTERPROOF* (*SHP*) maximum likelihood constraint. The best ML tree constructed from nucleotide data (-ln = 42499.24251) with GARLI 2.0. Spruce *SAL* clade constrained with the *SHP* clade. Tree search was conducted with GTR+I+Γ substitution model, estimated state frequencies, proportion of invariant sites was estimated, 4 rate categories.

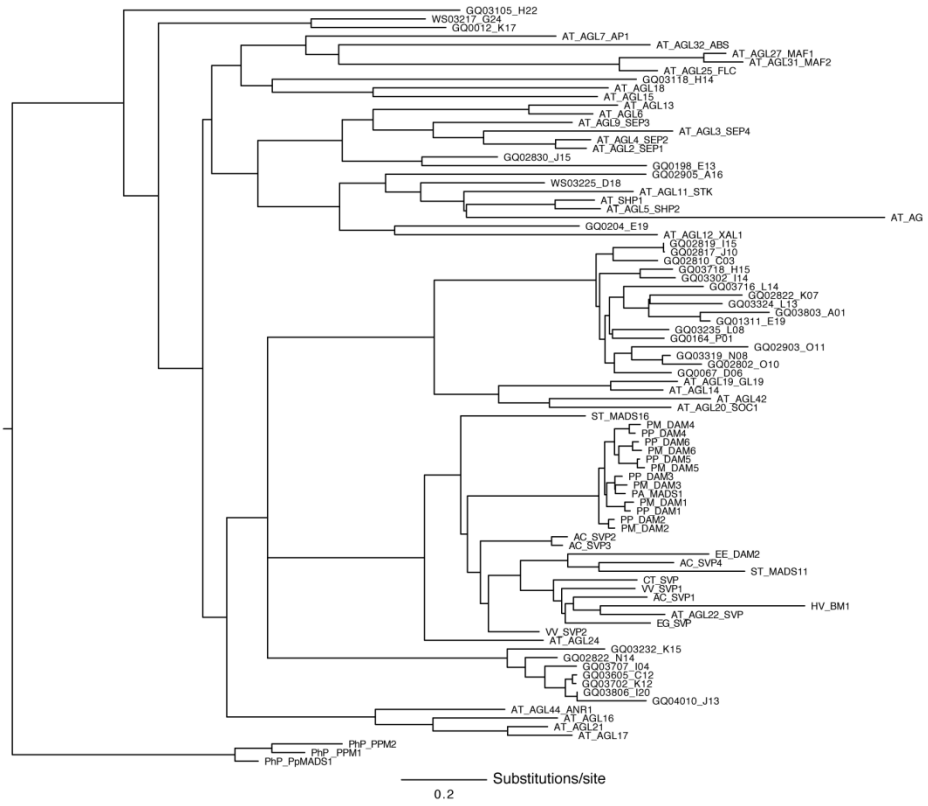


Figure S2.11 *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) maximum likelihood constraint. The best ML tree constructed from nucleotide data (-ln = 42143.22345) with GARLI 2.0. Spruce *SAL* clade constrained with the *SOC1* clade. Tree search was conducted with GTR+I+Γ substitution model, estimated state frequencies, proportion of invariant sites was estimated, 4 rate categories.

Table S2.1 List of nucleotide and amino acid sequences of 88 MIKC sequences from 14 different species used for phylogenetic trees. GenBank and GenPept accession numbers obtained from NCBI. Locus identity included in parenthese for *Arabidopsis* sequences. Authorities found from tropicos.org. Lineages are listed as A = angiosperm, B = bryophyte, C = conifer.

Species	Lineage	Initials	Genes	GenBank accession no.	GenPept accession no.
Actinidia chinensis Planch.	A	AC	SVP1	JF838216.1	AFA37967.1
Actinidia chinensis Planch.	A	AC	SVP2	JF838217.1	AFA37968.1
Actinidia chinensis Planch.	A	AC	SVP3	JF838218.1	AFA37969.1
Actinidia chinensis Planch.	A	AC	SVP4	JF838219.1	AFA37970.1
Arabidopsis thaliana (L.) Heynh.		AT	SEP3 (AT1G24260)	NM_102272.3	AEE30503.1
Arabidopsis thaliana (L.) Heynh.	A	AT	AP1 (AT1G69120)	NM_105581.2	AEE34887.1
Arabidopsis thaliana (L.) Heynh.	A	AT	XAL1 (AT1G71692)	NM_105825.2	AEE35216.1
Arabidopsis thaliana (L.) Heynh.	A	AT	MAF1 (AT1G77090)	NM_180648.3	AEE35931.1
Arabidopsis thaliana (L.) Heynh.	A	AT	SEP4 (AT2G03710)	NM_126418.2	AEC05738.1
Arabidopsis thaliana (L.) Heynh.	A	AT	ANR1 (AT1G08090)	NM_126990.3	AEC06290.1
Arabidopsis thaliana (L.) Heynh.	A	AT	SVP (AT2G22540)	NM_127820.3	AEC07320.1
Arabidopsis thaliana (L.) Heynh.	A	AT	AGL17 (AT2G22630)	NM_127828.2	AEC07331.1

Arabidopsis thaliana (L.) Heynh.	A	AT	SHP2 (AT2G42830)	NM_180046.2	AEC10175.1
Arabidopsis thaliana (L.) Heynh.	A	AT	AGL6 (AT2G45650)	NM_130127.1	AEC10582.1
Arabidopsis thaliana (L.) Heynh.	A	AT	SOC1 (AT2G45660)	NM_130128.3	AEC10583.1
Arabidopsis thaliana (L.) Heynh.	A	AT	SEP2 (AT3G02310)	NM_111098.3	AEE73791.1
Arabidopsis thaliana (L.) Heynh.	A.	AT	AGL16 (AT3G57230)	NM_115583.5	AEE79629.1
Arabidopsis thaliana (L.) Heynh.	A	AT	AGL18 (AT3G57390)	NM_115599.3	AEE79650.1
Arabidopsis thaliana (L.) Heynh.	A	AT	SHP1 (AT3G58780)	NM_001203201.1	AEE79831.1
Arabidopsis thaliana (L.) Heynh.	A	AT	AGL13 (AT3G61120)	NM_115976.1	AEE80158.1
Arabidopsis thaliana (L.) Heynh.	A	AT	STK (AT4G09960)	NM_001084895.1	AEE82817.1
Arabidopsis thaliana (L.) Heynh.	A	AT	AGL14 (AT4G11880)	NM_117258.5	AEE83062.1
Arabidopsis thaliana (L.) Heynh.	A	AT	AG (AT4G18960)	NM_118013.2	AEE841121.1
Arabidopsis thaliana (L.) Heynh.	A	AT	GL19 (AT4G22950)	NM_118424.2	AEE84684.1
Arabidopsis thaliana (L.) Heynh.	A	AT	AGL24 (AT4G24540)	NM_118587.5	AEE84922.1
Arabidopsis thaliana (L.) Heynh.	A	AT	AGL21 (AT3G37940)	NM_119955.2	AEE86856.1
Arabidopsis thaliana (L.) Heynh.	A	AT	FLC (AT5G10140)	NM_121052.2	AED91498.1
Arabidopsis	A	AT	AGL15	NM_121382.3	AED91941.1

thaliana (L.) Heynh.			(AT5G13790)		
Arabidopsis thaliana (L.) Heynh.	A	AT	SEP1 (AT1G34360)	NM_001125758.1	AED92208.1
Arabidopsis thaliana (L.) Heynh.	A	AT	ABS (AT5G23260)	NM_203094.1	AED93144.1
Arabidopsis thaliana (L.) Heynh.	A	AT	AGL42 (AT5G62165)	NM_125610.3	AED97574.1
Arabidopsis thaliana (L.) Heynh.	A	AT	MAF2 (AT5G65050)	NM_001126026.1	AED97992.1
Citrus trifoliate L.	A	CT	SVP	FJ373210.1	ACJ09169.1
Eucalyptus grandis W. Hill	A	EG	SVP	AY263809.1	AAP33087.1
Euphorbia esula L.	A	EE	DAM2	EU339320.1	ABY60423.1
Hordeum vulgare L.	A	HV	BM1	AJ249142.1	CAB97350.1
Physcomitrella patens (Hedw.) Bruch & Schimp	B	PhP	PPM1	XM_001769810.1	AAG09136.2
Physcomitrella patens (Hedw.) Bruch & Schimp	B	PhP	PPM2	AF150933.1	EDQ72735.1
Physcomitrella patens (Hedw.) Bruch & Schimp	B	PhP	PPMADS1	XM_001779819.1	EDQ55286.1
Picea glauca (Moench) Voss	C	PG	GQ02822_K07	BT105450.1	-
Picea glauca (Moench) Voss	C	PG	GQ03806_I20	BT116779.1	-
Picea glauca (Moench) Voss	C	PG	GQ0164_P01	BT102045.1	-
Picea glauca (Moench) Voss	C	PG	GQ03235_L08	BT111301.1	-
Picea glauca (Moench) Voss	C	PG	GQ03105_H22	BT107302.1	-
Picea glauca (Moench) Voss	C	PG	GQ0012_K17	BT100378.1	-
Picea glauca (Moench) Voss	C	PG	GQ0067_D06	BT101090.1	-
Picea glauca (Moench) Voss	C	PG	GQ01311_E19	EX309542.1	-

Picea glauca (Moench) Voss	C	PG	GQ0198_E13	BT102624.1	-
Picea glauca (Moench) Voss	C	PG	GQ0204_E19	BT102975.1	-
Picea glauca (Moench) Voss	C	PG	GQ02802_O10	BT103840.1	-
Picea glauca (Moench) Voss	C	PG	GQ02810_C03	BT104415.1	-
Picea glauca (Moench) Voss	C	PG	GQ02819_I15	BT105191.1	-
Picea glauca (Moench) Voss	C	PG	GQ02830_J15	BT105966.1	-
Picea glauca (Moench) Voss	C	PG	GQ02903_O11	BT106143.1	-
Picea glauca (Moench) Voss	C	PG	GQ02905_A16	BT106210.1	-
Picea glauca (Moench) Voss	C	PG	GQ03118_H14	BT108213.1	-
Picea glauca (Moench) Voss	C	PG	GQ03232_K15	BT111101.1	-
Picea glauca (Moench) Voss	C	PG	GQ03302_I14	BT111713.1	-
Picea glauca (Moench) Voss	C	PG	GQ03319_N08	BT112706.1	-
Picea glauca (Moench) Voss	C	PG	GQ03324_L13	BT113029.1	-
Picea glauca (Moench) Voss	C	PG	GQ03605_C12	BT114920.1	-
Picea glauca (Moench) Voss	C	PG	GQ03707_I04	BT115854.1	-
Picea glauca (Moench) Voss	C	PG	GQ03716_L14	BT116336.1	-
Picea glauca (Moench) Voss	C	PG	GQ03718_H15	BT116425.1	-
Picea glauca (Moench) Voss	C	PG	WS03217_G24	DR550143.1	-
Picea glauca (Moench) Voss	C	PG	WS03225_D18	DR553148.1	-
Picea glauca (Moench) Voss	C	PG	GQ04010_J13	EX439444.1	-
Picea glauca (Moench) Voss	C	PG	GQ02822_N14	BT105463.1	-
Picea glauca (Moench) Voss	C	PG	GQ02817_J10	BT105004.1	-
Picea glauca (Moench) Voss	C	PG	GQ03803_A01	BT116601.1	-

Picea glauca (Moench) Voss	C	PG	GQ03702_K12	BT115613.1	-
Prunus avium (L.) L.	A	PA	MADS1	EU196362.1	ABW82562.1
Prunus mume (Siebold) Siebold & Zucc.	A	PM	DAM1	AB576350.1	BAK78921.1
Prunus mume (Siebold) Siebold & Zucc.	A	PM	DAM2	AB576351.1	BAK78922.1
Prunus mume (Siebold) Siebold & Zucc.	A	PM	DAM3	AB576352.1	BAK78923.1
Prunus mume (Siebold) Siebold & Zucc.	A	PM	DAM4	AB576353.1	BAK78924.1
Prunus mume (Siebold) Siebold & Zucc.	A	PM	DAM5	AB576349.1	BAK78920.1
Prunus mume (Siebold) Siebold & Zucc.	A	PM	DAM6	AB437345.1	BAH22477.1
Prunus persica (L.) Batsch	A	PP	DAM1	DQ863253.2	ABJ96361.2
Prunus persica (L.) Batsch	A	PP	DAM2	DQ863257.1	ABJ96370.1
Prunus persica (L.) Batsch	A	PP	DAM3	DQ863256.1	ABJ96370.1
Prunus persica (L.) Batsch	A	PP	DAM4	DQ863257.1	ABJ96365.1
Prunus persica (L.) Batsch	A	PP	DAM5	DQ863251.1	ABJ96366.1
Prunus persica (L.) Batsch	A	PP	DAM6	DG863252.1	ABJ96367.1
Solanum tuberosum L.	A	ST	MADS11	AF008652.1	AAB94006.1
Solanum tuberosum L.	A	ST	MADS16	AF008651.1	AAV65504.1
Vitis vinifera L.	A	VV	SVP1	JQ387569.1	AFC96914.1
Vitis vinifera L.	A	VV	SVP2	XM_002285651.2	XP_002285687.1

3.0 Chapter 3: *Picea glauca SHORT VEGETATIVE PHASE/AGAMOUS-LIKE 24*-like 1 regulation may have evolved from a common angiosperm pathway, while *SHORT VEGETATIVE PHASE/AGAMOUS-LIKE 24-like 5* may be regulated by novel pathways

3.1 Introduction

Perennial trees of the boreal forest undergo seasonal changes in growth and development to protect against the harsh environmental conditions of winter. In preparation for the phase transition from active growth to dormancy, white spruce (*Picea glauca* (Moench) Voss*, Pg*), trees form a terminal bud and meristematic growth ceases (Rohde and Bhalerao 2007). The transition from active growth to dormancy is regulated by environmental cues such as photoperiod (Garner and Allard 1923, Nitsch 1957, Garris *et al.* 2009) and temperature (Kalcsits *et al.* 2009, Tanino *et al.* 2010, Rohde *et al.* 2011). Photoperiod has a more pronounced influence than temperature in the regulation of growth cessation and terminal bud formation in trees displaying indeterminate growth such as poplar (Nitsch 1957, Heide 1974). Low temperature delays the rate of bud set in white spruce trees in both long day (LD) and short day (SD) treatments (Hamilton *et al.* 2016), however the combinatorial effect between temperature and photoperiod is species dependent (Heide and Prestrud 2005, Junttila 1980). White spruce trees are able to form terminal buds in the absence of both SD and low temperatures, although both of these environmental cues affect the rate of bud formation (El Kayal *et al.* 2011, Hamilton *et al.* 2016). In white spruce, a determinate species, SD in combination with warm temperatures accelerates terminal bud formation and growth cessation in above ground tissue (Hamilton *et al.* 2016). Terminal bud formation and growth cessation proceed more slowly in LD and/or low

temperatures conditions, with the combination of SD and low temperatures displaying the slowest rate of progression (Hamilton *et al.* 2016).

In addition to environmental cues, a number of hormones are implicated in regulating the events that make up the activity to dormancy transition in angiosperms (Eriksson and Moritz 2002, Ruonala *et al.* 2006, Ruttink *et al.* 2007, Baba *et al.* 2011). Abscisic acid (ABA) content increases in response to SDs in poplar (Rohde *et al.* 2002) to cease growth, and may be involved in dormancy establishment in hybrid aspen (Tylewicz *et al.* 2015). During white spruce bud formation ABA content is low during bud development, which suggests ABA may have a role in growth cessation in conifers (El Kayal *et al.* 2011). ABA is also involved in the abiotic stress response to salinity, drought and cold. Low temperatures may contribute to ABA accumulation, growth cessation (Welling and Palva 2006) and the circadian clock during dormancy (Ramos *et al.* 2005). Decreased accumulation of gibberellins (GA) contributes to cessation of growth leading up to dormancy establishment in hybrid aspen (*Populus tremula* L. *x Populus tremuloides* Michx.; Eriksson *et al.* 2000, Eriksson and Moritz 2002). Timing of dormancy induction is shown to be linked to ethylene in birch and poplar (Ruonala *et al.* 2006, Ruttink *et al.* 2007). In birch trees, ethylene is not a requirement for the transition to endodormancy, but it can affect the timing of transition (Ruonala *et al.* 2006). Furthermore, ethylene may play a role in mediating correct developmental processes at the shoot apical meristem (SAM) since ethylene insensitive birch trees displayed altered bud structures (Ruonala *et al.* 2006). Auxin participates in a wide variety of plant development pathways, and is known to inhibit lateral bud outgrowth apical dominance in trees. Auxin sensitivity is also involved in halting cell division of cambial cells of hybrid aspen, and thereby participates in growth cessation and dormancy (Resman *et al.* 2010, Baba *et al.* 2011). Auxin levels in white sprue apical buds were found to decrease in

response to SD (El Kayal *et al.* 2011). A clear role for cytokinin (CK) in seasonal growth regulation have yet to be well established, however CKs are well established in the regulation of cell division and in stimulating the outgrowth of angiosperm lateral buds (Cline and Dong-Il 2002, Ferguson and Beveridge 2009). Increased CK levels were also found to correlate with Norway spruce bud size (Chen *et al.* 1996).

Despite the evolutionary divergence between conifers and angiosperms, angiosperms serve as a reasonable model to base our assumptions of developmental processes involved in white spruce dormancy. There is an accumulating body of evidence, mainly from angiosperms, that there is a conserved network involved in regulating the transition from vegetative to reproductive growth and the transition from active growth to dormancy (Böhlenius *et al.* 2006, Gyllenstrand *et al.* 2007, Mohamed *et al.* 2010, Karlgren *et al.* 2011). For example, genes orthologous to flowering time regulators, such as *Populus trichocarpa* Torr. & A. Gray *CONSTANS 2* and *FLOWERING LOCUS T 1* and *Populus tremuloides Michx. CENTRORADIALIS 1* and *CENTRORADIALIS 2,* have been shown to regulate bud formation in forest trees (Böhlenius *et al.* 2006, Hsu *et al.* 2006, Mohamed *et al.* 2010). Building on the findings of Böhlenius *et al.* (2006), *PaFTL2*, a gene sister to *FLOWERING LOCUS T* and *TERMINAL FLOWER 1* (*TFL1*) in *Arabidopsis thaliana (*L.) Heynh. (*At*), has been implicated in bud formation and growth cessation of Norway spruce (*Picea abies* (L.) H. Karst.; Gyllenstrand *et al.* 2007, Karlgren *et al.* 2011, Klintenäs *et al.* 2012). Expression of *PaFTL2* is upregulated after treatment with reduced day length, which contrasts expression of *Arabidopsis FT*, which is downregulated in response to SD (Gyllenstrand *et al.* 2007, Suárez-López *et al.* 2001). This difference in pattern suggests the role of *PaFTL2* may be suppression of growth, a function more similar to *Arabidopsis TFL1* (Karlgren *et al.* 2011).

Several studies have focused on the functional roles of *MADS-box* genes in bud formation and dormancy acquisition (Mazzitelli *et al.* 2007, Jiménez *et al.* 2009, Horvath *et al.* 2010, Sasaki *et al.* 2011, Wu *et al.* 2011). An important paper on bud development and dormancy induction investigated a natural occurring knock-out mutant of *Prunus persica* (L.) Batsch *DORMANCY ASSOCIATED MADS-BOX* (*DAM*) genes, an *SVP-like* gene, which resulted in the *EVERGROWING* mutant (Bielenberg *et al.* 2004, Jiménez *et al.* 2009). The *EVERGROWING* mutant does not produce terminals buds or enter a dormant state (Bielenberg *et al.* 2004, Jiménez *et al.* 2009), suggesting genes with functions of similar importance may exist in other species. In contrast, expression of *Arabidopsis SVP* is unaffected by changes in temperature and photoperiod (Hartmann *et al*. 2000). However, there is evidence that environmental conditions, such as warmer temperatures, result in the degradation of the SVP protein, thereby reducing function and impacting flowering time (Lee *et al.* 2007, Lee *et al.* 2013, Lee *et al.* 2014, Fernández *et al.* 2016). *AtSVP* is stable at low temperatures and acts to inhibit the transition to flowering by dimerizing with *FLOWERING LOCUS M* (*FLM*) (Lee *et al.* 2013). *At*SVP additionally forms a repressive complex with *FLOWERING LOCUS C* (*FLC*) to prevent the transcription of key flowering genes *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) and *FLOWERING LOCUS T* (*FT*) (Li *et al.* 2008, Jang *et al.* 2009, Searle *et al.* 2006). Expression of *AtFLC* is mediated by the vernalization pathway through activation by FRIGIDA (FRI) complexing with *FLOWERING C EXPRESSOR* (*FLX*) genes (Ding *et al.* 2013). *At*SOC1 an *At*SVP act to mutually inhibit one another's transcription in the transition to flowering (Li *et al.* 2008, Immink *et al.* 2012). There is also evidence that suggests components of the circadian clock, *CIRCADIAN CLOCK-ASSOCIATED1* (*CCA1*) and *LATE ELONGATED*

HYPOCOTYL (*LHY*), may downregulate *SVP* expression in the morning to alleviate SVP repression of *FT* (Fujiwara *et al.* 2008).

As demonstrated in Chapter 2, we identified a set of white spruce genes that are sister to *Arabidopsis SVP*, the closely related *AGL24* and the *Prunus persica DAM*. Based on the homology of *PgSVP-like* genes with *Arabidopsis SVP* and *AGL24*, we named these white spruce genes *SHORT VEGETATIVE PHASE/AGAMOUS-LIKE 24-like* (*PgSAL*). Transcript profiling data support that *PgSAL1*, *PgSAL2*, *PgSAL3*, *PgSAL4*, and *PgSAL5* may be involved in the early staged of bud development. To test the hypothesis that *PgSAL* genes are involved in bud formation and potentially other events that take place during the transition from active growth to dormancy, we investigated the upstream pathways regulating the activation of *PgSAL1* and *PgSAL5*. *PgSAL1* and *PgSAL5* were chosen to explore further based on their expression profiles and our ability to clone large regions of these promoters for further analyses. We wished to investigate the functional roles and upstream regulators of *PgSAL1* and *PgSAL5* related to bud development in white spruce, since few studies have looked into this topic in conifers.

To perform these analyses, we cloned the putative promoters for these two genes, and used available sequence data from draft assemblies of the white spruce genome (Birol *et al.* 2013, Warren *et al.* 2015) to further characterize the cloned sequences. Based on these analyses, they were designated as *PgSAL1* and putative *PgSAL5* promoters. We used these promoter sequences in two experimental approaches. In the first approach, we identified transcription factor binding sites (TFBS) in the *PgSAL1* and putative *PgSAL5* promoters with purpose of cataloguing the breadth of the possible interactions involved. In the second approach, we identified transcription factors (TF) that bind the *PgSAL1* and putative *PgSAL5* promoters. Together, these two approaches allowed us to identify components of upstream regulatory

networks involved in regulating *PgSAL1* and *PgSAL5*, which in turn reveal new insight about the functional roles of these genes.

3.2 Materials and Methods

3.2.1 Plant Material

White spruce terminal bud mRNA from two experiments were pooled to construct the cDNA library used for yeast one-hybrid screening. Combined, these two experiments cover the development of terminal buds from initiation to dormancy. Prior to exposure to treatments trees were grown under LD conditions for active growth. In the short-term time course trees experienced three weeks of active growth, whereas trees used in extended time course had approximately eight weeks of active growth. The short-term time course used two-year-old white spruce seedlings were grown under SD (8 h days/16 h nights) at 20°C. Terminal buds were harvested at five time points (Day 0, 7, 14, 28, 70). Trees in the extended time course experiment were two-year-old white spruce trees obtained from the experiment conducted in Hamilton *et al.* (2016). Trees were grown under SD at 22°C day, and 16°C night temperatures. Terminal buds harvested at three later time points (Day 92, 106, 126) were used for cDNA library construction. All terminal buds from both experiments were immediately frozen in liquid nitrogen upon harvest and stored at -80 °C.

3.2.2 *PgSAL1* and putative *PgSAL5* promoter isolation

PgSAL1 (GQ03605_C12, BT114920.1) and putative *PgSAL5* (GQ03806_I20, BT116779.1) promoters were isolated using the GenomeWalker™ Universal Kit (Clontech, Mountain View CA, USA) from a single white spruce sample. White spruce gDNA was extracted from white spruce needles using a modified cetyltrimethylammonium bromide (CTAB)

protocol (Chang *et al*. 1993, Roe *et al.* 2010). We used the fragments obtained using GenomeWalker™, 426 bp for *PgSAL1* and 862 bp for putative *PgSAL5*, to identify the corresponding genomic scaffolds in v1.0 of the Norway spruce genome sequence (Nystedt *et al.* 2013), the most complete genome assembly at the time, using the BLASTn function in ConGenie (congenie.org, Sundell *et al.* 2015). Primers were designed against the upstream region of the Norway spruce sequences and used to obtain final promoter sequences from white spruce (Table 3.1). Complete promoter sequences are given in Appendix 1 (A1).

Primers for Genome Walker ™ cloning, and cloning using the Norway spruce gDNA as a guide, were generated using Integrated DNA Technologies (IDT) PrimerQuest Tool (idtdna.com/Primerquest/Home/Index), primers (Table 3.1). To adhere to the Genome Walker™ recommendations, all primers were designed to be 26-30 bp in length with a GC content between 40-58%, with a maximum of three G's and C's in the primer's 3' end to prevent self-annealing. Primers were also designed to limit self-dimers and hetero-dimers, and hairpin strength. Promoter fragments were cloned into the pGEM®-T Easy vector (Promega, Madison, USA) and sequenced with the T7 and SP6 universal primers.

3.2.3 Yeast One-Hybrid (Y1H) Assay

Terminal bud mRNA was pooled to generate one cDNA library. Total RNA was extracted from two to four white spruce terminal buds per time point using the small scale CTAB protocol described in Pavy *et al.* (2008). RNA quality and quantity was assessed with an Infinite® M200 NanoQuant (Tecan, Männerdorf, Switzerland) and gel electrophoresis. mRNA was isolated from total RNA using NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, Ipswich, MA, USA). A subset of four mRNA samples were run on an

Agilent 2100 Bioanalyzer (Aglient, Waldbronn, Germany) to ensure purification and sample integrity.

The terminal bud cDNA library was created with the CloneMiner^{TM}II cDNA Library Construction Kit (Life Technologies, Carlsbad, CA, USA) according to the manufacturer instructions. Approximately 390 to 1174 ng of mRNA was used from each of the eight time points to create a total of 5720 ng of pooled enriched mRNA from which to create the cDNA library. A sample of 2 µg control mRNA provided by the kit was used as a positive control for all steps of the cDNA library construction to ensure the procedure had been carried out correctly. White spruce mRNA was split into two reaction tubes so that 2860 ng of enriched mRNA was used as a starting material for the cDNA synthesis, adaptor ligation steps and column chromatography. For cDNA synthesis, we were concerned that the oligo dT primers provided in the kit would not perform as well as the anchored oligo $dT_{(18)}N$, and therefore performed one cDNA synthesis reaction with the oligo dT primers in the kit, and the other with our own anchored oligo $dT_{(18)}N$. A first priming step was carried out over 18.5 min, over which the temperature declined from 70°C to 45°C at approximately 1°C/45 sec. The protocol for first stand synthesis was as follows: 45°C for 20 sec, 50°C for 20 sec, 55°C for 20 sec, then immediate removal of tubes from the machine onto ice to prevent temperature from increasing past 16°C. Second strand synthesis protocol was as follows: 16°C for 2 hours, addition of 2 µl T4 DNA polymerase to create blunt ends, 16°C 5 min, add 10 µl 0.5 M EDTA (pH 8) to stop reaction.

Following cDNA synthesis, samples were purified using phenol:choloroform:isoamyl alcohol (in proportions of 25:24:1). Following column cleanup, yields were quantified on an Infinite® 200 NanoQuant (Tecan Group Ltd., Männedorf, Switzerland) to ensure product

recovery. Yield was very low (18.2 ng for the oligo dT reaction and 120 ng anchored oligo $dT_{(18)}N$, 196.56 ng for the control) so the two white spruce reactions were pooled for a total of 138.84 ng to use for the remainder of the protocol. Library quality was assessed by checking 30 randomly sampled plasmids, and analyzing cDNA fragments by agarose gel electrophoresis. The cDNA library had a 100% recombinants (i.e. 100% of cDNA fragments recombined into the destination vector) with cDNA fragment sizes ranging from about 650 bp to 2.25 kb, with an average cDNA fragment size of approximately 1.68 kb. According the manufacturer's instructions, a standard cDNA library should have a minimum of 95% recombinants and the average insert size should be greater than or equal to 1.5 kb. As a further assessment of library quality, a titer of the number of colony forming units (cfu) per mL was also performed with the *Escherichia coli* (*E. coli*) colonies transformed with the cDNA fragments (see calculation in Appendix 3, A3) and was above the minimum of 5×10^6 cfu/mL recommended by the manufacturer.

A subsample of the cDNA library culture was grown in 50 mL culture of LB media to an OD600 of approximately 1, and plasmids extracted using a Qiagen Midiprep Kit (Qiagen, Hilden, Germany). Extracted plasmids (50 ng) were then cloned into the pDEST™22 plasmids (450 ng) using Gateway® LR recombination. Reactions were carried out at 25°C for 16-20 hours, inactivated with 2 µl of Proteinase K at 37°C for 15 min and then a final step of 75°C for 10 min. In this method, the cDNA sequence in the donor plasmid (pDONR™222) are flanked by sites known as "attL1 and attL2", and the lethal ccdB gene in the destination vector (pDEST™22) is flanked by "attR1 and attR2" sites. The LR Clones™ recognizes the "L" and "R" sites and will transfer the lethal ccdB gene into the donor vector, resulting in the cDNA

sequences residing in the destination vector. The pDEST™22 plasmid contains the GAL4 activation domain, which is necessary for promoter activation in yeast one-hybrid interactions.

Cloning of bait constructs was carried out according to Deplancke *et al.* (2006). Based on *SAL1* and putative *SAL5* upstream sequences cloned from Genome Walker™ we trimmed the cDNA portion of the sequence, leaving the untranslated region as part of the promoter sequence. *PgSAL1* and putative *PgSAL5* promoters were first cloned into the 476 p5E-mcs Gateway vector (purchased from addgene.org) using KpnI, SAlI and/or SmaI restriction enzymes and promoter-specific primers. *PgSAL1* and putative *PgSAL5* promoters were then LR cloned into the pMW#2 vector containing the histidine reporter gene (Deplancke *et al.* 2006, purchased from addgene.org) to generate promoter baits. To generate baits, promoters were cloned from the 476 p5E-mcs Gateway vector into pMW#2 using Gateway LR Clonase II enzyme mix (Invitrogen, Carlsbad, USA). Cloning reactions were incubated at 25°C for one hour, followed by addition of 1 µl of Proteinase K incubated at 37°C for 10 minutes to inactivate the Clonase enzyme. After each round of cloning, inserts were verified by PCR. pMW#2 vectors containing *PgSAL1* or putative *PgSAL5* promoters were linearized in order to be integrated into the yeast genome. See Appendix 3 (A3) for a detailed cloning protocol.

Saccharomyces cerevisiae YM4271 strain (Cerdarlane, Burlington, CA) cells were grown and transformed based on a protocol from Matchmaker Gold Yeast One-Hybrid Library Screening System (Clontech, Mountain View CA, USA). Linearized plasmids were transformed into the yeast genome using freshly prepared yeast cultures that had reached a minimum optical density (OD600) of 0.4-0.5. Yeast cultures were harvested by centrifugation. Prior to transformation yeast cells were incubated at 30°C for 30-45 minutes. Yeast cells were transformed at 42°C for 15-20 minutes. Following transformation, yeast cells were incubated at

30°C for 1.5 hours at 200 rpm of shaking. Yeast samples were harvested by centrifugation, diluted and plated onto selective media. Plates were incubated at 30°C for three to five days. See Appendix 4 (A4) for detailed protocols on cloning and yeast transformations.

For PCR screening, DNA was extraction from each selected yeast colony using Zylomase (Clontech, Mountain View, USA). PCR was conducted with pDEST™22-specific primers. PCR products were run on agarose gels to determine if single or multiple transformations occurred, and to determine relative cDNA size. DNA from yeast colonies yielding a single band were then extracted using the Qiagen PCR Purification Kit (Qiagen, Hilden, Germany) and the insert sequenced using pDEST™22-specific or oligo dT primers. In instances where nucleotide bases could not be called by the sequencing threshold, the corresponding bases were obtained by examining the raw sequencing data (characters in bold in appendix sequences A1). If the raw sequencing data were ambiguous, the sequences substituted by comparison to the white spruce cDNA sequence in NCBI with the highest sequence similarity (underlined characters in the sequence of Appendix 1, A1).

A subset of positive Y1H colonies subjected to sequencing were then used for a BLASTn search to confirm sequence identity. A subset of six sequences were selected for additional analyses. A more robust search was conducted by translating the six nucleotide sequences obtained from NCBI into the longest open reading frame (ORF) amino acid sequence. Longest ORF was determined by ORF Finder (ncbi.nlm.nih.gov/orffinder/). ORF start codons were identified as an "ATG" codon or alternative initiation codons. Sequences were then resubmitted to BLASTn to find similar sequences to determine identities. TFs identified from Y1H were determined to be full length or partials based on an alignment of their longest amino acid ORF with the highest similarity full-length spruce clone.

To provide additional evidence of sequence identity, a subset of Y1H TFs were submitted to a motif search or pairwise sequence comparisons with close relatives. To ensure the correct *PgMYB* were identified, sequence similarities for the *PgMYB* sequence similarity table (Supplemental Table 3.1) were restricted to the *PgMYB1-13* genes identified by Bedon *et al.* (2007). Sequence similarities are based on pairwise comparisons from sequence similarity determined with EMBOSS NEEDLE global amino acid alignment (ebi.ac.uk/Tools/services/web_emboss_needle/toolform.ebi). To confirm the presence of NBS-LRR (nucleotide-binding site/leucine-rich repeat) and WRKY domains, we performed additional motif analysis for this sequence. The NBS-LRR and WRKY domains were identified using TF domain database searches with Plant TFDB (planttfdb.cbi.pku.edu.cn/blast.php). Nucleotide sequences identified by Y1H can be found in Appendix 2 (A2).

A 3-amino-1,2,4-triazole (3-AT) screen was performed to assess the strength of the promoter-TF interaction of six Y1H TFs that we decided to pursue further based on identities obtained from the BLASTn search, using a protocol similar to the manufacturers protocol from the ProQuest^{TM} Two-Hybrid System (Invitrogen, Carlsbad, CA, USA). The chemical 3-AT inhibits the transcription of histidine, therefore colonies that continue to grow on higher concentrations of 3-AT represent a stronger DNA-protein interaction. The six Y1H TFs were extracted from yeast and transformed again into yeast using the above described methodology, to confirm this was a true interaction. Yeast was grown on plates lacking histidine and tryptophan to confirm the presence of the promoter and TF in the yeast cells. The histidine reporter gene is adjacent to the promoter sequences integrated into the yeast genome, whereas the tryptophan reporter gene is encoded within the vector containing the TF cDNA. The negative control yeast

lines contain the promoter being screened as well as the corresponding empty vector (pDEST™22), which was used for cDNA library construction.

A detailed description of replica plating is found in Appendix 5 (A5). Colonies were grown for two to three days at 28°C on non-selective YPDA media and then replica plated with sterile velvets onto amino acid drop-out media (containing: yeast nitrogen base without amino acids, amino acid media of choice, agar, glucose) with increasing concentrations of 3-AT on separate plates (See Appendix A3 for media recipes). One velvet was used for up to five replica plates. Each replica plate was cleaned with a minimum of five fresh velvets, and grown for two to three days at 28°C. Plates were photographed and visually observed for signs of yeast growth.

3.2.4 *In silico* promoter analysis

To identify TFBS present in promoter sequences of *PgSAL1* and putative *PgSAL5* promoters, sequences were submitted promoter sequences to rVista through the zPicture alignment tool (rvista.dcode.org, Loots *et al.* 2002, Loots and Ovcharenko 2004). rVista uses a comparative sequence analysis approach to identify putative plant TFBS based on sequences of previously described TFBS (Loots *et al.* 2002). We used the TRANSFAC V10.2 plant library and imposed a 0.75 matrix cut off, which has been shown to be a sufficient and acceptable cutoff to detect similarities while balancing the possibility of false positives (Loots *et al.* 2002, Loots and Ovcharenko 2004, Donner and Scarpella 2013).

To further confirm the cloned promoters belonged to the corresponding *SAL* gene, we performed an *in silico* BLASTN search using default parameters of the cloned promoters, and known cDNA sequences against the PG29 v.4.0 (Birol *et al.* 2013) and WS77111 v1.0 (Warren *et al.* 2015) white spruce genome assemblies on ConGenie (congenie.org/). Default parameters of the ConGenie search included: BLOSUM62 scoring matrix, e-value cutoff of 1e-3, standard

query genetic code, standard database genetic code, and 10 results returned. Higher e-value cutoffs were imposed for cDNA sequences of *SAL1* (90%) because contigs of higher e-values had good alignments. We submitted the entire cloned portion of the promoter, including the UTR and the cloned portion of the coding region (regions distinguished in Appendix A1). The known cDNA sequences of *SAL1* and *SAL5* also contained a UTR region. Preliminary alignments against queries were conducted using EMBOSS Needle nucleotide alignment (ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html). Contigs determined to be the most likely to be representative of the query sequences based on preliminary alignments. The final alignment figure of *SAL1* containing the most representative contig, cDNA sequence and cloned promoter containing a cloned portion of the coding sequence were aligned using MAFFT (mafft.cbrc.jp/alignment/server/). MAFFT parameters were as follows: auto alignment method, unalign level 0, gap open penalty 1.53, offset value 0, score of 0 assigned to "N" regions, and the default guide tree. The putative *SAL5* final alignment figure consisting of the most representative contigs, cDNA sequence and cloned promoter containing a cloned portion of the coding sequence were less conserved and contained large insertions, and thereby the alignment was conducted using Geneious v10.2.3 (geneious.com, Kearse *et al.* 2012) with the Mauve plugin.

3.3 Results

3.3.1 Isolation and *in silico* analyses of two *SAL* promoters

We were able to clone 923 and 1798 bp upstream of the *PgSAL1* and *PgSAL5* transcriptional start sites, respectively, combining GenomeWalker™ and mining of v1.0 of the white spruce draft genome (Sundell *et al.* 2015, see promoter sequences in Appendix A1). Both sequences were cloned as single fragments, verifying the integrity of the sequence. The *SAL1*

and *SAL5* promoters shared a 30% sequence identity (Figure 3.1). In both cases, the cloned sequences contained a portion of the coding sequence, allowing us to verify that both sequences are upstream of *MIKC* sequences. In both cases, this cloned region of the coding sequence only contained a portion of the M domain, which is highly conserved between *MIKC* genes and therefore did not allow us to verify that the promoter fragments were upstream of the targeted *SAL* genes. To investigate if the cloned sequences were upstream of the *SAL1* and *SAL5* genes, we used both the cloned promoter fragments and the cDNAs to query the PG29 v4 and WS77111 v1 white spruce genome assemblies using the ConGenie BLASTn function (www.congenie.org, last accessed January 8 2018). We used the highest hit contigs from these queries to construct preliminary alignments in order to determine which contigs produced a comparatively long and relatively continuous alignment (Table S3.1). Contigs determined to most likely represent the genomic portions of the *SAL1* or *SAL5* promoter and/or cDNA (i.e. the "best contigs") were submitted to a reciprocal BLASTn search as a method of confirmation if this contig would result in *SAL1* or *SAL5* as the highest BLASTn hit (Table S3.2). The best contig or contigs were aligned to the *SAL1* or *SAL5* promoter containing the cloned cDNA and *SAL1* (Figure S3.1) or SAL5 cDNA (Figure S3.2, Figure 3.3).

From this analysis, the best contigs aligning to both the *SAL1* promoter and *SAL1* cDNA were determined to be Pg-01r141201s2137277 from the PG29 v4 assembly, and Pg-02r141203s0882372 from the WS77111 v1 assembly (Table S3.1). This determination was based on the degree of alignment between the sequences, and that in the reciprocal BLAST using the contigs as the query, GQ03605_C12 (*SAL1*) was returned as the highest hit. This analysis lends confidence that the cloned *SAL1* promoter is indeed upstream of the *SAL1* coding sequence.

Using both the putative *SAL5* promoter and *SAL5* cDNA as queries produced less certain results. The best contigs for the *SAL5* promoter and *SAL5* cDNA were determined to be Pg-01r141201s0119707, Pg01r141201s23567302, and Pg-01r141201s2765746 from the PG29 v4 assembly, and Pg-01r141201s0119707 from the WS77111 v1 assembly (Table S3.1). However, the alignments for these contigs contained fewer stretches of continuous nucleotide alignments in comparison to to the *SAL1* alignment, which gives us less confidence that the contigs retrieved with the *SAL5* cDNA are not bona fide *SAL5* sequences. Furthermore, when the best contigs for *SAL5* are used as the BLAST query, the *SAL5* cDNA is not returned as the highest hit. This was true even for the contigs found by the *SAL5* cDNA best contig. Therefore, the current genome assemblies were not sufficient for us to determine whether the cloned *SAL5* promoter is in fact upstream of the *SAL5* coding region. For this reason, we refer to this promoter as the putative *SAL5* promoter.

3.3.2 TFBS search suggests *PgSAL* promoters are regulated by similar networks

The *PgSAL1* and putative *PgSAL5* promoter sequences were submitted to rVista v2.0 to identify putative TFBS within each of these promoters. The rVista search identified 64 TFBS for the *PgSAL1* promoter, and 66 for the putative *PgSAL5* promoter (Figure 3.2). Of these, 44 (69%) *PgSAL1* and 46 (69%) *PgSAL5* TFBS were found to be legitimate, nucleus-based TFs that had a defined function according to TAIR or UniProt (Tables 3.2 and 3.3, Figure 3.2). 42 of these motifs were shared between *PgSAL5* and *PgSAL1* promoters (Figure 3.2). TFBS were categorized as legitimate if literature searches revealed TFBS were valid, if the TF associated with the binding site was located in the nucleus, and in the TF associated with the binding site was found to have a function that was not limited to having a general role in transcription and extended for example to development or environmental responses. Some of the identified TFBS

were general motifs, and did not implicate a specific gene. In these instances, only one gene in that family of motifs was selected to be represented in Tables 3.1 and 3.2. After applying the above criteria, we determined that the *PgSAL1* and putative *PgSAL5* promoters each possessed TFBS associated with response to hormones (GA, ethylene, ABA, auxin, cytokinin), defense/wounding response, abiotic factors (such as cold temperatures, light and water stress), root development, development of reproductive structures, meristem development, cell division, cell differentiation, pigment biosynthesis and cell wall biosynthesis (Figure 3.3). These identified TFBS were not necessarily the same motif. Some differences in TFBS identified between the two promoters was the presences of MYB80, PIF3 (POLYCHROME INTERACTING FACTOR 3) and ZAP1/WRKY1 (ZINC-DEPENDENT ACTIVATOR PROTEIN-1) in putative *PgSAL5*, and ATHB9 (ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 9) and NAC (NAM, ATAF1/2, CUC2) in *PgSAL1* (Figure 3.3). The putative *PgSAL5* promoter had a greater number of motifs associated with the response to GA (PIF3), abiotic stress (PIF3) and defense/wounding (ZAP1/WRKY1, Figure 3.3). The *PgSAL1* promoter was found to have a greater number of motifs associated with the develop of seeds (NAC), meristems (ATHB9), leaves (ATHB9), as well as cell differentiation (ATHB9) and cell wall biosynthesis (NAC, Figure 3.3). The *PgSAL1* and putative *PgSAL5* promoters had 11 motifs associated with reproductive structure development, 10 of these motifs shared between the two promoters. Additional reproductive structure motifs included ATBH9 for *PgSAL1* and MYB80 for putative *PgSAL5*.

We further looked at the distribution of motifs within promoter sequences by visualizing the type and number of motifs with pie graphs (Figure 3.4). Both the *PgSAL1* and putative *PgSAL5* promoters contains the following motifs: AP2/B3, AP2/ERF, ARF, ARR, BHLH,

BZIP, DOF, E2F, EIL, GATA, HD-ZIP, MADS, MYB, PHD, RITA, TALE/KNOX, TCP, TRIHELIX, and ZNF. For the *PgSAL1* promoter, the most abundant motifs were BZIP (16%), MYB (16%), and MADS (11%, Figure 3.4A). For the putative *PgSAL5* promoter, the most abundant motifs were BZIP (18%), MYB (18%), MADS (11%, Figure 3.4B). The unique motifs for the *PgSAL1* promoter were ATBH9 (50%) and ABI4 (50%, Figure 3.4C), and for the putative *PgSAL5* promoter were MYB (33%), ZAP1/WRKY1 (33%) and PIF (33%, Figure 3.4D).

3.3.3 Six putative regulators bind *PgSAL1* and putative *PgSAL5* promoters in weak and strong interactions

Y1H assays were conducted to identify white spruce proteins interacting with the cloned *PgSAL1* and putative *PgSAL5* promoters. Approximately 800 yeast colonies were screened for *PgSAL1* and putative *PgSAL5* promoter interactions (refer to A3 for a detailed protocol). Approximately 317 yeast colonies screened by PCR possessed a single band, and as a result these PCR products were chosen for sequencing. Sequencing results were used to query the NCBI database using BLASTn, and pared down to six Y1H TFs for further analysis based on BLASTn identities which represented plausible TFs (Table 3.5). A list of other putative TFs that were identified in the Y1H screen but were not used for further analysis are listed in Appendix 5. Most of these did not produce a significant alignment when used to query the NCBI database.

From this full set of interacting proteins, we focused on a subset of interacting TFs whose putative functions shed some light on the signaling networks regulating *PgSAL1* and *PgSAL5*, and therefore offer clues as to the functions of *PgSAL1* and *PgSAL5* (Table 3.4). From the *PgSAL1* promoter-interacting proteins, we selected SUPPRESSOR OF OVEREXPRESSION OF CONSTANS-like (SOC1-like), FLC EXPRESSOR-like (FLX-like), and ABSCISIC ACID STRESS RIPENING-like (ASR-like) for the Y1H interaction strength assay. *PgSOC1-like* and

PgASR-like cDNAs were full-length, while *PgFLX*-like was a partial cDNA, with approximately 500 nucleotides truncated from the 3' end. *PgSOC1*-like sequences shared 64.8% sequence similarity with *SOC1/PTM5* from *Populus tremuloides*, a characterized *SOC1*-like gene from a perennial species (Cseke *et al.* 2003). *PgSOC1*-like has 100% sequence similarity to *Pg*GQ023235_L08, a spruce cDNA included in our phylogenetic analysis (Chapter 2). The phylogenetic analysis showed that *PgSOC1/Pg*GQ02335_L08 is sister to the clade containing *Arabidopsis SOC1*. *PgFLX*-like had a 29.8% sequence similarity to *FLX-like 3* gene in *Cicer arietinum* L. *PgASR*-like had a 22.7% similarity to *Solanum lycopersicon* L. *ASR4* and 33.3% similarity to *Solanum lycopersicon ASR1*.

From the putative *PgSAL5* promoter-interacting proteins, we screened CAPRICE/ENHANCER OF TRYPTYCHON AND CAPRICE-like (CPC/ETC-like), an R2R3 MYB, PgMYB1, and nucleotide binding site-leucine rich repeat NUCLEOTIDE BINDING SITE-LEUCINE RICH REPEAR/WRKY (NBS-LRR/WRKY) for the Y1H interaction strength assay (Table 3.4). *PgMYB1* was a full-length sequence, *PgCPC/ETC*-like was a partial sequence with 46 amino acids absent from the 5' end. The *PgNBS-LRR/WRKY*-like sequence was a partial sequence based on its shorter length (560 bp nucleotide, 147 amino acid) relative to its corresponding full length white spruce cDNA sequence *Pg*GQ0033_E20 (710 bp nucleotide, 171 amino acid), and relative to the *Arabidopsis* full length sequences showing the highest sequence similarity, AT1G69550.1 (5244 bp, 1400 amino acids). The *PgNBS-LRR/WRKY*-like partial cDNA also appeared to contain deletions and insertions. The Y1H *Pg*MYB1 TF was of interest to us because of the phylogenetic and functional characterization that has been carried out for this gene (Bedon *et al.* 2007, Bomal *et al.* 2008, Bomal *et al.* 2014). *PgMYB1* identity was validated by sequence comparison to *PgMYB1-PgMYB13*, a subset of the *PgMYB* TFs identified

by phylogenetic analysis as belonging to the same clade (Bedon *et al.* 2007). Whereas the *Pg*SAL1 promoter-interacting *PgMYB* had a 98.5% amino acid sequence similarity to *PgMYB1*, the sequence similarities of *PgSAL1* promoter-interacting *PgMYB* to *PgMYB2-13* ranged from 27.2% to 37.6%. This level of sequence similarly indicates that the *PgMYB* sequence identified by Y1H is *PgMYB1*.

The Plant TFDB domain search identified the *PgNBS-LRR/WRKY*-like TF sequences to have a WRKY domain (e-value 1e-09). The NBS-LRR domain was also identified by the Plant TFDB domain search (e-value 1e-06 to 6e-04). A globally optimized alignment of *PgNBS-LRR/WRKY*-like and *Arabidopsis* WRKY19 yielded a low sequence similarity of 3.5%, and 4.6% with *Arabidopsis WRKY16* (Table 3.4). The truncated *PgNBS-LRR*-like likely contributed to the low sequence similarity, as the *AtWRKY16* and *AtWRKY19* amino acids sequences were approximately 1200 to 1750 bp longer than *PgNBS-LRR*-like. Using alignment (Supplemental Figure A3.6, Figure S3.7) and conserved domain identification (ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi, Supplemental Figure S3.8), we identified that the *PgNBS-LRR/WRKY* domain aligned to separate LRR domains in the *AtWRKY16* and *AtWRKY19* sequence. I believe *PgNBS-LRR/WRKY*-like may be a truncated version of the spruce clone *Pg*GQ0033_E20. *PgNBS-LRR/WRKY*-like appeared to be a hybrid of NBS-LRR and WRKY, since both domains are present and share sequence similarity with previously identified *Arabidopsis* NBS-LRR and WRKY hybrids (Rinerson *et al.* 2015). Despite the low sequence similarity of *PgNBS-LRR*-like to *AtWRKY19* and *AtWRKY16*, the identified WRKY domain appeared to be conserved even in this truncated sequence.

All six of the above proteins were determined to interact with the *PgSAL1* or putative *PgSAL5* promoter to a greater degree than the negative controls (Figure 3.5). Interactions were

identified as “strong” or “weak” based on the amount of visible yeast growth on increasing concentrations of 3-AT. The *PgSAL1* promoter was identified to have a weak interaction with PgSOC1-like, and strong interactions with PgFLX-like and PgASR-like (Table 3.4). The *PgSAL1* promoter negative control exhibited almost no growth at 5 mM 3-AT. At 5 mM 3-AT, PgSOC1-like showed weak growth, while PgFLX-like and PgASR-like had more pronounced growth. PgFLX-like and PgASR-like colonies also exhibited some weak growth on 10 mM 3-AT plates, while PgSOC1-like and the negative control had no growth. Growth of the putative *PgSAL5* promoter negative control was dramatically reduced at 10 mM 3-AT and absent at 20 mM 3-AT. PgMYB1 appeared to have a weak interaction, since its growth at 10 mM was only slightly better than that of the negative control. PgMYB1 also appeared to have very small amounts of growth at 20 mM 3-AT. PgCPC/ETC-like and PgNBS-LRR/WRKY-like had strong interactions because they grew well on 10 mM 3-AT, and formed visible colonies on 20 mM 3-AT.

The putative *PgSAL5* promoter yeast line appeared to have a higher baseline activation of the histidine reporter gene compared to the *PgSAL1* promoter line. Promoters are integrated into the yeast genome independently, and therefore can have different baseline levels of expression based on their location in the genome. Different genomic integration sites may explain why the putative *PgSAL5* promoter yeast line was able to grow on higher levels of 3-AT than the *PgSAL1* promoter yeast line.

3.4 Discussion

We demonstrated in Chapter 2 that white spruce SAL1 and SAL5 genes are homologous to both angiosperm SVP and AGL24 genes. PgSAL1 and PgSAL5 showed similar but not identical patterns of expression (Chapter 2), peaking within the first two weeks of bud formation

and declines at later time points. Here, we investigated the possible regulatory networks that *SAL1* and *SAL5* function in, using both in silico identification of TFBS and *in vivo* identication of promoter-interacting protens via Y1H. These approaches are useful for identifying possible functions for these genes. These complementary approaches allowed us to examine whether the identified putative TFs regulating *PgSAL1* and *PgSAL5* are the same or different, as well as provide evidence to support a role for either or both *PgSAL1* and *PgSAL5* in processes associated with bud formation.

In the absence of a quality reference white spruce genome assembly at the time that this research was conducted, I cloned the promoters for *PgSAL1* and *PgSAL5* using Genome Walker. I used the two cloned promoters and the *SAL1* and *SAL5* cDNA sequences to query the PG29 v4 and WS77111 v1 white spruce draft assemblies to ascertain whether the cloned promoter sequences and targeted cDNA sequences could be aligned with confidence to the same genomic contig, thereby providing evidence that the cloned promoters are upstream of the targeted *SAL* coding sequences. From these analyses, we have confidence that the *SAL1* promoter is upstream of the *SAL1* coding sequence, and is therefore a bona fide *SAL1* promoter (Supplemental Figure 3.1, Supplemental Table 3.1, 3.2). In contrast, our analyses suggested that neither the *SAL5* promoter nor *SAL5* cDNA sequence are not represented in the current PG29 or WS77111 draft assemblies, since the alignment of the contigs with the *SAL5* promoter or cDNA sequences had large gaps and large regions of mismatches. The finding that the *SAL5* locus is not represented in these draft assemblies is not surprising, given that very few contigs contain a fully assembled locus, and only a small number of contigs contain multiple loci (Pavy *et al.* 2017). We have some evidence that the cloned promoter is mostly likely *SAL5* based on the high sequence similarity and low e-value to GQ03806_I20 (*SAL5*) when we use the cloned portion of the *SAL5*

UTR and coding sequence as BLAST queries. However, this region of the coding sequence contains the conserved M domain, and thus does not have unique *SAL5* signatures. For this reason, we have less evidence that the cloned SAL5 promoter is upstream of the *SAL5* coding sequence and therefore have referred to this promoter as the "putative SAL5 promoter". The inability to identify a high confidence contig from the *SAL5* cDNA demonstrates the limitations of the current white spruce genome assemblies. Considerable improvements to the assembly by additional sequencing and improved assembly methods are required in order for the white spruce genome to reach reference status.

The majority of putative TFBS identified by rVista were shared between *PgSAL1* and putative *PgSAL5* promoters. The Y1H identified six promising TFs for future investigations. Many of the putative TFBS and Y1H TFs suggested functions related to growth and development. In addition, many TFBS related to hormone regulation were identified.

3.4.1 *PgSAL1* and putative *PgSAL5* promoters are regulated by external stimuli and hormones

Pathways that may be involved in regulating the promoter activity of both *PgSAL1* and putatively *PgSAL5* include response to water stress, cold temperatures, light quality, biotic and abiotic stress, ethylene, GA, ABA, auxin, and CK. Both promoters also contained TFBS associated with cell division, cell differentiation, the development of reproductive structures, and seed development. Even though the promoters only shared ca. 30% sequence identity, almost all of the TFBS with defined functions were identified in both the *PgSAL1* and putative *PgSAL5* promoters, suggesting that the genes may be regulated by similar signaling pathways. The TFBS database searched demonstrated that the putative *PgSAL5* promoter had additional TFBS associated with the response to ABA, light, GA, and defense. In contrast, the *PgSAL1* promoter

was regulated by two additional TFs linked to development of the leaf, the seed, the meristem, cell wall biogenesis and cell differentiation. Based on these differences, the putative *PgSAL5* promoter may be associated with more regulatory responses related to hormonal control and external stimuli, while the *PgSAL1* promoter had some TFs linked to cell cycle control and structure development.

Potential roles development and phase transitions at the meristem are demonstrated by the presence of GAMYB and WRKY TFBSs in both the *PgSAL1* and putative *PgSAL5* promoters. GAMYB has been shown to be involved in the response to GA and ABA, as well as seed storage, floral initiation, stem elongation, anther development and seed development (Washio 2003, Woodger *et al.* 2003). The WRKY TF family has a variety of roles in plant development and biotic and abiotic stress (Ciolkowski *et al.* 2008). This is in agreement with our Y1H, which also suggested that the putative *PgSAL5* promoter interacted with a defense-related WRKY TF. Altered flowering time in *Arabidopsis* plants by soybean WRKYs further demonstrates that transition from flowering is affected by regulatory networks involved in the stress response, and appears to be mediated by WRKYs (Yang *et al.* 2016).

Multiple DNA binding motifs involved in light perception were identified in the *PgSAL1* and putative *PgSAL5* promoters, indicating that *PgSAL* genes are a downstream target of light perception and/or light quality. The *PgSAL1* and putative *PgSAL5* promoter possess four TFBS for the light response: CPRF2, CPRF3, RAV1 and TAV1 (Table 3.2, 3.3). This finding is in agreement with the knowledge that bud formation is accelerated under SD in white spruce (Hamilton *et al.* 2016) and that growth cessation and bud formation in other species such as *Populus spp*. is influenced by light signals and day length (Olsen 2010).

A greater number of TFBS were identified in the putative *PgSAL5* promoter, likely because the putative *PgSAL5* promoter sequence submitted for the motif search was about 800 bp longer than the promoter sequence submitted for *PgSAL1*. Larger fragments of promoter sequence are naturally more likely to increase the number of TFBS identified. We must acknowledge that TFBS need to further be validated and we predict that based on the nature of the search and conservation of functional TFBS that the number of true TFBS will likely be fewer than the number originally identified. Also, in order to confirm that the cloned promoters are upstream of the *SAL1* or *SAL5* gene, the promoter and the entire respective *SAL* coding sequence should be cloned from white spruce gDNA.

3.4.2 Yeast One-Hybrid Assay identified six proteins that may regulate *PgSAL* genes

Based on the Y1H assay, we have at least four putative TFs that appear to be good candidates to interact physically with *Pg*SAL promoters (Figure 3.5, Table 3.4): *Pg*ASR-like, *Pg*CPC/ETC-like, *Pg*FLX-like, and *Pg*NBS-LRR/WRKY-like. Here, I will discuss the potential roles of these strong and weak (*Pg*SOC1-like, *Pg*MYB1), interactions with *PgSAL1* or putative *PgSAL5*. Interactions could have been more accurately characterized as "strong" and "weak" if we possessed a positive Y1H control to compared our interactions. However, due to the resources available in the white spruce system, a positive Y1H control was not available. Further experimental validation of these interactions could be carried out with targeted deletions of portions of the promoter sequences to determine the area the TFs bind.

3.4.2.1 Transcription factors that interact with the putative *PgSAL5* promoter suggest roles in development and beyond bud formation

Since *Pg*NBS-LRR/WRKY-like protein possesses similarities to both LRR and WRKY domains, we speculate that this protein may help facilitate different protein interactions and binding partners related to transcription. *WRKYs* have several roles in response to stress pathways, including plant defense, MAP kinase signaling, activation of ABA signaling and promotion of salt and drought tolerance (Phukan *et al.* 2016, Rushton *et al.* 2010). WRKYs also have several developmental roles, and may act to inhibit GA signaling during seed dormancy (Phukan *et al.* 2016, Rushton *et al.* 2010). For example, soybean (*Glycine max* L.) defense-related *WRKYs* (*GmWRKY58* and *GmWRKY76*) accelerate time to flowering when overexpressed in *Arabidopsis* (Yang *et al.* 2016). These results demonstrate that TFs related to defense can influence other developmental pathways, such as the transition from vegetative to reproductive growth. Furthermore, it was demonstrated through ChIP-seq that these soybean WRKYs bound to the promoter region of multiple *Arabidopsis* flowering time genes, including *AtSVP* (Yang *et al.* 2016). This range of results leads us to believe that WRKYs may function outside of their traditional roles in plant defense, and participate in the regulation of flowering, which is intriguing because of the link between flowering time and bud formation.

WRKY TFs bind the W-box, (T)(T)TGAC(C/T), in promoters. The putative *PgSAL5* promoter does possess a traditional WRKY motif (TTGACT, +1769), but the *PgSAL1* promoter does not. The presence of a WRKY motif in the putative *PgSAL5* promoter is supported by the rVista database search which identified W-boxes from WRKY TFs in the promoters of both *PgSAL1* and putative *PgSAL5* (Table 3.2, 3.3). The pattern of both analyses suggests the presence of a potentially functional WRKY motif provides further evidence that there is a true interaction between putative and the WRKY TF.

PgNBS-LRR/WRKY-like may be a chimeric NBS-LRR/WRKY TF, demonstrated by the identified domains and high DNA binding affinity, and may have regulatory functions. Chimeric proteins containing both the NBS-LRR and WRKY domains have been identified in flowering plants (Rinerson *et al.* 2015, Rushton *et al.* 2010). In *Arabidopsis*, three *NBS-LRR/WRKY* genes have been identified: *AtWRKY16*, *AtWRKY19* and *AtWRKY52* (Rinerson *et al.* 2015). The *AtWRKY16* and *AtWRKY19* genes have DNA-binding capabilities, in addition to roles in signaling in the innate immune response (Rinerson *et al.* 2015, Rushton *et al.* 2010). Since our *PgNBS-LRR/WRKY*-like gene has a strong interaction strength on 3-AT plates, it seems unlikely its only purpose is in intracellular signaling. If *PgNBS-LRR/WRKY*-like only possessed a role in intracellular signaling and its interaction with the putative *PgSAL5* promoter is non-specific, we would predict would be more likely to have a weak interaction. In angiosperms, there have been at least eight types of *NBS-LRR/WRKY* genes identified, which possess unique combinations of NBS-LRR domains, WRKY domains and additional protein domains (Rinerson *et al.* 2015). Since our *PgNBS-LRR*-like sequence appears to be a partial sequence, in combination with the fact it appears to be similar to the relatively newly characterized *NBS-LRR/WRKY* gene hybrid, it is difficult to classify this protein with absolute certainty.

Despite well characterized NBS-LRR roles in disease resistance, members of this family have also been shown through yeast-two hybrid assays to interact with *MADS-box* genes (Cseke *et al.* 2007, Acevedo *et al.* 2004, Gamboa *et al.* 2001). In poplar, the *PTM5* and *PtLRR* protein-protein interaction was proposed to represent a novel *LRR*-type gene to mediate protein-protein interactions (Cseke *et al.* 2007). Furthermore, other LRR proteins in *Arabidopsis*, such as CLAVATA1 and ERECTA, have demonstrated developmental roles through protein-protein interactions and intercellular signaling (Clark *et al.* 1997, Torii *et al.* 1996).

Similar to most of the other interacting proteins that we identified with Y1H, functional characterization of the R3 MYB *CPC/ETC* genes in other species revealed possible roles in development. I could not find a direct link between *CPC/ETC* genes and *SVP*-like genes in the current literature. *Arabidopsis CPC* and *ETC* have contrasting roles in root development. CPC acts as a positive regulator of hair-cell differentiation and is involved in cell fate determination in epidermal cells (Wada *et al.* 1997). *ETC1* genes are believed to act as negative regulators of trichome differentiation, and a positive regulator of the development of root hairs (Kirik *et al.* 2003). Interestingly, single loss of function *ETC3/CAPRICE-LIKE MYB3 (CPL3)* mutants exhibit delays in flowering, which suggests *ETC3* has a role in the transition from vegetative to reproductive growth at the meristem (Tominaga-Wada *et al.* 2013). These findings lead us to propose that putative *PgSAL5* plays a role in cell fate determination and development. Additional *in vivo* or *in vitro* analysis is necessary to confirm the interaction of the *SVP*-like promoters and their proposed interaction partners.

PgMYB1, found to interact with the putative *PgSAL5* promoter, is perhaps the best characterized of all conifer transcription factors (Bedon *et al.* 2007, Bomal *et al.* 2008, Bomal *et al.* 2014). The presence of the MYB recognition sequence (TAACTG) in putative *PgSAL5* (+235 to 240) and *PgSAL1* (+209 to 212) is in agreement with the MYB TFBS identified by the database search. *PgMYB1* is hypothesized to be involved in lignin biosynthesis by regulating phenylpropanoid metabolism (Bedon *et al.* 2007). *PgMYB1* also has a high sequence similarity to *MYB1* in loblolly pine (*Pinus taeda* L.), which is suggested to be a positive regulator of lignin-synthesizing enzymes (Patzlaff *et al.* 2003). Bedon *et al*. (2007) found *PgMYB1* to be homologous to *Arabidopsis MYB20*, which is involved in cell differentiation and regulating fiber development (Ehlting *et al.* 2005). Our finding that *PgMYB1* interacts with the putative *PgSAL5*

promoter suggests that *PgSAL5* may possess roles outside of vegetative or reproductive bud formation, possibly in regulation of secondary cell wall formation. If true, it would suggest that *PgSAL5* has acquired functions distinct from *PgSAL1*, which seems more likely to regulate events associated with bud formation.

3.4.2.2 Transcription factors that interact with the *PgSAL1* promoter suggest role in bud formation

FLX is a component of the flowering pathway. It has been demonstrated in yeast and transient *in planta* assays that *At*FLX complexes with FRIGIDA (FRI) in order to promote the baseline expression of *FLC* (Choi *et al.* 2011, Ding *et al.* 2013). FLC in turn inhibits the expression of genes such as *SOC1*, *FT*, and *TWIN SISTER OF FT* (*TSF*), which repress flowering (Borner *et al.* 2000, Lee *et al.* 2000, Samach *et al.* 2000, Hepworth *et al.* 2002, Michaels *et al.* 2005). Vernalization, an extended period of cold temperatures required to initiate flowering in some species, alleviates the repression of *FLC* to promoter flowering (Michaels and Amasino 1999, Sheldon *et al.* 1999). The interaction of the *PgSAL1* promoter with FLX suggests that *PgSAL1* could be regulated by cold temperatures, and that TFBSs involved in the cold response are conserved within the promoter. This evidence that *PgSAL1* may be cold regulated is in agreement with our database search, which identified TFBS regulated by cold temperatures, including RELATED TO ABI1/VP1 (RAV1). *AtRAV1* is under circadian clock regulation (Fowler *et al.* 2005), is upregulated by low temperatures (Fowler and Thomashow 2002), and may negatively regulate growth (Hu *et al.* 2004). In *Arabidopsis*, *RAV1* also positively regulates leaf senescence and may act as an integrator of environmental cues with leaf maturity (Woo *et al.* 2010). Overall these results show evidence that *PgSAL1* may be involved in *CO/FT* regulatory network and may have a role in the control of bud formation.

Additional evidence linking *PgSAL1* to regulating developmental events at the SAM is the interaction of *Pg*SOC1-like with the *PgSAL1* promoter. SOC1 is a MADS-box protein that, in *Arabidopsis* has been found to bind the *AtSVP* promoter to allow flowering to proceed (Li *et al.* 2008, Immink *et al.* 2012). SOC1 binds to the CArG-box motif (CC[A/T]6GG). No motifs with 100% similarity to CArG-box motifs were identified by our promoter motif search in the either promoter. However, several MADS-box TFBSs were identified by the rVista search in both the *PgSAL1* and putative *PgSAL5* promoters, which include AGAMOUS (AG), AGAMOUS-LIKE 1, i.e., SHATTERPROOF1 (AGL1), AGAMOUS-like 15 (AGL15), AGAMOUS-LIKE 2, i.e. SEPALLATA1 (AGL2), and AGAMOUS-LIKE 3, i.e. SEPALLATA4 (AGL3). Kaufmann *et al.* (2009) identified through ChIP-seq in *Arabidopsis* that CArG-box-like motifs can be sufficient for interacting with MADS-box TFs. We propose that although angiosperm motifs can be useful tools when searching conifer promoters, it is possible these motifs may not be fully conserved in conifers. Lack of motif conservation in conifers increases the likelihood that conifer promoters have diverged and may only possess partially conserved angiosperm motifs. Additionally, MADS-box TFs bind as hetero- or homo-dimers in order to form TF complexes (Egea-Corines *et al.* 1999, Honma and Goto 2001). We speculate that SOC1 may have had a stronger interaction with the *PgSAL1* promoter in the 3-AT screen if the necessary accompanying TFs were also present to facilitate this interaction.

We propose *SOC1*-like genes regulate *SVP*-like genes not just in the annual *Arabidopsis*, but also in white spruce. Poplar *SOC1/PTM5* has not been identified as a target or regulator of poplar *SVP*-like genes. *PTM5* demonstrates a seasonal variation in expression, and is believed to have a role in both xylem and phloem differentiation, and in the vascular cambium (Cseke et al. 2003). *At*SOC1 has been shown to be a regulator of *SVP* expression (Immink *et al.* 2012). In the

perennial woody species kiwifruit (*Actinidia chinensis* (A. Chev.) C.F. Liang & A.R. Ferguson) *AcSOC1*-like gene is proposed to function in partnership with *AcSVP*-like to impart transcriptional regulation (Voogd *et al.* 2015). This proposed function is supported the overlapping location and timing of expression of *AcSOC1*-like and *AcSVP*-like genes in kiwifruit (Wu *et al.* 2012). Voogd *et al.* (2015) proposes that *AcSVP*-like genes may also regulate transcription of *SOC1*-like genes, which would complement the *Arabidopsis* model (Immink *et al.* 2012, Tao *et al*. 2012, Gregis *et al.* 2013). *SOC1* is a part of the photoperiodic regulation of flowering time in *Arabidopsis*, and is under the regulation of the circadian rhythm. If *PgSOC1*-like is also under the control of the circadian clock, then the interaction with the *Pg*SAL1 promoter suggest that *PgSAL1* may be regulated light and circadian rhythm.

ABA has been shown to have a role in bud formation and maturation, as well as the onset of ecodormancy and growth cessation, making this hormone and related TFs interesting candidates for regulators of *PgSAL1* (Rohde *et al.* 2002, Horvath *et al.* 2003, Ruttink *et al.* 2007). The observations that ABA content increased in white spruce buds under shortened photoperiods (El Kayal *et al.* 2011) and that shortened photoperiods increased ABA in poplar apical buds (Rohde *et al.* 2002), suggest that ABA may function in a similar manner in angiosperm and conifer bud development and growth cessation. Potato (*Solanum tuberosum* L.) *ASR* regulates tuber development, which is triggered by external stimuli similar to dormancy inducing conditions, such as shortened day length, cool temperatures, and increased ABA levels (Xu *et al.* 1998, Rodríguez-Falcón *et al.* 2006). *ASR* orthologs regulate the abiotic stress response (e.g. drought and salinity) in addition to fruit ripening and tuber development in potato (Golan *et al.* 2014, Frankel *et al.* 2004). Environmental conditions such as drought, salinity induce expression of *Ginkgo biloba* L. *Asr* and *Asr* orthologs (Shen *et al.* 2005). Overexpression

of *ASR* in tomato caused an increased tolerance to abiotic stress, including salinity, drought and cold (Golan *et al.* 2014). ABA content showed a slight increase in white spruce buds near the finalization of bud formation, which may be associated with drought tolerance (El Kayal *et al.* 2011). Regulation of *PgSAL1* by ASR suggests *PgSAL1* is regulated by ABA, which could be linked to the perception of dormancy-inducing desiccation conditions. Our speculation that ASR is involved in the ABA response is further supported by the identification of motifs from our promoter databased search that are regulated by the ABA response, including ABA INSENSITIVE 4 (ABI4), ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 5 (ATBH5), and EARLY METHIONINE BINDING PROTEIN-1(EmBP1).

3.4.3 Conceptual models of *PgSAL1* and putative *PgSAL5* regulatory networks

Based on our findings in this paper and previous research we inferred upstream pathways regulating *PgSAL1* and putative *PgSAL5* (Figure 3.6). Specifically, we propose that genetic interactions as well as ethylene, light, auxin, GA, ABA, CK, defense and abiotic factors play an important role in regulation of both genes.

The TFBS search yielded very similar TFs involved in activating both *PgSAL1* and putative *PgSAL5* promoters, suggesting that they are predominantly regulated by the same pathways. The TFBS identified from the database search were identified to be involved in the hormone response (ethylene, auxin, GA, CK, ABA), the response to light, as well as the defense response and the abiotic (water stress, cold) response. However, these hormones and environmental cues also induce transcriptional changes in the regulation of pathways outside of bud formation and dormancy induction, including senescence, dormancy maintenance, dormancy release, and bud burst.

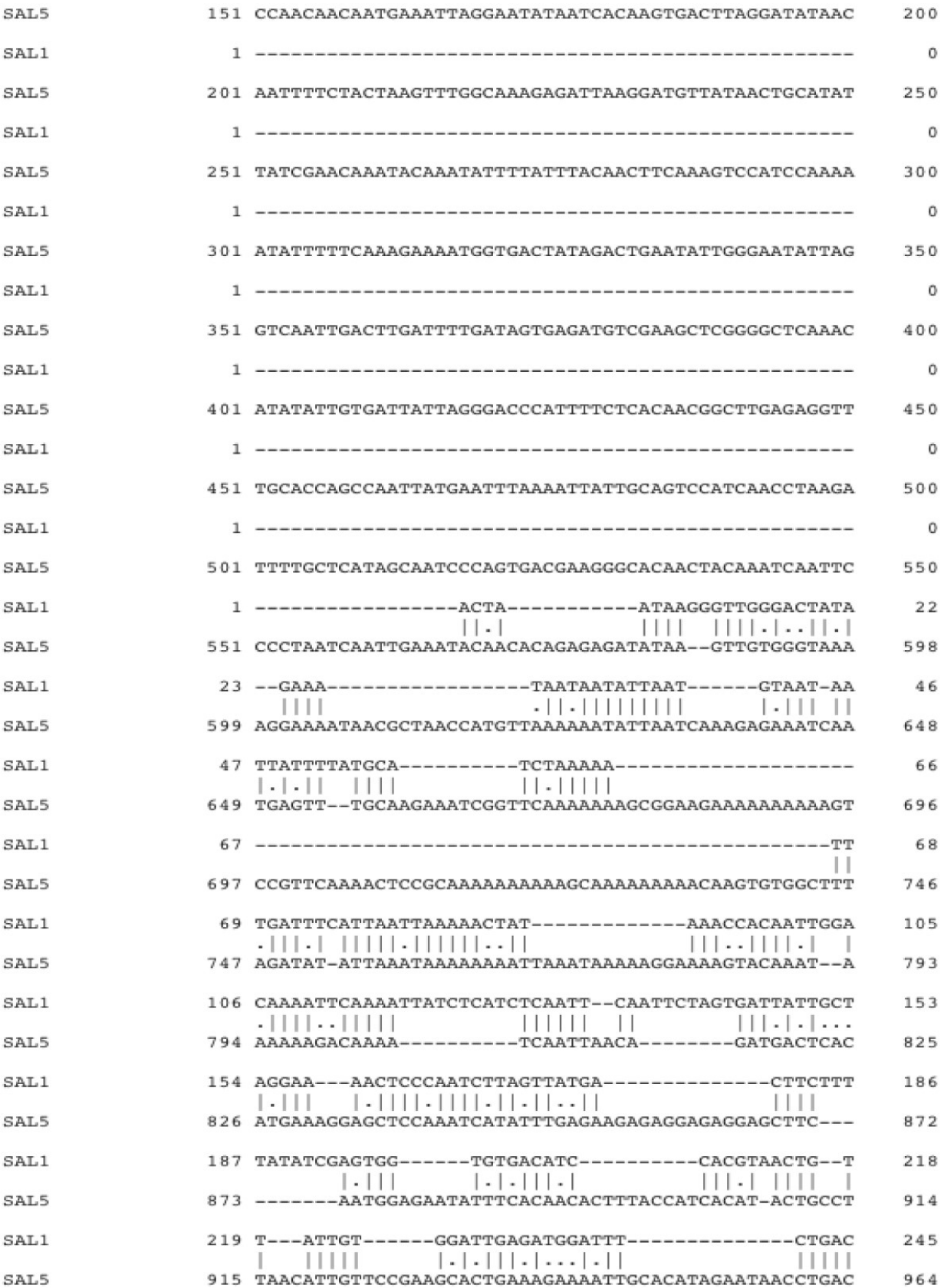
We hypothesized that *PgSAL1* was involved in processes in the early stages of bud formation and/or growth cessation based on expression data (Chapter 2). These findings are further supported here based on the DNA interacting partners that were identified through Y1H: *PgSOC1*-like, *PgASR*-like and *PgFLX*-like. *ASR*-like genes are believed to be ABA-responsive (Shen *et al.* 2005), while *FLX*-like genes may be regulated by cold temperatures (Ding *et al.* 2013). *PgSOC1*-like has not been functionally characterized in spruce; however, in poplar, S*OC1/PTM5* has been suggested to regulate the formation of wood tissues (Cseke *et al.* 2007).

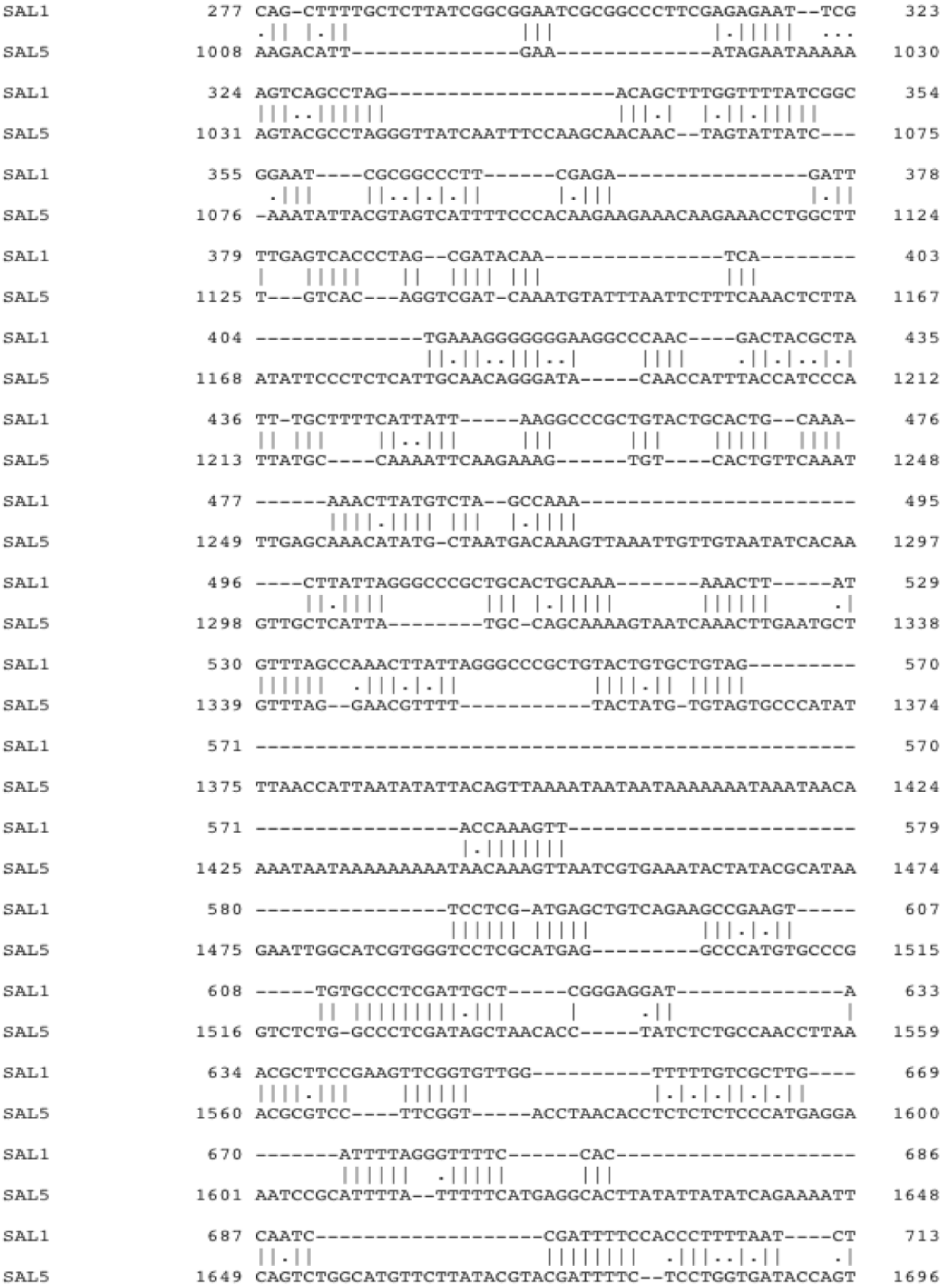
Like *PgSAL1, PgSAL5* showed an expression pattern consistent with roles in early bud development and/or growth cessation (Chapter 2). However, interacting TFs identified by Y1H *in vivo* interactions and through TFBS *in silico* analyses suggest that *PgSAL5* has a role distinct from *PgSAL1*, and perhaps different from *SVP/AGL24*-like genes characterized from other species to date. The known functions of *PgMYB1* in regulating phenylpropanoid biosynthesis and the hypothesized functions of *PgNBS-LRR*-like allow us to hypothesize that *PgSAL5* could participate in roles outside of bud initiation, and regulate development of wood tissue/secondary growth.

Our work has demonstrated that conifers may have a conserved regulatory pathway for bud formation that is similar to angiosperms. We also observed that *PgSALs* may have acquired or maintained roles that extend beyond bud formation into other areas of development not previously anticipated based on angiosperm models. An alternative theory is that angiosperms have lost part of the ancestral repertoire, since *PgSAL* are sister to angiosperm *SA*. These previously unanticipated functions may be a reflection of the evolutionary divergence between conifers and angiosperms, and our results required further experimentation to be conclusive. Additional experiments include electrophoretic mobility shift assay, tobacco co-infiltrations, or

ChIP, in order to validate the TF-promoter interactions identified by Y1H. Further work is required to verify proposed functions include RNA interference experiments in spruce for *PgSAL* genes and select TFs that regulate *PgSAL*s. Transgenics are needed to uncover the regulators and moderators of these pathways in relation to different aspects of seasonal growth. Additional experiments are also required to unequivocally link the *Pg*SAL1 and *Pg*SAL5 promoters upstream to their respective coding sequences. This can be done by cloning the entire promoter region and cDNA sequence as one piece, to confirm these promoters are upstream of the intended *SAL* genes. Through these experiments we hope to uncover if *PgSAL* genes share similar functions to their *DAM* homologs, and furthermore that these genes are non-functionally redundant/demonstrate functional divergence.

Chapter 3 Figures


SAL5 151 CCAACAACAATGAAATTAGGAATATAATCACAAGTGACTTAGGATATAAC 200
SAL1 1 ---------- 0
SAL5 201 AATTTTCTACTAAGTTTGGCAAAGAGATTAAGGATGTTATAACTGCATAT 250
SAL1 1 ---------- 0
SAL5 251 TATCGAACAAATACAAATATTTTATTTACAACTTCAAAGTCCATCCAAAA 300
SAL1 1 ---------- 0
SAL5 301 ATATTTTTCAAAGAAAATGGTGACTATAGACTGAATATTGGGAATATTAG 350
SAL1 1 ---------- 0
SAL5 351 GTCAATTGACTTGATTTTGATAGTGAGATGTCGAAGCTCGGGGCTCAAAC 400
SAL1 1 ---------- 0
SAL5 401 ATATATTGTGATTATTAGGGACCCATTTTCTCACAACGGCTTGAGAGGTT 450
SAL1 1 ---------- 0
SAL5 451 TGCACCAGCCAATTATGAATTTAAAATTATTGCAGTCCATCAACCTAAGA 500
SAL1 1 ---------- 0
SAL5 501 TTTTGCTCATAGCAATCCCAGTGACGAAGGGCACAACTACAAATCAATTC 550
SAL1 1 ----------ACTA----------ATAAGGGTTGGGACTATA 22
SAL5 551 CCCTAATCAATTGAAATACAACACAGAGAGATATAA--GTTGTGGGTAAA 598
SAL1 23 --GAAA----------TAATAATATTAAT------GTAAT-AA 46
SAL5 599 AGGAAAATAACGCTAACCATGTTAAAAAATATTAATCAAAGAGAAATCAA 648
SAL1 47 TTATTTTATGCA----------TCTAAAAA---------- 66
SAL5 649 TGAGTT--TGCAAGAAATCGGTTCAAAAAAAGCGGAAGAAAAAAAAAAGT 696
SAL1 67 ----------TT 68
SAL5 697 CCGTTCAAAACTCCGCAAAAAAAAAAGCAAAAAAAAACAAGTGTGGCTTT 746
SAL1 69 TGATTTCATTAATTAAAAACTAT----------AAACCACAATTGGA 105
SAL5 747 AGATAT-ATTAAATAAAAAAAATTAAATAAAAAGGAAAAGTACAAAT--A 793
SAL1 106 CAAAATTCAAAATTATCTCATCTCAATT--CAATTCTAGTGATTATTGCT 153
SAL5 794 AAAAAGACAAAA----------TCAATTAACA--------GATGACTCAC 825
SAL1 154 AGGAA---AACTCCCAATCTTAGTTATGA----------CTTCTTT 186
SAL5 826 ATGAAAGGAGCTCCAAATCATATTTGAGAAGAGAGGAGAGGAGCTTC--- 872
SAL1 187 TATATCGAGTGG------TGTGACATC----------CACGTAACTG--T 218
SAL5 873 -------AATGGAGAATATTTCACAACACTTTACCATCACAT-ACTGCCT 914
SAL1 219 T---ATTGT------GGATTGAGATGGATTT----------CTGAC 245
SAL5 915 TAACATTGTTCCGAAGCACTGAAAGAAAATTGCACATAGAATAACCTGAC 964


SAL1 277 CAG-CTTTTGCTCTTATCGGCGGAATCGCGGCCCTTCGAGAGAAT--TCG 323
SAL5 1008 AAGACATT--------------GAA-------------ATAGAATAAAAA 1030
SAL1 324 AGTCAGCCTAG--------------------ACAGCTTTGGTTTTATCGGC 354
SAL5 1031 AGTACGCCTAGGGTTATCAATTTCCAAGCAACAAC--TAGTATTATC--- 1075
SAL1 355 GGAAT----CGCGGCCCTT------CGAGA----------------GATT 378
SAL5 1076 -AAATATTACGTAGTCATTTTCCCACAAGAAGAAACAAGAAACCTGGCTT 1124
SAL1 379 TTGAGTCACCCTAG--CGATACAA---------------TCA-------- 403
SAL5 1125 T---GTCAC---AGGTCGAT-CAAATGTATTTAATTCTTTCAAACTCTTA 1167
SAL1 404 --------------TGAAAGGGGGGGAAGGCCCAAC----GACTACGCTA 435
SAL5 1168 ATATTCCCTCTCATTGCAACAGGGATA-----CAACCATTTACCATCCCA 1212
SAL1 436 TT-TGCTTTTCATTATT-----AAGGCCCGCTGTACTGCACTG--CAAA- 476
SAL5 1213 TTATGC----CAAAATTCAAGAAAG------TGT----CACTGTTCAAAT 1248
SAL1 477 ------AAACTTATGTCTA--GCCAAA----------------------- 495
SAL5 1249 TTGAGCAAACATATG-CTAATGACAAAGTTAAATTGTTGTAATATCACAA 1297
SAL1 496 ----CTTATTAGGGCCCGCTGCACTGCAAA-------AAACTT-----AT 529
SAL5 1298 GTTGCTCATTA--------TGC-CAGCAAAAGTAATCAAACTTGAATGCT 1338
SAL1 530 GTTTAGCCAAACTTATTAGGGCCCGCTGTACTGTGCTGTAG--------- 570
SAL5 1339 GTTTAG--GAACGTTTT-----------TACTATG-TGTAGTGCCCATAT 1374
SAL1 571 --- 570
SAL5 1375 TTAACCATTAATATATTACAGTTAAAATAATAATAAAAAAATAAATAACA 1424
SAL1 571 -----------------ACCAAAGTT----------------------- 579
SAL5 1425 AAATAATAAAAAAAAATAACAAAGTTAATCGTGAAATACTATACGCATAA 1474
SAL1 580 ----------------TCCTCG-ATGAGCTGTCAGAAGCCGAAGT----- 607
SAL5 1475 GAATTGGCATCGTGGGTCCTCGCATGAG---------GCCCATGTGCCCG 1515
SAL1 608 -----TGTGCCCTCGATTGCT-----CGGGAGGAT--------------A 633
SAL5 1516 GTCTCTG-GCCCTCGATAGCTAACACC-----TATCTCTGCCAACCTTAA 1559
SAL1 634 ACGCTTCCGAAGTTCGGTGTTGG----------TTTTTGTCGCTTG---- 669
SAL5 1560 ACGCGTCC----TTCGGT-----ACCTAACACCTCTCTCTCCCATGAGGA 1600
SAL1 670 -------ATTTTAGGGTTTTC------CAC-------------------- 686
SAL5 1601 AATCCGCATTTTA--TTTTTCATGAGGCACTTATATTATATCAGAAAATT 1648
SAL1 687 CAATC-------------------CGATTTTCCACCCTTTTAAT----CT 713
SAL5 1649 CAGTCTGGCATGTTCTTATACGTACGATTTTC--TCCTGGTGATACCAGT 1696

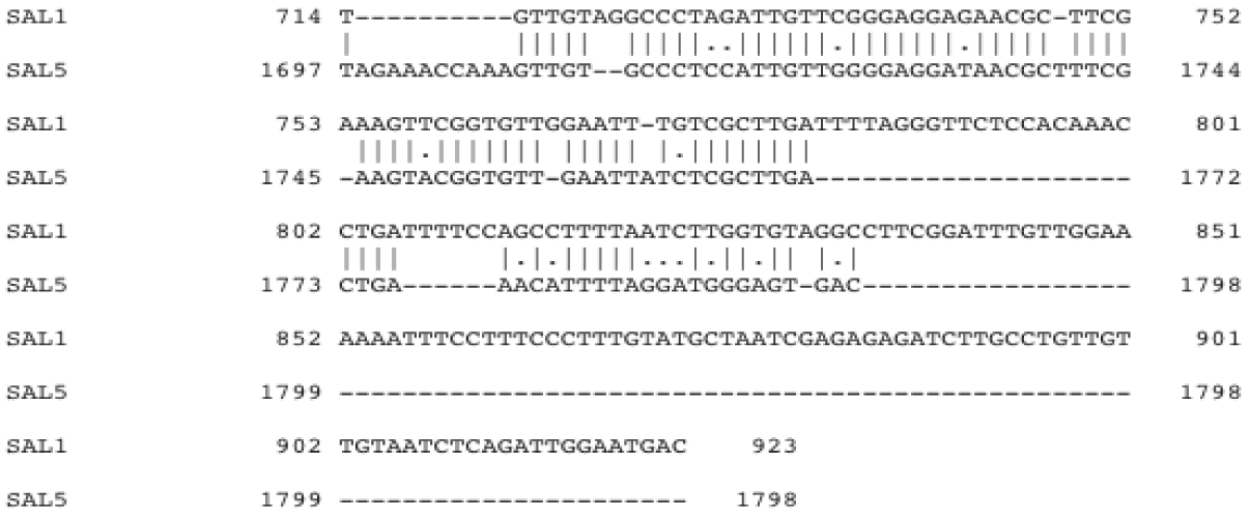


Figure 3.1 *PgSAL1* and putative *PgSAL5* promoter alignment. Promoters share an overall 30% sequence identity. Pairwise alignment was performed with EMBOSS Needle nucleotide alignment.

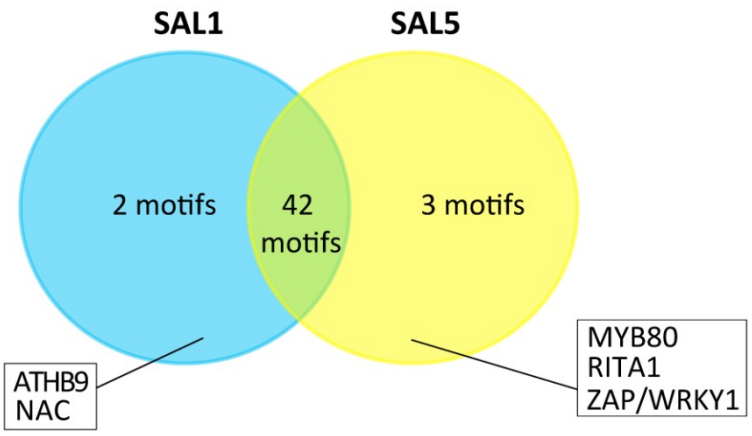


Figure 3.2 Transcription factor binding sites (TFBS) identified by rVista that are shared and distinct between the *PgSAL1* and putative *PgSAL5* promoters. The intersection of the Venn diagram contains both motifs that are identical between the two promoters (42 motifs), with unique TFBS for each promoter indicated in the non-intersecting portion. Identities of shared and distinct promoter motifs are given in Table 3.2 and Table 3.3

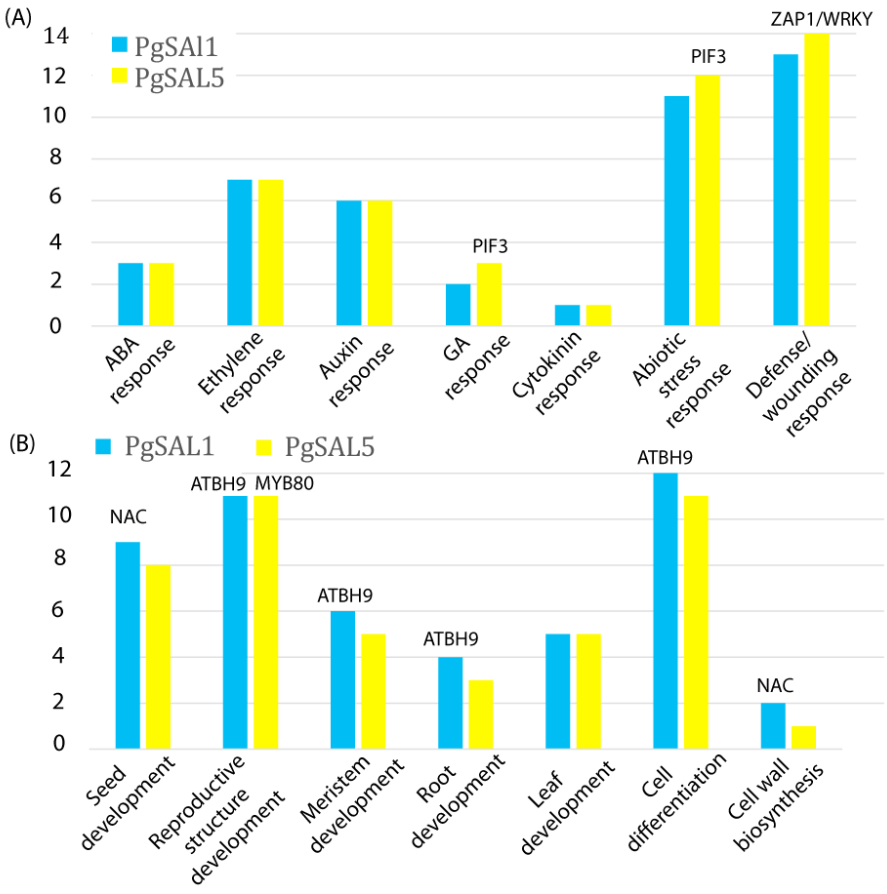


Figure 3.3 Frequency of functional category association with putative transcription factor binding sites TFBS) identified in the *PgSAL1* and putative *PgSAL5* promoters. (A) Total number of TFBS related to a hormone, abiotic or defense/wounding response pathway, identified for the *PgSAL1* or putative *PgSAL5* promoter. (B) Total number of TFBS related to a cell of structure development, identified for the *PgSAL1* or putative *PgSAL5* promoter. Most TFBS were annotated with more than one functional category, and thus multiple functional categories could be counted per TFBS.

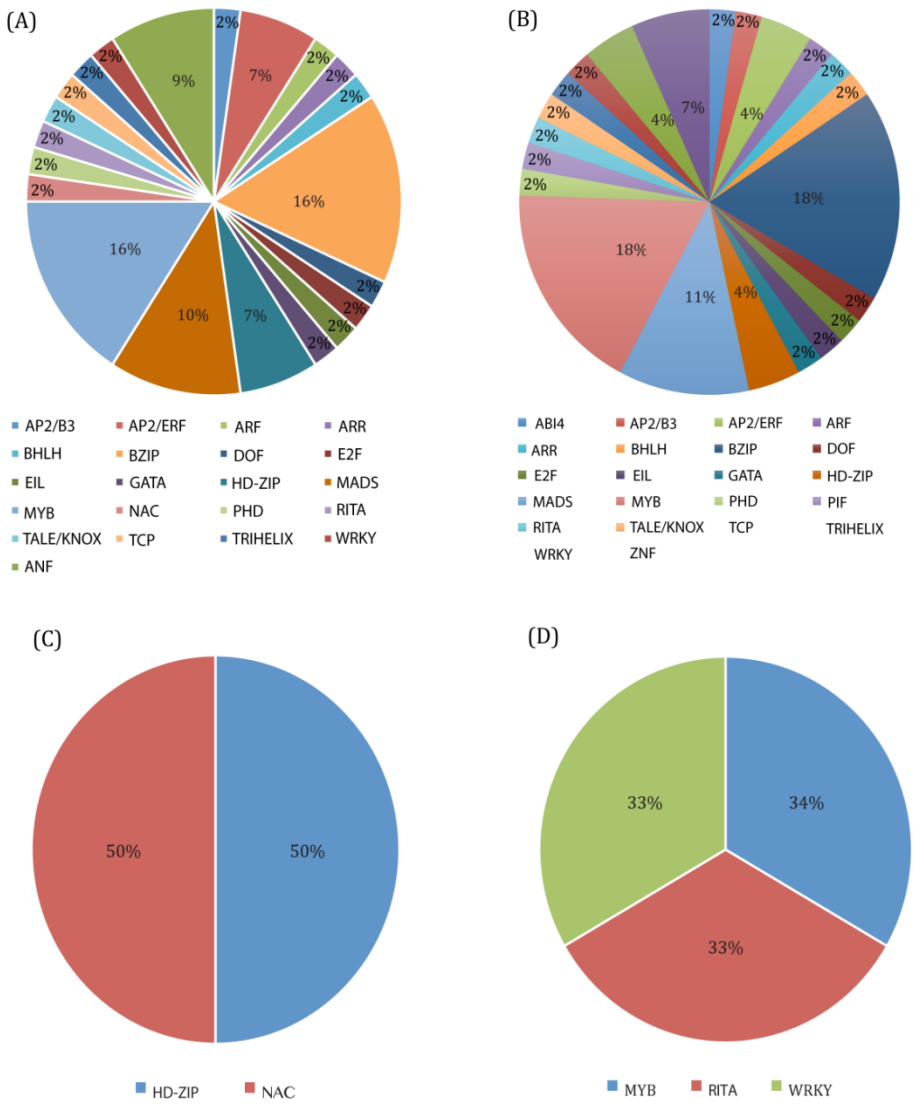


Figure 3.4 Pie charts depicting proportions of transcription factor (TF) gene families associated with putative transcription factor binding sites in the *PgSAL1* and putative *PgSAL5* promoters identified by rVista. (A) TF families putatively interacting with the *PgSAL1* promoter. (B) TF families putatively interacting with the putative *PgSAL5* promoter. (C) The set of identified TF families putatively interacting uniquely with the *PgSAL1* promoter. (D) The set of identified TF families putatively interacting uniquely with the putative *PgSAL5* promoter.

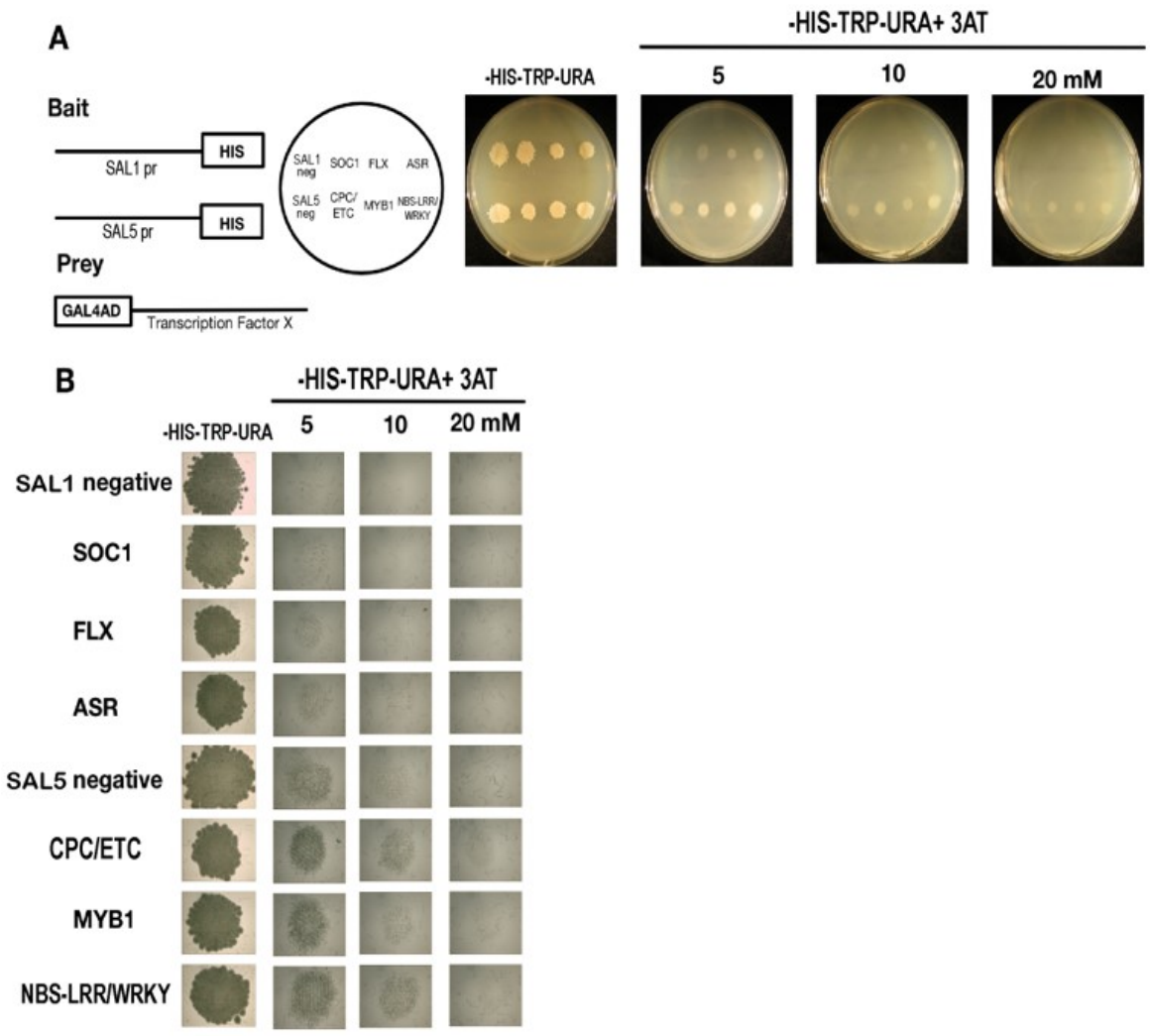


Figure 3.5 Yeast one-hybrid growth on selective 3-amino-1,2,4-triazole (3-AT) plates. (A) Whole plate photos of yeast growth on selective media with increasing concentrations of transcriptional inhibitor 3-AT. The vectors containing the promoter sequences possess the histidine reporter gene, whereas the transcription factor (TF) cDNA containing pDEST™22 vector possess the tryptophan reporter gene. The negative control yeast line contains the promoter being screened as well as the corresponding empty vector (pDEST™22), which was used for cDNA library construction. TF cDNAs chosen for 3-AT screen are white spruce: *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1*-like (*SOC1*-like), *FLOWERING LOCUS C OVEREXPRESSOR*-like (*FLX*-like), *ASCISIC ACID STRESS RIPENING*-like (*ASR*-

like), *CAPRICE/ENHANCER OF TRYPTYCHON AND CAPRICE*-like (*CPC/ETC*-like), MYB1, *NUCLEOTIDE BINDING SITE-LEUCINE RICH REPEAT/WRKY* (*NBS-LRR/WRKY*-like). The TFs were selected based on their sequence similarity to genes that likely have a promising role in the processes that may involve *PgSAL1* and putative *PgSAL5*. (B) Colony growth of corresponding colonies from whole plate under a dissecting microscope.

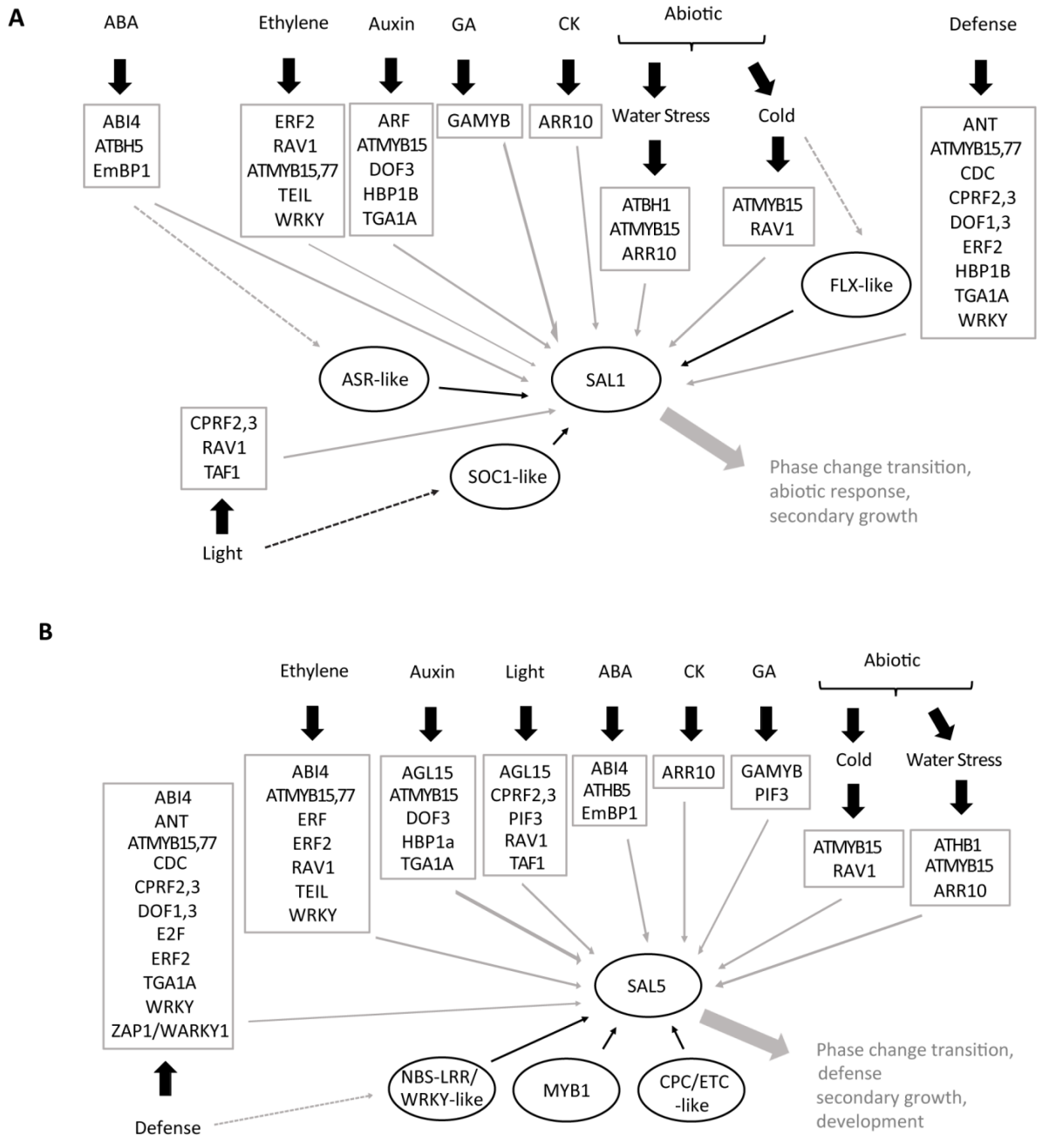


Figure 3.6 Summary figure of the *PgSAL1* and putative *PgSAL5* regulatory pathways. Based on previous transcriptional data and transcription factors identified from the yeast one-hybrid screen *PgSAL1* (A) may have roles in phase change transition, abiotic response, and/or secondary

growth, while putative *PgSAL5* (B) may have roles in phase change transition, defense, secondary growth and/or cell fate determination. Black arrows represent interactions we know exist based on rVista search and Y1H experiments identified in this paper. Grey arrows represent interactions we propose, but have not yet proven. Thin arrows represent direct interactions with *PgSAL* promoters. Thick arrows represent interactions which may be direct or indirect. Dashed lines represent interactions demonstrated in other organisms, see text for references.

Chapter 3 Tables

Table 3.1 Primers used for promoter cloning and insert verification. Promoters listed here were used for cloning promoter fragments using Genome Walker™ and the Norway spruce genome. Vector-specific primers listed here were used to confirm insert sequences by PCR and sequencing. Primer sequences are listed in the 5' to 3' orientation.

Genome Walker™		
Promoter	Primer Name	Primer Sequence
SAL1	GSP1	CTCGAGTAGTCGTACAGCTTCCCAGT
SAL1	GSP2	CGAGGGCTACATCTGCTTCACATAGA
SAL5	GSP1	AGGGCTACATCTGCTGCACATAGAAT
SAL5	GSP2	CTCCGCCTCTTCGAGAACGTCATCTG
Norway Spruce		
Promoter	Primer Name	Primer Sequence
SAL1	Fw border GSP1	AGATCATCTCAATACACCCATTTGACT
SAL1	Fw border GSP2	ACTAATAAGGGTGGGACTATAGAAA
SAL5	Fw border GSP1	ACTATCACCATTCCTTCAAAGTCCAGGAT
SAL5	Fw border GSP2	AGGCATCCAAATAATGATAGCCATAGAA
Vector-Specific		
Vector	Primer Name	Primer Sequence
476 p5E MCS	SAL1 Fw	GGGACCACCCTTTAAAGAGA
	SAL1 Rv	GGGACCACCCTTTAAAGAGA
	SAL5 Fw	GTCGACAGGCATCCAAATAATGATAGC
	SAL5 Rv	CCCGGGTGGTTTCTAACTGGTATCACC
pMW#2	M13Fw	GTAAAACGACGGCCAGT
	HIS293Rv	GGGACCACCCTTTAAAGAGA

Table 3.2 *Pg*SAL1 promoter rVista search. 923 bp of the SAL1 promoter sequence was analyzed for potential DNA binding motifs using rVista. TAIR ID (arabidopsis.org) or Uniprot entry ID (uniprot.org) is listed. Binding site name outputs from rVista are listed, with full length gene names or alternative gene names given in parentheses. Each binding has the corresponding gene family listed, along with the possible functional role of each associated transcription factor based on the description on TAIR or Uniprot. Upper case letters represent the conserved letters in the motif found in the *Pg*SAL1 promoter, while the lowercase letters are variable.

Binding site	Gene Family	Function	Motifs	Positions	Reference
ABI4 (ABA INSENSITIVE 4)	AP2/ERF	ABA response, defense response, ethylene response, root development	gttcGGTGTtg	645, 756	TAIR AT2G40220

AG (AGAMOUS)	MADS	Reproductive structure development, leaf development, cell differentiation	ctatCTATATACGGATTt, tAAACCACAATTGGacaa, ctagACAGCTTTGGTTTt, cttgCCTGTTGTTGTAAt, TTTCATTAATTAAAAA, TATCTATATACGGATT, TTGCTCTTATCGGCGG, TTGCCTGTTGTTGTAA	72, 91, 249, 250, 283, 331, 890, 891,	TAIR AT4G18960
AGL1 (AGAMOUS-LIKE 1, SHP1, SHATTERPROOF1)	MADS	Reproductive structure development	atTTCATTAATTAAAAac, taAACCACAATTGGACa, ctATCTATATACGGATtt, ctTGCCTGTTGTTGTAat	71, 91, 249, 890	TAIR AT3G58780
AGL15 (AGAMOUS-like 15)	MADS	Auxin response, GA response, seed development, reproductive structure development, light response	tTTCATTAATTAAAAa, aAACCACAATTGGACa, cTCCCAATCTTAGTTa, tATCTATATACGGATt, tTGCCTGTTGTTGTAa	72, 92, 161, 250, 891	TAIR AT5G13790
AGL2 (AGAMOUS-LIKE 2, SEP1, SEPALLATA1)	MADS	Reproductive structure development, cell differentiation	taaaCCACAATTGGacaa, actcCCAATCTTAGttat, ctatCTATATACGGattt, cttgCCTGTTGTTGTAat atttCATTAATTAAaaac,	91, 160, 249, 890	TAIR AT5G15800
AGL3 (AGAMOUS-LIKE 3, SEP4, SEPALLATA 4)	MADS	Cell differentiation, reproductive structure development	taaaCCACAATTGGacaa, caatTCAATTCTAGtgat, actcCCAATCTTAGttat, ctatCTATATACGGattt, ctagCCAAACTTATtagg, taatCTCAGATTGGaatg, aataATATTAATGTaata, cttgCCTGTTGTTGtaat,	28, 71, 91, 129, 160, 211, 249, 487, 890, 904	TAIR AT2G03710

			gtaaCTGTTATTGTggat		
AGP1 (ARABINOGALACTAN PROTEIN 1)	GATA	Cell differentiation, cell-cell recognition, seed development, programmed cell death	aTGCATCTaa, aTGGATTTct, tgACATCCAc, aGAGATCTtg	54, 201, 233, 884	UNIPROT Q8LCN5
ALFIN-like 1 (AL1)	PHD	Chromatin modification	ataaggGTTGGGact, tgaaagGGGGGGaag, AttTTCCACcctttt	5, 404, 405, 694	TAIR AT5G05610
ANT (AINTEGUMENTA)	AP2/ERF	Cell differentiation, cell proliferation, meristem development, reproductive structure development, defense response	ccttCGGATTTGTT	843	TAIR AT4G37750
ARF	ARF	Auxin response	TGGACAAa, aTTATCTC, GAGAGAAt, GATACAAt, tTTGTCGC	102, 117, 313, 394, 659, 770	TAIR AT1G19220 (many ARFs, one selected)
ARR10 (ARABIDOPSIS RESPONSE REGULATOR 10)	ARR, MYB-related	Cytokinin response, meristem development, root development, pigment biosynthesis, water stress	tgcATCT, aaaATTT, aaaAACT, caaTTCT, attATCT, AGTTatg, cgtAACT, ATATacg, AGCTttt, AGCTttg, AGAGatt, TGATttt, CGATttt, AGATtgt, TGATttt, caaACCT, cggATTT, gagATCT, AGATtgg	55, 62, 83, 117, 122, 134, 164, 72, 210, 255, 260, 278, 337, 372, 475, 520, 668, 692, 727, 779, 797, 803, 838, 852, 885, 911	TAIR AT4G31920

ATHB1 (ARABIDOPSIS THAIANA HOMEOBOX 1)	HD-ZIP	Leaf development, light response, water stress	tgTAATAATTAttt, tcTAAAAATTTgat, ttCATTAATTAaaa, ataCAATCATGAaa, tttTCATTATTAag	39, 59, 73, 395, 441	TAIR AT3G01470
ATHB5 (ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 5)	HD-ZIP	ABA response	tAATAAGGG, aAATAATAA, tAATAATTA, aAAAATTTG, cATTAATTA, cCACAATTG, TGATTATTg, cAATCTTAG, CTGTTATTg, cAATCATGA, TCATTATTa, aAAAAATTT	3,24, 41, 62, 75, 95, 143, 165, 215, 398, 444, 850	TAIR AT5G65310
ATHB9 (ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 9, PHV, PHAVOLUTA)	HD-ZIP	Leaf development, cell differentiation, seed development, reproductive structure development, meristem development	taaTGTAATAATTATTtta, tctAGTGATTATTGCTagg	36, 138	TAIR T1G30490
ATMYB15 (ARABIDOPSIS THALIANA MYB DOMAIN PROTEIN 15)	MYB	Cell differentiation, auxin response, cadmium response, defense response, cold response, ethylene response, water stress	CGAAGTTCGGTGT	641	TAIR AT3G23250

ATMYB77 (ARABIDOPSIS THALIANA MYB DOMAIN PROTEIN 77)	MYB	Cell differentiation, root development, ethylene response, defense response	taatAAGGGTTgg, aaacCACAATTgg, acgtAACTGTTat, gtAACTGTTattg	3, 92, 209	TAIR AT3G50060
BHLH66 (BASIC HELIX LOOP HELIX 66)	BHLH	Root development	CCACGTAA	207	UNIPROT Q9ZUG9
C1 (COLOURLESS 1)	MYB	Pigment biosynthesis	ataaggGTTgg, taAACcacaat, aaAACtcccaa, ccAACgactac, gccgaaGTTgt, cgggagGATaa, tccgaaGTTcg, gtgttgGTTtt, ccACCaatccg, tagattGTTcg, tcgaaaGTTcg, ttcggtGTTgg, ggatttGTTgg, cctgttGTTgt	91, 157, 421, 600, 624, 639, 650, 683, 726, 750, 839, 894	UniPROT P10290
CDC5 (CELL DIVISION CYCLE 5)	MYB	Cell differentiation, defense response	ccCGCTGTAct, ccCGCTGCAct, ctCGATGAGct	456, 506, 551, 582	TAIR AT1G09770
CPRF2 (COMMON PLANT RGULATOR FACTOR 2, BZIP17)	BZIP	Light response, defense response, pigment biosynthesis	tcCACGTAac	206	UniPROT Q99090
CPRF3 (COMMON PLANT RGULATOR FACTOR 3)	BZIP	Light response, defense response	atGACTTCtt, tcCACGTAac	176, 206	UniPROT Q99091
DOF1 (DOF	ZNF	Defense response	atGACTTCtt, tcCACGTAac	176, 206	TAIR

ZINC FINGER PROTEIN 1)					AT1G51700
DOF2 (DOF ZINC FINGER PROTEIN 2)	ZNF	Seed development	gactATAGaaa, tttCATTaatt, attaAAAActa, cttCTTTtata, taaCTGTtatt, cagCTTTtgct, ttgCTTTtcat, tattAAGGccc, aaaCTTAtgtt, tggTTTTtgtc, tcgCTTGattt, accCTTTtaat, agcCTTTtaat	16, 72, 80, 180, 212 277, 437, 448, 522, 654, 663, 701, 812	UniPROT B9F1L8
DOF3 (DOF ZINC FINGER PROTEIN 3)	ZNF	Cell wall modification, cell differentiation, auxin response, defense response	ttaTTTTatgc, gacaAAATtca, tgaCTTCtttt, cagCTTTtgct, ttgCTCTtatc, cagCTTTggtt, ttgGTTTtatc, catgAAAGggg, acgCTATttgc, ttgCTTTtcat, gaccAAAGttt, ttgGTTTttgt, tcgCTTGattt, accCTTTtaat, tcgCTTGattt, agcCTTTtaat, ttcCTTTccct	47, 104, 177, 277, 283, 336, 341, 402, 430, 437, 570, 653, 663, 701, 774, 812, 857	UniPROT Q39088
E2F	E2F	Cell division, cell development, glucosinolate metabolism	ggGGGGGAAgg, ttTTTGTCGct, atTTTCCACcc	409, 657, 694	TAIR AT1G47870 (many E2Fs, one selected)

EmBP1 (EARLY METHIONINE BINDING PROTEIN-1)	BZIP	ABA response	tcCACGTAAC	206	UniPROT P25032
ERF2 (ETHYLENE RESPONSE FACTOR- 2)	AP2/ERF	Ethylene response, cell division, defense response	ATCGGCG, GGCGGAA, CGCGGCC	291, 294, 302, 349, 352, 360	TAIR AT5G47220
GAMYB	MYB	Cell differentiation, reproductive structure development, GA response, amylase metabolism	aaggGTTg, tAAAccac, cttaGTTa, gtggATTg, tgccGTTt, tttgGTTt, cAACgact, ctcgATTg, gttcGGTg, gttgGTTt, gtcgCTTg, tcttGTTg, gttcGGTg, gtcgCTTg, atttGTTg, gcctGTTg	7, 91, 169, 223, 268, 340, 422, 645, 652, 662, 711, 756, 773, 841, 893,	UniPROT Q0JIC2

GT1 (GRASSY TILLERS 1)	TRIHELIX	Reproductive structure development, meristem development	CTAATAA, TAATAAG, ATAGAAA, GAAATAA, TATTAAT, GTAATAA, ATAATTA, TAATTAT, TTTATGC, CTAAAAA, TTAATTA, TAATTAA, ATTAAAA, TTAAAAA, TAAAAAC, CTATAAA, TATAAAC, TTATCTC, GTGATTA, TTATTGC, GTTATGA, TTATGAC, TTTATAT, CATCCAC, GTTATTG, GTGGATT, TTCTGAC, TATCTAT, CTATATA, TATATAC, TTTCTGC, GTTTACA, TTTACAG, TTTTATC, TTTTGAG, TATTTGC, TTTTCAT, TTATTAA, TATTAAG, GCAAAAA, CTTATTA, TTATTAG, GCAAAAA, CTTATTA, TTTCCTC, GTCAGAA, GAGGATA, TTTCCAC, TTTAAT, TCTCCAC, TTTCCAG, TTTTAAT, TTGGAAA, GGAAAAA, GAAAAAA, GTTGTAA, TAATCTC	2, 3, 20, 23, 26, 33, 40, 43, 44, 51, 60, 77, 78, 80, 81, 82, 88, 89, 118, 142, 146, 173, 174, 185, 204, 217, 223, 239, 250, 253, 254, 264, 272, 273, 345, 377, 434, 441, 447, 448, 472, 496, 497, 517, 541, 578, 593, 627, 680, 696, 705, 791, 807, 816, 846, 848, 849, 900, 904	UniPROT G1AQA5
HBP1a (HISTONE BINDING PROTEIN 1a)	BZIP	Histone modification	TCCACGTAac	206	UniPROT P23922

HBP1B (HISTONE BINDING PROTEIN 1B)	BZIP	Histone modification, auxin response, defense response	tcCACGTAAC, gTGGTGTGACatcc	195, 631	UniPROT P23923
KNOX3 (KNOTTED-1-LIKE 3)	TALE/KNOX	Meristem development	tggtGTGAcatc, cacgTAACtgtt, tttcTGACactc, tcgaGTCAgcct, gcctAGACagct, ttgaGTCAccct, agctGTCAgaag, ttttGTCGcttg	196, 208, 238, 321, 329, 379, 589, 658, 769	UniPROT Q43484
MYBAS1	MYB	Cell differentiation	ggaaAACtccc, acgtAACtgtt, aactGTTattg, cgttTACagct, agctTTTgctc, gcccAACgact, gccaAACttat, gccaAACttat, cgaaGTTgtgc, ggatAACgctt, cgaaGTTcggt, cggtGTTggtt, gattGTTcggg, ggagAACgctt, gaaaGTTcggt, ggaaTTTgtcg, atttGTTggaa, gcctGTTgttg	155, 209, 213, 271, 278, 419, 490, 535, 602, 629, 641, 648, 728, 740, 752, 766, 841, 893	UniPROT Q53NK6
NAC (NAM, ATAF1/2, CUC2)	NAC	Cell wall biogenesis, seed development	aGTGGTGTGACATCCACGTAACt	194	TAIR AT1G12260 (many NACs, one selected)
O2 (OPAQUE 2)	BZIP	Seed development	tcCACGTaac, tTATCTCATC, GATGAGCTGt, aATTTGATTTCAt, tTTATATCGAGTg, cTGTTGTTGTAAt	206, 118, 585, 65, 185, 895	UniPROT P12959
P	MYB	Pigment biosynthesis	ggTGTTGGt, ggTGTTGGt, aCCAATCcg, gtTGTAGGc, ggTGTTGGa, ggTGTAGGc	8, 649, 685, 715, 760, 796, 826	UniPROT P27898

PBF (PYRIMIDINE-BOX BINDING FACTOR)	DOF	Seed development	tattAATGtaa, tttCATTaatt, attaAAAActa, tgaCTTCtttt, cttCTTTtata, taaCTGTtatt, cagCTTTtgct, ttgCTCTtatc, cagCTTTggtt, ttgGTTTtatc, catgAAAGggg, ttgCTTTtcat, tattAAGGccc, ccgCTGTactg, ttgGTTTttgt, tggTTTTtgtc, tcgCTTGattt, accCTTTtaat, tcgCTTGattt, agcCTTTtaat, tccCTTTgtat	33, 72, 80, 177, 180, 212, 277, 283, 341, 402, 437, 448, 457, 552, 654, 663, 701, 774, 812, 863	UniPROT O24463
PCF2 (PROLIFERATING CELL FACTOR 2)	TCP	Meristem development	aaGGCCCAAC, taAGGCCCGC, ttGTGCCCTC, GTAGGCCCta	416, 451, 501, 546, 607, 718	UniPROT Q6ZBH6
RAV1 (RELATED TO ABI1/VP1)	AP2/B3	Cold response, ethylene response, brassinosteroid response, light response, root development, leaf development, reproductive structure development	aaACCACAATtg, taACTGTTATtg, ttATTGTGGAtt, ttTCTGCCGTtt, tcGGTGTTGGtt, ttTTTGTCGCtt, ttAATCTTGTtg, atCTTGTTGTag, tcGGTGTTGGaa, aaTTTGTCGCtt, ttAATCTTGGtg, gaTTTGTTGGaa, tgCCTGTTGTtg, ctGTTGTTGTaa, caAACCTGattt, atctCAGATTgg	92, 212, 218, 264, 657, 707, 710, 758, 768, 818, 840, 892, 895, 797, 906	TAIR AT1G13260
RITA1	RITA	Seed development	tgACATc, ccACGTa, tATGTct	177, 201, 207, 482	UniPROT Q6ETX0

TAF1 (TBP-ASSOCIATED FACTOR 1, HAF2, HISTONE ACETYLTRANSFERASE OF THE TAF11250 FAMILY 2)	BZIP	Histone modification, light response	tcCACGTaac	206	TAIR AT3G19040
TEIL (TOBACCO ETHYLENE-INSENSITIVE 3)	EIL	Ethylene response	ATTAATGT, ATGCATCT, AAATTCAA, ACGTAACT, AGATGGAT, ATGGATTT, ACGGATTT, CGATACAA, ACAATCAT, AAGTTCGG, ACAAACCT	34, 108, 209, 231, 233, 259, 393, 397, 643, 754, 796	UniPROT Q9ZWK1
TGA1A	BZIP	Auxin response, defense response, histone modification	tccACGTaac	206	UniPROT P14232
WRKY	WRKY	Defense response, ethylene response	ATTTTGAGTCA, ATGTTTAGCCA	376, 528	TAIR AT1G13960 (many WRKYs, one selected)

Table 3.3 Putative *PgSAL5* promoter rVista search. 1798 bp of the putative *SAL5* promoter sequence was analyzed for potential DNA binding motifs using rVista. TAIR ID (arabidopsis.org) or Uniprot entry ID (uniprot.org) is listed. Binding site name output from rVista listed, with full length gene names or alternative gene names in parentheses. Each binding has the corresponding gene family listed, along with the possible functional role of each associated transcription factor based on the description on TAIR or Uniprot. Upper case letters represent the conserved letters in the motif found in the putative *PgSAL5* promoter, while the lowercase letters are variable.

Binding site	Gene Family	Function	Motifs	Positions	Reference
ABI4 (ABA INSENSITIVE 4)	ERF/AP2	ABA response, defense response, ethylene response, root development	ttGCACCagcc, gtacGGTGTtg	450, 1747	TAIR AT2G40220
AG (AGAMOUS)	MADS	Reproductive structure development, leaf development, cell differentiation	gcatCCAAATAATGATAg, agtcCCTGTAATAGAAAa, aattCCAAATATTGCCAg, tctaCTAAGTTTGGCAAa, cTAACCATGTTAAAaaat, attaAATAAAAAGGAAAa, tCTTCCAATGAAAGaaaa, tTTCCCACAAGAAGaaac, ccatCCCATTATGCCAAa, aATAACAAAATAATaaaa, tctcTCCCATGAGGAAAt	5, 38, 58, 73, 206, 611, 767, 972, 1093, 1205, 1241, 1299, 1308, 1418, 1586	TAIR AT4G18960

AGL1 (AGAMOUS-LIKE 1, SHP1, SHATTERPROOF1)	MADS	Reproductive structure development	agTCCCTGTAATAGAAaa, aaTTCCAAATATTGCCag, tcTACTAAGTTTGGCAaa, ttGACTTGATTTTGATag, atTAAATAAAAAGGAAaa, atTAAATAAAAAGGAAaa, tcTTCCAATGAAAGAAaa	38, 73, 206, 356, 767, 972, 1093, 1205, 1241, 1299, 1307, 1586	TAIR AT3G58780
AGL15 (AGAMOUS-like 15)	MADS	Auxin response, GA response, seed development, reproductive structure development, light response	cATCCAAATAATGATa, gTCCCTGTAATAGAAa, tTACATGGAAAAGTAa, aTTCCAAATATTGCCa, tTAATATATTTAATTa, cTACTAAGTTTGGCAa, aTATTTTTCAAAGAAa, tAACCATGTTAAAAAa, aTATTAAATAAAAAAa, aTTAAATAAAAAGGAa, cTTCCAATGAAAGAAa, tTCCCACAAGAAGAAa, gTTCAAATTTGAGCAa, tTGCTCATTATGCCAg, aTGCCAGCAAAAGTAa, tTACTATGTGTAGTGc, tAACCATTAATATATt, aTAATAAAAAAATAAa	6, 39, 58, 74, 95, 207, 301, 612, 767, 973, 1094, 1241, 199, 1308, 1353, 1376, 1404	TAIR AT5G13790
AGL2 (AGAMOUS-LIKE 2, SEP1, SEPALLATA1)	MADS	Reproductive structure development, cell differentiation	gcatCCAAATAATGATag, agtcCCTGTAATAGAAaa, aattCCAAATATTGccag, tctaCTAAGTTTGGcaaa, ccatCCAAAAATATtttt, ccagCCAATTATGAattt, tcttCCAATGAAAGaaaa, tttcCCACAAGAAGaaac,	5, 38, 73, 206, 291, 455, 972, 1093, 1307, 1375	TAIR AT5G15800

			tatgCCAGCAAAAGtaat, ttaaCCATTAATATatta, tgttCAAATTTGAGCAaa		
AGL3 (AGAMOUS-LIKE 3, SEP4, SEPALLATA 4)	MADS	Cell differentiation, reproductive structure development	agtcCCTGTAATAGaaaa, aattCCAAATATTGccag, tttaATATATTTAAttag, atatATTTAATTAGactc, tctaCTAAGTTTGGcaaa, ccatCCAAAAATATtttt, ccagCCAATTATGAattt, ctaaCCATGTTAAAaaat, caaaTCATATTTGAgaag, tcttCCAATGAAAGaaaa, tttcCCACAAGAAGaaac, tgttCAAATTTGAGcaaa, tatgCCAGCAAAAGtaat, tttaCTATGTGTAGtgcc, ttaaCCATTAATATatta, ccctCCATTGTTGGggag, tgaaACATTTTAGGatgg	5, 38, 73, 94, 206, 291, 395, 455, 611, 839, 972, 1093, 1240, 1307, 1352, 1375, 1418	TAIR AT2G03710
AGP1 (ARABINOGALACTAN PROTEIN 1)	GATA	Cell differentiation, cell-cell recognition, seed development, programmed cell death	tcAAATCCAa, agAGATATAa, tTAGATATat	145, 577, 745	UniPROT Q8LCN5
ALFIN-like 1 (AL1)	PHD	Chromatin modification	aaacaaGTGTGGctt, agaggaGAGGAGctt, aagaaaGTGTCActg, agcTAACACctatct,	731, 857, 1227, 1532, 1574	TAIR AT5G05610

			accTAACACctctct		
ANT (AINTEGUMENTA)	AP2/ERF	Cell differentiation, cell proliferation, meristem development, reproductive structure development, defense response	CACCTCTTCCaatg	968	TAIR AT4G37750
ARF	ARF	Auxin response	aTTTTCTC, tTTCTCAC, GTGACGAa, AAGACAAa, GAAACAAg, tTTGTCAC, GATACAAc, aGTGTCAC, ATGACAAa, aTTTTCTC	425, 427, 521, 797, 1106, 1123, 1191, 1232, 1267, 1675	TAIR AT1G19220 (many ARFs, one selected)
ARR10 (ARABIDOPSIS RESPONSE REGULATOR 10)	ARR, MYB-related	Cytokinin response, meristem development, root development, pigment biosynthesis, water stress	AAATatt, AGACatt, aatATAT, TGATtcg, caaATCC, aggATAT, AGAGatt, aggATGT caaATAT, aatATTT, TGATttt, cgaAGCT, ATATatt, AGGTttg, aagATTT, caaATCA, AGATata, aaaATAT, caaAACT, AGATata, ATATatt, aaaATCA, aggAGCT, caaATCA, catATTT, aggAGCT, agaATAT, ACATact, agaATTT, AGACatt, caaATAT, caaATGT, caaATTT, ACATatg, aatATCA, catATTT, aatATAT, AAATact, AGAAttg, cctATCT, AAATccg, CGATttt	79, 89, 97, 129, 146, 191, 223, 231, 264, 267, 298, 300, 162, 182, 401, 446, 497, 540, 579, 624, 702, 747, 749, 802, 831, 839, 844, 864, 878, 904, 1000, 1009, 1075, 1139, 1244, 1288, 1370, 1384, 1458, 1474, 1540, 1600, 1673	TAIR AT4G31920

ATHB1 (ARABIDOPSIS THAIANA HOMEOBOX 1)	HD-ZIP	Leaf development, light response, water stress	ttcCAAATATTGcc, tgTTATAACTGcat, ataCAAATATTTta, ttgTGATTATTAgg, ttaAAATTATTGca, tatCAAATATTAcg, attTAATTCTTTca, gttAAATTGTTGta	75, 235, 261, 406, 471, 1072, 1146, 1275	TAIR AT3G01470
ATHB5 (ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 5)	HD-ZIP	ABA response	CAAATAATg, TAATGATAg, aAAAAATTA, cAAATATTG, AAAATATTt, gAATATTGG, CATATATTg, cAATTATGA, aAATTATTG, aAATCAATG, CAATGAGTt, aAAAAATTA, gAATATTTC, cAATGAAAG, cAAATATTA, TAATTCTTt, CAACCATTt, CCATTATGc, aAATTGTTG, CCATTAATa, aAATAATAA, tAAGAATTG	10, 14, 52, 78, 264, 298, 333, 400, 460, 474, 642, 646, 762, 879, 1075,1149, 1195, 1210, 1278, 1379, 1399, 1413, 1425, 1472	TAIR AT5G65310
ATMYB15 (ARABIDOPSIS THALIANA MYB DOMAIN PROTEIN 15)	MYB	Cell differentiation, auxin response, cadmium response, defense response, cold response, ethylene response, water stress	CCATTTACCATCC, ACACCTATCTCTG	1198, 1537	TAIR AT3G23250
ATMYB77 (ARABIDOPSIS THALIANA MYB DOMAIN PROTEIN 77)	MYB	Cell differentiation, root development, ethylene response, defense response	ctAGCAGTCcctg, atAACTGCAtatt, gaAAGTGTCactg, atatTACAGTTaa, tgatACCAGTTag	33, 239, 1229, 1387, 1687	TAIR AT3G50060

BHLH66 (BASIC HELIX LOOP HELIX 66)	BHLH	Root development	cCATTTACCATCc, gTACCTAACACCt	1198, 172	UniPROT Q9ZUG9
C1 (COLOURLESS 1)	MYB	Pigment biosynthesis	aaAACtagcag, ccAACaacaat, cgAACaaatac, ttgataGTGag, tgagagGTTtg, acAACtacaaa, ctAACcatgtt, caatgaGTTtg, tcAATtaacag, ccAAGcaacaa, gcAACaactag, caAACtcttaa, gacaaaGTTaa, ttAACcattaa, ccAACcttaaa	29, 151, 254, 367, 442, 533, 611, 646, 806, 1054, 1058, 1158, 1269, 1375, 1550	UniPROT P10290
CDC5 (CELL DIVISION CYCLE 5)	MYB	Cell differentiation, defense response	agCTCGGGGct, tgCTCATAGca, aaCGCTAACca	385, 504, 607	TAIR AT1G09770
CPRF2 (COMMON PLANT RGULATOR FACTOR 2, BZIP17)	BZIP	Light response, defense response	atTACATGga, gtCACTTGat, gaTTCGTGtt, caCAAGTGac, ttGACTTGat, gaGATGTCga, aaCAAGTGtg, ctCACATGaa, ttCACCTCtt, atTACGTAgt, caCAGGTCga, acTATGTGta, ccCATGTGcc, aaCACCTAtc, aaCACCTCtc, gtGACATGgc, ggCCCGTGag	57, 123, 130, 179, 356, 375, 732, 821, 966, 1080, 1128, 1355, 1504, 1536, 1578	UniPROT Q99090
CPRF3 (COMMON PLANT RGULATOR FACTOR 3)	BZIP	Light response, defense response	atTACATGga, gtCACTTGat, caCAAGTGac, ttGACTTGat, gaGATGTCga, aaCAAGTGtg, ctCACATGaa, ttCACCTCtt, caCAGGTCga, aaCACCTCtc, gtGACATGgc	57, 123, 130, 179, 356, 375, 732, 821, 966, 1080, 1128, 1355, 1504, 1536, 1578	UniPROT Q99091

DOF1 (DOF ZINC FINGER PROTEIN 1)	ZNF	Defense response	aaatAATGata, tggaAAAGtaa, aagtAAATtcc, tatATTTaatt, aatATTTtatt, tgaCTTGattt, ttgATTTtgat, ggtaAAAGgaa, aaatAACGcta, aatcAAAGaga, aaaaAAAGcgg, aagaAAAAaaa, aaaaAAAGtcc, aaaaAAAAagc, agcaAAAAaaa, tggCTTTagat, tattAAATaaa, aattAAATaaa, ataaAAAGgaa, aaatAAAAaag, gacaAAATcaa, gagCTTCaatg, actgAAAGaaa, actgAAAGaaa, aatgAAAGaaa, aagaAAAGtat, gtatAAAGata, atttAAAGaca, aataAAAAagt, tgtATTTaatt, aattCAAGaaa, agcaAAAGtaa, agttAAAAtaa, gcaCTTAtatt, aataAAAAaat, aaatAAATaac	11, 63, 68, 99, 266, 357, 361, 593, 603, 632, 671, 682, 688, 714, 722, 740, 752, 766, 772, 789, 799, 866, 891, 931, 978, 983, 990, 1003, 1023, 1143, 1222, 1313, 1394, 1406, 1413, 1421, 1429, 1434, 1606, 1625	TAIR AT1G51700
DOF2 (DOF ZINC FINGER PROTEIN 2)	ZNF	Seed development	tggaAAAGtaa, tggcAAAGaga, cttcAAAGtcc, tttcAAAGaaa, tgaCTTGattt, ggtaAAAGgaa, aaatAACGcta, aatcAAAGaga, aaaaAAAGcgg, aaaaAAAGtcc, caaaAAAAaaa, aaaaAAAAcaa, tggCTTTagat, tattAAATaaa, aattAAATaaa, ataaAAAGgaa, aaatAAAAaag, catgAAAGgag, gagCTTCaatg, acaCTTTacca, actgAAAGaaa, aatgAAAGaaa, aagaAAAGtat, gtatAAAGata, atttAAAGaca, taaaAAAGtac, tggCTTTgtca, attCTTTcaaa, aattCAAGaaa, agcaAAAGtaa,	63, 217, 282, 306, 357, 361, 593, 603, 632, 671, 682, 688, 712, 726, 740, 752, 766, 772, 789, 792, 825, 866, 891, 909, 932, 978, 983, 990, 1003, 1025, 119, 1151, 1222, 1268, 1313, 1334, 1394, 1405, 1409, 1413, 1429, 1433, 1441, 1606, 1625	UniPROT B9F1L8

			agttAAAAtaa, taatAAAAaaa, aaatAAATaac, aataAAAAaaa, aaaaAAAAtaa, taacAAAGtta, ccgCATTttat		
DOF3 (DOF ZINC FINGER PROTEIN 3)	ZNF	Cell wall modification, cell differentiation, auxin response, defense response	gaaaAAATtac, tggaAAAGtaa, tcaCTTGattc, tggcAAAGaga, cttcAAAGtcc, ggtaAAAGgaa, aatcAAAGaga, aaaaAAAGcgg, gaaaAAAAaaa, aaaaAAAGtcc, aaaaAAAGcaa, aaaaAAAAcaa, tggCTTTagat, aaaaAAATtaa, ataaAAAGgaa, taaaAAAGaca, gacaAAATcaa, catgAAAGgag, acaCTTTacca, actgAAAGaaa, aatgAAAGaaa, gtatAAAGata, atttAAAGaca, taaaAAAGtac, gaaaCAAGaaa, tggCTTTgtca, gccaAAATtca, caagAAAGtgt, tgacAAAGtta, agcaAAAGtaa, aaaaAAAAtaa, taacAAAGtta, gaggAAATccg, acgATTTctc	51, 63, 124, 217, 282, 593, 632, 671, 684, 688, 716, 726, 740, 761, 773, 778, 792, 799, 825, 891, 932, 937, 978, 983, 990, 1003, 1025, 1106, 1119, 1217, 1226, 1268, 1313, 1433, 1441, 1596, 1672	UniPROT Q39088
E2F	E2F	Cell division, cell development, defense response	agCGGAAGAaa, tcCGCAAAAaa, atTTTCCCAca, atATTCCCTct, gcCAGCAAAag	677, 708, 1091, 1168, 1310	TAIR AT1G47870 (many E2Fs, one selected)
EmBP1 (EARLY METHIONINE BINDING PROTEIN-1)	BZIP	ABA response	attACATGga, ttCACCTctt, attACGTAgt, caCAAGTGAC, gaAACCTGGC, caCAAGTTGC, GCATCGTGgg, ccCATGTGCC, taAACGCGTC	57, 966, 1080, 179, 1113, 1293, 1481, 1504, 1557	UniPROT P25032

ERF2 (ETHYLENE RESPONSE FACTOR- 2)	AP2/ERF	Ethylene response, cell division, defense response	GGCATCG, CTCTGCC	1480, 1545	TAIR AT5G47220
GAMYB	MYB	Cell differentiation, reproductive structure development, GA response, amylase metabolism	gtcaCTTg, cAACaaca, cAAGtgac, ggatGTTa, tAACtgca, gtcaATTg, cAACggct, cACCagcc, cAACctaa, cAACtaca, cAACacag, gtggGTAa, tAACgcta, tAACcatg, gtccGTTc, taggGTTa, tAACagat, cAAGcaac, cAACaggg, cAACcatt, tgctGTTt, tAACcatt, tacaGTTa, tAACaaaa, tAACaaag, cAACctta, ggcaCTTa, aAACcaaa	123, 152, 181, 232, 240, 351, 434, 453, 491, 534, 569, 590, 606, 612, 695, 811, 1039, 1055, 1184, 1195, 1335, 1376, 1391, 1420, 1441, 1551, 1624, 1700	UniPROT Q0JIC2
GT1 (GRASSY TILLERS 1)	TRIHELIX	Reproductive structure development, meristem development	ATAGAAA, GTAATAG, GAAAAAA, TATTGCC, TTTAATA, GTGTTTG, GTATTCA, GAAATTA, TAATCAC, GGATATA, TTTCTAC, GAGATTA, GTTATAA, TATTATC, CAAATAC, TTTTCAA, GTGACTA, GGGAATA, CTGAATA, TTTTCTC, TTTGCAC, TTTAAAA, TTATTGC, CATCAAC, CAACTAC, TAATCAA, GTGGGTA, GAAAATA, GTTAAAA, TATTAAT,	25, 45, 51, 55, 62, 70, 82, 94, 101, 105, 135, 141, 162, 170, 175, 192, 204, 224, 236, 237, 248, 258, 273, 305320, 331, 340, 351, 408, 426, 449, 470, 488, 534, 554, 590, 601, 619, 628, 667, 681, 699, 711, 723, 745, 752, 768, 772, 785,	UniPROT G1AQA5

			GTTCAAA, GAAGAAA, GTTCAAA, GCAAAAA, TTAGATA, ATAAAAA, TTAAATA, GTACAAA, AATTAAC, GAGAATA, TTTAAAG, GAATAAA, GTTATCA, TATTATC, ATATTAC, TTTTCCC, GAAGAAA, TTGTCAC, GTATTTA, CTTAATA, CATTTAC, TAATGAC, GTTAAAT, ATATCAC, CTCATTA, GTAATCA, CTGTTTA, GTTTTTA, TATTTAA, ATTAATA, ATATTAC, ATAAAAA, ATAATAA, GTTAATC, GTGAAAT, TATACGC, CTTAAAC, GAGGAAA, TTTTCAT, GTTCTTA, TTTTCTC, GTGATAC	791, 808, 877, 883, 964, 990, 1004, 1022, 1043, 1051, 1069, 1071, 1078, 1085, 1092, 1103, 1124, 1144, 1164, 1199, 1241, 1265, 1275, 1283, 1289, 1302, 1320, 1337, 1349, 1372, 1381, 1395, 1401, 1407, 1415, 1527, 1430, 1448, 1455, 1464, 1555, 1596, 1615, 1660, 1663, 1676, 1677, 1686, 1695	
HBP1a (HISTONE BINDING PROTEIN 1a)	BZIP	Histone modification	GTCACTTGat, gaAACCTGGC, caCAAGTGAC, CCCATGTGcc	123, 179, 1113, 1504	UniPROT P23922
HBP1B (HISTONE BINDING PROTEIN 1B)	BZIP	Histone modification, auxin response, defense response	gTGACTTAGGatat	184	UniPROT P23923

KNOX3 (KNOTTED-1-LIKE 3)	TALE/KNOX	Meristem development	ccctGTAAtaga, tgccAGACattt, ctaaGTCActtg, caagTGACttag, gataTAACaatt, ttagGTCAattg, agatGTCGaagc, ggctTGAGaggt, ccagTGACgaag, caatTAACagat, tgacTCACatga, atatTTCAcaac, taccATCAcata, gcacTGAAagaa, aaccTGACattc, ctttGTCAcagg, aagtGTCActgt, ctaaTGACaaag, tataTTACagtt	41, 85, 119, 181, 193, 347, 353, 376, 438, 518, 807, 818, 823, 881, 897, 930, 957, 1122, 1231, 1264, 1386, 1506, 1512, 1531, 1573	UniPROT Q43484
MYB80	MYB	Reproductive structure development	aGTAAATTCca, agGAATATAAt, aGGATATAAca, tGCATATTAtc, ggGAATATTAg, gaGAATATTTc, taGAATAACCt, gTCATTTTCcc, tTAATATTCcc, aGGAAATCCgc, gGCATGTTCtt	69, 168, 191, 244, 340, 877, 951, 1088, 1165, 1597, 1655	TAIR AT5G56110
MYBAS1	MYB	Cell differentiation	atccAACaaca, atatAACaatt, ggatGTTataa, ttacAACttca, tcacAACggct, agagGTTtgca, catcAACctaa, gcacAACtaca, ataaGTTgtgg, aaatAACgcta, aattAACagat, tcacAACactt, gaatAACctga, taggGTTatca, aagcAACaact, atacAACcatt, aattGTTgtaa, acaaGTTgctc, taggAACgttt, atttAACcatt,	149, 194, 232, 276, 431, 444, 488, 531, 583, 603, 808, 886, 953, 1039, 1056, 1192, 1279, 1294, 1342, 1373, 1391, 1532, 1548, 1574, 1624, 1691	UniPROT Q53NK6

			tacaGTTaaaa, agctAACacct, tgccAACctta, acctAACacct, ggcaCTTatat, accaGTTagaa		
O2 (OPAQUE 2)	BZIP	Seed development	attACATGga, tctAAGTCac, gtcACTTGat, attAGGTCaa, ttgACTTGat, gagATGTCga, ctcACATGaa, tgCACATaga, ttCACCTctt, attACGTAgt, caCAGGTcga, actATGTGta, gtgACATGgc, tctATGTGca, AATTACATGg, gCATATTATC, cCAAATCATA, AGTGACATGg, cAGTCCCTGTAAt, cATTTAATATATt, aTCAATTGAAATa, tGACTCACATGAa, aCCATGTTAAAAa, aTTATGCCAAAAt, aAATTGTTGTAAt, aCTATGTGTAGTg, cAGTTAAAATAAt, aTTATATCAGAAa	57, 118, 123, 346, 356, 375, 821, 945, 966, 1080, 1128, 1355, 1794, 1911, 56, 245, 838, 1793, 37, 92, 556, 614, 818, 1212, 1278, 1355, 1393, 1633	UniPROT P12959
P	MYB	Pigment biosynthesis	tCCAACAac, aCCAGCCaa, aTCAACCta, gCTAACCat, aATAACCtg, aACAACTag, caGGTCGAt, tACAACCat, aTTTACCat, caAGTTGCt, aCCTATCtc, aCCTAACac, gCCAACCtt	150, 454, 489, 610, 954, 1060, 1130, 1193, 1200, 1295, 1539, 1549, 1574	UniPROT P27898

PBF (PYRIMIDINE-BOX BINDING FACTOR)	DOF	Seed development	tccCTGTaata, tggaAAAGtaa, tcaCTTGattc, atcaCAAGtga, tggcAAAGaga, cttcAAAGtcc, tttcAAAGaaa, tgaCTTGattt, tgaATTTaaaa, ggtaAAAGgaa, aaatAACGcta, aatcAAAGaga, aaaaAAAGcgg, aaaaAAAAagt, caaaAAAAaaa, aaaaAAAAaca, tggCTTTagat, tattAAATaaa, aattAAATaaa, aggaAAAGtac, gacaAAATcaa, atgaAAGGagc, ttgaGAAGaga, acaCTTTacca, actgAAAGaaa, aatgAAAGaaa, gtatAAAGata, atttAAAGaca, aataAAAAagt, tggCTTTgtca, tgtATTTaatt, aattCAAGaaa, agcaAAAGtaa, taaaAAAAtaa, aaatAAATaac, aataAAAAaaa, taacAAAGtta, cttaAACGcgt, ccgCATTttat, ttaTTTTtcat, gcaCTTAtatt, cagaAAATtca	40, 63, 68, 124, 177, 217, 282, 306, 357, 361, 466, 591, 603, 632, 671, 682, 688, 712, 725, 730, 740, 752, 760, 770, 778, 790, 799, 826, 849, 891, 932, 978, 983, 990, 1003, 1023, 1119, 1143, 1222, 1236, 1313, 1408, 1413, 1429, 1434, 1555, 1604, 161, 1625, 1640	UniPROT O24463
PCF2 (PROLIFERATING CELL FACTOR 2)	TCP	Meristem development	agGGACCCAT, GTGGGTAAaa, taGTGCCCAT, GTGGGTCCtc, ATGAGGCCca, tcTGGCCCTC	417, 590, 1363, 1486, 1498, 1519	UniPROT Q6ZBH6
PIF3 (POLYCHROME INTERACTING FACTOR 3)	PIF	GA response, light response	attggCATCGTGGgtcc, gaggcCCATGTGCccgg, tctcCCATGAGGaaatc	1477, 1500, 1588	TAIR AT1G09530

RAV1 (RELATED TO ABI1/VP1)	AP2/B3	Cold response, ethylene response, brassinosteroid response, light response, root development, leaf development, reproductive structure development	taGCCATAGAaa, atCGAACAAAta, atCCAACAACaa, ctCAAACATAta, aaATTATTGCag, atACAACACAga, taTAAGTTGTgg, ccGCAAAAAAaa, aaGCAAAAAAaa, caAGTGTGGCtt, tcACAACACTtt, aaGCAACAACta, ttCCCACAAGaa, ttGCAACAGGga, agCAAACATAtg, aaATTGTTGTaa, tgCCAGCAAAag, tgAATGCTGTtt, tgCCAACCTTaa, ccATTGTTGGgg, atTATCTCGCtt, ctGAAACATTtt, ttCCAAAAAAaa, gtGCAGCAGAtg, gtCACTTGattc, ttaaCAGATGac, atAACCTGacat, ttCACCTCttcc, gaAACCTGgctt, ctCGCATGaggc, ggccCATGTGcc, ttCTCCTGgtga, cggcCAGGAAgg	20, 149, 252, 394, 474, 566, 582, 709, 721, 734, 886, 1056, 1094, 1181, 1252, 1278, 1309, 1331, 1417, 1548, 1718, 1759, 1773, 1882, 1916, 123, 810, 955, 966, 1113, 1126, 1293, 1502, 1678	TAIR AT1G13260

RITA1 (RICE TRANSCRIPTION ACTIVATOR - 1)	RITA	Seed development	ttACATg, tAAGTca, tTCGTgt, acAAGTg, tAGGTca, tgACTTg, gATGTcg, tgACGAa, acAAGTg, tcACATg, tcACATa, cACATag, tcACCTc, tACGCct, ttACGTa, cAGGTcg, ctATGTg, ccATGTg, acACCTa, tACCTaa, atACGTa	58, 120, 132, 180, 185, 348, 357, 522, 733, 822, 902, 947, 967, 1033, 1081, 1130, 1356, 1505, 1537, 1580, 1666	UniPROT Q6ETX0
TAF1 (TBP-ASSOCIATED FACTOR 1, HAF2, HISTONE ACETYLTRANSFERASE OF THE TAF11250 FAMILY 2)	BZIP	Histone modification, light response	gtcACTTGat, caCAAGTgac, aaCAAGTgtg, gaaACCTGgc, caCAAGTtgc, gcaTCGTGgg, ccCATGTgcc, aaCACCTctc	123, 179, 732, 1113, 1293, 1481, 1504, 1578	TAIR AT3G19040
TEIL (TOBACCO ETHYLENE-INSENSITIVE 3)	EIL	Ethylene response	AATTACAT, AAGTAAAT, AGACTCAT, TGATTCGT, ATGAAATT, AACTGCAT, AAATACAA, AAGTCCAT, CTGAATAT, ACATATAT, ATGAATTT, ATCAACCT, AAGTCCGT, AGATATAT, AAGTACAA, AGCTTCAA, TTGCACAT, ACATTCAC, AAGTACGC, AGGTCGAT, ATGTATTT, GGATACAA, AGGAACGT, AAATTCAA, AGGCCCAT, ACCTATCT, CGGTACCT, AAATTCAG, ATGTTCTT, ACGTACGA	56, 68, 110, 129, 160, 241, 259, 287, 331, 399, 465, 489, 564, 665, 693, 747, 783, 867, 944, 963, 1030, 1131, 1142, 1190, 1221, 1343, 1501, 1507, 1539, 1570, 1644, 1658, 1668	UniPROT Q9ZWK1

TGA1A	BZIP	Auxin response, defense response, histone modification	cacAAGTgac, gtcACTTgat, aacAAGTgtg, ctcACATgaa, attACGTagt, gtgACATggc, ggcCCGTgag, gtcACGTtct, aAGAGATTAAGGATGTTATaac, aATATTTTATTTACAACTTcaa, tggAGAATATTTCACAACACTt, ctaATGACAAAGTTAAATTGT	123, 179, 732, 821, 1080, 1794, 1801, 1853, 222, 266, 875, 1264	UniPROT P14232
WRKY	WRKY	Defense response, ethylene response	ATTCTAAGTCA, AGACTGAATAT, ATATTAGGTCA, AGACAAAATCA, TTACGTAGTCA, TGGCTTTGTCA, TGACAAAGTTA, TAACAAAATAA	116, 328, 344, 798, 1081, 1119, 1268, 1420	TAIR AT1G13960 (many WRKYs, one selected)
ZAP1/WRKY1 (ZINC-DEPENDENT ACTIVATOR PROTEIN-1)	WRKY	Defense response	aTTAGGTCAA, gTCCGTTCAA, gGTCGATCAA, TTAAACGCGt	346, 695, 1132, 1556	TAIR AT2G04880

Table 3.4 Yeast one-hybrid interaction strength table. Promoter-protein interactions are categorized as "weak" or "strong" based on yeast colony growth on increasing concentrations of 3-AT (Figure 3.5).

Promoter	Interacting Transcription Factor	Strength of Interaction
PgSAL1	*PgSOC1*-like	+, weak
PgSAL1	*PgFLX*-like	++, strong
PgSAL1	*PgASR*-like	++, strong
Putative *PgSAL5*	*PgCPC/ETC*-like	++, strong
Putative *PgSAL5*	*PgMYB1*	+, weak
Putative *PgSAL5*	*PgNBS-LRR/WRKY*-like	++, strong

Table 3.5 Yeast one-hybrid (Y1H) identified transcription factors sequence similarity to other plant species. Sequences obtained from the yeast one-hybrid search are indicated in parentheses "(Y1H)". Interacting promoter column is based on the transcription factors that interacted with a *Pg*SAL promoter in the yeast one-hybrid analysis. The transcription factor sequence obtained from the yeast one-hybrid screen was submitted to BLAST to assist in sequence identification. Selected BLAST hits are listed with the corresponding species, gene, and/or NCBI accession number. Pairwise amino acid sequence similarity of BLAST hit sequence and transcription factor obtain from the Y1H screen is shown below. Pairwise alignment performed with EMBOSS Needle amino acid alignment.

Interacting Promoter	BLAST Hit	NCBI Accession #	Amino Acid Length (bp)	Sequence similarity to Y1H identified sequence (%)
*Pg*SAL1	*PgSOC1*-like (Y1H)	-	218	-
-	*Picea glauca* GQ03235_L08	BT111301.1	218	100
-	*Picea abies SOC1*	KM516089.1	218	99.5
-	*Populus tremuloides SOC1/PTM5* (*SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1/POPULUS TREMULOIDES MADS-box 5*)	AF377868.1	220	64.8
*Pg*SAL1	*PgFLX*-like (Y1H)	-	151	-
-	*Picea glauca* GQ04104_P24	BT119390.1	288	39.8
-	*Cicer arietinum FLX-Like 3* (*FLOWERING LOCUS C EXPRESSOR-LIKE3*)	XR_001144004.1	284	29.8

-	*Arabidopsis thaliana FLX4*	NM_125585.2	23.8	23.5
*Pg*SAL1	*PgASR*-like (Y1H)	-	254	-
-	*Picea glauca* GQ04113_F22	BT119966.1	240	90.2
-	*Solanum lycopersicon ABSCISIC ACID STRESS RIPENING 1* (*ASR1*)	NM_001247208.2	297	22.7
-	*Solanum lycopersicon ABSCISIC ACID STRESS RIPENING 4* (*ASR4*)	NM_001282319.1	113	33.3
Putative *PgSAL5*	*PgCPC/ETC*-like (Y1H)	-	73	-
-	*Picea glauca* GQ03207_J20	BT109362.1	111	13.4
-	*Amborella trichopoda* R3 MYB-like *ETC1* (*ENHANCER OF TRY AND CPC 1*)	XM_006842642.2	78	29.6
-	*Camelina sativa* R3 MYB-like *ETC3*	XM_010429541.2	78	33.7
-	*Arabidopsis thaliana ETC3/CPL3* (*CAPRICE-LIKE MYB3*)	NM_116336.4	77	38.9
-	*Morus notabilis CPC* (*CAPRICE*)	XM_010111223.1	73	28.3
-	*Arabidopsis thaliana CPC*	NM_130205.2	117	25
-	*Arabidopsis thaliana ETC1*	NM_100020.4	115	24

Putative *PgSAL5*	*PgMYB1* (Y1H)	-	385	-
-	*Picea glauca MYB1*	EF601064.1	398	98.5
-	*Arabidopsis thaliana MYB20*	NM_105294.3	282	44.9
Putative *PgSAL5*	*PgNBS-LRR/WRKY*-like (Y1H)	-	147	-
-	*Picea glauca* GQ0033_E20	BT100632.1	171	46.4
-	*Arabidopsis thaliana WRKY19*	NM_001125496.2	1895	3.5
-	*Arabidopsis thaliana WRKY16*	NM_180802.2	1372	4.6

Chapter 3 Supplemental Data

```
1          11         21         31         41        50
|          |          |          |          |         |
---------- ---------- ---------- ---------- ----------     0 PgSAL1
tgaattaaca ctaatattaa acctccctct aattacttag gtattcccat    50 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     0 PgSAL1
tctccctctt agagagtatg ctagtttaat gtattatgtt ttatggacat   100 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     0 PgSAL1
ctccctttaa taaaatataa gaatatatag aactaataat attaatgtaa   150 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- -------act aat------- ---------- ----------     6 PgSAL1
taatttttt taaggtgact aattaattta tgcatctaaa aatttgattt   200 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
cattaataaa aaactataaa ctacaattgg acaatattca aaattatata   250 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
gctcatctca attcaattct agtgattatt gctaggaaaa ctcccaatct   300 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
tagtttttac ttcttttata tcgagtggtg tgacatccac gtaactgtaa   350 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
ttgtggattt aaatggattt ctgacactgt atctatatac ggatttctgc   400 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
cgtttacagc tttggtctta tcggcggaat cgcggccctt cgagagattt   450 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
cgagtcagcc tagaaagctt tggttttatt ggcggaatcg cggcccttcg   500 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
agagatttcg agtcacccta gtgatacaat catgaaaggg agggaaggcc   550 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
caacgaccac actatttgct tttcattatt aaggcccgct gtactgcact   600 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
gcaaaaaaat ttagacaaac ttattagggc ccgctgcact gcaaaaaact   650 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1
```

```
---------- ---------- ---------- ---------- ----------     6 PgSAL1
tatgtctagc caaacttatt agggcccgct gtactgttct gtagaccaaa   700 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
gtttcctcga tgagctgtca gaagccgaag ttgtgccctc gattgctcgg   750 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
gaggataacg cttccgaagt tcggtgttgg tttttgtcgc ttgattttag   800 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
ggttttccac ccttttaatc ttgttgtagg ccctagattg ttcgggagga   850 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
gaacgcttcg aaagttcggt gttggaattt gtcgcttgat tttagggttc   900 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
tccacaaacc cgattttcca gccttttaat cttggtgtag gccttcggat   950 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     6 PgSAL1
ttgttagaaa ttttttcctt tccctgtgta tgatttgtat gctaatcgag  1000 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ----aagggt tgggactata    22 PgSAL1
agagatcttg catgttgttg taatctcaga ttggagtgac atggcccgcg  1050 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


gaaataataa tattaatgta ataattattt tatgcatcta aaaatttgat    72 PgSAL1
agaaaataaa aattaagaga atagctaacg cttcggctag gcaggtcacg  1100 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


ttc------- ---------- attaattaaa aactataaac cacaattgga   105 PgSAL1
ttctcgaaga ggcgcagggg gttgttnaaa aactataaac cacaattgga  1150 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


caaaattcaa aattatctca tctcaattca attctagtga ttattgctag   155 PgSAL1
caaaattcaa aattatctca tctcaattca attctagtga ttattgctag  1200 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


gaaaactccc aatcttagtt atgacttctt ttatatcgag tggtgtgaca   205 PgSAL1
gaaaactccc aatcttagtt atgacttctt ttatatcgag tggtgtgaca  1250 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


tccacgtaac tgttattgtg gattgagatg gatttctgac actctatcta   255 PgSAL1
tccacgtaac tgttattgtg gattgagatg gatttctgac actctatcta  1300 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1
```

```
tatacggatt tctgccgttt acagcttttg ctcttatcgg cggaatcgcg   305 PgSAL1
tatacggatt tctgccgttt acagcttttg ctcttatcgg cggaatcgcg  1350 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


gcccttcgag agaattcgag tcagcctaga cagctttggt tttatcggcg   355 PgSAL1
gcccttcgag agaattcgag tcagcctaga cagctttggt tttatcggcg  1400 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


gaatcgcggc ccttcgagag attttgagtc accctagcga tacaatcatg   405 PgSAL1
gaatcgcggc ccttcgagag attttgagtc accctagcga tacaatcatg  1450 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


aaaggggggg aaggcccaac gactacgcta tttgcttttc attattaagg   455 PgSAL1
aaaggggggg aaggcccaac gactacgcta tttgcttttc attattaagg  1500 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


cccgctgtac tgcactgcaa aaaacttatg tctagccaaa cttattaggg   505 PgSAL1
cccgctgtac tgcactgcaa aaaacttatg tctagccaaa cttattaggg  1550 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


cccgctgcac tgcaaaaaac ttatgtttag ccaaacttat tagggcccgc   555 PgSAL1
cccgctgcac tgcaaaaaac ttatgtttag ccaaacttat tagggcccgc  1600 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


tgtactgtgc tgtagaccaa agtttcctcg atgagctgtc agaagccgaa   605 PgSAL1
tgtactgtgc tgtagaccaa agtttcctcg atgagctgtc agaagccgaa  1650 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


gttgtgccct cgattgctcg ggaggataac gcttccgaag ttcggtgttg   655 PgSAL1
gttgtgccct cgattgctcg ggaggagaac gcttccgaag ttcggtgttg  1700 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


gtttttgtcg cttgatttta gggttttcca ccaatccgat tttccaccct   705 PgSAL1
gtttttgtcg cttgatttta ggg------- ---------t tttccaccct  1734 Pg-01r141201s2137277
---------- ---------- ---------- ---------- ----------     0 SAL1


tttaatcttg ttgtaggccc tagattgttc gggaggagaa cgcttcgaaa   755 PgSAL1
tttaatcttg ttgtaggccc tagattgttc gggaggagaa cgcttcgaaa  1784 Pg-01r141201s2137277
---------- ---------- ---------- ---aggagaa cgcttcgaaa    17 SAL1


gttcggtgtt ggaatttgtc gcttgatttt agggttctcc acaaacctga   805 PgSAL1
gttcggtgtt ggaatttgtc gcttgatttt agggttctcc acaaacctga  1834 Pg-01r141201s2137277
gttcggtgtt ggattttgtc gcttgatctt agggttctcc acaaacccga    67 SAL1


ttttccagcc ttttaatctt ggtgtaggcc ttcggatttg ttggaaaaaa   855 PgSAL1
ttttccagcc ttttaatctt ggtgtaggcc ttcggatttg ttggaaaaaa  1884 Pg-01r141201s2137277
ttttccagcc ttttaatctt ggtgtaggcc ttcggatttg ttggaaaatt   117 SAL1


tttcctttcc c--------- tttgtatgct aatcgagaga gatcttgcct   896 PgSAL1
tttcctttcc c--------- tttgtatgct aatcgagaga gatcttgcct  1925 Pg-01r141201s2137277
tttcctttcc ctgtgtatga tttgtatgct aatcgagaga gatcttgcat   167 SAL1
```

```
gttgttgtaa tctcagattg gaatgacatg gcccgagaga aaatagagat   946 PgSAL1
gttgttgtaa tctcagattg gaatgacatg gcccgagaga aaatagagaa  1975 Pg-01r141201s2137277
gttgttgtaa tctcagattg gagtgacatg gcccgcgaga aaataaaaat   217 SAL1


gaagagaata gctaacgctt cggcgaggca gatggcgttc tcgaagaggc   996 PgSAL1
gaagagaata gctaacgctt cggcgaggca gatgacgttc tcgaagaggc  2025 Pg-01r141201s2137277
taagagaata gctaacgctt cggctaggca ggtcacgttc tcgaagaggc   267 SAL1


ggaggggtt gttcaaaaaa gctgaggagc tatcgattct atgtgcagca  1046 PgSAL1
ggaggggtt gttcaaaaaa gctgaggagc tatcgattct atgtgcagca  2075 Pg-01r141201s2137277
gcaggggtt gttcaaaaaa gctcaagagt tatcgattct atgtgaagca   317 SAL1


gatgtagccc tcgtcgtttt ttcttccact gggaagctgt acgactactc  1096 PgSAL1
gatgtagccc tcgtcgtttt ttcttccact gggaagctgt acaactactc  2125 Pg-01r141201s2137277
gatgtagccc tcgtcgtttt ttcttccact gggaagctgt acgactactc   367 SAL1


gag------- ---------- ---------- ---------- ----------  1099 PgSAL1
gagctccag- ---------- ---------- ---------- ----------  2134 Pg-01r141201s2137277
gagctccagt gtggaggtga tattggacaa gtacgttttg tatccgagca   417 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
---------- ---------- ---------- ---------- ----------  2134 Pg-01r141201s2137277
caattcaaaa ggatggacaa caaattctcg agttcgagag tcaagatccc   467 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
---------- ---------- ---------- ---------- ----------  2134 Pg-01r141201s2137277
aaaaggataa tacaacattt tgaagatgct agtcaagatt tgagggagga   517 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
---------- ---------- ---------- ---------- ----------  2134 Pg-01r141201s2137277
acttgaagga ttaactttaa aagatttaga aaaactagaa gaacaatttg   567 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
---------- ---------- ---------- ---------- ----------  2134 Pg-01r141201s2137277
aaatggagtt gagttgtatt cgatcacaaa aggtggaaca tcttgttaag   617 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
---------- ---------- ---------- ---------- ----------  2134 Pg-01r141201s2137277
aagataaatg agcttcaaca aaaggtaata caaatgatag aggagaatac   667 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
---------- ---------- ---------- ---------- ----------  2134 Pg-01r141201s2137277
aaaactccgt ggacagctaa atgaaggaga tggagaatga tggatgcgaa   717 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
---------- -gtactcatt aaaacattac tccc------ ----------  2157 Pg-01r141201s2137277
tctttcttca ttcagtcatt ggaaaatcag acccccccgtc atctgaatct   767 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
---------- ---------- ---------- ---------- ----------  2157 Pg-01r141201s2137277
ataactactt atgcttttaa attcaaatta cataaattgc ctatcaaaga   817 SAL1
```

```
---------- ---------- ---------- ---------- ----------  1099 PgSAL1
---------- ---------- ---------- ---------- ----------  2157 Pg-01r141201s2137277
ctgaggattc tgttacttcc ttgcagttag ggtatgcaat tatacttatc   867 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
---------- ---------- ---------- ---------- ----------  2157 Pg-01r141201s2137277
aaattctgtt ttttatttat ttttaacggg ggttgtaaca ttttatgtta   917 SAL1


---------- ---------- ---------- ---------- -----       1099 PgSAL1
---------- ---------- ---------- ---------- -----       2157 Pg-01r141201s2137277
gggtattttt ttgtttaaat ttggacgaaa ttttttttta aaaaa        962 SAL1
```

Figure S3.1 Alignment of PG29 Pg-01r1412s213727 contig, promoter cDNA alignments for *PgSAL1* were conducted in MAFFT (mafft.cbrc.jp/alignment/server/) using default parameters.

```
1          11         21         31         41        50
|          |          |          |          |         |
---------- ---------- ---------- ---------- ----------     0 PgSAL1
attttaatgt aatatattat atggacatct ctctttaata aactaataag    50 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     0 PgSAL1
aatctataga actaataata ttaatgtaat aatttttta aggtgtctaa   100 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     0 PgSAL1
ttaatttatg tatctaaaat ttttatttca ttaattaaaa actataaatt   150 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     0 PgSAL1
acaattgaac aaaattcaaa attatatagc tcatgtaaat tcacccattt   200 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     0 PgSAL1
attcaaaatt atatagatca tctcaataca cccatttgac tatatatatt   250 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- ---------- ---------- ----------     0 PgSAL1
tttatatttt atatattttt attattattt taatatatta tataatatta   300 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


---------- ---------- --actaataa gggttgggac tatagaaata    28 PgSAL1
tatggacatt tccccttaat aaactaataa ggg-tgggac tatagaaata   349 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


ataatattaa tgtaataatt attttatgca tctaaaaatt tgatttcatt    78 PgSAL1
ataatattaa tgtaataatt attttatgca tctaaaaatt tgatttcatt   399 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


aattaaaaac tataaaccac aattggacaa aattcaaaat tatctcatct   128 PgSAL1
aattaaaaac tataaaccac aattggacaa aattcaaaat tatctcatct   449 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


caattcaatt ctagtgatta ttgctaggaa aactcccaat cttagttatg   178 PgSAL1
caattcaatt ctagtgatta ttgctaggaa aactcccaat cttagttatg   499 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


acttctttta tatcgagtgg tgtgacatcc acgtaactgt tattgtggat   228 PgSAL1
acttctttta tatcgagtgg tgtgacatcc acgtaactgt tattgtggat   549 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


tgagatggat ttctgacact ctatctatat acggatttct gccgtttaca   278 PgSAL1
tgagatggat ttctgacact ctatctatat acggatttct gccgtttaca   599 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


gcttttgctc ttatcggcgg aatcgcggcc cttcgagaga attcgagtca   328 PgSAL1
gcttttgctc ttatcggcgg aatcgcggcc cttcgagaga attcgagtca   649 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1
```

```
gcctagacag ctttggtttt atcggcggaa tcgcggccct tcgagagatt   378 PgSAL1
gcctagacag ctttggtttt atcggcggaa tcgcggccct tcgagagatt   699 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


ttgagtcacc ctagcgatac aatcatgaaa ggggggggaag gcccaacgac   428 PgSAL1
ttgagtcacc ctagcgatac aatcatgaaa ggggggggaag gcccaacgac   749 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


tacgctattt gcttttcatt attaaggccc gctgtactgc actgcaaaaa   478 PgSAL1
tacgctattt gcttttcatt attaaggccc gctgtactgc actgcaaaaa   799 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


acttatgtct agccaaactt attagggccc gctgcactgc aaaaaactta   528 PgSAL1
acttatgtct agccaaactt attagggccc gctgcactgc aaaaaactta   849 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


tgtttagcca aacttattag ggcccgctgt actgtgctgt agaccaaagt   578 PgSAL1
tgtttagcca aacttattag ggcccgctgt actgtgctgt agaccaaagt   899 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


ttcctcgatg agctgtcaga agccgaagtt gtgccctcga ttgctcggga   628 PgSAL1
ttcctcgatg agctgtcaga agccgaagtt gtgccctcga ttgctcggga   949 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


ggataacgct tccgaagttc ggtgttggtt tttgtcgctt gattttaggg   678 PgSAL1
ggagaacgct tccgaagttc ggtgttggtt tttgtcgctt gattttaggg   999 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


ttttccacca atccgatttt ccacccttttt aatcttgttg taggccctag   728 PgSAL1
---------- ------tttt ccacccttttt aatcttgttg taggccctag  1033 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------     0 SAL1


attgttcggg aggagaacgc ttcgaaagtt cggtgttgga atttgtcgct   778 PgSAL1
attgttcggg aggagaacgc ttcgaaagtt cggtgttgga atttgtcgct  1083 Pg-02r141203s0882372
---------- aggagaacgc ttcgaaagtt cggtgttgga ttttgtcgct    40 SAL1


tgattttagg gttctccaca aacctgattt tccagccttt taatcttggt   828 PgSAL1
tgattttagg gttctccaca aacctgattt tccagccttt taatcttggt  1133 Pg-02r141203s0882372
tgatcttagg gttctccaca aacccgattt tccagccttt taatcttggt    90 SAL1


gtaggccttc ggatttgttg gaaaaaattt cctttccc-- -------ttt   869 PgSAL1
gtaggccttc ggatttgttg gaaaaaattt cctttccc-- -------ttt  1174 Pg-02r141203s0882372
gtaggccttc ggatttgttg gaaaattttt cctttccctg tgtatgattt   140 SAL1


gtatgctaat cgagagagat cttgcctgtt gttgtaatct cagattggaa   919 PgSAL1
gtatgctaat cgagagagat cttgcctgtt gttgtaatct cagattggaa  1224 Pg-02r141203s0882372
gtatgctaat cgagagagat cttgcatgtt gttgtaatct cagattggag   190 SAL1


tgacatggcc cgagagaaaa tagagatgaa gagaatagct aacgcttcgg   969 PgSAL1
tgacatggcc cgagagaaaa tagagatgaa gagaatagct aacgcttcgg  1274 Pg-02r141203s0882372
tgacatggcc cgcgagaaaa taaaaattaa gagaatagct aacgcttcgg   240 SAL1


cgaggcagat ggcgttctcg aagaggcgga gggggttgtt caaaaaagct  1019 PgSAL1
cgaggcagat gacgttctcg aagaggcgga gggggttgtt caaaaaagct  1324 Pg-02r141203s0882372
ctaggcaggt cacgttctcg aagaggcgca gggggttgtt caaaaaagct   290 SAL1
```

```
gaggagctat cgattctatg tgcagcagat gtagccctcg tcgttttttc  1069 PgSAL1
gaggagctat cgattctatg tgcagcagat gtagccctcg tcgttttttc  1374 Pg-02r141203s0882372
caagagttat cgattctatg tgaagcagat gtagccctcg tcgttttttc   340 SAL1


ttccactggg aagctgtacg actactcgag ---------- ----------  1099 PgSAL1
ttccactggg aagctgtaca actactcgag ctccaggtac tcattaaaac  1424 Pg-02r141203s0882372
ttccactggg aagctgtacg actactcgag ctccag---- ----------   376 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
attactccct tgttctttct tcttctggtt cgaattcata gtctccataa  1474 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   376 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
agaataaaga aacacagata aagctacttg tatttcgttg ggcatgtttt  1524 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   376 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
cagtaagttg gtagcacatg acagtaagtt gtgtagaaac aactattaaa  1574 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   376 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tctgtgaggt gcgcgcacaa agcttttcca aataaaaaga gagagcatct  1624 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   376 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
attctttctt agcaacgaag ataatactag tgaaaagtat acagcgatat  1674 Pg-02r141203s0882372
---------- ---tgtggag gtgatattgg acaagtacgt ----------   403 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tctttcttga caaagaagct ctaatctgta ttccaggagt gagtgagcag  1724 Pg-02r141203s0882372
---------- ---------- ----tttgta tcc------- ----------   412 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tataaagttc aaatctgttc atgtgggcca ttacagcttt tccaaagtgt  1774 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   412 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ttttggatca atttagtaat ggatcgtgaa cttttaatta aagtgtgtta  1824 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   412 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
gaataaaatt gctcttgaat cctagaaaag cacaagtcgg cagaaagaaa  1874 Pg-02r141203s0882372
---------- ---------- -------gag cacaattcaa aagga-----   430 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tccgaacaat aaaagggaaa ataatgaggg tttgttactt gatctgctct  1924 Pg-02r141203s0882372
--tggacaac aaa------- ---------- ---------- ------ttct   445 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
cattgttgga gctggtaaaa tcaacagaat aaatgtatgc ttaattcata  1974 Pg-02r141203s0882372
cgagttcgag agtcaagatc ccaaaaggat aa-------- ----------   477 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
cgtacaaagt agtttgatcg ctatgttctt ccgtattgat aatgaatcgc  2024 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   477 SAL1
```

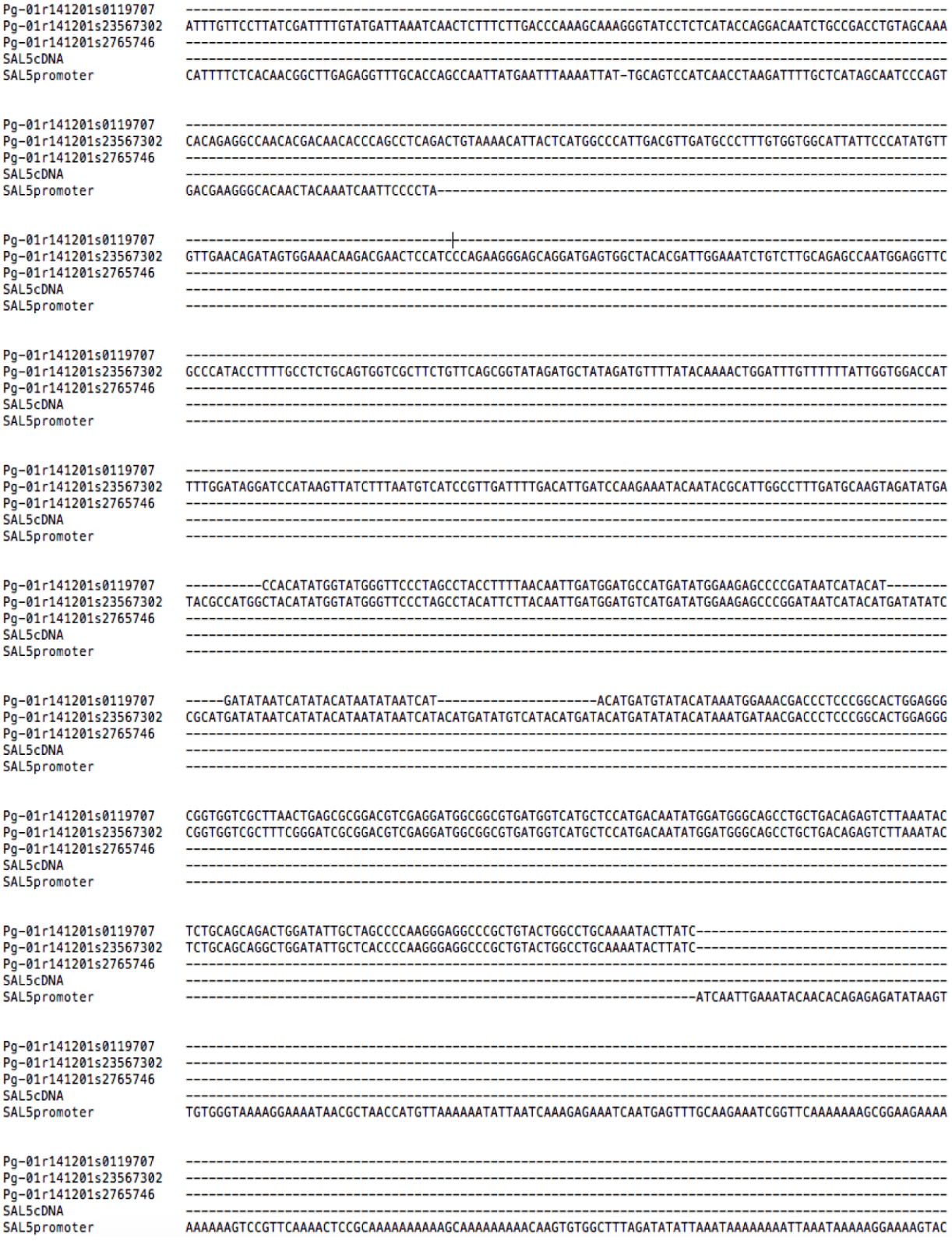
```
---------- ---------- ---------- ---------- ----------  1099 PgSAL1
attttaataa ataattatta aagctgctcg atatattctg ttccgtattc  2074 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   477 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ataatgaatg acatcttgct aaataattat taaaataata ttgttaaggc  2124 Pg-02r141203s0882372
---------- ---------t acaacatttt gaagatgcta --gtcaagat   506 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ttaatagtat gtgcgatgaa acaatagcgc tgggagtaac ctcagttctc  2174 Pg-02r141203s0882372
ttga------ ---------- ---------- -gggaggaac tt--------   521 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
cacattcatc tgatatcccg gcacagaagg actacgttga ggatctttat  2224 Pg-02r141203s0882372
---------- ---------- -----gaagg attaacttta aaagattta-   545 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tctgatttat actgggtaat ctcttcgtag tacttatgca tctgaattga  2274 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   545 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ttcaaggact aagaccaagg aaaatcttga atagaaattg atcgtaccca  2324 Pg-02r141203s0882372
--gaaaaact ag----aaga acaatttgaa atggagttga gttgtattc-   588 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
taaatccaat tttaatatag aatcggtggc tacgttttta aagatctttt  2374 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   588 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
gaatctttgc agagataaaa gcaagatgat tctttattct gatttatact  2424 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   588 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
gggtaatctc ttggcagtac ttacgcatct gagttgattc actaagacca  2474 Pg-02r141203s0882372
---------- ---------- ---------- ---------- -----gatca   593 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
agaaaaatat ttaacagaaa tttatcatac ccataaatcc aattttaaca  2524 Pg-02r141203s0882372
caaaaggt-- ---------- ---------- ---------- ---------g   602 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tagaatcttt gctgagagaa aaacaaggtg atatagtcat gggtttgata  2574 Pg-02r141203s0882372
gaacatcttg ttaagaagat aaatgagct- ---------- ----------   631 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ttgtccatta aaaataactt taacagtgca gtatcaaatg caatatatta  2624 Pg-02r141203s0882372
---tcaacaa aaggtaatac aaatgataga gga------- ----------   661 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
atatacgact ccaaaattca gtatatattg agtgtaatga gtttaatgtt  2674 Pg-02r141203s0882372
------gaat acaaaactcc gtggaca--- ---------- ----------   682 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
gtatattgag tttatttta ttaataaaat aaaataaaaa atcatgtggg  2724 Pg-02r141203s0882372
---------- ---------g ctaaatgaag gagatggaga atgatggatg   713 SAL1
```

```
---------- ---------- ---------- ---------- ----------  1099 PgSAL1
cgagcatttt atagatctct gcattttata gtttataata agaaaattat  2774 Pg-02r141203s0882372
cgaatctttc tt-------- ---------- ---------- ----------   725 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
aaaacttaca tttataaatt cgatgatatc tttaccaccg ctcaataata  2824 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   725 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
agagcatttt atagatctat gcattttgtg gtttatgata aaaaaattat  2874 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   725 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
aaaacttaca tttataaatt tgatgatatc tttatcacat atagtaatta  2924 Pg-02r141203s0882372
---------- ---------- ---------- -------cat tcagtcattg   738 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ggcagaaaga cttataccta ttggagatgg gcacaattaa tgatctttta  2974 Pg-02r141203s0882372
gaaaatcaga ccccccgtca tctgaatcta taactactta tg--ctttta   786 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
atcttaaatt --ataatttg atggctacac attgaagatg tcggtatatt  3022 Pg-02r141203s0882372
aattcaaatt acataaattg cctatcaaag actgaggat- ----------   825 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
cttagtcatc ctctaaaatc aatgatttag atttattgtt gataacttcc  3072 Pg-02r141203s0882372
---------- ---------- ---------- --------tc tgttacttcc   837 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ttgcatttat gaatgcatca cttaacctgt ttatttttta acatttacct  3122 Pg-02r141203s0882372
ttgcagtta- ---------- ---------- ---------- ----------   846 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
aatcatatta agaaattatt gtaacatcat atgattgcca cgtttaaata  3172 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   846 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
attaaccttt gagattacca tgtttagcag attagcaaac actccttcaa  3222 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   846 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
cttcattctt ggaaattaca cgtattacac aagccacgtt tctatgctat  3272 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   846 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
cacagattta atcaattgtc cgcgcgtttc ctctacctcg aatttcctca  3322 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   846 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
gatttcttca actgtaaacc tttcaatgtt caccaacccc gagttagcgg  3372 Pg-02r141203s0882372
---------- ---------- ---------- ---------- --------gg   848 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ctgtgctagt atatgtattt tgataatatt ataaaacata tttaatataa  3422 Pg-02r141203s0882372
gtatgcaatt atacttat-- ---------- ---------- ----------   866 SAL1
```

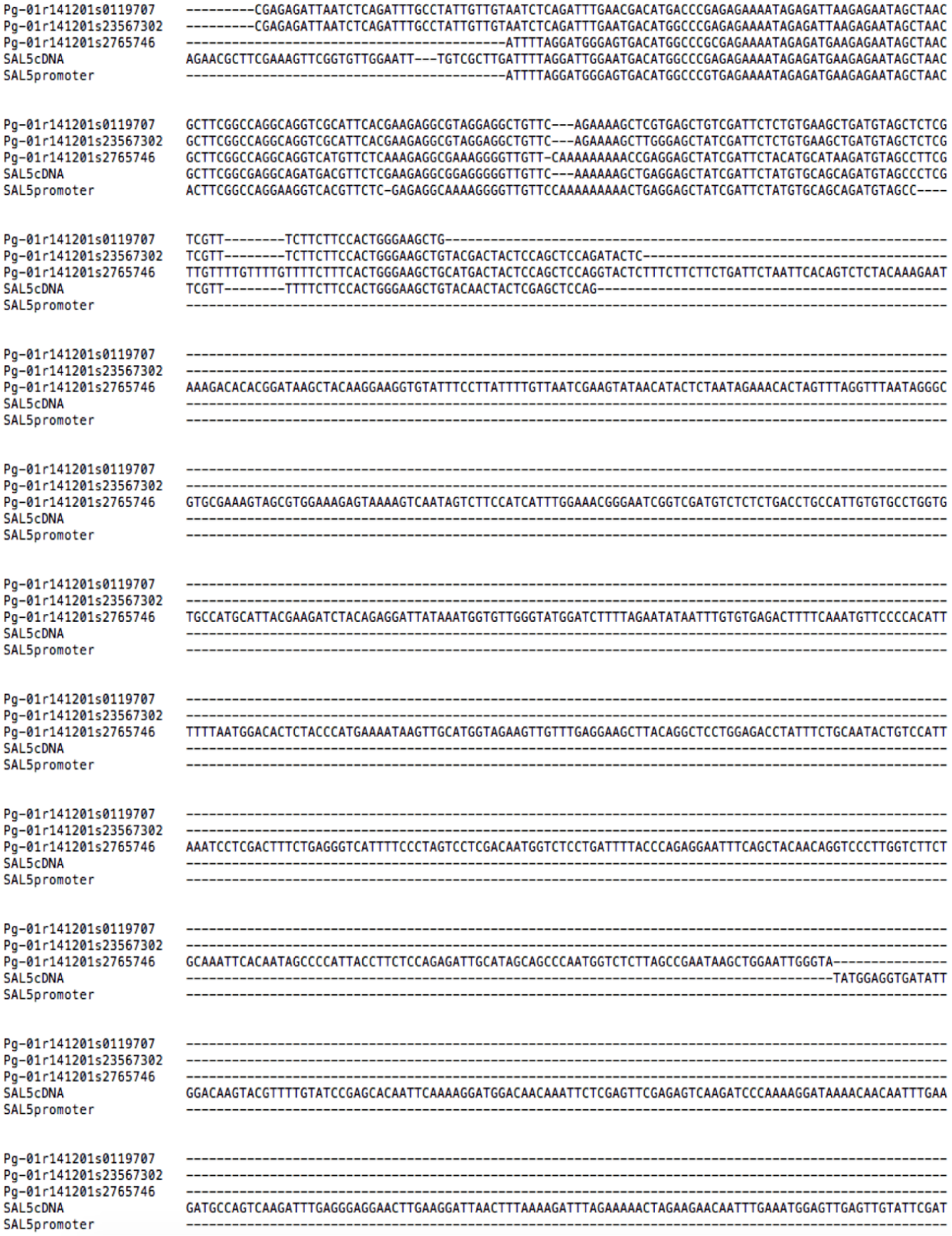
```
---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ttattatcat ttattttagt ggttttcgat ttacttaata aataatgtat  3472 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   866 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
gtatgagtgt atgcacctca ttcttgattc atccttgatt catcgttgtt  3522 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   866 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ttattgtttt taaatactta tatgtatgag tgcatgcact tcattcttga  3572 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   866 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tccatgattc atcattgttt tcaaaataag tatatgcact tcattcttga  3622 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   866 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tccatgattc atcattgttt tcaaaataag tatatggatg ctgttttttt  3672 Pg-02r141203s0882372
---------- ---------- ---------- ----caaatt ctgttttta-   882 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ttttttctct aggttattta gttaattgat aaagtaaagt ttttttaatg  3722 Pg-02r141203s0882372
tttattt--- ---------- -ttaacgggg gttgtaacat ttt----atg   914 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tatagagatt tttttggagg tactattgtt ctaactaaat aatagttgag  3772 Pg-02r141203s0882372
ttagggtatt ttttg----- ---------- ---------- ----------   930 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ttgatctcca attgatagaa tgtcattttt tattgagctt tttcacatat  3822 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   930 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tttctatgtt cagaattgag tacaatgaga cgaaacaaaa tatgacaaaa  3872 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   930 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
aaaaaaaatt gatgcacatt gtattgatat gaatttatag catgcacatt  3922 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   930 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
gtattgatat gaatttatag cacatggatt aacgagatct aagtctagat  3972 Pg-02r141203s0882372
---------- ---------- ---------- -------ttt aaatttgga-   942 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
attatgatgg ggacctacta ggtccaaatc ccagagacaa agttataata  4022 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   942 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
actttgtagt cgactccaac actccagcca tctaaaattc ttcaattggg  4072 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   942 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
cccaaagtca tgtcttaata gagtgattaa ctctctgtta cgcaccagtc  4122 Pg-02r141203s0882372
--cgaaattt ttttttaaaa aa-------- ---------- ----------   962 SAL1
```

```
---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tccaatcaat ttcagatagg gcgcttcctt tggcttggca ctaagccctt  4172 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   962 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
tctcccacac aaccgagtgt taaatccgcc acaagaccat gggaaaataa  4222 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   962 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
acagtggtct aatcagaccg ttggagagta aggtatcaga gttaagtgct  4272 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   962 SAL1


---------- ---------- ---------- ---------- ----------  1099 PgSAL1
ccgatgaatt ggagttgggt gctcctacga gttggagtca actgcttcga  4322 Pg-02r141203s0882372
---------- ---------- ---------- ---------- ----------   962 SAL1


---------- ----                                        1099 PgSAL1
tggtcgaagt taag                                      4336 Pg-02r141203s0882372
---------- ----                                         962 SAL1
```

Figure S3.2 Alignment of WS Pg-02r141203s0882372 contig, promoter cDNA alignments for *SAL1* were conducted in MAFFT (mafft.cbrc.jp/alignment/server/) using default parameters.


Pg-01r141201s0119707
Pg-01r141201s23567302
Pg-01r141201s2765746
SAL5cDNA
SAL5promoter
ATTTGTTCCTTATCGATTTTGTATGATTAAATCAACTCTTTCTTGACCCAAAGCAAAGGGTATCCTCTCATACCAGGACAATCTGCCGACCTGTAGCAAA
CATTTTCTCACAACGGCTTGAGAGGTTTGCACCAGCCAATTATGAATTTAAAATTAT-TGCAGTCCATCAACCTAAGATTTTGCTCATAGCAATCCCAGT
CACAGAGGCCAACACGACAACACCCAGCCTCAGACTGTAAAACATTACTCATGGCCCATTGACGTTGATGCCCTTTGTGGTGGCATTATTCCCATATGTT
GACGAAGGGCACAACTACAAATCAATTCCCCTA
GTTGAACAGATAGTGGAAACAAGACGAACTCCATCCCAGAAGGGAGCAGGATGAGTGGCTACACGATTGGAAATCTGTCTTGCAGAGCCAATGGAGGTTC
GCCCATACCTTTTGCCTCTGCAGTGGTCGCTTCTGTTCAGCGGTATAGATGCTATAGATGTTTTATACAAAACTGGATTTGTTTTTTATTGGTGGACCAT
TTTGGATAGGATCCATAAGTTATCTTTAATGTCATCCGTTGATTTTGACATTGATCCAAGAAATACAATACGCATTGGCCTTTGATGCAAGTAGATATGA
----------CCACATATGGTATGGGTTCCCTAGCCTACCTTTTAACAATTGATGGATGCCATGATATGGAAGAGCCCCGATAATCATACAT--------
TACGCCATGGCTACATATGGTATGGGTTCCCTAGCCTACATTCTTACAATTGATGGATGTCATGATATGGAAGAGCCCGGATAATCATACATGATATATC
-----GATATAATCATATACATAATATAATCAT----------ACATGATGTATACATAAATGGAAACGACCCTCCCGGCACTGGAGGG
CGCATGATATAATCATATACATAATATAATCATACATGATATGTCATACATGATACATGATATATACATAAATGATAACGACCCTCCCGGCACTGGAGGG
CGGTGGTCGCTTAACTGAGCGCGGACGTCGAGGATGGCGGCGTGATGGTCATGCTCCATGACAATATGGATGGGCAGCCTGCTGACAGAGTCTTAAATAC
CGGTGGTCGCTTTCGGGATCGCGGACGTCGAGGATGGCGGCGTGATGGTCATGCTCCATGACAATATGGATGGGCAGCCTGCTGACAGAGTCTTAAATAC
TCTGCAGCAGACTGGATATTGCTAGCCCCAAGGGAGGCCCGCTGTACTGGCCTGCAAAATACTTATC----------
TCTGCAGCAGGCTGGATATTGCTCACCCCAAGGGAGGCCCGCTGTACTGGCCTGCAAAATACTTATC----------
ATCAATTGAAATACAACACAGAGAGATATAAGT
TGTGGGTAAAAGGAAAATAACGCTAACCATGTTAAAAAATATTAATCAAAGAGAAATCAATGAGTTTGCAAGAAATCGGTTCAAAAAAAGCGGAAGAAAA
AAAAAAGTCCGTTCAAAACTCCGCAAAAAAAAAAGCAAAAAAAAACAAGTGTGGCTTTAGATATATTAAATAAAAAAAATTAAATAAAAAGGAAAAGTAC

```
Pg-01r141201s0119707    ----------------------------------------------------------------------------------------------------
Pg-01r141201s23567302   ----------------------------------------------------------------------------------------------------
Pg-01r141201s2765746    ----------------------------------------------------------------------------------------------------
SAL5cDNA                ----------------------------------------------------------------------------------------------------
SAL5promoter            AAATAAAAAAGACAAAATCAATTAACAGATGACTCACATGAAAGGAGCTCCAAATCATATTTGAGAAGAGAGGAGAGGAGCTTCAATGGAGAATATTTCA

Pg-01r141201s0119707    ----------------------------------------------------------------------------------------------------
Pg-01r141201s23567302   ----------------------------------------------------------------------------------------------------
Pg-01r141201s2765746    ----------------------------------------------------------------------------------------------------
SAL5cDNA                ----------------------------------------------------------------------------------------------------
SAL5promoter            CAACACTTTACCATCACATACTGCCTTAACATTGTTCCGAAGCACTGAAAGAAAATTGCACATAGAATAACCTGACATTCACCTCTTCCAATGAAAGAAA

Pg-01r141201s0119707    ----------------------------------------------------------------------------------------------------
Pg-01r141201s23567302   ----------------------------------------------------------------------------------------------------
Pg-01r141201s2765746    ----------------------------------------------------------------------------------------------------
SAL5cDNA                ----------------------------------------------------------------------------------------------------
SAL5promoter            AGTATAAAGATAGAATTTAAAGACATTGAAATAGAATAAAAAAGTACGCCTAGGGTTATCAATTTCCAAGCAACAACTAGTATTATCAAATATTACGTAG

Pg-01r141201s0119707    ----------------------------------------------------------------------------------------------------
Pg-01r141201s23567302   ----------------------------------------------------------------------------------------------------
Pg-01r141201s2765746    ----------------------------------------------------------------------------------------------------
SAL5cDNA                ----------------------------------------------------------------------------------------------------
SAL5promoter            TCATTTTCCCACAAGAAGAAACAAGAAACCTGGCTTTGTCACAGGTCGATCAAATGTATTTAATTCTTTCAAACTCTTAATATTCCCTCTCATTGCAACA

Pg-01r141201s0119707    ----------------------------------------------------------------------------------------------------
Pg-01r141201s23567302   ----------------------------------------------------------------------------------------------------
Pg-01r141201s2765746    ----------------------------------------------------------------------------------------------------
SAL5cDNA                ----------------------------------------------------------------------------------------------------
SAL5promoter            GGGATACAACCATTTACCATCCCATTATGCCAAAATTCAAGAAAGTGTCACTGTTCAAATTTGAGCAAACATATGCTAATGACAAAGTTAAATTGTTGTA

Pg-01r141201s0119707    ----------------------------------------------------------------------------------------------------
Pg-01r141201s23567302   ----------------------------------------------------------------------------------------------------
Pg-01r141201s2765746    ----------------------------------------------------------------------------------------------------
SAL5cDNA                ----------------------------------------------------------------------------------------------------
SAL5promoter            ATATCACAAGTTGCTCATTATGCCAGCAAAAGTAATCAAACTTGAATGCTGTTTAGGAACGTTTTTACTATGTGTAGTGCCCATATTTAACCATTAATAT

Pg-01r141201s0119707    ----------------------------------------------------------------------------------------------------
Pg-01r141201s23567302   ----------------------------------------------------------------------------------------------------
Pg-01r141201s2765746    ----------------------------------------------------------------------------------TAAGAATTGGCATCATG
SAL5cDNA                ----------------------------------------------------------------------------------------------------
SAL5promoter            ATTACAGTTAAAATAATAATAAAAAAATAAATAACAAAATAATAAAAAAAAATAACAAAGTTAATCGTGAAATACTATACGCATAAGAATTGGCATCGTG

Pg-01r141201s0119707    ----------------------------------------------------------------------------------------------------
Pg-01r141201s23567302   ----------------------------------------------------------------------------------------------------
Pg-01r141201s2765746    GGTCCTCGCATGAGGCCCACGTGCCCGGCCTCCGGCCCTCGGTATCTAACACCTATCTCTGCCAACCTTAAATGCGTCCTTCGGTACCTAACACCTCTCT
SAL5cDNA                ----------------------------------------------------------------------------------------------------
SAL5promoter            GGTCCTCGCATGAGGCCCATGTGCCCGGTCTCTGGCCCTCGATAGCTAACACCTATCTCTGCCAACCTTAAACGCGTCCTTCGGTACCTAACACCTCTCT

Pg-01r141201s0119707    ------------------------------------------------------------TCCAGCATGTGC---GGTAGACCAATTTCCCTCGATG
Pg-01r141201s23567302   ------------------------------------------------------------TCCAGCATGTGC---GGTAGACCAATTTCCCTCGATG
Pg-01r141201s2765746    CTCCCATGAGGAAATCCGCGTTTTATTTTTCGTGAGGCGCTTATATTATATCAGAAAATTTAGTCTGGCATGTTCTTATACGTACGATTTTCTCCCGGTG
SAL5cDNA                ----------------------------------------------------------------------------------------------------
SAL5promoter            CTCCCATGAGGAAATCCGCATTTTATTTTTCATGAGGCACTTATATTATATCAGAAAATTCAGTCTGGCATGTTCTTATACGTACGATTTTCTCCTGGTG

Pg-01r141201s0119707    TT--CTGTCAGAAACCTAAGGTGTGCCCTCGATTGTTCGGGAGGATAACGCTTCCTAAGTTTGGTGTTCGATTCTCTCGCTCGGTA-----TTAGGGTTC
Pg-01r141201s23567302   TT--CTGTCAGAAACCTAAGGTGTGCCCTCGGTTGTTCGGGAGGAAAACGCTTCCTAAGTTCGGTGTTGGATTCTCTCGCTCGGTA-----TTAGGGTTC
Pg-01r141201s2765746    ATACCAATTAGAAACCGAAGTTGTGCCCTCCATTGTTGGGGAAGATAACGCTTTCGAAGTACGGTGTTGAATTATCTCGCTTGACTGAAAC----------
SAL5cDNA                ----------------AAGTTGTGCCCTCGATTGCTCGGGAGGAGAACGCTTCCGAAGTTCGGTGTTGGTTTTTGTCGCTTGATT-----TTAGGGTTT
SAL5promoter            ATACCAGTTAGAAACCAAAGTTGTGCCCTCCATTGTTGGGGAGGATAACGCTTTCGAAGTACGGTGTTGAATTATCTCGCTTGACTGAAAC----------

Pg-01r141201s0119707    TCCACAAACCCGGTTTTCCTGCCGTTTGATCTTGTTGTGGGCCTTCGGATTTGATGGAAAATTTGCCCTTTCCGTGTGTATGATTTGTATGCTAATCG--
Pg-01r141201s23567302   TCCACAAACCCGGTTTTCCTGCCGTTTGATCTTGTTGTAGGCCTTCGGATTTGATGGAAAATTTGCCCTTTCCCGCTGTATGATTTGTATGCTAATCG--
Pg-01r141201s2765746    ----------------------------------------------------------------------------------------------------
SAL5cDNA                TCCAC----------------CCTTTTAATCTTGTTGTAGGCCCTAGATTGTTCGGGA--------------------------------------GG
SAL5promoter            ----------------------------------------------------------------------------------------------------
```


Pg-01r141201s0119707 ---------CGAGAGATTAATCTCAGATTTGCCTATTGTTGTAATCTCAGATTTGAACGACATGACCCGAGAGAAAATAGAGATTAAGAGAATAGCTAAC
Pg-01r141201s23567302 ---------CGAGAGATTAATCTCAGATTTGCCTATTGTTGTAATCTCAGATTTGAATGACATGGCCCGAGAGAAAATAGAGATTAAGAGAATAGCTAAC
Pg-01r141201s2765746 ----------ATTTTAGGATGGGAGTGACATGGCCCGCGAGAAAATAGAGATGAAGAGAATAGCTAAC
SAL5cDNA AGAACGCTTCGAAAGTTCGGTGTTGGAATT---TGTCGCTTGATTTTAGGATTGGAATGACATGGCCCGAGAGAAAATAGAGATGAAGAGAATAGCTAAC
SAL5promoter ----------ATTTTAGGATGGGAGTGACATGGCCCGTGAGAAAATAGAGATGAAGAGAATAGCTAAC
Pg-01r141201s0119707 GCTTCGGCCAGGCAGGTCGCATTCACGAAGAGGCGTAGGAGGCTGTTC---AGAAAAGCTCGTGAGCTGTCGATTCTCTGTGAAGCTGATGTAGCTCTCG
Pg-01r141201s23567302 GCTTCGGCCAGGCAGGTCGCATTCACGAAGAGGCGTAGGAGGCTGTTC---AGAAAAGCTTGGGAGCTATCGATTCTCTGTGAAGCTGATGTAGCTCTCG
Pg-01r141201s2765746 GCTTCGGCCAGGCAGGTCATGTTCTCAAAGAGGCGAAAGGGGTTGTT-CAAAAAAAAACCGAGGAGCTATCGATTCTACATGCATAAGATGTAGCCTTCG
SAL5cDNA GCTTCGGCGAGGCAGATGACGTTCTCGAAGAGGCGGAGGGGGTTGTTC---AAAAAAGCTGAGGAGCTATCGATTCTATGTGCAGCAGATGTAGCCCTCG
SAL5promoter ACTTCGGCCAGGAAGGTCACGTTCTC-GAGAGGCAAAAGGGGTTGTTCCAAAAAAAAACTGAGGAGCTATCGATTCTATGTGCAGCAGATGTAGCC----
Pg-01r141201s0119707 TCGTT--------TCTTCTTCCACTGGGAAGCTG----------
Pg-01r141201s23567302 TCGTT--------TCTTCTTCCACTGGGAAGCTGTACGACTACTCCAGCTCCAGATACTC----------
Pg-01r141201s2765746 TTGTTTTGTTTTGTTTTCTTTCACTGGGAAGCTGCATGACTACTCCAGCTCCAGGTACTCTTTCTTCTTCTGATTCTAATTCACAGTCTCTACAAAGAAT
SAL5cDNA TCGTT--------TTTTCTTCCACTGGGAAGCTGTACAACTACTCGAGCTCCAG----------
SAL5promoter ----------
Pg-01r141201s0119707 ----------
Pg-01r141201s23567302 ----------
Pg-01r141201s2765746 AAAGACACACGGATAAGCTACAAGGAAGGTGTATTTCCTTATTTTGTTAATCGAAGTATAACATACTCTAATAGAAACACTAGTTTAGGTTTAATAGGGC
SAL5cDNA ----------
SAL5promoter ----------
Pg-01r141201s0119707 ----------
Pg-01r141201s23567302 ----------
Pg-01r141201s2765746 GTGCGAAAGTAGCGTGGAAAGAGTAAAAGTCAATAGTCTTCCATCATTTGGAAACGGGAATCGGTCGATGTCTCTCTGACCTGCCATTGTGTGCCTGGTG
SAL5cDNA ----------
SAL5promoter ----------
Pg-01r141201s0119707 ----------
Pg-01r141201s23567302 ----------
Pg-01r141201s2765746 TGCCATGCATTACGAAGATCTACAGAGGATTATAAATGGTGTTGGGTATGGATCTTTTAGAATATAATTTGTGTGAGACTTTTCAAATGTTCCCCACATT
SAL5cDNA ----------
SAL5promoter ----------
Pg-01r141201s0119707 ----------
Pg-01r141201s23567302 ----------
Pg-01r141201s2765746 TTTTAATGGACACTCTACCCATGAAAATAAGTTGCATGGTAGAAGTTGTTTGAGGAAGCTTACAGGCTCCTGGAGACCTATTTCTGCAATACTGTCCATT
SAL5cDNA ----------
SAL5promoter ----------
Pg-01r141201s0119707 ----------
Pg-01r141201s23567302 ----------
Pg-01r141201s2765746 AAATCCTCGACTTTCTGAGGGTCATTTTCCCTAGTCCTCGACAATGGTCTCCTGATTTTACCCAGAGGAATTTCAGCTACAACAGGTCCCTTGGTCTTCT
SAL5cDNA ----------
SAL5promoter ----------
Pg-01r141201s0119707 ----------
Pg-01r141201s23567302 ----------
Pg-01r141201s2765746 GCAAATTCACAATAGCCCCATTACCTTCTCCAGAGATTGCATAGCAGCCCAATGGTCTCTTAGCCGAATAAGCTGGAATTGGGTA----------
SAL5cDNA ----------TATGGAGGTGATATT
SAL5promoter ----------
Pg-01r141201s0119707 ----------
Pg-01r141201s23567302 ----------
Pg-01r141201s2765746 ----------
SAL5cDNA GGACAAGTACGTTTTGTATCCGAGCACAATTCAAAAGGATGGACAACAAATTCTCGAGTTCGAGAGTCAAGATCCCAAAAGGATAAAACAACAATTTGAA
SAL5promoter ----------
Pg-01r141201s0119707 ----------
Pg-01r141201s23567302 ----------
Pg-01r141201s2765746 ----------
SAL5cDNA GATGCCAGTCAAGATTTGAGGGAGGAACTTGAAGGATTAACTTTAAAAGATTTAGAAAAACTAGAAGAACAATTTGAAATGGAGTTGAGTTGTATTCGAT
SAL5promoter ----------

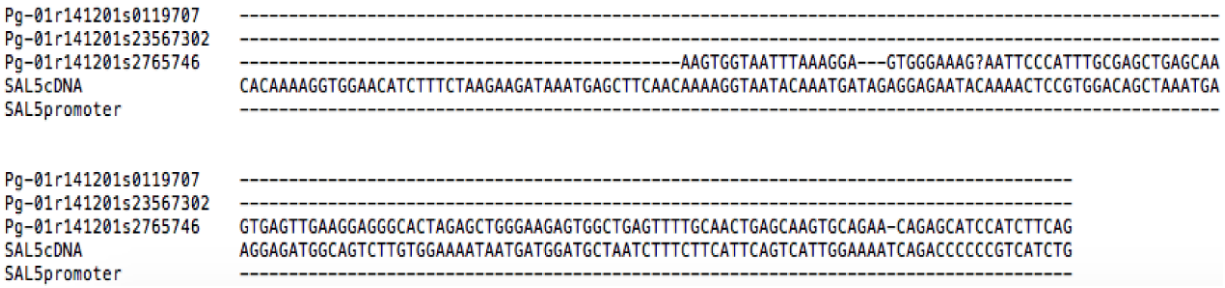


Figure S3.3 Alignment of PG29 Pg-01r141201s0119707, Pg-01r141201s2356730, Pg-01r141201s2765746 contigs, promoter cDNA alignments for putative *SAL5* were conducted in Geneious using the Mauve plugin. Coordinates for alignments are the following: Pg-01r141201s0119707 1010-1771, Pg-01r141201s2356730 7931-9262, Pg-01r141201s2765746 3375-4660, *SAL5* cDNA 1-762, and putative *SAL5* promoter 424-1932.

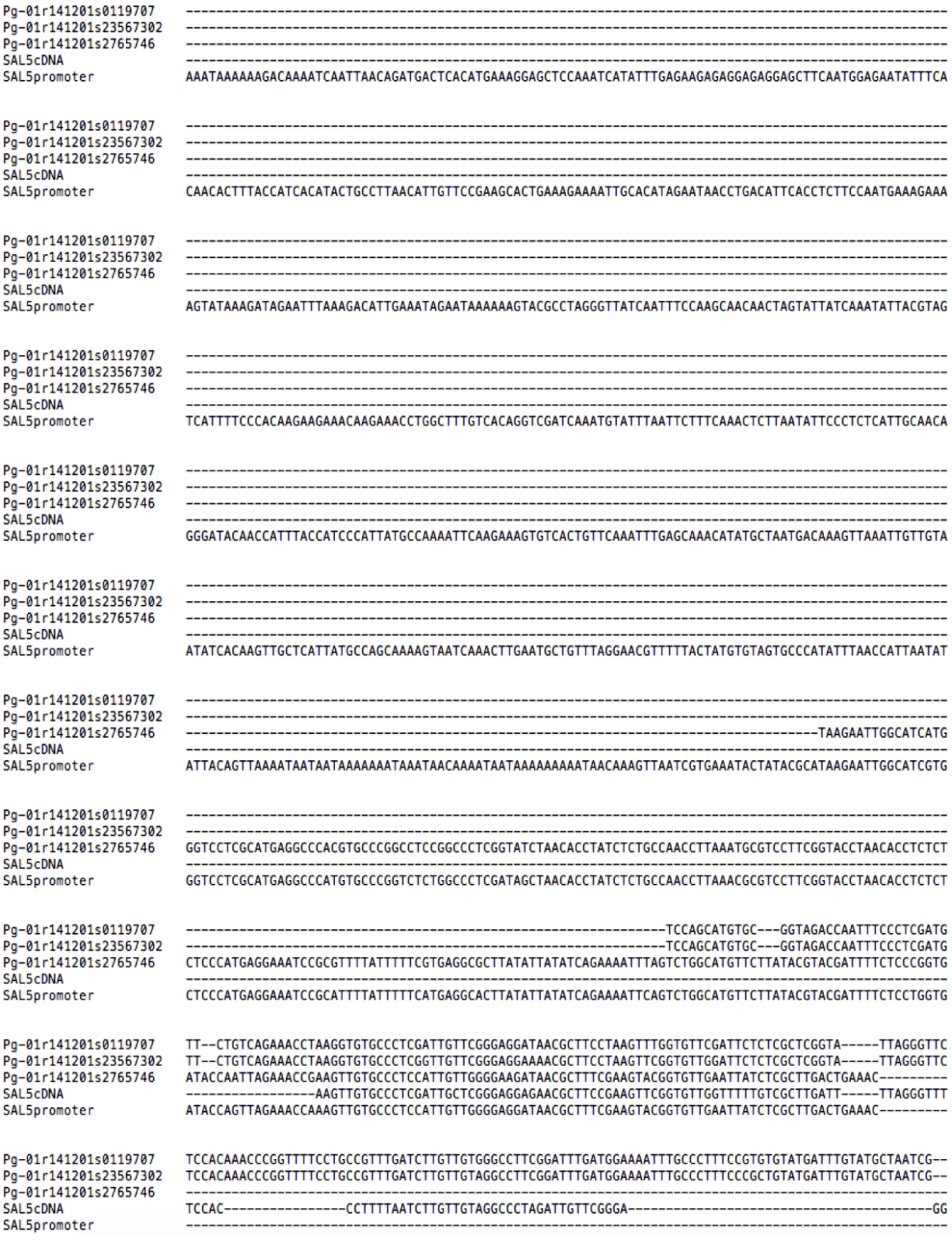

Pg-01r141201s0119707
Pg-01r141201s23567302
Pg-01r141201s2765746
SAL5cDNA
SAL5promoter

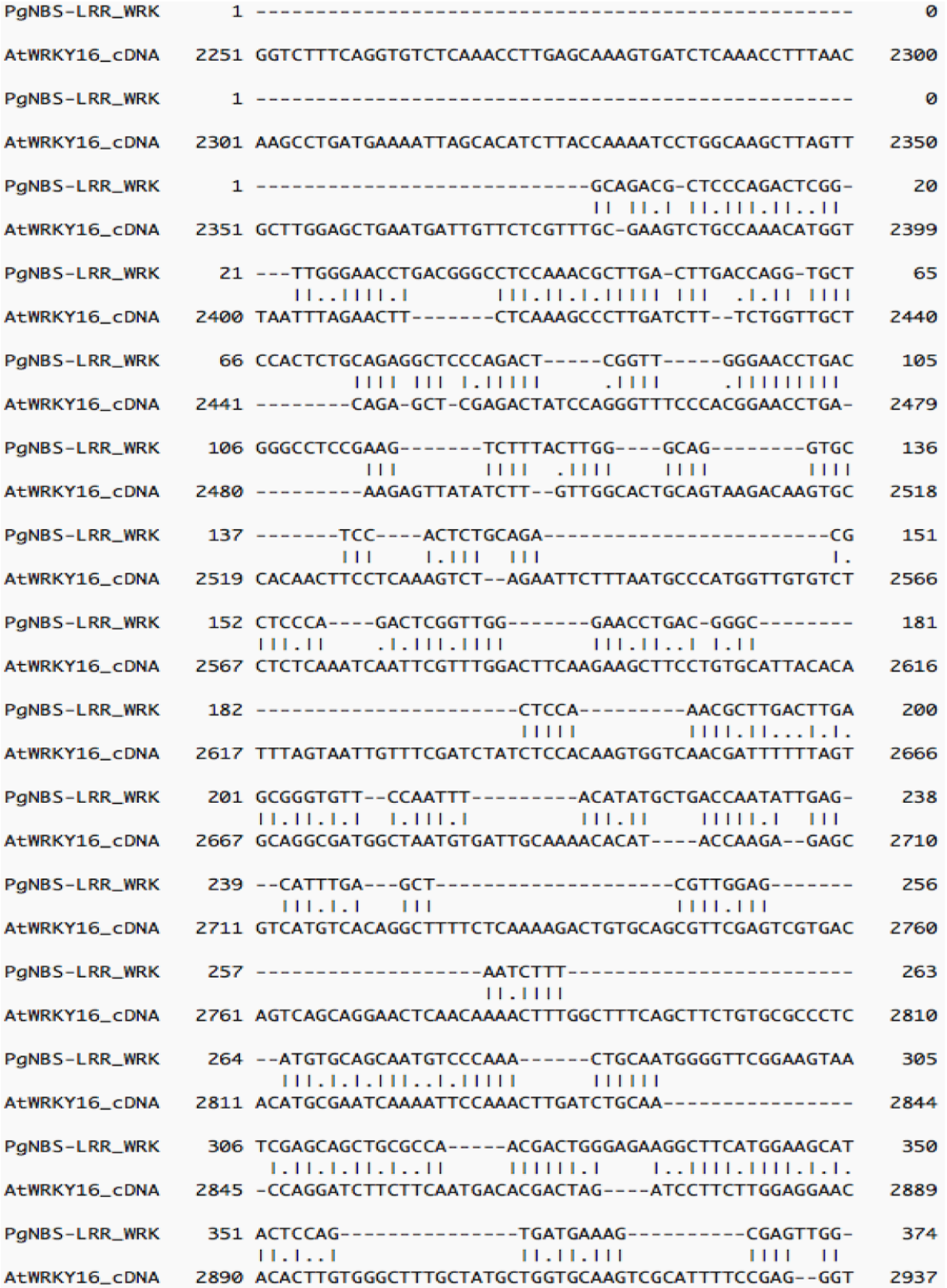
Figure S3.4 Alignment of WS77111 Pg-02r141203s0882372 contig, promoter cDNA alignments for putative *SAL5* were conducted in Geneious using the Mauve plugin. Coordinates

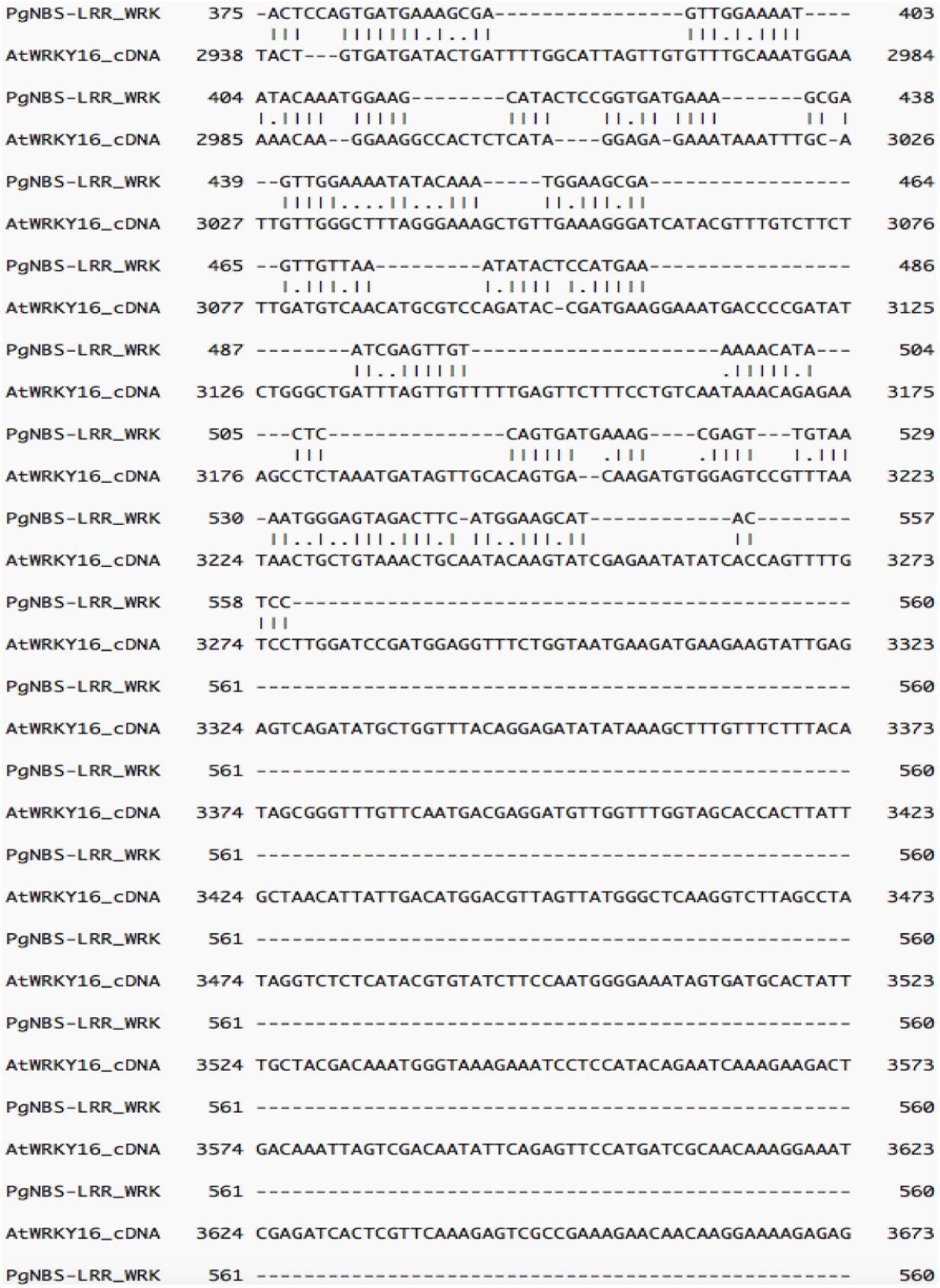
for alignments are the following: Pg-02r141203s0882372 925-3036, SAL5 cDNA 1-362, and putative *SAL5* promoter 1420-1588

PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1	GTTTTTAGCTAGGAATTGAGTGGTTCTCCTTCTCCCTCAACAACTGTTTG	50
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	51	AGAATGGTTGACTTGAGAATGGAGTGCGACTACCATTGGCCAACTTCAGT	100
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	101	GCTTAGGGACCACAATCGTTTTAGAGAGTTGGTTTCGGGTGCGACCCAGA	150
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	151	CGCACAGTTTCTAAAATAATGACCGAGAGTGAGCAAATCGTCTACATCAG	200
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	201	CTGCATAGAGGAGGTACGATACTCCTTCGTCAGCCACCTCTCCAAAGCTC	250
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	251	TCCAGCGAAAAGGTGTAAACGATGTCTTCATCGATAGCGATGATTCGCTT	300
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	301	TCCAACGAGTCTCAATCAATGGTCGAGAGAGCTAGGGTTTCTGTTATGAT	350
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	351	TTTACCAGGAAACCGTACGGTATCTCTTGACAAGCTCGTGAAGGTTCTCG	400
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	401	ATTGCCAGAAGAACAAAGATCAAGTGGTGGTTCCGGTGTTGTACGGTGTC	450
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	451	AGATCATCAGAGACCGAATGGCTTAGCGCGCTGGATTCGAAAGGATTCTC	500
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	501	ATCAGTACACCATTCCAGGAAAGAATGTAGTGACTCCCAGCTTGTAAAAG	550
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	551	AGACTGTTAGAGATGTGTATGAGAAGCTCTTTTATATGGAACGAATTGGA	600
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	601	ATATATTCGAAGCTGCTGGAGATTGAGAAAATGATTAACAAGCAACCGTT	650
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	651	GGACATCCGTTGTGTTGGAATTTGGGGTATGCCTGGCATAGGCAAGACTA	700
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	701	CACTTGCTAAAGCAGTCTTTGACCAAATGTCTGGTGAGTTTGATGCTCAT	750

PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	751	TGCTTTATTGAAGACTACACCAAAGCTATTCAAGAGAAGGGTGTTTATTG	800
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	801	TTTGCTGGAGGAACAGTTTTTGAAAGAAAATGCTGGTGCTAGTGGTACCG	850
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	851	TTACGAAATTGAGCTTGCTTAGGGATAGATTAAACAATAAGAGGGTTCTT	900
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	901	GTTGTTCTTGATGATGTCCGCAGTCCTCTGGTTGTGGAGTCTTTTCTTGG	950
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	951	AGGGTTTGACTGGTTTGGTCCCAAAAGTCTAATCATCATAACCTCCAAAG	1000
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1001	ATAAATCGGTGTTTCGCCTTTGTCGAGTCAATCAAATATACGAGGTTCAG	1050
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1051	GGTTTAAATGAGAAAGAGGCTCTTCAACTCTTCTCTTTGTGTGCGTCTAT	1100
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1101	AGACGATATGGCAGAGCAGAATCTCCACGAGGTGTCAATGAAAGTTATTA	1150
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1151	AATATGCTAATGGCCATCCATTAGCTCTCAATCTCTATGGCAGAGAACTG	1200
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1201	ATGGGGAAGAAAAGACCACCAGAAATGGAGATAGCATTCCTCAAACTCAA	1250
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1251	GGAATGTCCTCCAGCTATTTTTGTTGATGCAATCAAGAGCTCGTATGACA	1300
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1301	CACTCAATGACAGGGAAAAAAACATTTTTTTGGACATAGCTTGTTTCTTC	1350
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1351	CAGGGAGAAAATGTTGACTACGTGATGCAACTGCTTGAGGGTTGTGGTTT	1400
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1401	CTTTCCACATGTTGGAATTGATGTTCTTGTGGAGAAGAGTCTGGTGACTA	1450
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1451	TTTCAGAAAACCGAGTGCGGATGCATAACTTGATCCAAGATGTTGGCCGA	1500

PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1501	CAAATAATAAATAGAGAAACAAGACAGACTAAGAGGCGCAGCAGACTGTG	1550
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1551	GGAACCTTGCAGCATCAAATATTTATTAGAAGATAAGGAACAAAACGAAA	1600
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1601	ATGAAGAACAAAAAACAACTTTTGAACGTGCTCAGGTCCCTGAAGAGATC	1650
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1651	GAAGGCATGTTTCTGGACACATCAAACTTAAGTTTTGATATTAAGCATGT	1700
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1701	TGCCTTTGATAATATGTTGAACCTTAGATTGTTCAAGATTTACAGTTCCA	1750
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1751	ATCCTGAAGTCCATCATGTAAACAATTTCCTCAAAGGCTCTCTCAGTTCT	1800
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1801	CTTCCTAATGTGCTAAGACTCCTGCATTGGGAGAACTATCCTCTGCAGTT	1850
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1851	TCTGCCTCAAAATTTTGATCCTATACACCTTGTTGAAATCAACATGCCGT	1900
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1901	ACAGCCAACTTAAGAAACTTTGGGGTGGAACCAAGGACCTGGAGATGTTG	1950
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	1951	AAGACAATCAGGCTTTGTCATTCCCAACAACTAGTTGATATTGACGATCT	2000
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	2001	TTTAAAAGCTCAAAATCTTGAGGTAGTTGATCTCCAAGGCTGTACAAGAC	2050
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	2051	TGCAGAGTTTCCCAGCCACCGGTCAATTGCTACATTTACGAGTTGTAAAT	2100
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	2101	CTCTCAGGTTGCACAGAGATCAAAAGTTTCCCAGAAATTCCCCCAAATAT	2150
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	2151	TGAGACACTGAATCTACAGGGGACTGGTATAATAGAATTACCACTTTCCA	2200
PgNBS-LRR_WRK	1	----------	0
AtWRKY16_cDNA	2201	TTGTTAAGCCAAACTACAGAGAGCTTTTGAATCTTCTAGCTGAAATCCCG	2250


PgNBS-LRR_WRK 1 0
AtWRKY16_cDNA 2251 GGTCTTTCAGGTGTCTCAAACCTTGAGCAAAGTGATCTCAAACCTTTAAC 2300
PgNBS-LRR_WRK 1 0
AtWRKY16_cDNA 2301 AAGCCTGATGAAAATTAGCACATCTTACCAAAATCCTGGCAAGCTTAGTT 2350
PgNBS-LRR_WRK 1 GCAGACG-CTCCCAGACTCGG- 20
AtWRKY16_cDNA 2351 GCTTGGAGCTGAATGATTGTTCTCGTTTGC-GAAGTCTGCCAAACATGGT 2399
PgNBS-LRR_WRK 21 ---TTGGGAACCTGACGGGCCTCCAAACGCTTGA-CTTGACCAGG-TGCT 65
AtWRKY16_cDNA 2400 TAATTTAGAACTT-------CTCAAAGCCCTTGATCTT--TCTGGTTGCT 2440
PgNBS-LRR_WRK 66 CCACTCTGCAGAGGCTCCCAGACT-----CGGTT-----GGGAACCTGAC 105
AtWRKY16_cDNA 2441 --------CAGA-GCT-CGAGACTATCCAGGGTTTCCCACGGAACCTGA- 2479
PgNBS-LRR_WRK 106 GGGCCTCCGAAG-------TCTTTACTTGG----GCAG--------GTGC 136
AtWRKY16_cDNA 2480 ---------AAGAGTTATATCTT--GTTGGCACTGCAGTAAGACAAGTGC 2518
PgNBS-LRR_WRK 137 -------TCC----ACTCTGCAGA------------------------CG 151
AtWRKY16_cDNA 2519 CACAACTTCCTCAAAGTCT--AGAATTCTTTAATGCCCATGGTTGTGTCT 2566
PgNBS-LRR_WRK 152 CTCCCA----GACTCGGTTGG-------GAACCTGAC-GGGC-------- 181
AtWRKY16_cDNA 2567 CTCTCAAATCAATTCGTTTGGACTTCAAGAAGCTTCCTGTGCATTACACA 2616
PgNBS-LRR_WRK 182 ----------------------CTCCA---------AACGCTTGACTTGA 200
AtWRKY16_cDNA 2617 TTTAGTAATTGTTTCGATCTATCTCCACAAGTGGTCAACGATTTTTTAGT 2666
PgNBS-LRR_WRK 201 GCGGGTGTT--CCAATTT---------ACATATGCTGACCAATATTGAG- 238
AtWRKY16_cDNA 2667 GCAGGCGATGGCTAATGTGATTGCAAAACACAT----ACCAAGA--GAGC 2710
PgNBS-LRR_WRK 239 --CATTTGA---GCT--------------------CGTTGGAG------- 256
AtWRKY16_cDNA 2711 GTCATGTCACAGGCTTTTCTCAAAAGACTGTGCAGCGTTCGAGTCGTGAC 2760
PgNBS-LRR_WRK 257 -------------------AATCTTT------------------------ 263
AtWRKY16_cDNA 2761 AGTCAGCAGGAACTCAACAAAACTTTGGCTTTCAGCTTCTGTGCGCCCTC 2810
PgNBS-LRR_WRK 264 --ATGTGCAGCAATGTCCCAAA------CTGCAATGGGGTTCGGAAGTAA 305
AtWRKY16_cDNA 2811 ACATGCGAATCAAAATTCCAAACTTGATCTGCAA---------------- 2844
PgNBS-LRR_WRK 306 TCGAGCAGCTGCGCCA-----ACGACTGGGAGAAGGCTTCATGGAAGCAT 350
AtWRKY16_cDNA 2845 -CCAGGATCTTCTTCAATGACACGACTAG----ATCCTTCTTGGAGGAAC 2889
PgNBS-LRR_WRK 351 ACTCCAG---------------TGATGAAAG----------CGAGTTGG- 374
AtWRKY16_cDNA 2890 ACACTTGTGGGCTTTGCTATGCTGGTGCAAGTCGCATTTTCCGAG--GGT 2937


PgNBS-LRR_WRK 375 -ACTCCAGTGATGAAAGCGA----------GTTGGAAAAT---- 403
AtWRKY16_cDNA 2938 TACT---GTGATGATACTGATTTTGGCATTAGTTGTGTTTGCAAATGGAA 2984
PgNBS-LRR_WRK 404 ATACAAATGGAAG--------CATACTCCGGTGATGAAA-------GCGA 438
AtWRKY16_cDNA 2985 AAACAA--GGAAGGCCACTCTCATA----GGAGA-GAAATAAATTTGC-A 3026
PgNBS-LRR_WRK 439 --GTTGGAAAATATACAAA-----TGGAAGCGA---------- 464
AtWRKY16_cDNA 3027 TTGTTGGGCTTTAGGGAAAGCTGTTGAAAGGGATCATACGTTTGTCTTCT 3076
PgNBS-LRR_WRK 465 --GTTGTTAA---------ATATACTCCATGAA---------- 486
AtWRKY16_cDNA 3077 TTGATGTCAACATGCGTCCAGATAC-CGATGAAGGAAATGACCCCGATAT 3125
PgNBS-LRR_WRK 487 --------ATCGAGTTGT----------AAAACATA--- 504
AtWRKY16_cDNA 3126 CTGGGCTGATTTAGTTGTTTTTGAGTTCTTTCCTGTCAATAAACAGAGAA 3175
PgNBS-LRR_WRK 505 ---CTC----------CAGTGATGAAAG----CGAGT---TGTAA 529
AtWRKY16_cDNA 3176 AGCCTCTAAATGATAGTTGCACAGTGA--CAAGATGTGGAGTCCGTTTAA 3223
PgNBS-LRR_WRK 530 -AATGGGAGTAGACTTC-ATGGAAGCAT----------AC-------- 557
AtWRKY16_cDNA 3224 TAACTGCTGTAAACTGCAATACAAGTATCGAGAATATATCACCAGTTTTG 3273
PgNBS-LRR_WRK 558 TCC---------- 560
AtWRKY16_cDNA 3274 TCCTTGGATCCGATGGAGGTTTCTGGTAATGAAGATGAAGAAGTATTGAG 3323
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY16_cDNA 3324 AGTCAGATATGCTGGTTTACAGGAGATATATAAAGCTTTGTTTCTTTACA 3373
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY16_cDNA 3374 TAGCGGGTTTGTTCAATGACGAGGATGTTGGTTTGGTAGCACCACTTATT 3423
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY16_cDNA 3424 GCTAACATTATTGACATGGACGTTAGTTATGGGCTCAAGGTCTTAGCCTA 3473
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY16_cDNA 3474 TAGGTCTCTCATACGTGTATCTTCCAATGGGGAAATAGTGATGCACTATT 3523
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY16_cDNA 3524 TGCTACGACAAATGGGTAAAGAAATCCTCCATACAGAATCAAAGAAGACT 3573
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY16_cDNA 3574 GACAAATTAGTCGACAATATTCAGAGTTCCATGATCGCAACAAAGGAAAT 3623
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY16_cDNA 3624 CGAGATCACTCGTTCAAAGAGTCGCCGAAAGAACAACAAGGAAAAGAGAG 3673
PgNBS-LRR_WRK 561 ---------- 560

```
AtWRKY16_cDNA    3674 TGGTTTGCGTAGTGGATCGAGGCAGCCGGTCCAGTGACCTATGGGTTTGG   3723
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    3724 CGAAAGTATGGTCAAAAACCCATCAAAAGTTCTCCTTATCCAAGGAGTTA   3773
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    3774 CTATAGATGTGCCAGCTCGAAAGGTTGTTTTGCTAGGAAACAAGTCGAAC   3823
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    3824 GTAGCCGCACTGATCCAAATGTTTCAGTAATTACTTACATCTCTGAGCAT   3873
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    3874 AACCATCCATTCCCCACTCTACGCAATACTCTTGCCGGCTCCACTCGTTC   3923
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    3924 CTCTTCCTCCAAATGCTCAGATGTAACTACTTCTGCCTCATCGACAGTCT   3973
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    3974 CCCAAGACAAAGAAGGACCGGATAAATCCCATTTGCCTTCCTCCCCTGCT   4023
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    4024 TCTCCTCCTTATGCGGCCATGGTGGTTAAGGAGGAGGACATGGAGCAATG   4073
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    4074 GGACAATATGGAGTTCGATGTTGACGTTGAAGAAGATACTTTCATACCCG   4123
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    4124 AATTATTTCCAGAGGATACCTTCGCTGATATGGACAAGCTTGAGGAAAAT   4173
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    4174 TCTCAGACTATGTTTCTCTCTCGCAGAAGCAGCGGAGGCAACATGGAAGC   4223
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    4224 CCAAGGGAAGAACTCTAGTGATGATAGGGAGGTCAATTTACCTAGTAAAA   4273
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    4274 TTCTGAATAGATAGTTACTATTATGCAATGTTAATAATAATCTGTTTGAT   4323
PgNBS-LRR_WRK     561 --------------------------------------------------    560
AtWRKY16_cDNA    4324 TTTTTAACATTTGTTCGGACATCCAAACCTGTGGGACACAATTTATTACT   4373
PgNBS-LRR_WRK     561 ----------------     560
AtWRKY16_cDNA    4374 TTAAAGATAATATATC   4389
```

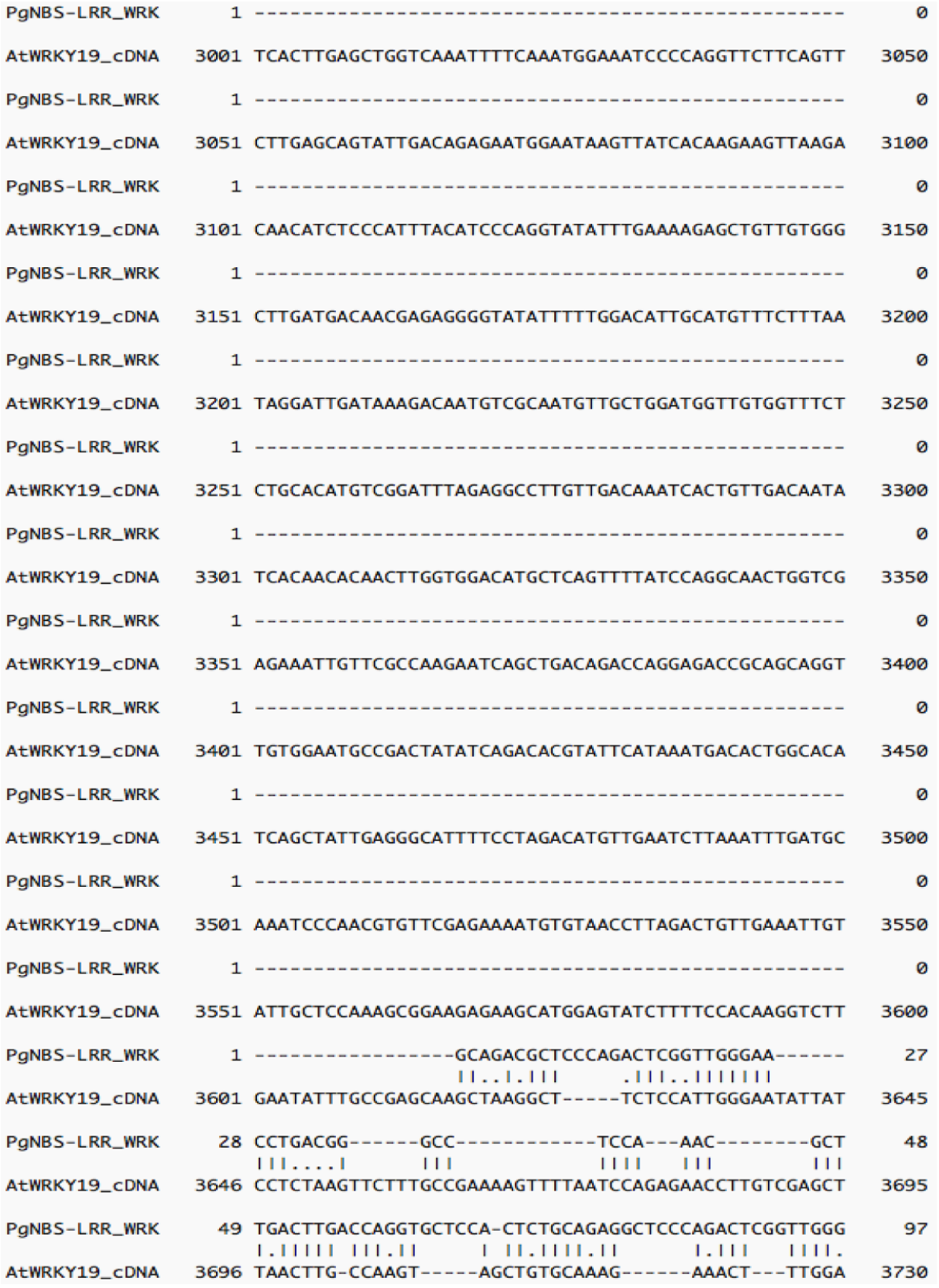
Figure S3.5 Alignment of *PgNBS-LRR/WRKY* and *Arabidopsis WRKY16* conducted in EMBOSS Needle nucleotide alignment (ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html).

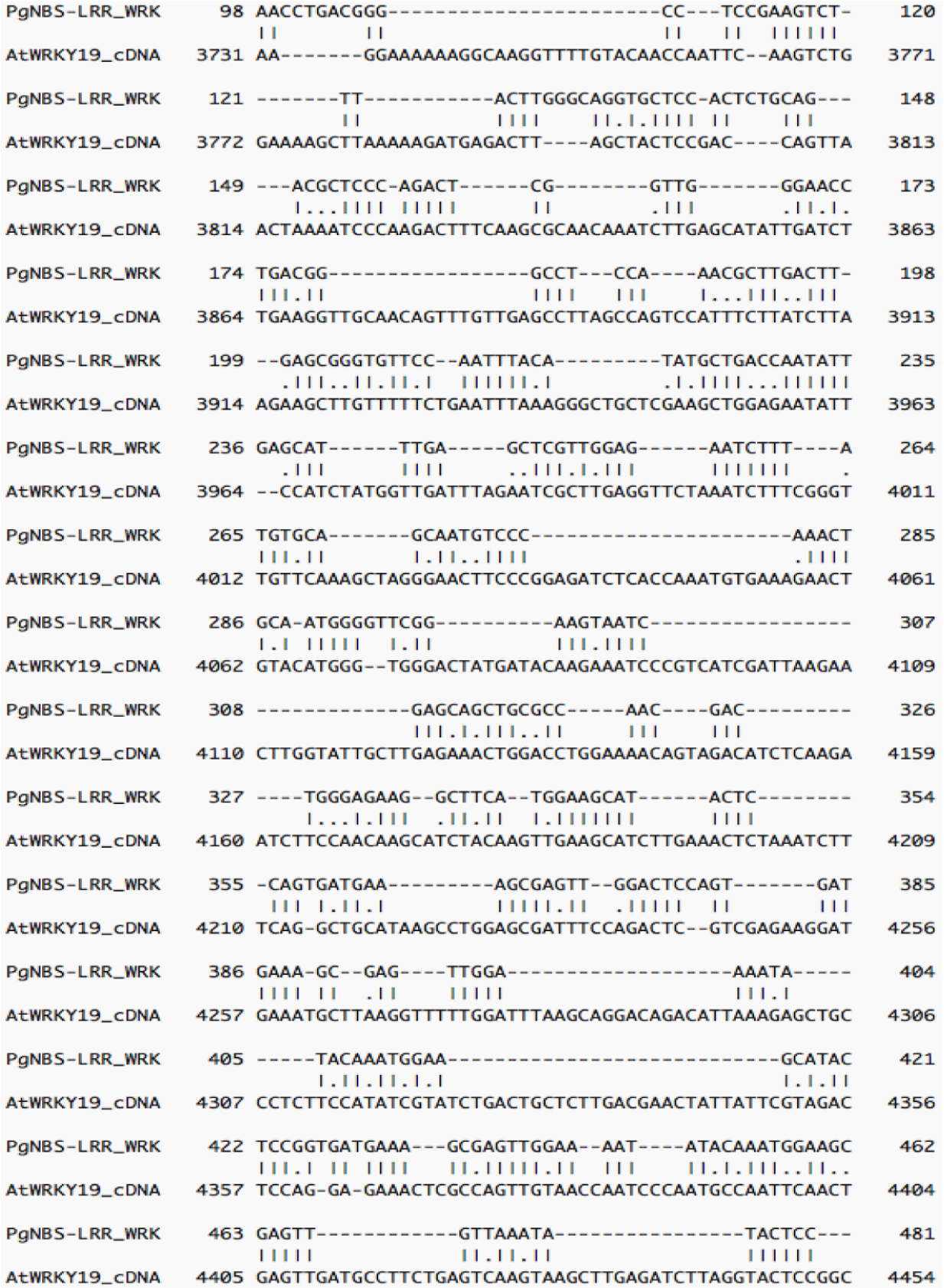
```
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA      1 ATGTCGGAGAAGGAAGAACTTCCGTTGACATTGACGTCCATCGGAGCGGC     50
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA     51 CACCGCGACTAGTGATTATCATCAGAGAGTAGGAAGTTCCGGTGAAGGGA    100
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    101 TTAGTAGCTCGAGTAGTGATGTTGACCCGAGGTTCATGCAGAATAGCCCC    150
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    151 ACGGGTTTGATGATTTCCCAATCGTCGTCGATGTGCACCGTACCGCCTGG    200
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    201 CATGGCAGCAACACCACCAATAAGCTCAGGTTCCGGTTTATCTCAGCAGC    250
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    251 TTAATAATTCTTCTAGTTCCAAGTTATGTCAAGTGGAAGGATGTCAAAAA    300
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    301 GGAGCAAGAGATGCATCTGGTCGTTGCATTTCCCATGGCGGTGGACGTAG    350
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    351 ATGCCAGAAACCTGATTGCCAGAAGGGAGCTGAAGGTAAAACAGTGTACT    400
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    401 GTAAAGCCCACGGAGGTGGTCGCAGATGTGAATATCTTGGATGCACCAAA    450
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    451 GGCGCAGAAGGCAGTACTGATTTTTGTATAGCTCATGGAGGTGGTCGAAG    500
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    501 ATGCAACCATGAAGATTGCACACGATCTGCTTGGGGAAGAACAGAATTCT    550
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    551 GTGTCAAGCACGGTGGAGGAGCGAGATGCAAAACATACGGCTGCGGAAAA    600
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    601 AGCGCTAGTGGTCCTTTGCCATTCTGCCGAGCCCATGGTGGTGGTAAAAA    650
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    651 ATGCAGCCATGAAGATTGCACAGGATTTGCTAGGGGAAGATCAGGACTCT    700
PgNBS-LRR_WRK      1 --------------------------------------------------      0
AtWRKY19_cDNA    701 GTCTCATGCACGGTGGGGGAAAGAGATGCCAAAGAGAGAACTGCACTAAA    750
```

PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	751	AGCGCTGAAGGTCTTTCGGGACTCTGCATATCCCATGGTGGTGGTCGGCG	800
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	801	ATGTCAATCTATTGGATGCACAAAAGGAGCGAAAGGGAGCAAAATGTTCT	850
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	851	GCAAAGCATGCATAACTAAAAGGCCTCTAACGATTGATGGAGGAGGAAAT	900
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	901	ATGGGAGGGGTAACAACAGGTGATGCCTTGAACTATCTCAAAGCTGTGAA	950
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	951	GGACAAGTTTGAAGACAGTGAGAAATATGACACTTTCCTTGAAGTCTTGA	1000
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1001	ATGACTGTAAACATCAGGGAGTTGACACTAGTGGCGTCATAGCCAGATTA	1050
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1051	AAAGATTTGTTCAAGGGCCATGACGACTTACTTTTGGGTTTTAATACCTA	1100
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1101	CTTGTCAAAGGAGTACCAAATAACCATTCTGCCCGAGGATGATTTCCCTA	1150
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1151	TCGATTTTCTTGACAAGGTTGAGGGACCTTATGAAATGACATATCAGCAA	1200
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1201	GCTCAAACAGTTCAAGCCAATGCCAATATGCAACCTCAAACTGAGTACCC	1250
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1251	TTCTTCCTCTGCGGTTCAATCATTTTCATCGGGTCAACCTCAGATCCCCA	1300
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1301	CCTCAGCTCCGGATTCTTCACTACTAGCTAAAAGTAATACCTCAGGTATA	1350
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1351	ACTATCATCGAGCACATGTCACAACAGCCTCTAAATGTTGACAAACAAGT	1400
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1401	TAATGATGGCTATAACTGGCAAAAGTATGGGCAAAAGAAAGTTAAAGGCA	1450
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1451	GCAAGTTTCCTCTAAGCTATTACAAGTGCACATATCTAGGATGTCCTTCC	1500

PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1501	AAGAGGAAGGTTGAGAGATCTCTTGATGGACAAGTAGCAGAAATCGTCTA	1550
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1551	CAAAGATCGACACAATCACGAACCTCCTAACCAAGGAAAAGATGGTAGCA	1600
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1601	CCACATATCTAAGTGGGAGTTCGACACACATCAATTGCATGAGCTCTGAA	1650
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1651	TTGACAGCATCACAGTTTAGCTCCAACAAGACTAAGATAGAGCAACAGGA	1700
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1701	AGCAGCAAGTCTAGCTACGACAATAGAGTACATGTCTGAGGCAAGTGACA	1750
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1751	ATGAAGAAGACAGTAATGGAGAAACTAGTGAGGGAGAGAAAGATGAAGAC	1800
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1801	GAGCCTGAACCAAAGAGAAGAATTACAGAAGTTCAGGTTTCGGAACTAGC	1850
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1851	TGATGCTTCAGATAGAACCGTGAGAGAGCCTAGGGTTATTTTCCAAACAA	1900
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1901	CGAGTGAAGTTGATAATTTAGATGATGGATATAGGTGGCGGAAATATGGA	1950
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	1951	CAGAAAGTTGTTAAAGGGAATCCTTATCCAAGGTTTTCCTCCTCTAAAGA	2000
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2001	TTATGATGTCGTAATCAGATACGGAAGAGCAGATATAAGCAATGAGGATT	2050
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2051	TCATTAGCCATCTTCGTGCTTCCCTCTGCCGGAGAGGGATTTCTGTCTAT	2100
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2101	GAAAAATTTAATGAAGTGGATGCACTTCCAAAATGTAGGGTTTTGATTAT	2150
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2151	AGTATTAACAAGCACATATGTCCCTTCGAACCTCTTAAACATTCTTGAAC	2200
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2201	ACCAACATACAGAGGATCGAGTGGTTTATCCAATTTTCTACAGACTATCA	2250

PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2251	CCATATGATTTTGTCTGTAACAGCAAGAATTATGAGAGATTTTATCTCCA	2300
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2301	AGATGAGCCAAAAAAATGGCAAGCTGCTTTGAAGGAAATAACTCAGATGC	2350
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2351	CTGGCTACACATTGACAGATAAGTCTGAATCTGAACTTATAGATGAGATT	2400
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2401	GTAAGAGATGCTTTAAAGGTGCTATGTTCTGCTGATAAGGTGAACATGAT	2450
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2451	TGGGATGGATATGCAAGTAGAGGAGATTTTGTCACTGCTATGCATTGAGT	2500
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2501	CCCTTGATGTTCGCAGCATTGGTATATGGGGTACAGTTGGTATAGGAAAA	2550
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2551	ACAACCATTGCTGAAGAGATCTTTCGCAAAATCTCTGTCCAATATGAGAC	2600
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2601	CTGTGTCGTCCTTAAGGACCTCCACAAAGAAGTTGAGGTAAAAGGTCACG	2650
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2651	ATGCTGTGAGAGAGAATTTTCTGTCTGAAGTTTTAGAGGTAGAACCTCAT	2700
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2701	GTTATCCGGATATCTGACATTAAAACAAGCTTCTTGAGAAGTCGGCTTCA	2750
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2751	GCGTAAAAGGATCCTTGTTATTCTTGACGATGTGAATGATTACAGAGATG	2800
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2801	TTGACACCTTTTTGGGGACGCTTAACTATTTTGGTCCAGGAAGCAGAATA	2850
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2851	ATCATGACCTCTAGAAATAGACGTGTTTTCGTACTATGTAAAATCGATCA	2900
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2901	TGTCTATGAGGTTAAGCCATTAGATATTCCTAAGTCTCTACTACTTCTTG	2950
PgNBS-LRR_WRK	1	----------	0
AtWRKY19_cDNA	2951	ATCGTGGGACATGTCAAATTGTTTTGTCACCTGAGGTTTACAAGACATTG	3000


PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3001 TCACTTGAGCTGGTCAAATTTTCAAATGGAAATCCCCAGGTTCTTCAGTT 3050
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3051 CTTGAGCAGTATTGACAGAGAATGGAATAAGTTATCACAAGAAGTTAAGA 3100
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3101 CAACATCTCCCATTTACATCCCAGGTATATTTGAAAAGAGCTGTTGTGGG 3150
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3151 CTTGATGACAACGAGAGGGGTATATTTTTGGACATTGCATGTTTCTTTAA 3200
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3201 TAGGATTGATAAAGACAATGTCGCAATGTTGCTGGATGGTTGTGGTTTCT 3250
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3251 CTGCACATGTCGGATTTAGAGGCCTTGTTGACAAATCACTGTTGACAATA 3300
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3301 TCACAACACAACTTGGTGGACATGCTCAGTTTTATCCAGGCAACTGGTCG 3350
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3351 AGAAATTGTTCGCCAAGAATCAGCTGACAGACCAGGAGACCGCAGCAGGT 3400
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3401 TGTGGAATGCCGACTATATCAGACACGTATTCATAAATGACACTGGCACA 3450
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3451 TCAGCTATTGAGGGCATTTTCCTAGACATGTTGAATCTTAAATTTGATGC 3500
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3501 AAATCCCAACGTGTTCGAGAAAATGTGTAACCTTAGACTGTTGAAATTGT 3550
PgNBS-LRR_WRK 1 ---------- 0
AtWRKY19_cDNA 3551 ATTGCTCCAAAGCGGAAGAGAAGCATGGAGTATCTTTTCCACAAGGTCTT 3600
PgNBS-LRR_WRK 1 ----------GCAGACGCTCCCAGACTCGGTTGGGAA------ 27
AtWRKY19_cDNA 3601 GAATATTTGCCGAGCAAGCTAAGGCT-----TCTCCATTGGGAATATTAT 3645
PgNBS-LRR_WRK 28 CCTGACGG------GCC----------TCCA---AAC--------GCT 48
AtWRKY19_cDNA 3646 CCTCTAAGTTCTTTGCCGAAAAGTTTTAATCCAGAGAACCTTGTCGAGCT 3695
PgNBS-LRR_WRK 49 TGACTTGACCAGGTGCTCCA-CTCTGCAGAGGCTCCCAGACTCGGTTGGG 97
AtWRKY19_cDNA 3696 TAACTTG-CCAAGT-----AGCTGTGCAAAG------AAACT---TTGGA 3730


PgNBS-LRR_WRK 98 AACCTGACGGG----------CC---TCCGAAGTCT- 120
AtWRKY19_cDNA 3731 AA-------GGAAAAAAGGCAAGGTTTTGTACAACCAATTC--AAGTCTG 3771
PgNBS-LRR_WRK 121 -------TT----------ACTTGGGCAGGTGCTCC-ACTCTGCAG--- 148
AtWRKY19_cDNA 3772 GAAAAGCTTAAAAAGATGAGACTT----AGCTACTCCGAC----CAGTTA 3813
PgNBS-LRR_WRK 149 ---ACGCTCCC-AGACT------CG--------GTTG-------GGAACC 173
AtWRKY19_cDNA 3814 ACTAAAATCCCAAGACTTTCAAGCGCAACAAATCTTGAGCATATTGATCT 3863
PgNBS-LRR_WRK 174 TGACGG----------GCCT---CCA----AACGCTTGACTT- 198
AtWRKY19_cDNA 3864 TGAAGGTTGCAACAGTTTGTTGAGCCTTAGCCAGTCCATTTCTTATCTTA 3913
PgNBS-LRR_WRK 199 --GAGCGGGTGTTCC--AATTTACA---------TATGCTGACCAATATT 235
AtWRKY19_cDNA 3914 AGAAGCTTGTTTTTCTGAATTTAAAGGGCTGCTCGAAGCTGGAGAATATT 3963
PgNBS-LRR_WRK 236 GAGCAT------TTGA-----GCTCGTTGGAG------AATCTTT----A 264
AtWRKY19_cDNA 3964 --CCATCTATGGTTGATTTAGAATCGCTTGAGGTTCTAAATCTTTCGGGT 4011
PgNBS-LRR_WRK 265 TGTGCA-------GCAATGTCCC----------AAACT 285
AtWRKY19_cDNA 4012 TGTTCAAAGCTAGGGAACTTCCCGGAGATCTCACCAAATGTGAAAGAACT 4061
PgNBS-LRR_WRK 286 GCA-ATGGGGTTCGG----------AAGTAATC---------- 307
AtWRKY19_cDNA 4062 GTACATGGG--TGGGACTATGATACAAGAAATCCCGTCATCGATTAAGAA 4109
PgNBS-LRR_WRK 308 -------------GAGCAGCTGCGCC-----AAC----GAC--------- 326
AtWRKY19_cDNA 4110 CTTGGTATTGCTTGAGAAACTGGACCTGGAAAACAGTAGACATCTCAAGA 4159
PgNBS-LRR_WRK 327 ----TGGGAGAAG--GCTTCA--TGGAAGCAT------ACTC-------- 354
AtWRKY19_cDNA 4160 ATCTTCCAACAAGCATCTACAAGTTGAAGCATCTTGAAACTCTAAATCTT 4209
PgNBS-LRR_WRK 355 -CAGTGATGAA---------AGCGAGTT--GGACTCCAGT-------GAT 385
AtWRKY19_cDNA 4210 TCAG-GCTGCATAAGCCTGGAGCGATTTCCAGACTC--GTCGAGAAGGAT 4256
PgNBS-LRR_WRK 386 GAAA-GC--GAG----TTGGA-------------------AAATA----- 404
AtWRKY19_cDNA 4257 GAAATGCTTAAGGTTTTTGGATTTAAGCAGGACAGACATTAAAGAGCTGC 4306
PgNBS-LRR_WRK 405 -----TACAAATGGAA----------------------------GCATAC 421
AtWRKY19_cDNA 4307 CCTCTTCCATATCGTATCTGACTGCTCTTGACGAACTATTATTCGTAGAC 4356
PgNBS-LRR_WRK 422 TCCGGTGATGAAA---GCGAGTTGGAA--AAT----ATACAAATGGAAGC 462
AtWRKY19_cDNA 4357 TCCAG-GA-GAAACTCGCCAGTTGTAACCAATCCCAATGCCAATTCAACT 4404
PgNBS-LRR_WRK 463 GAGTT------------GTTAAATA----------------TACTCC--- 481
AtWRKY19_cDNA 4405 GAGTTGATGCCTTCTGAGTCAAGTAAGCTTGAGATCTTAGGTACTCCGGC 4454


PgNBS-LRR_WRK 482 -----ATGAAATCGAGTTGT----------AAAAC-------AT 503
AtWRKY19_cDNA 4455 AGATAACGAAGT--AGTTGTTGGTGGTACGGTAGAGAAAACCCGTGGTAT 4502
PgNBS-LRR_WRK 504 -------ACTCCAG-------TGATGAA----AGCGAGT-TGTAA----- 529
AtWRKY19_cDNA 4503 TGAACGAACGCC-GACTATTTTGGTGAAGTCGAGAGAGTATCTAATTCCC 4551
PgNBS-LRR_WRK 530 ----------AATGGG----------AGT 538
AtWRKY19_cDNA 4552 GATGATGTTGTGGCGGTTGGTGGTGATATTAAGGGGCTAAGACCACCAGT 4601
PgNBS-LRR_WRK 539 AGACTTC----------ATGGAA-----GCATACTCC--------- 560
AtWRKY19_cDNA 4602 --ACTTCAGCTCCAACCAGCAATGAAACTATCTCATATTCCTCGAGGATC 4649
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 4650 AACTTGGGATTTCGTTACGCATTTCGCTCCACCTGAAACAGTTGCGCCGC 4699
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 4700 CGAGTTCCTCTTCAGAAGCCAGGGAAGAGGAAGTGGAAACGGAAGAGACG 4749
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 4750 GGAGCTATGTTTATCCCATTGGGGGATAAGGAGACATGCTCATTCACTGT 4799
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 4800 AAACAAGGGTGACTCCTCAAGGACAATATCTAATACGTCGCCGATTTATG 4849
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 4850 CCTCCGAAGGATCTTTCATCACGTGTTGGCAGAAGGGTCAACTTCTGGGA 4899
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 4900 CGAGGATCATTAGGGTCCGTATATGAAGGCATTTCAGCAGACGGGGACTT 4949
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 4950 CTTTGCTTTCAAGGAAGTTTCACTACTTGATCAGGGAAGTCAGGCACATG 4999
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 5000 AATGGATACAACAAGTCGAGGGGGGGATTGCGCTACTTAGTCAGCTTCAG 5049
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 5050 CATCAGAATATCGTGCGATATCGTGGCACAACTAAGGACGAGTCGAATTT 5099
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 5100 GTACATTTTTCTTGAACTTGTAACCCAAGGGTCCCTTCGAAAACTCTACC 5149
PgNBS-LRR_WRK 561 ---------- 560
AtWRKY19_cDNA 5150 AAAGAAACCAGCTTGGGGACTCTGTAGTCTCCTTATACACAAGACAGATT 5199

PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5200	CTTGATGGATTGAAATATCTCCACGATAAAGGTTTTATACACAGGAACAT	5249
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5250	TAAATGTGCAAATGTATTGGTGGACGCTAATGGAACAGTTAAACTTGCAG	5299
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5300	ATTTTGGATTGGCTAAGGTAATGTCCCTCTGGCGAACTCCGTATTGGAAT	5349
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5350	TGGATGGCTCCAGAGGTTATTCTTAACCCGAAGGATTATGATGGTTATGG	5399
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5400	AACTCCAGCTGATATATGGAGCCTTGGGTGTACTGTGCTAGAAATGTTGA	5449
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5450	CTGGTCAGATTCCCTACTCCGATCTGGAAATCGGTACAGCCTTGTATAAC	5499
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5500	ATTGGAACGGGTAAGCTTCCGAAAATACCTGATATTCTATCGCTAGACGC	5549
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5550	CCGGGATTTCATACTTACGTGTCTCAAAGTGAACCCGGAAGAGCGGCCAA	5599
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5600	CTGCAGCTGAGCTGCTTAACCATCCATTTGTGAATATGCCATTACCATCC	5649
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5650	TCGGGCTCAGGTTCAGTATCTTCGCTCCTCCGTGGATGAGGCTAATTTTA	5699
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5700	GAAGATTGCCTTCGATAGGTGAAACTTTGGGCTTTGTTGTCATGAGAATA	5749
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5750	TATTAGAATCAAAAGTTGTATATATATTCTAACATATGCCCTAATATGCT	5799
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5800	TTGCACTAAAATCGTTTGAATCTTCTGTTCCATACTTCACGTTGACAAAT	5849
PgNBS-LRR_WRK	561	----------	560
AtWRKY19_cDNA	5850	CACAATTTTTTTTATGTCTTTTGTTTCTTTTTTTTTGCTGAA	5891

Figure S3.6 Alignment of *PgNBS-LRR/WRKY* and *Arabidopsis WRKY19* conducted in EMBOSS Needle nucleotide alignment (ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html).

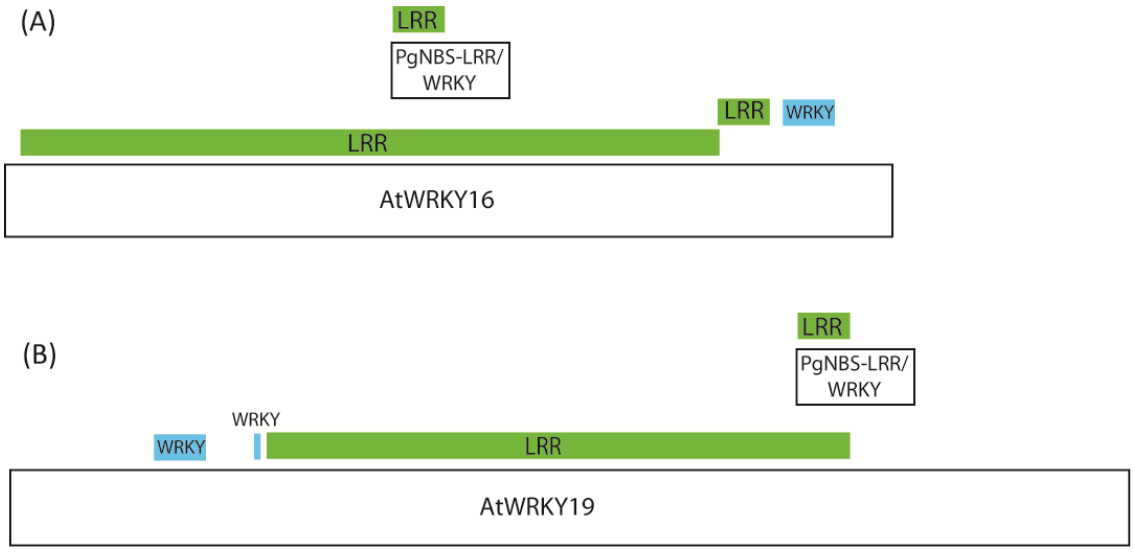


Figure S3.7 Overview of LRR and WRKY domains in *PgNBS-LRR/WRKY*, *Arabidopsis WRKY16* and *Arabidopsis WRKY19*. Domains were determined using NCBI domain finder (ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Table S3.1 Similarity of white spruce contigs promoters and coding sequences of *PgSAL1* and *PgSAL5*. Queries were submitted to PG29 v4.0 and WS77111 v1.0 assemblies through ConGenie (congenie.org) BLASTN. Number of overlapping nucleotides and percent identity determined by EMBOSS Needle nucleotide alignment (ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html). Top contigs were selected based on preliminary alignments performed in EMBOSS.

Query	PG29 Assembly				
	Contig	Average e-value from ConGenie BLAST (lowest)	Average identity (average similarity) from ConGenie BLAST	# of overlapping nucleotides (bp), out of total nucleotides (bp)	% identity
SAL1 cDNA 962 bp	Pg-01r141201s2137277	5.49e-9 (7.84e-147)	92.91 % (92.91 %)	699/2292	30.5
	Pg-01r141201s2554914	1.27e-111 (7.84e-147)	91.61 % (91.61 %)	767/3173	24.2
	Pg-01r141201s2305140	6.98e-101 (7.84e-147)	90.94 % (90.94 %)	719/35701	2

SAL1 promoter (containing cloned region of UTR and cDNA) 1099 bp	Pg-01r141201s2137277	0.00e+0 (0.00e+0)	94.47 % (94.47 %)	1048/2184	48
	Pg-01r141201s2137278	9.81e-45 (0.00e+0)	93.16 % (93.16 %)	941/1524	61.7
	Pg-01r141201s2498919	1.71e-11 (0.00e+0)	94.26 % (94.26 %)	686/3018	22.7
	Pg-01r141201s2356730	8.98e-19 (0.00e+0)		787/13508	5.8
SAL5 cDNA 1005 bp	Pg-01r141201s2137278	7.70e-14 (7.30e-101)	95.54 % (95.54 %)	549/1745	31.5
	Pg-01r141201s2554914	7.84e-70 (7.30e-101)	92.78 % (92.78 %)	722/3216	22.5
	Pg-01r141201s2356730	1.14e-8 (7.30e-101)	87.63 % (87.63 %)	363/13935	2.6
	Pg-01r141201s2305140	4.52e-8 (7.30e-101)	87.44 % (87.44 %)	718/35695	2
	Pg-01r141201s2880671	4.35e-5 (7.30e-101)	87.16 % (87.16 %)	691/3750	18.4
	Pg-01r141201s2577914	1.36e-40 (7.30e-101)	88.55 % (88.55 %)	639/117492	0.5
	Pg-01r141201s2613660	3.32e-38 (7.30e-101)	100.00 % (100.00 %)	673/10911	6.2
	Pg-01r141201s0119707	2.79e-6 (7.30e-101)	89.14 % (89.14 %)	287/2452	11.7
	Pg-	8.10e-36 (7.30e-101)	87.80 % (87.80 %)	681/5196	13.1

	01r141201s2765746				
Putative *SAL5* promoter (containing cloned region of UTR and cDNA) 1932 bp	Pg-01r141201s2084611	3.03e-143 (0.00e+0)	93.48 % (93.48 %)	1333/47202	2.8
	Pg-01r141201s1228192	4.09e-167 (0.00e+0)	95.01 % (95.01 %)	854/2162	39.5
	Pg-01r141201s2214380	9.97e-165 (0.00e+0)	94.87 % (94.87 %)	1141/2581	44.2
	Pg-01r141201s2513257	6.61e-132 (0.00e+0)	94.34 % (94.34 %)	1231/2987	41.2
	Pg-01r141201s2765746	2.41e-179 (0.00e+0)	94.69 % (94.69 %)	1471/5209	28.2
	Pg-01r141201s1684675	3.52e-155 (0.00e+0)	95.26 % (95.26 %)	770/2200	35
	Pg-01r141201s2814414	8.18e-167 (0.00e+0)	94.75 % (94.75 %)	867/2229	38.9
		WS77111 Assembly			

	Contig	Average e-value from ConGenie BLAST (lowest)	Average identity (average similarity) from ConGenie BLAST	# of overlapping nucleotides (bp), out of total nucleotides (bp)	% identity
SAL1 cDNA 962 bp	Pg-02r141203s2614426	1.41e-23 (2.83e-23)	94.38 % (94.38 %)	720/39358	1.8
	Pg-02r141203s0882372	4.81e-11 (5.19e-145)	91.91 % (91.91 %)	793/4373	18.1
	Pg-02r141203s2747360	8.70e-113 (5.19e-145)	92.62 % (92.62 %)	707/72371	1
	Pg-02r141203s2388593	1.36e-111 (5.19e-145)	91.61 % (91.61 %)	444/17123	2.6
	Pg-02r141203s2902978	1.83e-7 (5.19e-145)	90.37 % (90.37 %)	732/38308	0.2
	Pg-02r141203s0871761	1.21e-102 (5.19e-145)	90.15 % (90.15 %)	654/1105	59.2
	Pg-02r141203s0652671	1.22e-11 (5.19e-145)	98.09 % (98.09 %)	675/2033	33.2
	Pg-02r141203s2509160	3.33e-14 (5.19e-145)	99.31 % (99.31 %)	739/35118	2.1
SAL1 promoter (containing cloned region of UTR and cDNA) 1099 bp	Pg-02r141203s0882372	0.00e+0 (0.00e+0)	97.43 % (97.43 %)	1079/4353	24.8
	Pg-02r141203s2614426	3.18e-36 (0.00e+0)	93.10 % (93.10 %)	750/3766	19.9
	Pg-02r141203s2828653	9.66e-15 (0.00e+0)	96.39 % (96.39 %)	736/70040	1.1
Putative	Pg-	5.79e-13 (3.21e-103)	100.00 % (100.00	706/5580	12.7

SAL5 cDNA 1005 bp	02r141203s2780164		%)		
	Pg-02r141203s2391889	5.81e-77 (3.21e-103)	94.33 % (94.33 %)	731/16958	4.3
	Pg-02r141203s0871761	3.45e-72 (3.21e-103)	93.30 % (93.30 %)	545/1350	40.4
	Pg-02r141203s2614426	2.23e-15 (3.21e-103)	91.13 % (91.13 %)	677/3812	17.8
	Pg-02r141203s2388593	8.41e-70 (3.21e-103)	92.78 % (92.78 %)	684/16816	4.1
	Pg-02r141203s2747360	1.99e-10 (3.21e-103)	88.24 % (88.24 %)	738/72353	1
	Pg-02r141203s2902978	2.34e-5 (3.21e-103)	89.29 % (89.29 %)	712/383863	0.2
	Pg-02r141203s3008384	2.56e-48 (3.21e-103)	88.14 % (88.14 %)	744/2836	26.2
	Pg-02r141203s2920848	1.85e-4 (3.21e-103)	87.39 % (87.39 %)	772/79115	1
Putative *SAL5* promoter (containing cloned region of UTR and cDNA) 1932 bp	Pg-02r141203s2464582	0.00e+0 (0.00e+0)	93.96 % (93.96 %)	1390/208246	0.7
	Pg-02r141203s2713577	0.00e+0 (0.00e+0)	94.11 % (94.11 %)	1416/32715	4.3
	Pg-02r141203s2958320	3.79e-151 (0.00e+0)	94.21 % (94.21 %)	1225/5389	22.7
	Pg-02r141203s2699671	0.00e+0 (0.00e+0)	93.40 % (93.40 %)	1260/3592	35.1

	Pg-02r141203s1417763	2.61e-162 (0.00e+0)	94.90 % (94.90 %)	1297/3804	34.1
	Pg-02r141203s1829115	6.36e-160 (0.00e+0)	94.77 % (94.77 %)	880/2242	39.3
	Pg-02r141203s2913339	1.51e-137 (0.00e+0)	93.27 % (93.27 %)	1641/15863	10.3
	Pg-02r141203s3274604	4.23e-164 (0.00e+0)	95.18 % (95.18 %)	1064/2159	49.3
	Pg-02r141203s0593557	6.36e-160 (0.00e+0)	94.58 % (94.58 %)	1065/2163	49.2
	Pg-02r141203s2554061	3.78e-155 (0.00e+0)	93.26 % (93.26 %)	1333/5085	26.2

Table S3.2 Query contigs and cloned portions of genomic DNA to BLAST. Contigs denoted by "PG-####" were submitted to a MEGABLASTn discontiguous search. Cloned portions of the *SAL5* cDNA and /or cloned portion of UTR submitted to MEGABLASTn discontiguous BLAST only against *Picea glauca* sequences. Output of the BLAST searches are listed below.

Original query used to search ConGenie against the PG29 of WS77111 assembly (If applicable)	**Query**					
***SAL1* cDNA query PG29 assembly**	**Pg-01r141201s2137277**					
Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03605_C12 mRNA sequence	547	1211	37%	8.00E-154	92%	BT114920.1
Picea glauca clone GQ03232_K15 mRNA sequence	399	704	30%	3.00E-109	83%	BT111101.1
Picea glauca clone GQ03806_I20 mRNA sequence	390	1369	41%	1.00E-106	82%	BT116779.1
Picea glauca clone GQ02822_N14 mRNA sequence	279	784	30%	6.00E-73	92%	BT105463.1
Picea glauca clone GQ03707_I04 mRNA sequence	257	399	13%	2.00E-66	89%	BT115854.1
Picea glauca clone	192	192	6%	7.00E-47	92%	BT115613.1

GQ03702_K12 mRNA sequence						
Picea glauca clone GQ04008_C02 mRNA sequence	132	441	20%	6.00E-29	86%	BT118602.1
Picea glauca clone GQ0012_K17 mRNA sequence	57.2	57.2	6%	4.00E-06	70%	BT100378.1
Picea glauca clone GQ02817_J10 mRNA sequence	53.6	53.6	5%	4.00E-05	70%	BT105004.1
Picea glauca clone GQ0204_E19 mRNA sequence	53.6	53.6	7%	4.00E-05	70%	BT102975.1
Picea glauca clone GQ02810_C03 mRNA sequence	44.6	44.6	6%	0.022	67%	BT104415.1
Picea glauca clone GQ0063_K04 mRNA sequence	42.8	42.8	5%	0.078	68%	BT101011.1

***SAL1* cDNA query WS77111 assembly**	**Pg-02r141203s0882372**					
Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03605_C12 mRNA sequence	551	637	10%	1.00E-154	92%	BT114920.1
Picea glauca clone GQ03232_K15 mRNA sequence	405	405	8%	1.00E-110	83%	BT111101.1
Picea glauca clone GQ03806_I20 mRNA	396	799	11%	6.00E-108	82%	BT116779.1

sequence						
Picea glauca clone GQ03224_F06 mRNA sequence	333	333	6%	6.00E-89	88%	BT110609.1
Picea glauca cultivar PG29 clone BAC PGB02 3-carene synthase gene, complete cds, complete sequence	284	284	5%	3.00E-74	84%	FJ60917

***SAL5* cDNA query, PG29 assembly**		**Pg-01r141201s2356730**				
Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03619_H08 mRNA sequence	462	841	6%	2.00E-128	80%	BT115517.1
Picea glauca clone GQ03605_C12 mRNA sequence	457	457	2%	7.00E-127	86%	BT114920.1
Picea glauca clone GQ03232_K15 mRNA sequence	383	383	2%	1.00E-104	81%	BT111101.1
Picea glauca clone GQ03806_I20 mRNA sequence	259	259	3%	3.00E-67	73%	BT116779.1
Picea glauca clone GQ03707_I04 mRNA sequence	253	253	1%	1.00E-65	89%	BT115854.1
Picea glauca clone GQ02822_N14 mRNA sequence	251	251	1%	5.00E-65	89%	BT105463.1
Picea glauca clone	179	179	1%	3.00E-43	90%	BT115613.1

GQ03702_K12 mRNA sequence						
Picea glauca clone GQ04008_C02 mRNA sequence	159	159	1%	2.00E-37	84%	BT118602.1
Picea glauca clone BAC PGB09 (-)-ent-kaurene synthase gene, complete cds	120	209	1%	2.00E-25	80%	GU059905.1

***SAL5* cDNA query PG29 assembly**	**Pg-01r141201s0119707**					
Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03605_C12 mRNA sequence	408	408	20%	4.00E-112	84%	BT114920.1
Picea glauca clone GQ03232_K15 mRNA sequence	329	329	20%	3.00E-88	79%	BT111101.1
Picea glauca clone GQ03707_H11 mRNA sequence	286	286	10%	3.00E-75	95%	BT115852.1
Picea glauca clone GQ03619_H08 mRNA sequence	269	269	18%	3.00E-70	76%	BT115517.1
Picea glauca clone GQ0165_L23 mRNA sequence	242	242	11%	4.00E-62	85%	BT102074.1
Picea glauca clone GQ0026_J17 mRNA sequence	242	242	11%	4.00E-62	87%	BT100589.1
Picea glauca clone GQ03806_I20 mRNA	210	301	14%	2.00E-52	87%	BT116779.1

sequence						
Picea glauca clone GQ02822_N14 mRNA sequence	210	210	9%	2.00E-52	87%	BT105463.1
Picea glauca clone GQ03707_I04 mRNA sequence	208	208	9%	7.00E-52	86%	BT115854.1
Picea glauca clone GQ04008_C02 mRNA sequence	150	150	8%	2.00E-34	83%	BT118602.1
Picea glauca cultivar PG29 clone BAC PGB02 3-carene synthase gene, complete cds, complete sequence	150	150	7%	2.00E-34	87%	FJ609174.2

***SAL5* cDNA query WS77111 assembly**		**Pg-02r141203s2780164**				
Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03605_C12 mRNA sequence	551	637	10%	1.00E-154	92%	BT114920.1
Picea glauca clone GQ03232_K15 mRNA sequence	405	405	8%	1.00E-110	83%	BT111101.1
Picea glauca clone GQ03806_I20 mRNA sequence	396	799	11%	6.00E-108	82%	BT116779.1
Picea glauca clone GQ03224_F06 mRNA sequence	333	333	6%	6.00E-89	88%	BT110609.1
Picea glauca cultivar PG29	284	284	5%	3.00E-	84%	FJ609174.2

clone BAC PGB02 3-carene synthase gene, complete cds, complete sequence				74		
Picea glauca clone GQ02822_N14 mRNA sequence	282	443	8%	1.00E-73	92%	BT105463.1
Picea glauca clone GQ0204_I03 mRNA sequence	280	280	6%	3.00E-73	83%	BT102990.1
Picea glauca clone GQ03707_I04 mRNA sequence	262	262	4%	9.00E-68	90%	BT115854.1
Picea glauca 2S albumin (pgi2S) pseudogene	210	210	5%	5.00E-52	81%	U92078.1

***SAL5* cDNA query WS77111 assembly**	**Pg-02r141203s3008384**					
Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03605_C12 mRNA sequence	426	426	14%	2.00E-117	84%	BT114920.1
Picea glauca clone GQ03232_K15 mRNA sequence	347	347	14%	2.00E-93	80%	BT111101.1
Picea glauca clone GQ03806_I20 mRNA sequence	246	246	7%	4.00E-63	88%	BT116779.1
Picea glauca clone GQ02822_N14 mRNA sequence	242	242	7%	5.00E-62	88%	BT105463.1

Picea glauca clone GQ03707_I04 mRNA sequence	239	239	7%	7.00E-61	87%	BT115854.1
Picea glauca clone GQ03702_K12 mRNA sequence	179	179	4%	5.00E-43	90%	BT115613.1
Picea glauca clone GQ04008_C02 mRNA sequence	147	147	5%	3.00E-33	83%	BT118602.1
Picea glauca clone GQ03709_H13 mRNA sequence	82.4	82.4	2%	1.00E-13	91%	BT115958.1
Picea glauca clone GQ0201_K10 mRNA sequence	78.8	78.8	2%	1.00E-12	85%	BT102849.1
Picea glauca clone GQ04107_C21 mRNA sequence	77	77	2%	5.00E-12	91%	BT119559.1

Putative *SAL5* promoter containing cloned cDNA portion query PG29 assembly	**Pg-01r141201s2765746**					
Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03209_C03 mRNA sequence	412	412	5%	9.00E-114	96%	BT109496.1
Picea glauca clone GQ03806_I20 mRNA sequence	262	262	5%	1.00E-68	80%	BT116779.1
Picea glauca clone GQ03232_K15 mRNA	259	361	5%	1.00E-67	88%	BT111101.1

sequence						
Picea glauca clone GQ02822_N14 mRNA sequence	259	259	6%	1.00E-67	78%	BT105463.1
Picea glauca clone GQ03401_E02 mRNA sequence	219	219	2%	1.00E-55	96%	BT113201.1

-	**Putative *SAL5* promoter containing cloned cDNA portion**					
Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03806_I20 mRNA sequence	235	235	11%	6.00E-60	83%	BT116779.1
Picea glauca clone GQ02822_N14 mRNA sequence	219	219	13%	4.00E-55	79%	BT105463.1
Picea glauca clone GQ03232_K15 mRNA sequence	210	316	10%	2.00E-52	93%	BT111101.1
Picea glauca clone GQ03605_C12 mRNA sequence	170	170	7%	2.00E-40	86%	BT114920.1
Picea glauca clone GQ03707_I04 mRNA sequence	165	165	7%	8.00E-39	86%	BT115854.1
Picea glauca clone GQ04008_C02 mRNA sequence	82.4	82.4	3%	8.00E-14	85%	BT118602.1

-	**Putative *SAL5* only cloned UTR portion and cloned cdna**

Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03806_I20 mRNA sequence	235	309	100%	6.00E-61	83%	BT116779.1
Picea glauca clone GQ02822_N14 mRNA sequence	212	212	83%	7.00E-54	86%	BT105463.1
Picea glauca clone GQ03232_K15 mRNA sequence	210	316	90%	2.00E-53	93%	BT111101.1
Picea glauca clone GQ03605_C12 mRNA sequence	170	170	64%	2.00E-41	86%	BT114920.1
Picea glauca clone GQ03707_I04 mRNA sequence	165	165	64%	9.00E-40	86%	BT115854.1
Picea glauca clone GQ03702_K12 mRNA sequence	80.6	80.6	38%	3.00E-14	82%	BT115613.1
Picea glauca clone GQ04008_C02 mRNA sequence	69.8	69.8	28%	5.00E-11	86%	BT118602.1
Picea glauca clone GQ02902_K19 mRNA sequence	39.2	39.2	10%	0.088	96%	BT106085.1
Picea glauca clone GQ0205_L18 mRNA sequence	35.6	35.6	8%	1.1	100%	BT103051.1

-	**Putative *SAL5* cloned UTR only**

Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03232_K15 mRNA sequence	105	105	63%	2.00E-22	100%	BT111101.1
Picea glauca clone GQ03806_I20 mRNA sequence	73.4	73.4	71%	1.00E-12	85%	BT116779.1
Picea glauca clone GQ04008_C02 mRNA sequence	69.8	69.8	69%	2.00E-11	86%	BT118602.1
Picea glauca clone GQ02822_N14 mRNA sequence	55.4	55.4	59%	4.00E-07	84%	BT105463.1
Picea glauca clone GQ0205_L18 mRNA sequence	35.6	35.6	20%	0.35	100%	BT103051.1
Picea glauca clone GQ02816_C17 mRNA sequence	31.9	31.9	24%	4.2	91%	BT104878.1

-	**Putative *SAL5* cloned cDNA portion only**					
Description	Max score	Total score	Query cover	E value	Ident	Accession
Picea glauca clone GQ03232_K15 mRNA sequence	192	192	100%	3.00E-48	93%	BT111101.1
Picea glauca clone GQ03806_I20 mRNA sequence	176	176	100%	3.00E-43	90%	BT116779.1
Picea glauca clone GQ02822_N14 mRNA	158	158	100%	7.00E-38	87%	BT105463.1

sequence						
Picea glauca clone GQ03605_C12 mRNA sequence	152	152	100%	3.00E-36	86%	BT114920.1
Picea glauca clone GQ03707_I04 mRNA sequence	149	149	100%	4.00E-35	85%	BT115854.1
Picea glauca clone GQ03702_K12 mRNA sequence	80.6	80.6	64%	2.00E-14	82%	BT115613.1
Picea glauca clone GQ02902_K19 mRNA sequence	39.2	39.2	17%	0.048	96%	BT106085.1
Picea glauca clone GQ03603_F04 mRNA sequence	31.9	31.9	12%	7.1	100%	BT114832.1

Table S3.3 Sequence similarity of yeast one-hybrid PgMYB1 to previously identified PgMYB genes. MYB sequences were obtained from Bedon *et al.* (2007), and alignments and sequence similarities were conducted in EMBOSS Needle nucleotide alignment.

MYB Gene	NCBI Accession #	Amino Acid Sequence Similarity to Y1H *PgMYB* (%)
PgMYB1	EF601064.1	98.5
PgMYB2	EF601065.1	30.8
PgMYB3	EF601066.1	31.9
PgMYB4	EF601067.1	37.6
PgMYB5	EF601068.1	36.3
PgMYB6	EF601069.1	29.4
PgMYB7	EF601070.1	29.6
PgMYB8	EF601071.1	30.4
PgMYB9	EF601072.1	24.4
PgMYB10	EF601073.1	32.9
PgMYB11	EF601074.1	27.2
PgMYB12	EF601075.1	35.6
PgMYB13	EF601076.1	27.9

4.0 Conclusions

Dormancy related research to date has predominantly focused on angiosperm species, with some progress in Norway spruce (*Picea abies* (L.) H. Karst, *Pa*) and white spruce (*Picea glauca* (Moench.) Voss, *Pg*; Gyllenstrand *et al.* 2007, Holefors *et al.* 2009, Karlgren *et al.* 2011, Asante *et al.* 2011, El Kayal *et al.* 2011, Karlgren *et al.* 2013, Galindo-Gonzalez *et al.* 2012, Galindo-Gonzalez *et al.* 2015, Hamilton *et al*. 2016, Opseth *et al.* 2016). Our experiments aimed to elucidate the regulatory elements that play a role in the processes involved in the transition from active growth to dormancy. We used the well-established *CONSTANS/FLOWERING LOCUS T* (*CO/FT*) regulatory module (Koorneef *et al.* 1991, Böhlenius *et al.* 2006) to frame our initial hypotheses that *PgSVP*-like genes may play a similar role to the angiosperm orthologs of *SHORT VEGETATIVE PHASE* (*SVP*) that have been shown to participate in the initiation of bud formation, growth cessation, and/or dormancy induction (Li *et al.* 2009, Jiménez *et al.* 2009, Wu *et al.* 2011). Our objectives in this study were to (1) identify white spruce genes that may share functional conservation with angiosperm *SVP* genes through phylogenetic analysis, (2) to determine if these white spruce *SVP*-like genes displayed distinct or similar transcriptional profiles across the stages of bud development, and (3) investigate the upstream regulatory pathways that may control white spruce *SVP*-like genes by employing yeast one-hybrid and identifying conserved promoter motifs.

To investigate possible functions, we identified candidate *PgSVP*-like genes through a robust phylogenetic analysis of angiosperm and gymnosperm MADS-box genes. We determined *PgSVP*-like genes share a common ancestor with angiosperm *SVP/AGAMOUS-LIKE 24* (*AGL24*) genes, and thereby white spruce genes have been named *SVP/AGL24* (*SAL*). Transcript profiles of the seven *PgSAL* genes identified to be sister to angiosperm *SVP/AGL24* were

examined across developmental stages of bud formation in terminal buds under short day (SD) and long day (LD) conditions to hypothesize potential functions of these genes in white spruce. Transcriptional evidence suggests *PgSAL* genes may have roles across the different stages of bud development, including early (*PgSAL1-4, 5*) to mid (*PgSAL3, 7*) and late-phase development (*PgSAL*6), indicating a divergence in function from one another.

To provide additional evidence for possible functional roles of *PgSAL*, we investigated the upstream regulatory pathway by analysing the promoter sequences cloned from two *PgSAL* genes, *PgSAL1* and *PgSAL5*. These genes were chosen because they were shown to be homologous to the angiosperm *SVP/AGL24* clade, and transcriptional data suggested *PgSAL1* and *PgSAL5* may function in the early stages of bud development. Furthermore, we were able to clone substantial fragments (920 bp or greater) of the upstream regulatory regions of *PgSAL1* and *PgSAL5*. Based on *in silco* analysis of the cloned regions of promoters against the *SAL1* and *SAL5* cDNA sequences, it was determined the *SAL1* promoter is most likely upstream of the *SAL1* gene, whereas were cannot conclusively demonstrate with currently available genomic resources that the *SAL5* promoter is upstream the *SAL5* gene. For this reason, have named this promoter the "putative *SAL5* promoter", until further experimentation or resources are available to establish this relationship between promoter and coding sequence. Although we cannot conclusively confirm the identity of the putative SAL5 promoter, we can be certain that this putative *SAL5* promoter is not the promoter of *SAL1*, because the two promoters share little sequence identity, and the TFBS and Y1H analyses demonstrate that the *SAL1* promoter and putative *SAL5* promoter have different binding partners. This putative *SAL5* promoter may be upstream of *SAL5*, or this promoter may be upstream of a different *SAL* gene.

A transcription factor binding sites (TFBS) database search of *Pg*SAL1 and putative *Pg*SAL5 promoters revealed that these genes may be regulated by hormones and environmental cues, some of which have been linked to short day-induced growth cessation and/or bud development in angiosperms (Juntilla and Jensen 1988, Olsen *et al.* 1995a, b, Olsen *et al.* 1997a, Rohde *et al.* 2002, Ruonala *et al.* 2006, Ruttink *et al.* 2007, Kalcits *et al.* 2009, Davies 2010, Tanino *et al.* 2010). DNA-protein interactions identified from yeast-one hybrid experiments reinforce some of the regulatory pathways established by the TFBS search. The established TFBS [ABA INSENSITIVE 4 (ABI4), ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 5 (ATHB5), EARLY METHIONINE BINDING PROTEIN-1 (EmBP1), ARABIDOPSIS THAIANA HOMEOBOX 1 (ATHB1), COMMON PLANT RGULATOR FACTOR 2 (CPRF2), COMMON PLANT RGULATOR FACTOR 3 (CPRF3), RELATED TO ABI1/VP1 (RAV1), TBP-ASSOCIATED FACTOR 1 (TAF1), ARABIDOPSIS THALIANA MYB DOMAIN PROTEIN 15 (ATMYB15)] paired with the interaction of PgASR1-like, PgSOC1-like and PgFLX-like proteins with the *PgSAL1* promoter suggest that *PgSAL1* expression is linked to the ABA, light and low temperatures pathways. The PgSOC1-like and PgFLX-like interactions demonstrate a possible link between *PgSAL1* and the *FT/CO* regulatory pathway. The interaction of PgNBS-LRR/WRKY-like, PgMYB1, PgCPC/ETC-like in addition to the determined TFBS [ABA INSENSITIVE 4 (ABI4), AGAMOUS (AG), AGAMOUS-LIKE 1 (AGL1), AGAMOUS-LIKE 15 (AGL15), AGAMOUS-LIKE 2 (AGL2), ARABINOGALACTAN PROTEIN 1 (AGP1), AINTEGUMENTA (ANT), ARABIDOPSIS RESPONSE REGULATOR 10 (ARR10), ARABIDOPSIS THALIANA MYB DOMAIN PROTEIN 77 (ATMYB77), BASIC HELIX LOOP HELIX 66 (BHLH66), DOF ZINC FINGER PROTEIN 2 (DOF2), DOF ZINC FINGER PROTEIN 3 (DOF3), E2F, GAMYB, GRASSY TILLERS 1 (GT1), KNOTTED-1-LIKE 3

(KNOX3), MYB80, OPAQUE 2 (O2), PYRIMIDINE-BOX BINDING FACTOR (PBF), PROLIFERATING CELL FACTOR 2 (PCF2), RELATED TO ABI1/VP1 (RAV1), RICE TRANSCRIPTION ACTIVATOR -1 (RITA1)] provide evidence that *PgSAL5* may be involved in the developmental regulatory pathway.

Novel regulatory pathways suggested by the promoter analyses has led us to further expand on the possible functions of *PgSAL1* and *PgSAL5* beyond the model of activity-dormancy regulation proposed by Singh *et al.* (2017). The interaction of PgMYB1 with the putative *PgSAL5* promoter indicates a possible link to the regulation of cell development and cell wall biosynthesis during growth cessation. As photoperiods shorten, the tree's active growth cycle terminates, which results in the cessation of growth at meristems (Rohde and Bhalerao 2007), including secondary growth at the vascular cambium (Little and Bonga 1974, Espinosa-Ruiz *et al.* 2004). Potential functions of *PgSAL5* extend to cell fate determination and defensive roles, as determined by promoter analyses. Cell fate determination could be broadly interpreted to mean that *PgSAL5* may have a role in cell development in the terminal bud. *PgSAL5's* role in plant defense may function independently of *PgSAL5's* role in bud formation and/or growth cessation. Evidence of the involvement of *PgSAL5* in the defense pathway appear in both the TFBS search and the Y1H analysis, which leads us to believe this may be a genuine role of this gene. If this finding is valid then this function would be a departure from the roles *PgSAL* genes are traditionally assumed to participate in, based on the angiosperm model. Furthermore, *PgSAL1* appears to be a component of the abiotic response, specifically the response to low temperatures and ABA. Low temperatures are an environmental cue commonly correlated with reduced day length, which can influence bud set (Mølmann *et al.* 2005). The ability of *Arabidopsis thaliana* (L.) Heynh. (*At*) SVP to inhibit transcription of *FT* by complexing with

FLC is greater at lower temperatures (16°C), demonstrating the involvement of *SVP* in the thermosensory pathway (Lee *et al.* 2007, Li *et al.* 2008). Increased ABA content has been linked to bud development and maturation in poplar and birch (Rohde *et al.* 2002, Ruttink *et al.* 2007, Ruonala *et al.* 2006, Maurya and Bhalerao 2017). ABA levels can increase as a result of water stress, and limited access to water is an additional external cue that precedes dormancy and can result in growth cessation (Horvath *et al.* 2003). These findings suggest the influence ABA plays in signalling bud development/maturation, which could be a consequence of the water limiting conditions of the environment.

The results of the promoter analyses and expression profiles suggest that the roles of *PgSAL1* and putative *PgSAL5* are some overlapping functions, but other roles are non-redundant. The potential conserved functions of *PgSAL1* and novel functions of *PgSAL5* are consistent with another gene pair in *Picea abies* L. (*Pa*): *FTL1* and *FTL2*, with the former potentially functioning to regulate timing of the transition to reproductive growth through inhibition (Karlgren *et al.* 2011, Klintenäs *et al.* 2012), and the latter functioning in SD-induced growth cessation (Gyllenstrand *et al.* 2007, Asante *et al*. 2011, Karlgren *et al.* 2011, Klinetäs *et al.* 2012). Differences in homolog functions in *SVP*-like and *DAM*-like genes in raspberry (Mazzitelli *et al.* 2007), peach (Li *et al.* 2009, Yamane *et al.* 2011), kiwifruit (Wu *et al*. 2012), potato (Carmona *et al.* 1998) and trifoliate orange (Li *et al.* 2010) also demonstrate the functional diversification within this gene group.

Based on the findings presented in this thesis, I constructed a figure to demonstrate how *PgSAL1* and *PgSAL5* may function in molecular regulatory pathways (Figure 4.1). Since *PgSAL1* may be linked to the *CO/FT* pathway, I based this model on evidence from *PaFTL2* (Gyllenstrand *et al.* 2007, Karlgren *et al.* 2011) and the current poplar model of photoperiod

growth presented in Singh *et al.* (2017). In the poplar model *FT2* is downregulated in response to SD (Hsu *et al.* 2011), a contrasting effect of SD on *PaFTL2* (Gyllenstrand *et al.* 2007, Karlgren *et al.* 2011), to upregulate *LIKE-APETELA 1* (*LAP1*) (Azeez *et al.* 2014) and eventually lead to continued growth (Randall *et al.* 2015). Similar to the Norway spruce *PaFTL2* (Gyllenstrand *et al.* 2007, Asante *et al*. 2011, Karlgren *et al.* 2011, Klinetäs *et al.* 2012), I predict that there is a white spruce *FTL2* gene which is upregulated under short days. *PgFTL2* would go on to upregulate *PgAP1*-like, similar to the poplar model. SD treatment in white spruce accelerates bud set (El Kayal *et al.* 2011, Hamilton *et al.* 2016), which we predict is positively regulated by *PgSAL1*; therefore, this may indicate an undiscovered *PgAP1*-like gene regulates *PgSAL1* to initiate bud set and/or growth cessation. I hypothesize that *PgAP1*-like would function in a similar manner to *AtAP1* by inhibiting the expression of *AtSOC1* upregulation (Liu *et al.* 2007), and therefore *PgAP1*-like would inhibit *PgSOC1*-like. I believe the inhibitory role of *AtSOC1* on *AtSVP* (Immink *et al.* 2012) is conserved in white spruce, indicating *PgSOC1*-like would function to prevent the transcription of *PgSAL1* and the transition to bud set and/or growth cessation. Low temperatures may influence the *PgSAL1* and rate of bud set and/or growth cessation through interaction with PgASR-like and PgFLX-like. *PgFLX*-like may interact with a yet to be classified *PgFLC*-like gene; however, since less evidence exists in conifers that this pathway is conserved, I have limited the linkage of *PgSAL1* in this pathway. The proposed molecular regulatory pathway for *PgSAL5* is less complex since no connections that we know of have been made between the transcription factors (TFs) identified by yeast one-hybrid (Y1H) and a *SVP/AGL24*-like genes. DNA motifs identified by *in silico* techniques suggest *PgSAL5* may be regulated by photoperiod and low temperatures, although it is unknown if this regulation is directly or indirectly linked to the Y1H TFs. All TFs identified in the Y1H

screen however appear to participate in development, making *PgSAL5* a candidate for a more general role in shoot tip development.

Growth cessation and bud formation are complex traits, similar to the flowering pathway in angiosperms. The complexity of these traits is a result of the function of several TFs and multiple converging pathways being imparted on phase transitions. Consequently, a one-size fits all model of bud formation and growth cessation does not appear to hold true for angiosperm and conifer perennials. This lack of a single model can be due to differences in growth patterns, with angiosperms demonstrating indeterminate growth and conifers exhibiting determinate growth. Bud set induction in angiosperms is regulated by environmental cues such as photoperiod or temperature, while conifer bud initiation is influenced by photoperiod or temperature, and endogenous signals play an increasingly important role as the tree matures (Cooke *et al.* 2012). These endogenous cues are demonstrated by the ability of white spruce to form terminal buds in the absence of shortened photoperiods and low temperatures (El Kayal *et al.* 2011, Hamilton *et al.* 2016). These differences in growth and the varying strength of environmental cues between conifers and angiosperms likely result in differential molecular regulation of growth and development of structures such as buds. The lack of conservation is demonstrated by the fact *PaFTL2* is believed to control bud set and growth cessation (Gyllenstrand *et al.* 2007, Karlgren *et al.* 2011), whereas the poplar *PtFT2* is a positive regulator of growth (Hsu *et al.* 2011).

To investigate the hypothesized roles of *PgSALs*, further additional functional analyses are required. An approach may include creating transgenic white spruce over- and under-expressing *PgSAL* genes to understand their function in the activity-dormancy cycle and/or bud formation. Due to the difficulty associated with silencing of transgenics in white spruce, an alternative would be to perform overexpressing experiments into a species more amenable to this

experimentation, such as poplar. It is possible the experiments performed in poplar may not be a definitive examination of the function of *PgSALs*, as it is possible that the function of a white spruce gene in poplar may not be identical to its role in white spruce. An alternative or accompanying experiment would be the creation of *PgSAL* RNA interference (RNAi) lines to knock down gene expression in white spruce to elucidate the role of these genes in the species of origin. Knock-down experiments are more suited to experiments involved white spruce, since knock-out experiments would be difficult and time consuming to create in a perennial like white spruce. Additional experiments may also focus on confirming the DNA-protein interactions identified in the Chapter 3 Y1H screen. These DNA-protein interactions may be validated through electrophoretic mobility shift assay (EMSA) or tobacco co-infiltrations, the latter being more desirable since this would demonstrate an *in planta* interaction. Once these interactions have been confirmed, targeted deletions can be carried out to determined which nucleotides are essential for promoters to complex with specific TFs. Potential roles of *PgSALs* can be expanded on by further analyses of the upstream regulators using chromatin immunohistochemical precipitation (ChIP). Since ChIP is an *in planta* experiment, this approach would uncover actual physical protein-DNA interactions that occur in white spruce and therefore have greater functional implications. Additional experiemnts should be carried out to confirm the linkage of the SAL promoters to their respective genes. For this linkage to be confirmed future experiments should be designed to cloned the promoter and coding sequence as one fragment.

Future studies should be focused on investigating the roles of all *PgSAL* in processes surrounding dormancy and reproductive bud formation. Terminal bud formation and growth cessation are precursors for the entrance into dormancy (Rohde and Bhalerao 2007). Following dormancy release, preformed needle primordia elongate to push apart the previously formed bud

scales. Accordingly, it seems logical that *PgSAL* genes may also play a role in the entrance, maintenance and/or release of dormancy, and possibly bud burst. Although the research in this thesis focuses on the role of *PgSAL* genes in vegetative bud formation, we must consider that *PgSAL* genes may have a role in the formation of reproductive buds. *PaFTL1* and *PaFTL2* expression suggests these genes have a role in reproductive buds (Karlgren *et al.* 2011), so it seems possible that genes within the vegetative bud regulatory pathways would also play a role in the formation of male and female cone development.

Based on findings from experiments in this thesis, I would recommend that researchers keep an open mind when investigating these gene roles, as it is likely they may have undergone sub- or neo-functionalization, which may give rise to unexpected roles that may not appear cohesive with preconceived angiosperm models. Although angiosperm models are a valuable tool to guide initial hypotheses, they are incapable of providing an accurate explanation of the endogenous function of the corresponding conifer ortholog. A revised version of the molecular regulatory network involved in conifer growth cessation and bud development will inevitably have to be synthesized once this field has a greater understanding and breadth of the key components of this pathway.

Chapter 4 Figures

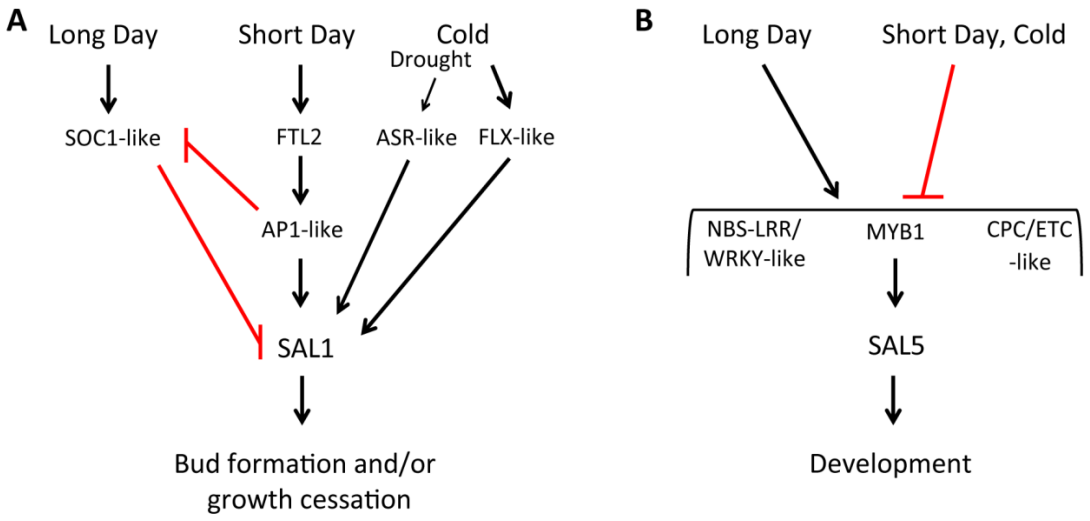


Figure 4.1 Proposed model of white spruce *SHORT VEGETATIVE PHASE/AGAMOUS-LIKE 24-like* (*SAL*) genes in molecular regulatory pathways. (A) There is evidence to support white spruce *SAL1* is directly or indirectly regulated by photoperiod, low temperatures and drought. SOC1-like, ASR-like and FLX-like white spruce proteins have been shown to physically interact with *SAL1*. Based on transcriptional data *SAL1* appears to participate in the regulation of bud formation and possibly growth cessation. (B) White spruce putative *SAL5* is bound by NBS-LRR/WRKY-like, MYB1, and CPC/ETC1-like TFs, all of which are linked to roles in development. We identified photoperiod and low temperature associated DNA motifs in the putative *SAL5* promoter, which may influence the transcription or function of *NBS-LRR/WRKY*-like, *MYB1*, and *CPC/ETC1*-like function and/or regulation in a direct or indirect manner.

References

Abascal F, Zardoya R, Posada D. (2005) ProtTest: Selection of best-fit models of protein evolution. Bioinformatics, 21(9):2104-05.

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T. (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science, 309(5737):1052-6.

Acevedo FG, Gamboa A, Paez-Valencia J, Jimenez-Garia LF, Izaguirre-Sierra M, Alvarez-Buylla ER. (2004) FLOR1, a putative interaction partner of the floral homeotic protein AGAMOUS, is a plant-specific intracellular LRR. Plant Science, 167:225-31. doi:10/1016/j.lantsci.2004.03.009.

Ali Z, Sarwat SS, Karim I, Faridi R, Jaskani MJ, Khan AA. (2016) Functions of plant's bZIP transcription factors. Pakistan Journal of Agricultural Sciences, 53(2):303-14.

Alvarez-Buylla, ER Pelaz, S, Liljergren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martínez-Castilla L, Yanofsky MF. (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proceedings of the Natural Academy of Sciences USA, 97(10):5328-33.

Amasino R. (2010) Seasonal and developmental timing of flowering. The Plant Journal, 61(6):1001-13.

Andrés F, Porri A, Torti S, Mateos J, Romera-Branchat M, García- Martínez JL, Fornara F, Gregis V, Kater MM, Coupland G. (2014) SHORT VEGETATIVE PHASE reduces gibberellin biosynthesis at the *Arabidopsis* shoot apex to regulate the floral transition. Proceeding of the National Academy of Sciences of the USA, 111(26):E2760–E2769. doi: 10.1073/pnas.1409567111.

Armstrong GA, Weisshar B, Hahlbrock K. (1992) Homodimeric and heterodimeric leucine zipper proteins and nuclear factors from parsley recognize diverse promoter elements with ACGT cores. The Plant Cell, 4:525-37.

Arora R, Wisinewski M, Rowland LJ. (1996) Cold acclimation and alterations in dehydrin-like and bark storage proteins in the leaves of sibling deciduous and Evergreen peach. Journal of the American Society of Horticultural Science, 121(5):915-9.

Asante SKA, Yakovlev IA, Fossdal CG, Holefors A, Opseth L, Olsen JE, Junttila O, Johsen Ø. (2011) Gene expression changes during short day induced terminal bud formation in Norway spruce. Plant, Cell & Environment, 34(2):332-46.

Azeez A, Miskolczi P, Tylewicz S, Bhalerao RP. (2014) A tree orthologs of *APETALA1* mediates photoperiodic control of seasonal growth. Current Biology, 24(7):717-24.

Baba K, Karlberg A, Schmidt J, SchraderJ, Hvidsten TR, Bako L, Bhalerao RP. (2011) Activity-dormancy transition in the cambial meristem involves stage-specific modulation of auxin response in hybrid aspen. Proceedings of the National Academy of Sciences USA, 108(8):3418-23.

Bangerth F. (1994) Response of cytokinin concentration in the xylem exudate of bean (*Phaseolus vulgaris* L.) plants to decapitation and auxin treatment, and relationship to apical dominance. Planta, 194(3):439-42.

Baskin JM, Baskin CC. (2004) A classification system for seed dormancy. Seed Science Research, 14(1):1-16.

Bazinet AL, Cummings MP. (2011) Computing the Tree of Life - Leveraging the Power of

Desktop and Service Grids. In Proceedings of the Fifth Workshop on Desktop Grids and Volunteer Computing Systems (PCGrid 2011).
Becker A, Theissen G. (2003) The major clades of MADS-box genes and their role in the developmental and evolution of flowering plants. Molecular Phylogenetics and Evolution, 29(3):464-89.
Bedon F, Bomal C, Caron S, Levasseur C, Boyle B, Mansfield SD, Schmidt A, Gershenzon J, Grima-Pettenati J, Séguin A, MacKay J. (2010) Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses. Journal of Experimental Botany, 61(14):3847-64.
Bielenberg DG, Wang Y, Fan S, Reighard GL, Scorza R, Abbott AG. (2004) A deletion affecting several gene candidates is present in the *Evergrowing* peach mutant. Journal of Heredity, 95(5):436-44.
Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MMS, Keeling CI, Brand D, Vandervalk BP, Kirk H, Pandoh P, Moore RA, Zhao Y, Mungall AJ, Jaquish B, Yanchuk Y, Ritland C, Boyle B, Bousquet J, Ritland K, MacKay J, Bohlmann J, Jones SJM. Assembling the 20Gb white spruce (Picea glauca) genome from whole-genome shotgun sequencing data. Bioinformatics, 29(12):1492-1497.
Blázquez MA, Green R, Nilsson O, Sussman MR, Weigel D. (1998) Gibberellins promote flowering of Arabidopsis by activating the *LEAFY* promoter. The Plant Cell, 10(5):791-800.
Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O. (2006) *CO/FT* regulatory module controls timing of flowering and seasonal growth cessation in trees. Science, 312(5776): 1040-3.
Bomal C, Bedon F, Caron S, Mansfield SD, Levasseur C, Cooke JE, Blais S, Temblay L, Morency MJ, Pavy N, Grima-Pattenati J, Séguin A, MacKay J. (2008) Involvement of *Pinus taeda MYB1* and *MYB8* in phenypropanoid metabolism and secondary cell wall biogenesis: a coaprative *in planta* analysis. Jounral of Experimental Botany, 59(14):3925-39.
Bomal C, Duval I, Giguère I, Fortin E, Caron S, Stewart D, Boyle B, Séguin A, MacKay JJ. (2014) Opposite action of R2R3-MYBs from different subgroups on key genes of the shikimate and monolignol pathways in spruce. Journal of Experimental, Botany 65: 495–508
Borner R, Kapmann G, CHandler J, Gleissner R, Wisman E, Apel K, Melzer S. (2000) A MADS domain gene involved in the transition to flowering in *Arabidopsis*. The Plant Journal, 24:591-9.
Bünning E. (1936) Die endogene tagesperiodik als grundlage der photoperiodischen reaktion. Berichte der Dutschen Botanischen Gesellschaft, 54:590-608.
Bünning E. (1960) Circadian rhythms and the time measurement in photoperiodism. Cold Spring Harbour Symposia on Quantitative Biology, 25:249-56.
Buckley TR, Simon C, Shimodaira H, Chambers GK. (2001) Evaluating hypotheses on the origin and evolution of the New Zealand alpine cicadas (*Maoricicada*) using multiple-comparisons tests of tree topology. Molecular Biology and Evolution, 18(2):223–34.
Campbell M, Segear E, Beers L, Knauber D, Suttle J. (2008) Dormancy in potato tuber meristems: chemically induced cessation in dormancy matches the natural process based on transcript profiles. Functional and Integrative Genomics, 8(4):317-28.

Canadian Wildlife Federation (2017) Spruce trees of Canada. Date accessed: 2017/08-04. http://cwf-fcf.org/en/resources/encyclopedias/flora/spruce-trees-of-canada.html.
Carmona MJ, Ortega N, Garcia-Maroto F. (1998) Isolation and molecular characterization of a new vegetative MADS-box gene from *Solanum tuberosum* L. Planta, 207(2):181-8.
Chen HJ, Bollmark M, Eliasson L. (1996) Evidence that cytokinin controls bud size and branch form in Norway spruce. Physiologia Planatarum, 98:612-8.
Choi K, Kim J, Hwang HJ, Kim S, Park C, Kim SY, Lee I. (2011) The FRIGIDA complex activates transcription of FLC< a strong flowering repressor in *Arabidopsis*, by recruiting chromatin modification factors. Plant Cell, 23:289-303.
Ciolkowski I, Wanke D, Birkenbihl RP, Stomssich IE. (2008) Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. Plant Molecular Biology, 68:81-92.
Clark SE, Williams RW, Meyerowitz EM. (1997) The *Clavata 1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. Cell, 89:575-85.
Cline MG, Dong-Il K. (2002) A preliminary investigation of the role of auxin and cytokinin in sylleptic branching of three hybrid poplar clones exhibiting contrasting degrees of sylleptic branching. Annals of Botany, 90:417-21.
Coenen C, Lomax TL. (1997) Auxin-cytokinin interactions in higher plants: old problems and tools. Trends in Plant Science, 2(9):351-6.
Collier SM, Moffett P. (2009) NB-LRRs work a "bait and switch" on pathogens. Trends in Plant Science, 14:521-9.
Cooke JEK, Eriksson ME, Junttila O. (2012) The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. Plant, Cell & Environment, 35(10):1707-28.
Cseke LJ, Zheng J, Podilla GK. (2003) Characterization of PTM5 in aspen trees: a MADS-box gene expressed during woody/vascular development. Gene, 218, 55-67.
Cseke LJ, Podila GK. (2004) MADS-box genes in dioecious aspen II: A review of MADS-box genes from trees and their potential in forest biotechnology. Physiology and Molecular Biology of Plants, 10(1):7-28.
Cseke LJ, Ravinder N, Pandey AK, Podila GK. (2007) Identification of PTM5 protein interaction partners, a MADS-box gene involved in aspen tree vegetative development. Gene, 391(1-2):209-22.
Darriba D, Taboada GL, Doallo R, Posada D. (2012) jModelTest 2: more models, new heuristics and parallel computing. Nature Methods, 9(8), 772. doi: 10.1038/nmeth.2109.
Davies PJ. (2010) Plant hormones: biosynthesis, signal transduction, action! Rev. 3rd ed. Springer, Dordrecht Heidelberg, London, New York. 1-15pp.
Davies B, Egea-Cortines M, de Andrade Silva E, Saedler H, Sommer H. (1996) Multiple interactions amongst floral homeotic MADS box proteins. European Molecular Biology Organization Journal, 15(16):4330-43.
De Bodt S, Raes J, Van de Peer Y, Theissen G. (2003) And then there were many: MADS goes genomic. Trends in Plant Science, 8(10):475-83.
De Mendiburu F. (2014) agricolae: Statistical Procedures for Agricultural Research. R package version 1.2-1. http://CRAN.R-project.org/package=agricolae.
Ding J, Nilsson O. (2016) Molecular regulation of phenology in trees – because the seasons they are a-changin'. Current Opinion in Plant Biology, 29:73-9.

Donner TJ, Scarpella E. (2013) Transcriptional control of early vein expression of CYCA2;1 and CYCA2;4 in Arabidopsis leaves. Mechanisms of Development, 130:14-24.
Druart N, Johansson A, Baba K, Schrader J, Sjödin A, Bhalerao RR, Resman L, Trygg J, Moritz T, Bhalerao RP. (2007) Environmental and hormonal regulation of the activity-dormancy cycle in the cambial meristem involves stage-specific modulation of transcriptional and metabolic networks. The Plant Journal, 50(4):557-73.
Eagles CF, Wareing PF. (1964). The role of growth substances in the regulation of bud dormancy. Physiologia Planatarum, 17:697-709.
Ehlting J, Mattheus N, Aeschliman DS, Li E, Hamberger B, Cullis IF, Zhuang J, Kaneda M, Mansfield SD, Samuels L, Ritland K, Ellis BE, Bohlmann J, Douglas CJ. (2005) Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. The Plant Journal, 42:618-40.
Egea-Cortines M, Saedler H, Sommer H. (1999) Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. European Molecular Biology Organization Journal, 18(19):5370-9.
El Kayal W, Allen CCG, Ju CJ, Adams E, King-Jones S, Zaharia LI, Abrams SR, Cooke JEK. (2011) Molecular events of apical bud formation in white spruce, *Picea glauca*. Plant, Cell & Environment, 34(3):480-500.
Ellis JG, Lawrence GJ, Dodds PN. (2007) Further analysis of gene-for-gene disease resistance specificity in flac. Molecular Plant Pathology, 8:103-9.
Eriksson ME, Israelsson M, Olsson O, Moritz T. (2000) Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. Nature Biotechnology, 18:784-8.
Eriksson ME, Moritz T. (2002) Daylength and spatial expression of a gibberllin 20-oxidase isolated from hybrid aspen (*Populus tremula x tremuloides Michx*). Planta, 214()6:920-30.
Espinosa-Ruiz A, Saxena S, Schmidt J, Mellerowicz E, Miskolczi M, Bakó L, Bhalerao RP. (2004) Differential stage-specific regulation of cyclin-dependent kinases during cambial dormancy in hybrid aspen. The Plant Journal, 38(4):603-15.
Fan H, Hu Y, Tudor M, Ma H. (1997) Specific interactions between the K domains of AG and AGLs, members of the MADs domain family of DNA binding proteins. The Plant Journal, 12(5):999-1019.
Ferguson BJ, Beveridge CA. (2009) Role for auxin, cytokinin, and stigolactone in regulating shoot branching. Plant Physiology, 149:1929-44.
Fernandez DE, Heck GR, Perry SE, Patterson SE, Bleecker AB, Fang SC. (2000) The embryo MADS domain factor AGL15 act postembryonically: inhibition of perianth senescence and abscission via constitutive expression. The Plant Cell, 12(2):183-97.
Fernandez DE, Want C-T, Zheng Y, Adamczyk BJ, Singhal R, Hall PK, Perry SE. (2014) AGAMOUS-LIKE18, along with SHORT VEGETATIVE PHASE and AGAMOUS-LIKE24, are necessary to block floral gene expression during the vegetative phase. Plant Physiology, 165(4):1591-1603.
Fitch WM. (1971) Toward defining the course of evolution: minimum change for a specific tree topology. Systematic Zoology, 20(4):406-16.
Frankel N, Carrari F, Hasson E, Iusem ND. (2006) Evolutionary history of ASR gene family. Genes, 378:74-83.

Fowler S, Thomashow MF. (2002) Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. The Plant Cell, 14:1675-90.
Fowler SD, Cook D, Thomashow MF. (2005) Low temperature induction of Arabidopsis *CBF1*, *2*, and *3* is gated by the circadian clock. Plant Physiology, 137:961-8.
Fox J, Weisberg S. (2011). An {R} companion to applied regression, second edition. Thousant Oaks CA: Sage. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.
Gamboa A, Paez-Valencia J, Acevedo GF, Vazquez-Moreno L, Alvarez-Buylla RE. Floral transcription factor AGAMOUS interacts *in vitro* with a leucine-rich repeat and an acid phosphatase protein complex.
Galindo-González LM, El Kayal W, Ju CJT, Allen CCG, King-Jones S, Cooke JEK. (2012) Integrated transcriptomic and protemomic profiling of white spruce stems during the transition from active growth to dormancy. Plant, Cell & Environment, 35:682-701.
Galindo-González LM, El Kayal W, Morris JS, Cooke JEK. (2015) Diverse chitinases are invoked during the activity-dormancy transition in spruce. Tree Genetics & Genomes, doi:10.1007/s11295-015-0871-0.
Golan I, Dominguez PG, Konrad Z, Shkolnik-Inbar D, Carrari F, Bar-Zvi D. (2014) Tomato ABSCISIC ACID STRESS RIPENING (ASR) gene family revisited. PLoS One:9(10): e107117. doi:10.1371/journal.pone.010711.
Graeber K, Nakabayashi K, Miatton E, Leubner-Metzger G, Soppe WJ. (2012) Molecular mechanisms of seed dormancy. Plant, Cell & Environment, 35(10):1769-86.
Grandi V, Gregis V, Kater MM. (2012) Uncovering genetic and molecular interactions among floral meristem identity genes in Arabidopsis thaliana. The Plant Journal, 69(5):881-93.
Gregis V, Sessa A, Colombo L, Kater MM. (2006) AGL24, SHORT VEGETATIVE PHASE, and APETALA1 redundantly control *AGAMOUS* during early stages of flower development in Arabidopsis. The Plant Cell, 18(6):1373–82.
Gregis V, Sessa A, Colombo L, Kater MM. (2008) AGAMOUS-LIKE 24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis. The Plant Journal, 56(6), 891-902.
Gregis V, Sessa A, Dorca-Fornell C, Kater MM. (2009) The Arabidopsis floral meristem identity genes *AP1*, *AGL24* and *SVP* directly repress class B and C floral homeotic genes. The Plant Journal, 60(4):626–37.
Gregis V, Andres F, Sessa A, Guerra RF, Simonini S, Mateos JL, Torti S, Zambelli F, Prazzoli GM, Bjerkan KN, Grini PE, Pavesi G, Colombo L, Coupland G, Kater MM (2013) Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in *Arabidopsis*. Genome Biology, 14:R56.
Gocal GF, Pharis RP, Yeung EC, Pearce D. (1991) Changes after decapitation in concentrations of indole-3-acetic acid and abscisic acid in the larger axillary bud of *Phaseolus vulgaris* L. cv Tender Green. Plant Physiology, 95(2):344-50.
Government of Alberta. (2003) Management of white spruce. Last modified: 201706-19. Date accessed: 2017-08-04. http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/apa6261.
Government of Alberta. (2006) Facts on wood. Date accessed: 2017-08-04. https://www.albertacanada.com/files/albertacanada/AIS-BP_WhiteSpruce.pdf.

Government of Alberta. (2013) Facts of Alberta wood species. Date modified: 2013-01-14. Date accessed: 2017-08-04. http://www.albertacanada.com/business/industries/bp-wood-species.aspx.
Government of Alberta. (2016) Trees – White spruce (*Picea glauca*). Date modified: 2016-03-30. Date accessed: 2017-11-23. https://www.agric.gov.ab.ca/app21/forestrypage?cat1=Trees&cat2=White%20Spruce.
Government of Canada. (2017) Species at Risk Public Registry: caribou boreal population. Date modified: 2017/07/28. Date accessed: 2017-08-04. http://www.registrelep-sararegistry.gc.ca/species/speciesDetails_e.cfm?sid=636.
Guindon S, Gascuel O. (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Systematic Biology, 52(5): 696-704.
Gramzow L, Theissen G. (2010) A hitchiker's guide to the MADS world of plants. Genome Biology, 11:214.
Gyllenstrand N, Clapham D, Källman T, Lagercrantz U. (2007) A Norway spruce *FLOWERING LOCUS T* homolog is implicated in control of growth rhythm in conifers. Plant Physiology, 144(1):248-57.
Hajdukiewcz P, Svab Z, Maglia P. (1994) The small, versatile pPZP family of Agrobacterium binary vectors for plant transformation. Plant Molecular Biolog, 25(6):989.
Hamilton JA, El Kayal W, Hart AT, Runcie D, Arango-Velez A, Cooke JEK. (2016) The joint influence of photoperiod and temperature during growth cessation and development of dormancy in white spruce (*Picea glauca*). Tree Physiology 36(11):1432-48.
Hänninen H, Tanino K. (2011) Tree seasonality in a warming climate. Trends in Plant Science, 16(8):412-6.
Hartmann U, Hohmann S, Nettesheim K, Wiseman E, Saedler H, Huijser P. (2000) Molecular cloning of *SVP*: a negative regulator of the floral transition in Arabidopsis. The Plant Journal, 21(4):351–60.
Heide OM. (1974) Growth and dormancy in Norway spruce (*Picea abies*). I. Interaction of photoperiod and temperature. Physiologia Plantarum, 30(1):1–12.
Heide OM, Prestrud AK. (2005) Low temperature but not photoperiod controls growth cessation and dormancy induction and release in apple and pear. Tree Physiology, 25(1):109–14.
Hepworth SR, Valverde F, Ravenscroft D, Mouradov A, Coupland G. (2002) Antagonistic regulation of flowering-time gene SOC1 by CONSTANS and FLC via separate promoter motifs. EMBO Journal, 21:4327-37.
Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theissen. (2002) Two ancient classes of MIKC-type MADS-box genes are present in the Moss *Physcomitrella patens*. Molecular Biology and Evolution, 19(6):801-14.
Holefors A, Opseth L, Ree Rosnes AK, Ripel L, Snipen L, Fossdal CG, Olsen JE. (2009) Identification of *PaCOL1* and *PaCOL2*, two *CONSTANS-like* genes showing decreased transcript levels preceding short day induced growth cessation in Norway spruce.
Honma T, Goto K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature, 409(6819):525-9.
Horvath DP, Anderson JV, Chao WS, Foley ME. (2003) Knowing when to grow: signals regulating bud dormancy. TRENDS in Plant Science, 8(11):534-40.
Horvath D. (2009) Common mechanisms regulate flowering and dormancy. Plant Science, 177:523-31.

Horvath DP, Sung S, Kim D, Chao W, Anderson J. (2010) Characterization, expression and function of *DORMANCY ASSOCIATED MADS-BOX* genes from leafy spurge. Plant Molecular Biology, 73(1-2):169-79.

Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. (2003) From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. Canadian Journal of Botany, 81(22):1247-66.

Hsu CY, Adams JP, Kim H, No K, Ma C, Strauss SH, Dmevich J, Vandervelde L, Ellis JD, Rice BM, Wickett N, Gunter LE, Tuskan GA, Brunner AM, Page GP, Barakat A, Carlson JE, DePamphilis CW, Luthe DS, Yuceer C. (2011) *FLOWER LOCUS T* duplication coordinates reproductive and vegetative growth in perennial poplar. Proceedings of the National Academy of Sciences of the USA, 108(26):10756-61.

Hu YX, Wang YX, Liu XF, Li JY. (2004) Arabidopsis RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. Cell Research, 14:8–15.

Immink RGH, Posé D, Ferrario S, Ott F, Kaufmann K, Valentim FL, de Folter S, van der Wal F, van Dijk ADJ, Schmid M, Angenent GC. (2012) Characterization of *SOC1*'s central role in flowering by the identification of its upstream and downstream regulators. Plant Physiology, 160(1):433-49.

Ingvarsson PK, García MV, Hall D, Luquez V, Jansson S. (2006) Clinal variation in *phyB2*, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). Genetics, 172(3):1845-53.

Irish VF. (2010) The flowering of Arabidopsis flower development. The Plant Journal, 61(6):1014-28.

Jang S, Torti S, Coupland G. Genetic and spatial interactions between FT, TSF and SVP during the early stages of floral induction in Arabidopsis. (2009) The Plant Journal, 60:614-25.

Jian LC, Li PH, Sun LH, Chen THH. (1997) Alterations in ultrastructure and subcellular localization of Ca2+ in poplar apical bud cells during the induction of dormancy. Journal of Experimental Botany, 48(311):1195-1207.

Jiménez S, Lawton-Rauh AL, Reighard GL, Abbott AG, Bielenberg DC. (2009) Phylogenetic analysis and molecular evolution of the dormancy associated MADS-box genes from peach. BMC Plant Biology, 9(81). doi:10.1186/1471-2229-9-81.

Jonkers H. (1979) Bud dormancy of apple and pear in relation to the temperature during the growth period. Scientia Horticulturae, 10(2):149–154.

Junttila O. (1980) Effect of photoperiod and temperature on apical growth cessation in two ecotypes of *Salix* and *Betula*. Physiologia Plantarum, 48(3):347–52.

Junttila O, Jensen E. (1988) Gibberellins and photoperiodic control of shoot elongation in *Salix*. Physiologia Plantarum, 74(2):371-76.

Junttila O, Nilsen J, Igeland B. (2003) Effect of temperature on the induction of bud dormancy in ecotypes of *Betula pubescens* and *Betula pendula*. Scandinavian Journal of Forest Research, 18(3):208–217.

Kalcsits LA, Silim S, Tanino K. (2009) Warm temperature accelerates short photoperiod-induced growth cessation and dormancy induction in hybrid poplar (*Populus* x spp.). Trees, 23(5):971-9.

Karlberg A, Bako L, Bhalerao RP. (2011) Short day-mediated cessation of growth requires the downregulation of *AINTEGUMENTALIKE1* transcription factor in hybrid aspen. PLoS Genetics, 7(11): e1002361.

Karlgren A, Gyllenstrand N, Källman T, Sundström JF, Moore D, Lascoux M, Lagercrantz U. (2011) Evolution of the PEBP gene family in plants: functional diversification in seed plant evolution. Plant Physiology 156(4): 1967-1977.

Karlgren A, Gyllenstrand N, Clapham D, Lagercrantz U. (2013) *FLOWERING LOCUS T/TERMINAL FLOWER1*-like genes affect growth rhythm and bud set in Norway spruce. Plant Physiology, 162(2):792-803.

Katoh K, Standley DM. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution, 30(4):772-80.

Kaufmann K, Melzer R, Theissen G. (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene, 347(2):183-98.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, *28*(12), 1647-9.

Kirik V, Simon M, Huelskamp M, Schiefelbein J. (2003) The *ENHANCER OF TRY AND CPC1* gene acts redundantly with *TRIPTYCHON* and *CAPRICE* in trichome and root hair cell patterning in *Arabidopsis*. Developmental Biology, 268:506-13.

Klintenäs M, Pin PA, Benlloch R, Ingvarsson PK, Nilsson O (2012). Analysis of conifer *FLOWERING LOCUS T/TERMINAL FLOWER1*-like genes provides evidence for dramatic biochemical evolution in the angiosperm *FT* lineage. New Phytologist, 196(4): 1260–73.

Koomneef M, Hanhart CJ, van der Veen JH. (1991) A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. Molecular Genetics and Genomics, 229(1):57-66.

Kozlowski TT, Pallardy SG. (1997) Physiology of Woody Plants, 2nd edition, Academic Press, New York, USA.

Langercrantz U. (2009) At the end of the day: a common molecular mechanism for photoperiod responses in plants? Journal of Experimental Botany, 60:2501-15.

Lee H, Suh SS, Park E, Cho E, Ahn JH, Kim SG, Lee JS, Kwon YM, Lee I. (2000) The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. Genes and Development, 14(18):2366-67. doi:10.1101/gad.813600.

Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH. (2007) Role of *SVP* in the control of flowering time by ambient temperature in *Arabidopsis*. Genes and Development, 21(4):397-402.

Lee J, Oh M, Park H, Lee I. (2008) SOC1 translocated to the nucleus by interaction with AGL24 directly regulates *LEAFY*. The Plant Journal 55(5), 832–843.

Lee JH, Ryu H, Chung KS, Posé D, Kim S, Schmid M, Ahn JH. (2013) Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. Science, 342:10.1126/science.1241097.

Lee JH, Chung KS, Kim S, Ahn JH. (2014) Post-translational regulation of SHORT VEGETATIVE PHASE as a major mechanism for thermoregulation of flowering. Plant Signaling & Behaviour, e28193.

Leinonen I. (1996) Dependence of dormancy release on temperature in different origins of *Pinus sylvestris* and *Betula pendula* seedlings. Scandinavian Journal of Forest Research, 11(1-4):122–8.
Lehti-Shiu MD, Adamczyk BJ, Fernandez DE. (2005) Expression of MADS-box genes during embryonic phase in Arabidopsis. Plant Molecular Biology, 58(1):89-107.
Levy A, Guenoune-Gelbart D, Epel BL. (2007) b-1,3-glucanases: plasmodesmal gate keepers for intercellular communication. Plant Signaling & Behavior, 2(5):404-7.
Li D, Liu C, Shen L, Wu Y, Chen H, Robertson M, Helliwell CA, Ito T, Meyerowitz E, Yu H. (2008) A repressor complex governs the integration of flowering signals in *Arabidopsis*. Developmental Cell, 15(1):110-20.
Li Z, Reighard GL, Abbott AG, Bielenberg DG. (2009) Dormancy-associated MADS genes from the EVG locus of peach [*Prunus persica* (L.) Batsch] have distinct seasonal and photoperiodic expression patterns. Journal of Experimental Botany, 60(12):3521-30.
Li ZM, Zhang JZ, Mei L, Deng XX, Hu CG, Yao JL. (2010) *PtSVP*, a *SVP* homolog from trifoliate orange (*Poncirus trifoliata* L. Raf.) shows seasonal periodicity of meristem determination and affects flower development in transgenic *Arabidopsis* and tobacco plants. Plant Molecular Biology, 74(1-2):129-42.
Lifschitz E, Eviatar T, Rozman A, Shalit A, Goldshmidt A, Amsellem Z, Alvarez JP, Eshed Y. (2006) The tomato *FT* ortholog triggers systemic signals that regulate growth and lowering and substitute for diverse environmental stimuli. Proceedings of the National Academy of Sciences of the USA, 103(16): 6398–6403.
Liu C, Zhou J, Brach-Drori K, Yalovsky S, Ito T, Yu H. (2007) Specification of Arabidopsis floral meristem identity by repression of flowering time genes. Development, 134(10):1901-10.
Liu C, Xi W, Shen L, Tan C, Yu H. (2009) Regulation of floral patterning by flowering time genes. Developmental Cell, 16(5):711–22.
Liu G, Li W, Zheng P, Xu T, Chen L, Liu D, Hussain S, Teng Y. (2012) Transcriptomic analysis of 'Suli' pear (*Pyrus pyrifolia* white pear group) buds during the dormancy by RNA-Seq. BMC Genomics, 13:700.
Little CHA, Bonga JM. (1974) Rest in the cambium of *Abies balsamea*. Canadian Journal of Botany, 52(7):1723-30.
Loots GG, Ovcharenko I, Pachter L, Dubchak I, Tubin EM. (2002) rVista for comparative sequence-based discovery of functional transcription factor binding sites. Genome Research, 12:832-39.
Loots GG, Ovcharenko I. (2004) Nucleic Acid Research, 32:W217-W221. DOI: 10/1093/nar/gkh383.
Luquez V, Hall D, Albrectsen BR, Karlsson J, Ingvarsson P, Jansson S. (2008) Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. Tree Genetics & Genomes, 4(2):279-92.
Masiero S, Li M, Will I, Hartmann U, Saedler H, Huijser P, Schwarz-Sommer Z, Sommer H. (2004) *INCOMPOSITA*: a MADS-box gene controlling prophyll development and floral meristem identity in *Antirrhinum*. Development, 131(23):5981–90.
Maurya JP, Bhalerao RP. (2017) Photoperiod- and temperature-mediated control of growth cessation and dormancy in trees: a molecular perspective. Annals of Botany, 120(3):351-60.

Mazzitelli L, Hancock RD, Haupt S, Walker PG, Pont SD, McNicol J, Cardle L, Morris J, Viola R, Brennan R, Hedley PE, Taylor MA. (2007) Co-ordinated gene expression during phases of dormancy release in raspberry (*Rubus idaeus* L.) buds. Journal of Experimental Botany, 58(5):1035–45.

Michaels SD, Amasino RM. (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell, 11:949-56.

Michaels, SD, Ditta G, Gustafson-Brown C, Pelaz S, Yanofsky M, Amasino RM. (2003) *AGL24* acts as a promoter of flowering in *Arabidopsis* and is positively regulation by vernalization. The Plant Journal, 33(5):867-73.

Michaels SD, Himelblau E, Kim SY, Schomburg FM, Amasino RM. (2005) Integration of flowering signals in winter-annual Arabidopsis. Plant Physiology, 137, 149-56.

Mimida N, Saito T, Moriguchi T, Suzuki A, Komori S, Wada M. (2015) Expression of *DORMANCY-ASSOCIATED MADS-BOX* (*DAM*)-like genes in apple. Biologia Plantarum, 59(2):237–244.

Mohamed R, Wang C-T, Ma C, Shevchenko O, Dye SJ, Puzey JR, Etherington E, Sheng X, Meilan R, Strauss SH, Brunner AM. (2010) Populus CEN/TFL1 regulates first onset of flowering, axillary meristem identity and dormancy release in *Populus*. The Plant Journal, 62(4):674-88.

Mølmann JA, Berhanu AT, Stormo SK, Ernstsen A, Junttila O, Olsen JE. (2003) Metabolism of gibberellin A_{19} is under photoperiodic control in *Populus*, *Salix* and *Betula*, but not in daylength-insensitive *Populus* overexpressing phytochrome A. Physiologia Plantarum, 119(2):278-86.

Mølmann JA, Asante DKA, Jensen JB, Krane MN, Ernstsen A, Junttila O, Olsen JE. (2005) Low night temperature and inhibition of gibberellin biosynthesis override phytochrome action and induce bud set and cold acclimation but not dormancy in *PHYA* overexpressors and wild-type of hybrid aspen. Plant, Cell & Environment, 28(12): 1579–88.

Moon J, Suh SS, Lee H, Choi KR, Hong CB, Paek NC, Kim SG, Lee I. (2003) The *SOC1* MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. The Plant Journal, 35(5):613-23.

Mouradov A, Glassick TV, Hamdorf BA, Murphy LC, Marla SS, Yang Y, Teasdale RD. (1998) Family of MADS-box gene expressed early in male and female reproductive structures of Monterey pine. Plany Physiology, 117:55-61.

Nakamura Y, Andrés F, Kanehara K, Liu Y-C, Dörmann P, Coupland George. (2013) *Arabidopsis* florgen *FT* binds to diurnally oscillating phospholipids that accelerate flowering. Nature Communications, 5:3553.

Natural Resources Canada, Canadian Forest Services. (2014a) The State of Canada's Forests Annual Report 2014. Ottawa, ON, 6-8 pp. http://cfs.nrcan.gc.ca/pubwarehouse/pdfs/35713.pdf.

Natural Resources Canada, Canadian Forest Services. (2014b) Softwood Lumber - Background. http://www.nrcan.gc.ca/19603. Date modified: 2014/04/26. Date accessed: 2017/08/04.

Natural Resources Canada, Canadian Forest Services. (2015a) The State of Canada's Forests 2015 - 25th anniversary edition. Ottawa, ON. http://cfs.nrcan.gc.ca/pubwarehouse/pdfs/36553.pdf.

Natural Resources Canada, Canadian Forest Services (2015b) White spruce. Date modified: 2015-08-04. Date accessed: 2017-08-04. https://tidcf.nrcan.gc.ca/en/trees/factsheet/38.

Natural Resources Canada, Canadian Forest Services. (2017) Softwood Lumber Fact Sheet. Date modified: 2017-04-25. Date accessed: 2017-08-04. http://www.nrcan.gc.ca/19605.
Navarro C, Abelenda JA, Cruz-Oró E, Cuéllar CA, Tamaki S, Silva J, Shimamoto K, Prat S. (2011) Control of flowering and storage organ formation in potato by *FLOWERING LOCUS T*. Nature, 478:119–122.
Nienstaedt H, Zasada JC. (1990) *Picea glauca* (Moench) Voss. 1 Conifers. USDA Forest Service Agriculture Handbook. 654, USDA Forest Service, Washington, DC. 204-26 pp in R. M. Burns and B. H. Honkala, eds. Silvics of North America.
Nitsch JP. (1957) Photoperiodism in woody plants. Proceedings of the American Society of Horticultural Sciences, 70: 526-44.
Norman C, Runswick M, Pollock R, Treisman R. (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-Fos serum response element. Cell, 55(6):989-1003
Okuba H. (2000) Growth cycle and dormancy in plants. In dormancy in plants – from whole plant behavior to cellular control. (Viemont JD, Crabbe J, eds), pp. 1 – 22, CABI.
Olsen JE, Jensen E, Junttila O, Moritz T. (1995a) Photoperiodic control of endogenous gibberellins in seedlings of *Salix pentandra*. Physiologia Plantarum, 93, 639–44.
Olsen JE, Junttila O, Moritz T. (1995b) A localised decrease of GA1, in shoot tips of *Salix pentandra* seedlings precedes cessation of shoot elongation under short photoperiod. Physiologia Plantarum, 93(4):627–32.
Olsen JE, Junttila O, Moritz T. (1997a) Long-day induced bud break in *Salix pentandra* is associated with transiently elevated levels of GA1, and gradual increase in Indole-3-acetic acid. Plant and Cell Physiology, 38(5):536–40.
Olsen JE, Junttila O, Nilsen J, Eriksson M, Martinussen I, Olsson O, Sandberg G, Moritz T. (1997b) Ectopic expression of oat phytochrome A in hybrid aspen changes critical daylength for growth and prevents cold acclimation. The Plant Journal, 12(6):1339-50.
Olsen JE. (2010) Light and temperature sensing and signaling in induction of bud dormancy in woody plants. Plant Molecular Biology, 73:37-47.
Opseth L, Holefors A, Ree Rosnes A, Lee Y, Olsen JE. (2016) *FTL2* expression preceding bud set corresponds with timing of bud set in Norway spruce under different light quality treatments. Environmental and Experimental Botany, 121:121-31.
Pařenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L. (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. The Plant Cell, 15(7):1538-51.
Passmore S, Maine GT, Elble R, Christ C, Tye BK. (1988) *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MAT cells. Journal of Molecular Biology, 204(3):593–606
Patzlaff A, Newman LJ, Ross, CD, Whetten RW, Smith C, McInnis S, Bevan MW, Sederoff RR, Campbell MM. (2003) Characterisation of *Pt*MYB1, an R2R3-MYB from pine xylem. Plant Molecular Biology 53:597-608.
Pavy N, Boyle B, Nelson C, Paule C, Giguere I, Caron S, Parsons LS, Dallaire N, Bedon F, Bérubé H, Cooke J, Mackay J. (2008) Identification of conserved core xylem sets: conifer cDNA microarray development, transcript profiling and computational analyses. New Phytologist, 180(4):766-86.
Pavy N, Lamothe M, Pelgas B, Gagnon F, Birol I, Bohlmann J, Mackay J, Isabel N, Bousquet J

(2017) A high-resolution reference genetic map positioning 8.8 K genes for the conifer white spruce: structural genomics implications and correspondence with physical distance. The Plant Journal, 90:189-203.
Petterle A, Resman L, Bhalerao RP. (2011) ABA and chromatin remodeling regulate the activity-dormancy cycle in hybrid aspen. Doctoral Thesis, No. 2011:17, Swedish University of Agricultural Sciences, Umeå.
Phukan UJ, Jeena GS, Shukla R. (2016) WRKY transcription factors: molecular regulation and stress responses in plants. Frontiers in Plant Science, 7:760, doi: 10.3389/fpls.2016.00760.
Pin PA, Nilsson O. (2012) The multifaceted roles of *FLOWERING LOCUS T* in plant development. Plant, Cell & Environment, 35:1742-55.
R Core Team (2017). R: A language and environment for statistical computing R. Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
Qiu Y, Li L, Wang B, Xue J, Hendry TA, Li R, Brown JW, Chen Z. (2010) Angiosperm phylogeny inferred from sequences of four mitochondrial genes. Journal of Systematics and Evolution, 48(6):391-425.
Ramos A, Perez-Solis E, Ibanez C, Casado R, Collada C, Gomez L, Aragoncillo C, Allona I. (2005) Winter disruption of the circadian clock in chestnut. Proceedings of the National Academy of Science USA, 102(19):7037-42.
Randall RS, Miyashima S, Blomster T, Zhang J, Elo A, Karlberg A, Immanen J, Nieminen K, Lee JY, Kakimoto T, Blajecka K, Melnyk CW, Alcasabas A, Forzani C, Matsumoto-Kitano M, Mähönen AP, Bhalerao R, Dewitte W, Helariuta Y, Murray JAH. (2015) *AINTEGUMENTA* and the D-type cyclin *CYCD3;1* regulate root secondary growth and respond to cytokinins.
Riechmann JL, Krizek BA, Meyerowitz EM. (1996) Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. Proceedings of the National Academy of Sciences of the USA, 93(10):4793-8.
Rigault P, Boyle, B, Lepage P, Cooke JEK, Bousquet J, MacKay JJ. (2011) A white spruce gene catalogue resource for conifer genome analyses. Plant Physiology, 157(1):14-28.
Rinne PLH, van der Schoot C. (1998) Symplasmic fields in the tunica of the shoot apical meristem coordinate morphogenetic events. Development, 125(8):1477-85.
Rinne PLH, van den Boogaard R, Mensink MGJ, Kopperud C, Kormelink R, Goldbach R, van der Schoot C. (2005) Tobacco plants respond to the constitutive expression of the tospovirus movement protein NS(M) with a heat- reversible sealing of plasmodesmata that impairs development. The Plant Journal, 43(5):688-707.
Rinne PLH, Welling A, Vahala J, Ripel L, Ruonala R, Kangasjärvi J, van der Schoot C. (2011) Chilling of dormant buds hyperinduces FLOWERING LOCUS T and receruit GA-inducible 1,3-beta-glucanases to reopen signal conduits and release dormancy in *Populus*. The Plant Cell, 23(1):130-146.
Rinerson CI, Rabara RC, Tripathi P, Shen QJ, Rushton PJ. (2015) The evolution of WRKY transcription factors. BMC Plant Biology, 15(66). doi: 10.1186/s12870-015-0456-y
Rohde A, Howe GT, Olsen JE, Moritz T, Van Montagu, Junttila O, Boerjan W. (2000) Molecular aspects of bud dormancy in trees. Molecular Biology of Woody Plants, 1:89-134.

Rohde A, Prinsen E, De Rycke R, Engler G, Van Montagu M, Boerjan W. (2002) PtABI3 impinges on the growth and differentiation of embryonic leaves during bud set in poplar. The Plant Cell, 14(8):1885-1901.
Rohde A, Bhalerao RP. (2007) Plant dormancy in the perennial context. Trends in Plant Science, 12(5):217-23.
Rohde A, Bastien C, Boerjan W. (2011) Temperature signals contribute to the timing of photoperiodic growth cessation and bud set in poplar. Tree Physiology, 31: 472–82.
Rodriguez J, Sherman WB, Scorza R, Wisniewski M, Okie WR. (1994) Evergreen peach, its inheritance and dormant behavior. Journal of the American Society for Horticultural Science, 119(4):789–792.
Rogers JS, Swofford DL. (1998) A fast method for approximating maximum likelihoods of phylogenetic trees from nucleotide sequences. Systematic Biology, 47(1):77-89.
Ruonala R, Rinne PLH, Baghour M, Moritz T, Tuominen H, Kangasjärvi J. (2006) Transitions in the functioning of the shoot apical meristem in birch (*Betula pendula*) involve ethylene. The Plant Journal, 46(4):628-40.
Rushton PJ, Somssich IE, Ringler P, Shen QJ. (2010) WRKY transcription factors. Cell Press, 15(5):247-58. doi: 10.1016/j.tplants.2010.02.006.
Lenth RV. (2016) Least-squares means: The R Package lsmeans. Journal of Statistical Software, 69(1), 1-33. http://CRAN.R-project.org/package=lsmeans.
Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W, Rohde A. (2007) A molecular timetable for apical bud formation and dormancy induction in poplar. The Plant Cell, 19(8):2370-90.
Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G. (2000) Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. Science, 288:1613-16.
Sarvas R. (1974) Investigations on the annual cycle of development of forest trees. II. Autumn dormancy and winter dormancy. Communicationes Instituti Forestalis Fenniae, 84:1–101.
Sasaki R, Yamane H, Ooka T, Jotatsu H, Kitamura Y, Akagi T, Tao R. (2011) Functional and expressional analyses of *PmDAM* genes associated with endodormancy in Japanese apricot (*Prunus mume*). Plant Physiology, 157(1):485-97.
Schneider H, Schuettpelz E, Pryer KM, Cranfill R, Magallón S, Lupia R, (2004) Ferns diversified in the shadow of angiosperms. Nature, 428:553-7.
Schwarz-Sommer Z, Hue I, Huijser P, Flor PJ, Hansen R, Tetens F, Lönnig WE, Saedler H, Sommer H. (1992) Characterization of the *Antirrhinum* floral homeotic MADS-box gene *DEFICIENS*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. EMBO Journal, 11(1):251-63.
Shimodaira H, Hasegawa M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution, 16(8):1114-6.
Searle I, He Y, Turck F, Vincent C, Fornara F, Krober S, Amasino RA, Coupland G. (2006) The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. Genes & Development, 20:898–912.
Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES. (1999) The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. Plant Cell, 11(3):445-58.

Shen G, Pang Y, Wu W, Deng Z, Liu X, Lin J, Zhao L, Sun X, Tang K. (2005) Molecular cloning, characterization and expression of a novel *Asr* gene from *Ginko biloba*. Plant Physiology and Biochemistry, 43(9):836-43.

Singh RK, Svystun T, AlDahmash B, Jönsson AM, Bhalerao RP. (2017) Photoperiod- and temperature-mediated control of phenology in trees - a molecular perspective. New Phytologist, 213:511-24.

Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refuli-Rodriquez NF, Walker JB, Moore MJ, Carlsward BS, Bell CD, Latvis M, Crawley S, Black C, Diouf D, Xi Z, Rushworth CA, Gitzendanner MA, Sytsma KJ, Qiu Y, Hilu KW, Davis CC, Sanderson MJ, Beaman RS, Olmstead RG, Judd WS, Donoghue MJ, Soltis PS. (2011) Angiosperm phylogeny: 17 genes, 640 taxa. American Journal of Botany, 98(4):704-30.

Sommer H, Beltrán JP, Huijser P, Page H, Lönnig WE, Saedler H, Schwarz-Sommer Z. (1990) DEFICIENS, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. EMBO Journal, 9(3):605-13.

Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G. (2001) *CONSTANS* mediates between the circadian clock and the control of flowering in Arabidopsis. Nature, 410(6832):1116-20.

Sukumaran J, Holder MT. (2010) DendroPy: A Python library for phylogenetic computing. Bioinformatics 26:1569-71.

Sundell D, Mannaperuma C, Netotea S, Delhomm N, Lin Y, Sjödin A, Van de Peer Y, Jansson S, Hvidsten TR, Street N. (2015) The plant genome integrative explorer resource: PlantGenIE.org. New Phytologist, 208:1149-56. sdoi:10.1111/nph.13557.

Sundström J, Carlsbecker A, Svensson ME, Svenson M, Johanson U, Theißen G, Engström. (1999) MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. Developmental Genetics, 25:253-66.

Swofford DL. (2002) PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Taiz L, Zeiger E. (2010) Plant Physiology 5th edition. Sinaur Associated, Inc. Sunderland, MA, USA. Pp 732-4.

Tao Z, Shen L, Liu C, Yan Y, Yu H. (2012) Genome-wide identification of SOC1 and SVP targets during the floral transition in Arabidopsis. The Plant Journal, 70(4):549-61.

Tanino KK, Kalcsits L, Silim S, Kendall E, Gray GG. (2010) Temperature-driven plasticity in growth cessation and dormancy development in deciduous woody plants: a working hypothesis suggesting how molecular and cellular function is affected by temperature during dormancy induction. Plant Molecular Biology, 73(1-2):49-65.

Theissen G, Kim JT, Saedler H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. Journal of Molecular Evolution, 43(5):484-516.

Tominaga-Wada R, Nukumizu Y, Wada T. (2013) Flowering is delayed by mutations in homologous genes *CARPICE* and *TRYPTICHON* in early flowering *Arabidopsis cpl3* mutant. Journal of Plant Physiology, 170:1466-8.

Torii KU, Mitsukawa N, oosumi T, Mitsuura Y, Yokoyama R, Whitier R, Komeda Y. (1996) The *arabidopsis ERECTA* gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. Plant cell, 8:735-46.

Trevaskis B, Tadege M, Hemming MN, Peacock WJ, Dennis ES, Sheldon C. (2007) *Short Vegetative Phase*-like MADS-box genes inhibit floral meristem identity in barley. Plant Physiology, 143(1):225-35.

Tylewicz S, Petterle A, Englund M, Resman L, Bhalerao RP. (2015) Molecular analysis of growth cessation in hybrid aspen. Ph.D. thesis, Swedish University of Agricultural Sciences Umeå.

Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. (2004) Photoreceptor regulation of COSTANS protein in photoperiod flowering. Science, 303(5660):1003-6.

Voogd C, Wang T, Varkonyi-Gasic E. (2015) Functional and expression analyses of kiwifruit *SOC1*-like genes suggest that they may not have a role in the transition to owering but may affect the duration of dormancy. Journal of Experimental Botany, 66(15):4699-710.

Wada T, Tachibana T, Shimura Y, Okada K. (1997) Epidermal cell differentiation in *Arabidopsis* determined by a Myb homolog, CPC. Science, 277:1113-6.

Wareing P. (1970) Cold resistance and injury in woody plants. Science, 169:1269-78.

Warren RL, Keeling CI, Yuen MMS, Raymond A, Taylor GA, Vandervalk BP, Mohamadi H, Paulino D, Chiu R, Jackman SD, Robertson G, Yang C, Boyle B, Hoffmann M, Weigel D, Nelson DR, Ritland C, Isabel N, Jaquish B, Yanchuk A, Bousquet J, Jones SJM, MacKay J, Birol, Bohlmann J. (2015) Improved white spruce (Picea glauca) genome assembliesand annotation of large gene families of conifer terpenoidand phenolic defense metabolism. The Plant Journal, 83:189–212

Washio K. (2003) Functional dissections between GAMYB and Dof transcription factors suggest a role for protein-protein associations in the gibberellin-mediated expression of the *RAmy1A* gene in the rice aleurone. Plant Physiology, 133:850-63.

Weiser CJ. (1970) Cold Resistance and Injury in Woody Plants. Science, 169(3959):1270-8.

Welling A, Palva ET. (2006) Molecular control of cold acclimation in trees. Physiologia Planatarum, 127:167-81.

Wickett NJ, Mirarab S, Nguyen, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, Ruhfei BR, Wafula E, Der JP, Graham SW, Mathews S, Melkonian M, Soltis DE, Soltis PS, Miles NW, Rothfels CJ, Pkorny L, Shaw AJ, DeGironimo L, Stevenson DW, Surek B, Villarreal JC, Roure B, Philippe H, dePamphilis CW, Chen T, Deyholos MK, Caucom RS, Kutchan TM, Augustin MM, Wang J, Zhang Y, Tian Z, Yan Z, Wu X, Sun X, Wong GK-S, Leebens-Mack J. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. Proceedings of the National Academy of Sciences of the United States of America, 111(45):R4859-68.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. (2007) An "Electronic Fluorescence Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS One, 2(8): e719.

Woo HR, Kim JH, Kim J, Kim J, Lee U, Song I, Kim J, Lee H, Nam HG, Lim PO. (2010) The RAV1 transcription factor positively regulates leaf senescence in *Arabidopsis*. Journal of Experimental Botany, 61(14):3947-57. doi:10.1093/jxb/erq206.

Woodger FJ, Millar A, Murray F, Jacobsen JV, Gubler F. (2003) The role of GAMYB transcription factors in GA-regulated gene expression. Journal of Plant Growth Regulation, 22:176-84.

Worral J, Mergen F. (1967) Environmental and genetic control of dormancy in *Picea abies*. Physiologia Planatarum, 20(3):733-45. doi:10.1111/j.1399-3054.1967.tb07217.x.

Wu RM, Walton EF, Richardson AC, Wood M, Hellens RP, Varkonyi-Gasic E. (2012) Conservation and divergence of four kiwifruit *SVP*-like MADS-box genes suggest distinct roles in kiwifruit bud dormancy and flowering. Journal of Experimental Botany, 63(2):797-807.

Xu X, van Lammeren AAM, Vermeer E, Vreugdenhil D. (1998) The role of gibberellin, abscisic acid, and sucrose in the regulation of potato tuber formation in vitro. Plant Physiology, 117:575-84.

Yamane H, Tomomi O, Jotatsu H, Sasaki R, Tao R. (2011) Expression analysis of *PpDAM5* and *PpDAM6* during flower bud development in peach (*Prunus persica*). Scientia Horticulturae, 129(4):844-8.

Yang Y, Chi Y, Wang Z, Zhou Y, Fan B, Chen Z. (2016) Functional analysis of structurally related soybean GmWRKY58 and GmWRKY76 in plant growth and development. Journal of Experimental Botany, 67(15):4227-42.

Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM. (1990) The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. Nature, 346:35-9.

Yanovsky MJ, Kay SA. (2002) Molecular basis of seasonal time measurement in *Arabidopsis*. Nature, 419:308-12

Yoo SK, Chung KS, Kim J, Lee JH, Hong SM, Yoo SJ, Yoo SY, Lee JS, Ahn JH. (2005) *CONSTANS* activates *SUPPRESSOR OF OVER-EXPRESSION OF CONSTANS 1* through *FLOWERING LOCUS T* to promote flowering in Arabidopsis. Plant Physiology, 139(2):770–8.

Yu H, Xu Y, Tan EL, Kumar PP. (2002) AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals. Proceedings of the National Academy of Sciences of the USA, 99(25):16336-41.

Yu H, Ito T, Wellmer F, Meyerowitz EM. (2004) Repression of *AGAMOUS-LIKE 24* is a crucial step in promoting flower development. Nature Genetics, 36(2):157-61.

Zwicki D J. (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

5.0 Appendices

Appendix 1. *Picea glauca SHORT VEGETATIVE PHASE/AGAMOUS-LIKE 24-like* (*SAL*) promoter sequences generated from Genome Walker™ cloning, and extended using gDNA sequences from the Norway spruce genome. Cloned portion of the coding sequence is highlighted in blue.

>*Pg*SAL1 promoter
ACTAATAAGGGTTGGGACTATAGAAATAATAATATTAATGTAATAATTATTTTATGC
ATCTAAAAATTTGATTTCATTAATTAAAAACTATAAACCACAATTGGACAAAATTCA
AAATTATCTCATCTCAATTCAATTCTAGTGATTATTGCTAGGAAAACTCCCAATCTTA
GTTATGACTTCTTTTATATCGAGTGGTGTGACATCCACGTAACTGTTATTGTGGATTG
AGATGGATTTCTGACACTCTATCTATATACGGATTTCTGCCGTTTACAGCTTTTGCTC
TTATCGGCGGAATCGCGGCCCTTCGAGAGAATTCGAGTCAGCCTAGACAGCTTTGGT
TTTATCGGCGGAATCGCGGCCCTTCGAGAGATTTTGAGTCACCCTAGCGATACAATC
ATGAAAGGGGGGGAAGGCCCAACGACTACGCTATTTGCTTTTCATTATTAAGGCCCG
CTGTACTGCACTGCAAAAAACTTATGTCTAGCCAAACTTATTAGGGCCCGCTGCACT
GCAAAAAACTTATGTTTAGCCAAACTTATTAGGGCCCGCTGTACTGTGCTGTAGACC
AAAGTTTCCTCGATGAGCTGTCAGAAGCCGAAGTTGTGCCCTCGATTGCTCGGGAGG
ATAACGCTTCCGAAGTTCGGTGTTGGTTTTTGTCGCTTGATTTTAGGGTTTTCCACCA
ATCCGATTTTCCACCCTTTTAATCTTGTTGTAGGCCCTAGATTGTTCGGGAGGAGAAC
GCTTCGAAAGTTCGGTGTTGGAATTTGTCGCTTGATTTTAGGGTTCTCCACAAACCTG
ATTTTCCAGCCTTTTAATCTTGGTGTAGGCCTTCGGATTTGTTGGAAAAAATTTCCTT
TCCCTTTGTATGCTAATCGAGAGAGATCTTGCCTGTTGTTGTAATCTCAGATTGGAAT
GACATGGCCCGAGAGAAAATAGAGATGAAGAGAATAGCTAACGCTTCGGCGAGGC
AGATGGCGTTCTCGAAGAGGCGGAGGGGGTTGTTCAAAAAAGCTGAGGAGCTATCG
ATTCTATGTGCAGCAGATGTAGCCCTCGTCGTTTTTTCTTCCACTGGGAAGCTGTACG
ACTACTCGAGAATCGAATTCCCGCGGCCGCC

>*Pg*SAL5 promoter
TTAGGCATCCAAATAATGATAGCCATAGAAAACTAGCAGTCCCTGTAATAGAAAAA
ATTACATGGAAAAGTAAATTCCAAATATTGCCAGACATTTAATATATTTAATTAGAC
TCATTCTAAGTCACTTGATTCGTGTTTGTATTCAAATCCAACAACAATGAAATTAGG
AATATAATCACAAGTGACTTAGGATATAACAATTTTCTACTAAGTTTGGCAAAGAGA
TTAAGGATGTTATAACTGCATATTATCGAACAAATACAAATATTTTATTTACAACTTC
AAAGTCCATCCAAAAATATTTTTCAAAGAAAATGGTGACTATAGACTGAATATTGGG
AATATTAGGTCAATTGACTTGATTTTGATAGTGAGATGTCGAAGCTCGGGGCTCAAA
CATATATTGTGATTATTAGGGACCCATTTTCTCACAACGGCTTGAGAGGTTTGCACC
AGCCAATTATGAATTTAAAATTATTGCAGTCCATCAACCTAAGATTTTGCTCATAGC
AATCCCAGTGACGAAGGGCACAACTACAAATCAATTCCCCTAATCAATTGAAATAC
AACACAGAGAGATATAAGTTGTGGGTAAAAGGAAAATAACGCTAACCATGTTAAAA
AATATTAATCAAAGAGAAATCAATGAGTTTGCAAGAAATCGGTTCAAAAAAAGCGG

AAGAAAAAAAAAAGTCCGTTCAAAACTCCGCAAAAAAAAAAGCAAAAAAAAACAA
GTGTGGCTTTAGATATATTAAATAAAAAAAATTAAATAAAAAGGAAAAGTACAAAT
AAAAAAGACAAAATCAATTAACAGATGACTCACATGAAAGGAGCTCCAAATCATAT
TTGAGAAGAGAGGAGAGGAGCTTCAATGGAGAATATTTCACAACACTTTACCATCA
CATACTGCCTTAACATTGTTCCGAAGCACTGAAAGAAAATTGCACATAGAATAACCT
GACATTCACCTCTTCCAATGAAAGAAAAGTATAAAGATAGAATTTAAAGACATTGA
AATAGAATAAAAAAGTACGCCTAGGGTTATCAATTTCCAAGCAACAACTAGTATTAT
CAAATATTACGTAGTCATTTTCCCACAAGAAGAAACAAGAAACCTGGCTTTGTCACA
GGTCGATCAAATGTATTTAATTCTTTCAAACTCTTAATATTCCCTCTCATTGCAACAG
GGATACAACCATTTACCATCCCATTATGCCAAAATTCAAGAAAGTGTCACTGTTCAA
ATTTGAGCAAACATATGCTAATGACAAAGTTAAATTGTTGTAATATCACAAGTTGCT
CATTATGCCAGCAAAAGTAATCAAACTTGAATGCTGTTTAGGAACGTTTTTACTATG
TGTAGTGCCCATATTTAACCATTAATATATTACAGTTAAAATAATAATAAAAAAATA
AATAACAAAATAATAAAAAAAAATAACAAAGTTAATCGTGAAATACTATACGCATA
AGAATTGGCATCGTGGGTCCTCGCATGAGGCCCATGTGCCCGGTCTCTGGCCCTCGA
TAGCTAACACCTATCTCTGCCAACCTTAAACGCGTCCTTCGGTACCTAACACCTCTCT
CTCCCATGAGGAAATCCGCATTTTATTTTCATGAGGCACTTATATTATATCAGAAA
ATTCAGTCTGGCATGTTCTTATACGTACGATTTTCTCCTGGTGATACCAGTTAGAAAC
CAAAGTTGTGCCCTCCATTGTTGGGGAGGATAACGCTTTCGAAGTACGGTGTTGAAT
TATCTCGCTTGACTGAAACATTTTAGGATGGGAGTGACATGGCCCGTGAGAAAATAG
AGATGAAGAGAATAGCTAACACTTCGGCCAGGAAGGTCACGTTCTCGAGAGGCAAA
AGGGGTTGTTCCAAAAAAAAACTGAGGAGCTATCGATTCTATGTGCAGCAGATGTA
GCC

Appendix 2. Transcription factor nucleotide sequences identified from yeast one-hybrid search. Sequences were obtained by sequencing the cDNA library pDEST22 vector insert. Character in bold are from raw sequencing data. Underlined characters were substituted from the closest spruce clone with the highest sequences identity.

> *Picea glauca SOC1*-like
TGACGGTTTTGAGGGC**A**AAAAA**G**AG**A**GGG**GA**GGAGAATGGTGAGGGGAAAGACTC
AGATGAAAAGGATCGAGAAC**GC**CCACGAGCAGGCAGGTTACGTTTT**C**TAAGCGCAGG
AATGGGCTGCTGAAGAAAGCTTACGAGCTCTCGGTGCT**C**TGCGATGCCGAAGTGGG
GCTTATAGTTTTTTCTCCAAGAGGGAAGCTCTATGAATTCGCCAGTCCCAGCATGCA
GGAAATTTTGGAAAAGTATCAAGACCGGTCGCAAGAAAGTGACATATCTGTTAGAA
CGAAAGAGCAAGATACTCAGTGTTTGAGACGAGAACTTGCAAATATGGAGGAAAAG
ATCAGGATTCTTGATTCAACACAAAGAAAAATGTTGGGGGAAGGGTTGACATCGTG
TTCAATGGCAGAATTAAATAAGTTAGAGAGCCAAGCTGAACGAGGATTGAGCCATA
TACGGGCTCGAAAGACTGAAATATTGATGGACCAAATAGAATGTCTGAAAAGGAAG
GAACTGTTCTTAAGCGAGGAGAATGCCTTCCTCAGTAAAAAGTATGTTGATCGTCAA
TCCATGGACGGTTCAGTTTCAACATCACCTTCAATTGGATTGGGAAGCATTGACAAC
ATTGAAGTTGAAACTCAATTGGTTATAAGACCTCCAACCGCACAAGATCACTTTTCT

GTAAATAATTCCTATTGAGATTGTTGTTTCAAAATTCAGACATTATTAATTCATTATG
AATGTTTTCATATGATCAGAAATGGTAGGCCAAGGAACTTTTCCAGTACAGCCTCAT
TATTTGGCGAAATTCAAAAGAACAGTCGTCGCTGAATATGGAATTTGAACCAAAAG
CAGTACAAGAAGTCTTTTCTTTCATTTTCGTTTCAAAATCTGTAAAATAAACTTGTCA
CTTCGTTATATAAATAA

>*Picea glauca FLX*-like
CTGAAAAACCCTACTCCATACATATTTTAAATCCTCAGGATCCATATGGCGGGAAGA
AATCGTCTACCACGCCATGCTCTGAATGGTGGTCCACGTGGCTTCCCTCCTGGTCCTG
GTCCAATCCGCGATGGTCCCTACAGGCCAGGACCTCTGCCTCATCCAGCTCGTTTAG
AGGAAGAGCTAGAATTACAGTATGAAGAAATCCAAAGGCTTCTTGCAGAAAACAGG
CGACTTGCTGCGACACATGTGGATTTGCGCCGAGAACTTGCTGGTGCTCAGGATGAA
TTACACCGCCTCAATCAAATTGTGGGCAATGTAAAAGTTGACAAAGAACGGCAAGC
AAGGGATCTGGTCCATATGCAGAGGCCTATGGCTTACATCTGTCCCAAGGTGGTGTG
GAAAAAGGCTCTCAGTATGGGTCTGGATCTGATCCCTGGGGATCCTTTGAAAAGCAA
CGATCCCATGCTCGCAGATAAACAGAAGGGCATAATACCATCTTAGCAATGCAGTCT
TGCTTGCTACTTTTTAAATTTTAATCGGTTATCTTGTGCTGAGAGCTTCCAGTGTCAA
AAGGATTGATGTACTTTGTAGAGATCATGGAGTGCTTAACTCGTCGAAGCTTTTTTA
CATCTGACATTTGAATTTGAAGGAGAGGCTTGTTCTTTGTAATAGGTTTTGTCTTGTT
ATCTCTTTGACTTGATACTTATTTTAAGTAGAAAATTGTGAAACTCA

>*Picea glauca ASR*-like
TATCTG**CTC**T**G**C**G**T**C**GCTCTC**G**T**CG**TTTGTGATTA**A**T**C**TCTCTCTG**T**GTGTTGGGTTC**G**T**A**
ATCCAG**T**C**G**C**C**G**C**T**GT**A**T**AAT**C**T**CTGTC**TTT**CTCTGTGTGT**CCTAATGACT**GA**CG**G
AA**GTGCCACCA**C**CTCTT**C**CGC**CA**CC**GAGAG**GAGGA**CG**AA**T**A**CC**A**TT**GCGTCA**A**C
TCGG**G**C**TA**T**G**C**T**A**CT**C**T**GGGAC**CC**C**TA**T**GG**GT**C**C**T**CT**GAGTAC**CC**C**CACTGG**AT**C**T**GG
GTAT**CCT**TCCC**GGACTGTTC**ACC**ATA**C**TG**CC**TC**TCCTT**A**T**AA**TTTCCCGCGCG**A**C
TTA**TCATACT**G**GC**TC**T**G**GCTATAACGCCGGCAGT**GATT**AT**CC**T**A**C**T**G**GT**TCTG**GC**
TAT**A**GT**G**C**C**GG**AG**ATG**ATTA**C**CAG**A**C**T**G**G**CT**C**TGAT**TA**T**C**A**C**A**GCG**G**C**TC**T**C**G**TT
ATAATG**C**C**G**ATTC**G**GGCTATAACGCCAACACTG**G**TTATCA**G**CAA**G**ACCA**A**TCA**G**AT
GATTATGACA**GA**G**C**TCGACAGGAGGTTAAAAGCGACAAGCGTA**AG**GAGCACGTCG
GGGAGCTCGGA**G**CCATG**G**CTGCCG**GA**G**G**CTATGCACTGTATGA**G**AA**G**CATGA**G**TC**A**
CA**T**AA**G**GATCCTGA**G**AATGCTCGGAGGCACAGGATA**G**AGGA**G**GA**A**GTCGCTGCGA
CGCTGCTGTTGGCAGC**G**GTGGGTATGCATTCCACGAGCGCCACGAGAAGAAACAAG
ACGAG**G**AAGAAGCCGAGGAAGCTGAGGGTGGCCGCAAGCACC**G**CC**A**CCAT**C**T**C**TT**C**
T**A**AGCTTGGCCCATGCCTCT**C**CAATGGCGGGCATTTGGTGCAGTAGC**C**GTAGCAGAG
GGCTATTGCTCTGTGGGTGTCAAACAAATAAATGGAGGGCTATTATCTCCATATGTA
ATTGTTTCCAGTGGAGTTGTAGAGTGTGTATTAGGGTTTGTTATTACTTGAGTTAGAG
TTGAATGGAATTAAGCTAGGGTGTGTGTCT**T**TGCTTCAATAAATATTCCAGAGCCAT
GTATAGCTAGCAGCAGATATGTCTATGTATTTTGCCAAATCTATTTCCTGTGGTATTT
ATCTCATCTCTCATGAATGATGTGCAGGCCAACCTGGTTGTTGTAAAAGGGTTATTT
GCATATAGAAGGAAAATGGCAATCATTTGTTGGATAAAAAA

>*Picea glauca CPC/ETC*-like
GATCATACTCATTCATATATATCTGTCCNTGGAGGATAGCAAGCAACACTTCTCACC
GCCAAAAGCAGAGGGAGACTGCAACGTTACCCCGGGAGGAGGACTAATATCTTTAT

GTGGGAGCAGCGTGGAATATTGCCAGAAGAGTGCTTGTGCTGCTCATGATATCTCTG
CTGATGAACAAGATTTGATAAATAGACTTCACAATCTTCTGGGTGACAGGGCGCCTT
CCATGGAGAAGAGTTGAGGAAATTGAGAATTACTGTAAAATGAGATACACGCCCAG
TACCTCTTCTTCACGCTCTTGAATCTCCCTTCTCTGGCCAGGTCATGGAGTGAGCACC
AATGGTTTTGGAGGATGCGATCGTAATCAGAATATGTGTGGATTTTTTTTTTGTTTTG
TCAATTGAGCATATGTGTGGATGAAACAAAATTGCGGCCGTTTTATCTATTATTTAT
GATATTTCACCATGGAGACAGTTCAATATATACTCNCAGTATATATGAGTGTATACT
CTATTTATTTGTGATATTTCAACTCATACAGTACAGTATGAGTATATAACAGCGTTGG
TTCTATCTTTTAGTTTTTTTTTGGTTGTAAAAGATGTAAATTTGGATAAAAATTATAT
TTTCTGTTGTTTAAAAAAAAAAAAAAAAAAACCCAA

>*Picea glauca MYB1*
CAACACACTTTCAAGGTACACAGATTATAAATTATATACAGAGGAGATTGATCCACT
GAGAGGCAACTTCCTTCCTCTTTCCCGGCCGGCCGGCCCTTATTCTATCTCTGCCGGA
ATCTTCTCCACACAGAGGCCAGCCCTTGAGCCTTTAGAACGATCCCCGAATGTGGAA
AAATCGACGATTCTAAAGCGCGATTCCTGTCTCTTCTCAGATCCTCTCGGCAAAATG
GGAAGGCAGCCTTGCTGTGACAAAGTGGGATTGAAGAAGGGTCCATGGACGGCTGA
GGAAGACAGAAAACTGGTGAATTTTATCACCATGCACGGCCATGGATGCTGGCGTG
AAGTACCCAAGCTTGCCGGTCTGCTGAGATGCGGAAAAAGCTGTAGATTGCGTTGG
ACAAATTACTTACGGCCAGATTTGAAGCGTGGATTATTGTCTGAATCAGAAGAAAA
ACTCATCATCGATCTACATGCTGCCATAGGGAATAGGTGGTCACGAATCGCTGCACA
GTTGCCAGGGAGAACGGATAACGAGATCAAGAATTACTGGAACACGAGGATTAAGA
AGAAACTCCGCCAGATGGGAATCGATCCCGTGACCCACAAGCCTCTCANCCAAATG
CAAATGCAGAGCTCCCCGACCCAGANTCTGCTGCTGCAAGAAAATGATGAACAGCA
GCAGCAGCAGAATGAGCCTGATCAGAATCATAGTAATGGCTCTGCGGAGACATTGG
TGTTGACGGCGAGAGAACCAAACGACGATATAGAGCCTCTCGAGAATTTTAACATG
GAGGATTCCATGCAATTGTTCAATGTCTGCTCGCCCACCAGCGTAATAAGCCTGTCG
GGGAGAACCGAGGAAGTTGACTCGGATGACTCTGACCAGGTCTCCAAGAGCTTCGG
CAATGGCGGCCATGCTCAGTACATTGGCCGAGAAAGCTCTGGTGTGAAGGCCGAAT
GTGGTTTGTCTGTGTGGGATCAGATGGGTGGCGTTTTGGGTGATCCGCTCTCCGATT
ACAATTCGCAGTGGAATGTCGATTTGGAATCGTGGACGGCTGGATTGGACGCTCATG
CGGCTTCTGCTTCTGCGTGGATTCAGCAGCTTCCTGACTGCCAATGGAACGACTTCC
AAGGCGATTTTGAGATCTGCAGCAAGTCATGTCCGGAGACTCTGCAGAGACTGGGG
CCCTTCCTGGATGACGATGAAATGTGAAAAAAGGAGATCCCAACAATATCTCATAA
AGAGATTGTACATTAACCCAGTAAATATGGAGGAGGAGGATGAGATATATAGAATA
TATACATATAAGTTAATATGAATGAAGTTGTGTGTATCTAATTCATTCATTCATTCAT
CCTACGATGTTTTCATGTAACAGATACCACATGGTTTAAACTTTGCCATCTTCATAAA
ATCCNACTCTTAGTTTGTCA

>*Picea glauca NBS-LRR/WRKY*-like
GCAGACGCTCCCAGACTCGGTTGGGAACCTGACGGGCCTCCAAACGCTTGACTTGAC
CAGGTGCTCCACTCTGCAGAGGCTCCCAGACTCGGTTGGGAACCTGACGGGCCTCCG
AAGTCTTTACTTGGGCAGGTGCTCCACTCTGCAGACGCTCCCAGACTCGGTTGGGAA
CCTGACGGGCCTCCAAACGCTTGACTTGAGCGGGTGTTCCAATTTACATATGCTGAC
CAATATTGAGCATTTGAGCTCGTTGGAGAATCTTTATGTGCAGCAATGTCCCAAACT
GCAATGGGGTTCGGAAGTAATCGAGCAGCTGCGCCAACGACTGGGAGAAGGCTTCA

TGGAAGCATACTCCAGTGATGAAAGCGAGTTGGACTCCAGTGATGAAAGCGAGTTG
GAAAATATACAAATGGAAGCATACTCCGGTGATGAAAGCGAGTTGGAAAATATACA
AATGGAAGCGAGTTGTTAAATATACTCCATGAAATCGAGTTGTAAAACATACTCCAG
TGATGAAAGCGAGTTGTAAAATGGGAGTAGACTTCATGGAAGCATACTCC

Appendix 3. Cloning Promoter Baits into Destination Vector.

Materials:

E. coli DH5α cells
LB ampicillin plates
Gateway LR clonase II enzyme mix
pMW#2 and pMW#3 vectors containing promoters
TE buffer (pH 8.0)
AflII or XhoI
β-mercaptoethanol
X-gal
NcoI or ApaI
Whatman filters
15-cm Petri dish
Tweezers
Liquid nitrogen

Z-buffer

60 mM Na_2HPO_4 (anhydrous)
60 mM NaH_2PO_4
10 mM KCl
1 mM $MgSO_4$
Adjust the pH to 7.0 with NaOH.

1. Clone promoter sequences into the 476 p5E-mcs Gateway vector.

2. Take the Gateway LR clonase II enzyme mix from the -80°C freezer and place it on ice. Compose all reactions on ice.

3. Combine in sterile tubes: ~200 ng of pMW#2 to generate DNA bait::HIS3 constructs or ~200 ng of pMW#3 to generate DNA bait::lacZ constructs (should be ≤1 µl), 1 µl of Gateway LR clonase II enzyme mix, and enough DNA bait Entry clone miniprep to obtain a final volume of 5 µl. As a negative control, prepare an identical LR mix without Entry clone but with TE buffer instead of DNA.

4. Incubate overnight at 25°C.

5. Transform the entire reaction mix into 50 µl of E. coli DH5α cells plate onto LB-ampicillin plates (100 µg/ml), and incubate them overnight at 37°C. The negative control should give rise to no or only a few colonies (less than five).

6. Verify the insert size of the sequences in the Destination vector by PCR of at least two colonies per construct using vector-specific primers.

7. Purify the Destination clone DNA by miniprep for subsequent integration into the yeast genome.

8. Digest approximately 1 µg of DNA bait::HIS3 constructs with either AflII or XhoI, and bait::lacZ constructs with NcoI or ApaI in a 20 µl reaction volume. Make sure that the restriction enzyme of choice does not cut within the DNA bait sequence.

9. Verify linearization of constructs by running 1-2 µl of the restriction digest reaction mixture on an agarose gel.

10. Transform linearized DNA into YM4271 yeast and PCR confirm using the procedure below in section A2.

Promoter Self Activation Test:

1. Pick 12-24 individual colonies containing both pMW#2 and pMW#3 integrations, and spot them onto an Sc-His-Ura plate. Incubate them for 1-2 days at 30°C.

2. Replica-plate the spots (A4.) onto the following plates: a fresh Sc-His-Ura plate, Sc-His-Ura + 3AT plates (containing 20, 40, 60, 80, and 100 mM 3AT). Grow 3-10 days at 30°C and monitor the colony growth. Dense colony formation indicates there is a high level of self-activation. We want to select the colony with the least amount of self-activation.

3. To perform β-Gal assays, replica-plate the spots onto a nitrocellulose filter that has been placed on top of a YEPD plate. This ensures growth of the respective yeast colonies on the nitrocellulose filter. Incubate overnight at 30°C.

4. B-gal Assay: Incubate the 3AT-containing plates for 3-10 days at 30°C. Monitor colony growth: Strong growth is an indication of self-activation.

i. Place two Whatman filters in an empty 15-cm Petri dish for each plate to be assayed.
ii. Make a mix of 6 ml of Z-buffer, 11 µl of β-mercaptoethanol, and 100 µl of 4% X-Gal per plate. IMPORTANT: Make sure to do this in a hood.
iii. Pour ~200 ml of liquid nitrogen into an ice bucket, cover the bucket, and place it in the hood.
iv. Transfer 6 ml of the Z-buffer mixture onto each plate containing Whatman filters. Make sure the entire paper is soaked with buffer, and remove air bubbles using tweezers.
v. Take the nitrocellulose filter containing the yeast using the tweezers and place it in liquid nitrogen for 10 seconds.
vi. Thaw the filter at room temperature by holding it in the air using tweezers. Place the filter with the yeast facing up onto the Whatman filter, and remove air bubbles.
vii. Incubate at 37°C. Check for blue-white coloring regularly every hour during the first 4 hours, and take pictures. Continue the incubation overnight at 37°C, and check again

for blue-white coloring the next day.

Appendix 4. Yeast Transformation Protocol.

Materials:

10X TE (50 mL)

10 mL 1 M Tris-HCl pH 8
2 mL 0.5 M EDTA pH 8
Solutions should already be autoclaved, therefore no need to sterilize

1 M Tris-HCl pH8 (100 mL)

12.11 g Tris-HCl
90 mL water
pH to 8
Top to 100 mL

0.5 M EDTA pH 8 (50 mL)

9.3 g EDTA
40 mL water
pH to 8
top to 50 mL

1 M Lithium Acetate (LiAc, 50 mL)

5.1 g
50 mL water
pH to 7.5
Autoclave or filter sterilize

50% (w/v) PEG (poly ethylene glycol 3350, 50 mL)

25 g PEG
50 mL water
Autoclave, seal bottle with Parafilm around the lid to prevent evaporation (PEG is very volatile when dissolved in liquid)

1.1X Te/LiAc (fresh, 10 mL)

1.1 mL 10X TE
1.1 mL 1M LiAc (10X)
7.8 mL DIW

TE/LiAC/PEG (1X/1X/40% final conc., 10 mL)

1 mL 10X TE
1 mL 1M LiAc
8 mL 50% PEG

0.9% (w/v) NaCl (50 mL)

0.45 g in 50 mL -> filter sterilize

YPDA agar plates (1L, scale appropriately)

10 g yeast extract
20 g Difco bacto-peptone
0.5 g adenine hemisulfate
Top to 950 mL water (when adjusting for scale, adjust this volume but keep glucose at 50 mL)
pH to 5.8-5.9
20 g agar

*glucose needs to be autoclaved separately or else it will caramelize in the solution. After autoclaving the 2 solutions can be combined, and then poured into plates

40% (w/v) Dextrose/glucose (50 mL)

20 g glucose in 50 mL water
autoclave separately from rest of YPDA solution

* After pouring plates, leave dry them at RT for 2-3 days (i.e. leave plates unsealed in a bag) before storing at 4°C for several months. Alternatively, plates can be dried under a flow hood for approximately 2 hours.
YPDA liquid medium (will need 3 – 12 mL for start cultures, 150 mL for larger cultures) *use same recipe as YPDA plate but omit the agar

Appropriate SC (synthetic complete) selective medium (1L)

6.7 g Difco YNB (yeast nitrogen base) w/o amino acids, with adenine hemisulfate
2 g amino acid drop out media (e.g. –His, -His-Ura, or –His-Ura-Trp)
20 g agar if making plates (if making liquid media omit)

40% (w/v) Dextrose/glucose (50 mL)

20 g glucose in 50 mL water
autoclave separately from rest of SC solution

*combine glucose and rest of SC media after autoclaving, pour plates, allow to dry for 2-3 days at RT, store at 4°C

- Frozen stock of yeast cells
- Sterile, DIW
- 1 Oakridge tubes for every 50 mL yeast culture
- 1-4 culture tubes
- 1 250 mL flask (sterilized)
- 1 500 mL flask (sterilized)
- Heat block (100°C)
- Water bath (42°C)
- Incubator (30°C)
- Parafilm

1. Streak yeast strain (from glycerol stock) you wish to transform on YPDA agar plate. Incubate upside down at 30°C until colonies appear (~3 days).

2. You may break here by sealing the plate with Parafilm and storing it at 4°C (dark) for up to 4 weeks. If you wish to proceed (start early in the morning!), inoculate 1 colony (2-3 mm diameter < 4 weeks old) in 3 mL YPDA medium in a culture tube. (If you wish, you can set up 3 separate 3 mL cultures from 4 separate colonies and choose the culture that grows the fastest to proceed with for the transformation. The faster growing cultures tend to result in higher transformation efficiency).

3. Incubate at 30°C 250 rpm for 8-12 hr.

4. Transfer 20 µl of the culture to 50 mL of YPDA in a 250 mL flask. Incubate at 30°C 250 rpm until OD600 reaches 0.15-0.3 (16-20 hr). Do not over grow! If culture is not at 0.15 by 20 hours, just continue growing until it reaches the minimum OD600.

5. Centrifuge cells at 700 x g for 5 min RT in Oakridge tube. Discard the supernatant and resuspended the pellet in 100 mL of fresh YPDA in a 500 mL flask. (Discarding supernatant also discards yeast waster. Using fresh media will help the yeast grow faster and they will be healthier).

6. Incubate at 30°C until OD600 reaches 0.4-0.5 (3-5 hr). *Turn heat block on to 100°C, water bath to 42°C.

7. Harvest cells by centrifuging at 700 x g for 5 min RT. Can reuse corresponding Oakridge tube from previous step. *During this step remove tube of herring/salmon sperm DNA from freezer and put in heat block at 100°C for 5 min to boil. Afterwards, move tube to ice to cool. Right before transformation return tube to heat block to boil for 5 min.

8. Discard supernatant and resuspended pellet in 30 mL of sterile water.

9. Centrifuge at 700 x g for 5 min RT. Discard the supernatant and resuspend in 1.5 mL of 1.1XTE/LiAC.

10. Transfer cells to microcentrifuge tubes and centrifuge on high for 15 sec RT.

11. Discard supernatant and resuspend each pellet a volume of 1.1X TE/LiAc that corresponds to the original culture volume, multiplied by the final OD600, divided by 100.

e.g. $\frac{\text{50 mL culture * OD600 of 0.5}}{100}$ = 0.25 mL (i.e. 250 µl) 1.1XTE/LiAc

The cells are now ready to be transformed. For best result, transform immediately, although they can be stored on ice for a few hours without significant loss in efficiency.

12. In the following steps parenthesis are used to indicate differences to be performed in transformation steps whether you are doing the transformation of a linearized plasmid for yeast integration, versus when you are transforming your cDNA library into yeast.
In a prechilled tube (1.5 mL tube for small scale OR 15 mL tube for library-scale) combine in the following order:

- 100 ng to 1 µg of linearized plasmid for small-scale OR 15 to 25 µg of cDNA library for library-scale.
- Herring sperm DNA (carrier DNA 10 µg/µl) – 25 µl for small-scale OR 100 µl for library-scale
- Competent cells – 50 µl for small-scale OR 600 µl for library-scale; gently mix
- TE/LiAC/PEG – 500 µl for small-scale OR 2.5 mL for library-scale; gently mix

12. Incubate at 30°C – 30 min for small-scale (mix by tapping or gently vortexing every 10 min) OR 45 min for library-scale (mix every 15 min).

13. Add DMSO and mix – 20 µl for small-scale OR 160 µl for library-scale

14. Incubate at 42°C in water bath – 15 min for small-scale OR 20 min for library-scale.

15. Centrifuge to pellet east cells – high speed for 15 sec for small-scale OR 700 x g for 5 min for library-scale.

16. Remove the supernatant and resuspend in YPD Plus Medium – 1 mL for small-scale OR 3 mL for library-scale. YPD Medium Plus sold separately by Clonetech, but can purchase from Zymo Research cat# Y1003-50 for 50 mL and Y1003-100 for 100 mL. This is a specialized medium to increase east transformation efficiency y >50% compared to YPD medium alone.

17. Incubate at 30°C with shaking for 1.5 hr.

18. Centrifuge to pellet cells – high speed 15 sec for small-scale OR 700 x g for 5 min library-scale

19. Discard the supernatant and resuspend in 0.9% NaCl – 1 mL for small-scale OR 15 mL for library-scale.

20. Spread 100 µl of 1/10 and 1/1000 dilution onto 100 mm plate containing the appropriate SD selection medium (e.g. SD-HIS for pMW#2 vectors). Do not plate undiluted transformed cells, colonies will be too dense.

21. Incubate plates upside down at 30°C until colonies appear (3-5 days).

22. Calculate transformation efficiency

$$\text{transformation efficiency} = \frac{\text{colony forming units (cfu)} * \text{suspension volume (mL)}}{\text{volume plate (mL)} * \text{amount of DNA (µg)}} * \text{dilution factor}$$

example:

$$\text{Transformation efficiency} = \frac{300 \text{ cfu} * 1 \text{ mL} * 10}{0.1 \text{ mL} * 0.1 \text{ µg}} = 3\text{x}10\text{^}5 \text{ cfu/µg}$$

Yeast PCR Screen:

1. Aliquot 14 µl of Z-Buffer and 1 µl of resuspended Zylomase enzyme (diluted according to manufacturer's recommendation, Zymo Research cat # E1005) into 0.2 mL tube for each colony being screened. The buffer Zylomase enzyme is resuspended in contains β-mercaptoethanol, so perform this step in a fume hood. The enzyme has a low solubility so I usually quick vortex it 3x in the resuspension buffer before aliquoting it into tubes.
2. Use sterile pipette tips to carefully and gently remove a large chunk of the colony (I remove ~3/4 of the colony). As long as there is a tiny bit of the colony left on the original plate, you can regrow more of the colony to use for other purposes.
    - I found yeast colonies do not readily dissolve in this buffer so you must pipette the colony off your pipette tip several times

3. Incubate tubes (ZYLOMASE program):
    37°C – 30 min
    95°C – 10 min (heat inactivate enzyme)
    4°C – Hold

4. Pellet debris by centrifugation at 700 x g for 5 min (since the centrifuge that will hold 0.2 mL tubes is in the MBSU I usually just centrifuge tubes in out 0.2 mL bench top quick spin for ~3-5 min. The result is not as quite good as the machine, but it works).

5. Remove supernatant and transfer to a new tube.

6. Add 85 µl of sterile water to dilute template.

7. The lysate can be stored at -20°C for subsequent PCR reactions.

8. Set up 50 µl PCR reaction according to Neb taq polymerase protocol, with 5 µl of yeast lysate as template. Use primers specific to your promoter to confirm presence (e.g. same primers used for previous cloning techniques). Run PCR with the following program (note this is the program used by Deplanke *et al.* 2006. They did not explain their logic so some steps may be able to be shortened, but I did not spend time testing this):

i. 94°C – 5 min
ii. 94°C – 1 min
iii. 55°C – 1.5 min (they did 56°C, but my primers had low Tm's)
iv. 68°C – 3.5 min (could probably shorten this step based on your promoter length)
➔ Repeat from step ii. for 29X
v. 68°C – 5 min
vi. 4°C – hold

9. Run 10 µl of PCR reaction on 1% agarose gel to confirm insert.
Note: Deplanke *et al.* 2006 stated that they had trouble with PCR efficiency from colonies grown on selective media, therefore they restreak colonies they wish to test on YPDA, grow for 2-3 days and then PCR screen. I found my PCR's were successful with colonies grown on selective media, but if you are having problems you can try restreaking, it will just take longer to screen.

10. Once confirmed, restreak colony you wish to make a glycerol stock from, restreak remaining part of colony on a YPDA plate and allow to grow for 2-3 days at 30°C.

11. Inoculate 3 mL of YPDA with colony and grow overnight at 30°C at 200-250 rpm.

12. Add 900 µl of yeast colony to 900 30% glycerol and store at -80°C.

Appendix 5. Yeast Replica Plating Protocol.

Materials:

2 96 well plates
Selective media or YPDA media
Selective plates
YPDA plate for B-gal screen
Note: I did not find the B-gal reporter gene to be very informative, so this step is optional.
Selective plates+3-AT (low to high concentration depending on needs)
Replica plating tool
Sterilized velvets

Procedure:

****This technique should be performed after yeast has been screened against the cDNA library. Positive colonies will grow on the selective media. You will PCR screen these colonies until you have found enough unique/diverse sequences. After which you can purify the PCR reaction for sequencing. Part of remaining yeast colony can be used to make a culture for a glycerol stock (plate or tube type). The replica plating can then be performed. After which you can recover the transformed plasmid with the Zymo Research Plasmid Extraction Kit. This step is done last because you will want to thoroughly screen your positive colonies first before you deem them worthy of extraction. Meaning:*

- *you will want a unique sequence (we don't need 10 plasmids of the same sequence to be 3-AT screened),*
- *you will probably want interesting transcription factors (housekeeping transcription factors may be of less interest to you), and*
- *you will want true positive interactions/strong interactions (the 3-AT screen will reveal if there are any false or weak positive interactions).*

1. Using a 96 well plate, pipette 160 µl of media (selective or YPDA, depending on purposes) into each well. With a pipette tip inoculate each well with your desired colony.

2. Grow the plate O/N at 28-30°C with 200-250 rpm. (Some protocols say you can grow the culture up to 72 hours). Make a glycerol stock plate: add 80 µl of 30 – 40% glycerol to each well, combine with 80 µl of the respective well in the culture plate.
e.g. 80 µl 30-40% glycerol + 80 µl culture = 160 µl total in each well

3. The remaining volume of the cultures (~85 µl) can be used to create a plate that will be used for replica plating. Pipette 3-5 µl of the culture onto the selective (SC) plate (here after referred to as the "culture spot") of choice to later replica plate onto the 3-AT plates and the YPDA plate for the B-gal screen.

4. Allow to grow for 2-3 days. When ready, replica plate the colonies growth on the selective original plate. You may have to do more than one 3-AT screen in order to narrow down the range of 3-AT concentrations you should be using to discriminate against strong and weak DNA-protein interactions.

E.g. SC 0 mM 3-AT (control)
SC 5 mM 3-AT
SC 10 mM
SC 20 mM
Optional: YPDA (for β-gal screen)

* the 3-AT plates can be grown 3-5 days (Deplanke *et al.* 2016 says 3-10 days, but I found this causes the spotted colonies to become overgrown and begin to merge with adjacent spotted colonies)
(*β-gal plate should be grown overnight for 1 day)

5. To replica plate, place the sterilized velvets over the base of the plating tool. Apply the ring over the velvet to keep it in place. Ensure the ring is far enough down the plating took that it will not touch the plate when it is applied to the velvet (i.e. don't place the ring too high or else it will contaminate your plate).

6. Make a mark on the velvet/ring so that you can have a marker to late identify each colony (a simple line on the top of the plate will do). Make this mark on each plate and align the line on the plate, with the line on the metal ring so that each plate will be identical in orientation.

7. Place the original plate on the velvet face down. Pat gently with your fingertips. About 10 gentle pats in the middle of the plate (approximately one pat for each row of colonies), followed by 10 gentle pats along the edge of the plate. Repeat this action so in total the middle of the plate will have approximately 20 gentle fingertip pats, and the edge will have approximately 20 gentle fingertip pats.

8. Remove the original plate and place the lined-up replica plate on the velvet. Press down gently approximately 10 pats along the middle of the plate. Remove the replica plate. You can use one velvet to replica plate 5-6 plates.

9. Clean the replica plates with a fresh velvet. You will need to press down much harder to clean the plate. Ensure there is at most a haze of cells on the plates. Too many cells will create a boundary layer so yeast cells will grow on top of the primary yeast layer and lead to false positives. I used a minimum of 5 clean velvets to clean my replica plates, but the number of required velvets will vary based on your colony density and how hard you press down on your replica plates and the cleaning velvets. Continue to clean plates with a fresh velvet until not colonies are visible, or only a slight haze of colony spots is visible.

10. Allow plates to grow at 28-30°C for 3-5 days. Take photographs of plates using a hand-held camera for larger shots to display the difference in colony growth within a plate. Photos of

specific colony spots can be taken with a dissecting scope camera in order to have a more accurate picture of the growth of each colony spot for figure photos and comparisons.

11. Cleaning the velvets:

i. Use a bristle brush, with water or water and soap, to scrape away any colony residue.
ii. Autoclave the velvets in a large container (e.g. 2 L beaker) filled with water.
iii. Dry velvets.
iv. Stack velvets in a pile and wrap in aluminum foil. Autoclave velvets (include a dry time). You may wish to leave these autoclaved packaged velvets in an incubator over night to ensure all residual moisture evaporates.

Appendix 6. Additional sequences obtained from the yeast one-hybrid screen. Seqeunces were obtained by sequencing pDest22 Fw and Rv PCR products with pDEST22 Fw, pDEST22 Rv or oligo dTs. Well i.d. refers to the corresponding frozen glycerol stock identity for identification and use in future experiments.

Promoter	Well i.d. (Promoter - Plate - Well)	Sequence	Length (bp)	BlastN Search
SAL1	C12-P1-B3	ATCAAACAAGTTTGTACAAAAAAGTTGGGGAGTTTCTT CATCTGTTCAACTATTTTCAGGACTTTTGAAATCAAAGT ATAATTATGTAAGGTGTGGATTTATTATCGTTTTAGAAA GGCTTCTTCTGCGATGTCAGCTGTCAGTGTTGGATGAAC ATGAACTTGACCAAGGAGCAGGTGACAGAGAATTTAG GCTAGATGGCAGTGACAGCTCATACATCCAGGAGAAAG CAAATGCTGTTATTGGCTTGATGAATGGAGCTTTGTCAC AAATATTTCTGGCAAATGAGACTAACCGTATCAATATT TTGAAGATGTGTGACCTGCTTTTTTCTCAACTTTGTTTG AGAATATCTCCATCAATGGCATCTTCTATTTATAGTAAT GCATATATAAGAGATAACACAGCCTTTGATAGTCTCAG CAATAGTTATGTTGGCTATGATAGACCTGATAATTTACT GTTCGGTACTGANCAAGCGGANANNCATTATGGAGATG ATAGACATGCAAACAGATCTGTATTTGNAAGTGGTTTA ATCCCTTCTAATTGTANTTCNACATCNNTGGNAGCATTG CTACTAANGNGNNANGCTGCANCCCCTACGCAACTAGT AAAAATGNATTCCTACAGNTCTTTTGTATNGGCCACTG ATTNATCTGNNTGGGGNAGCCACTGANGATATGNNNTT NGGCNNTGNTGNTGGNAGCAAGGGAAGAGGANATANA CCTNNNNNNNCNTCCNGATACAC	751	unknown

SAL1	C12-P1-C3	TGCCTTCTGCTTCCTGCAGTTGTGAGCTTTGAGACAAGTC AACTCTTGGCCATTTTCGCCTTCTGTTCGGAGGGTTTTTC CAGGGTTTAAGGCGTTTTTTGTCAAGTTCAGAGGGGCTG CTACAACCGGTTTGTGTTAAATTATCATGGCAAGAGCTA TGGCAACTGACTTCGCAAGGAGTCTGAATGCCCAACAAA TGTTGGGCTCTTTTATGGAAATTATTGTGGCTGTGAGGCT TTGGAGAATATCAGAGGCTTTGGTTCTCATGTCCATGGG GATGTTACCTGAATTTGCACTAGAGGATCTTCTTGCAGC CAAGTCATTCAACATGGAAACTGGACATGAAGAAGCCG AAGAAAATTTACCAAAGGTTCAGGATGGGTCTAGTGTTT TATGTAGACCCAGTTCAGAAGAGGAACTTATTGATGGAA GTGATGACATAATTAGTTTGGGAAGCAAAACAACTATCA CATCTTCTGAAGTAGCTGAATGCAAAGATGATGCTGAAA GCGATGATGATGACGAAGATGATGATGATGAAGATGGA GAGGATGATGATCAGGAAGAAGAATGTGGTGATGAANA AGGTTTGTCANCGNATGAAGGTGCTGANGATGGACCNC ANGAGAATGCGGNCGANGAGGAANAANANGAGGAGGC TNATGGCAATGACGAANAGGAANANNANGATGATGATG ATGATGAGNAGGACGATGACNNTGATGANGNGNTGATG ANNTGACGAGNAGGANGAGGAGNNNGAANANCCTCCTG CCNNNANNAANAAATGANANACTANCCTT	803	unknown
SAL1	C12-P1-D2	GGAGGAACTGCTGAGGGAGGAACTGCTAAGAGAGGAAC TTCGTCGAGAGGAACTTCGCCGAGAGGAACTGCGCCGA GAGGAACTGCGCCGAGAGGAACTGCTAAGAGAGGAGCT GCGCCGAGAGGAATTGCGCCGAGAGGAAATAATGAGAG ATGATATATTACGTTATCAAGAGGAAACAAGACGTGCA GCACGTGCTGAGTATAATATTCCTCAGGCTCCATTAGCT GGATATGGTGCTGATCCTGCTATATCTGAGAGGGACCTG TTACGATATGGGGCTGGAAGGGATTATATTCCTCAGGCT CCTCTTCAAGCTCAAGATCCATATGCGAACCTTCAAACC CAAGATCCATATGCGAACCTTCGAGCACAAGATCCATAT GCGAGCCTTCGAGTACAAGATGCATATGCTAGTGCATTA CCGAAGCAACTCCCAAGGGATCAGCTGCCTGAGTTGGAT	779	unknown

		TATCAGAGCCGGCTTGATCCAAGCATTGAATCTTTATAT CGTCAAAGGTTGGATGTGGATTATCGAAAGCCGATAGTG GATGTTGCAACTGAAAGCTACTATGCAGACCCTCTCCTG CAAAGGGATTTACGCCGACCTGAGTTGGGANCTTCTGTT GCTGGGNCNCCTCCTGCATATCTTGNAGCATCTTCACTG NATCGGTAGTTNTATAANCTGCTGAATTTTGAATCCNNN NTTACTATCCCNTCGNNCTATACTTGTTAGTTGTGAATCA NNCTGAANAANTGANTTNNNTGGTACGGTGCTGGACTAT TTT		
SAL1	C12-P1-F9	ATATAACTATCTATTCGATGATGAAGATACCCCACCAAA CCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGT ACAAAAAAGTTGGAGCTTCATATAAACGCAGCAATAAG CGATGGGATGCGATCCGGTTACATAAGCTTCTCTCGGCT ATCTGATCTCCGGTACAAGGATCTTAATTTCAGGTTTTCT ACATTTCAGATTTGTGAAGTGTATTGGGTGTATTTCTGGT CTCCAATGCTGTTCTGCTTCTAAGAGTTGCATCTGGATAT ATTGTTGCCTTGCATANCTGCNGTGTATACTAATTTCAAC ATTTCTTCNNAANACTGNNAANAANGGG	342	unknown
SAL1	C12-P1-E11	GGAAACCCAGAGACCCAATAGCAAAACGGCAGGGAAGC AAAGAAAAATCGTCGATGGCGGAGGTTATGGGATCAAT TCTCCCTCGTACCTCCTTCCTTTCACACAAGGCATTTAAA GGCAAAGCAGCAACACCATACAGAGTGCCTTATACAAG ATCAATGCCGCAGATTACCATGCAAGCAGAGAGAACTG TCAGCTTCTCATCCGAACTTAGCACCGATCTTCCTCTTTA TGAGCCTTCTGAGGTTCCCTTTGAGCAATATTTAAGTGA CAGGGAAAGAATATTTCAAGCAATATTCCCAGACAAAA GGAGGAGCGAGAAACTAAATGATGAAGAATGGCGAATT CATATGTTGCCTATTGAGTTCCTTTTCTTGACTGCATTTC CAGTCATTGATATGAGTATTATAGTGAAAGCACCGGGGC AAGGATATCCCCCGGGTATTTCAAAAAATGTTAAAAAAG TGCTAACCTTGGAAGCTACAAGATGGGAGCTTCGAGGCT TAGACTATGTTTTGCAGCCATCAGACTTTGTACTCGGAG TTCGTGNAGCTCTTTACTCANAAAATAATGGGGGNNANN	772	unknown

		NCNAGANNAAAGGNANTGATGGANANGANTGTTANCTT TGNATTANCTCCAGCACTTGCTGNTATTCCTGAANANNT TNTNNNAAGCATTGGACACGCNNTTNNGATTNAANTGNT GGAGANCATGNNGGNANNANTCNATANNAAACTTCNTG NCNATTACNNAGATTANTCNNNNNNNCNNAAATTGC		
SAL1	C12-P2-B7	NAGACTACNNANNANCCATGTTTNNNATCACCNNNTAG GGNNNGNNAATTATTATTCANCCNCNTACATAAACGTCA NTTGNTNTCTNNNNNNANTNNCGGANTGANCCTCNNAA TGTGGCTTCCAAATGCAGGACTAAGAGAGTTCCACTTGA ATTCNAANTNNNNGGAGAGCCATTAGCCCTCACCTAATT TAATANNTCTACTACTAACATCAGCTTCTCTCTCNAACCT ACAGCTCNNNGCCTCCGGTGTANNCNNNTCATGGAGAT GCCAGGAACCAGAANANNAGGAGACGCACACAAACTGN NNNAGCTTAATTGNNTTCGGGCTTGTGNACGATGAANGA NATGCNCTGGACTTGGCGGACGNTGTCGAANCCNANGA CGCGAATGANNGCTTTCNGNNATGCCTTCNNNCNNTCNN TCACCNNNNNCAACACCNGNNANGNCTCTGTGCANNCG AACATTGNCAATTTCCNCNTNACCNANAATCTNNNNNNN NNNTGCCCAGGGNNNNTTCCNNACTTCCNCNTNANGAT NCCNCCCNCGTGNANTTCCAGGCNAGGNANCCATTTGTN CCTCANCNNGNANNCNNNCTCCNNNNNNANNTGNTCTT NANNGANCCNNNNGANANNCGNNNGANTNTNNNNNNN TGNNNNGNNNNANGNAGNCCACANNAGNNNGCNTCGN NNNNNCNCCNCCNNNGCTNGNANNCNNNTNTCTCCCAT NANGNGNCNACC	741	unknown

SAL1	C12-P2-C10	GCAATATTTCNNNNNGNACCCNTCTAACATCATACTCNA ACAAAATTGATTTCNCAACTGTACATCATTTATCCNCGTT GCTCTCATGATATACNGNGNNNNNANNTCCNAGATTCCA AGANNCNATAGTAACTACTCTTTAGGCTCCCAAGACTTT GTCTGGACACCTAAGTTGCTCTTGCAGCTTCTGCTGTAG AAGTTGATTCTCCTCGGTTAGTAGTAAAGCCTGTTTCCTC AGTCTTTCATTTTCTTCTACGATGCACTGATTCCGTTGAA AAAGCTGTAAATTTAACCTCTCNATGCGAGAAGCCTCTT TGTCTATCTTCCTCTTCGANCTTCTTCTGAACGCAAGCCA CGCTTTGGCCGCCCTAGGAGCCNNNGAGTACNAAATCG GACGCCTTTTAATACGTTGCTTAAAGGANAACGAACGCC GCTGGAAAAATNGGCCCANNACNACCTGCTCGCGNCCA ANCCAAAGCCTTTGCGCGGGTCTTCGGCGACTCNNNCTC NAGACTGTCCAAATGTNATCTAACGGNNNNNNANNNGN NANANCTATNCCCCTGNNAANNCCNCACCCTCNGNAAN ATTTTNNANAATNNCAGNNNNNTNTNNTGNACAAACTN GTTTGATTCGANNCACCCTCTTTTTT	649	unknown
SAL1	C12-P2-E12	AGTACCCACAACAACCTTATGAGAATTTCCCCTTAGAAT GGCATCAATCCNTATTGACAATCAACAGATCTCAGCAGA AACTGAAGAAAAAGGTGAANAATACNNCAGACATTTCN NTGGCCGCAAAATTCTACCTAAAGATGTAATACAAACGA AAGAAATCACTACTGTACATACCCAGCGAGCGTTTACTT CGGTAAAAGGTTTCCACGATCTTGTTTTGCCTTGGCATA GCATGTGTAGTAAAACAGGTTCAAAATCAAAGTAAATA CAGTGAAAGCGGCACCGGCCGCAAAAACACCCTTTCGC AGCGTCTCACAAGACAGGTCATGCATAGAAAAGTAACC CCTGTATTTTGTATGGTAGGCATTCCTCGCCGCCCCTGCT AAGAAGCACGCCTCGGCAATCAAGAACGTTACCCAGCA AATGAAGAATAATAAGATCGCCCAGACACGAGATCCAC CAGGCTTAAGAACGCTTCCGCAGCACAAGCACCTCGTGA AGGCCATGACAAGAGNCTGGNTTGCCAGCAGAAACAGA AACCCCCCGACTCCATACCAAGTAGAGATGTCCGAATTG TAGACGCAGTAAGTTCTCTCGTCGTACTGGNCCGGTTTC	867	unknown

		ACGACTCCAGTGCTCCTCCTGCGCTCTGCGGCCACTGCG ANGCCGAAAGCAATGANNNCGAAGATGAAGATGACGCA GATAACCNCCTTTGACGCCNTTGCTTTTTGCTTGTTAGCA GAAACTGNNATGAANANGGATGGACAGATACCCCACNT TTTTTGNACAAACTTGNNNGATTCNACCCNCCCTNNTTTT TTTGGNTTTGNNGGGNNATCNNCATCATCGAANAGATAG NNNNANACNTCATCC		
SAL1	C12-P2-G10	GTATATAACTATCTATTCGATGATGAAGATACCCCACCA AACCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTT GTACAAAAAAGTTGGCAAAAACAAGTATCAATCTACTG AGTGAGGTGAAGAGCAAACGAAGAACCAGTTGGAGAAA TGGAAGCAGTGGGGGTGTGCAGATATATGGCAGTGCAG GCTTCACCCTGCCAATTGGATCGTGTGAGACGTAATACC ACTTTCGCCATGGCTACCCAAAAGGAAGTTGGAGGCACC CGCCAGGTTAAGGTTTTTCCACTGGGTGATAAATCTGCT GTTTTACGACCNNNAAATGAATTTAATGGGTCATCAATA AAGTTGCTCTCAAGGGTGGAACAGCTGAAATTGNTATCC AAAGCAGAAAAGGCAGGCCTATTATCTGCTGCTGAAAA GTCTGGTCTCTCTCTTTCTAAAATTGAAAGCCTGGGTCTT CTTTCANAAGCAGAGGAGCTAGGTATACTTTCATCAGCC ACTGATCCTAACACACCTGGAGCGCTTCTAAATCTAGCA ATAGCTCTTCTGATTGCAGGGCCATTATGTGTCTATTTTG TTCCCGATGATTCAAGTTGGGAAGTAGCACTGCAAGTTG TAATAGCTCTACTGTCTGTTGTCGGTGGGCCTGCAGCATT TGCCGGATCGAATTTGGTGTCAAAGTTGCAGAAATCAAC CTAAGANTCCAATTNNNCATTTATTTATAGTCATCATTCN TTCTGTTTGGCTTTACTATATTTAATATTCCGAGGNAANA GCAAANTATATTTAATANTNCGAGGTAGAGANGTATGG NGTGAANNTTTTNATNGNTCATCCNATAANCTGATTTNA TATTTTTNANTANTTTNNTTTTNTC	882	unknown

SAL1	C12-P3-C6	GTATATAACTATCTATTCGATGATGAAGATACCCCACCA AACCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTT GTACAAAAAAGTTGGTGAAAATCCAGAAGATTCAGCCG TGCCCAAGAAAGACGTTGTCAATGAAAAACTGTCATCCG TGACTCTGGATCAAGAGCAGGGTGTTGTGGATATTGAGA ATGAAAAGTTATTAGATTTAGCCTTGCCTAAGGAAGAAG ATGTTGTCAATAAAAGACTGTCATCGGTAGCTCTGGAGC AAGAGCAGACTATTGAAGATATGGAGAATGTAAATGCC TCCGAGAAATCTGCCATAGTGGGCGAAGACAATTCAATT ATTACTGCACCTGAGGGTGAGAATGATCAGGAAAAAAT TGTTGAGATATGTACAGACCCTGTTTCTGATAGGAATGA AGCTGAAAGAGACATAATCCATGCTTTAACAGAGGAGA CAGAGGAGTGTCATGACAATGACGAGATGGAGTTNNCT GTGGAGGTTCCTTCATTGACGATATCTAATGTCATAGAA GAAAACAATTTGGTGAGAATGGAGGAGACTATACCTTC AAATGAGAATGCGGATGGGAAAGAACCTCCAGCAGCTG CTGAAACCCAGAGCATAGGNNNCCGGTGCAAACTGNTA ACTCTCTANAANCTGCTCTTAGATTCCAAAATGAGGACG ANCTTGNTGCTNAANAGGAANTATTGNTTCCANCNCTCG AACAGCCNGTTGAAGGNAAANAATCGACNCTCANANGT GGTGAAATTTTNCAGGAACANAGTTGANCATNGNNANN TGGANANAACGCANCATTGACTCGANNNNCTTGNANNA NNNNCNAGNNNNAACTTTNNANANNATGATCNNACNNN ATTT	888	unknown

SAL1	C12-P3-G7	ATATAACTATCTATTCGATGATGAAGATACCCCACCAAA CCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGT ACAAAAAAGTTGGCCAACTTTTTTGTACAAAGTTGTCCC CCTGAGTATTAAAATGTCTCAAGCGCCGGAGAACCAGG ACCGCGGTCTGTTCGGTCTGTTTGGCAAGAAAAATGAAG AGACGGAGGGGACGCAAAATGATCAGAAGATGCATCCT CCTACCACCCAGACCCACAATCAGCCTCAGGGTCATGCT CAAGGTGCTGGTTATTATCCCACTGCTGCTCAGCATGGA GGAGAACAATACGAGGGTCACGGTCAGCAGGGGCAAGT TACCTCTGAGGAAGCTGAGAAAAAAAAAACGCACCGGTT TGATGGGAAAACTTCACCCCGCACACGGCTCCGGCTCCG GTTCCGGCTCCAGCTCTTCGAGTGATGAAGAGGACGAAG GAAAGAAGAAAGAAGGGGGGAGAAAGAAGAAAGGTTC AAAGGAAAAACGTCAGGGGGGACGCGACTCTTCAGATC AGTGCGGCCGTGAAGGTGAATATGGGGATCAAGGTGTG AAGAAGGAGGGGATGATGGATAAAATCAAAGACAAGCT CCCTGGACACCGTAATGCTAATGAATAACTGGAAGAAA AGGAGAAAACAGAAGCCAATCTGTGAGGCTTCTGTCCA AACGTCCGGCTGTGTTTAATGNTTTTGGGGCATCGACTC NNNAGTCCTAATAATAANACGTGTTCGTCACTTTACGTN GNTNAATNATATGTTTGCCNGAGGATAGACCATAATNNN TCNNTGNATCCNATCTGCATNNCNNNNGCAGCGNNTTAN TAGTTNGATCNGAGNANANTTNTAGCNNNNGNGTTNNN NNNGGGGNTTTTAC	898	dehydrin 2
SAL1	C12-P3-F4	TATAACTATCTATTCGATGATGAAGATACCCCACCAAAC CCAAAAAAAGAGGGNGGGTCGAATCAAACAAGTTTGTA CAAAAAAGTTGGCCCTCTGTGATTCTGGTTTGTGAGCGC TAGCAGTCATGCCTTCCCTCAACATCTCAACAAACGNAC CCTTGGAGGGANNGAACACCTCCGANNNACTTTNAGAG	193	unknown
SAL5	I20-P1-E9	ATATAACTATCTATTCGATGANGAAGATACCCCACCAAA CCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGT ACAAAAAAGTTGGATCGAACAGACCAGGTGTCCAGAAC CTCAATGGCGCATATACGAAGGGCCAGATGTACGGTGCC	424	unknown

		GCCCTCCCCACATTCCGCCCGACGCCCCCCAAGAGACGC		
		AGAAATCGCTGCTGCCTCTGCTTCCTCTGCCTCGTCGCGT		
		TCCTCCTCGTCCTGATTTTGCTGGCGGGAATCGCTGCGCT		
		GGTTATATGGGTCATCTACAGGNCTCNNCANCCCAGTTT		
		CACACTGAATTCAGTGCAGATCCCCAAGTTCAATGTCAC		
		CNTNNATTCNCATCTCANCTACNANNTCNANNTGCAAAT		
		GGATGCCNNNAATCCCNNCNAGAANGNNANCTTT		
SAL5	I20-P1-G6	GTNTATAACTATCTATTCGATGATGAAGATACCCCACCA	852	unknown
		AACCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTT		
		GTACAAAAAAGTTGGAGATTTTTCGCTGTTTACCCACTT		
		GTAGGTCTGAATTTCTGCATTCACTCTGTTGTGGGTCTGT		
		TCAGGGTTTTGGATTCTTCTCTTTAGAACTAACGCAATAA		
		ACTGTAAGGGTGTAAGCGAATTGAAGACTAATTGTAGA		
		AGGGAGGAAGGGAACACATCAAAAGGGTGATAAATTTT		
		GTCACTTTCAATGGCCAGTGCAGTGGCAGGACAATGTGA		
		TTCAACCCTAATAAGCAGGAGAGGGGGACTGCTCTTATC		
		TTCTTCAAGCTCCACTTACAACAATGGAGGCATGAAACT		
		TGATTTGCGGGTTCCTCTGCCAATGCAAGGTTCTGCTATG		
		GTGAGAGCGCCTCTACTAATTCTGGCCATGGCACCCAAA		
		AAGAAGGTGAATAAATACGATGACAATTGGAAGAAACA		
		GTGGTTCGGGGCCGGGATCTTTCTCGAAGGTGATGAAGA		
		TGTGGATGTGGATATTGTCAAAAAGTTGGANANNNNGA		
		AGGTTCTAAGTGGAGTGGAGAANGCTGGATTGCTTTCNA		
		AGGCTGATGAATTANGCCTTTCTCTCTCATCTATTGAAA		
		AAATGGGCCTCCTCTCAAAAGCANAANANTTGGGCCTGC		
		TAAGCCTTGCANAGAAAGTCGCTTCCATATCACCTGCGG		
		CAATGGCATCTGTGTCANTGCCATTAGTTGNGGCCGCTA		
		TTGCCACTANTGNACTCATTCCANANGANNCCNCTGGAC		
		TGNNNNNNNNTCNGAANTTTCTGNNAACCATTTTT		

SAL5	I20-P2-H2	TATAACTATCTATTCNATGATGAAGATACCCCACCAAAC CCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGTA CAAAAAAGTTGGTTCAGAAGGCATAGCAGAGGATCGAG GTTTTGCAGTACGAGTCTCCNAGCCAAAATCNNTGGCGG CCGTCACTCCTGTATTGATCANTGGTGCTTCTACTAGTAA ATCTTTTGAATATGGTGTGGTAAAACTCACTCCATCAAG ATACTCCTTGTCTAACTTAGTGTGGCAGAACAACCGTAA GCCATATGGACACANGNNNTGCAAAAGGCACATTTCNN GTGCAGAATACGACAGTGGCAGAGGGAGAGGAAGTAAT GGAGGTGATTTTCTTGCTGGGTTCTTTTTAGGAGGAGCT GTGTTTGGAGCTCTCGGCTACTTGTTTGCACCACAGATC AGCAGAGCTTTGTGGACTGGATATGAAGATGGCCTGTGG AAGAAGTTGCCCAAACGTATGGACGATGATGCAAGCAT GGAGAAGACCAGGAAGACTTTGAATGAAAAAATAGCTC AACTAAATGCTGCAATTGATGAGGTTTCTTCCCAACTGA GAGCAGAGGANGATGCCAGTGAACCAGCAGTCACTGCT TCTGAAGGAGAACCTGCTACATAAATACCATTCAAGATT CTGACTGTCTGGGAGATGGTGTAATGTTAGCCTATGGTC CAGTCATGCANACCANAGGACTCCTAAATTGGGAATAC GCTATGTTATTGGNATTANTGTGCCATTTATGTGCTTCTT GAAATGTATGGTAAGTTNNAANTTNGAACTATGCAANTG TTTGNTNNNTCNCNTGCAACNAGCCNTTTTCCNTCACAN GCTTAAATANNCTATTTTT	871	unknown
SAL5	I20-P2-A6	ATATAACTATCTATTCGATGATGAAGATACCCCACCAAA CCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGT ACAAAAAAGTTGGATGTTTANTAGGCTAGTTTCCCAACT CGTTGATCACGGTTTGTTCCATATATTTTCCGCCGTGTAG GAATGGTCATGGCTTCTGGTTGCATTCTTCCAATTGCATT TCCATCCTCCAACACGATTAAAGGGCCACATTCATGGCT ACCAATTTACAAAAACTTTTCGAAGGGAAGAATATCAGA GAGGCATCGGCGCTTGAAAATGGTGGTCCTTGCAGAGA GCAGTGGAGGTGGCTGCTGTGGGGGCAGCAGCAGTAGT AGCAGCAGTGGAGGAAGCTGCANCAGCCATGGAAAATC	793	unknown

		TTCAGTTCCTGATCTTTCCAAAATTGGAAAAGAGTTTGA AACCATGGTTGCCAGAGCCACTTTGAGTGAACGTGAGAA GGAATATAATACTGTGGAGATGAAGGGGAACATTACTC GGGATGACTTTAAAGAAGTTATGAACATTGNGCCTTCAA GATTTGCTGAAGAAGGAAAAGGGGAAACAGTCATTGAT CTACAAGCGATGCTAAATGAATTAAAAAATGACAACTTT GCATTTGACAATCCCGAGGATGTGTTTATCTAAGTGACA ACTATAGTAANGGGGAAGTGAGTTCACTGAGATCACTAC AAGAGATTGTTCAATCCATACNTGTAAAAGANTTGATTT TGANACTGCNGTTTTCNCCCTTTANCAAGNAGTTTCTCA ACTCNATGTGAAGATAC		
SAL5	I20-P2-A11	ATATAACTATCTATTCGATGANGAAGATACCCCACCAAA CCCAAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGT ACAAAAAAGTTGGCCAACTTTTTTGTACAAAGTTGTCCC CCTATAGTCTTCACTGGCTATCTCAAATTCCCTCCGAAGT TTTAGAAAAGAATTCAGTGACATGGAACAAAGGCAAAA TCTCCGCGGGAGGATCTCCCCTTGTTGGAGAAGCTTATT TCCGTCNNTATCNNANNGATTTCAACGGGTTTCTGANNG CCCGGGCACNGGANCTGG	290	unknown
SAL5	I20-P2-B6	ATATAACTATCTATTCGATGATGAAGATACCCCACCAAA CCCAAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGT ACAAAAAAGTTGGGGGGTAATCGCGACTACAATGGCGGC ACAGCAAGCTCGATCCCTTGCCCAGATTCTCCGCCTTTC ATCATCACATACACAATGTGTATCCTCCCGGGCTTCTCA GCTGCAGCCAAGTCGCAGATTTTCAGCAGAACATCATGG CCCAGCAAAGGTTAACTTTTGGGAAGATCCAATGAGCCC TTCGAAATGGAAAGATGAGCATTTTGTACTCTGGTCGCT CTCTGGCTGGGGTGTACTCATTTATAGTGGCTATAAATTT TTCACCGGTGGGAAAAAGGATGCAACTGCTGAGGTTGGT GCATAATCTTAATTTAGTTCTAGACATGCACACTGAAAA AGCTTGTGGTTAATTGTAAACTAAGATACATATCTTTCTT TCATTTGGCCATTCTTTAATTTCGTTTTCATAAATGCAAT TTAATCATTGTGATATTTAATAAATATTCATTCTTGCTAC	729	unknown

		ATGATCATCGTGTGTTCTTGCAACATCTTTAGTGGGATCC TTCGTATTTTTCCCTATTCTATTTTAAATTAGTAATTACTG AAGTTCTATTGTCAGCATAAATGGTTGTGTGATAAACAT GNNTAGTTCTGTACAACCTTATGCATACGAACAACATTT TTAGTTCTTCTATAGATATTGG		
SAL5	I20-P2-C3	ATATAACTATCTATTCGATGANGAAGANACCCCACCAAA CCCAAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGT ACAAAAAAGTTGGAATTCNNGGGCGTTATTTCAGGTTTT GTGTTTCCTTCCNAACAAANTANNNGTCNTGNAAATGAG AAAGGCNNCNCNACTTTNGTGTTTATATACCGATATTTG TCTTTCATATATTTGANNAACAGGCATCTTCAATCATGTG TACTTATGGATATTGTGTTTCGGTTTGTCAAGTTTTTTCA ATCAATAAACTGCAATGATNATGAAACATGGCTTTGATA ATCAAAATTTTCTAACCTANNGGTACAGGCGAGTTTCTT GCAGTATATGGTTGTANTATGGGCGAGTTTCTTGCANNA TANGGTTGTANTATGGGCGAGTTTCTTGNATANANCTCT ATGGCATGAATTTNACTCTANTCATTGNGCTACCTCATA ANTGGTGCNNTCANAGTAGTGNCATATAATGTATGANGC ACGA	512	unknown
SAL5	I20-P2-C8	GTATATAACTATCTATTCGATGATGAAGATACCCCACCA AACCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTT GTACAAAAAAGTTGGGAGATGGCATCATTACTTAAGAGT GGCTCCTTTACCTTCTAACCCGGGTCTTCCATGTTTTAAC ACTTGTGACCAGAGCCTTCTATGTGGTAGCAGCTGGTAC ATCATAATGTCTTCCTCGCCCTTTTTGTAATGCCTTCTAG GGCAGTGTGTTTTTTATCTCACTTCCTCGCCCTTTTTGTA ATGCTGTAGGGCAGTGTTTTTATCCCACTTCTCTCGTGGC GTCGGCCATGGATTTAGTATTCCTGGAAATGGCTCCTCG TTGTCAAGTCTTGTGTTTCTTGAAGATGGGCTTTTTCATC TTGTGATCCGTTTTGTTATTCCGTTCAAGCGAGCGTGTTT GTTCTACCGGGTTTTCCCTTCAAGCGAGCGTGTTTGTTTT TATCATATCTTACTCCAGGCGAGCGTGAGAACAGTAAGC TGATATTGNAGGTACATTGCTATCAAGATCAGAAGAGAG	813	unknown

		CAGTAAGATCAGATTAGAGAGCAGTAACGCCAGAAATG CTTCTACTTCATTCATATTTCAACCNGNNGTCACTAATGG TAGCAGTAATTTTCATCACCCTTAGNNTTAATCGTCACTG GCATTCTGTCNTCTCTAGCGCCTTTCTGGGCATTGNCCCN ATCCTCCATATTTCTATTGATTTGTCCNTGGNNATCGGGA GGTCTCTTTANNNCGTGGCTTGNTTTTTTNCTCTACAGNA NNNGNCTTNNTCTNCNTNNCTAC		
SAL5	I20-P2-D1	GTATATAACTATCTATTCGATGATGAAGATACCCCACCA AACCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTT GTACAAAAAAGTTGGGTTTCGAGAGAAAATTTGGGAATT TGATGACGGAATTTTGTCCGTCATTTGACGTGAAGTATTT CTGTCGCGCAAAGGGTTCGTATATGGCAAATGTTGAAGC GGACGCTGCGGGTTTCGATTCCGAGACCTCCAGACTGGC GATGGATAACAGCTCGGTTCAGAAGCCCAAGCCTCTGGT GAAAATGAGCGTCAATATTTCCGGTCCTGACGATGGCGG ATTCACTGTCAATAGACAGGGGGAGATTTCTGTCAAGAA GGCCCGTGCTGTGCACATCCAGGTAATGAGAATTCAAGA AGAAGATGAGCATCTGGGCGAGGATTTAAGGGAGGGCG TGAATCCGAAAGACAGATTTGTGTTTTTCCCAATCGCGT CNCANATGAAGGACATGTTCTTTGACTATTCNAGGCCCA CGNTTCCGTCGCCGCTCGGCATGANCGCTGNAGTTCNCT CNCTCTGANAAGATTCCAACATGCNNCCNAGTTNTGAGA AGGNNNTGTGTGTANATATNNNNCTNNGAGNNNTGNNG NANNNGANGNCNTNTGTNAATTA	645	unkown
SAL5	I20-P2-F4	TATAACTATCTATTCGATGATGAAGATACCCCACCAAAC CCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGTA CAAAAAAGTTGGGTANGGGTTCGANANNGANGGNAAGA GTCTCTCACAGTCTGTCCCCATATACACAGCGACCATGT CGAGCAAGCAAGGTGGAAAGGCTAAGCCCCTCAAGCAG CCAAAGAAGGATAAGGCCGAGTATGACGAGGCTGATCT AGCTCACATTCAGAAGAAGAAGGACGAGGAGAAGGCTT TGAAGGAGTTAAAAGCTAAGGCATCACAGAAGGGTAGT TTTGGTGGAACAGGGTTGAAGAAAAGTGGCAAAAAGTA	553	unknown

		GTTTGCCATTGACTGCTTTCTCTACATCATGCCACTAAAA TTATACTTATGGGGACTGGTTAGGAGAGTGTTTACTCAA TTACAAAGTATGTGATTACGGTTAAGAATGGACCTTCTG AATCACACACTTGCTGTTTTAATGTGAATGGATATTTAAT GTTGAAATTACAATGTGATATAGGTTTTAATTTTTTTTAC NAAAANNAAAA		
SAL5	I20-P3-C9	CCCACCAAACCCAAAAAAAGAGGGTGGGTCGAATCAAA CAAGTNNNGTACAAAAANNNNGGGGCGGATCCTCATTT GGGTATCAACCATNNCCAAATCTCTGCAGACCAAGTTCG GCTGAAATTTCTTTCTCCAATGGCAGCTACTGCGTGCAC ATGGGTACCCACAGGCTTCACAGCCCCAAGAAGGCACC GTAAGCCCATTACTGCGTCTCAATCTCGGGCTTCATTTAT AGGACTAAGGCTTGGAAACACACTGGATTCAAAGGCCC AAAACAGCTTTCAGAGCCAAACAGCCGTTTGCAGATCAT TCTCTCGCATCACCTGTGCGCTGAATCCTTCACTGGTGAT CAGTCTGAGCACAGGGGCTTCACTGTTCCTTGGAAGGTT CGTGTTCTTGTCATTTCAGAGGGACAATGTGGCGAAACA AGGCCTGCCTTCGCAGAATGNACAGACCCACTTCGAGGC AGGAGACACCAGAGCCACCGAGTACGTGAATCTTCTCA AGAGCAATGACCCAGCTGGGTTTAACATTGTTGATGTGC TTGCATGGGGTTCAANTTGGCCACATTTGTGGCTTACTTC ATCTTGGNNACTTCAAGCAACGGANACNNNCCTANTTTC TTTNNAANTCTCTTCTGNCNNNANAAAAATNNTTGTTNA NGNNTGGACTTTTNNAAAGGNCAANANACTGNATCTTTC TGTANTCTTATATATGTTTNNNATTGAATTTTAACCNNGT TAAATTTNCNANCT	754	KH domain-containing protein/zinc finger (CCCH-type) family protein
SAL5	I20-P3-D12	GTATATAACTATCTATTCGATGATGAAGATACCCCACCA AACCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTT GTACAAAAAAGTTGGATGATTTCCACATCTTTCTTCAGG TTATTGCATTGACCAATGCCCTCAACTGTCTTCCAAATTT CTCATAGTCGTGTACTATGACTTCCACATTTGCCTCCATT TCATTGCATTGGCCACCGATAGTCTTCCAATCTCTCCCGA GTCTTCTGGCTCTACGGCATGGGTTTCTTTTCTGATTCCT	964	unknown

		TCTAACAACGGCCTGTGTTCACAGATATATCAGAATGTC TTCCTCACCCTTTTTGTGATGAAAAGATTAAGTTGGAGG TCTTTGTGTTTTATCTCTGGCTTCCCTTTATGGTGACCAT GGTTTCCGAATCTTTCAACATGGCCTCAATACTAAAGAG AGGTTTCTTCATCTTCTACCCAGGGCCTTCAATGTGTTCG CAGATATGTCCAATTGTCCTCCTCACCCTTTCTGTAATGA AATGAATAAGTGGACATGTGGGCAGTATGATTCTATGTC TGGTTTCCTCTTCTGACGACCGTGGTATCAGTATCTCGAG ATGGCATCATTACTTTAGAGTGGCTCCTTTACCTTCTAAC CCGGGTCTTCTATGTTTTAACACTTGTGACCAGAGCCTTC TATGTGGTAGCAGCTGGTACATCATAATGTCTTCCTCGC CCTTTTTGTAATGNCTTCTAGGGCAGTGATTTTTTTATCT CACTTCCTCGCCCTTTTTGTAATGNTGNAGGGCAGNGNT TTTATCTCACTTCTCTCGTNNNTCNGCCATGNNTTTANTA NTCCTGGANATGGCTNCTCGTTGTCNANCCTTTNNNGAN ANGGGCTTTTTCATCTNGNGATCCGTTNNNNNNTNCGNN NNCGANCGNGTNAAGCNANCGNGNTNNTNTACCGGNNT NNCNNNANCNANCGTGTT		
SAL5	I20-P3-G3	ATATAACTATCTATTCGATGATGAAGATACCCCACCAAA CCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGT ACAAAAAAGTTGGCTTAAAAAAGGAACAGATGCAGTGA GGAAGCTTGTTAATGAAGGAGAGTCTGGCCACTTTACTC AGGGGTGTCCAACCACACTAGGAGGCAATCGAACTAGG GAGTTCATTGAGAAAATCCCTGTGAAGGATAAGCATCTA AAATCTCGTATTATTGGATCTGGTGGATCAGTTATTCNG AAGANTGNNAAAGANACANGNNGNANNATTANGNTNG NNNATAATG	316	unknown

SAL5	I20-P3-G10	GTATATAACTATCTATTCGATGATGAAGATACCCCACCA AACCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTT GTACAAAAAAGTTGGGTGTGCCTGATGGAACTAAGCTTG CCTACTATGTAAAAGGACAGCGATTACTAGATGGATATA AACTGGGATCTGGTATATGCTGCAGTTGCTGCGATACTG AGATTAGCTGTTCTCAGTTTGAAGCACATGCTGGACGAT CCTCAAGGCGGAATCCTTACAATAGTATCTATCTTCCGG ATGGGCAGTCCCTGCATGAAGTGGCACTTTCCCTGACAA GCCAAAGAAGTTTGAAGGCAAAATCATGTGATGAGAAT GAAGATATTTGTACAGAATGTGGAGATGGAGGTGATCTG CTTCTTTGTGATGGCTGTCCAAGGGCCTTCCACACAGATT GTGCTGGAGAGCAACGTATTCCGGTGGGTGATTGGTATT GCTTAAATTGNCAGCATCATTCGAGAACAAGAAGAAAG NTGTCTGCTAGAAAANAGCCAAAACTTTTTGGAAAGGCA GCATCACTTGGGTATCAGGAAAANCCTTCCAATCGCTGT ACACGTGNTGNCAATGNCCCANAGAAAACAGNTGGTGG ANGNGTACTATGCAGNNTTCATGANTTTGATAAATGGGC ATTTGGGGATCGCACTGNTATGCTCTGNGACCAATNNGA NAAGNANTTCCACGTTGGCTGNTTAANANACCGGGGCA TGACNNANTTAAANANTGCCNNAGGGTNANTGGNTCT	774	unknown
SAL5	I20-P3-H1	GTATATAACTATCTATTCGATGATGAAGATACCCCACCA AACCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTT GTACAAAAAAGTTGGAATATGCCAACGCTTTCACTGCGG GCATCCGCTTCATAACTCCATGGCTCCATAACGGTTTGA ACCTGGGACTATCGAGATCCAGTTGGTGGGGGCAAAAA TGTTGAAAAACGGTGCTGTACAGGGCTGTCTGGACTATC TGTATGGAATGGGTAGAATTGGAAAGAAAGATAAGACT GCAAGGAACCGGTATGAGGGTTAACATCAACTTTTGAAA ATCATGGCGCCATGCAAAGCAGGAAATAAACAAGATTC ATTGACTGAGGAATCAAACCCAAATGCAAATATACTTTC AGATACTACTGCTCGATTGACAAACATCATCAAATGTTG GAACATTAATGGAACTACATTTACAATTTAAGGATCAGA AAAAACCAAGTACATCTACAAGCAAATAGAGGGCAAGT	810	unknown

		GAAGTGCCAGCAAGTAAAAGAAATCTCAACAAATCCAA AAATATCGTCATTGTATTTTTTAGTTTTGATTTATTCTATT TTGAGAACGAAGAATGTGGAATATCTAACTACTTTTGGT TAAGACATTTTTGTATAAATAAACAAATCAAAGATCTAT TCCTGGATTCAATCTCAACAAATCCNAAAATATCGTCAT TGTATTTTTTAGTTTTGATTTATTCTATTTTGANAACGAA NAATGTGGNATATCTAACTACTTTTGGTTAAGACATTTTT GTATAAATAAACAAATCANNATCTATTCCTGG		
SAL5	I20-P4-C1	GTATATAACTATCTATTCGATGANGAAGANACCCCACCA AACCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTT GTACAAAAAAGTTGGATAACTGTGGGGTTCAATTTAACA AGATCAATGATGCCTATGAGACTGTAATGTCCAGTTTGG AAAAGGCTAAACATCAAAACTGTTCTGCCGATTACCATG TGGAGGACCTTATGGAAGTCGGGGACGATTCATGGGAA GAATGGATGGGATGGGAAGGAGCTGGAACCCTTGATTA TTCCTCCCATATTAACATTTATGCCTGATAAGATCTTCAT CATAAACTCTGTATCCATGTCTTTTTTGCCTTGTAAACAG TTATTGCTTCTCCATGACCCATCCCCGTTTATGGGCTACC TATGAAACTCTTGCTTTTGACACTATAACCCTTAGAAAG CCCATAGCGCTGCTTCATTCATCTCTTGTACATACGATAT ATTCTTCTATTTGTATAGCTAATTTTATCTCTTACATTAA CTCCCAGTGCACAAAGGACTTCTCGACCCAGTTGATGCA ACAGCTTGTTGAGATCTTATTAAGTCATTTGCTTTGCTCC TGAGTTTTGATTATTAGATTACTTGTACATATGCTCTGTT TANTTGCATATCAACTACCAGTACACTTTTTTGAGATCTT TTTCTTTCGGGACAGTGTGAANATNATGGGGANTTGCTC TGAAAATGAANAGCTCCTAACCATATTTGTTGNAATTAC AGTGNTACTCAATTTATGGAAACCATCATTTTGANTAC	784	unknown

SAL5	I20-P4-G12	GTATATAACTATCTATTCGATGATGAAGATACCCCACCA AACCCAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTT GTACAAAAAAGTTGGGCGATATAAACAAAAGGGCAAAG ACTGAGCGCATCGATCAATCCAGCAAGGAAGCCATGAA CGGCCTAGCATCAGACGGGCCCCGGGTCCAGGGATCGA ATCGCAGAGTGAAGCTCGATGTCGGAGGCAAGATCTTCG AAACCACGACTTCGACCCTCCAATCCGCAGGGAAAACCT CCCTCCTCGCCCGTTCGGCTTTGTCGACAGACTCCGCTGA AATCTTCTTCGACAGAGACCCGCATCTGTTCGCGCTCCT ACTCGGCGTTCTCCGGACGGGAAAGCTCTCAGCGTCGAC ATGGGAAAAATTCGACATCGAAGCCCTTATAGACGAAG CCTCCTATTACGGAATACTGGAGCCCGTCAAGAAAGCCA TGGCTCCGGAAGCCCTAGACGGAATCGATGTCGAAAGA GTCTCAATGGTCGTTCCCAATGGCCGGGATTACCCTTTG GCCATCTGTTCTTCACACGACGGTTCCGTCTGGGTCGGC CATGGCAGCAAAATCACGCCATACGACTGGGCGCTCCG GAAGCAGACCACGACGTTGACGGATCTTCACAGTGTCGA CACCATGAACAGGATCTCAGAAACCCTAGCGGCCGTTGG CGCAGAGGANTTTCCGGGGTTGCACATATACGACACCAA GAACGCGGCGCATGTGAAGAGCCTGACTTGGTCGGACA AATCCGACACGCGCGTCTACAAACCCTGCGTTCGAGCCC TCGCCTCGTCGNATTCTTCAATCTTCGCGAGCTTCGAGA ACGGGCAGCGAACAGAGAANACGATCCTCGTTGTCGAC AAGGANAGGNTCGAGGTTTNTCGAGAGANCNGCCGGCN ANGCGGTAACTCTGCGCACTCNAANNTTCGACGANTTTN CAGNTNG	973	unknown

SAL5	I20-P4-H12	ATATAACTATCTATTCGATGATGAAGATACCCCACCAAA CCCAAAAAAAAGAGGGTGGGTCGAATCAAACAAGTTTGT ACAAAAAAGTTGGTTCAACAGAGGCATCGCANTTATGC GAGACTTGCCAACAAGTACGGGCCGGTGATGCATTTCTG CATTGAGAATGCAAATGTTATCGTGGTTGGAAGTCCAGA GGTTGCCTTGGAAGTCCTCAAAACGAAGGACGCCGAGT GGGCATCCAGGCCACCTACGCTTTCGGGGAAGTACATTG GGGTTGATTTCCACGCCCTTGATTTCGCACCCAATGGCC CTCACTGGCGCCACCTGCGGAAGATATNNNNNACCCAC ATATTCTCTCNNGNANGATTANNGNNGNAGTCTTATATC CGANNANNGNNNNNNCTCCNCNTTGTGGACNNNATCTT CNCCCNNCNCNNANANGNNNNNNNNNNNNNTTAANTTT NNNNNNCNGNGANNCCTNNNNNNTNNGAANCGTG	496	unknown

Appendix 7. Multivariate anlysis of variance (MANOVA) R script for transcript abundance.

```
> src_files = c('C12', 'H14', 'I04', 'I20', 'J13', 'K12', 'K15', 'N14')
> n_genes = length(src_files)
> data_list = vector(mode='list', length=n_genes)
> src = "/Users/amandagregoris/Documents/Phd/PgqPCR/Stats/"
> for(i in 1:n_genes)
+ {
+     x_filename = paste0(src, '/', src_files[i], '.csv')
+     x = read.csv(file=x_filename, fileEncoding='UTF-8-BOM', na.strings = ' ')[,1:5]
+
+     empty_rows = which(is.na(x$qty))
+     if( length(empty_rows) > 0 ) x = x[-empty_rows,]
+
+     names(x)[ which( names(x)=='qty' ) ] = as.character(x$gene[1])
+
+     data_list[[i]] = x
+ }
> data_full = do.call('cbind', data_list)
> View(data_full)
> data_full[,names(data_full)=='photo']
   photo photo.1 photo.2 photo.3 photo.4 photo.5 photo.6 photo.7
1   long    long    long    long    long    long    long    long
2   long    long    long    long    long    long    long    long
3   long    long    long    long    long    long    long    long
4   long    long    long    long    long    long    long    long
5   long    long    long    long    long    long    long    long
6   long    long    long    long    long    long    long    long
7   long    long    long    long    long    long    long    long
8   long    long    long    long    long    long    long    long
9   long    long    long    long    long    long    long    long
10  long    long    long    long    long    long    long    long
11  long    long    long    long    long    long    long    long
12  long    long    long    long    long    long    long    long
13  long    long    long    long    long    long    long    long
14  long    long    long    long    long    long    long    long
15  long    long    long    long    long    long    long    long
16  long    long    long    long    long    long    long    long
17  long    long    long    long    long    long    long    long
18  long    long    long    long    long    long    long    long
19  long    long    long    long    long    long    long    long
20  long    long    long    long    long    long    long    long
21  long    long    long    long    long    long    long   short
22 short   short   short   short   short   short   short   short
23 short   short   short   short   short   short   short   short
24 short   short   short   short   short   short   short   short
```

```
25 short   short   short   short   short   short   short   short
26 short   short   short   short   short   short   short   short
27 short   short   short   short   short   short   short   short
28 short   short   short   short   short   short   short   short
29 short   short   short   short   short   short   short   short
30 short   short   short   short   short   short   short   short
31 short   short   short   short   short   short   short   short
32 short   short   short   short   short   short   short   short
33 short   short   short   short   short   short   short   short
34 short   short   short   short   short   short   short   short
35 short   short   short   short   short   short   short   short
36 short   short   short   short   short   short   short   short
37 short   short   short   short   short   short   short   short
38 short   short   short   short   short   short   short   short
39 short   short   short   short   short   short   short   short
40 short   short   short   short   short   short   short   short
41 short   short   short   short   short   short   short   short
42 short   short   short   short   short   short   short   short
> data_full[,names(data_full)=='day']
         day      day.1      day.2      day.3      day.4      day.5      day.6
1       zero       zero       zero       zero       zero       zero       zero
2       zero       zero       zero       zero       zero       zero       zero
3       zero       zero       zero       zero       zero       zero       zero
4      seven      seven      seven      seven      seven      seven      seven
5      seven      seven      seven      seven      seven      seven      seven
6      seven      seven      seven      seven      seven      seven      seven
7      seven      seven      seven      seven      seven      seven      seven
8   fourteen   fourteen   fourteen   fourteen   fourteen   fourteen   fourteen
9   fourteen   fourteen   fourteen   fourteen   fourteen   fourteen   fourteen
10  fourteen   fourteen   fourteen   fourteen   fourteen   fourteen   fourteen
11  fourteen   fourteen   fourteen   fourteen   fourteen   fourteen   fourteen
12  fourteen   fourteen   fourteen   fourteen   fourteen   fourteen   fourteen
13 twnetyeight twnetyeight twnetyeight twnetyeight twnetyeight twentyeight twnetyeight
14 twnetyeight twnetyeight twnetyeight twnetyeight twnetyeight twentyeight twnetyeight
15 twnetyeight twnetyeight twnetyeight twnetyeight twnetyeight twentyeight twnetyeight
16 twnetyeight twnetyeight twnetyeight twnetyeight twnetyeight twentyeight twnetyeight
17 twnetyeight twnetyeight twnetyeight twnetyeight twnetyeight twentyeight twnetyeight
18   seventy   seventy   seventy   seventy   seventy   seventy   seventy
19   seventy   seventy   seventy   seventy   seventy   seventy   seventy
20   seventy   seventy   seventy   seventy   seventy   seventy   seventy
21   seventy   seventy   seventy   seventy   seventy   seventy   seventy
22      zero      zero      zero      zero      zero      zero      zero
23      zero      zero      zero      zero      zero      zero      zero
24      zero      zero      zero      zero      zero      zero      zero
25      zero      zero      zero      zero      zero      zero      zero
26     seven     seven     seven     seven     seven     seven     seven
```

27 seven seven seven seven seven seven seven
28 seven seven seven seven seven seven seven
29 seven seven seven seven seven seven seven
30 fourteen fourteen fourteen fourteen fourteen fourteen fourteen
31 fourteen fourteen fourteen fourteen fourteen fourteen fourteen
32 fourteen fourteen fourteen fourteen fourteen fourteen fourteen
33 fourteen fourteen fourteen fourteen fourteen fourteen fourteen
34 fourteen fourteen fourteen fourteen fourteen fourteen fourteen
35 twnetyeight twnetyeight twnetyeight twnetyeight twnetyeight twentyeight twnetyeight
36 twnetyeight twnetyeight twnetyeight twnetyeight twnetyeight twentyeight twnetyeight
37 twnetyeight twnetyeight twnetyeight twnetyeight twnetyeight twentyeight twnetyeight
38 twnetyeight twnetyeight twnetyeight twnetyeight twnetyeight twentyeight twnetyeight
39 seventy seventy seventy seventy seventy seventy seventy
40 seventy seventy seventy seventy seventy seventy seventy
41 seventy seventy seventy seventy seventy seventy seventy
42 seventy seventy seventy seventy seventy seventy seventy

day.7

1 zero
2 zero
3 zero
4 seven
5 seven
6 seven
7 seven
8 fourteen
9 fourteen
10 fourteen
11 fourteen
12 fourteen
13 twnetyeight
14 twnetyeight
15 twnetyeight
16 twnetyeight
17 twnetyeight
18 seventy
19 seventy
20 seventy
21 seventy
22 zero
23 zero
24 zero
25 zero
26 seven
27 seven
28 seven
29 seven

```
30    fourteen
31    fourteen
32    fourteen
33    fourteen
34    fourteen
35 twnetyeight
36 twnetyeight
37 twnetyeight
38 twnetyeight
39     seventy
40     seventy
41     seventy
42     seventy
> View(data_full)
> data_full = data_full[, c(5,10,15,20,25,30,35,40,1,2,4)]
> levels(data_full$photo)
[1] "long"  "short"
> View(data_full)
> View(data_full)
> gene_idx = 1:8
> gene_names = names(data_full)[gene_idx]
> data_full_trans = data_full
> data_full_trans[, gene_idx] = log(data_full_trans[, gene_idx])
> data_full_trans[, gene_idx] = scale(data_full_trans[, gene_idx])
> y_range = range(data_full_trans[, gene_idx])
> par(mfrow=c(2,3))
> par(mfrow=c(2,4))
> for(i in gene_idx) hist(data_full_trans[, i], breaks=10, xlim=y_range, main=gene_names[i])
> par(mfrow=c(1,1))
> res_form = as.formula(paste(gene_names[i], '~ day'))
> res_day = lm(res_form, data=data_full_trans)$residuals
> i=1
> res_form = as.formula(paste(gene_names[i], '~ day'))
> res_day = lm(res_form, data=data_full_trans)$residuals
> print(paste('testing day effect residuals in ', gene_names[i], '...'))
[1] "testing day effect residuals in  C12 ..."
> res_form = as.formula(paste(gene_names[i], '~ photo'))
> i=1
> res_form = as.formula(paste(gene_names[i], '~ photo'))
> res_photo = lm(res_form, data=data_full_trans)$residuals
> i=7
> y_form = as.formula(paste('cbind(', paste(gene_names, collapse=', '), ')', '~ photo/day'))
> for(i in gene_idx)
+ {
+     res_form = as.formula(paste(gene_names[i], '~ photo'))
+     res_photo = lm(res_form, data=data_full_trans)$residuals
```

```
+     print(paste('testing photoperiod residuals in ', gene_names[i], '...'))
+     print(levene.test(res_photo~data_full_trans$photo))
+ }
[1] "testing photoperiod residuals in  C12 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1   1.434 0.2382
      40
[1] "testing photoperiod residuals in  h14 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  1  4.5826 0.03845 *
      40
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
[1] "testing photoperiod residuals in  i04 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.2384  0.628
      40
[1] "testing photoperiod residuals in  i20 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  1.9492 0.1704
      40
[1] "testing photoperiod residuals in  j13 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  1  3.5554 0.06663 .
      40
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
[1] "testing photoperiod residuals in  k12 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  1.0441  0.313
      40
[1] "testing photoperiod residuals in  k15 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1   0.676 0.4158
      40
[1] "testing photoperiod residuals in  n14 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.4323 0.5146
```

```
     40
> shapiro.test(residuals(yman)[,"C12"])

        Shapiro-Wilk normality test

data:  residuals(yman)[, "C12"]
W = 0.9675, p-value = 0.2715

> shapiro.test(residuals(yman)[,"h14"])

        Shapiro-Wilk normality test

data:  residuals(yman)[, "h14"]
W = 0.98416, p-value = 0.8184

> shapiro.test(residuals(yman)[,"i04"])

        Shapiro-Wilk normality test

data:  residuals(yman)[, "i04"]
W = 0.97613, p-value = 0.5171

> shapiro.test(residuals(yman)[,"i20"])

        Shapiro-Wilk normality test

data:  residuals(yman)[, "i20"]
W = 0.96652, p-value = 0.2512

> shapiro.test(residuals(yman)[,"j13"])

        Shapiro-Wilk normality test

data:  residuals(yman)[, "j13"]
W = 0.98427, p-value = 0.822

> shapiro.test(residuals(yman)[,"k12"])

        Shapiro-Wilk normality test

data:  residuals(yman)[, "k12"]
W = 0.96885, p-value = 0.302

> shapiro.test(residuals(yman)[,"k15"])

        Shapiro-Wilk normality test
```

```
data:  residuals(yman)[, "k15"]
W = 0.9543, p-value = 0.09213

> shapiro.test(residuals(yman)[,"n14"])

        Shapiro-Wilk normality test

data:  residuals(yman)[, "n14"]
W = 0.96382, p-value = 0.202

> for(i in gene_idx)
+ {
+     res_form = as.formula(paste(gene_names[i], '~ day'))
+     res_day = lm(res_form, data=data_full_trans)$residuals
+     print(paste('testing day effect residuals in ', gene_names[i], '...'))
+     print(leveneTest(res_day~data_full_trans$day))
+ }
[1] "testing day effect residuals in  C12 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  1.4218  0.246
      37
[1] "testing day effect residuals in  h14 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  1.0897 0.3758
      37
[1] "testing day effect residuals in  i04 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  4  2.1731 0.09114 .
      37
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
[1] "testing day effect residuals in  i20 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  0.5803 0.6788
      37
[1] "testing day effect residuals in  j13 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  1.3585  0.267
      37
[1] "testing day effect residuals in  k12 ..."
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  4   2.343 0.07272 .
      37
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
[1] "testing day effect residuals in  k15 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  0.3857 0.8174
      37
[1] "testing day effect residuals in  n14 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  0.2769  0.891
      37

> for(i in gene_idx)
+ {
+     res_form = as.formula(paste(gene_names[i], '~ photo'))
+     res_photo = lm(res_form, data=data_full_trans)$residuals
+     print(paste('testing photoperiod residuals in ', gene_names[i], '...'))
+     print(leveneTest(res_photo~data_full_trans$photo))
+ }
[1] "testing photoperiod residuals in  C12 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1   1.434 0.2382
      40
[1] "testing photoperiod residuals in  h14 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  1  4.5826 0.03845 *
      40
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
[1] "testing photoperiod residuals in  i04 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.2384  0.628
      40
[1] "testing photoperiod residuals in  i20 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  1.9492 0.1704
      40
```

```
[1] "testing photoperiod residuals in  j13 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  1  3.5554 0.06663 .
      40
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
[1] "testing photoperiod residuals in  k12 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  1.0441  0.313
      40
[1] "testing photoperiod residuals in  k15 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1   0.676 0.4158
      40
[1] "testing photoperiod residuals in  n14 ..."
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.4323 0.5146
      40
> yman = manova(y_form, data=data_full_trans)
> summary(yman)
          Df  Pillai approx F num Df den Df   Pr(>F)
photo      1 0.58492   4.4037      8     25 0.002002 **
photo:day  8 2.95783   2.3465     64    256 1.36e-06 ***
Residuals 32
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary.aov(yman)
 Response C12 :
                Df  Sum Sq Mean Sq F value    Pr(>F)
photo            1  0.4147  0.4147  1.4282    0.2408
photo:day        8 31.2947  3.9118 13.4737 2.663e-08 ***
Residuals       32  9.2906  0.2903
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response h14 :
            Df  Sum Sq Mean Sq F value    Pr(>F)
photo        1  1.4169 1.41690  2.4375 0.1283011
photo:day    8 20.9818 2.62272  4.5119 0.0009572 ***
Residuals   32 18.6013 0.58129
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
 Response i04 :
          Df  Sum Sq Mean Sq F value    Pr(>F)
photo      1  0.2006  0.2006  0.5537    0.4623
photo:day  8 29.2026  3.6503 10.0727 6.923e-07 ***
Residuals 32 11.5968  0.3624
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response i20 :
          Df  Sum Sq Mean Sq F value    Pr(>F)
photo      1  0.7028 0.70283  1.4570    0.2363
photo:day  8 24.8611 3.10764  6.4424 5.463e-05 ***
Residuals 32 15.4360 0.48238
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response j13 :
          Df  Sum Sq Mean Sq F value    Pr(>F)
photo      1  0.4283  0.4283  1.4752    0.2334
photo:day  8 31.2801  3.9100 13.4660 2.681e-08 ***
Residuals 32  9.2916  0.2904
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response k12 :
          Df  Sum Sq Mean Sq F value    Pr(>F)
photo      1  0.2968 0.29684  0.5712 0.4552907
photo:day  8 24.0748 3.00935  5.7913 0.0001366 ***
Residuals 32 16.6284 0.51964
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response k15 :
          Df  Sum Sq Mean Sq F value Pr(>F)
photo      1  1.3111 1.31112  1.4943 0.2305
photo:day  8 11.6116 1.45145  1.6542 0.1486
Residuals 32 28.0773 0.87742

 Response n14 :
          Df  Sum Sq Mean Sq F value    Pr(>F)
photo      1  3.0755  3.0755  13.205 0.0009676 ***
photo:day  8 30.4714  3.8089  16.354 2.598e-09 ***
Residuals 32  7.4530  0.2329
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> lsm_man <- lsmeans(yman, "day")
> pairwisecomp <- test(contrast(lsm_man, "pairwise"), side="=",  adjust="fdr")
> View(pairwisecomp)
> View(pairwisecomp)
> residuals(yman)
          C12         h14         i04         i20         j13         k12         k15
1  -0.540451275 -0.89944003 -0.227676273 -0.01251174  0.581768766 -0.09500930 -
1.86514568
2   0.703584475  0.46468195  0.764254381  0.41342674 -0.452208136 -0.30685465
1.33284921
3  -0.163133200  0.43475808 -0.536578108 -0.40091501 -0.129560630  0.40186395
0.53229646
4   0.123553870 -1.01953283  0.545061485  0.28445134 -0.141870309  0.38114517
1.33922535
5   0.031773912  0.19234838  0.252614259  0.11441171 -0.168205469  0.31936440 -
0.20973403
6  -0.304798613  0.59219662 -0.400666301 -0.79557232  0.123707452  0.59462219 -
0.26224379
7   0.149470832  0.23498783 -0.397009443  0.39670926  0.186368326 -1.29513177 -
0.86724753
8  -0.574544484  0.35546844 -0.170610895 -0.61368781  0.094943540  0.31230401
0.52101669
9   0.129198601 -0.35790055 -0.116016368 -0.31261878  0.254612321 -0.65101287
0.68743805
10 -0.330583864 -0.65454834  0.759080931  0.31753535 -0.751952841  0.40639807
0.68522411
11  0.990658880  0.71231047 -1.541958326  0.02985017  0.659533741  0.06259301 -
2.31923607
12 -0.214729134 -0.05533003  1.069504659  0.57892107 -0.257136761 -0.13028222
0.42555721
13 -0.599186134 -0.76923625 -0.233482516 -0.12768112 -0.235263814 -0.32519368 -
1.22549627
14 -0.312242689  0.54009667 -0.145568924 -0.05667938 -0.034747888 -0.72072874 -
0.49716431
15  0.600926385 -0.66179051  0.545531729 -0.37322817 -0.169846768  1.15037062
0.91634084
16  0.106786680 -0.23685199  0.516993369  0.24221875  0.362675491 -0.34095327
1.11943860
17  0.203715757  1.12778209 -0.683473658  0.31536993  0.077182980  0.23650507 -
0.31311887
18 -0.223958308  0.53939685  0.103564357 -0.65758628  0.504865386 -0.49922232 -
0.53957139
19  0.032709592 -0.21572820 -0.146849779 -0.53141828  0.348512420 -0.06174115
0.32108123
20  0.119571031 -0.45981507  0.112856944  0.25842712  0.101668291 -0.33572703
0.30372676
```

21 0.071677684 0.13614643 -0.069571522 0.93057744 -0.955046097 0.89669049 -0.08523659
22 0.782868118 0.13081058 1.036268204 0.92713105 -0.632753103 -0.28197864 0.65692049
23 0.391726902 -0.65586328 0.161083279 1.10799942 0.016389681 -1.47399275 0.30987190
24 -1.491428517 0.20727260 -0.639064128 -0.10203430 -0.624440166 1.23159217 -0.86216848
25 0.316833496 0.31778010 -0.558287355 -1.93309617 1.240803588 0.52437922 -0.10462391
26 -0.287329307 -0.37482542 -0.230336800 0.07417839 0.116615175 -0.39676445 -0.38099496
27 0.099876774 1.06651464 -0.560132841 0.12544344 0.076717028 0.56348887 1.19582391
28 0.190913228 -0.50719672 0.264135149 0.42112365 -0.062484253 -0.22764517 -0.46889172
29 -0.003460695 -0.18449250 0.526334491 -0.62074548 -0.130847950 0.06092075 -0.34593723
30 -0.970996123 0.02138205 -0.712197781 -1.06886291 0.990631025 -0.13938412 -0.72661865
31 -0.081341402 0.39948029 0.054571016 -0.14163958 0.007857128 -0.10097971 -0.06361785
32 0.519083586 -0.45004544 0.732102650 0.54360821 -0.373545524 0.23538286 0.14162048
33 0.502678287 -0.39744937 0.041799640 -0.05872827 -0.359767697 -0.09915016 0.50835030
34 0.030575652 0.42663247 -0.116275525 0.72562254 -0.265174932 0.10413112 0.14026571
35 -0.270760950 1.40564260 0.600390726 0.23324110 -0.225091254 -0.34309884 0.73850113
36 0.085339529 -1.68936405 0.117715125 0.35662660 0.489671368 0.18967137 0.42349664
37 -0.296933335 1.08382753 -0.821430461 -1.17081741 0.199028201 0.91434125 0.53274338
38 0.482354756 -0.80010608 0.103324610 0.58094970 -0.463608315 -0.76091378 -1.69474115
39 -0.215706673 -0.19733685 -0.114332234 -0.17859709 0.792463411 0.35109206 -0.67565710
40 0.603774185 1.13846746 -0.003559054 -0.46182337 -0.979742795 1.08206234 0.29250045
41 -0.606445089 -0.79405701 0.162211044 0.86682645 -0.160822844 -1.58767141 0.46568568
42 0.218377577 -0.14707360 -0.044319756 -0.22640598 0.348102228 0.15451701 -0.08252903

	n14
1	-0.80476894

2 0.76194327
3 0.04282567
4 0.43937612
5 0.13944352
6 -0.26999058
7 -0.30882905
8 0.21176816
9 0.09778827
10 0.26341140
11 -1.11971124
12 0.54674340
13 0.05264853
14 -0.28193043
15 0.81183507
16 -0.24379706
17 -0.33875611
18 -1.00174151
19 0.03030705
20 0.50282203
21 0.46861242
22 0.38043449
23 -0.09695937
24 -0.67550934
25 0.39203422
26 -0.29618572
27 0.09282627
28 0.09155487
29 0.11180458
30 -0.54226143
31 0.18073182
32 0.21405783
33 0.33496337
34 -0.18749159
35 -0.19600770
36 0.01481720
37 0.35758288
38 -0.17639238
39 0.13544329
40 0.14969718
41 -0.35835212
42 0.07321165